UNIVERSAL
LIBRARY

OU_154638

UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No. 541.3452/A 37Co V.5 Accession No. 26848

Author Alexander. Jerome.

Title Colloid Chemistry. 1944

This book should be returned on or before the date last marked below.

COLLOID CHEMISTRY

THEORETICAL AND APPLIED

BY SELECTED INTERNATIONAL CONTRIBUTORS

COLLECTED AND EDITED BY

JEROME ALEXANDER

VOLUME V

THEORY AND METHODS
BIOLOGY AND MEDICINE

REINHOLD PUBLISHING CORPORATION

330 WEST FORTY-SECOND ST., NEW YORK, U. S. A.

1944

COPYRIGHT, 1944, BY
REINHOLD PUBLISHING CORPORATION

All rights reserved

This book has been manufactured in
accordance with wartime regulations
in every respect.

Printed in United States of America by
KINGSPORT PRESS, KINGSPORT, TENN.

Preface

The wide circulation and enduring usefulness of the first four volumes of this series have led the publishers to ask the Editor to plan two new volumes which would cover more recently developed theories, methods and procedures, as well as new topics or newer aspects of old topics. Following the lines of the earlier volumes, Volume V contains the papers dealing with Theory and Methods, and Biology and Medicine, while Volume VI will contain those dealing with Technological Applications.

War conditions have prevented the Editor from securing, as before, a world-wide representation of nations and nationalities, but the enormous development of American science and technology has enabled him to secure a rather complete and distinguished group of contributors, who represent a variety of national origins. Sincere thanks are due to them all, for without their effort and sacrifice, these books could not have been prepared.

The wide variety of topics considered in this series, and the different modes of approach to the same or interrelated subjects, will, it is hoped, give the user of the books a broader and more integrated view of natural phenomena than is obtainable from treatises of more limited scope. Though the number is naturally incomplete, yet enough of the fragments of the jig-saw puzzle of nature are included to give, when properly placed, some notion as to the beauty, complexity, and interrelation of the principles governing animate and inanimate things. We must not allow the diversities of scientific methods and jargons to obscure the ideas and principles they carry, nor must we allow the glib use of scientific terms to take the place of understanding. As the philosopher Thomas Hobbes truly remarked: "Words are wise men's counters—they do but reckon by them; but they are the money of fools" ("Leviathan").

These books should prove of use to scientists in a wide variety of different fields in attacking ever-recurrent problems. Not only are the data and points of view assembled in the various papers of value in themselves, but the reader is subjected to the influence of knowledge obtained in fields remote from his own; and this has a quickening influence. The all too commonly held notion that a chemical analysis or the isolation of an effective chemical substance is an end in itself vanishes when we consider the highly complicated phenomena of biology and of industry. The more we learn, the greater the problems that confront us; and it is hoped that this series will continue to be of practical value to those seeking solutions of these problems.

Thanks are due to the publishers for their continued coöperation, and to the typesetters, printers, binders and others who carried the burden of fabricating the books; to the various publishers and publications whose courtesy has permitted the reproduction of quotations, facts and figures; and also to the many individuals who gave the Editor help and advice. In each case, the Editor has tried to see that appropriate acknowledgment is made, directly in the text or through the bibliography.

New York, N. Y.
January, 1944

JEROME ALEXANDER

Contents

PAGE

PREFACE iii

THEORY AND METHODS

CHAPTER

1. SUCCESSIVE LEVELS OF MATERIAL STRUCTURE—Jerome Alexander 1
2. THE SURFACES OF SOLIDS AND LIQUIDS AND THE FILMS THAT FORM UPON THEM —William D. Harkins 12
3. SOLUTIONS OF SOAPS AND DETERGENTS AS COLLOIDAL ELECTROLYTES—James W. McBain 102
4. APPLICATIONS OF ELECTRON DIFFRACTION—Lester H. Germer 120
5. METHODS FOR THE ANALYSIS OF COMPLEX MOLECULAR STRUCTURES WITH THE AID OF X-RAYS—Maurice L. Huggins 131
6. MICRORADIOGRAPHY OF COLLOIDAL MATERIALS—George L. Clark 146
7. THE ELECTRON MICROSCOPE—Albert F. Prebus 152
8. RHEOLOGICAL PROPERTIES OF SIMPLE AND COLLOIDAL SYSTEMS—Henry Eyring and Richard E. Powell 236
9. HIGH-VACUUM DISTILLATION—K. C. D. Hickman 253
10. STATES OF AGGREGATION OF SOME PARAFFIN CHAIN COMPOUNDS—Robert D. Vold and Marjorie J. Vold 266
11. POLYMERIZATION—H. Mark 280
12. FUNDAMENTAL ASPECTS OF THE ELASTICITY OF RUBBER AND OF RUBBERLIKE MATERIALS—Eugene Guth 286
13. THE VITREOUS STATE—Maurice L. Huggins, Kuan-Han Sun and Alexander Silverman 308
14. SOME IRREVERSIBLE EFFECTS OF HIGH MECHANICAL STRESS—P. W. Bridgman . 327
15. SONIC AND ULTRASONIC WAVES IN COLLOID CHEMISTRY—Karl Sollner 337
16. THE CYCLOTRON AND SOME OF ITS USES—Jerome Alexander 373
17. THE BETATRON: A GENERATOR FOR HIGH-ENERGY ELECTRONS AND X-RAYS—D. W. Kerst 380
18. THE ELECTROPHORETIC STUDY OF PROTEINS AND RELATED SUBSTANCES—Duncan A. MacInnes and Lewis G. Longsworth 387
19. HIGH-SPEED CENTRIFUGATION—E. G. Pickels 411
20. MEASUREMENT OF THE SURFACE AREAS OF FINELY DIVIDED AND POROUS MATERIALS BY LOW-TEMPERATURE ADSORPTION ISOTHERMS—P. H. Emmett . . . 434
21. ADSORPTION AND CRYSTAL HABIT MODIFICATION—Wesley G. France 443
22. SOME APPLICATIONS OF SELECTIVE ADSORPTION AND DIFFERENTIAL DIFFUSION IN CHEMICAL ANALYSIS—Beverly L. Clarke 457

CHAPTER PAGE
23. Colloid Chemical Aspects of Photographic Development—S. E. Sheppard . 472
24. Liesegang Rings—Andrew Van Hook 513
25. Fixing the Dates of Past Events by Tree Rings and Varves—Chester A. Reeds 519

BIOLOGY AND MEDICINE

26. Proteins—W. T. Astbury 529
27. Catalysis as a Biological Factor—Jerome Alexander 545
28. Photosynthesis—Paul Rothemund 600
29. Plant Cell Membranes—Wanda K. Farr 610
30. Recent Advances in Starch Chemistry—R. M. Hixon and R. E. Rundle . . 667
31. Enzymes and the Biological Action of Vitamins—A. E. Axelrod and C. A. Elvehjem 683
32. Minerals and Vitamins in Applied Nutrition—Lela E. Booher 707
33. The Mechanism of Hormone Action—Oliver Kamm and D. K. Kitchen . . 739
34. The Molecular Organization of Visual Processes—George Wald 753
35. Chemical Pacemakers and Physiological Rhythms—Hudson Hoagland . . 762
36. The Colloid Chemistry of Purified Viruses—Max A. Lauffer and W. M. Stanley 785
37. The Gene—A Structure of Colloidal Dimensions—Jerome Alexander . . 808
38. The Gene as a Chemical Unit—Jack Schultz 819
39. Colloid Chemistry in Embryonic Development—L. G. Barth 851
40. The Physical States of Protoplasm Compatible with Life—B. J. Luyet . . 859
41. Some Physical Properties of Protoplasm—Robert Chambers 864
42. Some Surface-chemical Properties of Protoplasmic Proteins—M. J. Kopac 875
43. Physical Changes of Muscle Related to Activity—Otto Meyerhof . . . 883
44. The Capillary Circulation—E. M. Landis 900
45. Biochemical and Physiological Mechanisms in Inflammation—Valy Menkin 917
46. The Modern Outlook on Blood Coagulation—John H. Ferguson 951
47. Immunology—William C. Boyd 957
48. Allergy and Anaphylaxis—Carl A. Dragstedt 979
49. Homeostasis: The Maintenance of Steady States in the Organism—Walter B. Cannon 985
50. The Causes and Nature of Cancer—Leo Loeb 995
51. Gerontology: The How and Why of Ageing—A. J. Carlson 1051
52. The Formation of Concretions—L. Lichtwitz 1063
53. Aero-emphysema and Caisson Disease, a Problem of Colloid Chemistry—Jean Piccard 1082
54. The Action of War Gases—Chauncey D. Leake and David F. Marsh . . . 1094

CHAPTER PAGE
55. Infective Aerosols—M. W. Jennison 1099
56. Physical Chemistry of Lipides—Mona Spiegel-Adolf 1119
57. Physicochemical Mechanisms in Neuropsychiatric Disorders—E. A. Spiegel and M. Spiegel-Adolf 1136
58. Psychiatry—S. DeWitt Ludlum 1154
59. Changes in the Surrounding Medium Produced by Free-living Cells—H. S. Jennings 1162
60. The Primary Physicochemical Basis of Organic Evolution—C. R. Plunkett 1173
Appendix—Table of Nuclear Properties—Robley D. Evans 1199
Author Index 1219
Subject Index 1236

Part I

Theory and Methods

Successive Levels of Material Structure

JEROME ALEXANDER

The tangible things with which we deal in our daily lives, and even the much smaller units revealed by our microscopes, ultramicroscopes, electron microscopes and x-ray spectrometers, are complexes resulting from the aggregation of still smaller units at successively lower structural levels. William S. Gilbert put a profound truth into the song of "Little Buttercup" when he wrote: "Things are seldom what they seem."

In the early days of science, its great leaders were termed natural philosophers, for their breadth of view was sufficiently wide to enable them to understand, appraise, correlate and extend existing knowledge of natural phenomena in many fields. As the increasing armies of scientists were deployed over an ever-extending scientific terrain, they accumulated masses of detailed information, much of which was reported in localized scientific dialects, buried in *oubliette* journals, and unappreciated by the comparatively few generalists who remained at scientific headquarters. Specialists in various fields developed their own scientific journals and jargon. Botany and zoölogy were estranged. Chemistry and physics became different disciplines and each spawned a numerous progeny of subsidiary sciences. In medicine the general practitioner, who treated whatever ills his patients happened to have, was overshadowed by a host of highly trained specialists who often found patients suffering with the particular ills they treated. This often saved patients; but it might happen that an opera singer could have a hemorrhage on the stage if a throat specialist failed to recognize active lung trouble.

The reaction from over-specialization in science has been steadily proceeding, not by abolition of the special fields and their journals, but by an increasing appreciation of the close and inseparable interrelations among all scientific disciplines. Nature is oblivious of the subdivisions of knowledge created by university departments, by textbooks, or by any hiatus in personal information. A single natural phenomenon, *e.g.*, respiration, may cross the frontiers of biology, physics, organic, inorganic and colloid chemistry. Although this is well known, it all too frequently occurs that specialists whose attention is focused on conditions or happenings at some one structural level, or even limited to that level, fail to notice or fairly to appraise the relations of what they see to facts emerging at other structural levels. Because of educational hysteresis, rapidly advancing scientific discovery is generally far in advance of textbooks, and because of intellectual inertia it is all too often considerably in advance of teachers. Joint meetings and symposia, the development of borderline sciences, reviews and books of broad scope, and splendidly organized and comprehensive abstract journals, offer to all who are willing to make the effort the means of extending both the area and depth of their scientific vision.

Science and Philosophy

Knowledge of isolated facts, important though they may be, cannot be considered as *science,* although legal Latin terms *scienter* mere knowledge of facts that fix responsibility on an individual. The further we go in our efforts to correlate and colligate facts in various fields, the more closely do we approach *philosophy,* which deals with knowledge of the highest generality, *science* representing the intermediate steps of partially unified knowledge. Laboratory apparatus used to be termed "philo-

sophical instruments," when scientists were "natural philosophers." The "Philosophical Transactions of the Royal Society" and the "Transactions of the American Philosophical Society" contain papers in a wide variety of scientific fields. John Dalton, who was the first Secretary of the Manchester Literary and Philosophical Society, was the discoverer of color blindness * from which he suffered, as well as of the law of chemical combination in fixed proportions. It would be unfortunate if scientific specialists who tolerate brethren in remote fields should adopt a supercilious attitude toward philosophers, who put the hard-won facts of science to the uses for which they are intended—further unification. The motto of Phi Beta Kappa, "Philosophy is the guide or steersman of life," applies cogently to science.

In dealing with ultimate scientific ideas, Herbert Spencer [1] points out that we can form no real concept of an ultimate unit or particle. To say that matter is infinitely divisible commits us to a supposition not realizable in thought. The other alternative, that matter is *not* infinitely divisable, cannot be represented in thought; for every material unit, despite its minuteness, must be thought of as having a top and bottom and a right and left side, like any larger particle, so that a plane of section can be conceived of between them. Considering motion, if we try to follow mentally the decreasing velocity of a rising pendulum, we cannot pass, in thought, from infinitesimally slow motion to *no* motion. The infinitesimal calculus overcomes this impasse by assuming that motion proceeds in a quantum or dropwise fashion, and that there is a sudden minute jump from $v = ds$ to $v = o$.[2]

Similarly, we can have no notion as to the limits of space. We cannot imagine that space ends with the so-called Einsteinian Universe, estimated at 2×10^9 light-years, a light-year being approximately six million million miles. This estimate is based on our inability to *measure* any greater distance with the aid of light, the only measuring rod at present available. The fact that we cannot *measure* relative velocities greater than the speed of light does not mean that no such velocities exist. As Soddy pointed out,[3] if two particles (electrons) are shot out from a radioactive body in opposite directions and at speeds approximating the speed of light (say nine-tenths c), their motion relative to each other ($1.8c$) would be nearly double the velocity of light. As E. Cunningham [4] observed, "This is quite true, and the principle of relativity has nothing to say against it. The principle maintains that a velocity greater than c (the velocity of light) *relative to the observer* cannot be observed. . . . The position is not that velocities greater than c are not conceivable, but that real bodies become illusory in observation if they are conceived of to be moving faster than light."

The Relativity of All Knowledge

Spencer devotes the fourth chapter of his "First Principles" to a discussion of "the relativity of all knowledge," and was thus a philosophical precursor of Albert Einstein who demonstrated the relativity of physical measurements. Lest we be too confident of all our sensory knowledge, let it be recalled that Blakeslee and Fox [5] demonstrated that the ability of persons to taste phenylthiocarbamid is heritable as a Mendelian recessive, and that even those who get any taste at all from it (about 70 per cent) describe it variously as bitter, sweet, salty, or sour. This indicates that there is a relativity of sense impressions. H. C. Moir [6] tested sixty persons as to their ability to recognize by taste four simple flavors—orange, lemon, lime, and vanilla. Only one person had a perfect score. Five had records of over 75 per cent, but forty-eight failed to reach 50 per cent. Vanilla was variously identified as black currant, lime, apricot, greengage, damson, lemon, pineapple, orange, tangerine, almond, red currant, strawberry.† Only a limited number of persons can taste sodium

* Also known as Daltonism.

† Chemists handling products sold to the public must be sure that their senses for color, taste, odor, and touch are "normal"; that is, that they correspond to those of the average person.

benzoate, and wide differences exist in the ability to detect and recognize such odors as verbena and to distinguish between wines.[7] R. J. Williams reports (*Science*, Dec. 11th, 1931) that a man whose sense of smell appeared otherwise normal could not detect the odor of a skunk, while *n*-butyl mercaptan, the "perfume" carried by skunks, had no unpleasant odor for him. Laselle and Williams[7a] in attempting to identify a substance as creatinine, found it tasteless, though the literature states that creatinine is bitter. It was not until they had tried the sample on several others that they located someone who found it bitter. Since lean meat contains much creatinine (about 2 grams per pound), and soups made from lean meat contain extracted creatinine, we have another possible basis of differences in taste. Williams believes that the problem is associated with the more general one of individual metabolic idiosyncrasies, which crop up at times in medicine (*e.g.*, reactions to morphine, novocaine, iodoform) or in industry (reactions to cosmetics, "chemicals," etc.).

On the other hand the sense of smell of many animals is highly developed. Dogs follow a scent; deer, elk, and most wild animals can be approached only from the lee side. When offered a strange kitten, or sometimes even one of her own that has acquired a "foreign" odor, a mother-cat, after carefully sniffing it, will "spit" and box the kitten's ears to drive it off. Bees and ants are largely guided by odor, as are most insects in seeking food or mates. A Swiss scientist who had a female moth caged in his city laboratory far removed from parks or woodland, was astonished to see large numbers of males of the same species fly in. It has recently been shown[8] that bats in flight emit sounds too shrill to be heard by the human ear but audible by bats, which even in pitch darkness are enabled to avoid objects (*e.g.*, the walls of caves) by sensing the reflected supersonic waves—an aural radar, so to speak.

Granting that we have come to an agreement as to the result of an observation, we have not reached an end but merely a new position from which to begin another advance. As Tennyson phrased it:

> Yet all experience is an arch wherethro'
> Gleams that untravell'd world, whose margin fades
> For ever and for ever as I move.

Why is it that persons differ in their ability to taste phenylthiocarbamid? By what mechanism do extremely minute amounts of odoriferous substances enable cats and dogs, insects and wild animals, to draw such swift and certain conclusions? Is the so-called "sixth sense," commonly found in the blind and sometimes in the seeing, based on an unusually well developed sense of hearing, coupled perhaps with an extremely sensitive epicritic sense? Mice and bats ("flying mice") hear what to us is supersonic, and fish are believed to "feel" the presence of rocks by virtue of eddies or reflected waves of compression, not quite the same as "sound waves" in the water.

But it is not always easy to secure an agreement as to the "facts," especially if there is no obvious way of explaining a phenomenon on the basis of proven existing knowledge. A case in point is that of dowsers, who practice the art of locating underground streams of water. One branch of the divining-rod (a forked twig, usually of hazel) is held in the right fist with the clenched fingers up, the other branch being held in like manner in the left fist, with the forked center of the rod vertical. The dowser slowly walks across the area under investigation, and the fork is supposed to turn sharply downward when he passes over running water. Watching a successful dowser at work, I observed, as have others, that the downward turning of the fork is consequent upon a slight inward rotation of the clenched fists; and I found that this effect can be repeated voluntarily. When I traversed the same area as the dowser, the rod dipped in the same places as with him, and for the same reason—my wrists spontaneously turned inward. As a check, I had about ten persons, unacquainted with the dowser's findings, make the same experiment. Most of them got the same response as the dowser, with the same inward twist of the wrist.

On the other hand, dowsers tested in a laboratory with concealed pipes through which water could be made to flow at will, failed to locate flowing water; and many writers of repute consider the effect a conscious one, the dowser being a "good guesser," who depends upon observing types of vegetation, the lay of the land, etc. I am tempted to refer to this rather dangerous mooted question first, because it shows to what a remote level of material structure one may be forced to go in *an attempt* to explain a phenomenon, but especially because so eminent a scientist as Sir J. J. Thomson, Master of Trinity College, Cambridge, in discussing water-dowsing states:[9]

"There are some who can detect water without using the rod, since they experience peculiar sensations when they are in places where the rod would move. The rod may be a convenient indicator of a physiological effect. There is no doubt of the reality of the dowsing effect. In fact, in many agricultural districts the dowser is the man they call in when they want to find the right place to dig a well, and he very often succeeds. We had an example of this at Trinity College. The water supply to one of our farms was very defective and a new well was badly wanted. At first the Senior Bursar, who was a Fellow of the Royal Society, proceeded in the orthodox way and employed eminent geologists to report on where we ought to sink a well. Their advice, however, did not lead to the discovery of any water. Our land agent said, 'If I were you, I would try old X, who has found a good many wells in this county and who will sink the well on the terms 'no water, no pay.' As there seemed nothing else to be done, the Bursar employed him and he found water. For this we were assailed in an article in Nature, which lamented that Trinity College —the College of Newton—should have given countenance to such superstitious and unscientific practices.

"Although I think that most of the people who have paid any attention to the subject believe in the reality of dowsing, there is no agreement about its cause. . . . In some cases the movement of the rod occurs when the dowser has been blindfolded, or when he has not been near the part of the country before."

Thomson points out that all of the many English well waters he had examined contained radioactive emanation from radium which has a life of only four days, so that stagnant water would soon lose its radioactivity, and if this were connected with the effect on the dowser, it would explain why he can detect only *running* water.

Almost a century ago Quincke observed[10] that when water is forced by pressure through a porous membrane or a single capillary, a potential difference is set up between the opposite sides of the membrane or the opposite ends of the capillary. This is known as "streaming potential." But I know of no evidence connecting the reactions of a dowser with any kind of electrical or other radiation, although something might be shown by refined apparatus similar to that used by Sir Edgar D. Adrian and his collaborators to measure the tiny and evanescent electrical impulses in single nerve fibrils,[11] and Loomis and Harvey[12] to record the Berger and other electrical rhythms of the brain. Thomson concludes:

"The divining-rod is perhaps, of all phenomena which may be thought to be psychical, the one most favorable for experiment. The motion of the rod is a mechanical effect and gives an indication of the magnitude of the phenomenon. The conditions under which the effect occurs can be made definite, and there is no lack of trustworthy people who possess the dowsing power: for these reasons I think such experiments are well worth making." *

* "Since the above was written I have seen a paper by Mr. H. M. Budgett, published by the British Society of Dowsers, on the connection between effects on a divining-rod and the readings of an instrument designed to detect very penetrating radiation. There seemed to be indications of some connection, but these were very faint. It would be interesting to see if the dowser could, by his divining, detect a strong beam of penetrating γ-rays." (Thomson). See also "Natural Potentials in Sedimentary Rocks" by P. A. Dickey, Tech. Pub. 1625, Am. Inst. Min. and Met. Eng., (1943) J. A.

Most people do not think of "static" unless it affects radio reception; yet many fires and explosions * have been due to sparks developed upon the release of static charges arising from the friction of belts on pulleys, or of rubber tires on roads. Gasoline trucks have chains dangling in electrical contact with the earth, and at toll gates there are erect and grounded wires to carry off static charges which otherwise might give a sharp shock to both chauffeur and toll-taker. Modern printing presses have combs and other devices to "bleed off" frictional charges which would make the paper sheets unmanageable.

It might seem a far cry from sun spots to textile mill troubles; but a direct connection is claimed to be consequent upon electromagnetic disturbances set up by the spots in the earth's atmosphere,[12a] for these accentuate conditions which usually do not seriously disturb textile operations. One mill had to close down some of its processes at times during a sun spot period, because excess static would not allow the fibers to draft properly. Trouble was widespread, and both machinery and fiber quality were blamed, although no change had been made in either. Static seems to be an important "imponderable," and it is not unlikely that atmospheric conditions suitable for textile operations depend largely upon the presence of sufficient ionization and moisture in the air to prevent troublesome accumulations of static. Under suitable atmospheric conditions combing the hair makes it "fly" or "bristle."

Tiny quantities of electricity may initiate extensive effects. Thus in the course of an informal discussion, Dr. Willis R. Whitney estimated that the energy utilized by a fly in crawling up one inch against the force of gravity (one "fly-inch" power), could, if released in Los Angeles, activate a radio aerial in New York for over thirty years. And even though Hamlet correctly observed that there are more things in heaven and earth than are dreamed of in our philosophy, when we try to explain a phenomenon we must not overlook principles of common knowledge, no matter to what structural level we may have to go.

The Application of Mathematics to Physical and Chemical Problems

The data obtained by experiment and observation are often subjected to what is termed rigid or rigorous mathematical analysis in order to develop the nature of the relations between individual observations and to deduce logical corollaries therefrom. A consideration of some of the principles underlying this highly valuable application of mathematics reveals certain pitfalls yawning for the unwary scientist who steps over into mathematics, as well as for the mathematician who deals with scientific data.

Pure mathematics deals with real and imaginary quantities and their relations, and is based on abstractions drawn from practical experience. Thus we know that two plus two make four, because if we have four apples and eat or give away two, we have two left. Certain primitive savages cannot count beyond one, so that when they trade sheep for knives, they place one knife with each sheep. The decimal system seems to have had its prehistoric origin in the fact that tally can be kept by the fingers. Eskimos describe *fifteen* as "two hands and one foot," and in keeping count it is usual to make four vertical strokes with a cross-stroke "tally" for five.

Pure mathematics has undergone great development, but has maintained its basic principle of dealing logically with abstractions which may or may not correspond to physical realities. Consider $\sqrt{-1}$, often used in electrical calculations. Untrammelled by material facts, a mathematician may romance like a novelist, providing only that he adheres to the logical use of assumed premises. And though life is often stranger than fiction, the mathematician Sylvester has said that a mathematician is never so happy as when he does not know what he is talking about.

* A press dispatch dated Nov. 10th, 1942, stated: "An explosion of anesthetic gas, reportedly caused by a static spark, killed a patient to-day as he was receiving a blood transfusion after a chest operation. . . ."

Applied mathematics, on the other hand, concerns itself with physical units and their behavior, and utilizes the logical, impersonal theorems of pure mathematics to express and develop qualitative and quantitative relations between these units and their behavior. Despite the great differences in philosophical and metaphysical theories and beliefs, and the admission that we cannot even conceive what are the ultimate units of matter, every person in his daily life shows by his actions that he believes matter to be *real*.[13] The units of the applied mathematician may be electrons flying about atomic nuclei, or the levers of a machine, or the planets of the solar system; but whatever they are, the physicist and chemist are compelled to discover, by experiment and observation, certain facts and relations regarding them. It is usual to combine and reason mathematically with these findings, after expressing them in a condensed mathematical form, or *formula*. The word "formula" is a diminutive of the Latin *forma* (form), and therefore means simply "a little or concise form." The pure mathematician may assume facts to form a convenient equation, subject to mathematical manipulation. The applied mathematician, who is often a physicist or a chemist, must find equations to fit the facts, always remembering that equations are mute and make no protests against sins of commission or omission. St. Paul warned that we cannot be all things to all men; but a truly valid equation must be all things to all pertinent facts, known and unknown.

The history of science is full of instances where accepted notions have been discarded or modified because of new knowledge. Even in the days when diet was based on "fat, carbohydrate, protein, and total calories," any grandmother knew that cod-liver oil had a potency far greater than olive oil, its quondam theoretical dietary equivalent. The discovery of the importance of trace substances in the diet, including vitamins, iodine, copper, etc., and recognition of the fact that most proteins do not contain all of the amino acids essential to well-being, have led to a marked revision in feeding both humans and animals. Long before any scientific explanation was available for the use of protective colloids in modifying cows' milk for infant feeding, practical physicians had recommended and utilized the addition of gelatin, gum arabic, dextrinized starch and the like, and in folk practice such colloids as Irish moss, Iceland moss, and Bavarian beer (of high dextrin content) were used.[14]

Certain elements exist in allotropic forms that owe their differences to the way in which their constituent atoms are assembled. This raises the question as to the minimum number of atoms which must get together in a certain configuration to make the smallest conceivable particulate unit of a certain allotrope. Thus a single carbon atom can not be graphite or diamond. Theoretically, the smallest unit we could call graphite must have at least twelve carbon atoms, arranged in two hexagonal layers 3.40 Å apart, while in diamond we can imagine a minimum unit of five carbon atoms, four being arranged in the form of a regular tetrahedron whose side is 1.54 Å, with the fifth carbon atom centered within. Particles handled experimentally are usually very much larger than these suppositious minimum units, but rings of six carbon atoms have been demonstrated in the mass spectrograph. By electron diffraction methods Germer and White estimate that the smallest crystals formed by vaporizing copper on supporting foils of iron or of Invar gave ring intensities between those calculated for crystals of 55 and of 397 atoms each.[15]

In J. W. Mellor's monumental book,* Vol. V, p. 720 (1924), it is stated: "Boart and carbonado are usually regarded as forms intermediate between diamond and graphite." The following descriptions are condensed from the catalog of a large

* Mellor quotes Lucretius (60 B.C.): "It matters much with what others, and in what positions, the same atoms are held together. . . . When the configuration of the atoms is changed, the properties of the body which is formed from them must also change." H. A. Miers (article on "Diamond" in "Encyclopedia Britannica," 11th ed., 1910) stated: "Both bort and carbonado seem to be really aggregates of crystallized diamond, but the carbonado is so nearly structureless that it was till recently regarded as an amorphous modification of carbon."

dealer in industrial diamonds: *Boarts* (borts) are more or less transparent crystals of most varied shapes, sometimes twin and multiple formations. *Carbons* (carbonado), often called amorphous diamonds, differ essentially from boarts in structure; they are a porous cluster of minute diamond crystals, fine to close-grained. The fracture of a good carbon should resemble the structure of fine steel. Much harder than boarts, and much less fragile, they can be used until entirely consumed. *Ballas* * are roundish aggregations of innumerable minute crystals grouped concentrically about a nucleus. Being non-porous and devoid of cleavage planes, they are about as hard as good carbons, but tougher because of their structure. Brazilian

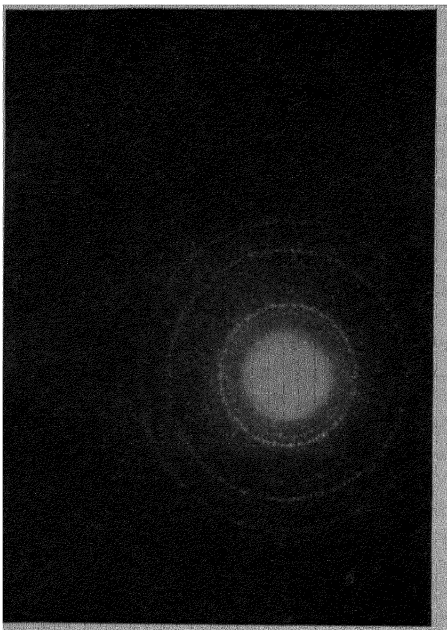


FIGURE 1. X-ray spectrogram of poor quality carbonado (unfiltered Mo radiation; distance 5 cm) Courtesy G. L. Clark.

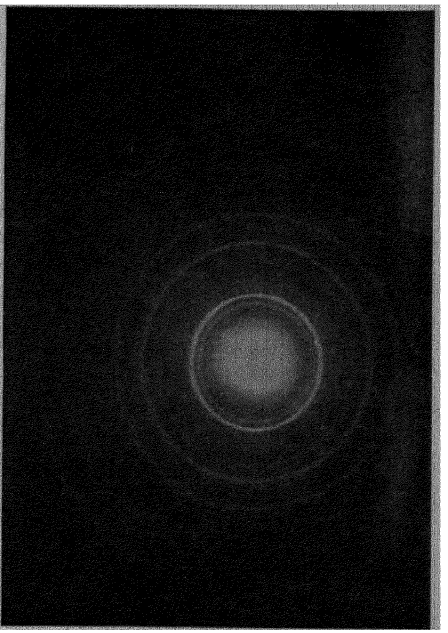


FIGURE 2. X-ray spectrogram of high quality carbonado (unfiltered Mo radiation; distance 5 cm) Courtesy G. L. Clark.

ballas is used to test the hardness of carbons. When the two are rubbed together, a white mark indicates that the carbon is harder, while a dark brown mark indicates that the ballas is harder. If both stones are of equal hardness, *no* mark results on either.

Early in 1928 I secured specimens of ballas, and of a fine and a poor grade of carbonado, and Professor George L. Clark kindly took x-ray spectrograms of them all. The results indicate that the ballas was composed of a great number of small diamond crystals, the poorer carbonado of still smaller and more numerous crystals, while the fine carbonado had still finer crystals, many being of colloidal dimensions so that the rings of its x-ray spectrogram were continuous. No other form of aggregation than that of diamond was apparent [16] (See Figs. 1 and 2).

* The word "ballas" appears to be a modified form of the Portuguese word *balas*, meaning *bullets*. The Spanish and Portuguese word *bola* means *ball*, whence is derived the Argentine *bolas*, the throwing-rope having ends weighted with heavy balls. . . . Through the Latin *ballista*, the word goes back to the Greek βαλλειν, meaning to *throw*. *Palla* is an old German form, and the golden balls of the house of Medici, now used by pawnbrokers, were known as *pallé*, which was the rallying cry of the Medici retainers.

Currently Recognized Structural Units and Levels

It is not so long ago that students were taught that atoms represent the lower limit of material structure: curious sphere-like bodies, all alike for each element, having potent but inscrutable idiosyncrasies or "affinities." Prout's hypothesis (1815) that all elements are built up from some single basic unit ("protyle" or hydrogen) was considered as having been completely disproved by the meticulous atomic weight determinations of Stas. We now know that only in relatively few of the elements (*e.g.,* He, Na, P, V, Mn, Co, As, I, Au, Bi, among the commoner ones) are the atoms all alike, for most elements have a variety of *isotopes*—atoms closely alike in chemical behavior but different in mass.* Thus chlorine is a mixture of about 76 per cent of an isotope of mass 35, with 24 per cent of an isotope of mass 37; tin has ten isotopes varying in mass from 112 to 124, and lead seven with masses from 203 to 210. The careful atomic weight determinations of T. W. Richards and others have shown differences in the atomic weight of lead from several sources. Furthermore, though the structure of atoms is not as simple as Prout had assumed, the basic work of Lord Rutherford and many others has shown that all the diverse elemental atoms are built up of a few units which we are at present unable to resolve further.

Starting with these basic units as our present null point, there is given below a triplex table which aims to illustrate the currently recognized structural levels or orders of material complexity commonly found in nature. On the left appear the basic material units and unit structures generally recognized by physicists and chemists. In the middle, at corresponding levels, appear the units recognized by biologists, geneticists, and medical men; and on the right, units demonstrable in cotton textiles, in starch, and in steel. As we rise to the larger and larger aggregates, the definiteness and strength of the forces holding together their proximate sub-units diminishes, and the boundaries between structure levels become less definite.[17]

The table demonstrates *the great importance of structure,* for from comparatively simple chemical elements and molecules, diverse and complicated structures develop, whose properties are greatly influenced by the manner in which their sub-units are assembled and held together.[18] It is impossible to do more than to suggest in the table the importance of secondary substances which, when present in sizable amounts or even in traces or as "impurities," may exercise an influence upon structure and function quite unexpected and out of proportion to the amount present.

Mixtures vs. Pure Substances

Mixtures often show curious properties. Thus "half and half" solder melts at 220° C, which is far below the melting points of its constituents, tin (m.p. 232° C) and lead (m.p. 327° C). The influence of colloids upon crystallization, long ago recognized,[19] is of great biological and technical importance. The quantity needed to produce a marked effect may be very small. Thus .01 per cent gelatin greatly retards the setting of plaster of Paris and prevents the formation of the ordinary visible crystals. Minute quantities of protective colloids may cause precipitates to fall out in the colloidal state, and small amounts of "addition compounds" greatly affect the electro-deposition of metals.[20]

H. G. Bungenberg de Jong and H. R. Kruyt[82] found that when the dispersed substance in hydrophilic sols is thrown out by a variety of methods, *e.g.,* salts in large or in small amounts, removal of solubilizing salts, temperature change, addition of an oppositely charged sol, the "precipitate" often forms viscous droplets which aggregate into a fluid mass called a *coacervate,* instead of forming a solid phase. The phenomenon, termed *coacervation,* had been observed by F. W. Tiebackx,[88] who commented on the resemblance to casein of the gelatin-gum arabic coagula. Strongly adsorbed shells of water are supposed to surround the droplets and act as a protective

* See Appendix in this volume for table of natural and artificial isotopes of the elements.

Successive Levels of Material Structure

Elementary Particles (at present (1943) recognized)

Order of Complexity 0 (?)	Charge	Size (Å)	Mass (gram)	Discoverer	Where Found
Electron	−e	3×10^{-5}	9×10^{-28}	Sir J. J. Thomson (G. Johnstone Stoney) (Sir Wm. Crookes)	Electric current Negative electrostatic charges Nuclear satellites and emissions
Positron	+e	3×10^{-5}	9×10^{-28}	C. D. Anderson	Emitted by atomic nuclei
Neutrino	0	2×10^{-5}		E. Fermi, W. Pauli	Existence inferred from general theory
Proton	+e	2×10^{-8}	1.66×10^{-24}	Sir Ernest Rutherford	Nucleus of protium (H)
Neutron	0	2×10^{-5}		James Chadwick	Atomic nuclei
Positive Mesotron	+e	2×10^{-5}	2×10^{-25}	C. D. Anderson and Seth Neddermeyer	Cosmic radiation
Negative Mesotron	−e	2×10^{-5}	2×10^{-25}	C. D. Anderson and Seth Neddermeyer	Cosmic radiation

Material Structures	Order of Complexity	Mode of Examination	Approx. Size	Chemical Units	Biological Units	Biology and Medicine: Cotton	Technology: Starch	Technology: Carbon Steel
Atomic Nuclei	1	Transmutation Cyclotron	H = 2×10^{-5} Å Au = 1×10^{-4} Å					
Atoms	2	Cyclotron X-Ray Chemical	5 Å ±	C, H, O, N, P, S, Fe, K, Na, Ca, Mg, Cl, Fl, B, I, Zn, Cu, Cr, Ni, V		C, H, O	C, H, O	Fe, C (0.1—1.0%) (S, P, Si, Mn, etc.) 0.5% C = ca 7.5% FeC_4
Molecules	3	Infrared and Raman Spectra, X-Ray, Chemical	50 Å ±	Amino acids, Lipoids, Nucleotides, Glucose, etc.	Vitamins, Hormones, Thiamin Pyrophosphate, Carriers† etc.	Glucose, etc.	Glucose	Fe, C, FeC_4
Macromolecules* Molecular Aggregates	4	Electron Microscope Ultramicroscope Chemical	100 mμ ±	Proteins, Glycogen, Cellulose, Starch, etc.	Enzymes, Chlorophyll, Cytochrome, Bacteriophages, Genes, Viruses, etc.	Molecular Chains and Groups	Molecular Chains and Groups	α-iron (Ferrite) γ-iron Cementite
Micells	5	Electron Microscope Ultramicroscope	1 μ ±		Cytoplasmic structures	Aggregates with impurities		
Microscopically Resolvable Units	6	Chemical Microscope	¼ μ ±		Chromosomes, Nuclei Cells, Bacteria	Fibrils	Starch grains	Austenite (α-iron) Martensite (γ-iron) Troostite Sorbite Pearlite
Visually Resolvable Units	7	Eye	50 μ ±		Tissues, Organs, Drosophila, Mouse, Whale	Fibers, Yarns, Fabrics, Clothing		Hair-spring, Saw, Engine, Bridge, Skyscraper

* Primary Colloidal Particles ..... 30 mμ ±
Secondary Colloidal Particles ..... 100 mμ ±

† Moleculobiont, hypothetical simplest living unit.

film along the lines suggested by Zsigmondy for the protective action of water.* In the case of many coacervates it seems likely that some impurity, or some fraction of one of the ingredients, may separate at the droplet interface and serve as a colloidal protector. It is well known that very pure materials, *e.g.*, iron, vanadium, may have properties quite different from even ordinary C.P. substances.

The effect of a product in a mixture varies with the specific natures of the substances mixed, or formed on mixing. Thus practical cooks use butter, hydrogenated oil, lard, etc., to make cakes and cookies "tender," and similar "shortening" materials are added to make pie crust "flaky." Such additions weaken the cohesiveness of the final product. On the other hand, technologists in various fields use small amounts of colloidal material to make the final product stronger; *e.g.*, rosin size in paper; starch size in warp yarns; thin layers of adhesives in building up cardboard and wood veneers. It often happens that thin layers of secondary materials become adsorbed at micellar surfaces, and act as *cohesive colloids,* binding the micellar or other small units into coherent structures.[21] This may be the case with starches, where the binder, "amylopectin," appears to have the same ultimate chemical analysis as "amylose." About twenty years ago Sponsler [22] showed that grinding destroys the x-ray pattern of starch which is due to some symmetry of structure in the granules; and recently Lampitt, Fuller and Goldenberg [23] ground starch for as long as 7,000 hours in a ball mill and concluded that grinding breaks the weaker lateral linkages without any significant breakdown in the "repeating units," identified by Brawn, Hirst and Young [24] as consisting of from 24 to 30 anhydro-β-glucose units. Glycogen ("animal starch") appears in similar "chains" composed of either 12 or 18 glucose units.[25] Farr [26] believes that ellipsoidal units of cellulose are cemented into fibrils by a cohesive substance which contains some pectin. Recently, Zweifach [27] stated that cells in the capillary wall are bound together by a calcium protein compound, and that diffusion takes place through this intracellular cement rather than through the cells themselves.

The Importance of Impurities

Although the importance of "impurities" has long been stressed, the mistaken popular notion still persists that an ordinary "chemical analysis" (usually presented in the form of atomic percentages) will tell all one needs to know about a manufactured or a natural product. The consulting chemist usually has to bring home the importance and delicacy of structure by pointing out, for example, that although dropping a watch on the floor may not change its chemical composition, the shock may ruin its value as a timepiece. Scientifically, recognition of the importance of structure has been a matter of gradual growth, mainly because scientists are prone to magnify and reason with the factor they can understand and explain, and fail to suspect that "rigorously exact" or "strictly quantitative" measurements may be vitiated, as von Weimarn emphasized,[28] by variations in structure and by the existence of unknown factors.

In plant and animal nutrition, nowadays, we hear much of secondary substances (*e.g.*, "roughage"), of impurities (*e.g.*, the average hen's egg contains about a milligram of zinc), and of trace substances essential to growth and life (*e.g.*, vitamins, copper, iodine, boron, etc.). Dr. R. A. Steinberg (U. S. Dept. of Agriculture, 1939) reported that a common mould needs traces of gallium to grow and reproduce—10 to 30 parts per billion of water are necessary, and no other element appears to serve as a substitute. In cases where such minute quantities are so potent, the presumption arises that the trace element enters into the formation of a catalyst.[29] Robinson, Whetstone and Scribner [30] investigated the abnormal behavior of alumina from leaves of a hickory tree growing in a Virginia pegamite vein, found that at least 0.2 per cent of the dry weight of the leaves consisted of rare-earth oxides,

* "Colloids and the Ultramicroscope," Chapter 18, translation by J. Alexander, New York, 1909.

mainly of cerium, lanthanum, praseodymium, neodymium, yttrium, samarium, and europium. Sir F. G. Banting[81] in speaking of the purification of insulin, said: "With the elimination of impurities the insulin was more rapidly absorbed and the duration of its effect was lessened. . . . Hagedorn (of Denmark) and his colleagues, 1933-35, found that the addition of protamine to their insulin so delayed the action that the day's supply of insulin could be given in one injection. This was the greatest advance in the treatment of diabetes since the discovery of insulin. Scott (of Toronto) found that when protamine was added to zinc-free amorphorus insulin, there was little or no delay in the rate of absorption. If, however, zinc was added, a combination occurred between the insulin and protamine with the characteristic slowing effect of the protamine. It would seem that the Danish insulin contained sufficient zinc or other specific metal to produce this combination." The effects of zinc-insulin (without protamine) seem to be intermediate between those of pure insulin and zinc-protamine-insulin.

For many years it had been taught as an elementary distinction that non-living things grow by *accretion,* whereas living things grow by *intussusception.* For example, Thomas Huxley stated in his "Anatomy of Invertebrates" [p. 10 (1888)]: "The increase in size, which constitutes growth, is the result of a process of molecular intussusception, and therefore differs altogether from the process of growth by accretion, which may be observed in crystals." But all growth is ultimately accretional, as was pointed out by J. Alexander and C. B. Bridges in Vol. II of this series, pp. 17-18. For on considering the table of material units, it at once becomes evident that intussusception in any unit of a higher order of complexity may be resolved into accretion at the surface of particles at some lower level; and conversely, accretion at the surface of particles grouped to form a unit of higher complexity results in the growth of this mass by intussusception.

Much of the "mystery of life" arose from inability to descend to levels of structure where basic life processes could be coupled up with chemical knowledge. Thus Goethe, himself a biologist, put the following words into the mouth of Mephistopheles (*Faust,* Part I, line 1581 ff):

> Who would know and describe a living thing,
> First kills it, the spirit severing;
> The parts he holds, without a kink,
> But lacks, alas, the spiritual link.
> *Encheiresin naturae* says Chemistry now,
> Mocking itself, for it knows not how.

And Alexander Pope wrote:

> Like following life through creatures you dissect,
> You lose it in the moment you detect.

Several papers in this volume show how clear a notion of physical life may be had at the catalytic level, in the colloidal zone.

In attempting to envisage the mechanism underlying a phenomenon, it is necessary to focus attention initially upon the structural level where the crucial factors are operative. Thus with radioactivity we consider the nuclear level; with chemical problems, the atomic and molecular levels. Structures of the most diverse chemical nature (iron, wood) may function as chairs, though tests may be devised to bring out the chemical differences. As a rule changes at lower levels exert potent influences at higher levels, and vice versa. Apparently conflicting observations may often be resolved when structural interrelations are properly appreciated and appraised.

References

1. Spencer, H., "First Principles," Chapter 3.
2. Alexander, J., *Scripta Mathematica,* **4,** 139-45 (1936).

3. Soddy, F., *Nature*, Nov., 1913.
4. Cunningham, E., "Relativity and the Electron Theory," p. 40 (1915).
5. Blakeslee, A. S., and Fox, A. L., *Proc. Nat. Acad. Sci.*, **18**, 115-120 (1932).
6. Moir, H. C., *Chemistry & Industry*, **55**, 146 (1936).
7. Miall, S., *Ibid.*, **51**, 769 (1932).
7a. Williams, R. J., and Laselle, P. A., *J. Am. Chem. Soc.*, **48**, 536 (1926).
8. Galambos, R., *Science*, **93**, 215 (1941); Griffin, D. R., and Galambos, R., *J. Expt. Zool.*, **86**, 481-506 (1941).
9. Thomson, Sir J. J., "Recollections and Reflections," (1937), pp. 158-63; see also "The Physics of the Divining Rod," C. J. Maby and T. B. Franklin (1937).
10. Quincke, G., *Pogg. Ann. der Phys.*, **107**, 1 (1859); **110**, 38 (1860).
11. Matthews, B. H. C., "Electricity in Our Bodies," London, 1931.
12. Loomis, A., Harvey, E. N., and Hobart, G., *Science*, **82**, 198 (1935); **83**, 239 (1936).
12a. Strang, P. M., "Textile Research," 1934; *Ind. Eng. Chem., News Ed.*, **19**, 1083 (1941).
13. There is a story that Bishop Berkeley saw one day the following limerick pinned on a tree in the University quadrangle:

A young Oxford man remarked: "God
Must think it exceedingly odd
That this sycamore tree
Continues to be
When there's no one about in the quad."

Next morning the philosopher's reply was found pinned below:

Dear Sir: Your bewilderment's odd:
I'm always about in the quad;
So this sycamore tree
Will continue to be
Since observed by, Yours faithfully, God.

Slightly different versions are given in "The Oxford Book of Light Verse" and in "The Peter Pauper Book of Limericks."
14. Jacobi, A., "Intestinal Diseases of Infancy and Early Childhood," 1889; Alexander, J., and Bullowa, J. G. M., *Arch. Pediatrics*, 1910; *J. Am. Medical Assn.*, 1910; also Vol. I, this series.
15. White, A. H., and Germer, L. H., *Phys. Rev.*, **9**, 492-7 (1941).
16. Alexander, J., *Science*, **73**, 185 (1931).
17. Wells, P. V., *J. Wash. Acad. Sci.*, **6**, 361 (1919); also Vol. I, this series.
18. Alexander, J., "Colloid Chemistry," 4th ed., Van Nostrand Co., 1937, discussion on "Residual Affinities and Molecular Assemblage."
19. Ord, W. M., "The Influence of Colloids on Crystalline Form and Cohesion," London, 1879; Alexander, J., *Kolloid Zeit.*, 1909; 1910.
20. Blum, W., Vol. IV of this series.
21. Alexander, J., *J. Soc. Chem. Ind.*, **55**, 206-9 (1936).
22. Sponsler, O. L., *Am. J. Botany*, **9**, 471 (1922).
23. Lampitt, L. H., Fuller, H. C., and Goldenberg, N., *J. Soc. Chem. Ind*, **60**, 1-6; 25-29; 69-72; 99-111 (1941).
24. Bawn, C. E. H., Hirst, E. L., and Young, G. T., *Trans Farad. Soc.*, **36**, 880 (1940).
25. Bell, D. J., *Biochem. J.*, **29**, 2031 (1935); **30**, 1612, 2144 (1936); **31**, 1683 (1937).
26. Farr, W. K., paper in this volume, and its references.
27. Zweifach, B. W., *Cold Spring Harbor Symposium*, **8**, 216 (1940).
28. Weimarn, P. P. von, *J. Chem. Education*, **3**, 373-81 (1928).
29. Alexander, J., this volume, "Catalysis as a Biochemical Factor."
30. Robinson, W. O., Whetstone, R., and Scribner, B. F., *Science*, **87**, 470 (1938).
31. Banting, Sir F. G., *Science*, **85**, 594-6 (1937).
32. Bungenberg de Jong and Kruyt, H. R. *Kolloid Zeit.*, **50**, 39 (1929).
33. Tiebacx, F. W., *Kolloid Zeit.*, **8**, 198, 238 (1911).

The Surfaces of Solids and Liquids and the Films That Form upon Them

PART I. LIQUIDS

WILLIAM D. HARKINS

University of Chicago and the Universal Oil Products Company

Introduction: Surfaces and Interfaces

The surface of a solid or a liquid, or the interface between any two three-dimensional phases, is a region of discontinuity which is extremely thin, except at temperatures very close to the critical point. The discontinuity in the density of the material is a manifestation of the extremely rapid variation in the intensity and shape of the electric fields with the distance from the surface. If this is plane there is a dissymmetry of the field, with reference to the plane, which results in more or

less orientation of the molecules in the surface region. In the special case in which long, straight-chain organic molecules are tightly packed in an interface, they are in general oriented either perpendicular to or parallel to the surface.

The perpendicular orientation is discussed in Volume I of this series, and the parallel orientation in a later section. The idea of molecular orientation in surfaces

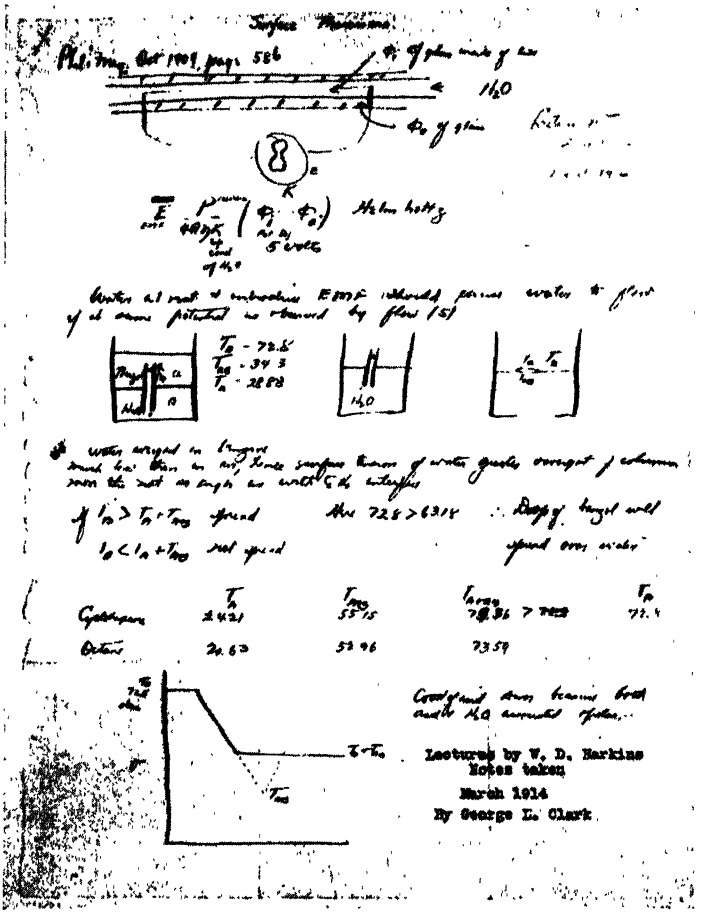


FIGURE 1. Earliest record of the direction of orientation of molecules at the surface of a liquid.

was developed independently by Sir William Hardy and by the writer, but the earliest series of lectures on this subject was presented by the writer at the University of Chicago in the winter of 1913-14. Since, however, Hardy did not give the direction of orientation of molecules in the surface, but merely indicated that the surface is a region of discontinuity, Fig. 1, a record of one of these lectures, is of considerable interest, since it indicates that polar groups orient themselves toward the water. This figure also represents a very early application of the idea of the spreading coefficient to show that benzene must spread on a clean water surface. It will be shown later that the denial of this fact has been made the basis of one of the most prominent theories of the spreading of films on water.

FILMS

Formation of Films

Films may form upon the surface of a liquid or a solid either by condensation from a vapor (adsorption) or by the spreading of material from a liquid or a solid body, which floats upon, or is in contact with, the surface. Films may also form by adsorption from a solution or by spreading at an interface, as, for example, between oil and water, between mercury and another liquid, or between a solid and a liquid. Thus either oleic or stearic acid spreads readily at an oil-water interface.

The type of film which spreads is determined by the interfacial free energy relations involved, as is shown later. Where the free surface energy would be decreased if a film of a certain type should spread, then it may spread; but if an increase of free energy is involved, it will not spread.

Three general types of films need to be considered:

(1) Monolayers: one molecule thick.

(2) Films thicker than monolayers, but thinner than duplex films.

(3) Duplex films: the film is sufficiently thick to give complete independence between the energy of the surface and that of the interface.

The Spreading of Oil Films on Water [21, 35, 81, 83] *

Oxidized Oils. If a drop of fresh Nujol, a heavy hydrocarbon oil, is put on a clean surface of water it does not spread, but remains as a thick lens. If the Nujol has been oxidized by air by stirring it in an evaporating dish over a flame, it is found to spread as a uniform polymolecular film over a considerable area. If, for example, the area of the surface of the water in a rectangular trough is 5×10^3 cm.^{2} a small drop of this oil spreads over the surface to form an invisible film, which, on moderate decrease of area on compression by the use of a movable barrier, exhibits a uniform light yellow color in white light. On compression the color changes to dark yellow, gold, red, purple, blue, and green, and these colors then repeat themselves as the spectrum moves through four or more orders, but in the higher orders the colors red and green become much the most prominent. On further thickening, the film becomes colorless. An experiment such as this involves films of a uniform optical thickness between a few hundred and several thousand Angström units. The limits of thickness are much greater than this.

Such polymolecular liquid films may be designated as duplex, since their upper and lower surfaces are far enough apart to make their surface energies entirely independent of each other. Thus the total free energy, γ_d of unit area of the duplex film of a liquid b on a liquid a is given by the expression:

$$\gamma_d = \gamma_b + \gamma_{ab} \tag{1}$$

Such a duplex film is, in general, unstable. Thus an oil film on water which exhibits a single diffraction color is always found to segregate into regions of varied thickness, and therefore of varied colors, at the end of a period which varies from a few seconds to a few weeks. In order to understand what occurs, a knowledge of the thermodynamics of films is essential.

If a crystal of an organic acid, an alcohol, or an ester is placed on the surface of clean water, it is found to spread as a monolayer. Oleic acid seems to spread to some extent as a duplex film, but the monolayer spreads more rapidly and therefore ahead of the duplex film, which changes rapidly to a monolayer and lenses.

Theories of Spreading. There are two *special* theories of the spreading of oil on water.

* A considerable part of the discussion of spreading, and of the tables of this chapter, have been taken from a paper by the writer in *Chemical Reviews* **29**, 385-417 (1941) by permission of W. A. Noyes, Jr., editor, and of Williams and Wilkins, publishers.

(1) Rayleigh's theory:[a,b,c,d] All liquids spread on the surface of clean water. To prevent spreading, it is necessary to have a foreign substance on the surface.

(2) Langmuir's theory:[e,f] Non-polar liquids, such as the hydrocarbons, do not spread on water, but if the molecules contain a polar group and a non-polar group, the liquid will spread.

It is obvious that these two theories are in disagreement, so both cannot be true. Rayleigh's theory is in a certain sense correct, namely to the same extent that all liquids and solids are volatile. The error in Langmuir's theory is that while polar groups aid in spreading on water, the process occurs with many oils which contain no polar groups; or may not occur if a polar group is present, provided the non-polar part of the molecule is sufficiently large.

In addition there is a single general theory of spreading as outlined below.

(3) The general thermodynamic theory of Harkins,[89] which is not limited, but applies to all liquids, to the spreading of solids on liquids, and to the spreading of mobile films on solids. The theory is based on the concept that there are two types of spreading:* (a) Duplex (D) spreading: a liquid, such as benzene or the oxidized Nujol already considered, may spread as a duplex film. The thermodynamic relations indicate that this is unstable and always transforms into a non-duplex film (a monolayer) and a lens or lenses. (b) Non-duplex (or M) spreading: a non-duplex film may spread directly from a three-dimensional solid or liquid, or it may be formed by the transformation of a duplex film. All non-duplex films of known thickness are monolayers.

Duplex spreading: The theory of the spreading of a duplex film may be stated in the following abbreviated form: If a drop of an insoluble oil (b) is placed on a clean water (a) surface, it may or may not spread. The condition which determines whether or not the oil spreads is exceedingly simple. If the oil likes itself (W_{cb}) better than water (W_a) it will not spread, while if it likes the water better than itself, it will spread. In order to use this relation in a quantitative sense it is essential to express the term "likes" as a thermodynamic quantity, which is done in the following equation:

$$S_{b/a} = -(\partial F/\partial \sigma)_{p,T} = W_A - W_{cb} \tag{2}$$

Here $-(\partial F/\partial \sigma)_{p,T}$ is the rate of decrease of the free energy (F) of the system with increase of area (σ) of the duplex film, and the concomitant decrease in the area of the clean surface of the water. W_a is the work required to pull the oil and water apart over an area of 1 sq cm, and W_{cb} is that required to rupture a bar of oil with a cross-section of 1 sq cm of surface of the oil (Figs. 2 and 3). A consideration of these processes gives values for W_a and W_c as follows:

$$W_A = \gamma_a + \gamma_b - \gamma_{ab} \text{ (Dupré's equation)} \tag{3}$$
$$W_{cb} = 2\gamma_b \tag{4}$$

so

$$S_{b/a} = \gamma_a - (\gamma_b + \gamma_{ab}) \tag{5}$$

Benzene. In order to obtain a better idea of the extent to which each of the three theories is valid, the discussion given below considers the spreading of three liquids —benzene, carbon disulfide, and a paraffin oil of high molecular weight.

[a] Lord Rayleigh, *Scientific Papers*, Vol. III (1887-92), pp. 351, 397, 513, 562, 572. See also:
[b] Marangoni, *Pogg. Ann.* 118, 348, 1871 (1865).
[c] Van der Mensbrugghe, "Essai sur la Theorie Mécanique de la Tension Superficielle," *Bull. Acad. roy. Belgique*, [3], 9, 12 (1885).
[d] Quincke, *Wied. Ann.* 2, 145 (1877).
[e] I. Langmuir, *J. Am. Chem. Soc.*, 39, 1864 (1917).
[f] I. Langmuir, *Cold Spring Harbor Symposia on Quantitative Biology*, Vol. 6 (1938), pp. 173, 182.

* Hardy also assumed two types of spreading: primary, which corresponds to non-duplex, and secondary, which corresponds to duplex. One of the principal differences is that Hardy assumed non-duplex films to be moderately thick, while here they are assumed to be monomolecular in general.

For dry benzene on a clean water surface, what may be termed the initial spreading coefficient has the value

$$S_{b/a} = \gamma_a - (\gamma_b + \gamma_{ab}) = 72.8 - (28.9 + 35.0) = 8.9 \text{ ergs cm}^2 \quad (5)$$

That is, if the area of the duplex film increases by 1 cm^2 and that of the clean surface of the water decreases by this amount, there is a decrease in free energy of 8.9

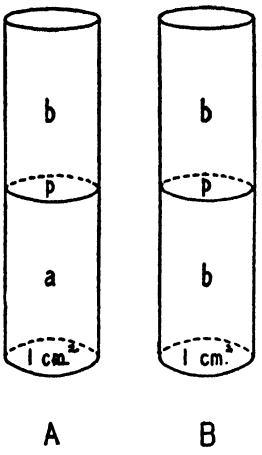


FIGURE 2. (A) If the liquid b is pulled from the liquid a at the plane p, the increase in free energy is $\gamma_a + \gamma_b$ in the ideal case in which the liquids are completely separated from each other. There is a concomitant decrease of free energy equal to γ_{ab}, due to the destruction of the interface of ab. The work done (W_A) equals the net increase in free energy, or $W_A = \gamma_a + \gamma_b - \gamma_{ab}$. (B) If the liquid b is pulled apart at the plane p, the increase in free energy is $2\gamma_b$, or $W_{cb} = 2\gamma_b$.

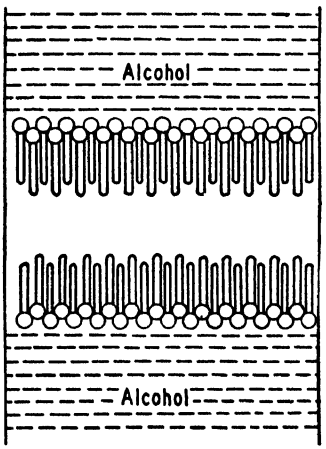


FIGURE 3. Idealized diagram illustrating the separation of a bar of liquid alcohol into two parts. The molecules are represented as already oriented at the interface, since the break is well under way. This diagram indicates that in such an ideal case the energy of separation of a moderately long-chain alcohol should be practically equal to that of a hydrocarbon of the same molecular size, except for the relatively small energy of orientation of the molecules at the interface before the actual separation has occurred.

ergs, which is a considerable one; and this indicates that benzene should spread readily to form a duplex film.

However, since the extremely thin duplex film of benzene becomes saturated with water very rapidly, the coefficient of spreading changes rapidly to the semi-initial coefficient, which is

$$S_{b'/a} = \gamma_a - (\gamma_{b'} + \gamma_{a'b'}) = 72.8 - (28.8 + 35.0) = 9.0 \text{ ergs cm}^{-2} \quad (6)$$

where the prime indicates that b is saturated with a. The value of this spreading coefficient indicates that wet benzene also spreads on a clean water surface to form a duplex film.

If b and a are mutually saturated, then

$$S_{b'/a'} = \gamma_{a'} - (\gamma_{b'} + \gamma_{a'b'}) = 62.2 - (28.8 + 35.0) = -1.6 \text{ ergs cm}^{-2} \quad (7)$$

This may be designated as the final spreading coefficient. Its negative value indicates that benzene will not spread over the surface of water if the liquids are mutually saturated. An inspection of the numerical terms involved shows that the whole change between $S_{b'/a}$ and $S_{b'/a'}$ is due to a decrease of 10.6 ergs cm^{-2} in the free surface energy of the surface of water. Now if the free surface energy of the water saturated with benzene is subtracted from that of the clean water, there is obtained as the difference (π_e)

$$\gamma_w - \gamma_e = \pi_e \quad (8)$$

where, if surface tensions instead of free surface energies are considered, π_e is by definition the film pressure, and the subscript in γ_e indicates that the film of the oil b is in equilibrium with a lens of the oil.

Now equation (6) minus equation (7) gives

$$S_{b'/a} - S_{b'/a'} = \gamma_a - \gamma_{a'} = \gamma_w - \gamma_e = \pi_e \quad (9)$$

or the difference between the semi-initial and the final coefficients of spreading is equal to the equilibrium film pressure. It is shown above that for benzene on water this difference is 10.6 ergs cm^{-2}, which is therefore the value of the film pressure of the non-duplex film of benzene on the water.

To obtain this thin film a clean surface of water in a large crystallizing dish may be partly covered with a layer, that is, a large lens, of benzene. If the curvature of the surface of the water is adjusted to give such a result, the lens or layer (*ca.* 2 mm thick) will form an annular ring in contact with the outer edge of the dish, while in the center a circular hole will be formed. According to Langmuir, there is no benzene film in this hole, but the vertical type of film balance, or the ring method, may be used to test this, and such a measurement gives a film pressure of 10.6 dynes cm.$^{-1}$ Thus the surface of the water in the hole is covered with a monolayer which gives this high pressure, corresponding to about 200 atmospheres. Obviously this monolayer is in equilibrium with the large lens.

Carbon Disulfide. (Fig. 4). For carbon disulfide on water the semi-initial coefficient of spreading is

$$S_{b'/a} = \gamma_w - (\gamma'_o + \gamma'_{wo}) = 72.8 - (31.8 + 48.6) = -7.6 \text{ ergs cm}^{-2}$$

and the final coefficient is

$$S_{b'/a'} = 70.5 - (31.8 + 48.6) = -9.9 \text{ ergs cm}^{-2}$$

or

$$\pi_e = S_{b'/a} - S_{b'/a'} = 2.3 \text{ ergs cm}^{-1}$$

Thus the values of the initial and semi-initial spreading coefficients predict that carbon disulfide will not spread as a duplex film; but the difference between the semi-initial and final spreading coefficients indicates that it does spread as a monolayer, which has an equilibrium film pressure of 2.3 dynes cm^{-1}. Presumably this monolayer is a gaseous film. Obviously it may have any film pressure less than 2.3 dynes cm^{-1} if it is not in equilibrium with a lens of the liquid and the vapor from the liquid.

Methylene iodide, a non-spreading oil in the sense that its spreading coefficient has a very large negative value ($S_{o'/w} = -23.6$ and $S_{o'/w'} = -24.2$), nevertheless gives a non-duplex film or monolayer with a film pressure of 0.55 dyne cm^{-1}.

Paraffin Oil of Very High Molecular Weight. A paraffin oil may have a molecular weight so high that at 20° C it gives almost no vapor pressure in three di-

mensions. Such an oil gives a high negative spreading coefficient, so it does not spread as a duplex film. Its cohesion is so large that it does not even spread as a gaseous monolayer, at least not sufficiently to give a measureable film pressure with the ordinary type of film balance. Such an oil may be considered as non-spreading.

Normal hexadecane (M. W. = 226) has an initial spreading coefficient of −9.31, so it does not spread as a duplex of film. The equilibrium pressure of its gaseous monolayer has not been determined, but is presumably small.

Free Energy Diagram for the Spreading of Oils on Water (Fig. 4)

The amount of work involved in forming one sq cm of a clean, plane surface of water at 20° C is 72.8 ergs. Thus in a diagram (Fig. 4) the surface of water may be considered as existing at a free energy level of 72.8 erg cm^{-2}. Any oil which can reduce this free-energy level by spreading to give any particular type of film may spread to give such a film. If the particular type of oil film lies at a higher free-energy level than that of water, it will not spread spontaneously, and if it could be spread by any mechanical means, would be unstable.

It is apparent from the figure that all four of the liquids represented spread on water to form monolayers. The spreading coefficient, that is the decrease in free energy involved when a lens of the oil spreads to form its equilibrium monolayer, is given by the value of the film pressure, π. The values in erg cm^{-2} are as follows: *n*-heptyl alcohol, 44.2; benzene, 10.7; carbon disulfide, 2.3, and methylene iodide, 0.6. Thus this diagram appears to justify that part of Rayleigh's theory which considers that all liquids spread on water. However, the equilibrium film pressure of a non-polar liquid oil of very high boiling point, such as octadecane, is presumably very low (at 30° C). The film pressure of a solid *n*-paraffin of very high molecular weight is extremely small and would be difficult to detect.

With respect to their spreading as duplex films all liquids fall into two classes: those for which the free energy of the duplex film is (1) below, and those for which it is (2) above, that of the subphase. The free-energy level of any duplex film is equal to the sum of its two interfacial energies, or $f = \gamma_o + \gamma_{ow}$. Thus for benzene, saturated with water, $f = 28.8 + 35.0 = 63.8$, while for dry benzene $f = 28.9 + 35.0 = 63.9$. Thus the duplex film does not exhibit a single free-energy level, but its different levels are close together.

The decrease of free energy for the spreading of a duplex film of benzene on water is 9.0 erg cm^{-2}, and for isopentane 9.4 erg cm^{-2}, which indicates that not only aromatic compounds, but also lower *n*-paraffin oils, spread readily on water to form duplex films.

While carbon disulfide and methylene iodide spread as monolayers, it is evident that they do not form duplex films.

The lower hydrocarbons are found to spread both as duplex films and monolayers, and those of higher (but not too high) molecular weight give monolayers alone.

The Non-Spreading of Water on Organic Liquids

From the relation for the initial spreading coefficient

$$S_{b/a} = W_A - W_{Cb} = \gamma_a - (\gamma_b + \gamma_a{}'_b{}')$$

it has been found that water will not spread on the organic liquids for which the relations are known; presumably this means that it will not spread on any organic liquid. Here *a* is an oil and *b* is water. For isopentane, $S_{w/o} = 13.7 - (72.8 + 49.6) = -108.7$, an extremely large negative value. For carbon disulfide $S_{w/o} = 31.4 - (72.8 + 48.6) = -90$. For water on primary isoamyl alcohol, $S_{b/a} = 23.7 - (72.8 + 5.0) = -54.1$. However the greater mutual solubility of a pair of liquids such as these reduces greatly the value of the final spreading coefficient. Thus for

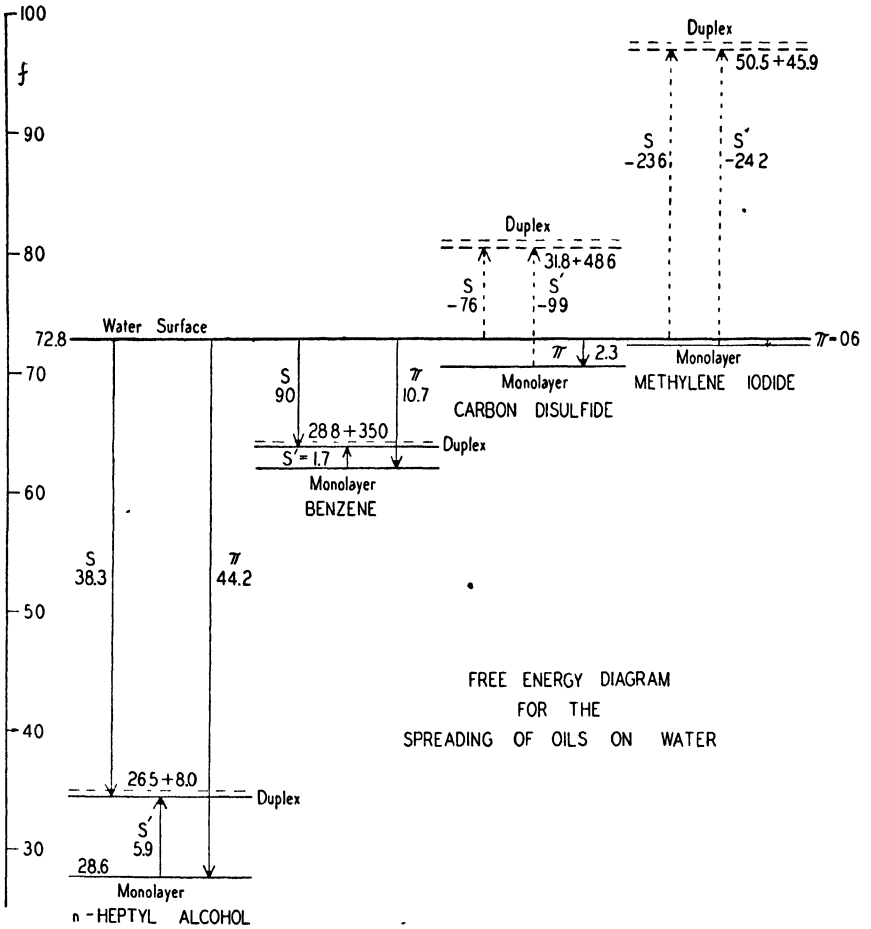


FIGURE 4. Free surface energy (f) diagram for the spreading of duplex films and of monolayers on the surface of water. The free surface energy of a duplex film is the sum of that for the interface (γ_{ww}) and for the surface of the oil ($f_a = \gamma_{wo} + \gamma_o$). The lower of the two energy levels of a duplex film represents oil saturated with water, while the upper level is that when no water is in the surface of the oil. S designates the semi-initial spreading coefficient $S_{b'/a}$, which may be positive (arrow pointing downward) or negative (arrow pointing upward). S', which represents the final coefficient ($S_{b'/a'}$) *is always negative.* Obviously if the free energy of a duplex film is above that of water, it will not form; if below that of water it may spread, *but is always unstable with reference to the formation of a monolayer (and a lens).* For the spreading of water on an oil, the free-energy levels for the duplex film are above that for water (usually by 40 to 110 erg cm^{-2}), so water does not spread on an oil as a duplex film. Thus, except for the greater height of the duplex level the diagram is like that for methylene iodide.

Any film the free-energy level of which is above that of water (72.8 ergs cm^{-2} at 20° C) cannot spread, and any film the level of which is below that of water may spread spontaneously on water. Benzene, a non-polar liquid (isopentane has practically the same levels), should, according to the diagram, be able to spread either as a duplex film or as a monolayer, but only the latter is stable. For carbon disulfide, a non-polar liquid, the duplex film is at a free-energy level 7.6 ergs cm^{-2} higher than that of a clean water surface, so a duplex film is not formed by spreading. The free-energy level of the gaseous monolayer is, however, 2.3 erg cm^{-2} below that of the water surface, which rapidly attains equilibrium with the lens and the vapor in a saturated atmosphere.

this alcohol $S_{b'/a'} = 23.6 - (25.9 + 5.0) = -7.3$, which, while still negative, is very small for water on an oil. However, it is obvious that the smallness of this value is due to the great reduction of the free surface energy of the water which is caused by the oil.

The Spreading of Liquids and Solids on Mercury (Fig. 5)

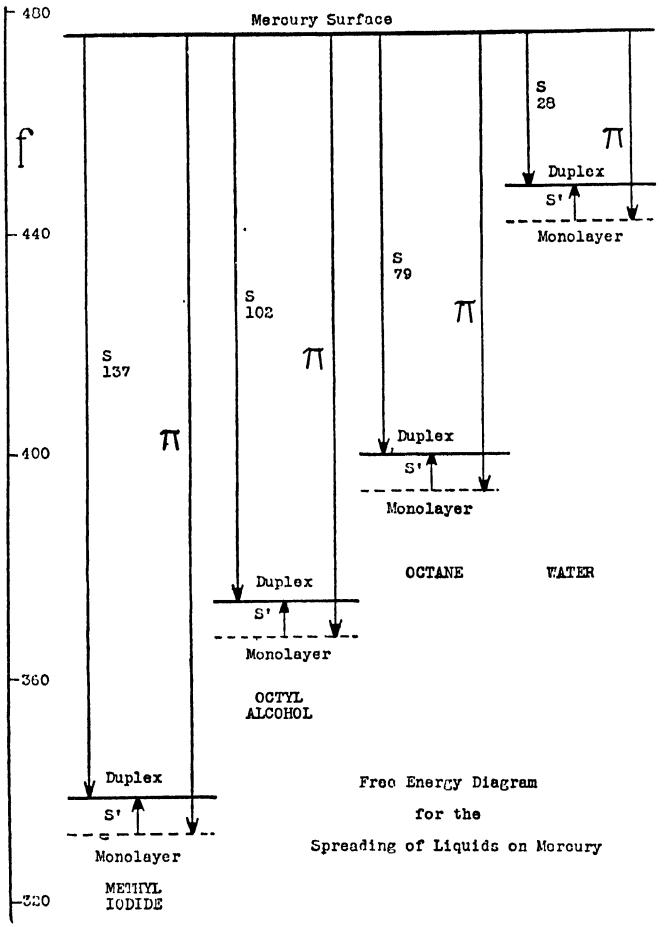


FIGURE 5. Water and all organic liquids have positive initial spreading coefficients on mercury, and thus may spread to form duplex films. The monolayers lie at a lower free energy level, so the duplex film, if formed, transforms into a monolayer and a lens. The free energy levels of the duplex films are known and are properly placed in the diagram. Those for the monolayers have not been determined, so only the sign of the final spreading coefficient (S') is known. While the film pressure is larger than the initial spreading coefficient, it is probable that it is not very much larger, so the diagram gives the general location of the free-energy level for each monolayer.

The free surface energy of mercury in erg cm^{-2} is: 480 at 0° C, 476 at 20° C, and 467 at 60° C. The total surface energy (h or e) is 540. Thus if a free-energy diagram, similar to Fig. 1, is constructed for the spreading of liquids or solids on

mercury, Fig. 5 is obtained and the level for the mercury surface at 20° C is very high (476). The result of this is that all duplex and all monolayer levels for water and for oils lie far below the mercury level. There is a large decrease in free energy if water or any liquid oil spreads as a duplex film on mercury. Thus water and all organic liquids spread on mercury.

There are no reliable determinations of the final spreading coefficient, $S_{b'/a'}$, or of the equilibrium film pressure for any liquid on mercury. However, the lenses which form on the non-duplex film, presumably a monolayer, indicate that this film pressure is somewhat greater than the initial spreading coefficient, $S_{b/a}$. In ergs per cm^{-2} the latter quantity has the values: water, 32; acetone, 60; hexane, 79; propyl alcohol, 85; carbon tetrachloride, 93; benzene, 90; chloroform, 97; octyl alcohol, 102; ethylene bromide, 116; ethyl iodide, 135; and methyl iodide, 137.

Many organic solids are found to spread on the clean surface of mercury.

Transformation of a Duplex Film into a Monolayer and a Lens

The examples given in this section show that in each case the free surface energy of the monolayer is less than that of the duplex film, provided the latter can spread at all. This is a general relation; the duplex film always changes into a monolayer and a lens (or lenses if the time is not sufficiently long to give a single lens). If a duplex film cannot form, any film that is produced is a monolayer. Such a monolayer may be formed even if the initial spreading coefficient is negative, but may not give a measurable film pressure if the oil is non-polar and does not give a measurable three-dimensional vapor pressure.

If the atmosphere above an oil film is supersaturated with oil, the material condensed on the monolayer does not thicken it; but any excess deposited on the monolayer either evaporates or goes into the monolayer. As a result of the latter process, either the monolayer expands, or an equal amount of material in the monolayer is forced into the lens.

The rate of spreading of an oil on water increases with the magnitude of the spreading coefficient and with a decrease in the viscosity of the layer which spreads. The values of the initial spreading coefficients of a few oils are given in Table 1. As the final spreading coefficient of an oil on water is always negative (Table 2),

Table 1

A. Initial Spreading Coefficients of a Few Oils on Clean Water at 20° C.

Molecules with Polar Groups	S	Paraffins	S	Non-polar Aromatic Hydrocarbons	S
n-Octyl alcohol	35.7	Isopentane	9.4	*p*-Cymene	10.1
Heptaldehyde	32.2	Hexane	3.4	Benzene	9.8
Oleic acid	24.6	Diisoamyl	3.7	*o*-Xylene	6.9
Ethyl nonylate	20.9	Heptane 30°	0.2	Toluene	6.8
o-Nitrotoluene	4.1	40°	0.5	*p*-Xylene	6.7
Nitrobenzene	3.8	50°	0.9	Mesitylene	5.6

Liquids with Negative Coefficients	S	Liquids with Negative Coefficients	S
Ethylene dibromide	−3.2	Monoiodobenzene	−8.7
o-Monobromotoluene	−3.3	Bromoform	−9.6
Carbon disulfide	−4.2	Acetylene tetrabromide	−15.6
		Methylene iodide	−26.5

Temp. (°C) =	10	20	30	40	50	60	70
n-Heptane	−0.3	−0.4					
n-Hexadecane	−9.56	−9.52	−9.40	−9.26	−9.13	−9.03	

B. Some Final Spreading Coefficients of Oils on Water

Oil	S	Oil	S
Benzene	−1.5	*n*-Heptyl alcohol	−5.9
Carbon disulfide	−10.0	Isoamyl alcohol	−2.6
Methylene iodide	−24.2		

All final spreading coefficients are negative.

Table 2. Equilibrium Film Pressures (π_e), and Initial ($S_{b/a}$), Semi-initial ($S_{b'/a}$), and Final($S_{b'/a'}$) Spreading Coefficients of Oils on Water at 20° C. (H_2O = 72.75.)

	γ_e	π_e	γ of Dry Oil	γ of Wet Oil	γ_i	$S_{b/a}$	$S_{b'/a}$	$S_{b'/a'}$
Methylene iodide	72.20	0.55	50.68	50.52	45.87	−23.80	−23.64	−24.19
Carbon disulfide	70.49	2.26	32.35	31.81	48.63	− 8.23	− 7.69	− 9.95
Benzene	62.36	10.39	28.88	28.82	35.03	8.84	8.90	− 1.49
n-Heptyl alcohol	28.53	44.22	26.97	26.48	7.95	37.83	38.32	− 5.90
Isoamyl alcohol	25.92	46.83	23.73	23.56	5.00	44.02	44.19	− 2.64

Reference 35, p. 25 (1928).

no oil will spread over its equilibrium monolayer. As the initial and final coefficients for water on oil are always negative, water will not spread over the surface of an oil.

Since (1) many hydrocarbon oils, which do not contain polar groups, nevertheless spread on water to duplex films and then to monolayers, and (2) substances whose molecules contain a strongly polar group do not spread if the hydrocarbon fraction is too large, it is apparent that the presence or absence of a polar group is not the true criterion for spreading or non-spreading. Examples of the second type are arachidic acid (20 carbon atoms) and acids having longer normal paraffin chains. These substances are solids because of the higher cohesion due to the longer chains.

Table 3. Effect of Molecular Constitution upon the Formation of Duplex Films and of Monolayers

Molecules or Groups	Attraction for Water W_A	Attraction for Like Molecules W_C	Duplex Film [1]	Stability of Monolayer
Hydrocarbons of high molecular weight	Weak 60	Weak 65	None	None or gaseous
$-CH_2I$, $-CH_2Br$, $-CH_2I$	Moderately strong 75	Strong 100	None	None or gaseous
Hydrocarbons of low molecular weight	Weak 45	Weak 40	Unstable	Gaseous
Esters	Strong 80	Weak 50	Unstable	Stable. Relatively high compressibility
$-CH_2OH$, $-COOH$, $-CN$, $-CONH_2$, $-C_6H_4OH$, $-NHCONH_2$, etc.	Very strong 100	Weak 50	Unstable	Stable liquid, solid and gaseous
$-SO_2H$, $-C_6H_4SO_3H$, $-SO_4H$	Extremely strong	Moderately weak	Forms if material liquid, but unstable	Form monolayers
Soaps (solids)	Extremely strong	Moderately weak	None	Form monolayers

[1] If the spreading material is a solid, a duplex film does not form, but only a monolayer spreads. See Section 12 for the spreading of a duplex film of an oil which, of itself, does not spread in this way, by the addition of a polar-non-polar oil as a second component.

Table 3 expresses extremely general relations of the process of spreading, but it should be remembered that these cannot be entirely accurate.

Phase Relations of Monolayers

General Relations and Order of a Phase Transition

Monolayers exhibit surface phases in many respects analogous to those found in three-dimensional systems: (1) gas, (2) liquid, and (3) solid. There are in addition three other phases which seem to show no direct analogy to three-dimensional

systems. The first of these, discovered by Labrouste, is a coherent phase, and in that sense analogous to a liquid, but it may occupy as much as two and a half times the area of the most tightly packed liquid phase. In three dimensions this would be as if a quart of milk could exist in another expanded liquid state to give a volume as large as four quarts. This two-dimensional phase has been designated by Adam as an *expanded liquid* film.

There is also a phase of very high compressibility, with an area between that of the condensed and the expanded liquid phases, designated by the writer as the *intermediate* liquid phase. Some workers believe this to consist of small islands of the condensed liquid phase in a sea of the expanded liquid.

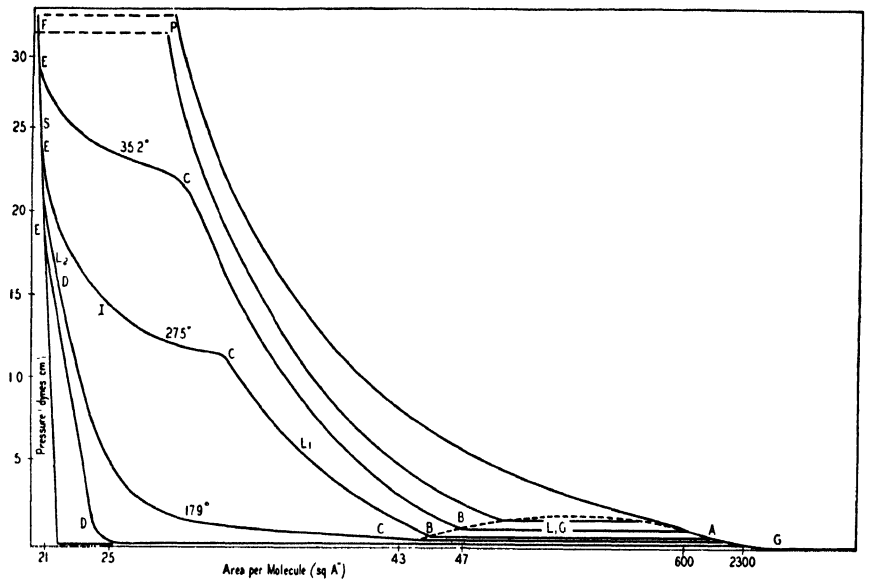
A sixth phase, discovered recently by Harkins and Copeland,[87, 88] exhibits remarkable and anomalous properties. It has the small compressibility of a solid or S phase, but at certain temperatures is much less viscous than the condensed liquid (L_c) phase from which it is formed by increase of pressure. Since it shows some properties characteristic of the liquid, and others of the solid state it has been designated by the writer as the LS phase.

Since, also, the "solid," which is the most highly condensed phase, *may*, under some conditions, have a viscosity as low as that of a liquid monolayer and may exhibit no rigidity, it will be designated merely as the S phase.

Table 4. Two-dimensional Phases and Probable Order of Changes Between Them

Phases	Order of Change Between Them
I. Gas, G. Includes compressed vapors such as the "vapor expanded" film of Adam Perfect gas: $\pi\sigma = kT$ Imperfect gas: minimum value of κ of same order as for L_1 films: e.g., 0.04 for ethyl palmitate at 25 C, 34A^2, and 14 dynes cm^{-1}	
	I and II: first (L_1G)
II. Liquid, L_e (BC). "Liquid expanded" of Adam. Liquid of high compressibility κ = 2 to 7 $\times$ 10^{-2} h_s = 50 ergs cm^{-2}	
	II (or I) and III: second (C)
III. Intermediate (I) or transition (CE). Liquid of extremely high compressibility Intermediate, I. Maximum κ of order of 2 to 5 $\times$ 10^{-1} for normal long-chain acids 2.2 $\times$ 10^{-1} for normal long-chain esters Minimum same as κ of L_2 film: h_s = 300 ergs cm^{-2}, which is nearly constant at the higher areas, but decrease rapidly to a low value as transition to next phase is approached	
	III and IV: third (D)
IV. Liquid, L_c (DE). "Liquid condensed" of Adam. Liquid of low compressibility κ = 5 to 10 $\times$ 10^{-3} π = b $-$ $a\sigma$ log η = log η_0 + $k\pi$: h_s small	
	IV and V: first (perhaps anomalous) changing to second at higher temperatures. (See Fig. 14)
V. LS. Superliquid phase of extremely low compressibility κ = 0.0005 to 0.0017. Viscosity becomes high at temperatures far above that of first-order charge π = b $-$ $a\sigma$	
	IV and VI: second V and VI: second
VI. S. Previously classed as solid, may have the rigidity of a solid, or a non-Newtonian or a Newtonian viscosity. The most highly compressed phase. κ = 0.0005 to 0.001 π = b $-$ $a\sigma$	

These phases and their most prominent characteristics are outlined in Table 4, while their general pressure (π)-area (σ) relationships are represented in Fig. 6. No distinction is made in the figure between S and LS films, since it is on too small



(In the figure L_1 is used for L_e and L_2 for L_c).

FIGURE 6. General phase diagram of a monolayer at a temperature below the critical temperature for the formation of a liquid expanded (L_e) film. Phases: FE or S, solid: ED or L_c liquid condensed; DC or I, intermediate; CB or L_e, liquid expanded; BA or L_eG, gaseous film and L_e film in equilibrium; G gaseous film. The temperatures given are those for pentadecylic acid. For longer chain acids and for an alcohol of the same chain length the temperatures are higher, etc. The scale of molecular areas is much shortened (above about 47 Å) and the pressure scale is much expanded.

a scale for this purpose; but these highly condensed monolayers are shown later in Fig. 14.

The temperatures given in Fig. 6 are correct for pentadecylic acid, but are lower for acids with a lower, and higher for acids of a higher chain length. With alcohols the temperatures are much higher.

If a three-dimensional solid is changed to the liquid state, heat is absorbed by the system, and it seems difficult to conceive of the melting of a solid without the accompaniment of a latent heat. Such a phase transition is said to be a first-order change.

A phase change is of the second order, in the terminology of Ehrenfest, if latent heat is absent, and there are discontinuities in the heat capacity and in the coefficients of thermal expansion and compressibility. In a third-order change there is a discontinuity only in the derivative of the specific heat.

The order of a phase change is determined, according to Ehrenfest, by the lowest derivative of the free energy of the system which exhibits a discontinuity. As applied to monolayers this is indicated by the equations given below.

The Gibbs free energy is defined as follows:

$$F = E - ST - pv - \gamma\sigma \tag{10}$$

$$dF = dE - S\,dT - T\,dS - p\,dv - v\,dp - \gamma\,d\sigma - \sigma\,d\gamma \tag{11}$$

The differential of the internal energy (dE) is

$$dE = dQ - dW = T\,dS - p\,dv + \gamma\,d\sigma \tag{12}$$

so

$$dF = -S\,dT + v\,dp - \sigma\,d\gamma \tag{13}$$

and

$$(\partial F/\partial\gamma)_{p,T} = -\sigma \tag{14}$$

or, since

$$d\pi = -d\gamma \tag{15}$$

$$(\partial F/\partial\pi)_{p,T} = \sigma \tag{16}$$

$$(\partial F/\partial T)_{p,\sigma} = -s = -q/T \tag{17}$$

Thus a discontinuity in area at constant film pressure and temperature, or in entropy or heat absorbed, is a first-order change.

Now

$$(\partial^2 F/\partial\pi^2)_{p,T} = (\partial\sigma/\partial\pi)_{p,T} \tag{18}$$

$$\kappa = -\frac{\partial\sigma}{\sigma\partial\pi} \tag{19}$$

so a discontinuity in the compressibility without one in the area indicates a second-order change

Also

$$(\partial^2 F/\partial T^2)_{p,\pi} = C_\pi/T \tag{20}$$

$$(\partial^3 F/\partial\pi^3)_{p,T} = (\partial^2\sigma/\partial\pi^2)_{p,T} \tag{21}$$

$$(\partial^3 F/\partial T^3)_{p,\pi} = \left[\frac{\partial(C_\pi/T)}{\partial T}\right]_{p,\pi} \tag{22}$$

and a discontinuity in the third derivative but not in the second, indicates a third-order change.

Gaseous or Vapor Films (G, Fig. 7)

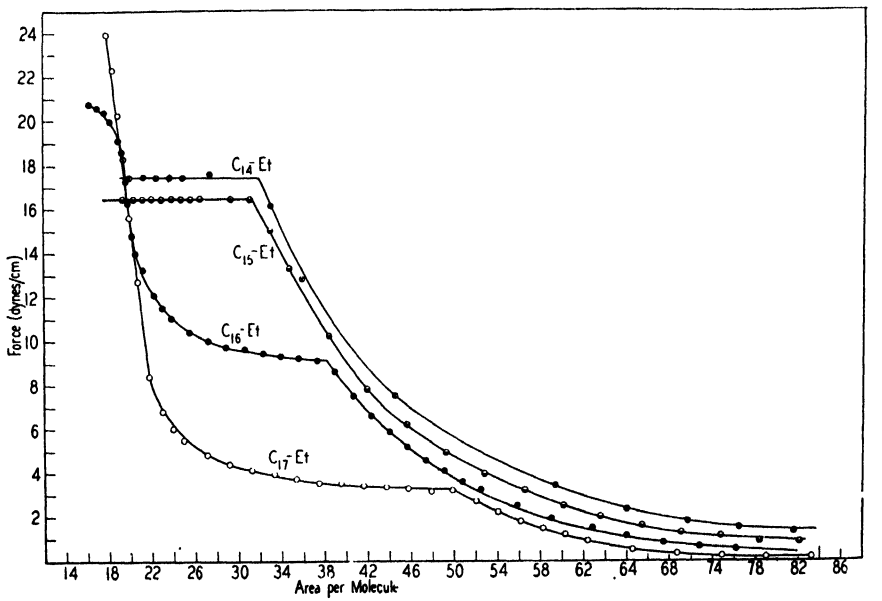


FIGURE 7. General phase diagram at temperatures above T_c for the liquid expanded phase: ethylesters at 25° C.

At very large molecular areas of the order 50,000 Å^2 the monolayer of pentadecylic acid seems to exhibit a behavior in accord with the perfect gas law $\pi\sigma = kT$, where σ is the molecular area and k is the Boltzmann constant of a three-dimensional gas. As the area is decreased, the two-dimensional gas becomes imperfect and the pressure increases in accord with a relation of the van der Waals type. If the temperature is 14.5° C the pressure becomes constant at a pressure of 0.11 dyne cm^{-2} at an area of about 2,400 Å^2 per molecule and remains constant until this is reduced to a relatively small area, above which the pressure increases. At 18° C this lower limit is 44 Å^2 per molecule, but at 14.5° C it is probably about 26 Å^2. Between these limits the film is heterogeneous, and at temperatures about 19 to 40° C and more, consists of islands of expanded liquid film in a sea of gaseous film.

The heterogeneity of this film in the region L_1G of Fig. 6 has been demonstrated in the work of several investigators by a method suggested by the writer. This consists in surveying the surface of the film by moving over it an electrode with which the surface potential may be determined. If the film is entirely gaseous, the surface potential is in general less than 5 mv; however, if there are islands or continents of a more condensed film this potential is much higher, since it increases in general with the mean number of film molecules under unit area of the surface of the electrode. The results of such a survey, as made by Harkins and Fischer, are shown later in Fig. 22. Here the maximum surface potential for myristic acid is at 56.5 Å^2, since the islands consist of patches of liquid-expanded film.

Compressed or High-pressure Vapor Films (Figs. 7 and 8)

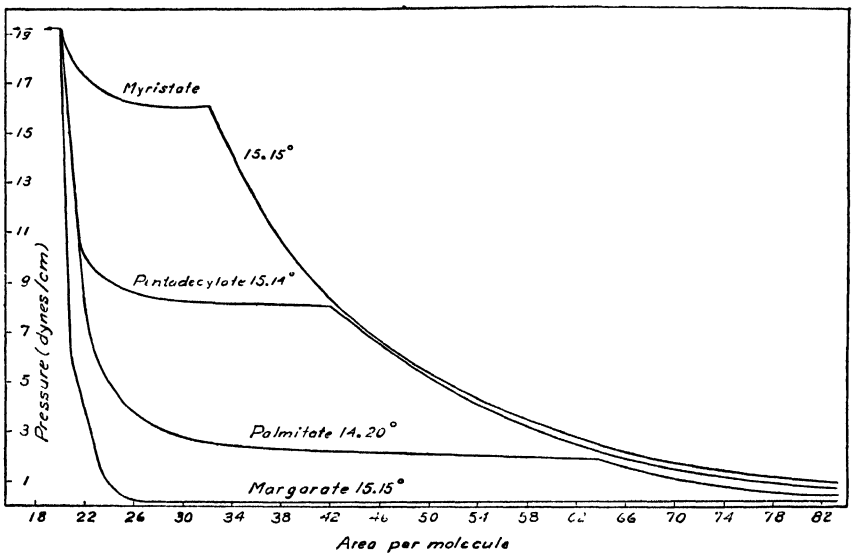


FIGURE 8. Esters at 15° C.

If the temperature is above the critical point for the formation of a liquid-expanded monolayers, the vapor film may be compressed to relatively low areas and high pressures, as shown in Figs. 7 and 8, where the righthand, nearly hyperbolic curve represents a gaseous film. Such films have been designated by Adam as *vapor-expanded,* since their π-σ curves resemble those of "liquid-expanded" monolayers. They are distinguished from these by the fact that they exhibit higher pressures at large areas, and no discontinuity on further expansion, while such liquid-expanded films as have been investigated show a discontinuity at a molecular area of from 40 to 50 Å^2 for single-chain molecules.

Compressed vapor films condense directly into the intermediate liquid state by a phase transformation of the second kind, that is without any latent heat of condensation. These relations show that a two-dimensional gas may condense into a liquid by:

(1) A change of the first order with emission of heat.

(2) A change of the second order without latent heat.

The Expanded Liquid Phase

The expanded liquid phase of a monolayer exhibits very nearly, but not quite, a hyperbolic relation between pressure (π) and area (σ), as shown in the curves on the right of Figs. 9, 10, 11 and 12, which represent the normal straight-chain paraffin

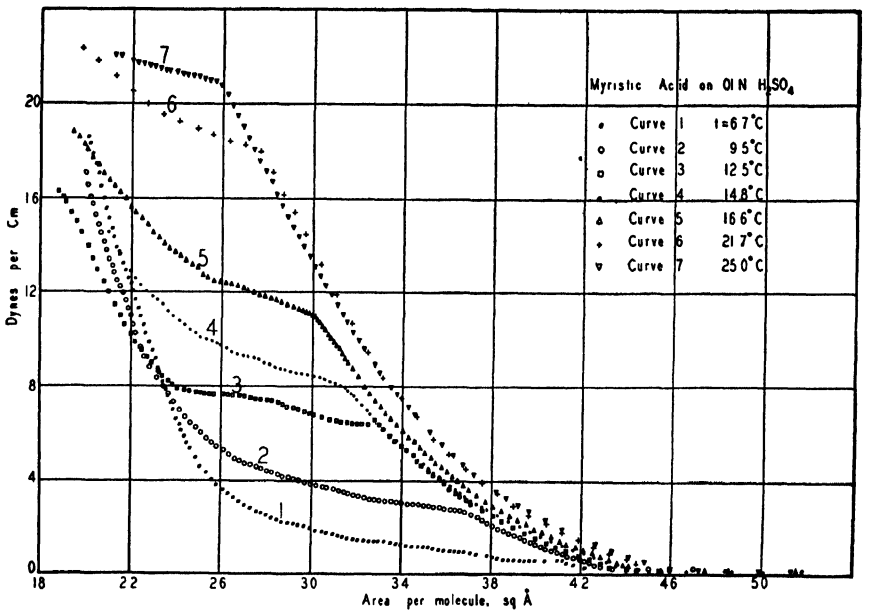


FIGURE 9. Liquid monolayers of myristic acid. Curve 7 for 25° C is shifted toward lower areas by the solubility of the film at this temperature.

acids of from 14 to 16 carbon atoms per molecule, and also octadecane nitrile. It is apparent that the areas of these hyperbolas lie at molecular areas above zero, and at film pressures below zero.

Langmuir has expressed this relation by the equation

$$(\pi + a)(\sigma - b) = nkT \tag{23}$$

but it is found that neither a nor b is a constant.

The Intermediate Liquid Phase. (Figs. 9, 10, 11 and 12)

On increase of pressure at constant temperature, decrease of area at constant pressure, or decrease of temperature at constant area, the expanded liquid phase is changed, by a transformation which seems to be of the second order, into the intermediate liquid phase.

The principal characteristic of this phase is its extremely high compressibility ($-\kappa$), especially at its higher areas. This is so high that the transition considered above is sometimes considered to be of the first order. However the maximum com-

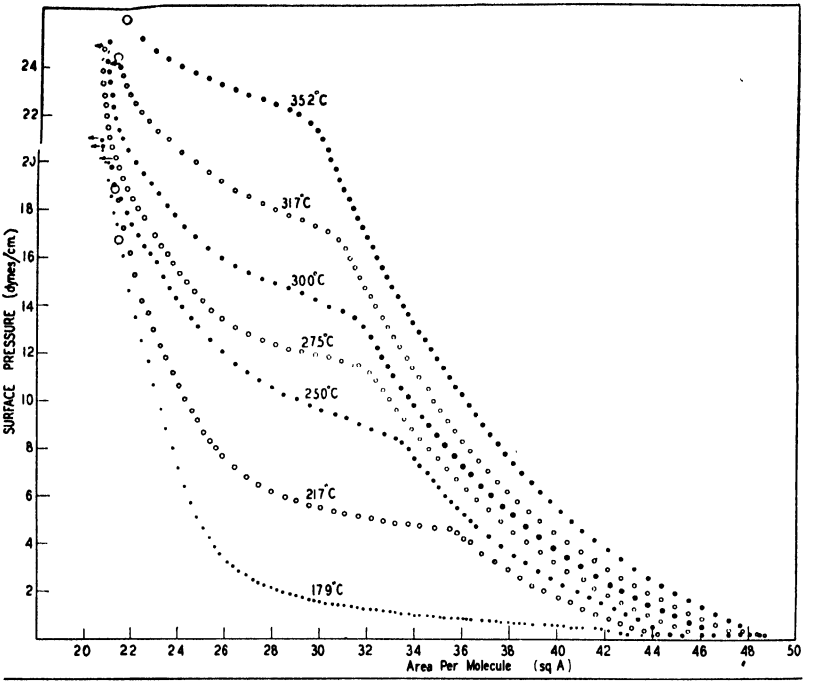


FIGURE 10. Liquid monolayers of pentadecylic (15-carbon) acid. The arrows indicate collapse.

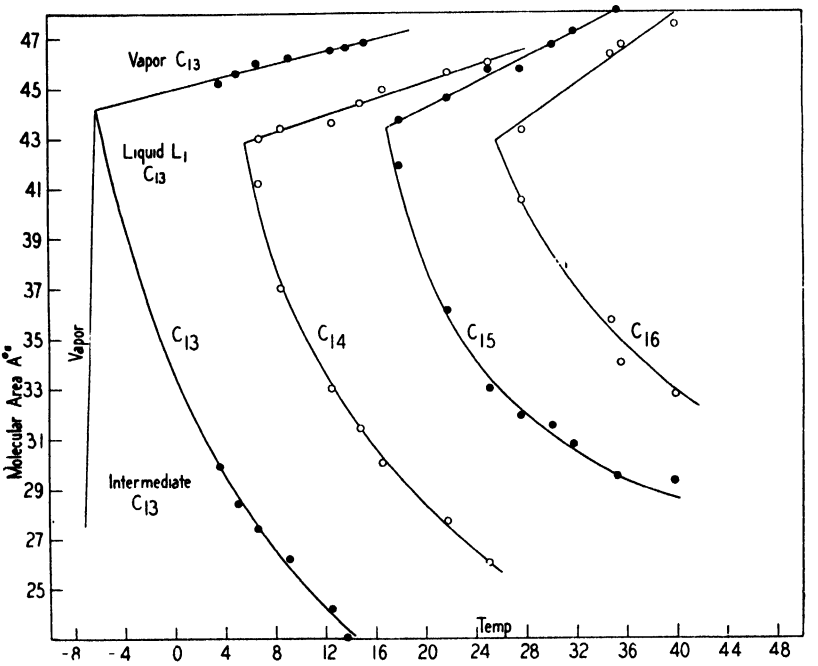


FIGURE 10a. Phase diagram showing the regions of stability for the liquid expanded films of the fatty acids C_{13} to C_{16}.

pressibility of this phase varies widely with the type of molecule. For octadecane nitrile it is relatively small (0.02 to 0.06); for long-chain acids moderately large (myristic acid 0.25, pentadecylic acid 0.35); and for esters large (ethyl margarate, 2.1).

Adam and Harding have found the intermediate phase of a margaric nitrile monolayer to be heterogeneous. This would indicate a first order change. How-

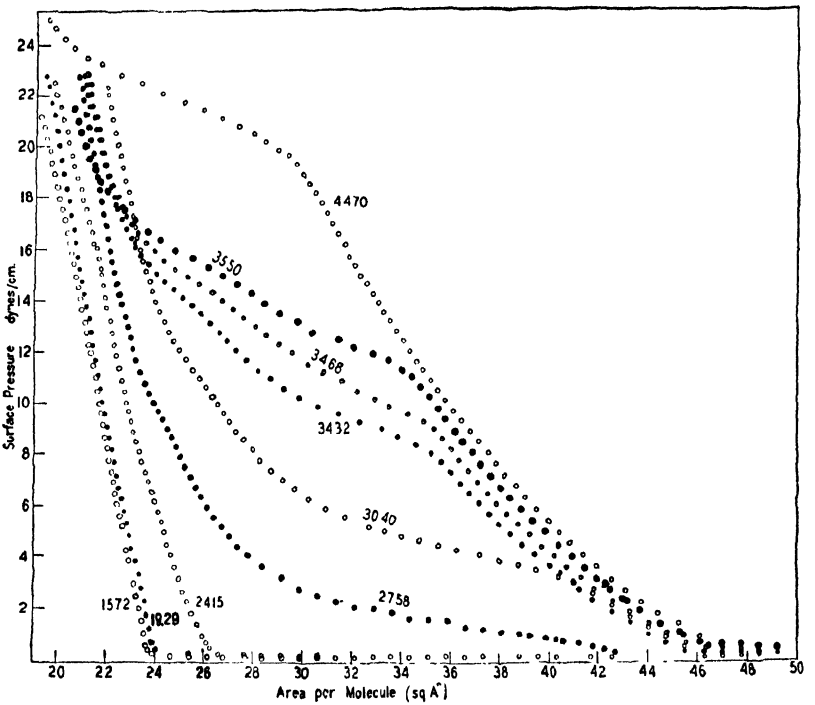


FIGURE 11. Monolayers of palmitic (16-carbon) acid. Liquid condensed (L_c) film at 15.79 and 19.29° C.

ever, this behavior, which would indicate an infinite compressibility, does not seem to accord with the relatively low value of the compressibility found for octadecane nitrile by Harkins and Copeland.

Effect of Absence of Hydrogen Bonds in a Monolayer

Since in monolayers of the long-chain alcohols and acids the polar groups are oriented toward the water, they are also brought close together, if the film is condensed. With either of these substances there is a possibility that there may be hydrogen bonding either in (1) the film itself, or (2) between polar groups in the film and molecules of water.

It is, therefore, of interest to learn the effect, produced upon the molecular area, by the elimination of the first possibility, that is, of hydrogen bonding in the film. This may be eliminated by the use of a long-chain nitrile.

The pressure-area-temperature relations obtained with monolayers of octadecane nitrile are exhibited in Fig. 12. From this figure it is evident that the —CN group gives a much higher area (with a minimum of 25.82 Å^2 at 15.1°) in the monolayer than either —CO·OH or —OH. Adam explains the relations on the basis of the assumption that the nitrile group is the largest, and the hydroxyl group the smallest

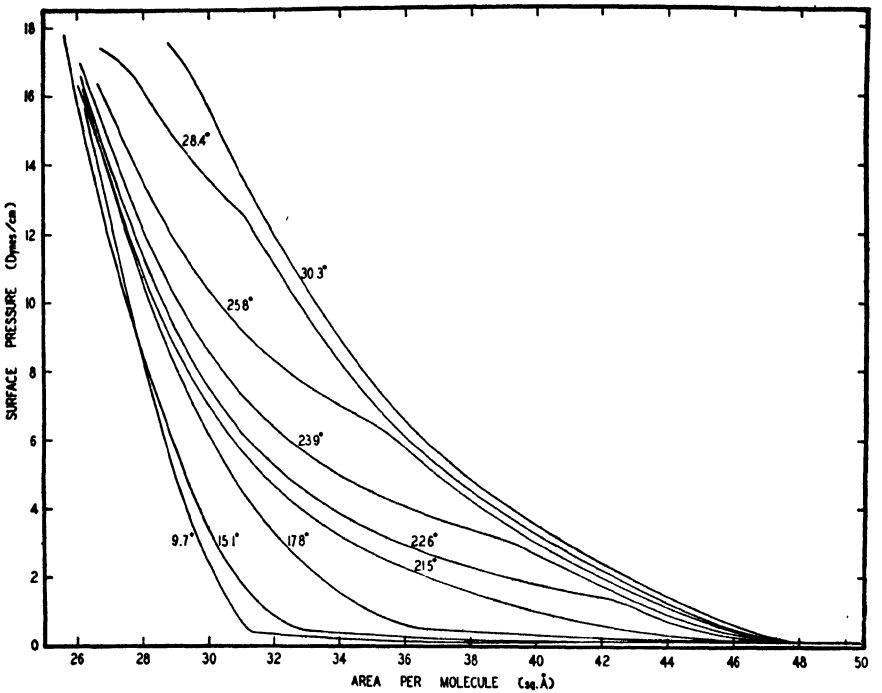


FIGURE 12. Monolayers of octadecane-nitrile on water. The intermediate phase exhibits a lower compressibility than that of the acids.

of the three. However, it does not seem possible that the nitrile group, with its two atoms and a carbon-nitrogen distance of only 1.15 Å, can be larger than the carboxyl group with its content of four atoms, even though one (H) of the latter is small. Therefore it seems necessary to reject Adam's hypothesis, and to seek another cause for the high area in the nitrile monolayers. Obviously if there is intramonolayer hydrogen bonding in the case of the acids and alcohols, its absence in the nitrile films would give a sufficient explanation of their higher areas.

Some investigators assume, without solving the statistical mechanics of the problem, that the dipoles in monolayers of this type always repel each other. If this were assumed, the higher dipole moment of the nitriles (3.56×10^{-18}) as compared with those of the acids (1.4×10^{-18}) and the alcohols (1.7×10^{-18}) would give a sufficient cause for the higher areas exhibited by the nitriles. However, as the presence of dipole repulsion has not been proved, the larger molecular areas of the nitrile monolayers may be due to either or both of the causes considered above.

The Condensed Liquid Phase (Fig. 13)

As the intermediate film is compressed, the pressure rises more and more rapidly and the compressibility decreases until at a point *D*, Fig. 4, the π-σ curve changes to a straight line. If it is considered that a change of phase occurs at *D*, then it is of the third order.

The compressibility of the liquid condensed film which exhibits this linear relation is low—usually between 0.005 and 0.01. The pressure-area relations of the *n*-long chain paraffin acids of from 16 to 20 carbon atoms are shown in Fig. 13, and it may be observed that the molecular area at a given pressure decreases with the length of the chain.

High-pressure Phases: The S, or Solid Phase, and the LS Phase of Highly Abnormal Viscosity

If a liquid condensed monolayer is compressed isothermally, it finally reaches an area sufficiently low to give a change of phase into what has usually been designated as a solid film. The phase which is formed is characterized by a compressibility only about an eighth as great as that of the liquid monolayer from which it is formed.

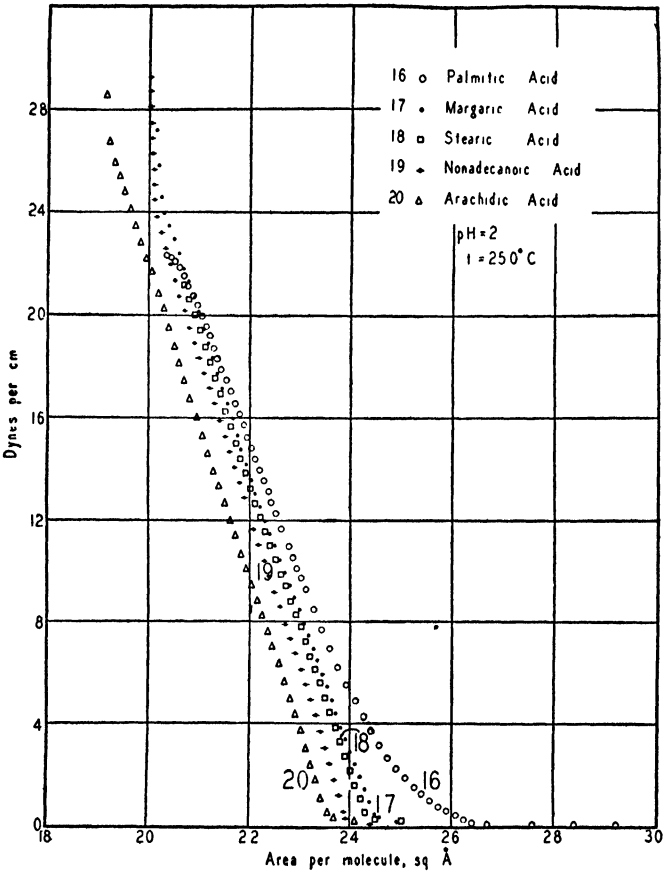


FIGURE 13. Effect of increasing chain length in decreasing the molecular area in liquid condensed (L_c) monolayers of the long chain acids of from 16 to 20 carbon atoms.

If this change of phase is considered as a freezing of the film, it is somewhat surprising to find that freezing or melting occur without latent heat. That is, the latent heat of fusion is zero, and this indicates that the change is a phase transformation of the second order.

The pressure at which this transition occurs depends upon the temperature, and in the case of an *n*-straight-chain paraffin derivative, is highly dependent upon the nature of the end group. Thus the pressure of the transition is much higher for acids than for alcohols.

Ideas concerning the "solid" phase were in a very confused state until recently, when Harkins and Copeland [88] discovered that there are two high-pressure phases

instead of one. Both of these have about the same compressibility; but the phase which the more nearly resembles a solid (*S* phase) is characterized by a lower molecular area than the new *LS* phase. The only organic compounds which have thus far been investigated sufficiently to show a distinction between the *S* and the *LS* phases are the *n*-long-chain paraffin alcohols.

While these two phases are most easily distinguished by the great differences exhibited by their viscosities, the probable existence of two phases instead of one was indicated in experiments designed to discover a first order change in a condensed monolayer.

A First Order Change in a Non-gaseous Monolayer

To those who are accustomed to consider only three-dimensional systems, the phase relations in two dimensions appear anomalous. Thus, a change from three to two dimensions, that is, a loss of one dimension, results in an increase of one in the order of a phase change of a condensed system.

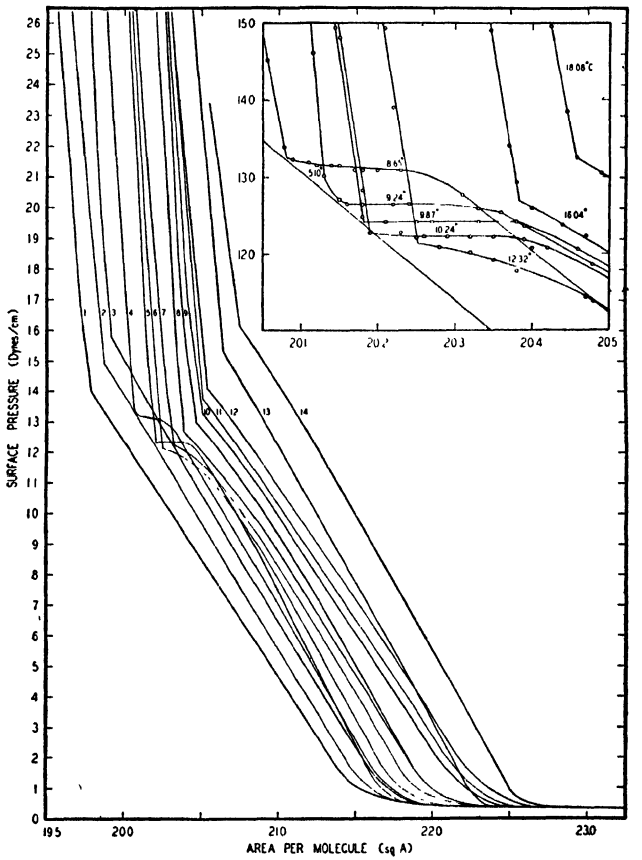


FIGURE 14a. Pressure-area temperature relations of condensed monolayers of octadecanol on water. The insert gives an enlarged section in the region of the first-order transition. Outside of this region all of the transitions are second-order. The temperatures (° C) for the curves in the principal diagram are: 1, 2.85; 2, 5.10; 3, 7.10; 4, 8.62; 5, 9.90; 6, 12.32; 7, 14.06; 8, 16.04; 9, 18.08; 10, 19.85; 11, 22.95; 12, 25.12; 13, 30.30; 14, 34.85. Curves for four other temperatures are given in the insert. Correction: Multiply Surface Pressures by 0.9315.

This would not be difficult to understand if a two-dimensional film or monolayer were suspended in space, since in such a case, molecular binding in the third direction would be absent, and this would be very likely to involve an increase in the order of a change of phase. The fact that this occurs, even when an aqueous subphase is present, indicates that the binding to the subphase is sufficiently loose to allow the increase of order which is observed.

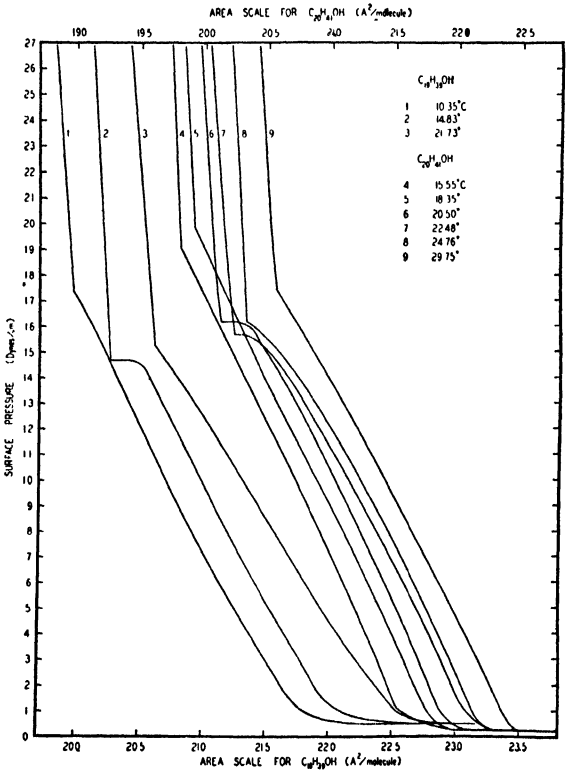


FIGURE 14b. Pressure-area temperature relations for the normal long-chain paraffin alcohols with 19 and 20 carbon atoms in the chain. The temperature of the first-order transition increases 5° while the pressure rises about 2 dyne cm^{-1} for each additional carbon atom. Correction: Multiply Film Pressures by 0.9315.

It is shown in Fig. 14a that in a narrow range of temperature near 10° C the π-σ curve is flat over a range of area of about 0.2 Å². This flatness is characteristic of a first order change. The latent heat of this change is very small and varies from 20 to 40 cal. per molecule. This is absorbed in passing from the low- to the high-pressure modification, as in the change from ice to water.

The regions of temperature and pressure in which the liquid condensed (L_c), the LS, and the S phases are stable are shown in Fig. 15, and those of temperature and molecular area in Fig. 16. The variation of pressure with temperature seems to be nearly linear for the second order changes, but is a region with a curved boundary for the first order.

The extremely interesting relation shown in Fig. 15 is that the boundary between the S and the LS phases of octadecanol lies at almost a *constant molecular area* of 19.98 Å². Thus whether the film is in the LS or the S state is *highly dependent upon the distance between the molecules.*

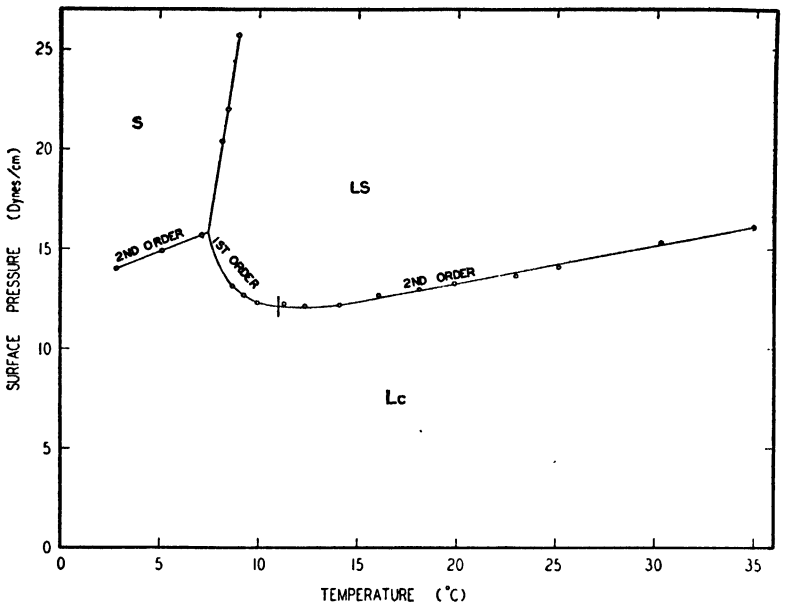


FIGURE 15. Pressure-temperature values for the phase boundaries of condensed monolayers of octadecanol. Transitions $L_e \rightleftharpoons LS$ are *first-order* from the triple point at 7.5° C on the descending curve to about 11° C, and become *second-order* above this. Transitions $L_e \rightleftharpoons S$ are *second-order*. Multiply pressure by 0.9315.

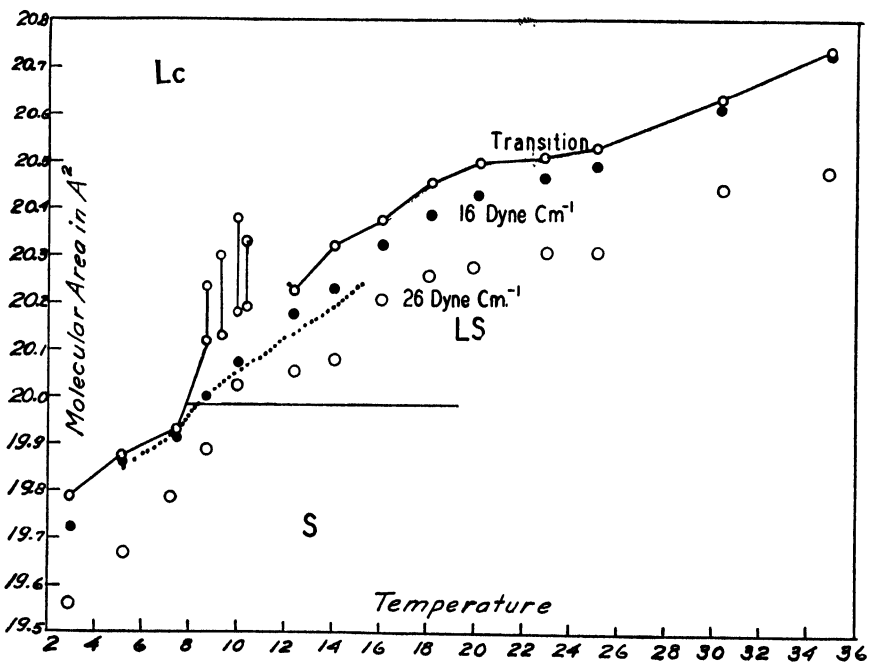


FIGURE 16. Molecular area-temperature values for the phase boundaries of condensed monolayers of octadecanol. The dotted line gives values at 16 dyne cm^{-1} as obtained by increase of the temperature of a monolayer. The black circles give values at the same pressure, from the isotherms. Multiply P by .93.

The *LS* phase exhibits highly abnormal viscosity relations, discussed in the section on viscosity, and both of the high-pressure monolayers are characterized by an extremely large heat of expansion per unit area, as shown in the section on thermodynamic relations.

The Film Balance and Accessories

The Trough

In 1891 Agnes Pockels initiated the use of a long rectangular trough, filled with water to the brim, for the investigation of the surface tension relations of films of oil. The surface of the water was cleaned by "sweeping" or "scraping" the surface by a transverse barrier which consisted of a narrow sheet of metal 1.5 cm wide. In the work of the writer the trough is shallow, about 6 mm deep, and the water rises about 6 mm above the level of its edges. It is made of stainless steel, and has a hollow bottom, through which an aqueous ethylene glycol solution is pumped from the thermostat which surrounds the trough. A trough which is inside an air thermostat, but not in contact with the thermostat water, is entirely unsuitable. Proper temperature control with such an arrangement is impossible, since (1) it takes an extremely long time to establish temperature equilibrium between a gas and a considerable body of liquid and solid, and (2) the water in the trough is subject to increases or decreases of temperature due to lack of equilibrium between the water vapor in the thermostat and the water in the trough, and this in turn is affected greatly by changes in temperature.

The barriers of paraffined glass or stainless steel are lifted or moved along the surface by the use of slots operated by levers and screws. The screws used for the measurement of the length of the film have a pitch of about 1 mm per turn, and are provided with a vernier to estimate the fractions of a turn; those used for sweeping have a pitch of about 1 cm per turn.

Film Balances

Film balances were used in the last decade of the 19th century by Agnes Pockels and by Lord Rayleigh. Any device which measures directly the difference between two surface tensions, $\gamma_1 - \gamma_2$, and commonly to obtain $\gamma_w - \gamma_f = \pi$, may be considered as a film balance. Here γ_w is the surface tension of the pure liquid, usually water, γ_f is the surface tension of the film-covered surface, and π is by *definition*, the film pressure.

Vertical Pull Film Balance (Fig. 17). The simplest, cheapest, and most widely applicable film balance is one in which the film-covered surface of the water in the trough exerts a vertical downward pull upon a sheet of solid material, such as glass, quartz, or metal, whose contact angle toward the surface of the liquid is zero. A convenient solid sheet is a microscopic slide or cover glass of quartz or glass, or a thin, plane sheet of platinum. Since the sensitivity of the method is (nearly) inversely proportional to the thickness of the sheet, an extremely thin sheet of platinum may be used when it is desired to obtain the film pressure to 0.001 dyne cm^{-1} or less.

The decrease in the downward pull on the sheet is measured by the deflection of the beam of an analytical balance, from one arm of which the slide or sheet is hung in such a way that its surface is partly below and partly above the surface of the water.

An apparatus of this type was used very early by Wilhelmy to measure the surface tension of liquids, but on account of the inherent difficulties of his method, it has remained almost unused. If, however, the process of measurement is changed into the differential form needed for the measurement of the surface pressure, all the serious difficulties and errors disappear.

Film balances of this type have been used both by Harkins and Myers, and by

Dervichian. The latter constructed such a balance in a form designed to give continuous rapid recording of the film pressure and area on photographic paper. This form is valuable in the study of monolayers which collapse rapidly. A different automatic type of recording has been used in this laboratory in the study of the penetration of films of cetyl alcohol by cetyl sodium sulfate. The phenomenon of penetration was discovered by Schulman, but he failed to realize that at large molecular areas of the alcohol the penetration is an extremely slow process, and as a consequence many of his values are extremely far from those attained at equilibrium.

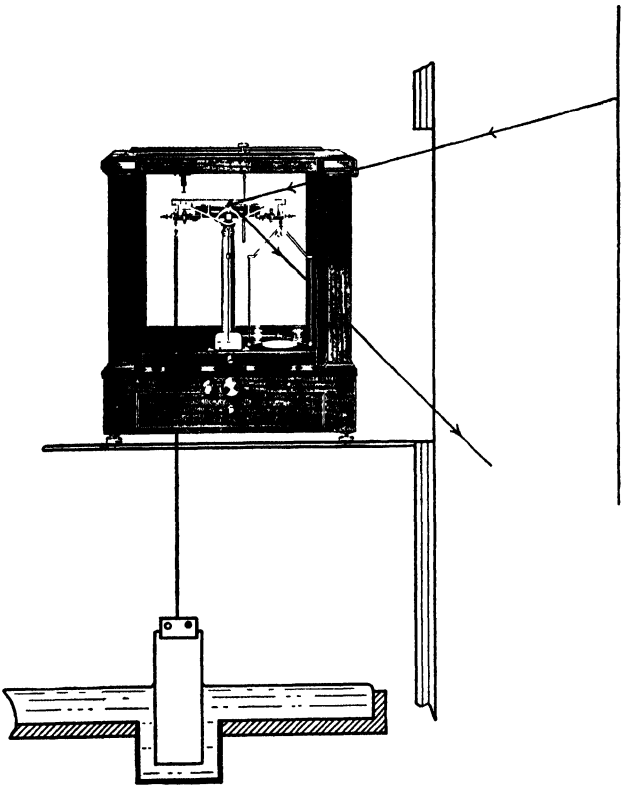


FIGURE 17. Vertical type of film balance for the measurement of the film pressure of soluble and insoluble films. (The balance is kept on the closed top of the thermostat, while the trough and film are inside.)

In order to obtain the rate at which equilibrium is attained in this slow process, Gordon and Harkins use a motion-picture camera which photographs the position of a line of light on a long, graduated scale at satisfactory intervals of time, *e.g.* every five minutes. Very accurate measurements are obtained in this way over periods of many hours. If the changes of pressure are rapid, the motion-picture camera may be operated at normal speed.

The vertical film balance exhibits the following advantages over the horizontal type commonly used:

It is much cheaper for any given accuracy.

It is much simpler and may be installed in a few minutes, since an ordinary balance and a microscope slide give results of a high degree of accuracy.

It is specially useful in investigations in which ions of the metals have a deleterious effect, since it is extremely easy to use a slide, barriers, and trough which consist wholly of quartz, so that the solution and its surface come into contact with quartz alone.

In a simple form it will give the film pressure of a soluble film, while the use of the horizontal types involves the very elaborate movable partition of the PLAWM * trough of McBain.

The perimeter p of a slide or sheet w cm wide and t cm thick, is 2 $(w + t)$, and, with a zero contact angle the downward pull of the surface on the slide is 2 $(w + t)$ or p. Let W be the apparent weight in grams of the slide partly immersed in water. If the surface tension is reduced by the spreading of a film, the slide rises until the decrease in downward pull $(-F_d)$ is balanced by the decrease of buoyancy $(-F_u)$ of the submerged portion of the slides, or

$$\Delta F_d = \Delta F_u \tag{24}$$

$$2p\Delta y = gplw\Delta h \tag{25}$$

where g is the acceleration of gravity, and Δh is the change in height of the silde measured by the use of a cathetometer. In practice it is much simpler to mount a galvanometer mirror with its center in the axis of rotation of the beam, and to observe the deflection of the beam by a telescope and illuminated scale, or to use a beam of light and observe its deflection on a scale. For a segment of a circular scale, Δh is proportional to ΔS, the change in the scale reading, or

$$-p\Delta\gamma = k\Delta S \tag{26}$$

where the value of the constant k is determined by observing the values of ΔS associated with different weights on the balance pan with the slide immersed in water which has a clean surface.

$$k = -g\Delta M/\Delta S \tag{27}$$

Thus the reduction in the surface tension, $\Delta\gamma$, which is the film pressure π, caused by spreading and compressing the film, is

$$\pi = \Delta\gamma = k\Delta S/2p \tag{28}$$

In the determination of the film pressure of a soluble substance the trough is filled with pure water, the surface is swept, and the scale reading of the beam of light is set to zero. Then either of two procedures may be followed:

(1) The solute is added and the solution stirred until the concentration is uniform throughout. The surface is swept again, and the scale reading, taken when equilibrium is attained, gives the film pressure.

(2) The pure water is removed from the trough, and replaced by exactly the same volume of solution. The equilibrium scale reading gives the film pressure, as in (1).

Surface-active solutes are often present at such great dilutions that no corrections for differences of density are needed, but with more concentrated solutions such corrections should be applied.

In a second type of vertical film balance a slide, hung from one end of the balance beam, is suspended in clean water, and another slide, from the other end of the beam, is suspended in the solution.

Horizontal Float Film Balance. In the horizontal float film balance designed by Langmuir, and improved by Adam and by others, the clean surface of the water is separated from that covered by the film by a floating barrier or "float" of paraffined metal, glass, or mica. The great improvement introduced by Adam was to prevent the film from flowing around each end of the float by an extremely thin sheet of gold

* PLAWM = Pockels, Langmuir, Adam, Wilson, McBain.

or platinum, each set edgewise through the surface and fastened to one end of the float and to a piece of metal in contact with the wall of the trough.

This type of film balance appeals greatly to the minds of those who do not understand it on account of the mistaken impression that it actually measures the magnitude of pressure ($\pi = F/L$) which acts *in the plane of the surface* against the edge

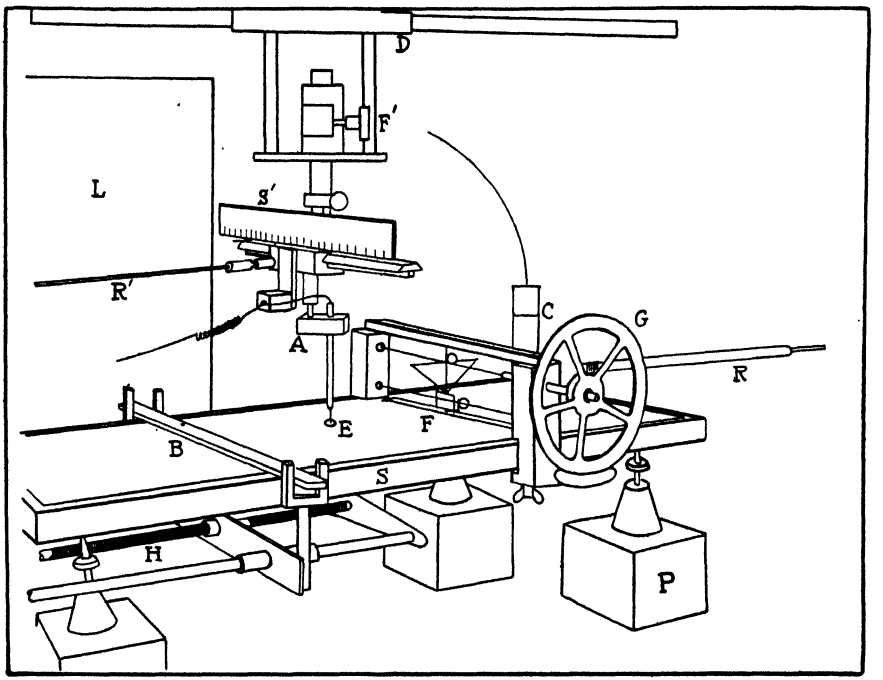


FIGURE 18. Schematic diagram of horizontal type of surface balance for researches with insoluble films.

A. Amber block supporting electrode
B. Compressing barrier
C. Calomel half-cell
D. Rack for horizontal movement of contact potential electrode
E. Silver electrode coated with polonium for measurement of surface potential
F. Movable float of torsion balance
G. Divided circle
H. Calibrated lead screw
P. Paraffin blocks
R,R' Control rods extending to outside of housing
S. Brass, stainless steel, or silica trough
S' Scale for electrometer

of the floating barrier. This is by no means true, since what it measures is the difference between two surface tensions, which meet the float *at two different angles* from the horizontal. Thus the primary assumption upon which this method was based is invalid; but a mathematical analysis of the problem, as given by Harkins and Anderson,[65] shows that, nevertheless, the method gives a correct value of $\gamma_w - \gamma_f$.

A diagram and photograph of a film balance used by the writer in 1932 are shown in Figs. 18 and 19. The electrode *E* was used to determine the surface

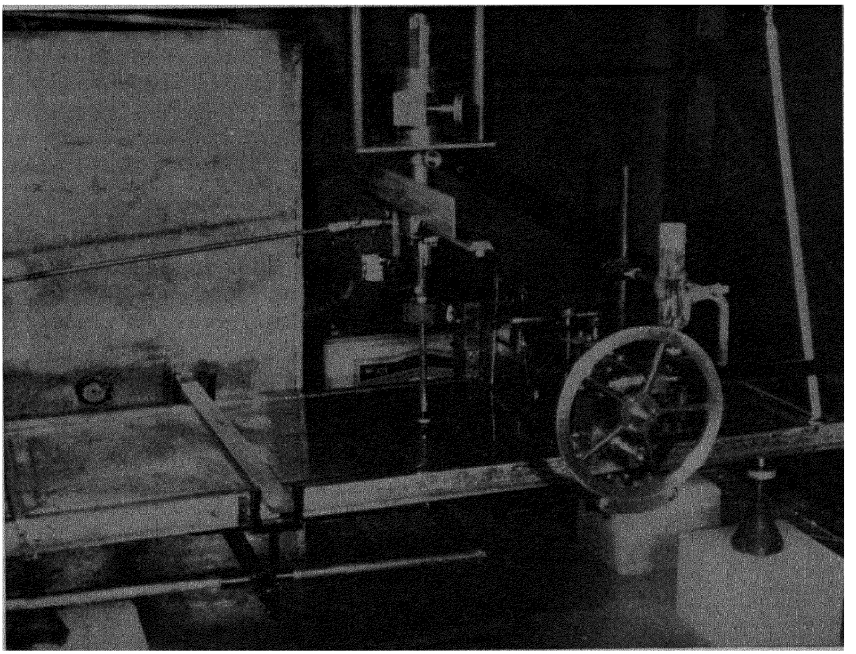


FIGURE 19. Film Balance (1932 model) with apparatus for surveying the surface with a polonium coated electrode.

potential at any part of the film, and a second electrode, not shown, served to determine the potential of the clean surface of the water. Fig. 20 is a photograph of the film balance of Harkins and Myers, which was inclosed in an air thermostat.

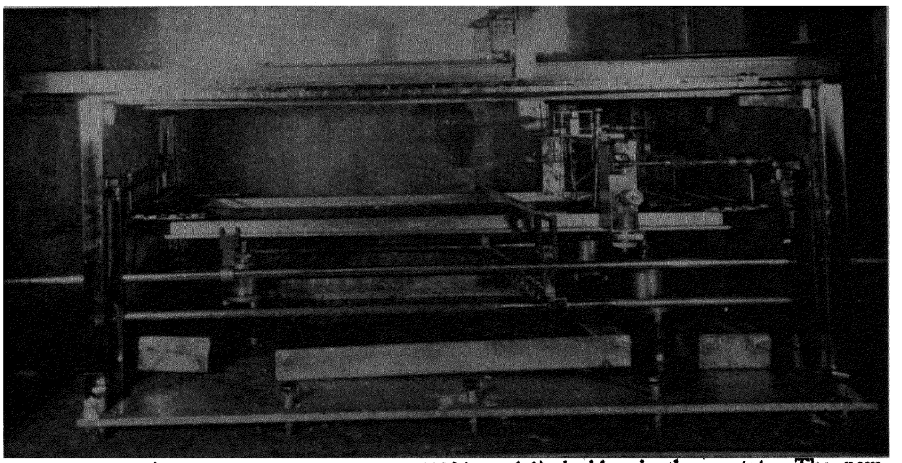


FIGURE 20. Trough and Film Balance (1934 model) inside air thermostat. The new model (1941) is buried inside a water thermostat and the sweeping is carried out by screws of 1 cm per turn which replace the rods at *D* while the barrier used to change the area of the film is moved by the fine screw shown here directly underneath the trough. The trough *E* contains a supply of clean water into which large sheets of filter paper (not shown) are dipped. This is done in order to saturate the air with water vapor.

The Thermostat (Fig. 21). Even though it is not possible to obtain accurate data on the pressure-area relations of monolayers without an accurate regulation of the temperature of the surface of the water in the trough, almost all the work on this subject has been carried out either without the use of a thermostat at all, or with only the bottom of the trough in contact with the water of a thermostat. It is not generally realized that the control of the temperature of the surface of water is very much more difficult than that of the water itself.

In the design of a thermostat for work on films it seems best to use an air thermostat which is entirely buried in the center of a liquid-filled thermostat, in order to cause the temperature of the radiation which strikes the surface of the liquid to be the same as that of the air chamber. This chamber should be saturated,

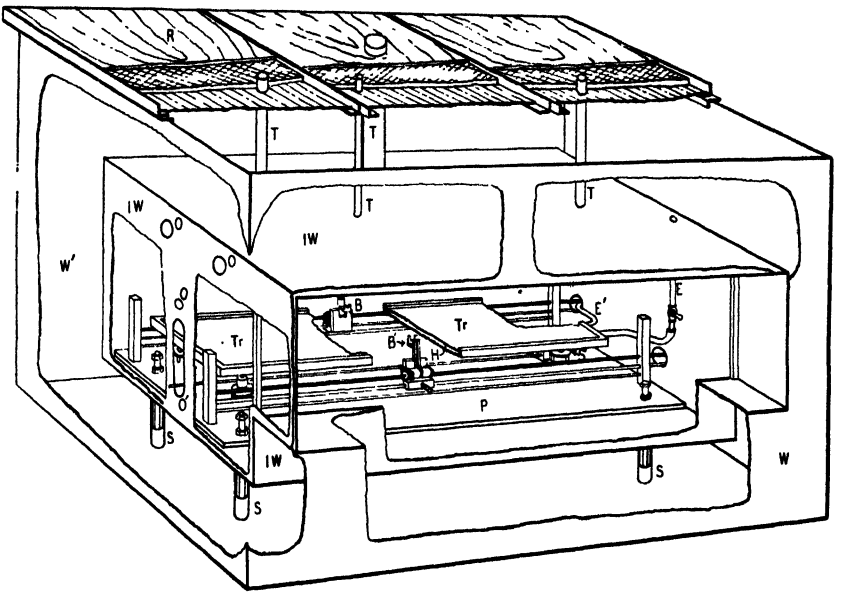


FIGURE 21. Liquid-filled thermostat, for accurate work at temperatures from −30° C to +70° C.

but not supersaturated, with water vapor. To produce this saturation water surfaces of large area, other than that of the water in the trough, should be used. Large sheets of filter paper which dip into clean water in a vessel of large area are used by the writer.

A thermostat of this type with an inner chamber 112 × 66 × 41 cm, was constructed by Harkins and Copeland (Fig. 21). This air chamber is surrounded on the bottom, top, and four sides by hollow metal walls 10 to 20 cm thick, and these are filled with an aqueous solution of ethylene glycol. The only thinner section is the glass door, which is hollow, but only 2.5 cm thick. The solution of the thermostat is pumped through this door, and also through the hollow bottom of the metal trough used for the film.

The solution in the thermostat is heated by a heating coil and cooled by a coil from a bath cooled by a refrigerating machine. This makes it possible to investigate monolayers on salt solutions at temperatures far below 0° C.

The temperature is kept constant to +0.001° C by a mercury regulator which operates a heater by a vacuum tube-relay circuit.

The film balances, one each of the horizontal and of the vertical type, are operated by rods or wires which pass through tubes between the outer and inner walls of the thermostat. The controls are elaborate and will be described elsewhere.

SURFACE OR FILM POTENTIAL (FIG. 22)

Palmitic Acid

Area						
48 Å²		68	217	315	330	
	Barrier→	35	59	141	200	←Float
		233	130	50	61	
282.2 Å²		345	315	324	315	
		297	310	311	312	
		319	314	308	305	
		331	313	313	331	

Pentadecylic Acid

48.5 Å²	155	163	170	175
	160	20	165	170
	165	24	101	170
39.7 Å²	179	183	183	180
	179	180	181	180
	175	179	180	179
	175	179	181	186

Myristic Acid

56.5 Å²	10	22	31	42	90	165	158
	0	15	30	38	164	160	156
	5	35	51	165	167	165	165
	48	98	166	174	171	176	164
	145	148	170	169	170	103	199
37.1 Å²	175	170	170	160	157		
	172	165	168	160	155		
	169	166	167	164	158		
	166	169	188	165	155		
	163	167	190	195	—		
25.0 Å²	255	248	255	281			
	255	248	255	245			
	255	252	258	243			
	254	277	261	238			
	253	254	261	256			

FIGURE 22. Surface potentials of monolayers of the long-chain paraffin acids on aqueous 0.01*N* myristic acid at 17° C, others at 22° C. The monolayers are held in place by the barrier on the left and the float on the right.

General Relations

If a small sheet of silver or some other metal is fastened parallel to and a short distance above a clean water surface, the potential between the two may be changed by as much as 0.95 volt, or approximately by the potential of an Edison cell, if a monolayer of oil spreads over the surface of the water underneath the electrode. This difference of potential may be determined by the use of a Compton electrometer with a potentiometric system, if the air under the electrode is made into an electrically conducting medium, as by a deposit of polonium on the electrode. It may also be obtained if the electrical capacity is changed rapidly, as by the use of the vibrating electrode of Zisman, who used a vacuum-tube circuit, with a frequency determined by a crystal.

If the electrode is placed at first over a clean water surface, and then over the same surface covered by a liquid condensed (L_c) monolayer of stearic acid, the change of potential is found to be close to 400 mo, and it is said that the *surface potential* is 400 m.

The earliest determination of the surface potential seems to be that of Bichat and Blondlot, while Kenrick extended the work in 1896. Frumkin and Guyot have made notable contributions to the subject.

The potential difference (ΔV) between two parallel plane *metal* plates of an ordinary parallel plate condenser, with a unit dielectric constant, is

$$\Delta V = 4\pi\sigma s \tag{29}$$

in which σ is the surface density of the charge and s is the distance between the plates.

Since $\sigma = ne$, where n is the number of electronic charges per sq cm and e is the electronic charge,

$$V = 4\pi n(es) = 4\pi n\bar{m} \tag{30}$$

in which $\bar{m}$ is the product of the value of the electronic charge by the distance between the plates.

Schulman and Rideal considered that this equation, which they attribute to Helmholtz, may be applied to calculate the dipole moment of the molecules of the film on the basis that $\mu_s = \bar{m}$ if n is considered as the number of molecules in one sq cm of film. Thus

$$\mu_s = \Delta V/4\pi n \tag{31}$$

It is found, however, that the value thus obtained is only about one-sixth, or even one-tenth that determined by the standard methods for the dipole moments of molecules which contain the same polar groups.

In 1914 the writer assumed that in monolayers of oleic or stearic acid, or other polar-nonpolar molecules, the polar groups are oriented toward the water, and it may be assumed that it is their polar groups which give a part of the contact potential. However, these polar groups, by their dipole moments, orient the water molecules underneath them. There is thus an interaction between the polar —COOH groups and the molecules of water, between the —COOH groups themselves and between the dipoles of the water molecules. Those who have applied this simple equation have not made any attempt to solve the statistical-mechanical problem involved in order to show whether or not the apparent positive charges of the —COOH dipoles lie in a single plane and the negative charges in another. Thus, at the present time the equation

$$\mu_s = \Delta V/4\pi n \tag{32}$$

has no special significance.

However, it is to be expected that on the whole the surface potential should increase with the number of dipoles per unit area of monolayer, so that in the equation

$$m = \Delta V/n \tag{33}$$

the value of m should be much more constant than that of ΔV; and since $1/4\pi$ is a constant, $\mu_\perp$ should also be very much more constant than ΔV, but this does not cause either m or $\mu_\perp$ to be a dipole moment.

In 1928 it was suggested to Frumkin by the writer that the great discrepancies in the values of the surface potential as obtained by various observers, or even a single observer, at that time, were not due to any fault of the method, but rather to the lack

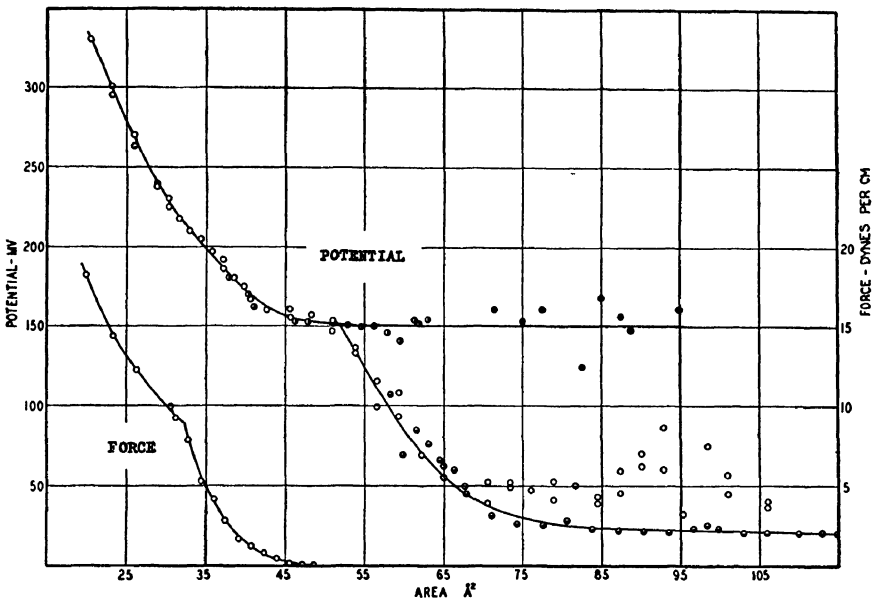


FIGURE 23. Pressure-area and potential-area of myristic acid monolayers. At areas greater than 52 Å^2 molecule^{-1} the film is heterogeneous (gaseous and liquid-expanded).

of knowledge of what phase of the film was present under the electrode. This idea led to the use of an electrode which could be moved in such a way as to survey the entire surface, as in the apparatus of Harkins and Fischer.[49]

The use of this type of electrode gave definite evidence that the region $L_\perp G$ in Fig. 6 represents a heterogeneous film, which consists of islands of liquid-expanded film in a sea of gaseous film at lower temperatures or with longer molecules; the islands may be in the liquid condensed, or in the solid state. The results of the surface potential survey of several monolayers are given in Fig. 22.

This survey of a monolayer of myristic acid at a mean molecular area of 56.5 Å^2 at 17° C gave a surface potential of zero at one position close to the movable barrier, and this value indicates that nothing but gaseous film was underneath the electrode (area *ca.* 0.7 cm^2) in this position. The most concentrated part of the monolayer was adjacent to the float and exhibited a potential of 199 mv. This is the potential of the liquid expanded phase, since at an area of 37.1 Å^2 the potential was not larger than this. At a molecular area of 25.0 Å^2, where the film is in the intermediate state, the potential was more constant but varied from 248 to 281 mv.

At 23° C the monolayer of palmitic acid at 48 Å² is partly in the gaseous state and partly in the liquid condensed state with a maximum surface potential of 330 mv. The lower potentials, as low as 35 mv, indicate that most of the area under the electrode is covered by gaseous film.

Fig. 23 exhibits the pressure-area and potential-area relations of a monolayer of myristic acid at about 15° C. At mean molecular areas above about 47 Å² the surface potential seems to have an upper limit of about 160 mv and a lower limit which decreases as the area rises. This behavior is due to the heterogeneity of the film above 47 Å², and the maximum potential possible is that of the liquid expanded

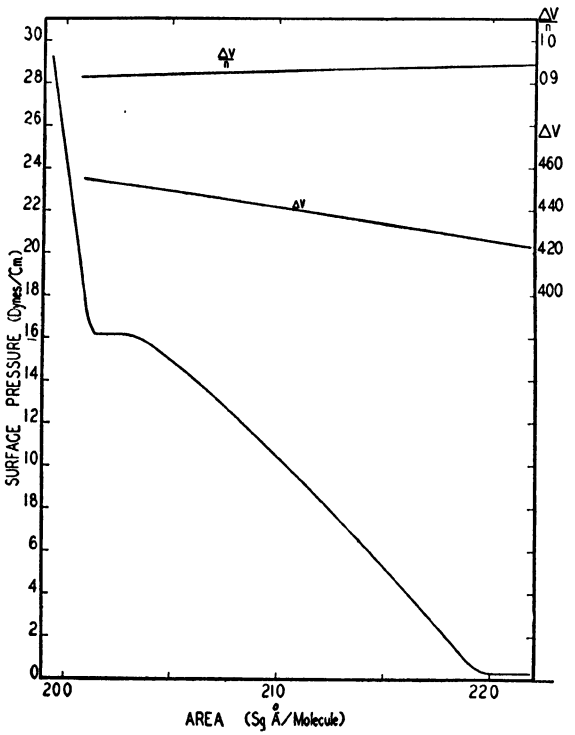


FIGURE 24. Potential (V)-area and pressure-area relations of monolayers of eicosanol (20-carbon alcohol) at 25° C.

film. The lower limit is determined by the lowest mean density of the film which can be obtained under the electrode. It is obvious that, as the mean molecular area increases, it is possible to get more and more gaseous film under the electrode, and thus obtain a smaller surface potential.

It is apparent from the figure that the curve for the variation of surface potential with area gives no indication of the second order transition between liquid expanded and intermediate film, which is so apparent in the pressure-area curve. A similar relation is exhibited in Fig. 24, which gives no indication of the first-order transition in the pressure (π)-potential (V) curve, though it is very apparent in the relation between pressure and area.

Effect of Double Bonds on the Surface Potentials of Polycyclic Compounds (*d*-Pimeric and Tetrahydro-*d*-Pimeric Acids)

The three-ring compounds, *d*-pimeric and tetrahydro-*d*-pimeric acids, are related in structure to cholesterol and to certain hormones and vitamins. In their condensed monolayers the molecules are oriented with their long axes nearly per-

pendicular to the surface of the water and the maximum thickness of the film is 12 Å, or half the length of the stearic acid molecule, while the molecular area of these compounds is, at film collapse, about 2.3 times higher than that of this acid. At a film pressure of 5 dyne cm^{-1} the molecular areas are equal (51 $Å^2$) but the compressibility of the monolayer of the unsaturated (0.0101) is much higher (Fig. 25) than that of the saturated (0.007) compound.

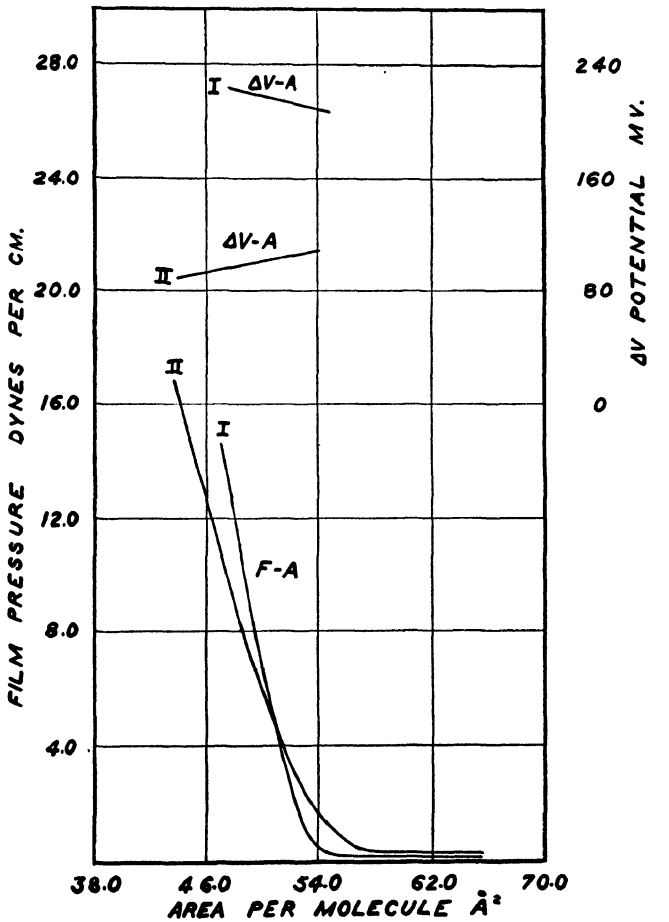
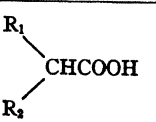


FIGURE 25. Pressure-area and potential-area relations of three ring compounds: I. *d*-pimeric acid (saturated) and II, tetrahydro-*d*-pimeric acid (see Fig. 26). Note that the double bonds of II reduce the surface potential to about 50 per cent of that of I, and that the double bonds greatly increase the compressibility.

The molecule of *d*-pimeric acid contains two double bonds which lie at the "top" of the molecule, if it is oriented as it is supposed to be in a tightly packed, condensed film. The effect of these double bonds on the surface potential is extremely interesting. Since the surface potential of unsaturated is less than half that of the saturated compound, the dipoles due to unsaturation seem to be opposed to those of the carboxyl group. Furthermore, while the surface potential of the saturated compound rises as the film becomes more tightly packed, that of the unsaturated compound

Table 6. Compressibility (κ) of Monolayers in Different States

Number of Carbon Atoms	Substance	Molecular Area (σ)	Pressure (π)	Temperature, t° C	Compressibility ($-\kappa$)
	Liquid Expanded State				
13	Tridecylic acid	26-31		12.5	0.019-0.027
	Tridecylic acid	32-41		12.5	0.026-0.06
14	Myristic acid	31-37.5		16.6	0.02-0.04
	Myristic acid	37.5-43.5		16.6	0.04-0.07
15	Pentadecylic acid	31.5-39.5		25.0	0.025-0.035
	Pentadecylic acid	39.5-44		25.0	0.035-0.045
57	Triolein	103	8-16	7	0.03
	Triolein	126	0-8	7	0.06
18	Ethyl palmitate (gas)	40	8.5	25	0.04
18	Octadecane-nitrile	35.5-45	0.25-6.5	25.8	0.03-0.06
	Intermediate (I) State				
18	Octadecane-nitrile	35.5-27	6.5-16	25.8	0.02-0.06
	Intermediate state: acids				
14	Myristic acid	35	3	9.5	0.25
15	Pentadecylic acid	40		17.9	0.35
	Intermediate state: esters				
18	Ethyl palmitate	34.7	9.3	25	0.53
19	Ethyl margarate	47	3.5	25	1.2
	Ethyl margarate	58	2.1	15.2	2.2
	Liquid condensed (L_2)				
15	Pentadecylic acid	21-23.2		27.5	0.005-0.027
16	Palmitic acid	21-23.7		25	0.008-0.01
17	Margaric acid	21.6-25		25	0.007-0.0085
18	Stearic acid	21-24		25	0.0063-0.0095
19	Nonadecanoic acid	20.4-23.6		25	0.0063-0.0075
20	Arachidic acid	20.6-24.2		25	0.0065-0.0084
	Solid films				
17	Margaric acid	20	30	25	0.0008
18	Stearic acid	20	40	22.5	0.0009
18	Octadecyl alcohol	20.3	20	20	0.0007
18	Calcium stearate	20	0.5	22.7	0.0008
19	Ethyl margarate	20.5	19.5	20	0.0020
	Molecules which lie flat (ω-hydroxydecanoic acid polymer)				
	ω-Hydroxydecanoic acid polymers		2	20	0.046-0.058
	Tertiary paraffin-chain compounds 				
	R_1 R_2				
15	1 C 14 C	30	18	18	0.013
15	7 C 8 C	62	14	18	0.020
16	8 C 8 C	55	15	18	0.017
12	6 C 6 C	58	7.5	18	0.026

three of the cases thus far investigated. Thus the slope $(\partial\kappa/\partial\sigma)_T$ of the compressibility exhibits two values (Fig. 27). For myristic acid at 16.6° C these are 0.0030

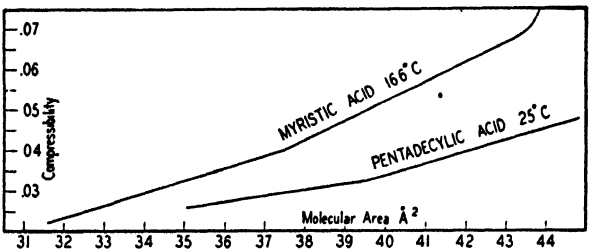


FIGURE 27. Variation of the compressibility of liquid-expanded (L_e) monolayers with molecular area.

in the low-area region (33 to 37.5 Å²) and 0.005 in the high-area region (37.5 to 43.5 Å²); and for pentadecylic acid at 25° C the values are 0.014 (35 to 39.5 Å²) and 0.003 (39.5 to 46 Å²). For tridecylic acid at 12.5° C the change of slope occurs at about 31.5 Å². In the high-area region the compressibility increases much more rapidly as the area approaches closely that at which vaporization begins.

If the above relation could be shown to be general it would indicate a change of a higher order at about the mean molecular area for the L_e film. However, for palmitic acid at 39.85° C the compressibility is linear with respect to area from $-\kappa = 0.016$ at 36 Å² to $-\kappa = 0.032$ at 44.5 Å², and thus does not exhibit the behavior described above. Therefore, it would not seem certain that this is a change of a higher order. However, an indication of a change at just this point has been observed by Dervichian [6, pp. 933-4] in the fluidity, apparent dipole moment, and compressibility of myristic acid at 38 to 39 Å², of triolein at 3×38.3 Å², and of tricaproin at about the same area. He assumes the change to be of the second order in the case of triolein, and of the third order for tricaproin.

Dervichian's general theory is "that areas corresponding to ordinary phase changes are found as points of discontinuity of higher order in those phases which exist at higher temperatures," while the observations of this laboratory indicate that these areas are not constant, but shift with the temperature or with the number of carbon atoms in the chain.

Dervichian assumes that the discontinuities at 38 to 39 Å² occur at just the area of the triple point, but we obtain the value 43 Å² as the area at the triple point Fig. 10a for the normal acids from tridecylic to palmitic, while as listed above, the apparent discontinuities occur at areas from 31.5 to 39.5 Å², and seem to exhibit an increasing value as the length of the chain increases.

In Table 4 it has been assumed, on the basis of the compressibility and the energy relations, that there is a third (possibly second)-order change at D (Fig. 6) from the liquid condensed (L_e) to the intermediate film. At this point the compressibility, on going from L_e to I, begins to rise with extreme rapidity and often exhibits the particular type of inflection characteristic of a cubic form. The behavior of the compressibility at this point is some evidence in favor of a change of a higher order at this point, as is the behavior of the energy of spreading and extension, as already been pointed out.

From the relations outlined above it is evident that the compressibility is a very important property in the determination of what phase or what change of phase is involved in any definite case. Kirkwood has developed a theory according to which, as a monolayer expands through the intermediate state, there is an increase in the

Table 7. Compressibility (κ) of Monolayers of the Normal Long Chain Alcohols in Different States

Compound	$t°$ C	A²/Molecule	Dyne/cm^{-1}	κ
		L_2 State		
Octadecanol	25.12°	22.6-20.54	1.5-14.05	0.0055-.008
Octadecanol	9.90	21.8-20.4	1.5-12.37	.004 -.016
Octadecanol	7.10	21.8-19.78	1.5-14.90	.0055-.008
Nonadecanol	21.73	22.7-20.71	1.5-15.27	.0055-.008
Nonadecanol	14.83	22.6-20.5	1.5-14.70	.0055-.008
Nonadecanol	10.35	22.3-19.97	1.5-17.40	.0047-.008
Eicosanol	24.76	22.3-20.33	1.5-16.20	.004-.009
Eicosanol	22.48	22.3-20.30	1.5-15.70	.004-.016
Eicosanol	18.34°	21.9-19.92	1.5-19.30	0.004-.009
		LS State		
Octadecanol	25.12°	20.54-col.	14.05-col.	0.0014-.0017
Octadecanol	9.90	20.2 -col.	12.37-col.	.0007-.0009
Nonadecanol	21.73	20.71-col.	15.27-col.	.0008-.0009
Nonadecanol	14.83	20.36-col.	14.7 -col.	.0005-.0009
Eicosanol	24.76°	20.33-col.	16.20-col.	.0005-.0009
Eicosanol	22.48°	20.24-col.	15.70-col.	0.0008-.0011
		S State		
Octadecanol	7.10°	19.78-col.	14.90-col.	0.0006-.0008
Nonadecanol	10.35	19.97-col.	17.40-col.	.0007-.0009
Eicosanol	18.34°	19.92-col.	19.30-col.	0.0007-.001

Col. = Collapse.

number of molecules which have free rotation around their long axis. From this theory it has been possible to calculate the change in compressibility which occurs in the second order transition.

$$\text{Intermediate} \rightleftharpoons \text{Expanded}$$

with good agreement between the calculated and the experimental values.

Monolayers Whose Molecules Lie Flat on the Surface of Water

In 1917 Harkins, Davies, and Clark [6] assumed that in the surfaces of pure liquids or in condensed monolayers on water the molecules are oriented perpendicular to the surface, provided they are long-chain paraffins with a polar group or groups at one end of the molecule. However, they also assumed that a proper distribution of such groups along the chain, or a symmetrical distribution in a ring compound, would give a monolayer in which the molecules lie flat. In order to test the latter hypothesis, long-chain linear ω-hydroxydecanoic acid polymers, with chains from 60 to 1970 Å (0.2μ) long were obtained from W. H. Carothers. Since in these chains there is one polar group at every tenth carbon atom, the molecules should lie flat on the surface and thus give a tightly packed monolayer about 4.5 Å thick, with about 4.5 Å distance (d) between the centers of the chains. The experiments showed that at a pressure of 3 dynes per cm^2 the apparent thickness of the films varies from 4.0 to 5.0 Å, while the "width" of the molecule, or the molecular spacing, varies from 4.0 to 4.8 Å. The product of the averages of these linear dimensions is about 20 Å^2, which may be considered as the area of the cross-section of the molecule perpendicular to its length. Such relationships together with those expressed in Tables 8 and 9 show that these molecules lie flat on the surface.

The pressure-area and potential area relations of these monolayers are shown in Fig. 28. The surface potentials of these extremely thin films are found to be high (*ca* 400 mv).

Table 8. Properties of Films with Different Orientations of the Molecule

Long Axis Vertical	Long Axis Horizontal
Molecular area independent of length	Molecular area increases with length
Withstand high pressures (to the order of 15 to 50 dynes per cm)	Collapse at low pressures (2 to 10 dynes)
Surface potential increases with length of molecules up to 18 carbon atoms	Surface potentials of the same order for long and short molecules
Compressibility decreases with increasing length of chain	Compressibility high and independent of length of molecules

Table 9. Films of Polymers of ω-Hydroxydecanoic Acid on Water

Mean values at temperatures from 21.5 to 26.0°. Concentration of hydrogen chloride in the water 0.01 *M*. Columns 2 and 3 are taken mostly from the paper of Carothers and van Natta

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Mol. Wt.	Density d^{25}_4	Length of Molecule (Å)	Vol. of Molecule (cu. Å)	Area per Molecule at a Film Pressure $f = 0$	$f = 3$	$f =$ max.	Maximum Force f_m	t = Thickness of Film at a Film Pressure $f = 0$	$f = 3.0$	$f =$ max.	d = Width of Molecular Space at $f = 0$	$f = 3.0$	$f =$ max.	$t \times d$ Molecular Cross Sect. at 3 Dynes (sq. Å)	Compressibility at $f = 0$
780	1.0957	60	1170	330	290	246	5.7	3.6	4.0	4.8	5.5	4.8	4.1	19.5	0.040
1715	1.0935	133	2590	700	604	556	4.5	3.7	4.3	4.7	5.3	4.5	4.2	19.4	.046
3190	1.0877	248	4840	1260	1080	970	4.2	3.9	4.5	5.0	5.1	4.4	3.9	19.6	.048
4170	1.0814	324	6360	1490	1300	1130	4.5	4.3	4.9	5.6	4.6	4.0	3.5	19.6	.043
5670	1.0751	440	8700	2050	1800	1688	4.4	4.2	4.8	5.2	4.7	4.1	3.8	19.7	.041
7330	1.0715	570	11330	2670	2277	2180	3.8	4.2	5.0	5.2	4.7	4.0	3.8	19.9	.049
9330	1.0668	730	14450	3800	3200	3000	4.0	3.8	4.5	4.8	5.2	4.4	4.1	19.8	.053
16900	1.0627	1320	26200	6440	5510	5270	3.4	4.1	4.8	5.0	4.9	4.2	4.0	19.9	.048
25200	1.0621	1970	39100	10430	8800	8700	3.2	3.8	4.4	4.5	5.3	4.5	4.4	19.9	.052

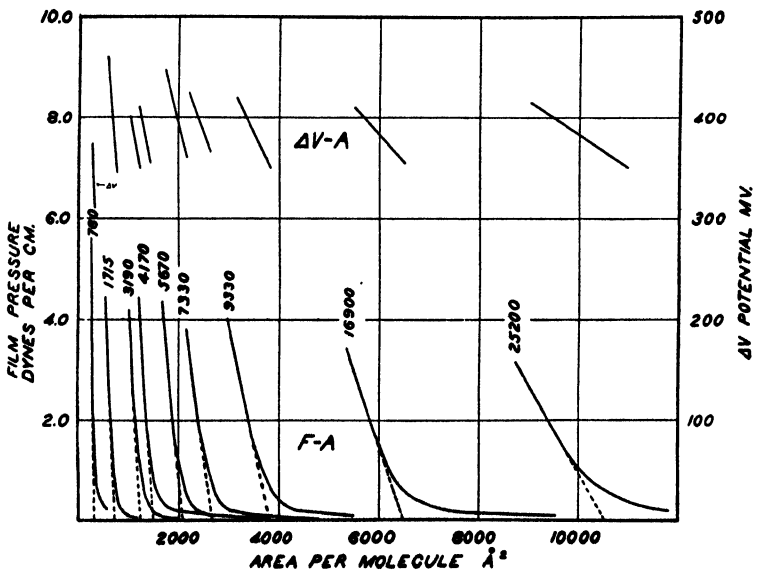


FIGURE 28. Pressure-area and potential area of the polymers of ω-hydroxy-decanoic acid. These molecules lie flat on the surface of the water.

THE VISCOSITY OF MONOLAYERS

Viscosity of Monolayers and Canal Viscosimeters

In the flow of a monolayer of oil under a difference of pressure, the first molecular layer of water under it is carried along with the film, and this in turn affects

the other adjacent layers underneath. Thus the flow of the film is retarded by the viscosity of the water. If the viscosity of the film itself is desired, as in the flow of the film through a narrow and deep canal, a correction term enters, and this involves the first power of η_0, the viscosity of the water. If the film spreads with sufficient rapidity, and excessive convection can be avoided, the water level can be seen to be momentarily lower at the place where a hexane solution of a polar-nonpolar oil was put on the surface than near the outer portion of the spreading film.

This drag on the water is indicated in the correction term in the equation of Harkins, Myers, and Kirkwood [67] for the viscosity (η) of a monolayer as determined by the flow of the film through a deep canal whose length (l) is large in comparison with its width (a):

$$\eta = \alpha a^3/12A - a\eta_{0/\pi}$$

Here η_0 is the viscosity of the subphase and α is the pressure gradient $(\pi_2 - \pi_1)/_l$ of the film, and A is the area flux. It may be noted that the law of Poiseuille is here modified since, as the flowing body loses one dimension, the power of a is reduced by unity from four to three.

The last or correction term is large in comparison with the viscosity of an expanded liquid film unless the canal is very narrow. On water at 20° C the corrections in surface poises are for canals of different widths:

1 mm = 0.00032
0.5 mm = 0.00016
0.1 mm = 0.000032

General Viscosity Relations [69, 81]

The viscosity of a monolayer is highly dependent both upon the closeness of the molecular packing in the film, and upon the structure of the monolayer otherwise. No measurement of the viscosity of a gaseous film has been made. Liquid-expanded monolayers have higher, but still low, viscosities; those of liquid-condensed films are high, and of S films are anomalous, but often even higher.

Normal Long-Chain Paraffin Acids. The viscosities of monolayers of these acids exhibit remarkable but normal relations, as exhibited in Fig. 29. At the lower pressures the monolayers are in the liquid-condensed state, but at the high pressures they are in the S state. The normal behavior is that of the acids of 18, 19, and 20 carbon atoms per molecule, since the films of the shorter-chain acids are not sufficiently removed from the *intermediate* state to have wholly the properties of the liquid-condensed state.

Use of Viscosity Measurements in the Detection of Phase Transitions. The second-order transition from the condensed liquid to the S state, in the case of these acids, occurs at what is the intersection of two straight lines in the pressure-area (π, σ) diagram. Thus, in approaching the transition from either side of the pressure-area diagram, there is no hint of its existence until the point itself is reached. In the case of the viscosity, however, it is apparent from Fig. 29 that the linear relation between log η and begins to disappear at a pressure several dynes away from that of the transition.

The most important viscosity relations of these acids on an aqueous subphase of pH 2 are given below:

(1) From the lowest pressures investigated up to 19 dynes per cm or more, the logarithm of the surface viscosity is proportional to the film pressure, or

$$\log \eta = \log \eta_0 + k\pi \qquad (33a)$$

for all the acids from 16 to 20 carbon atoms.

(2) The viscosity of the liquid films increases rapidly with the length of the hydrocarbon chain.

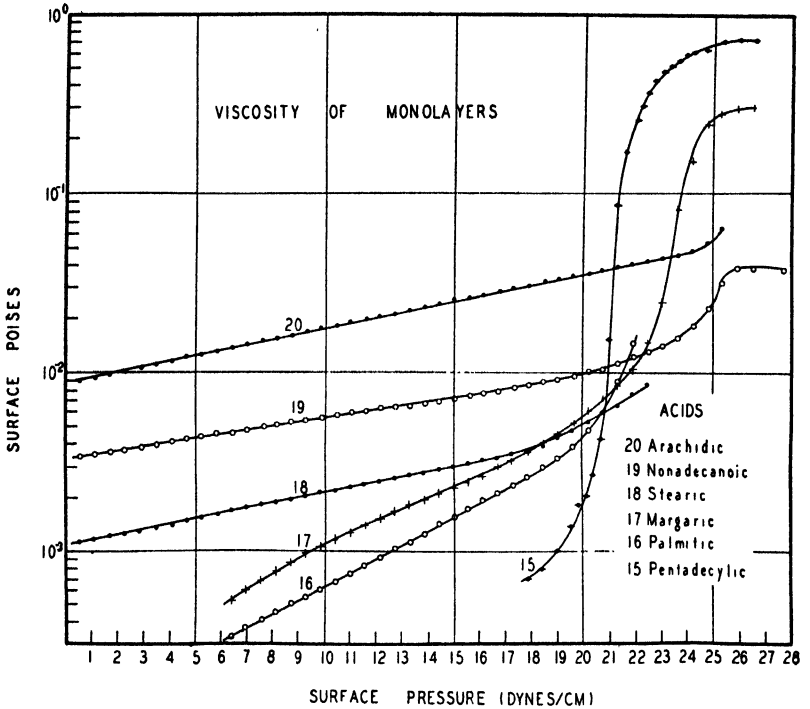


FIGURE 29. Viscosity-pressure relations of monolayers of the normal long-paraffin chain fatty acids.

(3) The liquid or low pressure-condensed films exhibit Newtonian viscosity.

(4) The viscosity of the "plastic" (or high pressure-condensed) films decreases with the length of the hydrocarbon chain.

(5) Films designated as "plastic" exhibit *non-Newtonian* viscosity, that is, *the viscosity varies with the rate of shear.*

(6) The film viscosity begins to increase more rapidly than corresponds to eq. (33a) at a pressure considerably below that of the sharp kink in the film-pressure-area curve. Also, the rise in viscosity between the low pressure-condensed film is larger, the shorter the hydrocarbon chain of the fatty acid.

(7) Above the transition pressure the surface viscosity increases less rapidly until it approaches some degree of constancy.

(8) The viscosities of the monolayers of the acids are very much less than those of the corresponding alcohols. Thus, at four dynes per cm pressure, with 16 carbon atoms in the molecule, the alcohol film is of the order of 33 times more viscous than that of the acid; with 17 carbon atoms the factor is the same; with 18 carbon atoms it falls to 17 times that of the acid.

Viscosity Relations of the n-Long Alcohols of from 14 to 20 Carbon Atoms per Molecule

An investigation of the monolayers of the *n*-long chain alcohols by Fourt and Harkins [69] at 25° C show that at this temperature the relations are very much the same as those exhibited by the acids, except that, as mentioned above, the viscosities are from 17 to 33 times higher than those of acids of the same chain length.

It was found that the viscosities of the alcohols of from 14 to 18 carbon atoms,

when calculated as bulk viscosities, are from ten thousand to a hundred million times larger than the viscosity of water at the same temperature.

The normal viscosity relations exhibited by these alcohols at 25° C (Fig. 30) become anomalous and remarkable when the temperature range of the experiments is extended to that between 5.0 and 50° C.

Abnormal Variation of the Viscosity of Liquid-Condensed Phase of an Alcohol with Temperature and Pressure (Figs. 31, 32, and 33 and Table 10.)[96]

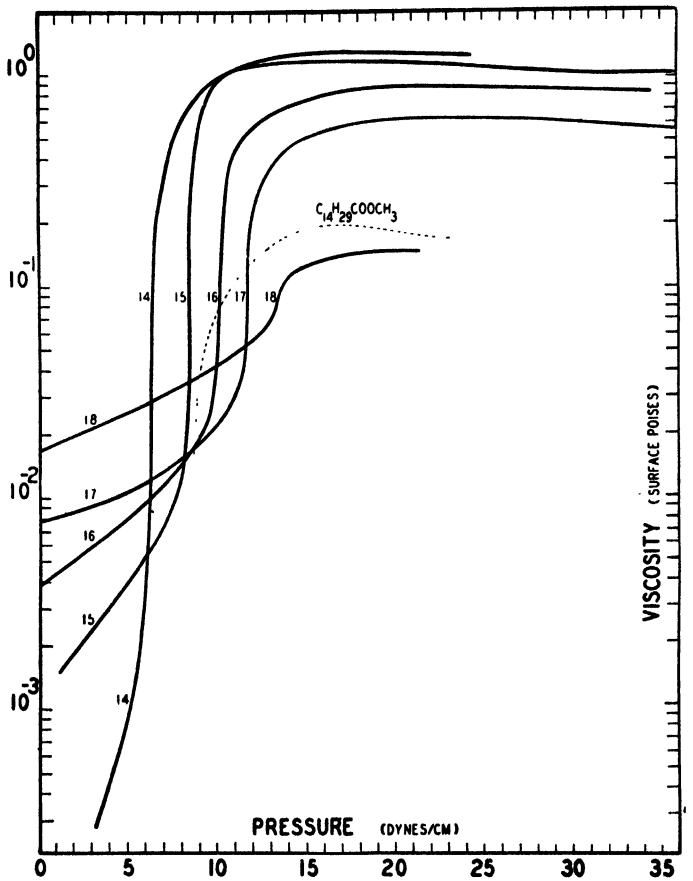


Figure 30. Viscosity-pressure relations of the normal chain alcohols at 25° C.

Effect of Temperature

At very low film pressures the viscosity of the liquid condensed (L_c) monolayer of a normal long-chain paraffin alcohol decreases with increasing temperature, as is normal. Thus the value is about seven times lower at 25° C than at 5° C. However, in Fig. 31 the curves for this phase are seen to cross each other at higher pressures, with the result that at the highest pressures at which this phase exists the viscosity in surface poises decreases from a little over 0.1 at 25° C to about 0.0045 at 8.85° C, or to about a twenty-fifth of its value in a *decrease* of temperature of 16° C. This is extremely remarkable, since in general the viscosity of a phase *rises* rapidly as the temperature is decreased.

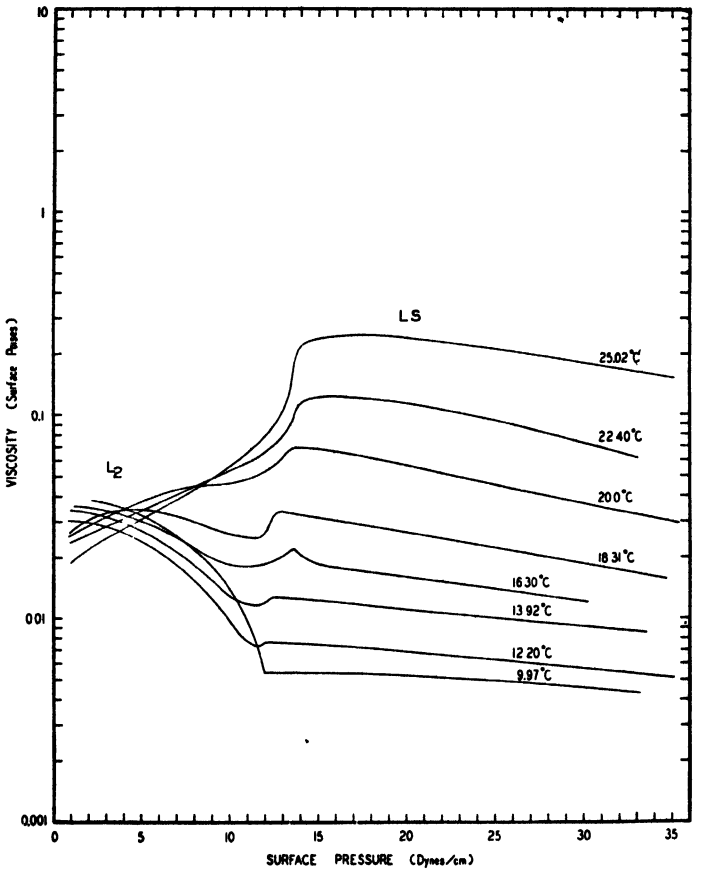


FIGURE 31.

Multiply pressures in Figure 31 and Table 10 by 0.931.

Table 10. Surface Viscosity of Octadecanol [a] (in Surface Poises)

	Pressure (dynes/cm)			
t° C	1	10	18	25
5.1°	0.150	0.352	2.65	—
7.2	.0700	.0620	0.0890	0.490
8.1	.0673	.0392	.00555	.126
8.4	.0560	.0250	.00450	.0655
8.8	.0370	.0196	.00440	.00515
10.0	.0390	.0136	.00530	.00493
12.2	.0302	.00955	.00710	.00620
13.9	.0340	.0127	.0115	.0102
16.3	.0357	.0187	.0170	.0140
18.3	.0270	.0260	.0292	.0207
20.0	.0260	.0457	.0612	.0452
22.4	.0238	.0532	.1020	.0935
25.0°	0.0192	0.0570	0.249	0.206

[a] Note that at a film pressure of one dyne per cm the viscosity decreases with increasing temperature, as is normal, from 0.150 poise at 5.1° to 0.0192 at 25° C. At 10 dynes per cm the viscosity decreases (normal) from 0.352 at 5.1° to 0.0096 poise at 12.2° C, but increases (abnormal) from this temperature up to 0.0570 at 25° C. In all of the above changes the monolayer is in the liquid

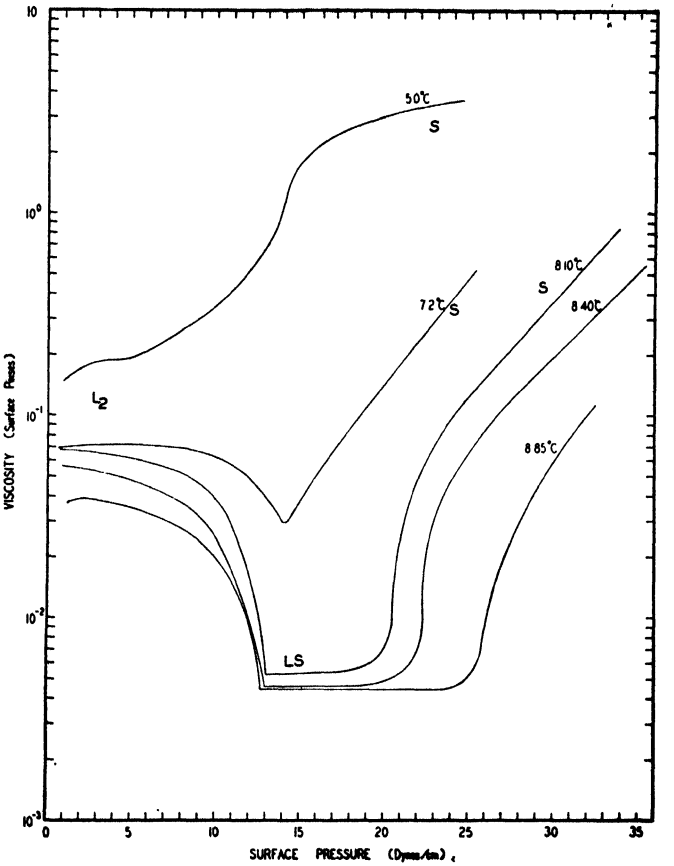


FIGURE 32. Viscosity relations of the liquid condensed (L_c), and the LS end phases of octadecanol between 5.0C and 8.85° C. Multiply all pressures for the alcohols by 0.931.

As the temperature is decreased still further (Fig. 32), the viscosity of the liquid condensed (L_c) phase, at the highest pressures at which it exists, increases from about 0.045 at 8.85° C to about one surface poise, or the increase is by a factor of about 200 for a temperature increment of 4° C.

Effect of Pressure

At both 5 and 25° C the effect of pressure upon the viscosity of the L_c film of octadecanol may be described as normal. As the temperature rises from the lower or falls from the higher value, the effect becomes increasingly abnormal as 9° C is approached. At 9.97° C the viscosity decreases with increasing rapidity, from about 0.04 at low pressure to 0.005 surface poise at a pressure of 12 dynes cm^{-1}. At 8.10° C the decrease is from 0.07 to slightly below 0.006 surface poise by an increase

condensed (L_2) state. At 18 dyne cm^{-1} the viscosity in the S state decreases with extreme rapidity with increase of temperature, from 2.65 poises at 5.1° to 0.0890 poise at 7.2° and still lower. Then a transition to the LS state occurs, after which the viscosity falls much more slowly to a minimum of not far from 0.0044 poise at about 8.8° C. Then the direction of the change changes and the viscosity of the LS phase increases rapidly with increase of temperature (abnormal) until it becomes 0.249 at 25° C.

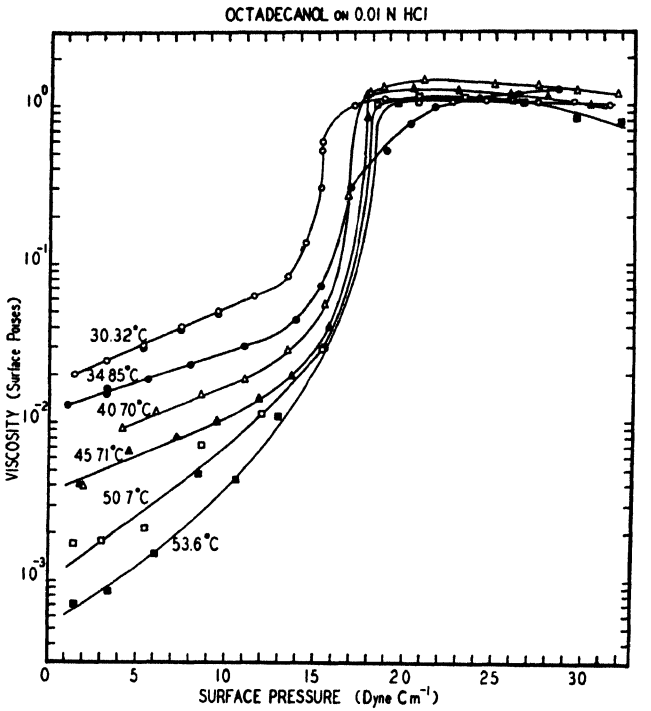


FIGURE 33. Viscosity relations of the liquid condensed and *LS* phases of octadecanol between 30° and 54° C. The extremely abnormal relations of Figures 31 and 32 have disappeared, since the viscosity again falls with rise of temperature. It is remerkable that identical values of viscosity and pressure occur at three different temperatures. Multiply all pressures for alcohols by 0.931.

of pressure to 13 dynes cm^{-1}. This behavior is extremely abnormal since the viscosity is reduced to an eighth or tenth of its value by an increase of pressure which should give threefold or fourfold increase.

A SUPERLIQUID PHASE WITH EXTREMELY ABNORMAL VISCOSITY RELATIONS

Figs. 31, 32 and 33 indicate the existence of a new phase which exhibits extremely remarkable viscosity relations as follows:

(1) Although the viscosity is Newtonian, it is essentially independent of pressure. In this constancy of viscosity with increase of pressure, it resembles solid monolayers, which, however, have non-Newtonian viscosity.

(2) Although the phase exists in the same range of high pressures as the S (usually designated as solid) phase, its viscosity at temperatures of 8 to 12° C is much lower than that exhibited at low pressures by the liquid (L_c) phase from which it is formed by increase of pressure.

(3) The viscosity of this phase is lowest at those temperatures at which it is formed by a *first-order* transition from the liquid-condensed (L_c) phase, by increase of pressure.

(4) As the temperature is increased in the range of the second-order transition $L_c \rightleftharpoons LS$, the viscosity rises very rapidly and takes on more and more the characteristics commonly exhibited by the *S* phase, although there seems to be no indication of any change of phase. At 25° C the viscosity of the *LS* phase has become non-Newtonian, high (*ca.* 0.25 surface poise), and decreases slightly with increase of

pressure. Since the viscosity is non-Newtonian, the value cited refers to the rate of shear employed in this work. A Couette type of apparatus has been designed and will be used later to determine the relation between viscosity and rate of shear.

The Transition $L_e \rightleftharpoons LS$ and $LS \rightleftharpoons S$ as Revealed by Abrupt Changes in Viscosity

Between 8 and 10° C (Figs. 31 and 32) the slope $(\partial \log \eta/\partial \pi)$ takes on an extremely large negative value as the pressure approaches about 13 dynes cm^{-1}, but at this pressure changes abruptly to zero. The pressures and temperatures of these abrupt changes correspond to phase transitions from the condensed liquid to a high pressure phase as revealed by the presure-area (π, σ) diagram of Fig. 15. At 8.10, 8.40, 8.85, and 9.97° C the viscosity remains constant for increases of pressure of 4, 5.5, 11, and more than 22 dynes cm^{-1}, respectively.

At all of these temperatures except the highest, the viscosity begins to increase rapidly at the end of these intervals, and at 8.10 and 8.40° C the slope of the curve takes on an almost infinite *positive* value, after which it decreases, but remains high.

The most obvious explanation of this extremely high slope is that it corresponds to a phase transition, which would seem to be from the LS to the S phase. This assumption is the basis for the inclusion in Figs. 15 and 16 of a boundary between these two phases, which, as obtained from the viscosity relations, is found to be linear, with a slope which indicates a rise in the transition pressure of 6.4 dynes cm^{-1} per ° C.

The evidence for the existence of the transition from the LS to the S phase is not that of the viscosity alone, since the effect of the transition is apparent in the entropy and energy relations. These exhibit discontinuities at this transition, and also at that between the L_e and S phases. At a molecular area of 19.98 A^2 the $L_e \rightleftharpoons S$ transition occurs at a pressure of 17.61 dynes cm^{-1}, and the $LS \rightleftharpoons S$ transition at 23.73 dynes cm^{-1}. These pressures agree with those obtained from the viscosity relations.

The alcohols of 18, 19 and 20 carbon atoms per molecule give similar pressure-area diagrams, but the temperature of the first-order transition rises about 5° C for each additional carbon atom.

Discussion

The molecular area at which the transition $LS \rightleftharpoons S$ occurs in octadecanol seems to be almost independent of pressure and temperature, and what little work has already been done with the 20-carbon alcohol seems to indicate that the area for its similar transition is almost the same. Also, the range in area of the first-order transition seems to be affected only slightly by an increase in the number of carbon atoms, while both the pressure and the temperature are increased materially. These facts seem to suggest that the mechanism involved is extremely sensitive to the distance between the molecules in the monolayer. Furthermore, since similar abnormalities in pure three-dimensional liquids involve only minor changes of viscosity, the enormous effects found here would seem to be associated with the presence of the aqueous subphase. This can effect the relation in two ways:

(1) By orientation of the molecules, the hydroxyl groups (also the hydrocarbon chains) of the alcohols are brought into proximity.

(2) The hydroxyl groups are brought into contact with molecules of water.

By (1) there could be some type of bonding between the hydroxyl groups of the alcohols, or by (2) between these groups and the water.

To explain the results, it is necessary to assume a type of bonding which is extremely sensitive to changes of intermolecular distance. In either (1) or (2) hydrogen bonding is likely to be involved. The transition $S \rightleftharpoons LS$ takes place at a molecular area not far from 19.98 $Å^2$ at any pressure at which both phases are stable.

This corresponds to a temperature close to 8.6° C. The curve nearest to this temperature in Fig. 32 is that at 7.2° C, at which temperature the viscosity of the L_e phase is almost constant, with increase of pressure up to 9 dynes cm^{-1} and then decreases more and more rapidly until at 14 dynes cm^{-1} the monolayer changes directly into the S phase. At the transition point the viscosity is slightly less than 0.03 surface poise, but that of the S phase rises rapidly with pressure, until at 25.5 dynes cm^{-1} it is 0.5 surface poise.

At 8.85° C and 16 dynes cm^{-1} the molecular area is 20.0 $Å^2$ and the viscosity, which is that of the LS phase, is 0.0044 surface poise. If at this pressure the molecular area of this phase is increased to 20.75 $Å^2$ by increase of temperature, the viscosity increases by a factor of 55 to 0.2400. *Thus the binding is increased greatly by an increase of molecular area and temperature, both of which should normally decrease it very greatly.*

The extremely high fluidity of the LS film is exhibited only when the molecules are tightly packed and when the temperature is relatively low. Increase of molecular distance is accompanied by an extremely rapid rise of viscosity. Thus at a pressure of 16 dynes cm^{-1} the viscosity is increased 55 times by *increase* of 0.07 Å in mean molecular distance.

While the viscosity of the condensed liquid (L_e) phase of the alcohols decreases at low pressures with the temperature, which is normal, this is reversed at high pressures. At higher and lower temperatures, increase of pressure gives the normal increase of viscosity with pressure, but at intermediate temperatures the viscosity decreases more and more rapidly as the pressure increases, until finally a phase transformation occurs.

Non-Equilibrium Relations Between a Slightly Soluble Monolayer and Its Aqueous Subphase as Exhibited by Film Pressure (π) and Surface Potential (V)[60]

If the monolayer of a slightly soluble monolayer is compressed rapidly, the film pressure is increased and the surface potential is changed. The sign of both these changes is reversed if the film is more or less completely removed by sweeping. If the monolayer consists of lauric acid the return to equilibrium at 25° C involves a period of time of the order of a half hour. However, Fig. 34 indicates that the film pressure (F) returns (increases) to its equilibrium value much more slowly than the surface potential. It should be noted that sweeping reduces the film potential to that of the clean surface of water.

Effect of Bivalent Positive Ions on Monolayers of the n-Long Chain Paraffin Acids and of Soaps: Biological Effects

The remarkable effects of sodium hydroxide, sodium chloride, and calcium chloride upon an interfacial film of oleic acid were discovered by Clowes,* and were found to be even more remarkable when investigated quantitatively by Harkins and Zollman.[29] The interfacial tension between an aqueous solution and a 0.001 molal solution of oleic acid in olive oil was found to be only 0.002 dyne cm^{-1}, when the aqueous phase contained 0.001 m NaOH and 0.15 m NaCl. This is the lowest interfacial tension between water and an oil thus far measured. (See Vol. I., pp. 243-4 of this series). However, if this same aqueous solution was made 0.0015 molal with respect to $CaCl_2$, the interfacial tension rose by a factor of about 3500 to 6.88 dyne cm^{-1}, an extraordinary increase.

In his early work on films of stearic and similar acids, Adam found anomalous relations, in that below 17 dyne cm^{-1} he obtained very different pressure-area relations with the three following aqueous subphases: (1) dilute HCl, (2) fresh

* Clowes, G. H. A., *J. Phys. Chem.*, 20, 408 (1916).

distilled water, and (3) old distilled water. On dilute HCl he found the ordinary "solid" phase at pressures above 17.5 dyne cm^{-1}, and the condensed liquid phase below this pressure. On *old* distilled water the "solid" phase alone was obtained at pressures above 2.5 dyne cm^{-1}, whereas on *fresh* distilled water the results were

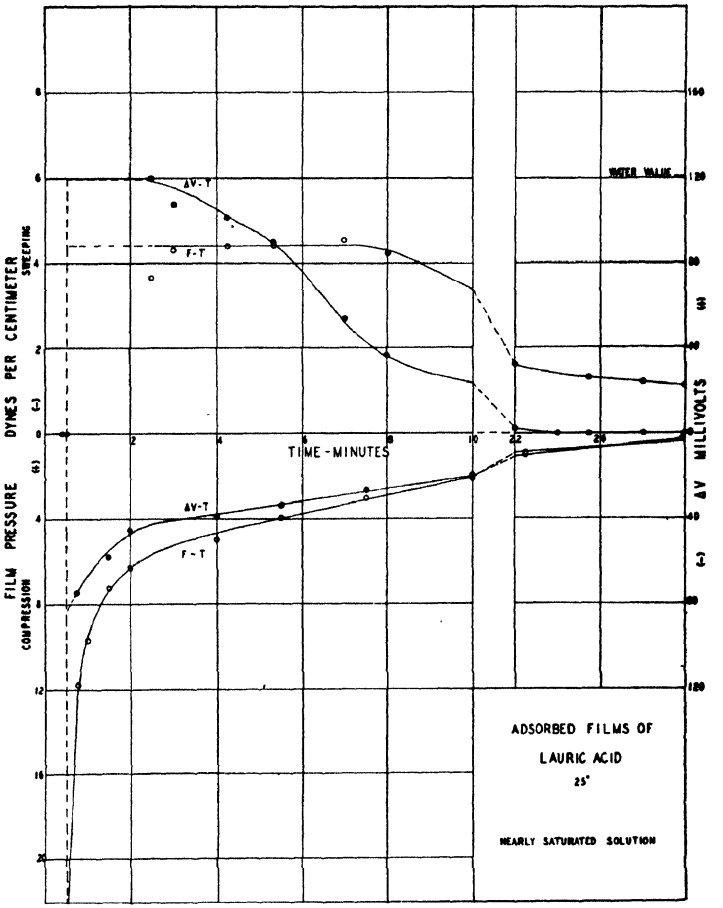


FIGURE 34. Effect of time on the film pressure and surface potential of solutions of lauric acid.

the same as on HCl, except that the condensed liquid phase was found to contract rapidly.

This behavior was explained by Harkins and Myers [60] as due to metallic ions, in this case principally calcium. On HCl the film consists of stearic acid, but in old distilled water there is enough Ca^{++} ion to convert this largely into calcium stearate, which gives a condensed *solid* film at all pressures except the lowest. The fresh distilled water evidently contained a minute concentration of metallic ions (largely Ca^{++}) but not sufficient to give a solid film immediately. However, such ions gradually diffused into the monolayer and thus gradually converted a part of the stearic acid into calcium stearate. Since at 25° C the extrapolated area for the liquid monolayer is 24.40 $Å^2$ (24.38 by Harkins and Copeland, and 24.41 by Nutting

and Harkins) and for the solid monolayer is 20.3 Å^2, the area per hydrocarbon chain is 16.8 per cent less for the calcium salt than from the acid. (Fig. 35a).

These last results are taken from the unpublished work of Harkins, Copeland, and Adinoff. The values for the calcium stearate film were obtained on a subphase of $NaHCO_3$ solution with a pH of 8.0. Above the transition pressure of 25.5 dyne

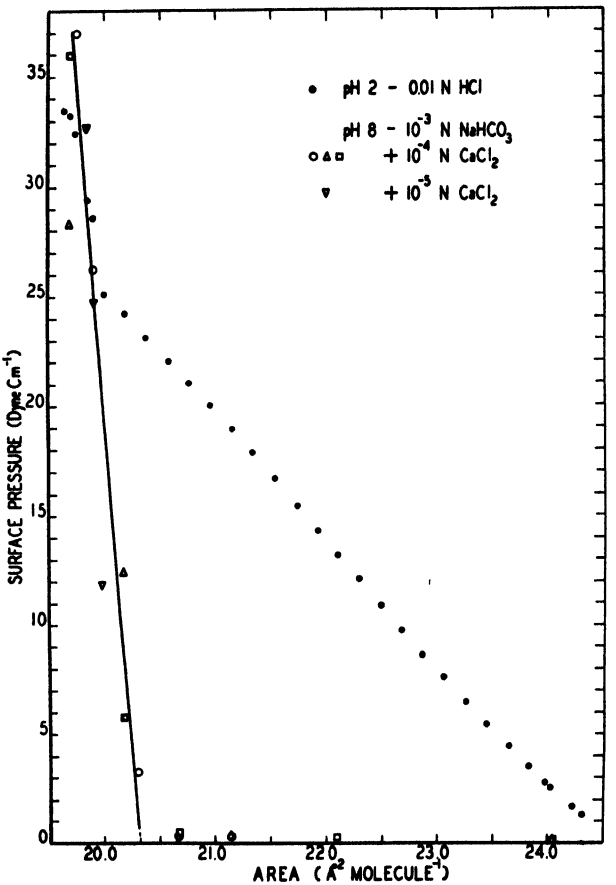


FIGURE 35a. Ca^{++} ions in a basic subphase transform the liquid condensed (L_c) film (dots) into a solid film (solid line).

cm^{-1} the acid and its calcium salt give practically the same area per chain at the same pressure. Thus the Ca^{++} ions do not seem to pack the chains more or less tightly, but at pressures below 25.5 dyne cm^{-1} *they transform the less condensed liquid phase into the more condensed solid phase.* At a pH of 7.6 with 10^{-4} molar Ca^{++} ions, remarkable time effects were observed. (Figs. 35b and c).

According to *one* theory developed by Harkins the change of phase produced by Ca^{++} ions may be attributed to their coördination number of 8. It is assumed that each Ca^{++} ion in the film is attached to four —COO— ions of the long-chain acid, and to four water molecules of the subphase. Since each —COO— ion is attached to two different Ca^{++} ions, this gives the proper ratio of one Ca^{++} ion to two —COO— ions. Another point of view also advanced by the writer is that the calcium stearate

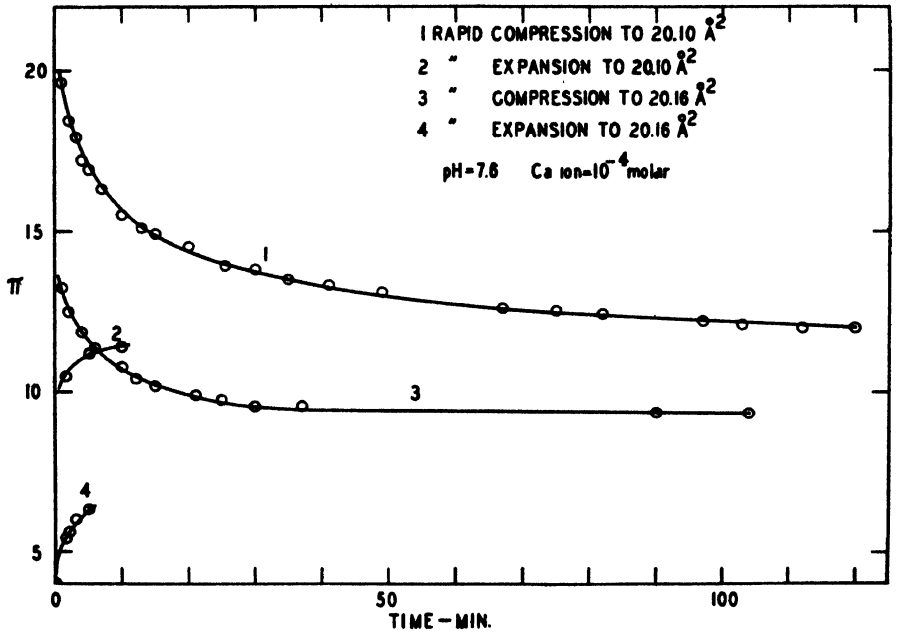


FIGURE 35b. Film pressure vs time.

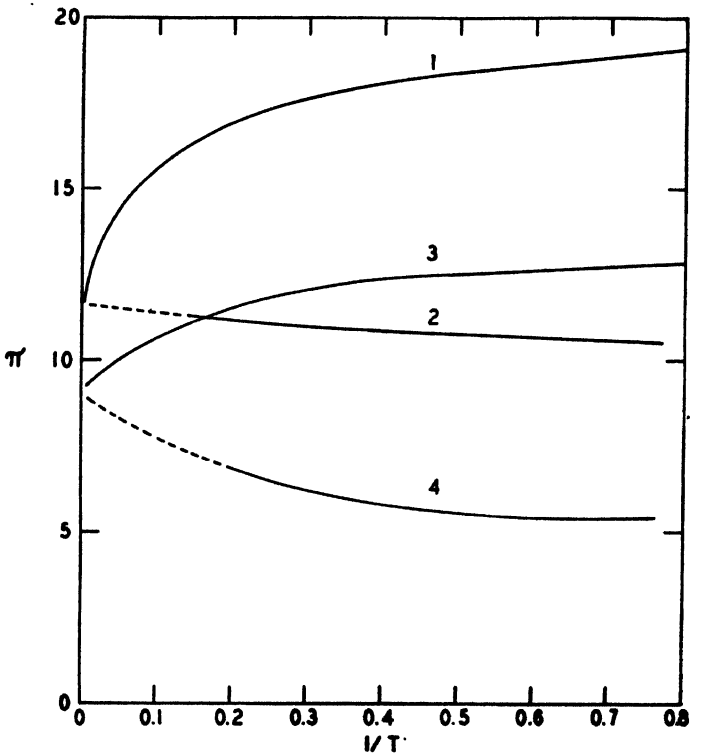


FIGURE 35c. Film pressure vs reciprocal of the time.

is insoluble and forms a solid monomolecular sheet in which two stearate ions are combined with one calcium ion.

Monolayers of this type exhibit considerable rigidity and may be classed as truly solid.

According to Langmuir and Schaefer,* the monolayer on a subphase of 10^{-4} molal calcium carbonate is the nearly neutral soap at pH 11, but at pH 5.1 is one-half neutral soap, while it consists entirely of the free fatty soap at pH 3, which is quite far on the acid side of the neutral point.

Mitchell, Rideal, and Schulman find that in the photochemical decomposition of α-hydroxy-stearic acid on a 0.01*N* sulfuric acid subphase, probably with the splitting off of carbon dioxide, the reaction occurs only if the wave length is below 2400 Å, but that 3γ of nickel ions per 500 cm^{-3} sensitizes the reaction to light of 2537 Å. Also A. H. Hughes found that even in 0.01*N* hydrochloric acid ferric ions affect both the surface potential and the physical state of the film. On the same subphase the surface potential of a stearic acid monolayer is increased from 50 to 60 mv by ferric ions at concentrations greater than 0.0002 molal; but copper, nickel, cobalt, chromic, or ferrous ions produced no such effects at *M*/300.

Mixed Monolayers

Introduction

Mixed monolayers on an aqueous subsolution exhibit a certain analogy in two dimensions to solutions and mixed crystals in three dimensions. Thus at fixed values of temperature and film pressure, a component of the film may possess a certain definite solubility in the monolayer.

The properties of mixed films, that is, those which consist of two or more components in addition to water, are of fundamental importance for biology, since membranes are undoubtedly structures built up on one or more mixed monolayers as a foundation. Indeed, the first investigation of mixed films was undertaken by Leathes† on account of the supposed importance in biology of monolayers of this type; but it seems preferable to consider the behavior of simpler molecules, such as the mixtures of alcohols and acids investigated in 1925 by Harkins and Morgan.[25] However, before this is done it is desirable to give a few of the relations exhibited by monolayers in which nonpolar molecules are mixed with those which are polar-nonpolar.

Solubility of Nonpolar Oils in Expanded Monolayers of Polar-Nonpolar Substances

The attraction between a nonpolar oil and water is due largely to dispersion forces, whose energy varies inversely as the sixth power of the distance. Polar groups exhibit an energy which varies as the inverse cube, and also with the cosine of the angle between the dipoles.

In the case of alcohols, organic acids, or amines, hydrogen bonds increase the attraction between the monolayer and water, or in the monolayer itself. These are highly directional bonds of low energy when compared with covalent bonds.

On account of its relatively low attraction for water, a nonpolar oil should not be able to remain in a liquid condensed monolayer of an acid, alcohol, etc.; but in a liquid expanded monolayer at low film pressure (and high area) there is area between molecules which is not completely occupied, and which is therefore available for occupation by molecules of nonpolar oils. Thus it is not surprising that, as found by Harkins and Myers,[56, 59] the shape of the pressure (π)—area (σ) curves for myristic, pentadecylic, and similar acids is greatly changed by the presence of a non-volatile hydrocarbon, such as Nujol (Fig. 36). Small amounts of Nujol increase the film

* Langmuir, I. and Schaefer, V. J., *J. Am. Chem. Soc.*, **58**, 287 (1936).
† Leathes *Z. Physiol. Chem.*, **130**, 113 (1923); *Lancet*, ii, 853 (1925).

pressure at all areas; large amounts decrease the film pressure at the lower areas. It may be noted that at a very low film pressure the presence of Nujol increases the area per molecule of *acid* from the standard value of 46 Å^2 to about 56 Å^2 with myristic or 53 Å^2 with pentadecylic acid. With some proportions the areas are intermediate.

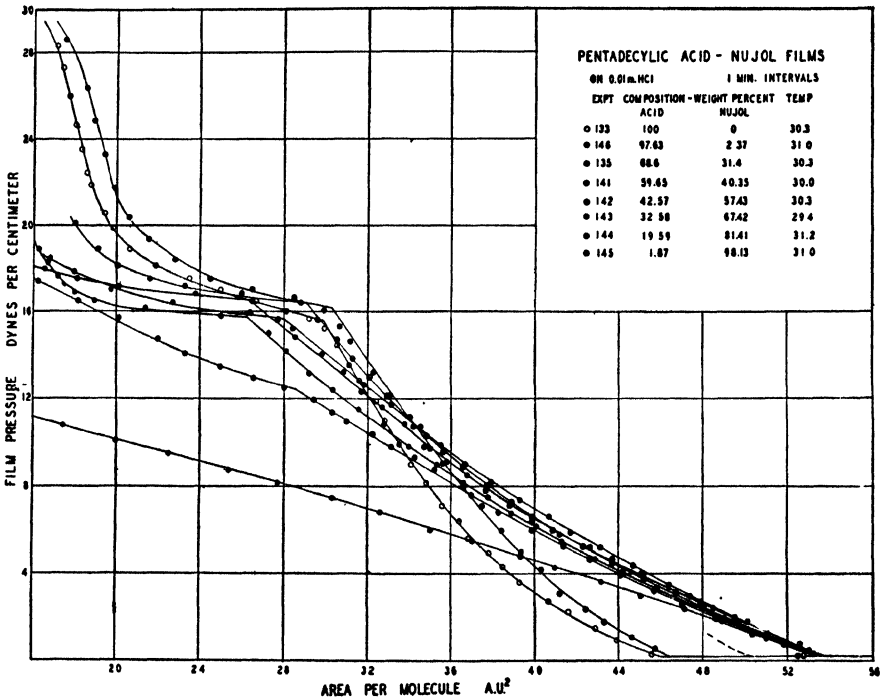


FIGURE 36. Mixed monolayers of pentadecylic acid and a hydrocarbon oil (Nujol).

It is of interest that a sufficiently large proportion of nonpolar oil, which is partly dissolved in the monolayer, and partly present as lenses on top of the monolayer, removes the second-order transition between the expanded and the intermediate states.

That the monolayer of acid dissolves hydrocarbon was shown by the use of an ultramicroscopic examination [64] of the monolayer. At 23° C a myristic acid monolayer which contained 35 per cent of the hydrocarbon oil was found to be optically empty at a very low film pressure, which shows that this much of the oil is soluble in the monolayer. At a film pressure of 5 dyne cm^{-1} the solubility was reduced to 25 per cent, and at 13.5 dyne cm^{-1} to 9.2 per cent (Fig. 37a). When the solubility was obtained by compressing the film it was found that too rapid compression appeared to give a slight supersaturation.

The phenomenon is an extremely beautiful one as observed in the ultramicroscope. If the film is "clean" the field is found to be almost optically empty; but when a slight supersaturation is obtained, a great number of brilliant scintillating droplets suddenly appear. These disappear just as suddenly if the area is increased slightly, and this procedure may be repeated as often as desired. The pressure-temperature and molecular area at this point give the values under which the solubility is represented by the composition of the film.

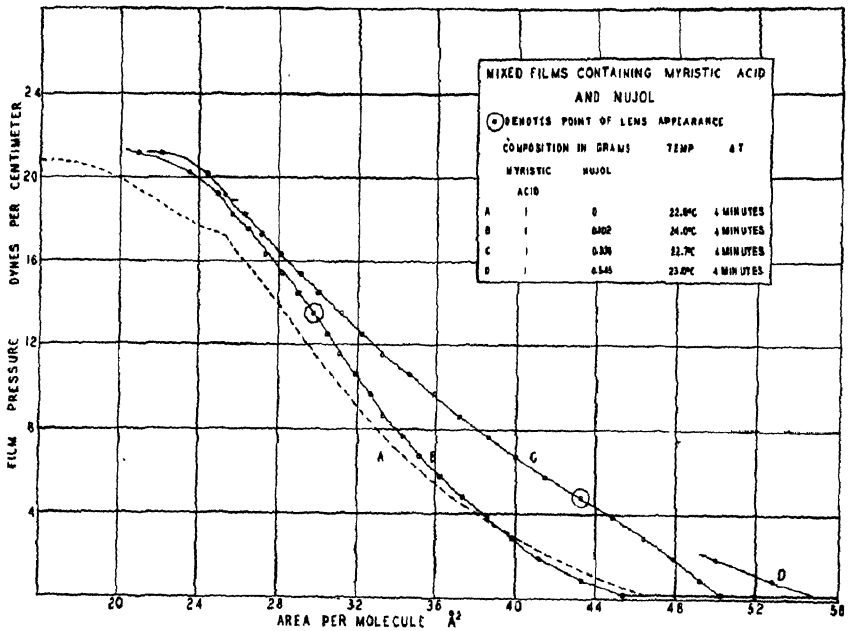


FIGURE 37a. Determination of Solubility in Two-dimensional Systems by the Ultramicroscope.

The Squeezing out of Crooked Molecules from Monolayers by Increase of Pressure: Separation of Oleic Acid from Monolayers of Stearic Acid, Stearyl Amine, etc.

Mixtures of oleic acid with cetyl alcohol, tripalmitin, and triolein have been investigated by Schulman,* who found that in every case the film was unstable above the collapse pressure of the component which has the lower collapse pressure.

Ultramicroscopic investigations on mixed films of this type by Florence and Harkins [73] have revealed extremely interesting relations, as follows:

(1) On an acid subphase at pH 3, oleic acid separated from a monolayer of either stearic acid or stearyl alcohol as visible lenses at a mean molecular area of 22.5 Å² on increase of film pressure to 24 dyne cm^{-1}. In these films half the molecules were of oleic acid and half were of the other constituent (Fig. 37b).

(2) After the lenses appeared, the film pressure increased much more slowly as the oleic acid separated, until at about 27 dyne cm^{-1} and a mean area of all of the molecules of 10.5 Å², the pressure began to rise rapidly. It is evident that here practically all the oleic acid has been squeezed out of the monolayer, and that the stearic acid or stearyl alcohol which remains has a molecular area of $10.5 \times 2 = 21$ Å². The linear pressure-area plot above this pressure is, for the alcohol, almost exactly that for the pure alcohol, although a trace of oleic acid may remain until the film pressure rises to 40 dyne cm^{-1}. The monolayer of stearic acid alone is not so stable and collapses at 32 dyne cm^{-1}.

If oleic acid is replaced by elaidic acid, lenses appear at about 24 dyne cm^{-1} if the other constituent is either stearyl alcohol or stearic acid, but the amount of elaidic acid which separates as lenses is *very small* (see Fig. 38), as is made evident by the fact that for the equimolar mixture at about 25° C the *mean* molecular area is 19.5 Å²

* Schulman, J. H., *J. Biochem.*, **29**, 1243 (1935).

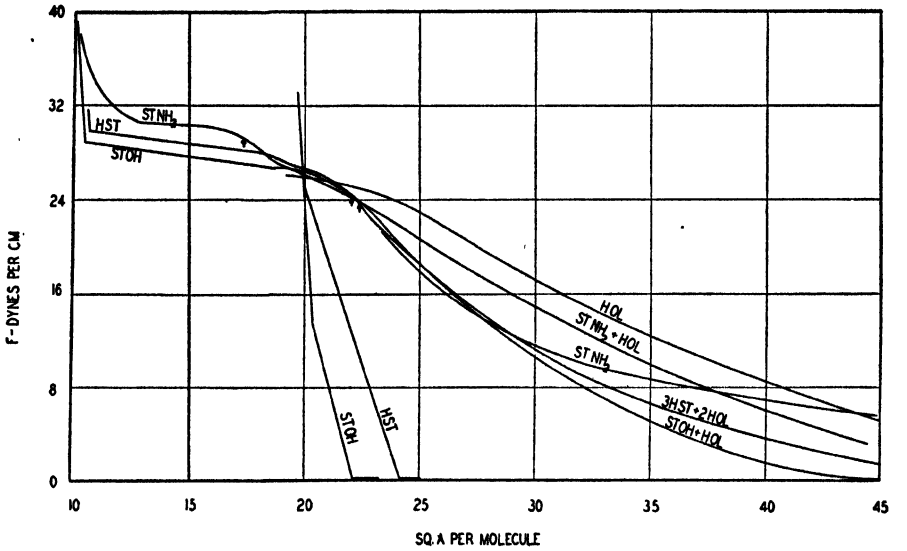


FIGURE 37b. Pressure-area relations of an oleic acid monolayer, and of mixtures of oleic acid (HOl) with stearyl alcohol (StOH), stearic acid (HSt), and stearyl amine ($StNH_2$) on a subsolution of pH 3. The arrows represent the points at which lenses of oleic acid become visible in the ultramicroscope. Since at the extreme left all of the oleic acid has been squeezed out the figure has been drawn in such a way that the values of the area multiplied by two give the areas for the constituent left in the film: *e.g.*, the almost vertical line at 10.5 Å^2 for the mixture of oleic acid and stearyl alcohol indicates that the molecular area for the alcohol left alone in the film is 21 Å^2.

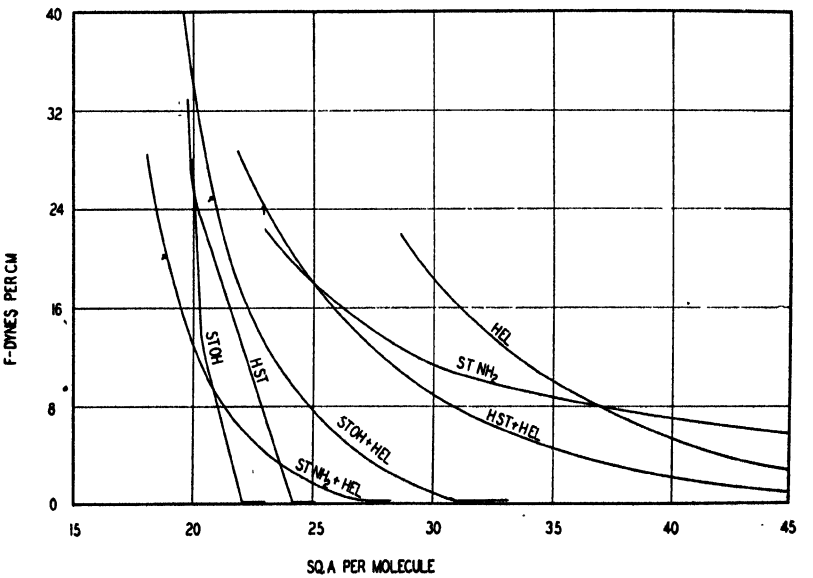


FIGURE 38. Same as Fig. 37b but oleic acid, the *cis* form replaced by elaidic acid, the *trans* form.

at 40 dyne cm^{-1} for the mixture of stearyl alcohol and elaidic acid. At this temperature and pressure the alcohol alone exhibits an area of 20.0 $Å^2$, which is not much higher.

The explanation of the fact that at the higher pressures almost all the elaidic acid molecules remain in these mixed films, while the oleic acid molecules are squeezed out (to form lenses on top of the film) is extremely interesting. Oleic and elaidic acids are space isomers with a double bond in the middle of the hydrocarbon chain. Since oleic acid is the *cis* isomer it has a much greater bend at the double bond than elaidic acid, which is in the *trans* form (Fig. 39). Thus when the orientation

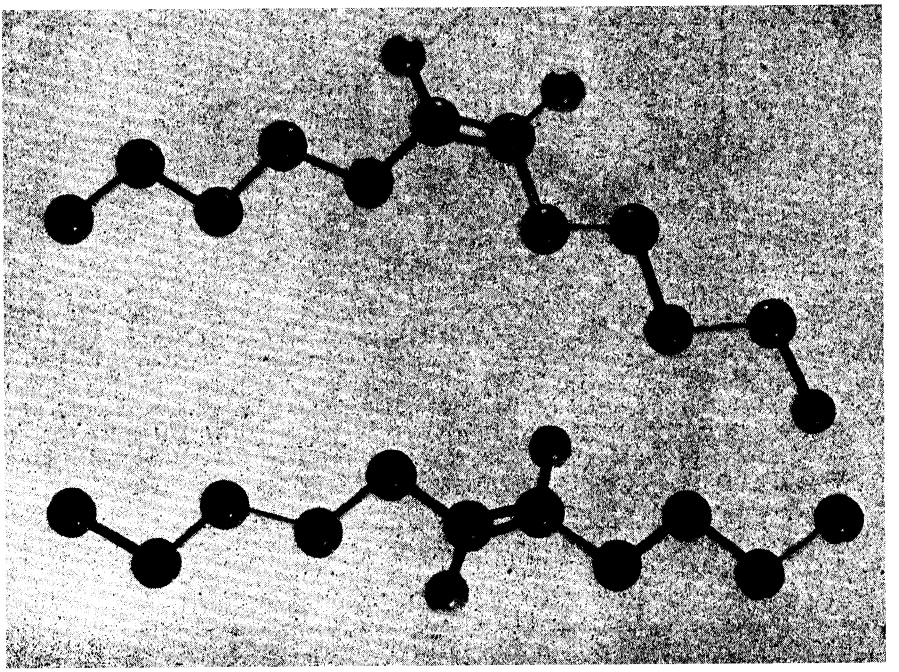


FIGURE 39. Arrangement of carbon atoms in *cis* and *trans* arrangements at the double bond.

of the molecules is approximately perpendicular to the surface of the water, an oleic acid molecule occupies more area than one of elaidic acid, and this in turn more than one of stearic acid, stearyl alcohol, etc. Since the energy of attraction between the chains varies inversely as the sixth power of the distance, oleic acid molecules are very much less tightly bound in the monolayer than the others. Since all the three acids have carboxyl groups adjacent to the water, it is not to be expected that the oleic acid can be less tightly bound to the water than the others.

Penetration of Insoluble Monolayers of Alcohols and Biological Derivatives by Capillary-active Substances (Fig. 40)

An interesting observation made by Schulman and Hughes * shows that great changes in the film pressure and potential of an insoluble monolayer may be produced by injecting beneath it certain capillary-active substances. Schulman and

* Schulman, J. H., and Hughes, *Proc. Roy. Soc., London* **A124**, 333 (1929).

Stenhagen* consider that the phenomenon is due to the formation of definite compounds: for example, that one molecule of cetyl alcohol combines with either one or three, but not with two, molecules of sodium cetyl sulfate, when an insoluble film of the alcohol is penetrated by the latter. However, the recent unpublished work of Harkins, Gordon, and Copeland, and the earlier work of Harkins and Florence, show that the *evidence* presented for the existence of these compounds is not correct, as will be shown later.

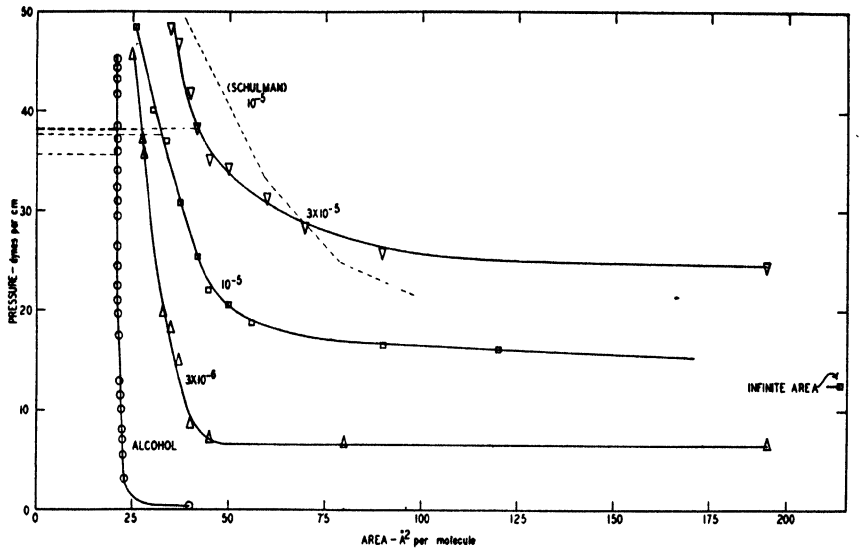


FIGURE 40. Pressure-area relations of equilibrium of monolayers of cetyl alcohol after penetration of molecules of sodium cetyl sulfate from aqueous solution of molal concentrations: 3×10^{-6}; 1×10^{-5}; and 3×10^{-5} at 25° C. Harkins, Gordon, and Copeland, with dotted curve by Schulman at 1×10^{-5} m for comparison. Schulman's values do not represent equilibrium, but only "one minute points."

The qualitative results of Schulman and co-workers indicate that the effects are highly specific: "for instance, saponin readily penetrates films of cholesterol or ergosterol, but not films of cholesterol acetate, cetyl alcohol, or calciferol. Similarly sodium cetyl sulfate penetrates cholesterol films, but not those of cholesterol acetate or calciferol."

The fact that sodium cetyl sulfate will penetrate monolayers of cetyl or elaidyl alcohol, but not readily one of oleyl alcohol, is due to the considerable bend in the latter molecule due to the *cis* form at the double bond. This indicates that London dispersion forces play a part in this phenomenon.

The change in the pressure-area relations at 25° C of a monolayer of cetyl alcohol, as produced by different concentrations of sodium cetyl sulfate, is illustrated by Fig. 40, taken from the work of Harkins, Gordon, and Copeland. The curve on the left is for a monolayer of cetyl alcohol, which at 25° C and 26 dyne cm^{-1} gives an area of 20.3 $Å^2$ $molecule^{-1}$. The first curve to the right of this exhibits the *equilibrium* relations when the aqueous subphase contains 3×10^{-6} mole of sodium cetyl sulfate per 1000 g water. The effect of the penetration by this latter substance is to increase the area per molecule of cetyl alcohol, plotted on the x-axis very slightly at high pressures. At low pressures and above 50 $Å^2$ per molecule of alcohol,

* Schulman, J. H. and Stenhagen, E., *Ibid.*, **B126**, 356 (1938).

the film pressure differs only very slightly from that given by the sodium cetyl sulfate alone (6.46 dyne cm^{-1}). However, at an area of 26.7 $Å^2$ for the alcohol, the pressure alcohol when alone is not far from 20 $Å^2$, and since if sodium cetyl sulfate had an area of 20 $Å^2$, the area 40 $Å^2$ would correspond to the molecular area of the 1:1 compound, Schulman concludes that since 38 $Å^2$ is close to 40 $Å^2$, the cause of the collapse at this area is the existence of this definite compound. However, since our work indicates that at a film pressure of 50 dyne cm^{-1} there is a continuous change of area from 20 $Å^2$ to more than 35 $Å^2$ per molecule of alcohol, if equilibrium is attained, his evidence is worthless, since it does not correspond to the facts. His evidence that a compound of 1 alcohol: 3 sodium cetyl sulfate exists is based on a kink at about 78 $Å^2$, which is close to 80 $Å^2$, which would allow area for a 1:3 compound, provided the proper area for the sodium sulfate molecule is 20 $Å^2$. However, since accurate work reveals no such kink, there is no evidence for this compound. This does not prove that compounds do not exist, but only that there is no *evidence* for their existence.

The curve which represents the penetration from a 10^{-5} molal solution of sodium cetyl sulfate gives an opportunity to compare this work with that of Schulman, whose curve for the same concentration is shown by the dotted line. The great difference between the two is due to the fact that Schulman used a single film, whose area was decreased each 60 seconds (1-minute points); thus time was not allowed to attain equilibrium by diffusion from the film into the subphase. In our work each point plotted represents a separate experiment. At an area of 45 $Å^2$ per molecule of sodium cetyl sulfate, the film pressure was still changing slowly at the end of six hours when 10^{-5} molal sodium sulfate was used. With 3×10^{-6} molal sodium cetyl sulfate equilibrium was attained more slowly.

On account of the long period needed to attain equilibrium, experiments of this type should be carried out with great care. For this work the highly accurate thermostat of Harkins and Copeland (see Fig. 21) was used. In this the trough and film balance are totally inclosed by a thermostat filled with aqueous diethylene glycol, regulated to 0.001° C. The air in the thermostat is kept very nearly saturated with water vapor. The vertical type of film balance was used. The film pressures, which for a single point were taken for periods as long as 12 hours or even more, were recorded automatically on the photographic film of a motion-picture camera by a method devised and an apparatus constructed by Gordon. The mirror of the film balance gave a fine line of light on a very long vertical Celluloid scale, graduated in millimeters. The mechanical device, operated by a synchronous clock, causes the motion-picture camera to take a photograph of the scale every five minutes during the whole time necessary to attain equilibrium.

From other work in this laboratory it has been found that these mixed films may give pressures as high as about 60 dyne cm^{-1}, but a peculiar phenomenon prevents the measurement of these high pressures in the ordinary type of experiment. This is because at about 50 dyne cm^{-1} these solutions begin to spread over the paraffin used to keep the aqueous phase from spilling over the edge of the trough.

For this type of work a new technique was developed six years ago. Instead of injecting the solution of the surface-active material under the insoluble monolayer, which leads to an inhomogeneous solution and monolayer, the first step is to fill the trough with a homogeneous solution. For example, the trough is filled with a solution of sodium cetyl sulfate. On the surface of this solution two barriers are placed in contact with each other, and perpendicular to the length of the trough. One of these is moved rapidly away from the other, thus sweeping off the adsorbed monolayer of sodium cetyl sulfate. As the moving barrier advances, drops of a hexane solution of cetyl alcohol, or another film-forming substance, are put on the surface just behind this barrier. Since the film pressure of the alcohol is greater than that of the sodium cetyl sulfate, the surface of the water is covered by the

insoluble film before appreciable penetration occurs. This technique leads to the formation of a homogeneous mixed film.

Mixtures of Components Each of Which Gives a Condensed Film (Fig. 41)

The following interesting relations are exhibited by the film which consists of an equimolar mixture of stearic acid and octadecyl (stearyl) alcohol.[72]

(1) The liquid condensed film exhibits a molecular area which is the mean of that given by the pure components. Thus the "limiting areas at zero pressure" as obtained by extrapolation are at 25° C: stearic acid = 24.3, octadecyl alcohol = 21.9, mean value calculated = 20.1, determined = 23.1.

(2) The S or "solid" film, has almost exactly the same molecular area in films of either the acid or the alcohol, but for the equimolar mixture the area is higher (by

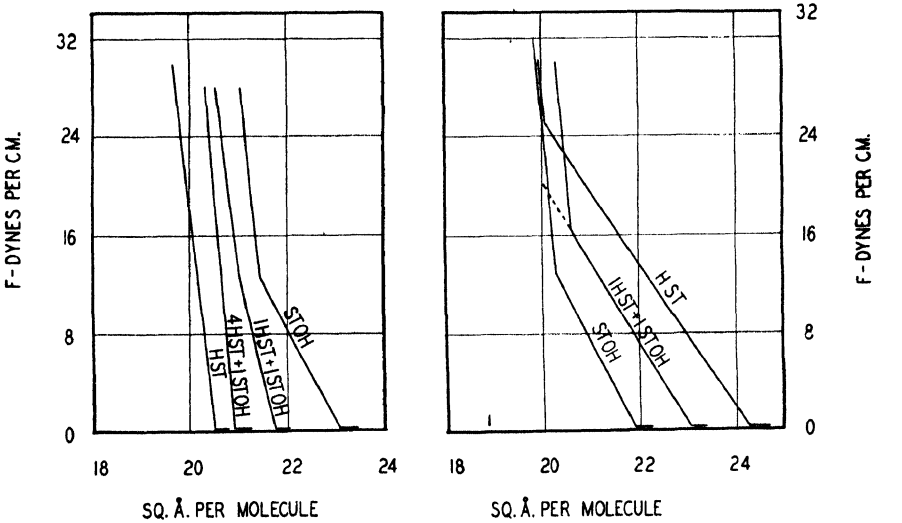


FIGURE 41. Pressure-area relations of mixtures of stearic acid and stearyl alcohol.

about 0.75 Å^2). This seems to be due to the fact that the transition $L_e \rightleftharpoons S$ occurs in the mixture at a pressure which is much (1.4 dynes) below the mean of the kink points for the films of one component. If S films were to be considered as two-dimensional solids, then it could be said that the mixed film freezes at a larger molecular area than that of either component.

The Condensing Effect of Alcohols upon the Expanded Monolayers of Acids

Alcohols with 14 or more carbon atoms per molecule form highly condensed S films at high and liquid-condensed films at low pressures. These substances reduce greatly the extent of the expansion in liquid-expanded and intermediate films. However, they may exhibit an area either higher (+) or lower (−) than the mean value, and in Table 11 this is represented as the interaction in Å^2 at the limiting area for the film. This is the area at which the pressure begins to rise rapidly above that of the two-dimensional vapor from the film.

In both Table 11 and Figs. 42 and 43 the numbers in the first column or on the curve represent the number of carbon atoms in the hydrocarbon chain, first of the alcohol and secondly of the acid. Thus 18–15 represents a mixture of octadecyl alcohol and pentadecylic acid. As compared with the mean value the relations which are obtained may be noted as follows:

(1) A contraction occurs if in mixtures with pentadecylic acid the alcohol chain is of the same length or longer than that of the acid.

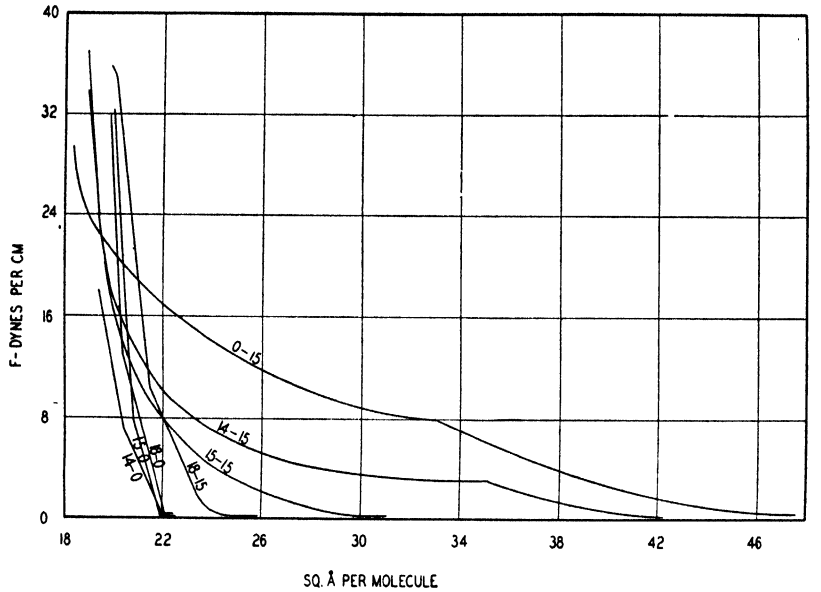


FIGURE 42. Mixtures (1:1) of condensed alcohol and expanded films of pentadecylic acid. Pentadecylic alcohol eliminates the expanded film of pentadecylic, but not that of myristic acid. The first number gives the number of carbon atoms in the molecule of alcohol, and the second, in the acid.

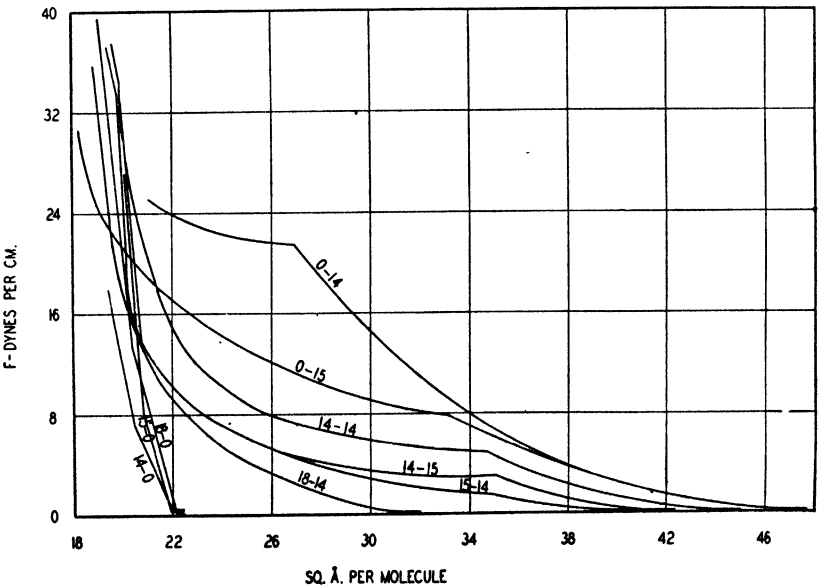


FIGURE 43. Mixtures (1:1) of condensed alcohol films and expanded films of myristic acid (pH = 3). The curve for myristyl alcohol with pentadecylic (acid (14-14) is added for comparison with the effect produced with interchange of number of carbon atoms in the alcohol and acid). Stearyl alcohol does not condense the myristic acid, while it does the pentadecylic acid film (Figure 42).

Table 11. Increase of Area or of Surface Potential (Interaction) Above the Mean Value for Mixtures of Acids, Alcohols, and Amines

Mixture	pH	Area Interaction at Limiting Area (A^2)	Limiting Area (A^2)	Dipole Interaction (mv)
Alcohols + Acids (1 : 1)				
18-18	3	+0.02		+33
18-18	9.5	−0.05		+20
18-15	3	−9.6		+2
15-15	3	−4.5		+50
14-15	3	+7.1		+28
18-14	3	−3.7		+29
15-14	3	+5.8		+63(+3)
14-14	3	+9.5		+42
Acids + Acids (1 : 1)				
20-15	3	−4.3		−27
20-14	3	+2.5		−35
18-14	3	+4.5		—
Alcohol + Amine (1 : 1)				
18-18	3	Approx. −25	21.45	+60
16-18	3	Approx. −25	22.20	+30
14-18	3	Approx. −25	23.60	+26
18-16	3	Approx. −25	22.90	+200
16-16	3	Approx. −25	23.20	+100
14-16	3	Approx. −25	26.25	+200
Acid + Amine (1 : 1)				
18-18	3	Approx. −25	21.6	+200
16-18	3	Approx. −25	(20.4)	+240
15-18	3	Approx. −36	22.8	+105
14-18	3	Approx. −36	22.8	—
18-16	3	Approx. −25	(20.4)	—
16-16	3	Approx. −25	(20.4)	—
14-16	3	Approx. −36	21.6	—
Pure Components				
Alcohols	3		22.0	
18 and 16 acids	3		24.5	
15 and 14 acids	3		47	
18 and 16 amines	3		(Area = 70 Å² at ½ dyne)	

STEARIC ACID—STEARYL AMINE MIXTURES ON SUBPHASE OF pH = 3

6 Mol. S + NH_3 : Mol. HSt	Area at F = 2 dynes ($Å^2$)	Dipole Interaction (mv)
(pure $StNH_2$)	55	—
4 : 1	29.2	+140
2 : 1	22.2	+112
1 : 1	20.1	+200
2 : 3	20.8	+103
1 : 3	21.3	+125
1 : 6	23.1	+108
0 (pure HSt)	23.9	—

OCTADECYL ALCOHOL—STEARYL AMINE MIXTURES ON SUBPHASE OF pH = 3

Mol. $StNH_3$: Mol. StOH	Area at F = 2 dynes ($Å^2$)	Dipole Interaction (mv)
(pure $StNH_2$)	55	—
2 : 1	24.5	+30
1 : 1	21.1	+60
1 : 2	21.9	+60
0 (pure StOH)	21.85	—

(2) Expansion is exhibited if the alcohol chain is the shorter.
(3) With myristic acid there is contraction if the alcohol chain is much longer (*i.e.* 18–14) but expansion if the alcohol chain is one atom longer or of the same length.

It is of considerable interest that in an equimolecular mixture any alcohol of greater length than 14 carbon atoms eliminates the expanded film and leaves only the intermediate film (15–15, Fig. 42) or the liquid condensed film (18–15). That is, the transition pressure ($I \rightarrow L_e$) or kink, is depressed to negative film pressures.

It is found (Fig. 43) that equimolecular mixtures of the 14 alcohol with 15 acid, and of 15 alcohol with 14 acid, give the same pressure-area curve for the transition film over a considerable region, but the expanded liquid film exhibits a considerably higher pressure if the acid chain is the longer.

With myristic acid, stearyl alcohol inhibits the expanded film, but pentadecyl alcohol permits its existence.

Since the molecular interaction is expressed in Å^2 at the limiting area, it is not surprising that there is an expansion with reference to the mean when the mixed film at this area is an expanded one, and a contraction when the film is of either the L_e or the I type.

If the mixed film is of the intermediate type and that of the acid expanded, the relation is reversed. Thus at 7 dynes the interaction is -6.2 Å^2 for the 14–15 mixture while it is $+7.1$ at the limiting area.

Thus the sign and magnitude of the interaction depends upon the effect of the mixing upon the state or class of the film which is produced.

Increasing the chain length of the condensing agent has the same effect as a lowering of temperature even though this effect is not apparent with the pure alcohols with from 14 to 18 carbon atoms. Increasing the chain length of the expanded acid has a similar effect.

The deviation from the mean of the surface potential for the alcohol-acid monolayers is positive in all cases, and is smallest in magnitude when the difference in chain length is greatest. The greatest interaction in mv is 63 for the 15 alcohol—14 acid, and 50 for the 15–15 mixture. These are the largest interactions as measured by the potentials.

Films of Mixed Acids

Films of mixed acids (Fig. 44) exhibit the same general relations as those which consist of an acid and an alcohol, except that a long chain acid condenses an expanded film of an acid much less than the alcohol with the same number of carbon atoms. For example, at an area of 26 Å^2 the equimolecular mixture of the 18 and 14 acids gives a film pressure of 8.5 dynes/cm, while with the 18 alcohol and the 14 acid it is only 3.3 dynes. This is related to the fact that liquid condensed alcohol films are much more condensed than those of the same type with acids.

One component films of arachidic (20) or stearic (18) acid exhibit practically the same pressure-area relations for the liquid film, and when mixed in equimolecular proportions with myristic acid give almost the same values. In the case of the mixtures the interaction in terms of the surface potential is negative (Table 11), which is the opposite sign from that for alcohol-acid mixtures.

In mixtures of the 22–14 acids Adam and Jessop * were unable to detect an appreciable condensation but in this work the mixture gave an increase from the mean of 2.5 Å^2 at the limiting area, but only 0.25 Å^2 at 8 dynes per cm. These involve comparisons of expanded films for the pure 14-component and the mixture. At 12 dynes the condensation is 6.3 Å^2, but here the mixture is in the intermediate state. These facts indicate that for an understanding of the relations between the films

* Adam, N. K., and Jessop, G., *Proc. Roy. Soc., London* **A120**, 473 (1928).

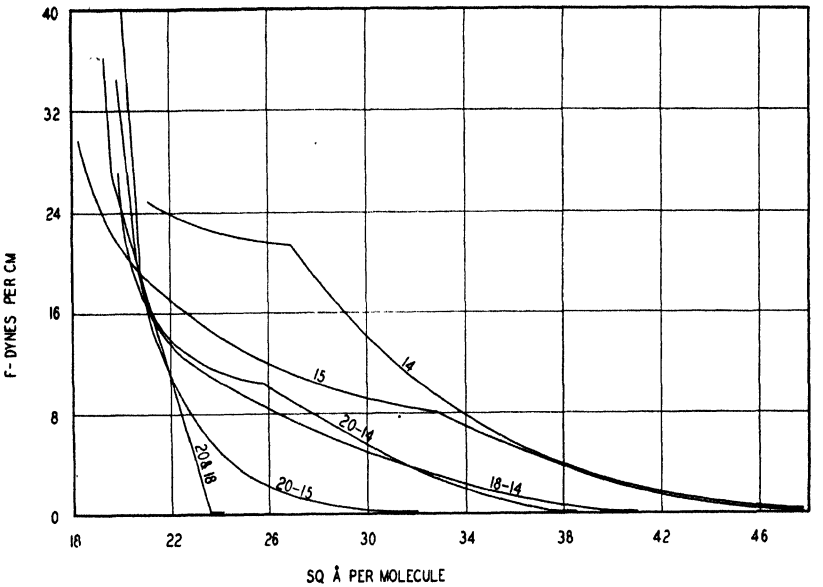


FIGURE 44. Mixtures of condensed and expanded acid films (pH = 3). Arachidic, stearic, pentadecylic, and myristic acids. The pressure-area values for the 20-14 and 18-14 mixtures are almost the same, but the position of the transition point expanded transition ($L_e \rightarrow I$) is affected by an extremely large shift.

of pure components and the mixtures in all of this work it is essential to study the entire curves as presented in the figures.

Positive Organic Ions in Mixed Films

Alcohol-Amine Mixtures: The use of amines on dilute acid subsolutions gives an opportunity to investigate the effect of a positive ion of an organic type. In alcohol mixtures this gives rise to an ion-dipole effect. The pure amines on acids give vapor-expanded films of extremely high expansion, but these are very greatly condensed by the addition of an equal number of molecules of the 14 to 18 alcohols (Fig. 45). Of the mixtures, that in which both the amine and the alcohol have 18 carbon atoms gives the closest approach to a solid film, with an extrapolated area at zero compression of 20.7 $Å^2$ which is the normal value for solid condensed films. All of the other mixtures give curves of the general type exhibited by intermediate films.

At the low pressure of the limiting area the 14 alcohol and the 18 amine give the highest, and the 18–18 mixture the lowest area, as might have been predicted, but at the higher pressures the order is shifted on account of the fact that the 14–16 mixture is by far the most compressible, as is normal.

Mixtures which contain the 16 carbon amine not only give abnormally low areas at the high pressures, but they also exhibit extremely high positive deviations (ΔV) of the surface potential from the mean values. Thus (Table 11) the interaction was +200 mv for the 18–16, 100 mv for the 16–16, and 200 mv for the 14–16 mixture, if the first number of each pair refers to the alcohol and the second to the amine. The mixtures with octadecyl- (stearyl-) amine give potential deviations of the same sign and order of magnitude as those found with the alcohol-acid mixtures, that is 25 to 60 mv. The large deviations with cetyl amine as listed above were the maxima, and it is not certain that the films were molecularly homogeneous.

The effect of changing the proportions in the alcohol-amine mixtures is exhibited

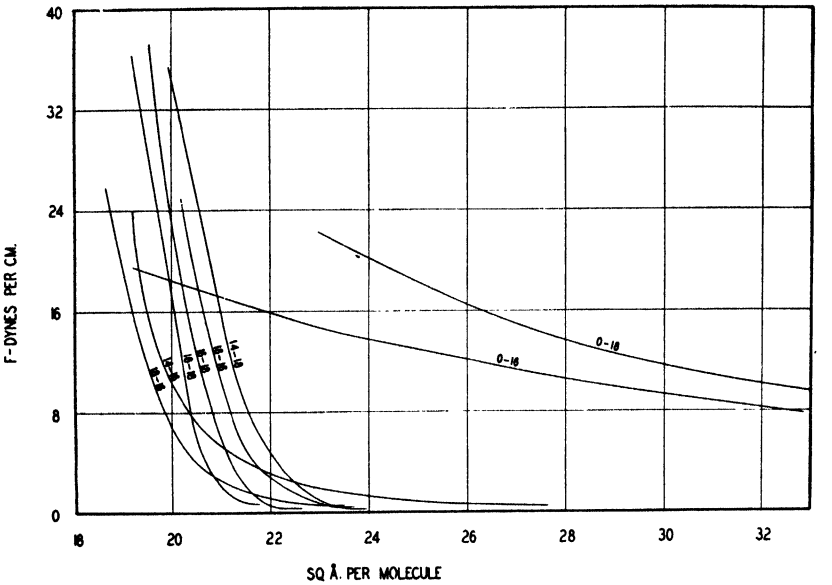


FIGURE 45. Mixtures (1:1) of condensed alcohol and vapor-expanded amine films (pH = 3). For the one-component alcohol films see Fig. 42 and Fig. 43. The curves at the lower pressures are of the transition type. The first number gives the number of carbon atoms in the alcohol, and the second in the amine, molecule. The 14-16 and 18-16 mixtures have a high compressibility.

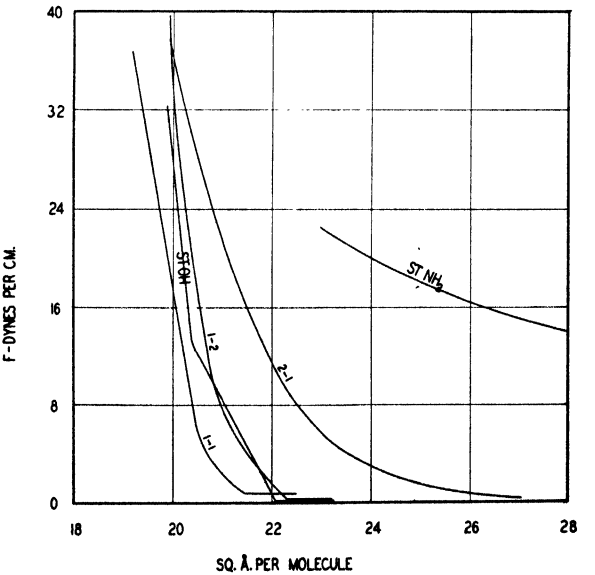


FIGURE 46. Effect of proportions in mixed stearyl alcohol and amine films (pH = 3). The first figure gives the number of molecules of amine and the second of alcohol. The 1:1 mixture exhibits the greatest condensation.

in Fig. 46. It may be noted that the 1:1 mixture gives the lowest area, and that one molecule of stearyl alcohol to two of the stearyl amine was not sufficient to give a solid film, though in Fig. 45 it has been shown that 1:1 produces the transformation.

Acid-Amine Mixtures (pH = 3): When an organic acid is present in the mixture there is the possibility of a chemical action with the positive ion of the amine, so an even more powerful condensing action is to be expected. Thus it is not surprising that a mixture of myristic acid with either cetyl or stearyl amine, both of which when alone form very highly expanded films at 25° C, gives a highly condensed film, but in no case is the relation between film pressure and molecular area linear (Fig. 47). The positive deviation of the surface potential from the mean is very high (Table 11), of the order of 200 mv, and is highest for the equimolar mixture.

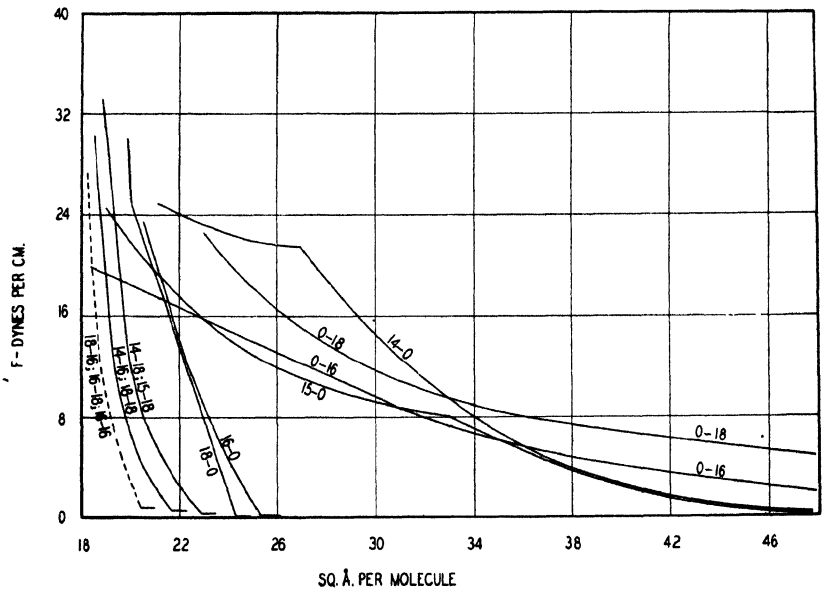


FIGURE 47. Mixtures (1:1) of acids and amines (pH = 3). The first figure gives the number of carbon atoms in the acid, and the second in the amine, molecule. All of the seven mixtures exhibit pressure-area curves of the same form, and with not very widely differing and very low areas. At the lower pressure all the mixtures give curves of the transition form.

In the acid-amine mixtures examined, 14–18, 15–18, 16–18, 18–18, 14–16, 16–16, 18–16, there is no evidence of a definite effect of chain length. All of the curves for the mixtures are seen to be practically parallel, and of the same form, and the difference in area is not very significant.

The effect of a change of the relative proportions of stearyl amine and stearic acid is presented in Fig. 48, where for example 4:1 means 4 molecules of amine to one of acid. The 1:1 mixture is found to give the minimum area and the greatest interaction of the potentials (200 mv).

The pressure of the transition $L_e \rightleftharpoons S$ for the mixtures is lowered, as the proportion of the amine is increased, from 25 dynes/cm for the pure acid, to 15.4 for one-seventh amine, to 9.8 for one-fourth amine, and the transition point disappears for two-fifths amine or more.

The effect of changes of the proportions in a mixed film of stearic acid and

stearyl amine, as given in Fig. 48, is to give a maximum in the potential interaction and a minimum in the area for a 1:1 proportion.

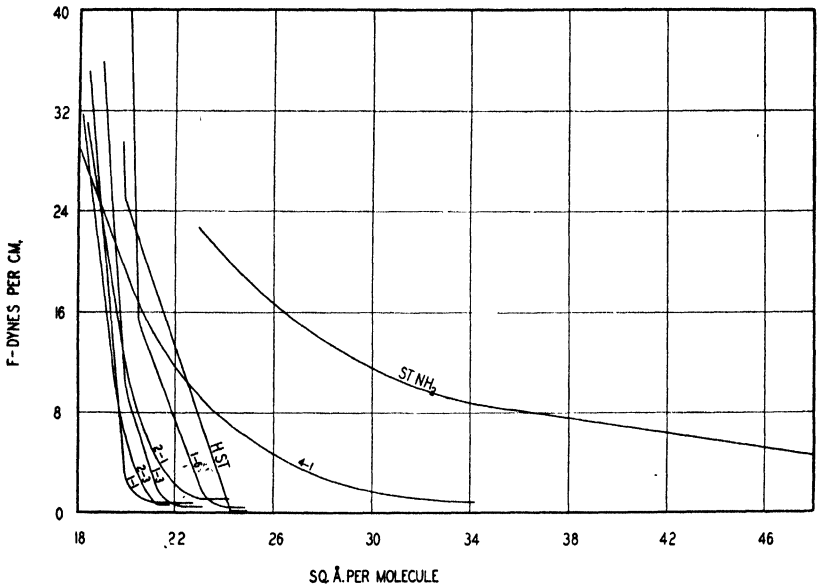


FIGURE 48. Effect of proportions in mixed films of stearyl amine and stearic acid (pH = 3). The numbers on the curves give the ratio mole of amine to acid. The 1:1 mixtures exhibits the greatest condensation with very low molecular areas, and above 3 dynes the low compressibility 0.0023 (at $\pi = 0$ extrapolated) or a Type S film, while the similar value for one component stearyl alcohol I film is 0.0024. With 75 molar per cent of acid the sharp transition II I or liquid to solid is preserved, but seemingly not with 66⅔ per cent.

THE ENERGY OF EXPANSION OF A MONOLAYER

Introduction

If the area of a water surface is increased by 1 cm^2 the energy of expansion, which may be designated as the energy of extension, is at 20° C 117 erg cm^{-2}, while for *n*-hexadecane the value is 53 erg cm^{-2} (*n*-heptane = 50) or less than half as much. It might seem logical, therefore, to conclude that if a monolayer of an oil is put on water, the energy of extension should be reduced. It will be shown later that it may be either unchanged by the presence of the monolayer, or very greatly increased, in certain instances to as much as 1830 erg cm^{-2}, which is an extremely high energy. This is, indeed, very remarkable, since it shows that the molecules are so arranged in the film as to give a high degree of interaction.

In a later section thermodynamic equations, which give the energy of expansion and of phase transition, are developed. In three dimensional systems there is a considerable change of volume (at constant pressure) *and of energy* during each phase transition, but in two dimensional systems this is the exception rather than the rule, though vaporization is accompanied by a large absorption of heat in either three or two dimensions.

Fusion in three dimensions involves a change of volume and an absorption of heat. In two dimensions the area is unchanged at the melting point, no heat is absorbed, and the change of phase is brought about by an increase of volume alone.

In general, changes of phase in two dimensional condensed systems involve no heat change. The energy of expansion is related in a remarkable way to the nature of the phase of the monolayer. Its magnitude and its variation with changes in phases or transitions between them, give important knowledge of the molecular interactions in two dimensional systems, and may be used to test the validity of theories of the molecular structure of the monolayer.

As an example of this we may test a well-known theory of the structure of a monolayer in the intermediate phase as proposed by Langmuir. Consider this phase as having just been formed by a transition from a liquid expanded phase by decrease of area. According to the theory the molecular structure is the same as in the expanded phase, except that a few micelles have now formed. With a given film forming substance every micelle contains the same small number (*e.g.*, 13) of molecules. This makes it essential that if in the expansion of the monolayer a certain micelle begins to dissociate, it must disappear completely, as an aggregate, before another micelle begins to disintegrate. Since the molecular area in the micelle is much less than that in the nonmicellular material, this phase should have a very high compressibility, as is true. However, Langmuir's theory predicts that, as the intermediate phase is compressed, the energy of compression should remain constant.

The measurements show, however, that for a normal long chain paraffin acid this energy is about 300 ergs per sq cm during the early part of the compression, but decreases gradually to zero before this phase changes into a liquid condensed phase. Thus if the micelle theory is to correspond with these facts the micelles would need to vary in size with the compression, instead of remaining of constant size. According to Kirkwood's theory, the structure is very different.

While a three-dimensional gaseous or condensed phase has only a single isothermal energy of expansion, this is not true of a monolayer. Consider a rectangular monolayer which is confined at one end of a rectangular trough by a single transverse barrier. If the barrier is moved in such a way as to (1) increase the area of the film by one cm^2 and to decrease the area of the water by the same amount the energy required is h_s, the energy of spreading of the monolayer. If, however, a ring or any other object in contact with the film is raised far enough to increase the area of the film by 1 cm^2 without affecting the area of the clean water, the energy required is designated by h_e, the energy of *extension* of the monolayer.

It is evident that $h_s = h_e - h_w$, since in spreading the energy of the film is increased by h_e, but the energy of the surface of the water is decreased by h_w, since unit area of the water disappears.

In other words the energy of extension of a monolayer is always greater than the energy of spreading by an amount equal to the surface energy of water at the given temperature. At 25° C this is 118.5 erg cm^{-2}.

Determination of the Energy of Spreading (h_s) and of Extension (h_e)

A monolayer of pentadecylic acid at a temperature of 20° C and a molecular area of 34.3 $Å^2$ per molecule has a film pressure (π) of 3 dyne cm^{-1}, so the surface tension of water covered by the film is $72.75 - 3 = 69.75$ dyne $cm^{-1} = \gamma_f$. Here 72.75 dyne cm^{-1} is the surface tension of clean water. If the film is of infinite extent its face energy is increased by 69.75 erg cm^{-2} if the area is increased by one cm^2. The film is being thus extended absorbs an amount of heat

$$q_e = -T\left(\frac{\partial \gamma_f}{\partial T}\right)_\sigma = Ts_e = 293 \times 1.106 = 324 \text{ erg cm}^{-2}$$

The total energy required for extension of the film is therefore:

$$h_e = \gamma_f - T\left(\frac{\partial \gamma_f}{\partial T}\right)_\sigma = 69.75 + 324 = 394 \text{ erg cm}^{-2} \qquad (34)$$

If the energy of spreading (h_e) is desired, it is only necessary to subtract from each of these values the corresponding value for clean water. Thus

$$f_e = \Delta\gamma_e = \gamma_e - \gamma_w = 69.75 - 72.75 = -3.0 \text{ erg cm}^{-2}$$
$$s_e = s_e - s_w = 1.106 - 0.148 = 0.958 \text{ erg cm}^{-2} \text{ deg}^{-1}$$
$$h_e = h_e - h_w = 394 - 116 = 278 \text{ erg cm}^{-2}$$
$$\gamma_e = \gamma_e - \gamma_w = 324 - 43 = 281 \text{ erg cm}^{-2}$$

These and other values are listed in Tables 12a and b.

Table 12a. Energy (erg cm^{-2}) of Extension and Spreading of Pentadecylic Acid.

	f	h	$q = Ts$	s
Intermediate film at 20° C and σ = 34.3A^2				
A. Clean water	72.75	116	43.25	0.148
B. Film (extension)	69.75	394	324	1.106
$C = B - A$ (spreading)	−3.0	278	281	0.959
Intermediate film at 18° and σ = 38A^2				
A. Clean water	73.05	115.5	42.49	0.146
B. Film (extension)	72.32	367	295	1.014
$C = B - A$ (spreading)	−0.73	252	253	0.869
Expanded film at 22° C and σ = 38A^2				
A. Clean water	72.46	116_4	44.00	0.14_9
B. Film (extension)	69.80	161	91.6	0.31
$C = B - A$ (spreading)	−2.64	45	47.6	0.16

Table 12b. Energy of Extension and Spreading of a Liquid and a Solid Film of Hexadecyl Alcohol at σ = 20.0A^2.

	f	h	$q = Ts$	s
Condensed liquid film at 9° C π = 7.7 dyne/cm				
A. Clean water	74.37	115.5	41.17	0.14_6
B. Film (extension)	66.7	166	99.6	0.35_8
$C = B - A$ (spreading)	−7.7	$50._7$	58.4	0.20_7
Solid film at 16° C π = 13.8 dyne/cm				
A. Clean water	73.34	116.1	42.77	0.148
B. Film (extension)	59.5	508	449	1.55
$C = B - A$ (spreading)	−13.8	392	406	1.40_4

The experimental determination of these energy quantities for an insoluble film is very simple. The surface tension of the film is simply the surface tension of water (γ_w) minus the film pressure (π). The other quantity needed is the rate of change of the surface tension of the film, at constant area, with the pressure. This rate of change is the slope of the curve which relates film pressure to temperature, as in Fig. 49, which represents the intermediate phase (steep curves) and the liquid expanded phase (1_e), with a more gentle slope, at higher values of π and T.

The energy changes in the expansion of a monolayer on water are remarkable in their variation with the phase involved, as indicated by the values in Table 13, which gives only the order of magnitude of the various energies (See Figs. 49b, c, and d).

Thermodynamics of the Spreading of Films [21, 85, 81, 88]

The maximum work (δW) done by a surface when its area (σ) is increased by $d\sigma$ is

$$\delta W = -\gamma d\sigma + p dv. \qquad (35)$$

The value of the free energy of Gibbs is defined by the equation

$$F = E - ST + pv \tag{36}$$

and the Helmholtz free energy by

$$A = E - ST. \tag{37}$$

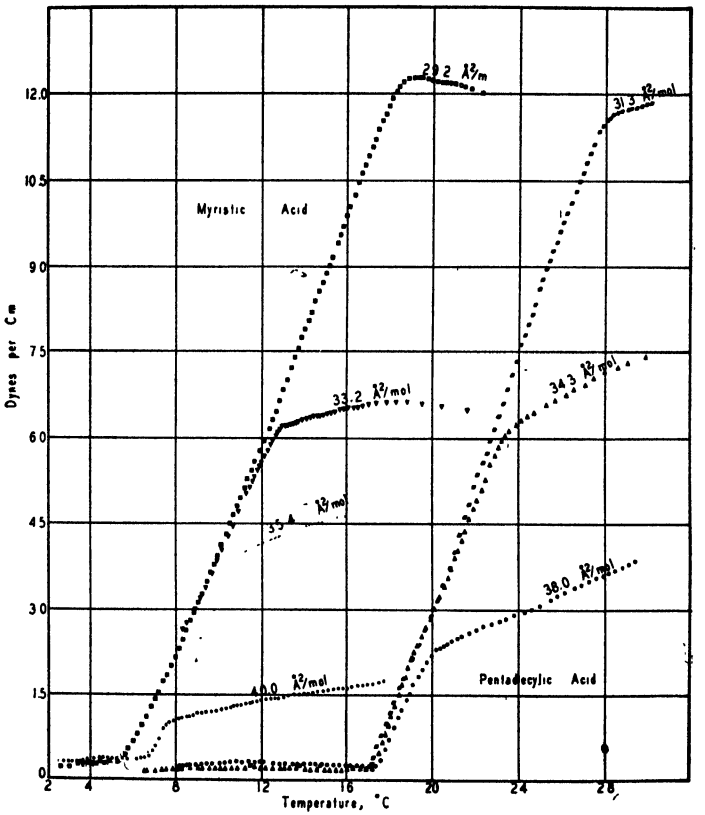


FIGURE 49a. Pressure-temperature relations of monolayers of myristic and pentadecylic acids. From left to right: L_c, I, and L_e phases. The upper part of the highest curve for myristic acid has a slope of incorrect sign, since at this high temperatures and pressures myristic acid monolayers are slightly soluble.

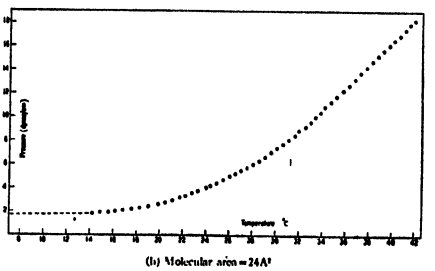
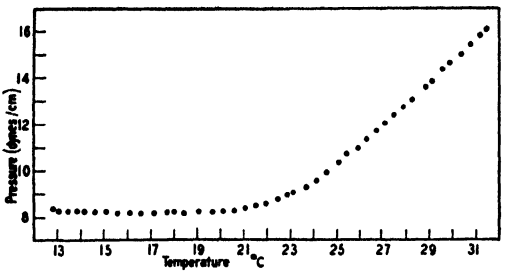


FIGURE 49b. Pressure temperature relations of palmitic acid monolayers (condensed and intermediate phases).



Molecular area = 21A².

FIGURE 49c. Pressure-temperature relations of pentadecylic acid.

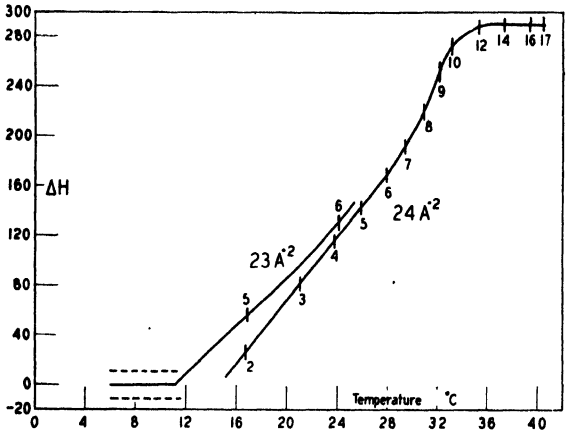


FIGURE 49d. Increase of the "Heat Content" for spreading $h_s = (\partial H_s/\partial T)_{p,T,\sigma}$ as a function of temperature and pressure; numbers on curves give the film pressure π.

It follows that

$$F = A + pv \tag{38}$$

and

$$dF = dA + pdv + vdp. \tag{39}$$

If a processes is reversible and isothermal, $T\Delta S = Q$ is the heat adsorbed and

$$dA = -dW, \tag{40}$$

so

$$dF = -dW + pdv + vdp. \tag{41}$$

From (41) and (35)

$$dF = \gamma d\sigma + vdp, \tag{42}$$

so that at constant pressure and temperature

$$(\partial F/\partial\sigma)_{p,\,T} = \gamma \tag{43}$$

and for a saturated surface γ is a function of p and T only, so

$$\Delta F = \gamma\Delta\sigma, \tag{44}$$
$$f = \gamma. \tag{45}$$

The condition that a liquid b shall by itself spread on a liquid a, at constant temperature, is thus that the summation of the terms $(\partial F/\partial\sigma')_{p,\,T}$ must be negative, that is less than zero. Thus, if b spreads on a, a certain area of the surface of a disappears and certain areas of the surface of b and of the interface ab, appear. The increase of free energy is

$$dF = \left(\frac{\partial F}{\partial \sigma_b}\right)_{p,\,T} d\sigma_b + \left(\frac{\partial F}{\partial \sigma_{ab}}\right)_{p,\,T} d\sigma_{ab} + \left(\frac{\partial F}{\partial \sigma_a}\right)_{p,\,T} d\sigma_a. \tag{46}$$

For a film of appreciable thickness the effects of gravitation must also be considered; but they may be neglected here.

The condition for spreading to occur starting from any given condition, is

$$dF < 0, \tag{47}$$

and for non-spreading

$$dF > 0. \tag{48}$$

If it is assumed that

$$d\sigma_b = d\sigma_{ab} = -d\sigma_a, \tag{49}$$

which is practically valid for a very thin layer, then

$$(dF/d\sigma)_{p,\,T} = \gamma_b + \gamma_{ab} - \gamma_a. \tag{50}$$

provided the film is *duplex*, that is, thick enough to cause the film of b to exhibit the surface tension of the pure liquid. Let $-dF/d\sigma$ be designated as the final spreading coefficient $(S_{b'/a'})$ then

$$S_{b'/a'} = \gamma_{a'} - (\gamma_{b'} + \gamma_{a'b'}) \tag{51}$$

(p, T = constant, the surfaces are saturated as is designated by the primes).
It will be shown later that

$$S_{b'/a'} < 0. \tag{52}$$

That is, the final spreading coefficient is always negative.

Now if a drop of the *pure* liquid b is put on the surface of the pure liquid a, and the surfaces of these liquids are considered to be still uncontaminated with each other, then the initial coefficient of spreading $S_{b/a}$ is given * by the equation

$$-(dF/d\sigma)_{p,\,T} = S_{b/a} = \gamma_a - (\gamma_b + \gamma_{a'b'}). \tag{53}$$

If b is saturated with a, but the surface of a is clean, the initial coefficient may be designated as the semi-initial spreading coefficient $(S_{b'/a})$, which, on account of the small effect of water on the surface tension of an organic liquid, is not very different from $S_{b/a}$ when these are the liquids involved.

Now

$$S_{b'/a} = \gamma_a - (\gamma_{b'} + \gamma_{a'b'}). \tag{54}$$

The work (W_A) required at constant T and p to pull the interface ab apart to give clean surfaces of a and b, is, for unit area of interface, given by:

$$W_A = \gamma_a + \gamma_b - \gamma_{a'b'} \tag{55}$$

and the work of cohesion (W_C) for the liquid b is *

$$W_{Cb} = 2\gamma_b \tag{56}$$

or

$$W_A - W_{Cb} = \gamma_a - (\gamma_b + \gamma_{a'b'}). \tag{57}$$

* W. D. Harkins and A. Feldman, *J. Am. Chem. Soc.*, 44, 2665 (1922); W. D. Harkins, Sixth Colloid Symposium Monograph (1928), p. 31.

From (53) and (57) the initial spreading of the liquid b over the clean surface of a is given by

$$-(\partial F/\partial \sigma)_{p,T} = S_{b/a} = W_A - W_{Cb} \gtreqless 0. \tag{58}$$

Thus spreading to form a duplex film occurs if $S_{b/a}$ is greater than zero, while b remains as a lens if it is less than zero.

Transformation of a Duplex Film to Form a Non-Duplex Thin Film (Presumably a Monolayer)

If a non-duplex thin film or a monolayer of b is present on the surface of the liquid a, the pressure (π) of the film is *defincd* as:

$$\pi = \gamma_a - \gamma_f, \tag{59}$$

where γ_a is the surface tension of the pure liquid a and γ_f is its surface tension when covered with a film of b. While there is nothing in the thermodynamics which indicates that such a thin non-duplex film is always a monolayer, certain relations, presented later, seem to indicate that this is true.

What is now desired is the relation between the film pressure and the spreading coefficient, which is obtained from equations developed earlier.

In the treatment of the transformation of a duplex film into a monolayer it is essential to consider the semi-initial spreading coefficient ($S_{b'/a}$; Eq. (54)).

Equation (51) may be written

$$S_{b'/a'} = \gamma_e - (\gamma_{b'} + \gamma_{a'b'}), \tag{51'}$$

where γ_e is the equilibrium value of γ_f and Eq. (54)—Eq. (51′) gives

$$S_{b'/a} - S_{b'/a'} = \pi_e. \tag{60}$$

Thus π_e, the pressure of the monolayer in equilibrium with a lens of b, is equal to the semi-initial minus the final spreading coefficient.

With an oil on water *the final spreading coefficient is always negative, so the film pressure* (π_e) *is always larger than the semi-initial coefficient* ($S_{b'/a}$), which has nearly the value of the initial coefficient ($S_{b/a}$).

Thermodynamics of Extension and of Spreading of a Monolayer

The heat content (H) of a system is given by the equation

$$H = E + pv \tag{61}$$

or

$$dH = dE + pdv + vdp, \tag{62}$$

and from 2

$$H = F + ST \tag{63}$$

or

$$dH = dF + TdS + SdT. \tag{64}$$

In the treatment which follows, energy values designated by small letters refer to the change of energy for unit area. Thus in

$$h = (\partial H/\partial \sigma)_T,$$

the h designates the increase of heat content of a film of very large area for unit increment of area.

Let us consider a trough filled with a body of water with a plane surface partly

covered with a monolayer. The area of the film together with that of the subphase may be increased by one sq cm by the use of some mechanical device while the area of the water surface is kept constant. Thus a ring of suitable dimensions, already lying in the surface, may be pulled upward until there is a unit increase of area of the film. If the film area is thus increased by $d\sigma_f$, instead of unity, the equation for the process which may be designated as *extension* (e), may be written

$$dF_e = (\partial F/\partial \sigma_f)_{T,\sigma_w} d\sigma_f = \gamma d\sigma_f \tag{65}$$

or

$$\Delta F(\Delta\sigma = 1 \text{ cm}^2) = (\partial F/\partial \sigma_f)_{T,\sigma_w} = f_e, \tag{66}$$

which gives the *free energy of extension* of the film. The entropy of extension is

$$s_e = (\partial S/\partial \sigma_f)_{T,\sigma_w} = -(\partial \gamma_f/\partial T)_{\sigma_f}. \tag{67}$$

The increment of heat content is given by

$$h_e = (\partial H/\partial \sigma_f)_{T,\sigma_w} = \left[\frac{\partial(\gamma_f/T)}{\partial(1/T)}\right]_{\sigma_f} \tag{68}$$

and the heat absorbed (q) by

$$q_e = T(\partial S/\partial \sigma_f)_{T,\sigma_w} = -T(\partial \gamma_f/\partial T)_{\sigma_f} = Ts_e. \tag{69}$$

Similar equations may be written for the energy of *spreading* of a film. This process (s) may be assumed to occur as follows: The surface of the water in the trough is separated into two parts by a movable barrier. On one side of the barrier the surface is covered by the film, on the other side the water surface is clean. Let the barrier be moved in such a way that the area of the film is increased by 1 sq cm, and that of the water decreased by this amount.

The equations for *spreading* are:

$$dF_s = (\partial F/\partial \sigma_f)_{T,\,\Sigma} d\sigma_f - (\partial F/\partial \sigma_w)_{T,\,\Sigma} d\sigma_w = -(\gamma_w - \gamma_f) d\sigma_f = -\pi d\sigma_f, \tag{70}$$

where $\Sigma = \sigma_w + \sigma_f$.

The increase of entropy which accompanies the spreading of the film over the water is from Eq. (50):

$$s_s = (\partial S/\partial \sigma_f)_{T,\,\Sigma} = -(\partial^2 F/\partial T \partial \sigma_f)_\Sigma = -(\partial \pi/\partial T)_{\sigma_f,\,\Sigma}. \tag{71}$$

The heat absorbed is

$$q_s = Ts_s \tag{72}$$

and the increase in heat content

$$h_s = (\partial H/\partial \sigma_f)_{T,\,\Sigma} = -\left[\frac{\partial(\pi/T)}{\partial(1/T)}\right]_{\sigma_f,\,\Sigma} \tag{73}$$

Equations (65) to (69) which give the energy and entropy of *extension* for the films are also applicable to pure liquids.

Thus using Eq. (68) Harkins calculated the increase of the heat content when the surface area of water is increased, as $h = 116$ ergs per cm^{-2} at 20°.

In Eq. (59) π is defined as

$$\pi = \gamma_w - \gamma_f, \tag{74}$$

and it can be shown that in Eq.

$$s_s = s_e + s_w \tag{75}$$

and in Eq. (67),

$$h_s = h_e + h_w. \tag{76}$$

The energy changes of the greatest interest are those associated with the isothermal expansion (or compression) of a film between any molal area σ_1, and any other molal area σ_2. The heat content and heat absorbed may be calculated for either the spreading or the extension of the film. For one mole of material the increase of heat content is

$$\Delta H_m = \int_{\sigma_1}^{\sigma_2} (\partial H/\partial\sigma)_T d\sigma_m = \int_{\sigma_1}^{\sigma_2} h d\sigma_m, \tag{77}$$

and the heat absorbed:

$$Q_m = T\int_{\sigma_1}^{\sigma_2} (\partial S/\partial\sigma)_T d\sigma_m = T\int_{\sigma_1}^{\sigma_2} s d\sigma_m, \tag{78}$$

where ΔH_m and h of Eq. (77) and Q_m and s of Eq. (78) stand for $(\Delta H_m)_e$, h_e, $(Q_m)_e$, and s_e when the process is one of extension, and for $(\Delta H_m)_s$, $(Q_m)_s$, and s_s when the process is one of spreading.

The heat absorbed may readily be calculated from ΔH and ΔF of any process. For *spreading*:

$$(Q_m)_s = (\Delta H_m)_s - (\Delta F_m)_s = (\Delta H_m)_s + \int_{\sigma_1}^{\sigma_2} \pi d\sigma_m. \tag{79}$$

For extension:

$$(Q_m)_e = (\Delta H_m)_e - (\Delta F_m)_e = (\Delta H_m)_e - \int_{\sigma_1}^{\sigma_2} \gamma d\sigma_m. \tag{80}$$

Table 13a. Energy of Spreading (h_s) of Two-dimensional Phases on an Aqueous Subphase at pH 2. (Values in erg cm^{-2})

Phase	*n*-Long Chain Acids
S	200 to 300
L_c	0
I	0 rising to from 260 to 300, then constant
L_e	50
n-Long Chain Alcohols	
S	600 to 1630
LS	76 to 750
L_c	36 to 220

Energy Relations in Transformations from Three to Two-Dimensional Systems

Heats of Spreading and Freezing. By the application of a Clapeyron equation in the form

$$\Delta H = \lambda_{S\rightarrow F} = T\left(\frac{\partial \pi_e}{\partial T}\right)(a_F - a_S)$$

it is possible to determine the latent heat of spreading $\lambda_{S\rightarrow F}$ of a solid S to form a film F upon an aqueous subsolution. Here π_e is the film pressure at equilibrium between the crystals and the film and a the area of the surface involved. In practice a_s may be made sufficiently small to be negligible.

Since the values of the temperature required to give the same molecular area at

equilibrium in a monolayer of either the 14 or 15 carbon atom acid are not very different, it is apparent that the differences in the latent heats of spreading, at any given value of a_F, are almost entirely determined by the value $\partial\pi_e/\partial_T$ of the slope of the π_e, T curve.

Values of these slopes have been determined by Cary and Rideal,* who find $\partial\pi_e/\partial T$ to be practically the same for both myristic and pentadecylic acid, that is, the π_e,T curves are parallel. The values of the latent heats of solidification or "settling" of these acids have been determined by King and Garner,† and these are, in kcal per mole, 10.7 for the even (14 C) and 10.30 for the odd (15 C) acid if the solid is in the α-form. The transformation $\alpha \rightarrow \beta$ of the odd acid evolves 1.20 kcal per mole in addition.

Thus the tightness of binding of the molecules of the even (14 C) acid may be said to be 10.7 kcal per mole greater in the crystals than in the three-dimensional liquid,

Table 13b. Latent Heat of Spreading of Myristic Acid on an Aqueous Subsolution of pH 2.0 ($\partial\pi_e/\partial T = 0.662$) ($\lambda_{s\rightarrow L} = 10.7$ kcal per mole)

Area per Molecule at Equilib. Press. (Å²)	T (°C)	Equilib. Press. (Dynes per cm)	Entropy (S), (Ergs per Molecule per °C × 10^{-16})	Entropy (S), Cal. per Mole per °C	Latent Heat of Spreading (Kcal per Mole)	Internal Latent Heat (Kcal. per mole)	State of Film
22.3	12.5	9.8	14.7	21.3	6.07	5.79	Condensed
23.4	14.8	11.4	15.4	22.3	6.46	6.08	Intermediate
23.6	9.5	7.7	15.5	22.5	6.40	6.13	Condensed
25.7	16.6	12.5	16.9	24.5	7.14	6.69	Intermediate
28.4	25.0	18.1	18.7	27.1	8.12	7.38	Expanded
28.9	21.7	15.9	19.0	27.6	8.17	7.51	Expanded

Table 13c. Values of the Constants in the Equation

Acid	Below a	M.P. b	Above a	M.P. −b	M.P. (°C)
Myristic	1.62	0.65	45.2	0.172	53.9
Pentadecylic	6.60	0.582	47.3	0.208	52.3
Palmitic			45.3	0.188	61.8
Margaric			42.7	0.137	60.6

while for the odd (15 C) acid the value is 10.3 or 11.5, as has been stated. The effects of the small changes of volume are neglected.

Energy values for myristic acid, as determined by the use of a film balance of the horizontal type are shown in Table 13b. The value of the slope $\partial\pi_e/\partial T$ was found by Harkins and Nutting to be 0.662, but later work by Harkins and Copeland by the ring method gave 0.650. That obtained by Cary and Rideal was 0.568.

Fig. 49b shows the increase of the film pressure (π_e) of the monolayer in equilibrium with crystals of palmitic acid rises rapidly with the temperature to 33.7 dyne cm^{-1} at 61.8° C, where the crystals melt, and then decreases as soon as the crystals change to liquid lenses. Below this temperature the spreading process is accompanied by absorption and above it by emission of heat.

With margaric acid the amount of heat emitted at 61° C by spreading from a liquid lens is about 1.66 kcal $mole^{-1}$. It is obvious that the difference in sign of ΔH between crystals and lenses has its origin in the latent heat of fusion of the crystals.

Increase of Area of a Monolayer Due to Salt in the Subphase

That ions of a polyvalent metal in a subphase of high pH often convert the liquid condensed monolayer of a long-chain acid into a solid monolayer has been shown previously. However, this is a purely chemical effect. Harkins and Morgan[25]

* Cary, A., and Rideal, E. W., *Proc. Roy. Soc. London*, **A109**, 328 (1925).
† King, A. M., and Garner, M. E., *J. Chem. Soc.*, 578 (1931).

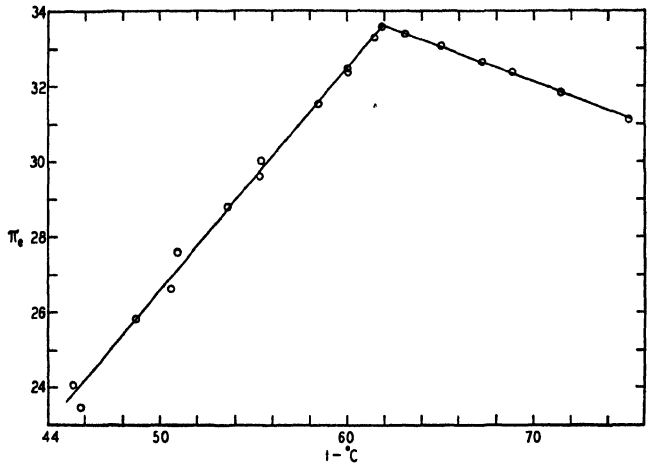


FIGURE 49b. Effect of Temperature on the equilibrium pressure of crystals ($t = 44°$ to 61.8° C) and of Lenses ($t = 61.8$ to 73° C) of palmitic acid.

showed in 1925 that when chemical effects are absent the effect of a salt, which increases the surface tension of water, is to increase the area of the monolayer. Phenanthrene, dibromanthracene, etc., which do not give a film on water, were found to give one on concentrated calcium chloride solution. These films were evidently monolayers studded with small three-dimensional crystals.

Since the surface tension of an 11.2 molal calcium chloride solution at 25° C is about 107 dyne cm^{-1}, or 35 dyne cm^{-1} greater than that of water, the pull on the film-covered surface by that of the clean solution is about 50 per cent greater than that of the surface of water. A very closely packed monolayer of oil reduces the surface tension of the salt solution very greatly. Thus, since

$$\pi = \gamma_s - \gamma_f$$

where γ_s is the surface tension of the salt solution, the value of π in such a case must be high.

Figure 50, from upublished work by Harkins and Copeland, represents M NaCl + 0.01M HCl as the subphase on a monolayer of octadecyl alcohol. This has a very much smaller effect than 11.2M $CaCl_2$, since it increases the surface tension of water by only 8.5 dyne cm^{-1}. However, it may be noted that at the lower film pressures the molecular area is about one $Å^2$, or nearly 5 per cent larger than on water. The presence of the salt was found to increase the pressure of collapse of the film to a very high value (about 61 dyne cm^{-1}, or about 250 atmospheres).

This figure shows, for the first time, three condensed phases on a single isotherm (7.42° C). The presence of the salt raises the pressure of transition, from the phase (L_c) of greatest compressibility to that with an intermediate value, from 15.7 to 20.0 dyne cm^{-1}, while the transition from this to the least compressible phase occurs at 45.3 dyne cm^{-1}.

Isotherms at 25° C for octadecylamine chloride on $10^{-2}N$ HCl were found by Hoffmann and others (in this laboratory)* to exhibit a considerable increase of area when the subphase was made 0.01N with respect to HBr, and about as much again when this was replaced by HI. These relations were reversed below about 4 dyne cm^{-1}. The use of hydrogen iodide eliminated the second-order transition between the expanded and intermediate phases.

* Hoffman, E. J., Boyd, G. E., and Ralston, A. W., *J. Am. Chem. Soc.*, **64**, 501 (1942).

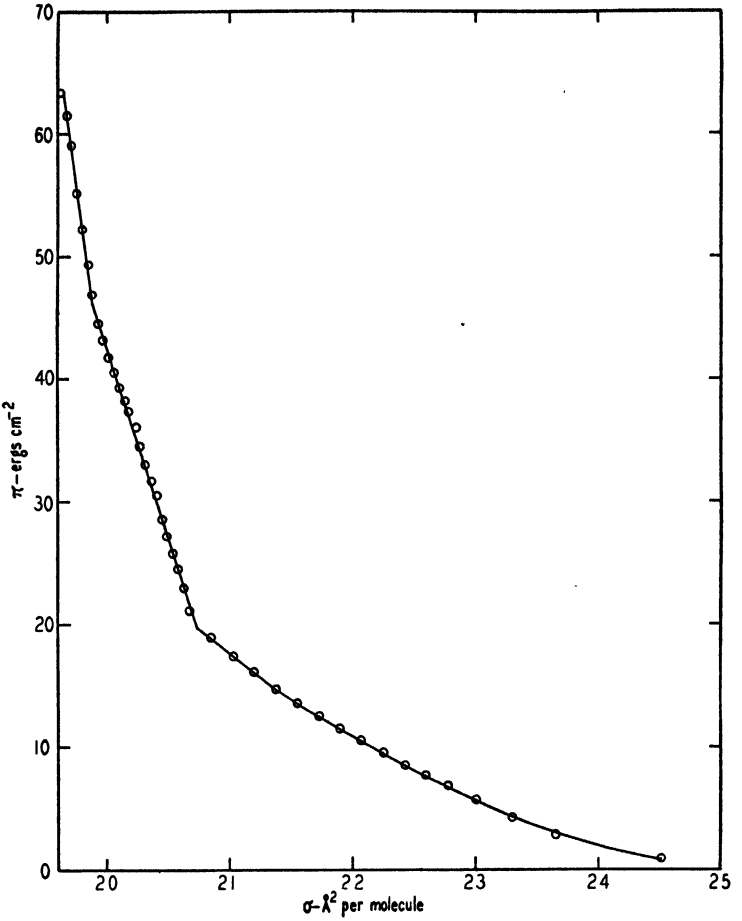


FIGURE 50. A graphic representation of M NaCl + 0.01M HCl as the subphase in a monolayer of octadecyl alcohol.

THE DEPOSITION OF MONOLAYERS TO FORM MULTILAYERS

It has been found by Blodgett and Langmuir * that calcium stearate and many other monolayers may be transferred from the surface of an aqueous solution to that of a solid. At certain high values of the pH of the solution the deposition on the solid occurs only on the down trip (X films), and at certain lower values on both the down and up trips (Y films), and in some instances only on the upward movement (Z films).

It is claimed by Bikerman † that strongly hydrophobic surfaces pick up monolayers on the down trip only. The relations are exhibited by him as in Fig. 51.

Blodgett and Langmuir ‡ measured the thickness of multilayers by the use of

* Blodgett, K. B., *J. Am. Chem. Soc.*, **57**, 1007 (1935); Blodgett, K. B., and Langmuir, I., *Phys. Rev.*, **51**, 964 (1937).

† Bikerman, J. J., *Proc. Roy. Soc. London*, **170A**, 130 (1939).

‡ Blodgett, K. B., *Phys. Chem.*, **41**, 975 (1937); Blodgett, K. B., and Langmuir, I., *Phys. Rev.*, **55**, 391 (1939).

interference colors with polarized white light, or by interference intensities with monochromatic light. Barium stearate films were found to be 24.4 Å thick per molecular layer.

In this laboratory this method was used to determine the thicknesses of multilayers of catalase [88] produced by another method. Each monolayer was adsorbed

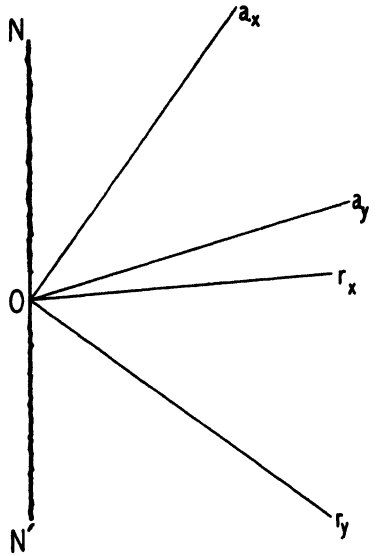


FIGURE 51. Illustrates Bikerman's theory, according to which if surface of slide is highly hydrophobic, both advancing and receding angles (between a_x or r_x and the slide) in the water are greater than 90°. If less hydrophobic the advancing angle (for a_y) is greater than 90° but the receding (up trip) is less (r_y). On withdrawing the slide deposition occurs for r_yON' but not for r_xON', but for the down trip for both a_xON and a_yON, since all these angles are acute and allow the monolayer and the surface of the slide to approach each other during the movement of the slide. Thus the hysteresis of the contact angle is involved.

from a catalase solution upon a slide. The layers in order of their adsorption, and their mean thicknesses in Å were: (1) catalase, 53; anticatalase, 56; catalase, 10; anticatalase *ca.* 45; catalase *ca.* 10; etc. Thus catalase on anticatalase gave a thin layer, but catalase on barium stearate or anticatalase on catalase a thick one.

Porter and Wyman * found that multilayers of the Y type deposited on a slide of metal give by the method of Zisman, a very low, *almost zero,* apparent contact potential, while the potential exhibited by X multilayers is nearly proportional to the number of layers, and of the order of 60 mv per layer (See Figs. 51 to 54).

According to Langmuir's first idea, this potential was supposed to be due to the dipoles of the molecules of the multilayer. Later he and Harkins and Matton † attributed the apparent contact potential to the presence of charges.

Charge Theory of Multilayer Potentials [68, 75]

The theory of these potentials, as developed by Harkins and Mattoon, is given below.

* Porter, E. F., and Wyman, P., *J. Am. Chem. Soc.*, 59, 2746 (1937); 60, 1083 (1938).
† Langmuir, I., *J. Am. Chem. Soc.*, 60, 1190 (1938); Harkins, W. D., and Matton, R. W., *Phys. Rev.*, 53, 911 (1938).

Let it be supposed that when any single X layer is deposited on the down trip of the slide it collects no charge, but that on the upward movement, when no film is deposited, the outer surface of the X layer takes up a positive charge. This would produce an induced charge in the surface of the metal on which the layer is de-

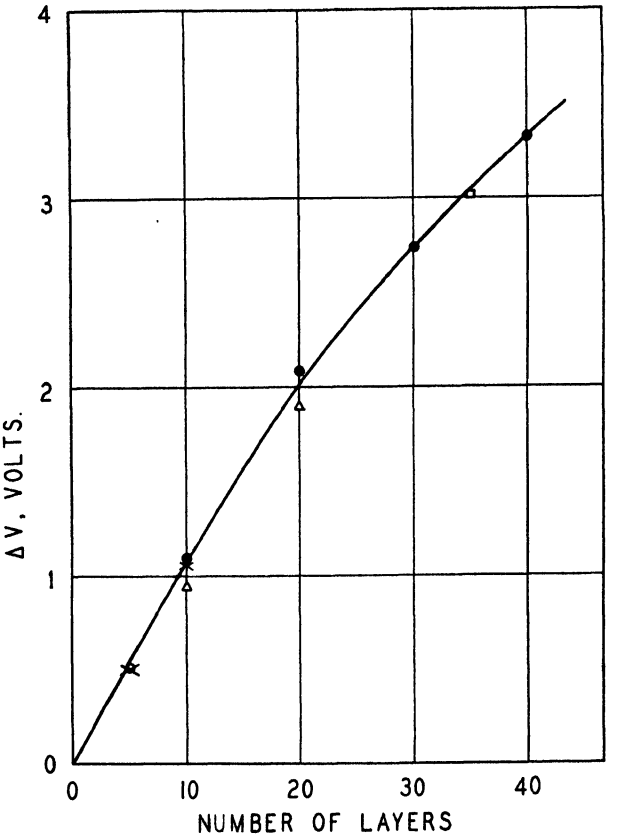


FIGURE 52. Film potentials of five different X multilayers on gold made from solutions 1(0, X), 2(△, □), and 3(●). The potential increment is about 100 mv/layer up to 20 layers.

posited. The general Gauss-Poisson equation for the potential is $\Delta^2 V = -4\pi neT/\epsilon$ for a condenser which consists of two equally charged parallel plates, in which n is the number of unit charges per sq cm, e is the electronic charge, T is the total thickness of the dielectric (X layers, paraffin, etc.), and ϵ is the dielectric constant of the dielectric under the plane of the charge. Thus if all the charge of a multilayer is associated with the *outer* layer only,

$$\Delta V = 4\pi neNt/\epsilon, \tag{1}$$

where t is the constant thickness of each X layer, and N is the total number of X layers. Thus the potential is proportional to the number of layers, as is nearly true (Figs. 52 and 53).

If, however, an equal charge is deposited on each X layer and persists, then the potential will rise much more rapidly than with the first power of the thickness, which

is contrary to a part of the experimental results presented later. The relation for equal charges retained on the surface of every layer is

$$\Delta V = 4\pi net N(N+1)/2\epsilon, \tag{2}$$

where ne is the surface density of the charge, t is the thickness of each X layer, and N is the total number of X layers. Thus if one X layer increases ΔV by 0.1 volt, then 20 X layers would give a potential of 21 volts. Experimentally, 20 X layers give

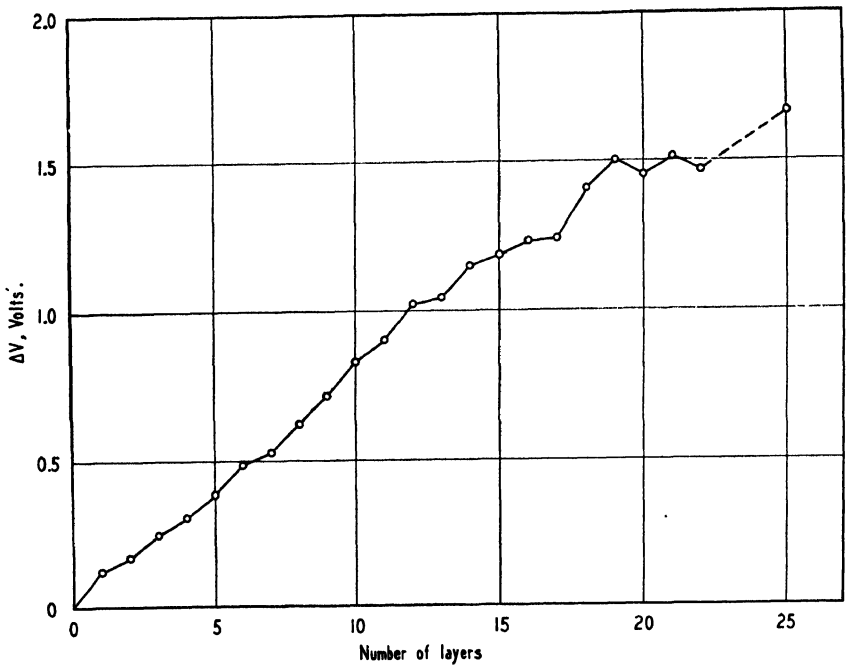


FIGURE 53. Film potential for each of 22 single layers of an X film on gold. Average slope 75 mv per layer.

a potential of about two volts. The relations are modified if X layers are deposited upon a dielectric which covers the metal. Two different distributions of charge are considered below.

(1) A definite charge, whose value is independent of the thickness of the initial dielectric and of the X multilayer, is present on the outside layer only. For simplicity the thickness (T) of the dielectric is represented by $T = Mt$, where t is the average thickness of the X layers in a multilayer. The dielectric constant of the dielectric is ϵ_D and that of the X multilayer ϵ. Then

$$\Delta V = 4\pi net(M/\epsilon_D + N/\epsilon). \tag{3}$$

(2) A charge of a certain constant value remains upon every layer of the X multilayer. Then

$$\Delta V = 4\pi net\left[\frac{NM}{\epsilon_D} + \frac{N(N+1)}{2\epsilon}\right]. \tag{4}$$

It is evident that if each X layer is to retain its charge and if the multilayer is deposited directly on the metal, the charge must decrease rapidly with the distance from the metal if the film potential is to be proportional to the number of X layers.

Thus to meet this condition the charge densities on successive layers must be $\rho/1$, $\rho/2$, $\rho/3$, . . . ρ/N.

The difference in the potential increments according to Eqs. (3) and (4) are illustrated by listing the potentials calculated for the case in which the thickness of the initial paraffin layer is equal to that of 100 X layers and the potential increment by Eq. (1) is 0.1 volt per layer. For simplicity the dielectric constants of the paraffin and the X multilayer are supposed to be the same. Actually that of the paraffin is somewhat lower (Fig. 54).

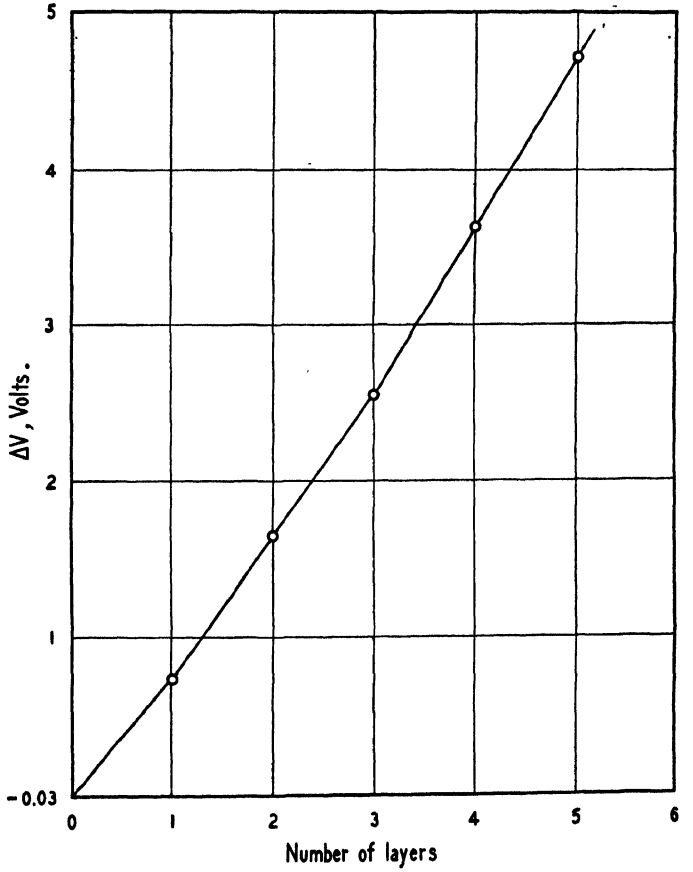


FIGURE 54. Film potential of an X multilayer on a moderately thin sheet of paraffin on gold. Average slope 900 mv per layer.

The way in which the charges are built up on X and Y multilayers is shown by work in this laboratory.[77]

The method used seems to offer promise of revealing much that occurs in the production of a film potential when X or Y multilayers are built upon a metal at various values of pH of the subsolution utilized. This consists in measuring the electric current, or charge, which passes between the metal slide upon which the film is deposited, during its up and down trips, and another (comparison) electrode of the same metal (gold), immersed in the solution. In using this method there are several difficulties which must be kept in mind.

(1) If one electrode has been in the solution for a long time (*e.g.* ½ hour), and the other is now dipped into the liquid, a relatively large current flows between them during the first few seconds, even after the slide has ceased its motion.

(2) If both electrodes have been (partly) immersed for a long time, a small current usually flows, probably because there are small differences in the metal surfaces or the adjacent solutions.

(3) A current passes between the electrodes while one of them is being dipped into the solution through a clean surface. During the down trip this flows in the opposite direction from that during the up trip. This current is due to the constant galvanic potential between the metal electrode and the solution, and the change in capacity during the dipping.

In the experiments current (2) is made zero by adjusting the depth of immersion of the comparison electrode. The disturbing current (1) disappears if both electrodes are immersed for a long period, or if several layers have been deposited on the slide, thus insulating it from the solution. The disturbing current (3) also decreases with the number of layers since the capacity between the electrode and the solution also decreases.

If it is desired to study the currents due to the first few layers, both electrodes should be dipped synchronously: one through the monolayer on the subsolution, and the other through a clean surface of the latter.

The first experiments consisted of dipping clean gold through the clean surface of a solution buffered at pH 9.4. Apart from currents (1), which become very small after the first few dips, there first appear, during both the up and down trips, currents of about the same magnitude but in opposite directions for the two. These currents show that the potential of the gold electrode is negative with respect to the solution. No currents flow if the slide emerges wet, as it may if the upward movement is very quick. This is because the electric double-layer of the gold and solution is removed with the wet slide, and is not disturbed as when the gold emerges dry.

During the deposition of X films at a pH of 9.4, there appear, apart from currents (1), quite the same currents as are produced by dipping clean gold through a clean surface of the solution, but the currents become smaller and change sign as the dippings increase. This reversal of sign takes place when the X film has attained a potential of about 0.4 volt. After this the currents have almost a constant value, or decrease slightly with the number of layers.

During the up trip the positive current flows from the comparison electrode through the solution to the X multilayer, and in the opposite direction on the down trip.

An X multilayer, with a positive charge on the outside layer may be dipped into the solution, when it is found that if the film potential before the dipping was greater than about 0.5 volt, say 2 volts, a positive current flows through the solution from the multilayer to the gold of the comparison electrode. If the system X multilayer-gold is now removed from the solution it exhibits a decrease of film potential, the magnitude of which increases with the time of immersion. Thus the current during the down trip with X deposition, has the same direction as the discharging current specified here.

The total charge transferred during the down or up trip is of the same order of magnitude, as closely as could be calculated from the preliminary data, as the charge on the outer surface of the multilayer would need to be to give the measured film potential.

From the above it is not necessary to conclude that the film is completely discharged during the down and charged again on the up trip. It is possible that a double layer is formed between the surface of the film and the solution. In any

event the potential with respect to the solution is reduced when the multilayer-metal system is in the solution.

If the slide is allowed to remain at rest at its deepest immersion, no current flows when the number of X layers is small, but with a larger number of layers a very small discharge current is observed. The increment of potential per layer of

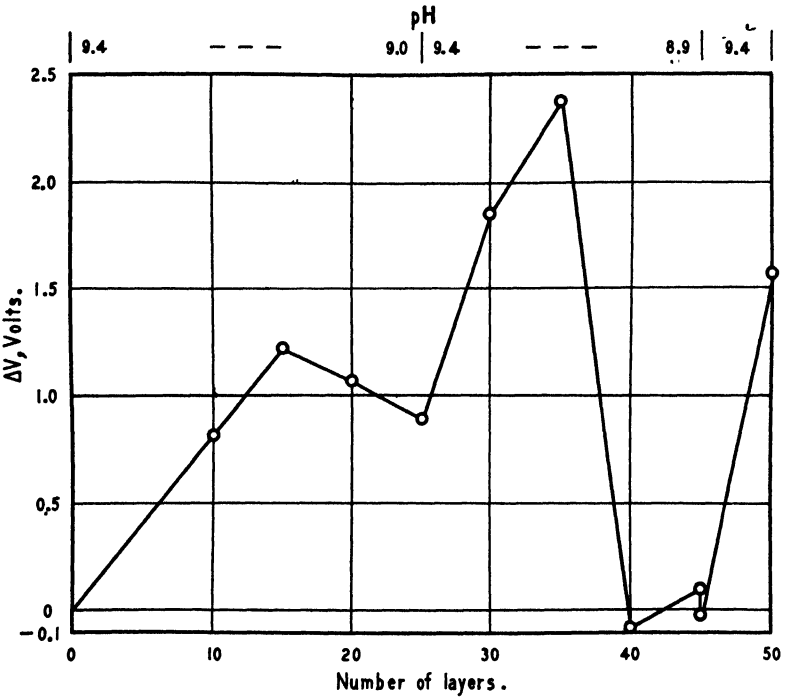


FIGURE 55. Effect of pH on film potentials.

an X multilayer is known to decrease with the number of layers, so if the potential is due to a charge on the outermost layer only, this charge must also decrease at the same rate. It is possible that this is due to a partial discharge of the film during its building, through the action of the small currents just described.

In the production of Y films at pH 7.0 the relations are different. During the down trip the current always flows through the solution from the comparison electrode to the slide plus multilayer, and in the opposite direction during the up trip, until, usually, the film potential attains a value of about 0.2 volt. After this the potential does not increase, since now both currents fall to almost zero.

INTERMOLECULAR FORCES AT SURFACES AND INTERFACES

The term "monolayer" and what is known as the monomolecular theory have led to certain false points of view according to which the energy of the film is restricted to the single layer of molecules in the oil film itself. In Volume VI of this series, in the chapter on solids, it will be shown that the attraction of anatase (TiO_2) for water remains appreciable through all of the five first layers of water molecules, although the energy falls off almost exponentially with the distance.

The attractive energy between the molecules in the monolayer, and between these and the water are of many different types, of which the dispersion and dipole forces are the most important in the more usual types of insoluble films.

Whenever there is homogeneous binding in a solid or a liquid body the general energy equation may be written

$$E = -aV + bA,$$

in which a and b are constants, V is the volume of the body, and A is the area of the surface. At the surface there are so few bonds that the binding energy is less.

The types of bodies to be considered are:

(1) Liquids
- (a) Non-polar
- (b) Semi-polar
- (c) Polar
- (d) Metallic

(2) Solids
- (a) Ionic Crystals (as salts)
- (b) Metals
- (c) Semi-conductors
- (d) Valence Crystals (diamond, carborundum)
- (e) Molecular Crystals

The energy of molecular attraction for parts of molecules may be expressed in terms of the potential (V) as outlined in a simplified form:

(1) Ion-Ion: $V = \frac{e_1 e_2}{R}$, a scalar

$$F = \Delta V = \frac{e_1 e_2}{R^2}, \text{ a vector.}$$

(2) Ion-Dipole: $V = \frac{e\mu \cos\theta}{R^2}$, where μ is the dipole moment.

(3) Dipole-Dipole (Head-on) $= \frac{2\mu_1 \mu_2}{R^3}$

General:

$$V = \frac{\mu_1 \mu_2}{R^3} (\text{Cos } \gamma - 3 \text{ Cos } \theta_1 \cdot \text{Cos } \theta_2),$$

where γ is the angle between the vectors of the dipole moments.

(4) Atom-Ion: $V = \frac{aze^2}{2R_4}$

(5) Dipole-Dipole Polarization or Dispersion Forces:

$$V \propto \frac{1}{R^6}$$

(6) Ion-Metal (Mirror Image): $V = \frac{e^2}{4a}$

(7) Dipole-Metal (Mirror Image):

$$V = \frac{\mu^2}{16a^3} -2 \cos\theta_1 \cos\theta_2 + \sin\theta_1 \sin\theta_2 \cos(\theta_1 - \theta_2),$$

or for a dipole directed toward the surface: $V = -\frac{\mu^2}{8a^3}$.

R = distance, a = distance from surface, e = electronic charge, V = dipole moment, α = polarizability, z = valence, F = force, θ = angle.

Since the dipole-dipole interaction energy varies as the cube of the distance, the

effect is still appreciable, in the case of films of alcohols, acids, and others of this type, in the second layer of water molecules.

A more pronounced effect is the orientation effect of the dipoles of the film upon those of the water.

The Hydrogen Bond in "Monolayers"

The energy of molecular interaction in many "monolayers" which contain hydrogen in the polar group, and between the molecules of the "monolayer" and the aqueous subphase, is considerably increased by a special type of interaction between the hydrogen and two other atoms. This interaction is commonly designated by the term "hydrogen bond."

The effect of this bond upon the structure of the surface region is very great, since it gives a bond between molecules which is highly directional, and in this sense, though its energy is much lower, resembles to some extent a covalent bond. The depth of the surface region may be considered to be the thickness of all of that layer in which either the intermolecular energy or the structure is different from that in the interior of the subphase itself. The depth to which the energy is different is very small, but certainly more than one molecular layer of water is affected. Thus

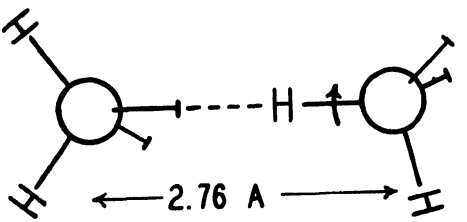


Figure 56. Two molecules of water held together by a hydrogen bond. The dashes represent a region of negative charge. Free rotation may occur about the hydrogen bond. The $O\begin{smallmatrix}H\\H\end{smallmatrix}$ angle is 105°.

the surface region is several molecules in thickness, since it includes the monolayer of oil and several layers of water.

The hydrogen bond was first recognized by Moore and Winmill, and later by Pfeiffer, and the application of the idea was widely extended by Latimer and Rodebush and by Huggins.*

The hydrogen bond has a relatively small energy, which is in calories per mole: alcohol 6.2, ice 4.5, acetic acid 8.2, benzoic acid 4.3, etc. The strength of hydrogen bonds between atoms decreases in the order: fluorine, oxygen, nitrogen, chlorine. The strength of the bond should increase with increase in the electronegativity of the atoms, and with their size.

Water may be assumed to be very nearly tetrahedrally coördinated as in the model of Bernal-Fowler † (Fig. 56). Free rotation but no bending at the hydrogen bond may be assumed, with an angle for $O\begin{smallmatrix}H\\H\end{smallmatrix}$ of 105°.

* Moore, T. S. and Winmill, T. F., *J. Chem. Soc.*, **101**, 1635 (1912); Pfeiffer, P., *Ann.*, **398**, 137 (1913); Latimer, W. M. and Rodebush, W. H., *J. Am. Chem. Soc.*, **42**, 1419 (1920).

† Bernal, J. D., and Fowler, *J. Chem. Phys.*, **1**, 515 (1933).

In alcohol (ROH) the bond angle (Fig. 57) may also be taken as 105°. The extent of molecular orientation is dependent upon the location of the charge distribution in a polar molecule. Thus in nitro derivatives and nitriles the charge distribution is much more shielded than when —OH groups are involved, and the effect is much smaller in comparison with the magnitude of the dipole moment.[100]

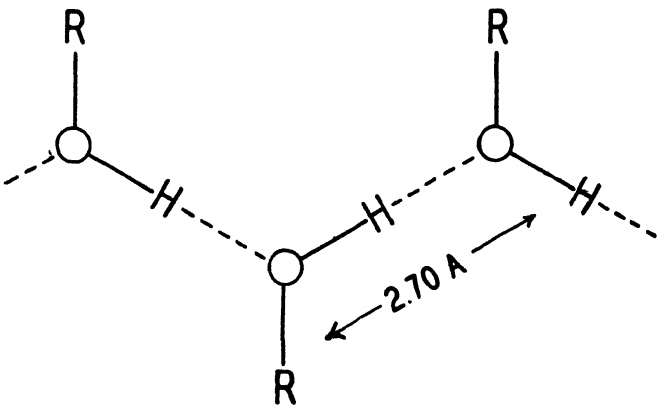


FIGURE 57. Hydrogen bonds in alcohol.

If ice is melted only about 15 per cent of the hydrogen bonds are ruptured, but if it is sublimed all of these bonds are broken. Of the resultant heat of sublimation, 12.2 kcal mole^{-1}, only about one-fourth can be attributed to van der Waals forces, and about 9 kcal mole^{-1} to the rupture of hydrogen bonds. The higher hydrogen bond energy in alcohol (6.2 kcal mole^{-1}) as compared with water (4.5) seems to be related to the smaller (2.70 Å) O—H—O distance in alcohols as compared with that in water (2.76 Å).

In formic acid, a simple carboxylic acid, this distance is still less (2.67 Å). In acetic acid the bond energy is 8.2 kcal mole^{-1}. In water the hydrogen bonds do not have a sufficiently high energy to give an appreciable concentration of polymeric molecules in the vapor phase, but the stronger hydrogen bonds of formic and acetic acids give stable double molecules which vaporize as such.

In Volume I, pages 222-6, of this series the writer showed that the value of h/λ for acetic acid (0.336) is much higher than for methyl (0.164) or ethyl (0.186) alcohol. Here h is the total energy absorbed when the molecule moves from the interior of the liquid into the surface, and λ that absorbed when the molecule moves from the interior of the liquid into the vapor. The values refer to a corresponding temperature of 0.7. In the case of the alcohols all of the hydrogen bonds must be broken in the process of vaporization, while this is not true for acetic acid, which makes λ much higher relatively for the alcohols. Since when the molecules move into the surface no hydrogen bonds need to be broken, the value of h/λ should be much smaller for the alcohols, as the data indicate to be true.

In the —SH group of the mercaptans the proton is buried more deeply in the electronic atmosphere of the sulfur atom than in that of the oxygen atom of —OH. Thus the hydrogen bonds in the monolayer of a mercaptan, and between the —SH groups and the oxygen of the water have a considerably lower hydrogen bond energy than in the case of the otherwise similar alcohols. This is made evident by the difference in the magnitudes of the work of adhesion (W_A) at 20° C. Thus to pull an alcohol from the surface of water requires about 92 erg cm^{-2} of work, while for a mercaptan the value is only 75 per cent as large. With nitriles the very much

higher value of the dipole moment (3.6×10^{-18} instead of 1.7×10^{-18} for alcohols) keeps the work of adhesion toward water about as high as with the alcohols.

An interesting paper by Alexander,* gives somewhat definite pictures of the hydrogen bonding in the monolayer and between the monolayer and the water. He concludes that intra-monolayer bonding in alcohol films cannot occur, but his ideas postulate much less flexibility in the alcohol chains and more exactness in orientation than seems probable. It seems preferable to assume that condensed alcohol monolayers are held together by (1) van der Waals forces between the chains, (2) hydrogen bonding between the polar groups in the monolayer, and (3) hydrogen bonding between an alcohol and a water molecule. This may then be bonded to either a second alcohol or another water molecule. At low film pressures alcohols give the smallest molecular areas of any of the substances thus far investigated, although some other substances such as amines on a basic subphase, have given rather low areas but have not been investigated with sufficient exactness to determine just what these areas are.

Adsorption

The section on Liquids is not complete without a consideration of the extremely important subject of adsorption. However, the thermodynamic theory of adsorption in liquids is the same as that when solids are involved, so both types of systems should be treated together. For this reason, adsorption will be treated in Volume VI of this series, after the Surfaces of Solids have been discussed.

On account of the importance of adsorption isotherms, and of their use in the determination of area, a bare outline of a subject which will be treated fully in Volume VI, is given in the three following sections.

A Method for Determining the Absolute Area of a Very Finely Divided Solid

Several important advances, which may be utilized in the determination of the area of either a crystalline powder or a porous solid, have resulted from the work (unpublished) of Dr. George Jura and the writer. These are described very briefly in this and the two following sections.

If a finely divided powder is placed in the saturated vapor of a liquid, it becomes covered with a film adsorbed from the vapor. If the liquid exhibits a zero contact angle with the solid, the adsorbed film at equilibrium is duplex (see p. 15), that is, the surface energy of the outer surface of the film is exactly that of the pure liquid. For example, if the vapor is that of water, the surface energy of the outer surface of the film is 118.5 erg cm^{-2} at 25° C. If the powder is now dropped into the pure liquid contained in an extremely sensitive calorimeter, heat is liberated, and the total amount of this heat Q, in ergs, is equal to the total area (Σ) of the powder multiplied by the surface energy of the liquid (118.5 erg cm^{-2} for water).

Thus
$$Q_{ergs} = 118.5\,\Sigma$$

or
$$\Sigma = \frac{Q_{ergs}}{118.5}\,cm^2$$

If q is the heat liberated per gram of powder, then

$$\Sigma = \frac{q}{118.5}\,cm^2\,g^{-1}$$

This method, applied to a sample of titanium dioxide (anatase) pigment, gave an area of 13.8 square meters per gram. The area of this powder, as calculated by the method of Emmett, Brunauer, and Teller, from an adsorption isotherm of nitrogen, is 13.9, or almost the same. The result of the calculation by their theory involves

* Alexander, A. E., *Trans. Faraday Society*, 37, 426 (1941).

the assumption that the mean area occupied by a nitrogen molecule is 16.2 Å^2, while our method involves no such assumption.

In our work, each small crystal becomes coated with about five molecular layers of water (or ten of nitrogen) and we have already measured the effect of distance upon the energy of attachment of the liquid to the solid. The energy of vaporization of water is 10,480 cal mole^{-1}, and for successive layers on anatase, the values are higher than this in the successive layers as follows: (1) 5400, (2) 500, (3) 110, (4) 60, (5) 20, and all beyond 5 about 30 cal mole^{-1}. Thus it is obvious that with the duplex film, no error is due to the effect of the solid.

In practice, a correction is made for the thickness of the water film, since this causes the area of the film coated crystals to be slightly greater (less than 1 per cent) than that of the solid. The thickness of the water film is given by the weight of water adsorbed, divided by the area.

An Adsorption Isotherm with a Wide Range of Validity

The equation which represents the pressure-area relation of a condensed monolayer, either liquid or solid, on water, is

$$\pi = b - a\sigma \tag{1}$$

Here π is the film pressure, σ is the molecular area in Å^2, and a and b are constants.

We have found that this *same equation* is valid for adsorbed films or solids. The *very remarkable* fact which emerges is that the equation is valid, not only for the condensed monolayer, but also when the film grows to 2, 3, 4, 5, and even more monolayers in thickness, *i.e.*, both for polymolecular films and condensed monolayers without a change of the values of the constants.

The application of this equation to a large number of solids involves a considerable amount of calculation, since the film pressure π must be found by an integration of the Gibbs adsorption equation.

On this account the equation was transformed into the following equivalent relation:

$$\log p = \frac{-A}{v^2} + \tag{2}$$

in which p is the pressure of the vapor which is in equilibrium with the surface of the solid, and v is the volume of gas, at standard pressure and temperature, which is adsorbed per gram of solid.

This is the simplest adsorption isotherm thus far discovered which is valid over any considerable range and it seems to be valid over a greater range of pressures than any other isotherm. Comparisons have been made for 65 solids, of areas from 2 to 600 square meters per gram.

The Langmuir isotherm is not valid for any of these solids. The isotherm of Brunauer, Emmett, and Teller,* gives an excellent fit over a considerable range, while our new isotherm exhibits agreement between its linear form and the data over a much greater range of pressures, and especially to very much higher pressures. Thus with nitrogen on barium sulfate the data fit the linear relation up to a pressure of 722 mm, or to 0.95 per cent saturation.

* Their equation is:

$$\frac{p}{V(p_0 - p)} = \frac{1}{V_m C} + \frac{(C-1)_p}{V_m C_{p^0}}$$

which is much more complicated than our isotherm in Eq. (2).

A New Method for Determining the Area of a Porous or a Non-porous Solid Without the Assumption of a Molecular Area

Until 1938 there was no good method for the determination of the area of a crystalline powder and a porous solid. At that time, the method of Emmett, Brunauer, and Teller, described by Emmett elsewhere in this volume, became fully developed. Its principal disadvantage is that the application of their theory to an adsorption isotherm does not give an area, but only the number (N) of molecules in the first monolayer. In order to determine the area, it is necessary for them to use the relation:

$$\Sigma = N\sigma$$

in which σ is the mean area per molecule of the substance used, usually nitrogen. Emmett's value for σ, as calculated from the density of liquid nitrogen, is 16.2 Å^2, while from that of solid nitrogen he calculates 13.4. Since there is no basis for the choice of one of these values rather than the other, or of some other value of the same order of magnitude, a considerable uncertainty is introduced.

The New Method. Values are obtained at constant temperature for the amount of nitrogen (or water, etc.) adsorbed on one gram of solid at pressures (p) up to saturation.

The value of log p is plotted on the Y-axis, and that of $\frac{1}{v^2}$ on the X-axis. The square root of the slope of the straight line gives the relative area of the solid. The area (Σ) per gram of solid is given by

$$\Sigma = ks^{\frac{1}{2}} = kA^{\frac{1}{2}} \tag{3}$$

where k is a constant for nitrogen and s is the slope of the straight line determined from any single solid by the method just described.

The best value of k is, at present, 4.06, so that the area of any solid is given by:

$$\Sigma = 4.06\, s^{\frac{1}{2}} = 4.06 \frac{\Delta \log p}{\Delta \left(\frac{1}{v^2}\right)} \tag{4}$$

This simple equation has been used to calculate the area of 85 solids of areas of from 2 to 600 square meters per gram, and has in every case given an area within 10 per cent of that which may be obtained from the equations of Brunauer, Emmett, and Teller, by the assumption that nitrogen has the same area on all surfaces of solids.

However, a comparison of the two methods shows that the assumption that the area of the nitrogen molecule is constant is certainly invalid.

References

Papers on the Physics and Chemistry of the Surfaces of Liquids and Solids, 1916-1942, by William Draper Harkins, University of Chicago

1. "The Drop Weight Method for the Determination of Surface Tension" (with E. C. Humphery), *J. Am. Chem. Soc.*, 38, 228 (1916).
2. "Apparatus for the Determination of the Surface Tension at the Interface Between Two Liquids" (with E. C. Humphery), *Ibid.*, 38, 236, (1916).
3. "The Surface Tension at the Interface Between Two Liquids, and the Effect of Acids, Salts, and Bases upon the Interfacial Tension" (with E. C. Humphery), *Ibid.*, 38, 242, (1916).
4. "A Simple Apparatus for the Accurate and Easy Determination of Surface Tension, With a Metal Thermoregulator for the Quick Adjustment of Temperature" (with F. E. Brown), *Ibid.*, 38, 246, (1916).
5. The Structure of the Surfaces of Liquids, and Solubility as Related to the Work Done by the Attraction of Two Liquid Surfaces as They Approach Each Other" (with F. E. Brown and E. C. H. Davies), *Ibid.*, 39, 354, (1917).
6. "The Orientation of Molecules in the Surfaces of Liquids, the Energy Relations at Surfaces, Solubility, Adsorption, Emulsification, Molecular Association, and the Effect of Acids and Bases on Interfacial Tension (Surface Energy VI)," (with E. C. H. Davies and G. L. Clark), *Ibid.*, 39, 541, (1917).

7. "The Determination of Surface Tension (Free Surface Energy), and the Weight of Falling Drops: The Surface Tension of Water and Benzene by the Capillary Height Method" (with F. E. Brown), *Ibid.*, **41**, 499 (1919).
8. "An Electromagnetic Hypothesis of the Kinetics of Heterogeneous Equilibrium, and of the Structure of Liquids," *Proc. Nat. Acad. Sci.*, **5**, 152 (1919).
9. "An Electromagnetic Hypothesis of the Kinetics of Heterogeneous Equilibrium, the Structure of Liquids, and Cohesion" (with H. H. King), *J. Am. Chem. Soc.*, **41**, 970 (1919).
10. "The Effects of Acids and Bases on the Surface Energy Relations of β, β-Dichloroethylsulphide (Mustard Gas), *Ibid.*, **41**, 1977 (1919).
11. "The Change of Molecular Kinetic Energy into Molecular Potential Energy: The Entropy Principle and Molecular Association," *Proc. Nat. Acad. Sci.*, **5**, 539 (1919).
12. "Cohesion, Internal Pressure, Adhesion, Tensile Strength, Tensile Energy, Negative Surface Energy, and Molecular Attraction," *ibid.*, **5**, 562 (1919).
13. "The Adhesion Betweeen Mercury, Water, and Organic Substances, and the Fundamental Principles of Flotation," *ibid.*, **5**, 569 (1919).
14. "An Apparent High Pressure Due to Adsorption, the Heat of Adsorption, and the Density of Gasmask Charcoals" (with D. T. Ewing), *Proc. Nat. Acad. Sci.*, **6**, 49 (1920).
15. "The Orientation of Molecules in Surfaces, Surface Energy, Adsorption, and Surface Catalysis. V. The Adhesional Work Between Organic Liquids and Water" (with G. L. Clark and L. E. Roberts), *J. Am. Chem. Soc.*, **42**, 700 (1920).
16. "Surface Tension and Molecular Attraction: On the Adhesional Work Between Mercury and Organic Liquid" (with E. H. Grafton), *J. Am. Chem. Soc.*, **42**, 2534 (1920).
17. "Surface Energy of Mercury and the Energy Relations at the Interface Between Mercury and Other Liquids" (with E. H. Ewing), *J. Am. Chem. Soc.*, **42**, 2539 (1920).
18. "The Orientation of Molecules in Surfaces. VI. Cohesion, Adhesion, Tensile Strength, Tensile Energy, Negative Surface Energy, Interfacial Tension, and Molecular Attraction" (with V. C. Cheng), *Ibid.*, **43**, 35 (1921).
19. "A High Pressure Due to Adsorption, and the Density and Volume Relations of Charcoal" (with D. T. Ewing), *ibid.*, **43**, 1737 (1921).
20. "Vaporization in Steps as Related to Surface Formation" (with L. E. Roberts), *Ibid.*, **44**, 653 (1922).
21. "Films: The Spreading of Liquids and the Spreading Coefficient" (with A. Feldman), *ibid.*, **44**, 2665 (1922).
22. "Surface Energy in Colloid Systems" (in Colloidal Behavior, Ed. by Bogue, McGraw-Hill Book Co., New York, 1924).
23. "Monomolecular Films on Water: The Oriented Adsorption of Derivatives of Benzene" (with E. H. Grafton), *J. Am. Chem. Soc.*, **47**, 1329-1335 (1925).
24. "The Oriented Wedge Theory of Emulsions" (with N. Beeman), *Proc. Nat. Acad. Sci.*, **11**, 631-637 (1925).
25. "Polymolecular and Monomolecular Films" (with J. W. Morgan), *ibid.*, **11**, 637-643 (1925).
26. "Monomolecular Films Between Liquids: Butyric Acid Between Water and Hexane, and Acetic Acid Between Water and Benzene" (with H. M. McLaughlin), *J. Am. Chem. Soc.*, **47**, 1610-1613 (1925).
27. "The Effect of Sodium Hydroxide Upon the Surface Tension of a Solution of Sodium Nonylate" (with G. L. Clark), *ibid.*, **47**, 1854-1856 (1925).
28. "The Structure of Films of Water on Salt Solutions. I. Surface Tension and Adsorption for Aqueous Solutions of Sodium Chloride" (with H. M. McLaughlin), *ibid.*, **47**, 2083-2089 (1925).
29. "Interfacial Tension and Emulsification" (with H. Zollman), *ibid.*, **48**, 69-80 (1926).
30. "The Structure of Films of Water on Salt Solutions. II. The Surface Tension of Calcium Chloride Solutions at 25°" (with E. C. Gilbert), *ibid.*, **48**, 604-607 (1926).
31. "The Drop-Weight Method for the Determination of Surface Tension and the Weight of the Ideal Drop," *Nature*, **117**, 690-691 (1926).
32. "Application of the Drop Weight Method to the Determination of the Surface Tension of Colloidal Solutions," *Nature*, **118**, 732-733, (1926).
33. "Surface Energy and Surface Tension:" W. D. Harkins, in "Colloid Chemistry," Edited by Jerome Alexander, I: 192-264, Chemical Catalog Company (Reinhold Publishing Corp.) 1926.
34. "The Stability of Emulsions, Monomolecular and Polymolecular Films, Thickness of the Water Film on Salt Solutions, and the Spreading of Liquids." "Colloid Symposium Monograph," V: 19-48, Chemical Catalog Company (Reinhold Publishing Corp.) 1927.
35. "Electrical Relations at Surfaces, the Spreading of Liquids, the Thickness of Surface Films, and the Drop Weight and Ring Methods for the Determination of Surface Tension," *Ibid.*, Vol. 6, 17-40, 1928.
36. "Surface Energy and the Orientation of Molecules in Surfaces as Revealed by Surface Energy Relations," *Z. Physik. Chem.*, **A139**, 647 (1928).
37. "Emulsions: Stability, Area Per Molecule in the Interfacial Film, Distribution of Sizes and the Oriented Wedge Theory" (with Norvil Beeman), *J. Am. Chem. Soc.*, **51**, 1674 (1929).
38. "Surface Structure and Atom Building" (17th Willard Gibbs Medal Address), *Science*, **70**, 433, 463 (1929).
39. "The Surface Tension of Blood Serum, and the Determination of the Surface Tension of Biological Fluids" (with Henry N. Harkins), J. Clinical Investigations, **7**, 263 (1929).
40. "The Shapes of Drops and the Determination of Surface Tension" (with B. B. Freud), *J. Phys. Chem.*, **33**, 1217 (1929).
41. "The Drop Weight Method for the Determination of Surface Tension. The Effect of an Inclination of the Tip Upon the Drop Weight" (with David M. Gans), *J. Am. Chem. Soc.*, **52**, 2287 (1930).
42. "A Method for the Determination of Surface and Interfacial Tension from the Maximum Pull on a Ring" (with Hubert F. Jordan), *J. Am. Chem. Soc.*, **52**, 1751 (1930).
43. "Wetting of Pigments and Other Powders" (with Roy Dahlstrom), *Ind. Eng. Chem.*, **22**, 897 (1930).
44. "The Direct Measurement of the Adsorption of Soluble Substances by the Bubble Method" (with David M. Gans), *J. Phys. Chem.*, **35**, 722 (1931).
45. "The Activity Coefficients and the Adsorption of Organic Solutes. I. Normal Butyl Alcohol in Aqueous Solution by the Freezing Point Method" (with Roy W. Wampler), *J. Am. Chem. Soc.*, **53**, 850 (1931).
46. "An Adsorption Method for the Determination of the Area of a Powder" (with D. M. Gans), *Ibid.*, **53**, 2804 (1931).
47. "Flocculation, Dispersion and Settling of Pigments in Relation to Adsorption" (with L. W. Ryan and D. M. Gans), *Ind. Eng. Chem.*, **24**, 1288 (1932).
48. "Monomolecular Films. The Solid-Liquid Interface and Sedimentation and Flocculation of Powders in Liquids (with David M. Gans), *J. Phys. Chem.*, **36**, 86 (1932).

49. "Monomolecular Films. The Liquid-Liquid Interface and the Stability of Emulsions" (with Earl K. Fischer) *Ibid.*, **36**, 98 (1932).
50. "Contact Potentials and the Effects of Unimolecular Films on Surface Potentials. I. Films of Acids and Alcohols" (with E. K. Fischer), *J. Chem. Physics*, **1**, 852 (1933).
51. "Surface Pressures and Potentials of Films of Long Molecules: Polymers of ω-Hydroxy Decanoic Acid."
52. "Monomolecular Films of Molecules which Lie Flat on the Surface of Water. I. Surface Pressures and Potentials of Films of Long Molecules: Polymers of ω-Hydroxydecanoic Acid" (with H. E. Ries and E. F. Carman), *J. Chem. Phys.*, **3**, 692 (1935).
53. "Surface Potentials and Force-Area Relations of Monomolecular Films. II. *d*-Pimaric and Tetrahydro-*d*-pimaric Acid" (with H. E. Ries and E. F. Carman), *J. Am. Chem. Soc.*, **57**, 2224 (1935).
54. "The Rearrangement of Molecules in Monomolecular Films: Polycyclic Compounds of the Five Ring Series" (with H. E. Ries and E. F. Carman), *J. Chem. Phys.*, **3**, 228 (1935).
55. "The Rearrangement of Molecules in Plastic Monomolecular Films: Pressure-Area and Potential Relations for Polycyclic Compounds of the Five Ring Series" (with E. F. Carman and H. E. Ries), *J. Am. Chem. Soc.*, **58**, 1377 (1936).
56. "Polymolecular Films" (with R. J. Myers), *Ibid.*, **58**, 1817 (1936).
57. "An Electronic Analysis of Some Surfaces by Means of Slow Electrons" (with R. J. Moon), *J. Phys. Chem.*, **8**, 941 (1936).
58. "Hydrogen Ion Concentration and the Behavior and Measurement of Monomolecular Films on Water" (with R. J. Myers), *J. Chem. Phys.*, **4**, 716 (1936).
59. "Polymolecular Films: Mixed Films With Two or More Components. I. Fatty Acids and Nonpolar Substances" (with R. J. Myers), *J. Phys. Chem.*, **40**, 959 (1936).
60. "Effect of Traces of Metallic Ions at Interfaces and on the Surface of Water" (with R. J. Myers), *Nature.*
61. "Contact Potentials of Reversible Films of Lauric Acid" (with R. J. Myers), *Nature*, **138**, 405 (1936).
62. "Effects of Traces of Metallic Ions on Monolayers" (with R. J. Myers), *Nature*, **139**, 367 (1937).
63. "Linear or Edge Energy and Tension as Related to the Energy of Surface Formation and of Vaporization," *J. Chem. Phys.*, **5**, 135 (1937).
64. "Ultramicroscopic Examination of Mixed Films" (with F. M. Fowkes and R. J. Myers), *J. Am. Chem. Soc.*, **59**, 593 (1937).
65. "I. A Simple Accurate Film Balance of the Vertical and Type for Biological and Chemical Work, and a Theoretical and Experimental Comparison with the Horizontal Type. II. Tight Packing of a Monolayer by Ions" (with T. F. Anderson), *Ibid.*, **59**, 2189 (1937).
66. "The Surface Tension of Aqueous Soap Solutions as a Function of Hydrogen Ion and Salt Concentration. I. Sodium Laurate and Sodium Nonylate" (with F. A. Long and G. C. Nutting), *Ibid.*, **59**, 2197 (1937).
67. "The Viscosity of Monolayers: Theory of the Surface Slit Viscosimeter" (with J. G. Kirkwood), *J. Chem. Phys.*, **6**, 53 (1938).
68. "The Contact Potential of Solid Films Formed by Evaporation and by Solidification and of Built-up Multilayers on Metals" (with R. W. Mattoon), *Phys. Rev.*, **53**, 911 (1938).
69. "Surface Viscosity of Long-Chain Alcohol Monolayers" (with L. Fourt), *J. Phys. Chem.*, **42**, 897 (1938).
70. "Protein Monolayers: Films of Oxidized Cytochrome C" (with T. F. Anderson), *J. Biol. Chem.*, **125**, 369 (1938).
71. "Effects of Space Isomerism on the Squeezing Out of an Unsaturated Compound From a Mixed Monolayer on an Aqueous Subsolution" (with R. T. Florence), *Nature*, **142**, 913 (1938).
72. "Molecular Interaction in Mixed Monolayers on Aqueous Subsolutions. I. Mixtures of Alcohols, Acids and Amines" (with R. T. Florence), *J. Chem. Phys.*, **6**, 847 (1938).
73. "Molecular Interaction in Mixed Monolayers. II. Unstable Mixtures with Unsaturated Acids" (with Robert T. Florence), *Ibid.*, **6**, 856 (1938).
74. "Pressure-Area Relations of Monolayers at the Solid-Liquid Interface" (with F. M. Fowkes), *J. Am. Chem. Soc.*, **60**, 1511 (1938).
75. "Film Potentials of Stearate Multilayers and Other Dielectrics on Metal Surfaces" (with R. W. Mattoon), *J. Chem. Phys.*, **7**, 186 (1939).
76. "Viscosity of Two-Dimensional Systems: Effect of Pressure and Temperature, and Detection of Phase Transitions in Monolayers" (with G. E. Boyd), *Ibid.*, **7**, 203 (1939).
77. "A New Method for the Investigation of the Electrical Properties of Multilayers" (with Gerhart Groetzinger), *Ibid.*, **7**, 203 (1939).
78. "Energy Relations in Transformations from Three to Two-Dimensional Systems. I. The Latent Heat and Entropy of Spreading of Myristic and Pentadecylic Acids" (with G. C. Nutting), *J. Am. Chem. Soc.*, **61**, 1702 (1939).
79. "The Pressure-Area and Pressure-Temperature Relations of Expanded Monolayers of Myristic and Pentadecylic Acids" (with G. C. Nutting), *Ibid.*, **61**, 2040 (1939).
80. "Some Aspects of Surface Chemistry Fundamental for Biology," in "Recent Advances in Surface Chemistry and Chemical Physics," Publication No. 7 of the American Association for the Advancement of Science, pp. 19-46.
81. "Molecular Interaction in Monolayers: Viscosity of Two-Dimensional Liquids and Plastic Solids. V. Long-Chain Fatty Acids" (with E. Boyd), *J. Am. Chem. Soc.*, **61**, 1188 (1939).
82. "Pressure-Area Relations of Fatty Acid and Alcohol Monolayers" (with G. C. Nutting), *Ibid.*, **61**, 1180 (1939).
83. "Immunochemistry of Catalase. II. Activity in Multilayers" (with P. C. Fourt and L. Fourt), *J. Biol. Chem.*, **132**, 111 (1940).
84. "The Change with Time of the Surface Tension of Solutions of Sodium Cetyl Sulfate and Sodium Lauryl Sulfate" (with G. C. Nutting, F. A. Long), *J. Am. Chem. Soc.*, **62**, 1496 (1940).
85. "The Viscosity of Monolayers: A Test of the Canal Viscosimeter" (with G. C. Nutting), *Ibid.*, **62**, 3155 (1940).
86. "The State of Monolayers Adsorbed at the Interface Solid-Aqueous Solution" (with F. M. Fowkes), *Ibid.*, **62**, 3377 (1940).
87. "The Thermodynamics of Films: Energy and Entropy of Extension and Spreading of Insoluble Monolayers" (with T. F. Young and E. Boyd), *Ibid.*, **8**, 954 (1940).
88. "The Thermodynamics of Films: Energy and Entropy of Extension and Spreading of Insoluble Monolayers" (with T. F. Young and E. Boyd), *Ibid.*, **8**, 954 (1940).
89. "The States of Monolayers" (with E. Boyd), *J. Phys. Chem.*, **45**, 20 (1941).
90. "A General Thermodynamic Theory of the Spreading of Liquids to Form Duplex Films and of Liquids or Solids to Form Monolayers," *Ibid.*, **9**, 552 (1941).
91. "Surface Films of Fatty Acids, Alcohols, and Esters," *Chem. Rev.*, **29**, 385 (1941).

92. "Surface Chemistry," *Nature*, **148**, 743 (1941).
93. "The Energy of Immersion of Crystalline Powders in Water and Organic Liquids," (with G. E. Boyd), *J. Am. Chem. Soc.*, **64**, 1190 (1942).
94. "The Binding Energy between a Crystalline Solid and a Liquid: the Energy of Adhesion and Emersion. Energy of Emersion of Crystalline Powders" (with G. E. Boyd), *Ibid.*, **64**, 1195 (1942).
95. "A Superliquid in Two Dimensions and a First-Order Change in a Condensed Monolayer. I. Energy, Compressibility, and Order of Phase Transformations," (with L. E. Copeland), *J. Chem. Phys.*, **10**, No. 5, 272-286 (1942).
96. "A Superliquid in Two Dimensions and a First-Order Change in a Condensed Monolayer. II. Abnormal Viscosity Relations of Alcohol Monolayers in Condensed Liquid Phases," (with L. E. Copeland and G. E. Boyd), *J. Chem. Phys.*, **10**, No. 6, 357-365 (1942).
97. "Energy Relations of the Surfaces of Solids. I. Surface Energy of the Diamond," *J. Chem. Phys.*, **10**, No. 5, 268-272 (1942).
98. "Energy Relations of the Surfaces of Solids, II. Spreading Pressure as Related to the Work of Adhesion Between a Solid and a Liquid," (with H. K. Livingston), *Ibid.*, **10**, No. 6, 342-356 (1942).
99. "The Film Balance As an Analytical Tool for Biological and Food Research," (with G. E. Boyd), *Ind. Eng. Chem.*, **14**, 496 (1942).
100. "The Pressure-Area Temperature and Energy Relations of Monolayers of Octadecanenitrile," (with L. E. Copeland), *J. Am. Chem. Soc.*, **64**, 1600 (1942).

Papers Quoted But Not Yet Published

101. "The Effect of Calcium Ions on the Area of Monolayers of the Long-Chain Paraffin Acids" (with L. E. Copeland and B. Adinoff).
102. "The Adsorption of Organic Acids from Non-polar Solvents on the Surfaces of Polar Solids."
103. "An Accurate Method for the Determination of the Adsorption of Vapor on the Surface of a Solid" (with G. Jura).
104. "A Method for the Determination of the Area of a Finely Divided Polar Solid without the Assumption of a Molecular Area: A Combination of Adsorption and Measurement of the Energy of Immersion" (with G. Jura).
105. "Thermodynamic Relations of Solid-Vapor and Solid-Liquid Interfaces" (with G. Jura).
106. "The Lowering of the Free Energy of the Surface of a Crystalline Solid by an Adsorbed Film" (with G. Jura).
107. "The Distance to which a Crystalline Solid Attracts an Adjacent Liquid: Variation of the Energy of Attraction with Distance" (with G. Jura).
108. "A New Point of View of the Process of Monolayer Penetration" (with L. E. Copeland and S. Gordon).
109. "The Abnormal Viscosity Relations of Monolayers of the Long-Chain Alcohols" (with L. E. Copeland).
110. "The Latent Heat and Entropy of Spreading of *n*-Long Chain Paraffin Acids" (with L. E. Copeland).
111. "The Viscosity of Monolayers of Octadecyl Alcohol between 25° C and 50° C," (with L. E. Copeland.
112. "A New Linear Form of the Adsorption Isotherm (log (p/p_0) vs $1/v^2$)" (with G. Jura).

Solutions of Soaps and Detergents as Colloidal Electrolytes

JAMES W. McBAIN

Department of Chemistry, Stanford University

It was found by McBain and his collaborators [1] that, in very dilute solution, a pure salt of a higher fatty acid—a soap—behaves as an ordinary electrolyte, but that as the concentration is increased it exhibits typical colloidal properties. McBain therefore defined colloidal electrolytes as electrolytes in which one ion is partially or wholly replaced by conducting colloidal particles, spontaneously formed in the solution. The soaps were the first and most thoroughly studied of this great class of colloidal electrolytes, but it was recognized from the outset that there must be innumerable other examples, including dyes, detergents, and biocolloids—indeed all ionizable organic compounds of not too low molecular weight. There are also several inorganic representatives, such as sodium silicates. Krafft [2] in 1896 pointed out that cetyl amine hydrochloride resembled soaps in its properties.

Hence we recognize anion-active and cation-active colloidal electrolytes, and a few cases in which, like Reychler's [3] triethyl cetyl ammonium cetyl sulfate, or the more recent commercial lauryl pyridinium laurate or lauryl pyridinium lauryl sulfate, both anion and cation are active. The great majority of the modern synthetic

detergents are anion-active sodium salts, though some are cation-active. There are a few, moreover, which are not electrolytes at all, such as nonyl glucoside, or such commercial non-electrolytic detergents as the Peregals, Emulfors, and Igepals, which are long-chain compounds of high molecular weight, exhibiting close similarity to the surface properties of the colloidal electrolytes.

Historical

Thirty years ago it was assumed that colloids had no very definite properties, and that they were exceedingly subject to change, depending upon their history. It was supposed that a colloid could not conduct; nor could it have a definite quantitative, easily measurable effect on lowering of freezing point or vapor pressure. Such of these properties as were observed were ascribed to impurities or products of hydrolysis or decomposition.

However, the most important result of the study of pure soaps was to establish that their solutions have definite, reproducible properties, easily rendered independent of their history, like those of sugar or potassium chloride. This laid the sound, permanent foundation for the study of this subject.

The main result, then, was to show that certain colloidal particles form spontaneously from ions, molecules or crystals; that they enter into true reversible equilibrium with these; and that therefore colloidal electrolytes are truly stable in the strictest thermodynamic sense.

So novel was this finding that when in 1925 some of the evidence for it was presented to the Colloid Committee for the Advancement of Science in London, it was dismissed by the Chairman, a leading international authority, with the words, "Nonsense, McBain."

These stable colloidal particles have a self-organizing structure based on the principle of like to like. Polar groups are exposed to water, and hydrocarbon groups to hydrocarbon groups, to minimize interfacial energy and leave, instead, the maximum number of water molecules in mutual contact. Thus each molecule has its soluble groups outside, and its insoluble portions exposed as little as possible. Several structures have been suggested in which this purpose may be accomplished.

Soap Solutions

The proof that soaps are colloidal electrolytes and that such a class had to be recognized lies in the quantitative comparison of different kinds of physical chemical data, since any one kind, standing alone, might be explained by some other hypothesis.

First it had to be shown that hydrolysis is almost negligible, except in very dilute solution, so that the properties are those of the soap itself. Hydrolysis alkalinity never exceeds $N/1000$, and where desired it is readily suppressed by adding a slight excess of free hydroxide, which does not combine with soap.[4]

Then it was found that soap solutions conduct excellently; in fact, in concentrated solution their equivalent conductivity is almost as great as that of an organic salt such as sodium acetate. On the other hand, their osmotic effects, such as lowering of freezing point, of vapor pressure, of dew point or of osmotic pressure itself, are only moderate, being less than half what one would expect for an ordinary salt.

This contrast between electrical and osmotic behavior could be explained only by making two assumptions; first, that when these two physical chemical properties departed from the normal behavior of an ordinary salt, and also from each other, a colloid must have formed in the solution; secondly, that the colloid must exhibit quite appreciable conductance. There is universal agreement on these points.

This conclusion follows from a quantitative consideration of two different kinds of physical chemical measurements, which could not have been arrived at with certainty through either alone; and it illustrates the great importance of using all available methods of study. Upwards of two dozen have so far been applied.

Typical osmotic behavior is illustrated by Figure 1 (due to S. A. Johnston and A. P. Brady) for solutions of potassium oleate. The osmotic coefficient g is defined as the ratio of the observed lowering of freezing point to that which would be expected for an ideal fully dissociated electrolyte forming two ions:

$$g = i/2 = 1 - j = \frac{\theta}{2 \times 1.858\, m}$$

where g is the osmotic coefficient, i is van't Hoff's coefficient, j is the function of Lewis and Randall, m is the molality, 1.858 is the molal lowering for any ideal solute, and θ is the observed lowering of freezing point in degrees.

For comparison, the behavior of a real electrolyte, potassium chloride, is also shown in Figure 1.

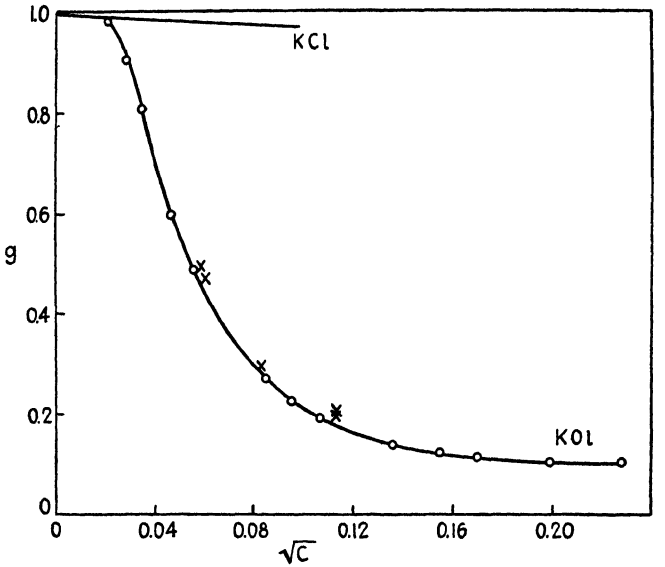


FIGURE 1. Comparison of the Osmotic Behavior of Potassium Oleate as Colloidal Electrolyte with that of Potassium chloride as Typical Strong Electrolyte. The osmotic coefficient, g, is the lowering of freezing point observed divided by 3.716 times the molality, m.

It is seen that the behavior of potassium oleate and potassium chloride is similar in extreme dilution; but that over the range $0.0004m$ to $0.06m$, a 150-fold change in concentration, the value of g falls away to one-tenth of that for potassium chloride of the same concentration. The activity coefficient, of course, falls to a much lower value and continues to fall, whereas the g value somewhat abruptly begins to rise again in higher concentration.

Figure 2 (due to O. E. Bolduan) shows the osmotic behavior of the potassium salts of the saturated fatty acids of twelve, ten and four carbon atoms. The distinctive behavior of the soaps is entirely lost in potassium butyrate, where hydration effects obscure all others in high concentration.

Figure 3 (taken from McBain, Laing and Titley[5]) reproduces a series of conductivity curves of soaps at 90° and at room temperature, showing the family relations between the fall in conductivity in rather dilute solution, followed by a

minimum between 0.05 and 0.1*N* solutions, followed by a large increase in conductivity in higher concentration.

It is generally agreed that the only possible explanation of the great decrease in osmotic behavior and conductivity is that ions are replaced by colloidal particles, which, although conducting, have a lower equivalent conductivity than the ions from which they are formed. There is disagreement only as to the kinds and shapes

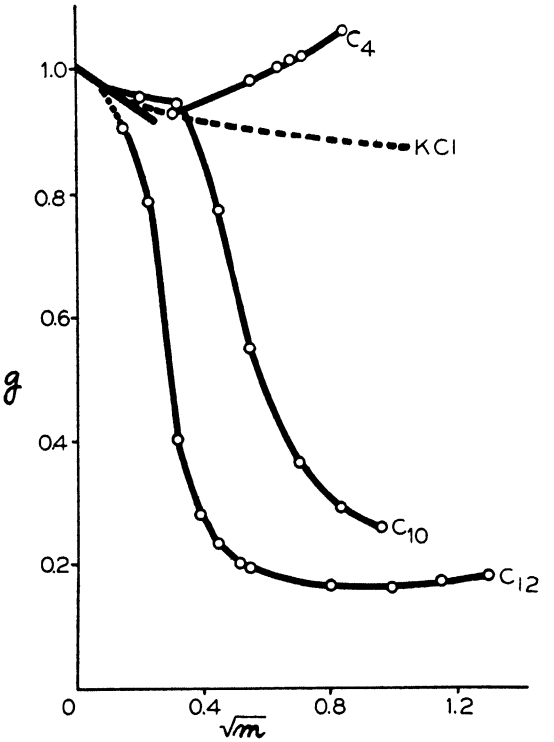


FIGURE 2. Comparison of the Osmotic Behavior of Potassium Salts of the Saturated Fatty Acids, with twelve, ten and four carbon atoms.

of colloidal particles involved. The experimental data have been substantiated by ultrafiltration [6] and by measurements of electromotive force using hydrogen [4] and sodium and potassium amalgam [7] electrodes.

The main conclusions as to the constitution of colloidal electrolytes as outlined above remain unaffected by the fact that, at the time they were deduced, strong electrolytes were considered to be incompletely ionized until after the adoption of the Debye-Hückel theory (1923).

Solutions of Dodecyl Sulfonic Acid

Lottermoser and Püschel [8] were the first to publish data on nonhydrolyzable colloidal electrolytes. Their extensive conductivity measurements include the alkyl sulfates of sodium, potassium, hydrogen, silver, magnesium, calcium, barium, zinc, copper and iron, including the straight-chain alkyl groups containing between twelve and eighteen carbon atoms, supplemented by measurements of electromotive force as well as of surface tension. They found that in very dilute solution the behavior

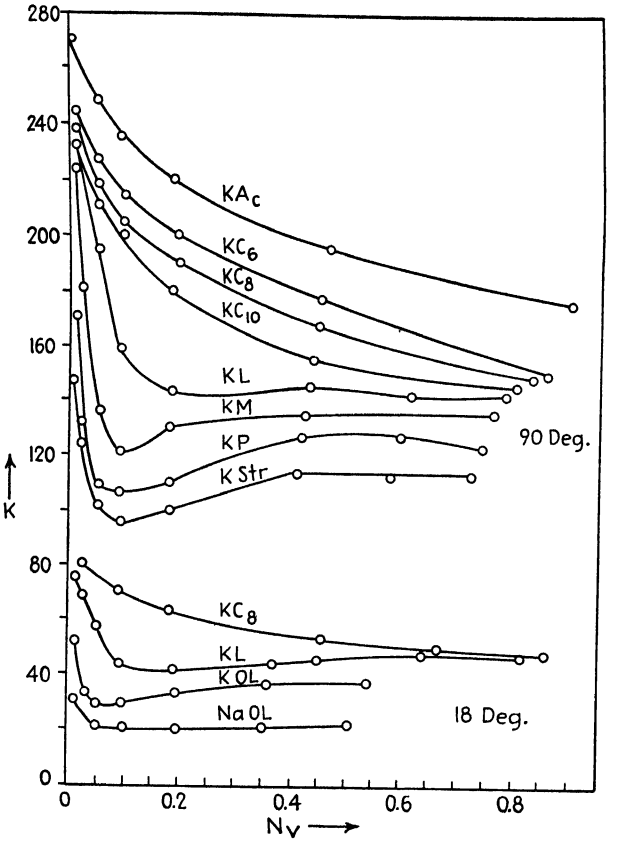


FIGURE 3. Equivalent Conductivity of Soap Solutions at 18° and at 90°.

is like that of a strong electrolyte; then at the following critical concentrations for univalent salts the conductivity and activity suddenly fall to low values by formation of colloid:

C_{12} 0.003 –0.006*m*
C_{14} 0.0016–0.0025*m*
C_{16} 0.0004–0.0006*m*
C_{18} 0.0003*m*

Thereupon there is a minimum in the curve of equivalent conductivity at 60° observed for:

C_{12} 0.1 –0.2*m*
C_{14} 0.1*m*
C_{16} 0.05*m*
C_{18} 0.025–0.05*m*

Finally, there is a maximum in the conductivity at 60° for:

C_{12} 0.75*m*
C_{14} 0.5*m*
C_{16} 0.25*m*
C_{18} 0.25*m*

Similar conductivity measurements were later supplied by Hartley,[9] Howell and Robinson,[10] Ward,[11] Tartar and collaborators,[12] and several others. McBain and Betz [13] were the first to study the straight-chain sulfonic acids, measuring conductivity, lowering of freezing point, and electromotive force.

By far the most carefully studied colloidal electrolyte is lauryl sulfonic acid, a highly soluble, non-hydrolyzable, free acid, based upon a paraffin chain of twelve carbon atoms.

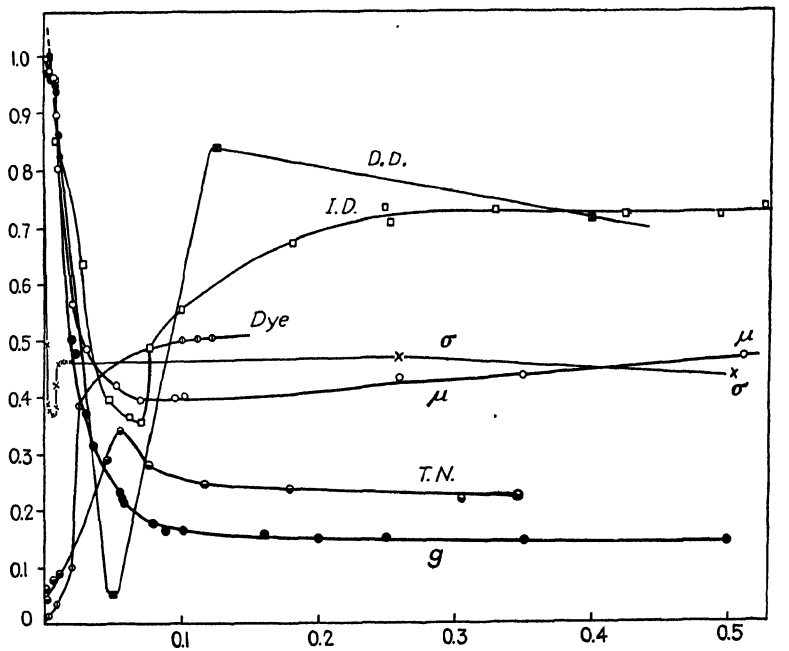


FIGURE 4. Quantitative Data for Solutions of the Typical Colloidal Electrolyte; Lauryl Sulfonic Acid; including Conductivity, μ, Transport Number, T.N., Osmotic Coefficient, g, Integral Diffusion Coefficient, 1D, Differential Diffusion Coefficient DD, Dye Solubilization, Dye, Surface Tension, σ, to show the shape of each curve and the actual concentrations at which maxima and minima occur for different properties.

Figure 4 shows data for lauryl sulfonic acid, all from the Stanford University laboratory, comprising conductivity,[13, 14] transport number,[15] lowering of freezing point,[13, 14] diffusion,[16] dye solubility,[17] and surface tension.[18] (The other surface properties are presented in Figure 12, since surface anomalies occur only in the lowest range of concentration, before the fall in conductivity and freezing point begins).

The equivalent conductivity of lauryl sulfonic acid at 25° falls rapidly at between 0.008 and 0.009m to a minimum, where it is 39 per cent of its value for infinite dilution; then it rises again steadily.

The osmotic coefficient falls off at 0.00105m at 0° and becomes nearly constant at 0.085m, at a value between 14.5 and 16 per cent of that for infinite dilution, as compared with the 39 per cent for conductivity.

The transport number of the alkyl group or sulfonic radical is but little affected from its value for complete dissociation at infinite dilution, 0.059 equivalent per Faraday of current, until 0.016m (where g and conductivity have both fallen to

65 per cent). Then the transport number rises to a sharp maximum at 0.055*m* with a value of 0.35 equivalent per Faraday of current. Thus its rise lags far behind the formation of colloid as shown by conductivity and freezing point, but its maximum is reached well before the minimum in conductivity is reached. Thereafter it falls again, rapidly at first, to a value of 0.22.

The diffusion coefficient likewise passes through a sharp minimum. For the integral diffusion coefficient this is at 0.07*m*, but the true or differential diffusion coefficient shows the true minimum to be at 0.05*m*, the same as the sharp maximum

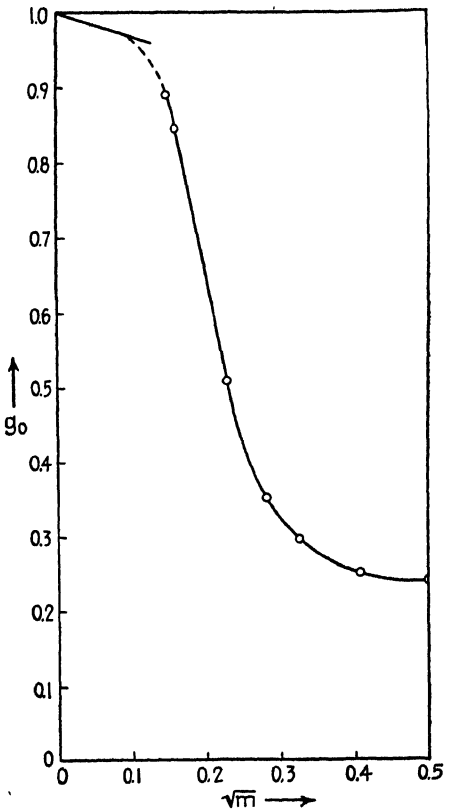


FIGURE 5. The osmotic coefficient of the Wetting Agent and Detergent "Tergitol 4," sodium tetradecylsulfate.

for transport number. Both diffusion coefficients then rise greatly to a value about 73 per cent that for infinite dilution, whereas the minimum for integral diffusion was 36 per cent and for differential diffusion coefficient 5 per cent of that for infinite dilution.

The dye solubility, to be discussed later, anticipates the fall in conductivity and osmotic coefficient, but otherwise is almost a mirror image of the *g* value, not changing after 0.1*m*.

The surface tension has a sharp minimum at 0.005*m*, 36 per cent of the value of water, then remaining constant above 0.01*m* at 47 per cent.

Attention should be drawn to the rather abrupt change in osmotic coefficient from a rapidly falling value to constancy, to the minimum in conductivity rising markedly with further increase in concentration, to the very deep and sharp mini-

mum in diffusion coefficient followed by a high value for all concentrated solution, and finally to the definite discrepancies between the concentrations at which these various effects are observed.

The writer finds it impossible to explain all these quantitative data without as-

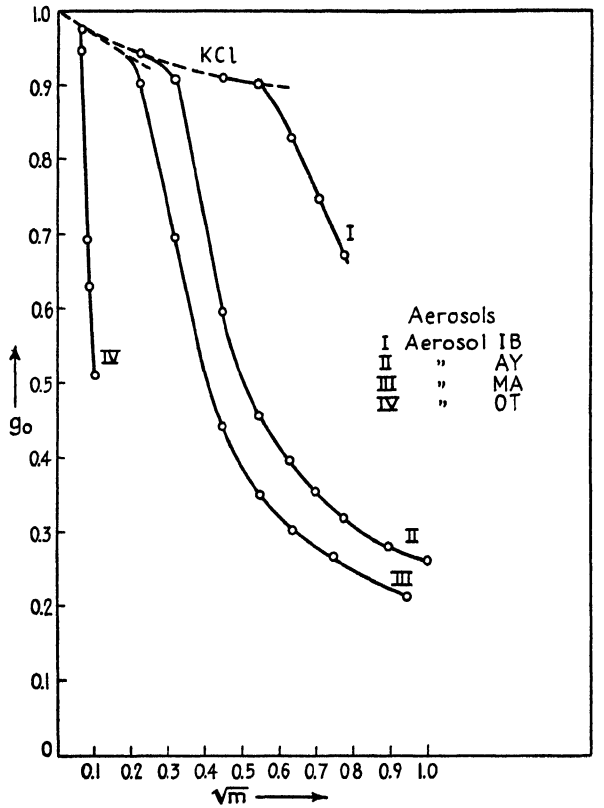


FIGURE 6. The osmotic Coefficients of a series of the Wetting Agents and Detergents, Aerosol OT (dioctyl, 2 ethylhexanol), Aerosol MA (dihexylmethylamyl) Detergents, Aerosol 1B (diisobutyl), and Aerosol AY (diamyl), all esters of sodium sulfosuccinate.

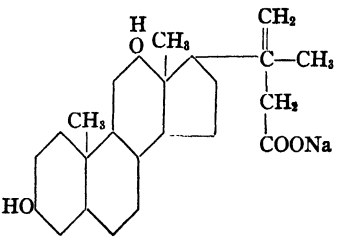
suming the presence of many sizes and kinds of colloidal particles, shifting according to the concentration.

Osmotic Coefficients of Colloidal Electrolytes

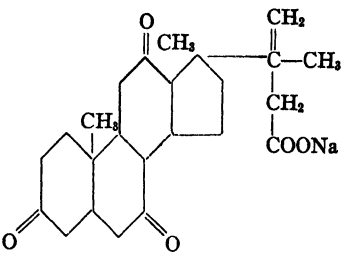
Tergitol 4, the sodium salt of the highly branched secondary tetradecyl sulfate, 7-ethyl, 2-methyl, 4-decanol sodium sulfate, is a typical colloidal electrolyte as shown by Bolduan's freezing point data in Figure 5.

Likewise the Aerosols, esters of mono-sodium sulfonate of succinic acid, are all colloidal electrolytes, the colloid forming at higher and higher concentrations as the alkyl groups of the ester decrease in size. It has been computed that this trade name could include over a million such compounds of which only a few dozen have been prepared. Bolduan's data for the Aerosols are given in Figure 6.

Two bile salts have also been measured by Johnston and Bolduan and are shown in Figure 7. Their formulas are



Sodium Desoxycholate



Sodium Dehydrocholate

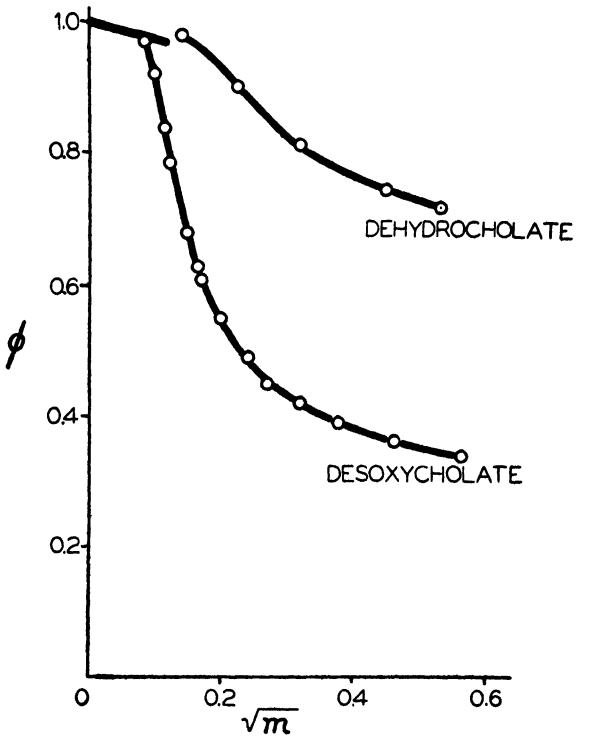


FIGURE 7. The Osmotic Coefficient of the Bile Salts; the detergent sodium desoxycholate, and the non-detergent sodium dehydrocholate.

They are remarkable in that this minor difference in formula has such a profound effect on their properties. Sodium desoxycholate, like the other colloidal electrolytes, solubilizes otherwise insoluble dyes and hydrocarbons. Practically speaking, sodium dehydrocholate is not a colloidal electrolyte; it dissolves no dye, and it salts out hydrocarbons from water.

A. P. Brady at Stanford University has discovered that all existing freezing-point data for colloidal electrolytes may be placed upon one of three curves on the

same diagram, provided that the concentrations are brought to comparable values by multiplying with a constant for each substance. There is one normal curve on Figure 8 for all the straight-chain compounds, such as the saturated soaps, potas-

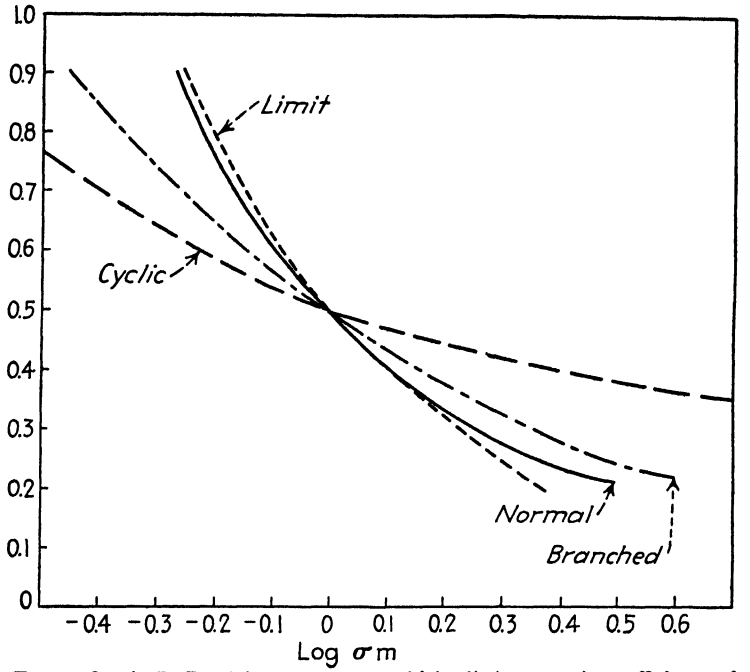


FIGURE 8. A. P. Brady's curves upon which all the osmotic coefficients of colloidal electrolytes may be placed by merely multiplying the concentration by an arbitary constant in each case. The three curves are for straight-chain, branched and cyclic compounds, respectively; the dotted curve is the physical limit for any solution remaining homogeneous instead of separating into two layers; the lowering of freezing point and activity would then remain constant over this range.

sium octoate, decylate, and laurate, lauryl sulfonic acid, sodium decyl sulfonate, sodium dodecyl sulfonate and the cation-active dodecyl amine acetate.[19] There is another curve for the branched-chain compounds: Aerosols, Tergitol, potassium oleate, and $C_{11}H_{23}COOC_3H_4NHCOCH_2N(C_2H_5)_2Cl$. The third curve is for the polycyclic bile salt. All the existing data fall on the curves within the experimental error. A fourth dotted line has been drawn in to show how steeply the osmotic coefficient would fall in the extreme limit, where the activity and lowering of freezing point remained actually constant, and the solution would fall apart into two layers of constant composition. It is seen that all the curves have less slope than this, and that the straight-chain compounds come nearest to it.

Other Examples of Colloidal Electrolytes

The danger of using only one physical chemical method is illustrated[20] by Figure 9, for the conductivities of methylene blue, meta benzopurpurin, and cetyl pyridinium chloride. The curve for cetyl pyridinium chloride resembles that of the soaps. That for methylene blue rises strikingly in very dilute solution, whereas that for meta benzopurpurin resembles an ordinary salt.

Hartley, Collie and Samis [21] comment: "This is well illustrated by the equivalent conductivity-concentration curves for the two dyes, methylene blue and 'meta' benzopurpurin, and for the paraffin-chain salt, cetyl pyridinium chloride. The first has an equivalent conductivity rising *above its infinite dilution value* in very dilute solution, a phenomenon which seems to admit no other interpretation than the formation of (probably small) micelles. The second, which transport number measurements have conclusively shown to contain micelles, has an *apparently* normal conductivity.

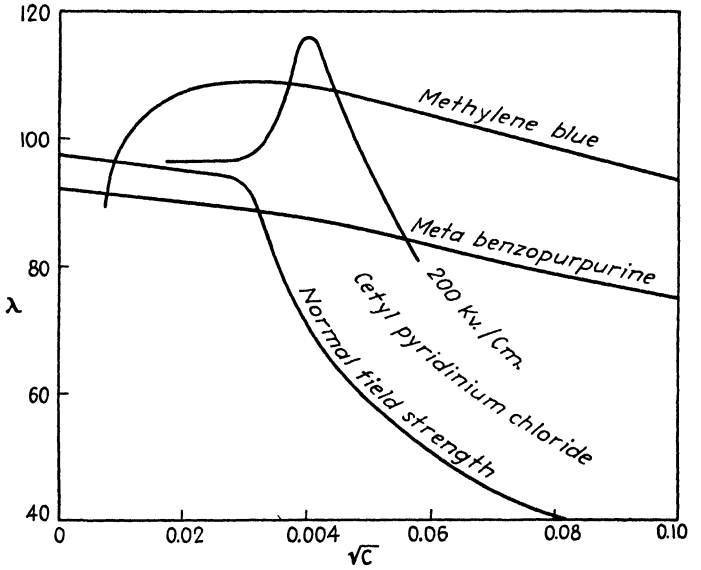


FIGURE 9. The Conductivity of 3 Colloidal Electrolytes, Methylene Blue, Benzopurpurin and Cetyl Pyridium Chloride, showing Three Types of Behavior all caused by Colloidal Particles.

The third shows, at a concentration of about $N/1000$, a very sudden *fall* of equivalent conductivity which, there is nevertheless very strong evidence to show, is accompanied by the formation of large micelles."

The conductivity of the bile salt sodium desoxycholate [22] falls to the extent of only two-thirds, which is not a very pronounced departure from that of a salt, and the osmotic coefficient is lowered to one-third. This points to smaller, better-conducting micelles in the bile salt solution, presumably on account of the greater difficulty of close packing them in the colloidal particle.

The equivalent conductivities of some cation-active detergents [23] are shown in Figure 10. They resemble a homologous series of anion-active colloidal electrolytes.

Figure 11 includes data [21] for the conductivity and the transport number (of the cation) of two cation-active detergents, cetyl pyridinium bromide and cetyl trimethylammonium bromide. It is noteworthy that, in these instances, the fall in conductivity is accompanied by a large increase in transport number, rising to a maximum of over 1.6 equivalents of cetyl radical per Faraday of current in 0.014*m* solution. Thereafter, the transport number decreases considerably while the conductivity continues to fall gently. The significance [15] of such high transport numbers in excess of unity, like those obtained for sodium oleate,[24] is not that such an ion has a high conductance, but rather the opposite. This general misconception was contributed to by a well known diagram [21] in which the conductivities shown here in Figure 11 were multiplied by the transport numbers and the results entitled the

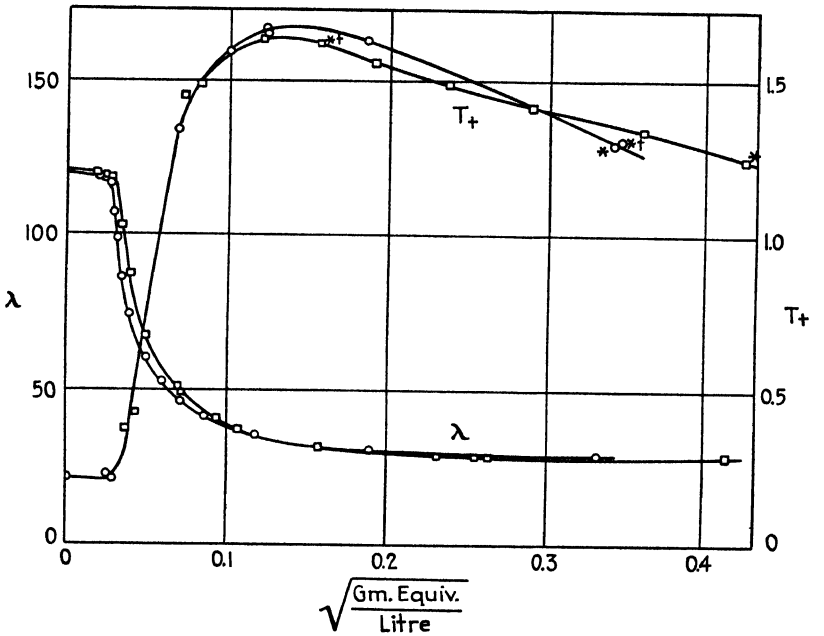


FIGURE 10. Equivalent Conductivity, λ, and Transport Number, T_+ of Cetyl Pyridinium Bromide and of Cetyl Trimethylammonium Bromide at 35°.

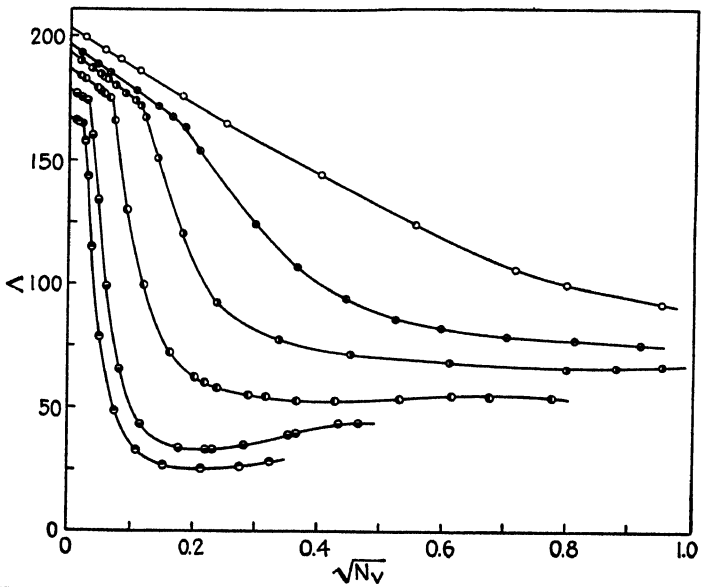


FIGURE 11. Conductivities of Cation-active Detergents, the hydrochlorides of octyl-, decyl-, tetradecyl- and hexadecylamines.

"equivalent conductivities of the component radicals." This needs a note of explanation.

Always, in any electrolytically conducting solution, the total current is distributed among the carriers present, that is, all ions and colloidal particles. Hence each kind carries a fraction of the total current, and that fraction is the true or electrical transport number of that particular species of ion or particle. Hence true or electrical transport numbers are always less than unity. The conductivity of the part is always less than that of the whole. The transport of material, on the other hand, is equal to the true transport number multiplied by the number of equivalents of material associated with each electrical charge. If the same kind of material, such as bromine, is contained in more than one carrier, then the observed transport of that substance is the algebraic sum of the material transport numbers of all the kinds of carriers containing it.

Here, some bromine must be carried toward the cathode in the positively charged colloidal micelles or particles, out-weighing the bromine being carried toward the anode by the faster simple bromide ion. Hence the true significance of the data is not that the conductivity of the bromine is negative, as alleged, but that bromine constitutes a part of the positively charged colloid; and furthermore, to produce such a large gross transport number as 0.6 equivalent of total bromine toward the cathode per Faraday of current, there must be several atoms of bromine in the positive colloid for each free positive charge thereon. M. E. Laing [24] has shown how to combine conductivity and transport data so as to derive the actual composition of the average colloidal particle.[25]

The conductivity and the osmotic pressure fall on account of the withdrawal of ions into the colloidal particle. The conductivity is most affected by the withdrawal of ions of opposite sign into the same colloidal particle.

Kinds of Colloidal Particles in the Solution

McBain originally tried to predicate only one kind of colloidal particle, but was forced to abandon this position entirely when it was necessary to take into account such independent quantitative physicochemical data as have been referred to above. He then adopted a spherical ionic micelle together with the much more poorly conducting lamellar micelle. There is direct proof of the existence of lamellar micelles by x-ray diffraction,[26] although this is not exhibited over the whole range of concentration in which colloid is present. The molecules are arranged parallel, side by side, in two layers showing side spacing, and the long spacing equal to double the length of the molecule. This evidence is ignored without reference by the English writers.

McBain's spherical ionic micelle was postulated as being comprised of not more than ten like ions retaining their charges. Such a particle would exhibit a corresponding lowering of osmotic coefficient, but must have an enhanced equivalent conductivity to the extent of $n^{2/3}$, where n is the number of like ions merged into one spherical particle. The principle is that of Stokes' Law, or like the rapid fall of a large raindrop compared to a similar weight of small drops of mist. The total force in this case, charge multiplied by applied electrical field, was the same before and after aggregation, but the resistance to movement is $n^{2/3}$ smaller for the one large sphere than for n small spheres. McBain [27] points out that the concentrations of ions must increase, however slightly, over the whole range of concentration, and that therefore the formation of ionic micelles must gradually increase from great dilution to high concentration. Small micelles, as in Figure 9 for methylene blue, would serve the same purpose, if the like ions were not over-burdened with too many opposite ions.

A complex of three ions, including two of like charge, will lower their contribution to the osmotic coefficient, g, to a third of the value for three free ions; but it

will also reduce the conductivity, since the one charge is now loaded. A complex of five ions, including two of like charge, will contribute greatly to lowering of g, but may enhance the conductivity. Thus a small micelle of whatsoever shape, if it has several charges, may resemble McBain's ionic micelle in its effect upon osmotic activity and conductivity and even upon transport number. Likewise, Hartley's suggested micelle of 50 ions of one sign and 30 of the other enhances conductivity instead of diminishing it, while lowering the contribution to g about five fold. Van Rysselberghe [28] calculates the composition of the average micelle which is required to agree with both conductivity and freezing point of lauryl sulfonic acid solutions. Taking both kinds of data together necessitates that the average composition of the micelle vary over the whole range of concentration, and that the first average product of aggregation be an ion pair.

The lamellar micelles, with their compensating ions, were shown by McBain [24, 29] to conduct, in the case of sodium soaps, about one-fourteenth as well as an equal concentration of sodium chloride.

The majority of investigators, such as Adam, Stauff,[30] Hess,[26] Bury,[31] Lawrence,[32] and Meyer and van der Wyk [33] are of the opinion that several kinds of colloidal particles are to be found in the solution of a colloidal electrolyte, as was seen by Hartley to be necessary for methylene blue, Figure 9, p. 112.

In an important contribution, Meyer and van der Wyk [33] discuss the formation of large micelles by the successive addition of ions and ion pairs, pointing out that after the first few these involve approximately equal energy. Hence there would be a range of statistical distribution of sizes in each concentration. McBain [34] pointed out that, for the lower soaps, the first and simplest aggregate is just an ion pair and is there the first to occur. Bury *et al.*[35] showed that, for a large micelle, the application of the Mass Law would predict a fairly abrupt, although never a perfectly sharp, initiation of the formation of an appreciable quantity of micelles. This they call the critical concentration for micelles. We have seen that the micelles continue to be formed over a very wide range, at least until the minimum conductivity is reached.

Tartar [12] and Ward [11] adhere to an earlier suggestion of Hartley's that there is only one kind of micelle—spherical and of constant size—whose radius is determined by the length of the hydrocarbon group.

Aggregation of pre-existing micelles has likewise been proved to occur.[36] This may even lead to gelatinization of the type of a brush heap structure. Association is the key principle of colloids.

Most colloidal electrolytes whose conductivity has been carefully examined exhibit less falling off of conductivity in extreme dilution than is demanded by the Debye-Hückel theory of ions. (Compare the striking departure from methylene blue in Hartley's Figure 9.) For myristal sulfonic acid the conductivity in the initial range of concentration remains constant, as has been shown by unpublished work by M. E. L. McBain. McBain believes that, although in extreme dilution the colloidal electrolyte is practically all dissociated into free ions, nevertheless some ionic micelle must begin to form. Small micelles would behave similarly.

There remains for discussion only the rise in conductivity in concentrations above the minimum. Here, as was earlier done, instead of abandoning the Mass Law, which was relied upon for the critical concentration, Hartley [37] and Ward [11] attempt to explain it in terms of overlapping ionic atmospheres and changes in dielectric constant of water to comparatively great distances from the micelles. They consider that all the charges upon the surface of a particle are equivalent to the same charges concentrated in a point in the center of the particle or mass. However, the free electric charges on all micelles hitherto proposed are so far apart that McBain regards them as being practically independent, each having its own ionic atmosphere. It has been many times observed that the ionic strength is that of a uni-univalent electrolyte, since additions of indifferent electrolyte affect conductivity, freezing

point, etc., to about the same extent as if they were measured in the presence of uni-univalent electrolytes.[38]

Surface Properties of Colloidal Electrolytes

Solutions of colloidal electrolytes exhibit remarkable surface properties. They greatly lower the surface tension of water and even lower the surface tension of hydrocarbons [39, 40] in which they are soluble. A very dilute soap solution, for example, may have a surface tension as low as 25 dynes as compared with 72 dynes for water. This is far lower than 40 dynes for a close-packed film of oleic acid on water or 32.5 dynes for pure oleic acid itself. This fact constitutes one of the many proofs that surface tension is not completely determined by the exposed atoms or molecules, but that the surface is hundreds or even thousands of Ångstroms deep. In many cases the surface tension in very dilute solution passes through a minimum, as was seen in Figure 4, followed almost immediately by a shallow maximum. Non-electrolytic detergents exhibit no minimum, but instead, after the first great fall in surface tension, remain at a constant value throughout all subsequent concentrations. Curves such as these for detergents have been classified as Type III curves, and well over 100 examples have been cited.[41, 42] Similar curves are found for interfacial tension between oil and water.[43]

All the surface anomalies occur in the neighborhood of this minimum, even although it lies below the so-called critical concentration for the formation of micelles.

For all concentrations after the surface tension has recovered from its minimum value, the surface is almost completely reversible and the surface tension independent of time. For all lower concentrations, the surface tension requires extremely long periods of time to come to a constant or equilibrium value.[18, 44] In contrast, the actual adsorption in the surface is almost immediate, and independent of time.

The greatest anomaly, to which many authors have given anxious thought without finding a valid explanation, is the glaring contradiction between the adsorption and the surface tension data. The Gibbs theorem, according to which the adsorption, Γ, is proportional to the slope of surface tension, σ

$$\Gamma = -\frac{d\sigma}{d\mu} = -\frac{d\sigma}{dF} = -\frac{a}{RT}\cdot\frac{d\sigma}{da} = \text{approx} -\frac{c}{RT}\cdot\frac{d\sigma}{dc}$$

predicts positive adsorption for the first downward slope of surface tension; no adsorption whatsoever for the minimum where the surface tension of water is most lowered and both surface and solution are supposed to be 99.9 per cent water; negative adsorption just after the minimum, followed again by no adsorption at all! This is compared with the actual measurements by microtome and interferometer given below.[18]

m	Interferometer	Microtome	Gibbs
0.0020	+3.0	+2.7	+2.4
0.0060 (σ min.)	....	+5.4	0
0.0062	+3.1	....	0
0.0084	+5.3	+3.7	−8.7
0.0087	....	+3.9	−11.2 *
0.0113	....	+5.2	−0.3
0.0117	+1.8	+5.2	−0.2
0.0122	....	+5.7	−0.2

* Here the Gibbs theorem predicts that there can be no lauryl sulfonic acid at all in a surface layer 12700 Å. deep, or correspondingly deeper if this is not pure water.

To save the situation, which is unparalleled elsewhere in science, many workers have made suggestions to save the Gibbs theorem, but unfortunately at the expense of the main principles of thermodynamics. The derivation of the theorem has been repeated by many writers and found to be rigorous back to the point where Gibbs

made his choice of terms or energies to be included in the first law of thermodynamics as applied to the case he had in mind. The latest suggestion, again thermodynamically unsound, is to use, instead of the activity of the colloidal electrolyte, the activity only of its simple ions.[45] However, it was overlooked, as is clear from Figure 4, that for laurylsulfonic acid, the phenomena occur below the critical concentration for micelles, and the calculated result would remain the same.

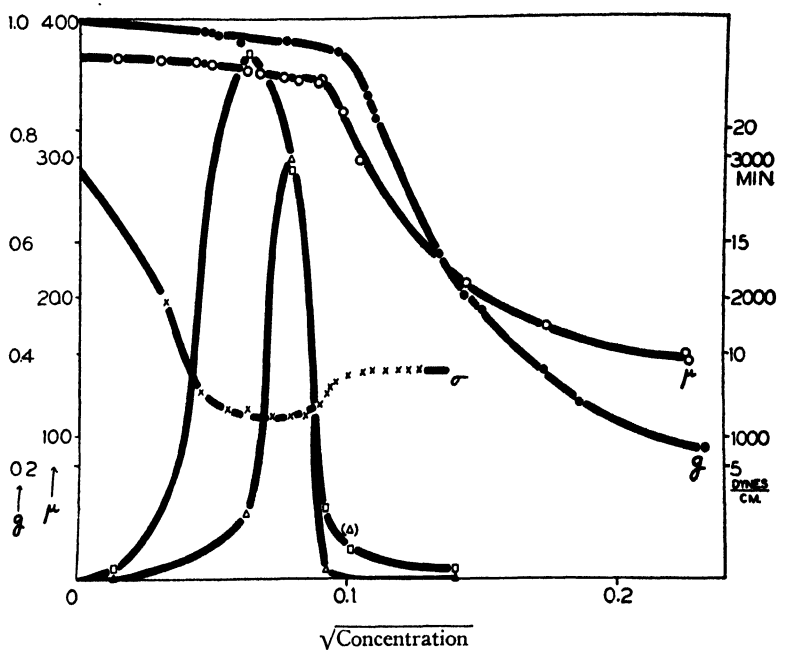


FIGURE 12. Surface properties of very dilute solutions of laurylsulfonic acid, such as surface tension, σ, showing that the marked anomalies occur below the "critical concentration for micelles" as deduced from conductivity μ and coefficient g. The curve with broad maximum is the maximum effect upon surface tension produced by suddenly enlarging the surface by sweeping. The sharp maximum represents the time required to recover from such sweeping, amounting to 3000 minutes.

Solutions of potassium chloride, bromide and iodide resemble each other in surface tension. Nevertheless,[46] in 0.0002*m* solutions the chlorides, bromides and iodides of cetyl pyridinium and of cetyl tri-methyl ammonium are 62, 52 and 37; 63, 52 and 44 dynes respectively.

The other anomaly[47, 48] is that in these very dilute solutions which contain largely or only simple ions, not only is the time required to establish surface tension many orders of magnitude greater than that calculated from diffusion, but if the surface is compressed, a pellicle is formed upon it. The pellicle is a semi-denatured film which requires a definite pressure to be exceeded before the film will re-dissolve, whereas for small compression there is no pellicle and no anomaly in more concentrated solution where colloid predominates.

Figure 12 reproduces on a larger scale than Figure 4 the more dilute region of solutions of laurylsulfonic acid. The curves for conductivity μ and osmotic coefficient g, are given for comparison with the curve for surface tension, σ, to show that the minimum in surface tension at 1.55 grams per liter is well below the "critical concentration for micelles."

Only two other curves are drawn in Figure 12. Both exhibit a high maximum at or near the minimum for surface tension. Both refer to the effect on the float of a Pockels-Langmuir-Adam film balance filled with clean solution of laurylsulfonic acid. When the surface is enlarged by sweeping a barrier *away* from the float, a pull on the float is observed, a temporary increase in surface tension of the swept solution. The maximum effect, 23.5 dynes per cm, is produced at 1 g per liter, but the time required for this pull to disappear again is sharply maximum at 1.58 g per liter, being several thousand minutes; it has fallen to less than thirty minutes at the "critical concentration for micelles."

To avoid complicating Figure 12, some other results may be briefly summarized in words. In all solutions more dilute than the critical concentration, even a slight compression of the surface of laurylsulfonic acid solution, by moving a sweep a short distance toward the float, causes an initial pressure or push upon the float (a temporary reduction in surface tension) by crowding the molecules or ions that were in the surface. The effect steadily increases with dilution, and far greater initial pressures are produced below the critical concentration than above it. A permanent pressure is produced near and above the "critical concentration for micelles" which has a characteristic final value of about 0.5 dyne when nearly all the surface is compressed together.

Some Factors of Detergency

Since the discussions of this subject have recently been summarized,[48] it is sufficient to recall that the equation for detergent action is:

$$\text{fabric} \times \text{dirt} + \text{soap} = \text{fabric} \times \text{"soap"} + \text{dirt} \times \text{"soap."}$$

The action is a replacement of the unknown or undesired dirt by soap or a detergent, leaving the dirt suspended and protected by soap. The quotation marks indicate that very frequently acid or basic constituents of the soap predominate in the sorption compound.

Ordinary soap is the most universal detergent, whereas the soapless detergents are advantageous for specific purposes; such as washing wool and not cotton, etc.

The prerequisite for washing or detergency is wetting, as is easily seen when trying to wash a greasy dish. Hence we may select four factors;

1. Wetting: Due to surface tension and interfacial tension relations and to orientation of the ions or molecules in the surface.

2. Protective Action: Due to adsorbed film of detergent both upon the skin or fabric and upon the dirt removed.

3. Suspending Action: Due to the n free compensating ions of opposite charge which surround any charged particle, such as the remote dirt, and so in effect reduce its particle weight in the proportion $1/(1+n)$. This is only true for ionizing media and a further principle, such as orientation of solvent molecules, or cybotaxis, has to be sought in non-ionizing media.

4. Solubilization: Due to colloidal particles sorbing or including otherwise insoluble matter.

The solubilization of otherwise insoluble dye is an excellent test for colloidal particles. In critical experiments care has to be taken to avoid the mere suspending and protecting of pre-existing colloidal particles of dye by utilizing only large clean crystals of dye. The dye number for laurylsulfonic acid was given in Figure 4, and it was pointed out that it showed some colloid below the "critical" concentration for colloid. Addition of electrolytes favors the formation of colloid, and hence increases the dye number unless the colloid is completely formed already, after which further electrolyte depresses the dye number. Thus by suitable small addition of electrolyte, colloid can be produced in the most dilute solutions.

Numerical data for dye numbers have recently been summarized.[49]

The dye number may be used to settle the question as to whether the dye dissolves

in the hydrophobic part of the micelle, as was suggested by Hartley for his spherical micelle; or alternatively that the dye is sorbed in and upon the micelle or included in layers in the lamellar micelle, as indicated by the x-ray studies of Hess and his collaborators.[27] If the dye is merely dissolved in the hydrocarbon part of the micelle, the amount so dissolved will be proportional to the amount of that hydrocarbon. In a homologous series, then, the dye solubility, in concentrations sufficiently high for the complete formation of colloid, should increase in proportion to the number of CH_2 groups in the paraffin chains.

This has been tested at Stanford University by Johnson. He found, for example, that the number of milligrams of dye solubilized per mole of the potassium salts of the saturated fatty acids were for potassium octoate 80, for potassium decylate 240, for potassium dodecylate 400, and for potassium tetradecylate 1400. This apparently disproves the suggestion of solution in the hydrocarbon portion of the micelle and lends weight to the x-ray evidence for the layers in the lamellar micelles.

Applications

Soap is used for cleansing, for making and breaking emulsions and foams, and for a number of specific purposes. Dyes need only mention.

Soapless detergents are used in a multitude of ways in aqueous and non-aqueous systems, apart from the detergent principles summarized above. For example, cation-active detergents are among the best germicides and fungicides, the effectiveness of all detergents depending upon pH. They are used in insecticides, and for drowning mosquito larvae. Detergents are used to make paints and other materials adhere to, or be indifferent toward, various solids, such as metals, quartz, earth, etc. The applications in the petroleum, food, dye and textile, and paint and lacquer industries—in fact in practically every industry—are so varied and new, and so specific for different small classes of these new materials, that few have yet found their way into published literature.

Colloidal electrolytes include a vast number of substances in which extreme surface activity in very dilute solution and ability to conduct colloidal particles in higher concentrations play a dominant role, which not only makes them of great interest in themselves, but gives them a most varied usefulness in all branches of applied science.

References

1. McBain, J. W., *et al.*, *J. Chem. Soc.*, **101**, 2042 (1912); **105**, 417, 957 (1914); **113**, 825 (1918).
2. Krafft, F., *Ber.*, **29**, 1330 (1896).
3. Reychler, A., *Kolloid-Z.*, **13**, 252 (1913); *Bull. Soc. Chim. Belg.*, **27**, 217 (1913).
4. McBain, J. W., and Martin, H. E., *J. Chem. Soc*, **105**, 957 (1914).
5. McBain, J. W., Laing, M. E., and Titley, A. F., *Trans. Chem. Soc.*, **115**, 1279 (1919).
6. McBain, J. W., and Jenkins, W. J., *Trans. Chem. Soc.*, **121**, 2325 (1922).
7. Salmon, C. S., *Trans. Chem. Soc.*, **117**, 530 (1920).
8. Lottermoser, A., and Püschel, F., *Kolloid-Z.*, **63**, 175 (1933).
9. Hartley, G. S., *Trans. Faraday Soc.*, **31**, 31 (1935); Hartley, G. S., and Murray, R. C., *ibid.*, 183; Hartley, G. S., "Aqueous Solutions of Paraffin-Chain Salts," Actualités Scientifiques et Industricles, No. 387, Hermann et Cie, Paris, 1936.
10. Howell, O. R., and Robinson, H. G. B., *Proc. Roy. Soc. London*, **155A**, 386 (1936).
11. Ward, A. F. H., *Proc. Roy. Soc. London*, **176A**, 412 (1940); *J. Chem. Soc.*, **1939**, 522.
12. Wright, K. A., Abbott, A. D., Sivertz, V., and Tartar, H. V., *J. Am. Chem. Soc.*, **61**, 550 (1939).
13. McBain, J. W., and Betz, M. D., *J. Am. Chem. Soc.*, **57**, 1905, 1909, 1913, 1916 (1935).
14. McBain, M. E. L., Dye, W. B., and Johnston, S. A., *J. Am. Chem. Soc.*, **61**, 3210 (1939).
15. McBain, M. E. L., read at the Colloid Symposium at Boulder, Colo., June 1942, communicated to *J. Phys. Chem.*
16. McBain, M. E. L., *Proc. Roy. Soc. London*, **170A**, 415 (1939).
17. McBain, J. W., and Merrill, R. C., *J. Phys. Chem.*, **46**, 10 (1942).
18. McBain, J. W., and Wood, L. A., *Proc. Roy. Soc. London*, **174A**, 286 (1940).
19. Ralston, A. W., Hoerr, G. W., and Hoffman, E. J., *J. Am. Chem. Soc.*, **63**, 2578 (1941).
20. Moillet, J. L., Collie, B., Robinson, C., and Hartley, G. S., *Trans. Faraday Soc.*, **31**, 120 (1935).
21. Hartley, G. S., Collie, B., and Samis, C. S., *Trans. Faraday Soc.*, **32**, 796 (1936).
22. McBain, J. W., and Johnston, S. A., *Proc. Roy. Soc. London*, **181A**, —— (1942).
23. Ralston, A. W., and Hoerr, C. W., *J. Am. Chem. Soc.*, **64**, 773 (1942).
24. Laing, M. E., *J. Phys. Chem.*, **28**, 680 (1924).
25. See also McBain, J. W., and Thomas, W. M., *J. Phys. Chem.*, **40**, 997 (1936).

26. Kristnamurti, K., *Indian J. Phys.*, **3**, 307 (1929); Hess, K., and Gundermann, J., *Ber.*, **70**, 1800 (1937); Hess, K., Kiessig, H., and Philippoff, W., *Naturwiss.*, **26**, 184 (1938); Hess, K., Philippoff, W., and Kiessig, H., *Kolloid-Z.*, **88**, 40 (1939); Stauff, J., *Naturwiss.*, **27**, 213 (1939); Kiessig, H., Philippoff, W., *Naturwiss.*, **27**, 594 (1939).
27. McBain, J. W., *Nature*, **148**, 702 (1940).
28. Van Rysselberghe, P., *J. Phys. Chem.*, **43**, 1054 (1939).
29. McBain, M. E. Laing, *Trans. Faraday Soc.*, **31**, 153 (1935).
30. Stauff, J., *Kolloid-Z.*, **89**, 224 (1939).
31. Bury, C. R., *Trans. Faraday Soc.*, **31**, 195 (1935).
32. Lawrence, A. S. C., *Trans. Faraday Soc.*, **31**, 196 (1935).
33. Meyer, K. H., and van der Wyk, A. J. A., *Helv. Chim. Acta*, **20**, 1321 (1937).
34. McBain, J. W., *J. Phys. Chem.*, **43**, 671 (1939).
35. Jones, F. E., and Bury, C. R., *Phil. Mag.*, **4**, 841 (1927); Grindley, J., and Bury, C. R., *J. Chem. Soc.*, **1929**, 679; Davies, D. G., and Bury, C. R., *J. Chem. Soc.*, **1930**, 2263.
36. McBain, J. W., and M. E. Laing, *Trans. Chem. Soc.*, **117**, 1507 (1920).
37. Hartley, G. S., *Kolloid-Z.*, **88**, 22 (1939).
38. For example, McBain, J. W., and Searles, J., *J. Phys. Chem.*, **40**, 493 (1936).
39. McBain, M. E. Laing, and Perry, L. H., *J. Am. Chem. Soc.*, **62**, 989 (1940).
40. Hauser, E. A., and Grossman, A. J., paper presented at the 50th Anniversary Symposium, University of Chicago, Sept. 1941.
41. McBain, J. W., and Mills, G. F., *Reports on Progress in Physics*, **5**, 30 (1930).
42. McBain, J. W., "Advances in Colloid Science," **1**, 99, Interscience Publishers, Inc., New York, 1942.
43. Powney, J., and Addison, C. C., *Trans. Faraday Soc.*, **33**, 1248, 1254, 1510 (1937).
44. Adam, N. K., and Shute, H. L., *Trans. Faraday Soc.*, **34**, 758 (1938).
45. Alexander, A. E., *Trans. Faraday Soc.*, **38**, 248 (1942).
46. Hauser, A. E., and Niles, G. E., *J. Phys. Chem.*, **45**, 958 (1941).
47. Doss, K. S. G., *Current Science*, **4**, 415 (1935); *Proc. Ind. Acad. Sci.*, **4**, 11, 97 (1936); *Current Science*, **5**, 645 (1937); *Kolloid-Z.*, **84**, 138 (1938).
48. McBain, J. W., and Perry, L. H., *Ind. Eng. Chem.*, **31**, 35 (1939); McBain, J. W., and Spencer, W. V., *J. Am. Chem. Soc.*, **62**, 239 (1940); McBain, J. W., and Sharp, E. R., *J. Am. Chem. Soc.*, **63**, 1422 (1941).
49. Merrill, R. C., and McBain, J. W., *J. Phys. Chem.*, **46**, 10 (1942); *Ind. Eng. Chem.*, **34**, 915 (1942).

Applications of Electron Diffraction

LESTER H. GERMER

Bell Telephone Laboratories, New York, N. Y.

The diffraction of electrons by crystalline and pseudocrystalline substances has become a method of procedure for various chemical and physical investigations. Electron diffraction is thus an immediate "practical" application of the wave nature of matter. This is a position which it shares with the electron microscope, although the latter would probably have been invented independently of wave mechanics, and its present technical development would not have led to the discovery of electron waves.

The effectiveness of the electron diffraction method in obtaining desired information can be judged from the fact that it is being used in this country in thirty or forty laboratories, divided about evenly between educational and industrial institutions. Much of the electron diffraction work of these laboratories has been published in about 300 scientific papers, and about 700 papers have been published dealing with similar work carried out in foreign laboratories.

Of all these investigations, many are pioneer studies connected with the discovery of the diffraction of electrons and with the observation of the phenomenon under various conditions; others, and especially the more recent investigations, are more properly the application of the electron diffraction technique to various problems of more or less immediate and practical importance. Researches of the latter type only will be considered here. Since, furthermore, there exist comprehensive summaries of the literature on the subject, it seems undesirable to include any extensive bibliography or any account of more than a very few of the researches which have been made.[1] This paper will deal first very briefly with the fundamentals of the experimental technique; then with a consideration of the fields of usefulness of electron diffraction; and finally with some selected researches upon the structure of exceedingly small particles and exceedingly thin films.

Experimental Method

It is customary to produce electron diffraction patterns by the scattering of a beam of electrons of uniform speed. The speed, v, determines the electron wave-length, λ, through the de Broglie formula $\lambda = h/mv$, and the wave-length usually employed is of the order of 5×10^{-10} cm corresponding to electrons which have been accelerated by a potential difference of about 50 kv. The electrons of the primary beam must receive this acceleration in a vacuum chamber. After acceleration and collimation, the electron beam is scattered by the experimental specimen and the diffraction pattern of the specimen is recorded on a photographic plate placed perhaps 50 cm from it and normal to the beam. The evacuated chamber in which the electrons are produced, accelerated, collimated, scattered and recorded is commonly called an electron diffraction camera. It is attached to vacuum pumps, and is equipped with means for adjusting the specimen and the photographic plate while the camera is exhausted.

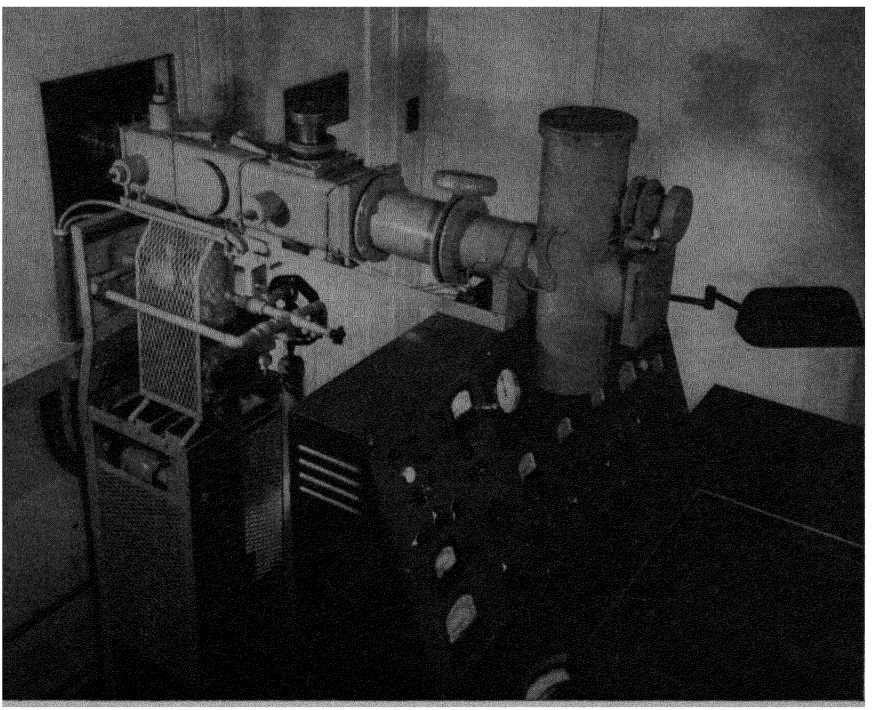


FIGURE 1. Electron diffraction camera used in the Bell Telephone Laboratories.

A photograph of the camera used in the Bell Telephone Laboratories appears as Figure 1, and a description of an earlier model of this apparatus was published several years ago.[2] Most experimenters have built their own cameras and many of these have been described in the literature (see, for example, the bibliography of Thomson and Cochrane [1]). Since an electron microscope can be modified to function as a diffraction camera,[3] this type of equipment may become widely used.

Fields of Usefulness of Electron Diffraction

Electron diffraction has been applied to the determination of the arrangement of atoms both in molecules in the gaseous state and in solid bodies. Although the

former is the more systematically developed field of usefulness,[4] it will not be considered here. We shall be concerned only with the diffraction of electrons by solid bodies, or rather with conclusions regarding the structure of solid bodies which can be drawn from the way in which they diffract electrons.

The atomic structure of solid bodies is precisely the type of information obtained by the methods of x-ray diffraction. Yet, because x-rays are very penetrating, and because electrons of the speeds commonly used are strongly scattered and thus penetrate very slightly, the fields of usefulness of the two techniques do not greatly overlap. If one attempts to describe the weakening of a primary beam of electrons of 50 kv equivalent speed in a mica crystal by means of an absorption coefficient, μ, one obtains[5]: $\mu \sim 2 \times 10^5$ cm^{-1}. For heavy metals it would be considerably larger. In empirically useful language, one can state that information regarding the structure of a substance made up predominantly of heavy atoms (atomic number of the order of 75) can be obtained only when the experimental conditions are such that the path of the electrons in the substance is considerably less than 10^{-5} cm in length. Suitable *upper limits* for length of path vary from about 500 A for gold to 2000 A for magnesium.

From these estimates of penetrating power it is clear that electron diffraction can be used to investigate only the structure of extremely thin films and extremely thin surface layers. Thus its field of usefulness is limited, but it is limited to a region which is in many cases of great fundamental and practical interest and which cannot be so adequately studied by any other known method.

Within its restricted field electron diffraction can be put to various uses. The most widely developed of these, leaving out of account investigations of the structure of gaseous molecules, is in microchemical analysis. An electron diffraction pattern from a solid block is characteristic of the substance at its surface. Comparison of the pattern from an unknown substance on the surface of a block with patterns obtained from known substances can enable one to identify the unknown. For this type of analysis to be effective, it is necessary that patterns from a large number of substances be observed and recorded in accessible form. A tabulation of the crystal spacings, and relative intensities of the corresponding reflections, for 1000 substances in powder form has been published by Hanawalt, Rinn and Frevel.[6] A committee of the National Research Council, under the direction of Dr. Wheeler P. Davey of Pennsylvania State College, is engaged in assembling and publishing sets of spacings for a larger number of compounds. The actual structures are available[7] for many compounds for which powder patterns happen never to have been recorded; identification of one of these appearing as an unknown substance at the surface of a block can frequently be made, although not so readily as if the powder pattern data were available.

This method of chemical identification by means of an electron diffraction pattern is similar to identification from an x-ray pattern. Either diffraction method yields more detailed information than any other analytical procedure; it determines not only what chemical elements are present but also just how they are combined, and, if the crystal structure of the substance is already known, precisely how the atoms are arranged in the crystals. In addition, either method will in general give a great deal of information in regard to the crystals themselves: information in regard to their average size; whether or not they are randomly oriented and, if not, quantitative data on the distribution in orientation; whether or not the crystals are seriously strained; even in some cases data regarding their shapes.

Since an x-ray pattern is produced by the body of a block and an electron pattern by its surface layers only, in some cases only the former yields useful information, and in other cases only the latter. In many instances the two diffraction patterns are characteristic of different compounds, or the same compounds in different proportions, and the patterns supplement each other; both are necessary for the knowledge

desired. As an illustration of the need for both types of analysis, one can cite studies of the oxidation and reduction of iron under conditions which are industrially important. For example, the electron diffraction pattern of Figure 2 from the surface

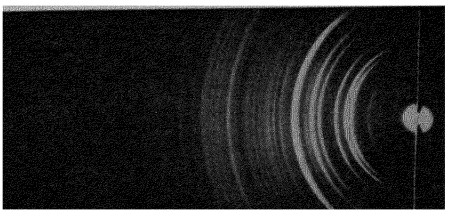


FIGURE 2. Electron diffraction pattern of Fe_2O_3 obtained from the surface of an iron plate.

of an iron plate can be identified as due entirely to the oxide Fe_2O_3, but the strongest lines of an x-ray pattern from this plate are those of Fe_3O_4, with weaker lines of iron and still weaker lines of Fe_2O_3. From the two patterns one can conclude that the iron was covered by a layer of Fe_3O_4 about 10^{-8} cm thick, on the surface of which was a 10^{-4}-cm layer of Fe_2O_3.

In the analysis of an unknown substance from its x-ray pattern the identification can be greatly aided by a supplementary spectroscopic analysis to determine the chemical elements which are present. In fact, in a laboratory where such analyses are frequently made it may be the usual procedure to make the spectroscopic analysis first, and to follow this by the x-ray analysis. On the other hand, identification from electron patterns cannot in general be aided by spectroscopic analysis. The surface layers which produce the electron patterns are frequently too thin to permit a spectroscopic analysis; if they were *not* too thin, then probably the x-ray method would be applicable; the latter is always more convenient if the layer to be studied is sufficiently thick. Because of technical difficulties such as those arising from an unfavorable surface contour of the specimen, and because aid cannot, in all cases at least, be obtained from spectroscopic or ordinary microchemical analysis, identification of a surface film from its electron diffraction pattern is in general more difficult than the identification of the bulk material from its x-ray pattern. Yet the electron diffraction method is being widely used in industrial laboratories and elsewhere for the identification of surface films because of the precise knowledge which it is capable of giving and because in a great many cases there is no other available method.

Although microchemical analysis of surface layers is at the present time the most widely developed use of electron diffraction from solid bodies, analyses of this sort are not the subject of the bulk of the papers on electron diffraction, nor is microchemical analysis as such by any means the most interesting application. Others are extremely diverse and poorly systematized, but they seem at present to offer potentialities of great future development.

One thinks at once of using electron diffraction to get information about catalytic processes. Although catalytically active surfaces are being studied by electron diffraction in some industrial laboratories, one cannot judge how extensive and how successful this work is because very few of the results have been published.[8]

From a great number of diverse applications of electron diffraction I have chosen to describe here three which have to do with particles of sub-colloidal dimensions, or with films only a few molecules thick. The choice was arbitrary and was dictated by familiarity with the researches. These studies can perhaps be thought of as belonging to a branch of science appropriately described as surface chemistry or surface physics. Thought of in this way, they are representative of the most interesting, and although the least well developed perhaps ultimately the most important of the uses which can be made of electron diffraction. They illustrate furthermore the extreme sensitivity of the method in its application to very small amounts of material.

Electron Diffraction Studies

Diffraction of Electrons by Small Crystals.[9] By using an extremely thin organic foil as a support, very small amounts of various elements and inorganic compounds have been studied by electron diffraction by the transmission method. With a foil as thin as 100 Å the electron scattering from the foil makes an inappreciable addition to the scattering pattern from a deposited layer of heavy metal atoms as thick or thicker on the average than two or three atomic layers. The method is therefore adaptable to the determination of the arrangement of atoms in exceedingly thin films.

In one series of experiments diffraction patterns have been obtained from films of widely different mean thicknesses and of various substances deposited in vacuum upon a Formvar supporting foil by vaporization from a tungsten ribbon. A pattern from a cesium iodide film of 30 Å mean thickness is shown in Figure 3. From esti-

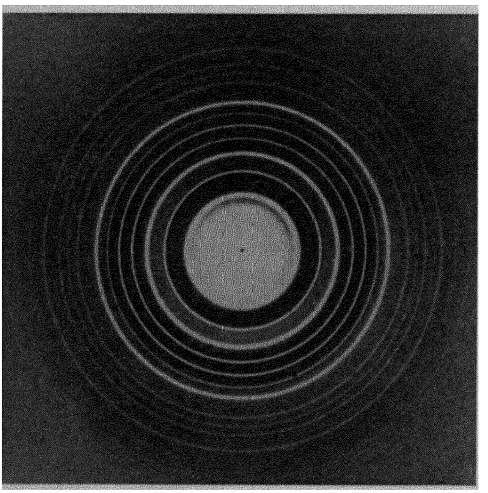


FIGURE 3. Diffraction pattern from a film of cesium iodide of 30 Å mean thickness.

mated breadths at half maximum ΔR of the rings of this pattern, one can calculate the mean crystal size C by means of the well-known Scherrer formula, which can be written in the form $C = L\lambda/\Delta R$, where $L\lambda$ is the product of the specimen-plate distance and the electron wave-length. In this manner one determines that the crystals of cesium iodide in this film have average linear dimensions of about 200 Å. This value, together with the fact that the amount of deposited material was sufficient to form a layer only 30 Å thick on the average, leads to the inevitable conclusion that, under the conditions of formation of the film, the molecules are unstable when spread out uniformly, and that they cluster at once into quite large crystals.

Somewhat similar deductions have been made regarding conditions of stability in films of many different substances, although in many cases the mean crystal size has been found to be much smaller than 200 Å. In particular, metal atoms in general form smaller crystals than do molecules of most inorganic compounds which have been investigated.

A rather extensive study of copper films has been carried out, the metal in these experiments being vaporized upon supporting Formvar foils within the diffraction camera itself in order that in each test a diffraction pattern could be obtained without exposing the freshly prepared film to oxidation in the air. The films of copper studied have had various mean thicknesses, the greatest being 780 Å, and the least 3 Å, which is only slightly larger than the diameter of a copper atom. All films

thicker than 20 Å produce diffraction patterns characteristic of the face-centered cubic structure in which copper crystallizes. The sharpness of the rings varies somewhat from one pattern to another, the rings of the patterns from the thicker films being narrower than those of the patterns from the thinner films. This variation of sharpness arises from differences in mean crystal size in the different films, the calculated size varying from about 70 Å in the thickest films to 50 Å in 20-Å films. It is clear in the first place that crystals are never larger than 70 Å no matter how thick the film, and in the second place that in films as thin as 20 Å the originally uniform distribution of atoms is unstable, and they draw together to make 50-Å crystals.

Diffraction patterns from films thinner than 20 Å appear at first to be qualitatively different. The rings are not only broader than those from thicker films, but they appear to be altered in relative intensity, so that a pattern can no longer be recognized as that of face-centered cubic crystals. The interpretation of this is quite simple. A spherical crystal of copper of 20 Å diameter contains somewhat less than 400 atoms. The resolving power of crystals of this size is sufficiently good so that a resulting powder pattern will be readily recognized as that of face-centered cubic crystals. If, however, the number of atoms in the individual crystals is much less than 400, the resolving power will be so poor that the characteristic "face-centered cubic reflections" are no longer distinguishable. In Figure 4 are given calculated scattering curves for randomly oriented spherical or near-spherical face-centered cubic crystals containing various small numbers of atoms. (In order that these curves shall be applicable to crystals of different metals, the atomic scattering factor of copper has not been included in the calculation; this would introduce a factor which is continuously decreasing with increasing angle.) The locations, Miller indices, and relative intensities of reflections from extended face-centered cubic crystals are indicated at the bottom of the figure. By comparing these with the calculated scattering curves above, it is clear that crystals containing 379 atoms each will produce a pattern from which the face-centered cubic arrangement of atoms can be readily deduced, but that crystals of 55 atoms each, or any smaller number, cannot be interpreted in this manner. It would certainly be utterly impossible to interpret an experimentally observed pattern like the calculated pattern for 55-atom crystals, unless one had independent reasons for suspecting a face-centered cubic structure.

When the electron scattering pattern from the film of copper of 3 Å mean thickness is compared with the calculated curves of Figure 4 it is discovered that the experimental pattern is intermediate in character between the curve for the 379-atom crystals and that for crystals of 55 atoms. Thus it appears clearly that the anomalous patterns obtained from extremely thin copper films are accounted for by the small number of atoms in the individual crystals and their resulting poor resolving power. It is possible to estimate from the *character* of the pattern from the 3 Å film that the individual crystals contain on the average about 200 atoms and that they are more or less equi-axed. Thus, for this thinnest film also, the originally uniformly distributed layer of atoms is unstable and crystals of appreciable dimensions are formed. For such excessively small crystals the *character* of the scattering pattern gives a quantitative measure of mean size, which is a considerable extension of the range of crystal size which can be determined.

Arrangement of Molecules in Monolayers and Multilayers. Layers of various long-chain organic molecules can be built upon the surface of a block, one layer at a time, by the well-known Langmuir-Blodgett technique, in which molecules in a single layer floating upon water are transferred to the block when it moves through the water surface. Electron diffraction patterns of such built-up layers of molecules yield information regarding the structure of the layers. The patterns are simple and readily interpreted, because in each molecule the separation between alternate carbon atoms is repeated over and over, because the carbon atoms of each molecule are coplanar, and because the molecules of any one layer align themselves with their

axes parallel. The information regarding the structure of these built-up layers of long molecules is of interest in itself; but perhaps of more importance is the ex-

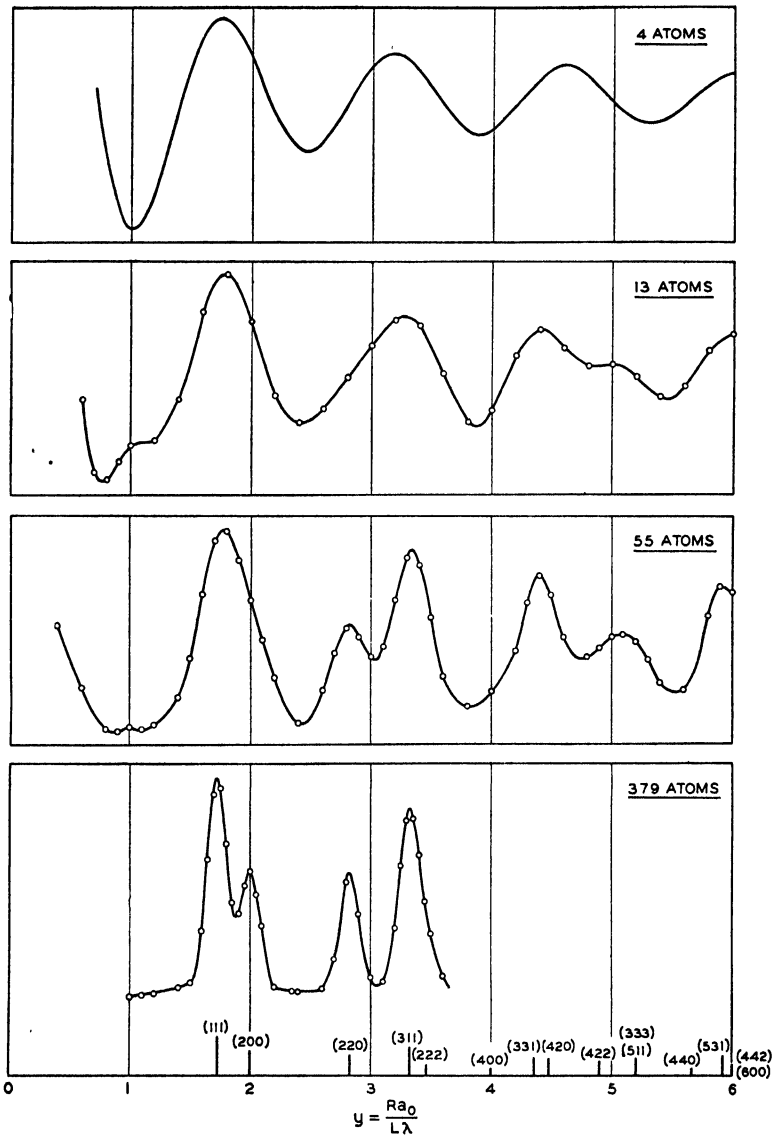


FIGURE 4. Theoretical scattering curves from randomly oriented face-centered cubic crystals made up of various numbers of atoms.

perience which experimenters have gained by interpreting these patterns. This has enabled them to attack successfully diffraction patterns of surface films made up of organic molecules having less simple shapes, although the latter investigations are still of a rather rudimentary character. It will suffice here to reproduce patterns obtained from a few layers of long-chain molecules, to indicate how these patterns

are interpreted, and to state conclusions drawn from a systematic study of built-up films containing various numbers of layers of two simple organic molecules.[10]

The diffraction patterns of Figure 5 were produced respectively by a single layer of barium stearate molecules upon the surface of a metal block, and by three layers

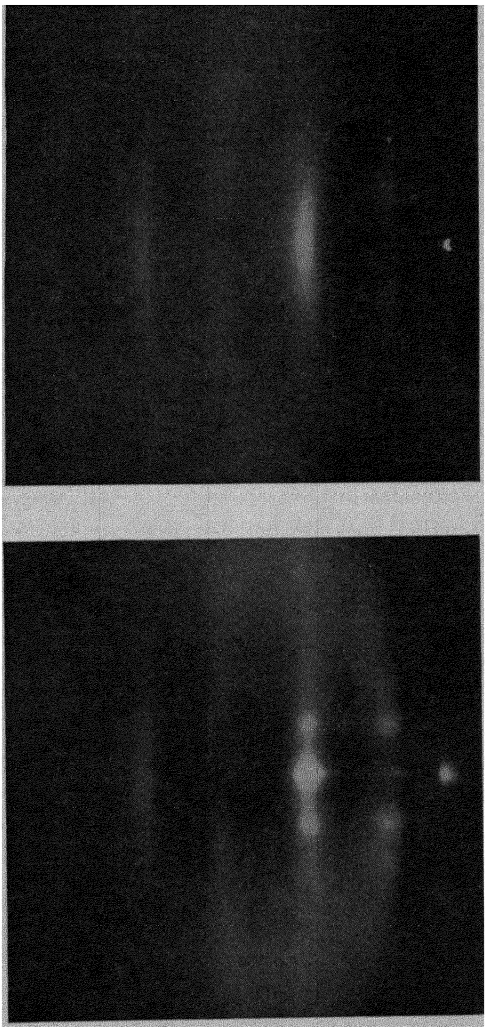


FIGURE 5. Diffraction patterns from barium stearate.
A. A single layer of molecules.
B. Three layers of molecules.

of molecules. The separation, R, between the diffuse vertical bands of these patterns corresponds by the Bragg formula, which can be written $Rd = L\lambda$, to a spacing $d = 2.50$ Å, repeated many times in a direction normal to the surface; observation of this spacing is conclusive proof that the bands are due to hydrocarbon chains standing perpendicular to the surface. The distribution of scattering intensity is markedly different along the bands of even order from that along the bands of odd order. This difference arises from the zig-zag nature of the chains and from the

fact that all the carbon atoms of each chain lie in one plane. The diffraction features lying upon the bands of Figure 5B are produced by regularities in the lateral arrangement of the molecules, and it is easy to calculate, from the positions of these features, that the molecules are built into a number of structures in each of which the axes of the molecules intersect the corner-points of an extended, two-dimensional, close-packed hexagonal net lying parallel to the surface, the side of the net being 4.85 Å. From the absence of any diffraction features on the bands of the pattern of Figure 5A we must conclude that, in the first layer of barium stearate molecules upon the surface, there is no regularity of lateral arrangement; in the first layer the molecules are packed closely together but without definite regularity.

Electron diffraction patterns produced by stearic acid molecules built upon a metal surface show some similarities to patterns from barium stearate molecules, but there are also differences. The stearic acid molecules produce diffuse bands corresponding to the same separation of 2.50 Å; but, whereas stearic acid molecules in the first layer stand normal to the supporting surface as do the molecules of barium stearate, in the second and subsequent layers stearic acid molecules stand with their axes inclined to the surface. The inclination to the surface is always by a definite angle which lies in the plane that was vertical when the block was dipped through the water surface. The stearic acid molecules of all layers except the first are built, furthermore, into true crystals having monoclinic symmetry. Some significant parameters of these crystals can be determined, and these parameters agree with previous knowledge of stearic acid crystals and extend such knowledge.

It is interesting to note the differences between the soap or acid molecules of the first layer and similar molecules in layers beyond the first. Molecules which touch the metal surface stand normal to it and do not possess any regularity in their lateral arrangement; but in subsequent layers molecules are built into compact crystalline arrays determined by the habit of the substances in bulk. The simplest interpretation seems to be chemical attraction between the surface and the metal atoms of the soap or the carboxyl groups of the acid, this attraction being strong enough to hold the first layer of molecules fast in certain positions, which are determined by the surface atoms rather than by the natural habit of the built-up layers.

This interpretation is supported by a series of electron diffraction experiments upon built-up layers which have been rubbed. Rather light rubbing takes off all except the first layer of barium stearate or stearic acid molecules; but the first layer is exceedingly resistant to removal by rubbing with material which does not abrade the underlying metal itself. It is possible by extremely light rubbing in one direction to "upset" stearic acid molecules of all layers except the first, so that they then lie with their axes nearly, but not quite, parallel to the surface. In this condition the molecules are moderately resistant to continual light rubbing in the same direction, but they are removed by almost the lightest possible rub in the opposite direction.

The Structure of Black Carbon.[11] Electron diffraction studies of what was formerly called amorphous carbon have made significant additions to our knowledge of its structure. The most informative of these studies have been carried out upon carbon deposited upon quartz blocks by pyrolysis of methane. The resulting carbon films are found in many cases to exhibit marked preferential orientation, and it is to this orientation that the ease of interpretation of the unusual diffraction patterns must be attributed. These patterns are more readily interpreted than those from randomly oriented carbon having the same structure, for the reason that a diffraction pattern from a single crystal, which is being rotated about an important crystal direction, can yield more information than a Debye-Scherrer pattern from many small crystals.

In Figure 6 is reproduced one of the electron diffraction patterns from pyrolytically deposited carbon having very strong orientation. The uniformly spaced, heavy, diffuse spots along the horizontal center line correspond to a separation of

3.66 Å normal to the surface of the supporting block. This distance is about 9 per cent greater than the separation between adjacent hexagonal planes in graphite, and from the result of this comparison it is natural to infer that the pattern of Figure 6 was produced by graphite-like crystallites oriented with their hexagonal planes predominantly parallel to the supporting quartz surface.

Along the vertical line at the shadow edge one can find on the original plate 5

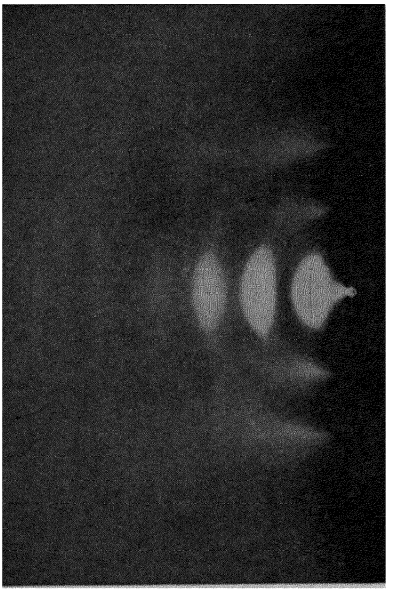


FIGURE 6. Diffraction pattern from very small oriented pseudo-crystals of black carbon.

pairs of diffraction features having radii which are proportional, as closely as they can be measured, to the square roots of the integers 1, 3, 4, 7 and 9. These ratios are just those of (h k o) reflections from graphite crystallites, and the separation between atoms turns out to be 2.46 Å, which is just that between atoms in the hexagonal planes of graphite. Yet the relative intensities of these reflections are not those to be expected from graphite. It is easy to discover that the intensities correspond to reflections from parallel, but otherwise utterly unrelated, planes of carbon atoms, in each of which the atoms are arranged as in the basal plane of graphite. The intensities are quite different, on the one hand from what they would be if the planes of atoms were built into true graphite crystals, and on the other hand from what they would be if the atoms in the parallel, but otherwise unrelated, planes had a close-packed arrangement (*e.g.*, the atoms in the hexagonal planes of magnesium or zinc). This deduction from the intensities of the (h k o) reflections is borne out by the fact that no trace whatever can be discovered of any general reflections of the type (h k l).

These two sets of diffraction features on the pattern of Figure 6 lead thus to a fairly clear qualitative picture of the structure of the carbon film: It is made up of sharply oriented pseudo-crystals in each of which the atoms are arranged in hexagonal planes which are exactly like those of graphite. In each pseudo-crystal these planes are parallel and equi-spaced by distances about 9 per cent greater than corresponding planes in graphite. Yet there are no regularities in the lateral positions of planes belonging to the same crystallite, and they are therefore not true crystals at all.

These deductions, which can be made in such a very simple manner, are sub-

stantiated by transmission patterns produced by electrons which have passed *through* similar films of carbon, as well as by a more penetrating analysis of the pattern of Figure 6. This more detailed analysis, which will not be given here, leads also to quantitative values for average dimensions of the pseudo-crystals and to a mathematical expression representing approximately the law of distribution of orientation of the individual psuedo-crystals about the surface normal. The average dimensions discovered in this manner are very small—rather astonishingly small in fact. Parallel to the hexagonal planes it is found that the average dimension is 20 Å, or about 5 unit hexagons, and normal to these planes it is 10 or 15 Å, or 3 or 4 layers of atoms.

This structure represents the most degenerate form of pseudo-graphitic carbon. Previous x-ray investigations had revealed an extended series of carbon starting with graphite and having progressively less regular structures. By an extrapolation process in this series of x-ray studies an inferred structure had been previously assigned to this very degenerate carbon. Although this postulated structure agrees with that ascertained in the present investigation, the electron diffraction studies give the first reliable proof of the true structure.

These three researches were chosen for mention here primarily because they illustrate the extreme sensitivity and delicacy of the electron diffraction technique. Each of these studies was concerned with the structure of units of matter so very small that their dimensions are given in terms of atom diameters or of lengths of molecules more appropriately than in ordinary units of length. Many investigations have been carried out in this sub-colloidal region, but the possibilities are still scarcely realized.

References

1. An up-to-date account of most electron diffraction investigations is contained in the book by Thomson and Cochrane "Theory and Practice of Electron Diffraction," Macmillan, 1939. The many determinations of molecular structure by electron diffraction are well described in three papers by Brockway, L. O., "Electron Diffraction by Gas Molecules," *Rev. Mod. Phys.*, **8**, 231-266 (1936); Beach, J. Y., "Electron Diffraction Investigation of the Structure of Gas Molecules," Recent Advances in Surface Chemistry and Chemical Physics, Publications of the A. A. A. S., No. 7, 88-97 (1939); and Maxwell, L. R., "Symposium on Optical Methods for Study of Molecular Structure. II. The Electron Diffraction Method," *J. Optical Soc. Am.*, **30**, 374-395 (1940). From among a number of other general treatises on the subject of electron diffraction I wish to list here only Kirchner, F., "Elektroneninterferenzen und Roentgeninterferenzen," *Erg. d. exakt. Naturwiss.*, **11**, 64-132 (1932) and Finch, G. I., and Wilman, H., "The Study of Surface Structure by Electron Diffraction," *Ibid.*, **16**, 353-436 (1937).
2. Germer, L. H., *Rev. Sci. Inst.*, **6**, 138-142 (1935).
3. Hillier, J., Baker, R. F., and Zworykin, V. K., *J. App. Phys.*, **13**, 571-577 (1942).
4. See the reviews of Brockway, Beach, and Maxwell.[1]
5. Kirchner, F., *Ann. d. Phys.*, **13**, 38 (1932).
6. Hanawalt, J. D., Rinn, H. W., and Frevel, L. K., *Ind. Eng. Chem., Anal. Ed.*, **10**, 457 (1938).
7. See Ewald, P. P., Hermann, C., *et al.*, "Strukturbericht," Vol. 1 to 6.
8. See, however, papers by Beeck, O., and his collaborators, *Phys. Rev.*, **55**, 601-602 (1939); *Proc. Roy. Soc. London*, **A177**, 62 (1941).
9. Germer, L. H., *Phys. Rev.*, **56**, 58-71 (1939); Germer, L. H., and White, A. H., *Phys. Rev.*, **60**, 447-454 (1941).
10. Germer, L. H., and Storks, K. H., *Proc. Nat. Acad. Sci.*, **23**, 390-397 (1937); *J. Chem. Phys.*, **6**, 280-293 (1930); *Phys. Rev.*, **55**, 648-654 (1939).
11. White, A. H., and Germer, L. H., *J. Chem. Phys.*, **9**, 492-497 (1941).

Methods for the Analysis of Complex Molecular Structures with the Aid of X-Rays

MAURICE L. HUGGINS

Eastman Kodak Company, Rochester, N. Y.

The analysis of a complex structure with the aid of x-rays is possible only if the structure is composed of many like (or very similar) units, repeated at uniform intervals. A few very large virus molecules may have internal structures which satisfy this requirement, but in general it is satisfied only by crystalline aggregations of many molecules. In any case, the problem is essentially that of determining the structure of a complicated *crystal,* and so we shall treat it here. At best, this is not a simple problem and one cannot expect, without a prohibitive amount of labor, to deduce accurately the relative positions of all of the atoms in really complex structures. Yet the general methods of procedure have now been developed to such a degree that much more can be accomplished in this direction than a few years ago, and there is every reason to expect further progress in the future.

The aim of this paper is to outline the general procedures now in use for complex structure analyses. For details, one must of course go to other sources.[6, 8, 22]

Outline of Structure Analysis Methods

When a beam of x-rays impinges on a crystal, diffracted beams are produced in many directions. Each beam is customarily designated by a set of three indices ($h\ k\ l$) which give the orientation in the crystal of a set of parallel planes which may be visualized as "reflecting" that beam. (The indices are integers, proportional to the reciprocals of the intercepts of each reflecting plane on the crystal axes, these intercepts being measured as fractions or multiples of the unit distances.) Given a large number of reflections, the crystal structure analyst determines, by procedures we shall not take space to discuss, the smallest unit cell, consistent with the external symmetry of the crystal, which will account for these reflections—that is, the smallest cell, on the basis of which integral indices can be assigned to all the reflections observed.

When the dimensions of the unit cell have been determined, the numbers of molecules and of atoms of each kind that it contains are readily computed from the chemical formula of the substance, the atomic weights, and the density.

A considerable amount of information about the internal symmetry of the crystal —the existence and spatial distribution in the unit cell of centers, planes and axes of symmetry—can be obtained from a study of the indices of the observed reflections, noting classes of reflections (*e.g.,* those having all indices odd) which are entirely absent, etc. Sometimes, but not always, the complete distribution of symmetry elements in the structure (the "space group") can be determined in this way, with the assistance of tables[1, 16] of the requirements for each space group. If the correct combination of symmetry elements (space group) cannot be determined uniquely by this procedure, certain combinations can frequently be eliminated from further consideration by virtue of the fact that they will not permit any arrangement having the known numbers of atoms in the unit cell, or having the atoms grouped together

into molecules of the known chemical formula and the known or assumed molecular size, shape, and symmetry.

After as much as possible has been determined about the internal symmetry of the crystalline structure, the next step is to consider what arrangements of atoms and molecules are possible, consistent with that symmetry and with other known or assumed facts, such as the numbers of atoms of each kind per unit, the distribution of these atoms into molecules, minimum interatomic distances, etc. With the simplest types of structures, such as those found for many inorganic substances, the exact positions of all the atomic centers are determined by the symmetry properties and dimensions of the unit cell. In slightly more complicated structures, the positions of some or all of the atoms are not so determined, but can be deduced by comparisons of relative intensities of different reflections, with or without considerations of probable or reasonable interatomic distances. For still more complicated structures, complete structure analyses can sometimes be obtained by trial-and-error methods, guided by the experimental intensity data. Once a structure has been deduced, it is of course checked by comparison between calculated and experimental intensities of the observed reflections. It must also account satisfactorily for other known physical properties of the crystals.

The relative intensities of different reflections from a crystal depend not only on the kinds and distribution of atoms in it, but also on various other factors, such as the shape of the specimen, whether it is a powder or a single crystal, the wave-length of the x-rays used, the angles between incident and reflected beams, the position of the photographic film relative to the crystal sample and the incident beam, the characteristics of the photographic emulsion, etc. In analyzing a complicated crystal structure it is useful to compute from the observed intensities, as accurately as possible, a set of "structure factor squares," F^2_{hkl}, these giving the effect of the crystal structure on the intensities, after correcting for the other factors related to the experimental technique. For any assumed distribution of atoms in the unit cell one can calculate (by addition, in a prescribed manner, of appropriate cosine and sine terms for each atom) the magnitudes of the structure factor squares to be expected and then compare these calculated values with the experimental ones.

Fourier Syntheses

If the structure is too complicated for satisfactory solution by trial-and-error methods, or if, after a partial or approximate solution, a more complete or more accurate knowledge of the structure is desired, one can add together the structure factor squares in such a way as to yield directly much useful information, as we shall see. Even more useful are summations of the structure factors themselves, in those cases for which it is possible to determine the sign of each.

The structure factor is the product of a (positive) amplitude factor and a complex phase factor, the latter being undeterminable directly from the experimental data. If the structure has a center of symmetry at the origin of coördinates (any corner of the unit cell), the phase factor must have the value $+1$ or -1. The magnitude of the structure factor can then be obtained by taking the square root of the corresponding structure factor square. Its sign remains unknown, unless one has some information about the structure, other than that embodied in the experimental F^2 values. If the positions of the heavier atoms (those having the greatest scattering power for x-rays) in the unit cell are approximately known, the signs of all the more important structure factors can be computed with sufficient certainty. For example, if it is known that most of the electrons in the unit cell are concentrated in a single atom, having its center at a center of symmetry, one can choose the origin there; the signs of all the structure factors (or at least all of those not having very large indices) are then positive.

Knowing the magnitudes and signs of the structure factors, (ideally for all pos-

sible sets of indices, actually for all those corresponding to observed diffractions), the Fourier series summation

$$\rho_{xyz} = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F_{hkl} \cos 2\pi (hx + ky + lz), \tag{1}$$

in which each structure factor is multiplied by the appropriate cosine, gives [4, 9, 10, 12] the (time-average) electron density at the point in the unit cell having coördinates x, y, z. The V in this equation represents the volume of the unit cell. Summations for a sufficient number of points enable one to determine the atomic distribution in the unit. The center of each atom is obviously located at an electron density peak, and the higher the peak (in general) the more electrons in the atom.

As indicated in Equation 1, the electron density is given by an *infinite* Fourier series. The set of F values available experimentally being far from infinite, the results obtained from the corresponding incomplete series are in error. Error also results from inaccuracy in the experimental intensities and in the corrections involved in changing intensities to F^2 values. Fortunately, it can be shown that with reasonable care and by adopting certain procedures which we need not discuss here, the resulting errors can usually be made negligible (or nearly so), if one is interested only in the positions and identification of the atoms in the structure and not in the exact electron distribution within each atom and between the atoms.

To cover the complete unit cell thoroughly in the way just outlined would take a great deal of time and labor, but this can be obviated by certain modifications of the procedure. The desired information can frequently be obtained by computing the electron density only for certain planes or along certain lines cutting through the unit cell. For example, the density distribution in the XY plane (the face of the unit parallelopiped which includes the X and Y axes) is given by the equation

$$\rho_{xy0} = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F_{hkl} \cos 2\pi (hx + ky) \tag{2a}$$

$$= \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} G_{hk} \cos 2\pi (hx + ky) \tag{2b}$$

where

$$G_{hk} = \sum_{l=-\infty}^{\infty} F_{hkl} \tag{2c}$$

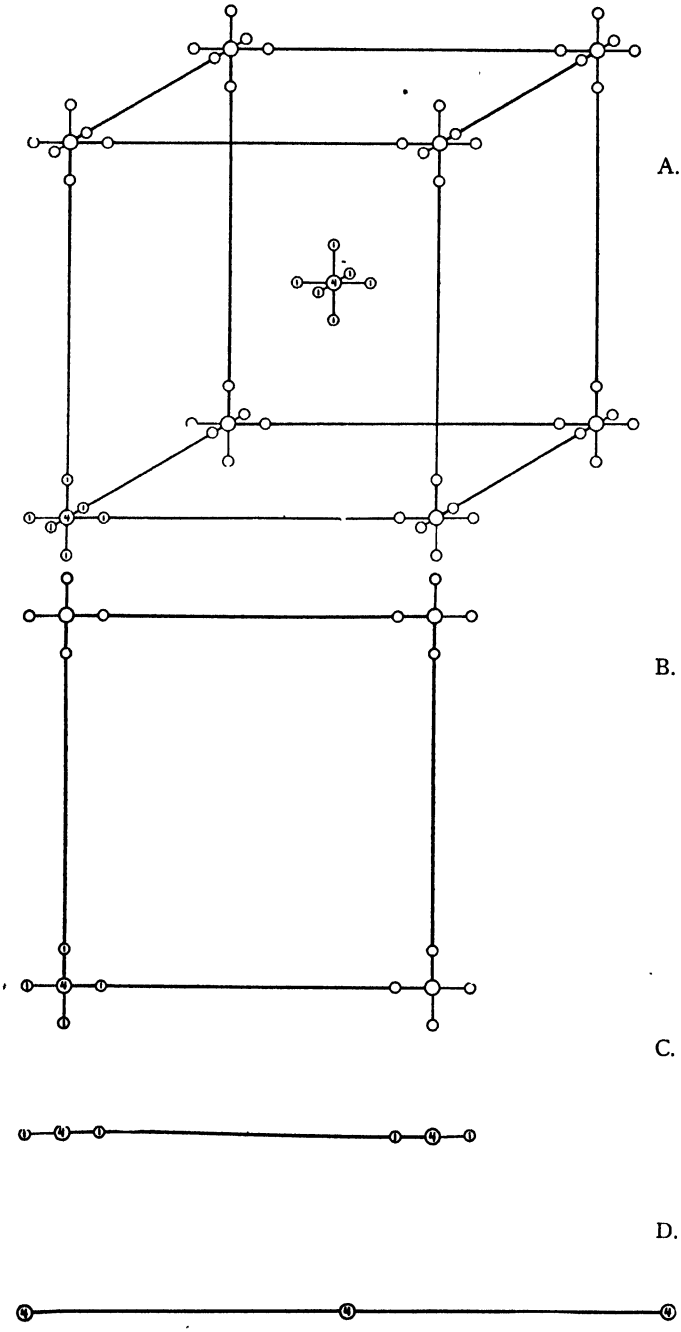
Likewise, the density along the X axis is given by

$$\rho_{x00} = \frac{1}{V} \sum_{h=-\infty}^{\infty} H_h \cos 2\pi (hx), \tag{3a}$$

where

$$H_h = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F_{hkl} \tag{3b}$$

In practical structure analysis, two-dimensional summations,[4] including only terms for reflections corresponding to planes in certain zones (*e.g.*, only $h\ k\ 0$ reflec-



A. A little more than the unit cell of the structure. This same electron distribution would be obtained by calculating the electron density, using Equation 1, at a sufficient number of points.

B. Electron distribution in the XY plane ($z = 0$). This would be obtained by the application of Equation 2. (For this structure, the distribution in the XZ or YZ plane is exactly the same.)

C. Electron distribution along the X axis ($y = z = 0$), obtainable by Equation 3. (The distribution along the Y or Z axis would be the same).

D. Electron distribution along the cube diagonal ($x = y = z$). (The distribution along each of the other cube diagonals would be the same).

FIGURE 1. Electron distribution in a simple hypothetical crystal structure. Atoms are assumed to be points with scattering powers equal to the numbers of electrons given in the circles.

tions), are more frequently used. Such a summation yields a projection of the electron density in the unit cell onto a plane normal to the zone axis. For instance,

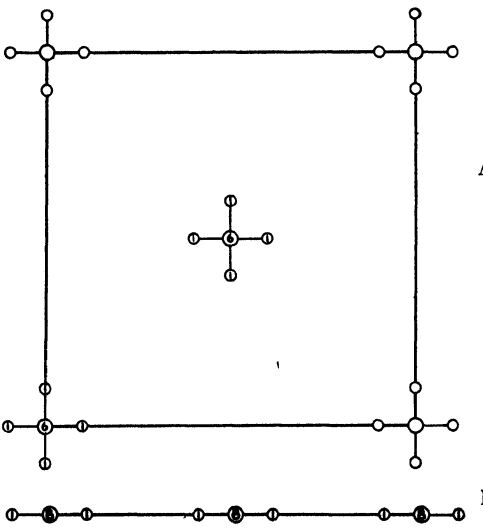
$$\rho_{xy} = \frac{1}{A} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} F_{hk0} \cos 2\pi (hx + ky) \tag{4}$$

gives a projection of the electron density (parallel to the Z axis) onto the XY plane. A is the area of the projection of the unit cell, in this case equal to the area of the XY face of the unit cell. With such a projection for each of the three axial planes (XY, XZ, YZ) the complete distribution of atoms in a structure can be determined, at least theoretically.

One-dimensional summations,[9] giving projections of the unit cell onto a line through it, are also useful. For a projection onto the X axis, the equation is

$$\rho_x = \frac{1}{L} \sum_{h=-\infty}^{\infty} F_{h00} \cos 2\pi (hx) \tag{5}$$

L is the length of the projection of the unit cell, in this case equal to the unit distance (a_0) in the X direction.



A. Projection onto the XY (or XZ or YZ) plane (Equation 4).

B. Projection onto the X (or Y or Z) axis (Equation 5).

FIGURE 2. Projections of the electron distribution in the hypothetical structure of Figure 1A.

Figures 1 and 2 show what would be obtained ideally, using infinite series and absolutely accurate data, from a simple hypothetical structure (Figure 1A) consisting of point atoms. In an actual case, the atoms show up as peaks (rather than points) of electron density. The projections and sections are therefore represented as contour maps. Figures 6B and 9C are typical examples.

If enough is known about the structure from a preliminary trial-and-error analysis to make possible the determination of the signs of the structure factors (assuming the structure to have a center of symmetry) for the more intense reflections, a summation including only terms corresponding to these more intense reflections may enable one to obtain the signs of the structure factors for additional

reflections and so, by successive approximations, to arrive at the final detailed structure.

The following summations, which are like those of Equations 1 to 5 inclusive except for the replacement of F_{hkl} by F^2_{hkl} and the omission of the $\frac{1}{V}$, $\frac{1}{A}$ and $\frac{1}{L}$ factors, are also useful.

$$P_{xyz} = \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F^2_{hkl} \cos 2\pi (hx + ky + lz) \tag{6}$$

$$P_{xy0} = \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F^2_{hkl} \cos 2\pi (hx + ky) \tag{7a}$$

$$= \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} G'_{hk} \cos 2\pi (hx + ky) \tag{7b}$$

$$G'_{hk} = \sum_{l=-\infty}^{\infty} F^2_{hkl} \tag{7c}$$

$$P_{x00} = \sum_{h=-\infty}^{\infty} H'_h \cos 2\pi (hx) \tag{8a}$$

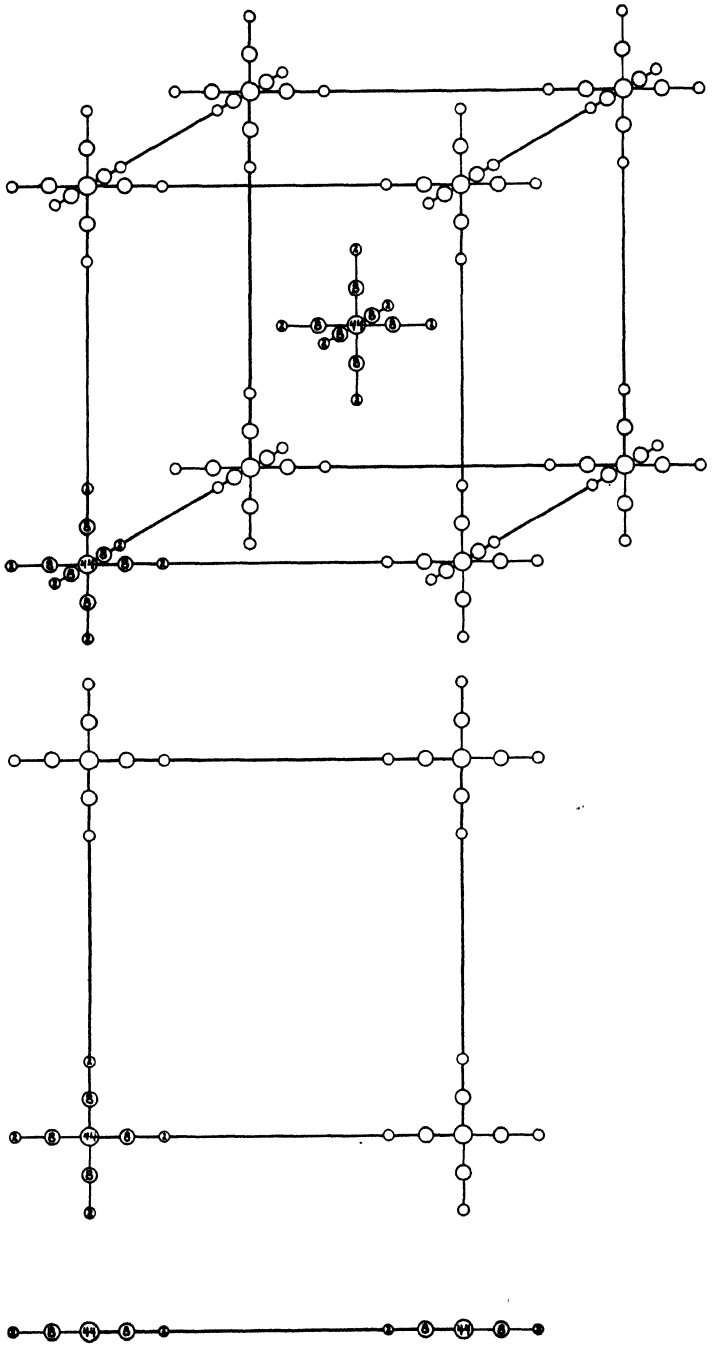
$$H'_h = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F^2_{hkl} \tag{8b}$$

$$P_{xy} = \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} F^2_{hk0} \cos 2\pi (hx + ky) \tag{9}$$

$$P_x = \sum_{h=-\infty}^{\infty} F^2_{h00} \cos 2\pi (hx) \tag{10}$$

They require no knowledge of the phases and so no preliminary knowledge of the structure. Such summations have been shown by Patterson [20] and Harker [11] to yield "vector structures" [20] and projections [20] and sections [11] of such structures (Figures 3 and 5), from which one can deduce the principal interatomic distances and directions in the structure. This knowledge may suffice to determine the phases of the reflections for electron density summations or to enable the complete structure to be worked out in other ways.

The physical meaning of the Patterson summations can be explained by reference to Figure 4. In this figure, *a* represents a distribution of point atoms along a line. In *b*, all the vectors *from* each atom in the unit *to* each atom (including itself) in the unit are represented, each vector terminus being given a weight equal to the product of the numbers of electrons in the atoms at the two ends of the vector. In order to show these vectors individually, the lines representing them are here separated by shifting them vertically with respect to each other. In *c*, these vectors are all shifted horizontally, so as to make their origins coincide, except for the vertical displacement. Without this vertical displacement, the result is as indicated by the heavy line (and the circles it connects) in *d*. (If two vectors terminate at the same point, their weights are added.) These circles represent the Patterson



A. A little more than the unit cell (Equation 6).

B. Vector distribution in the XY (or XZ or YZ) plane (Equation 7).

C. Vector distribution along the X (or Y or Z) axis (Equation 8).

FIGURE 3. Patterson vector structure corresponding to the hypothetical structure of Figure 1A.

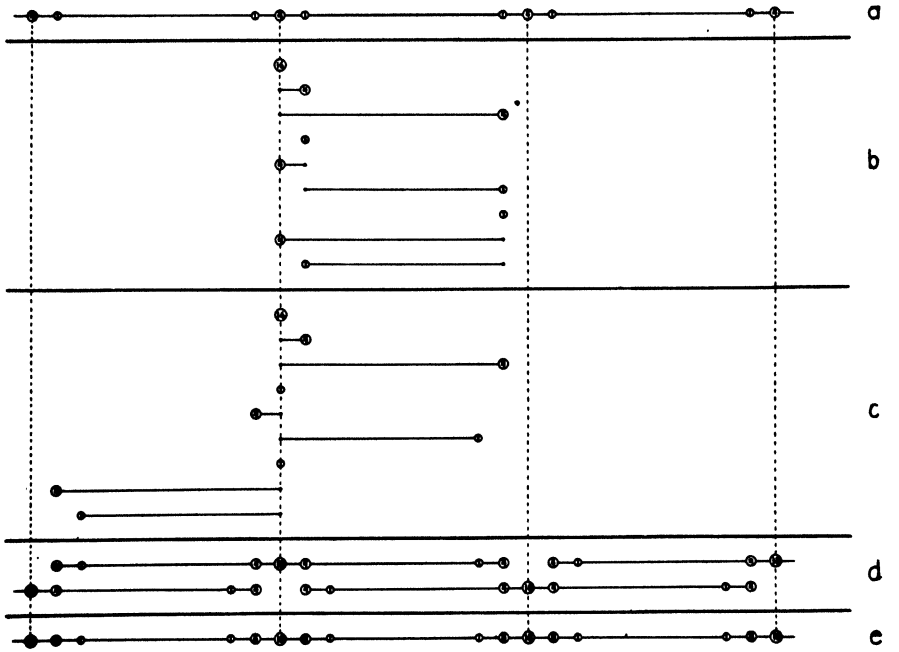
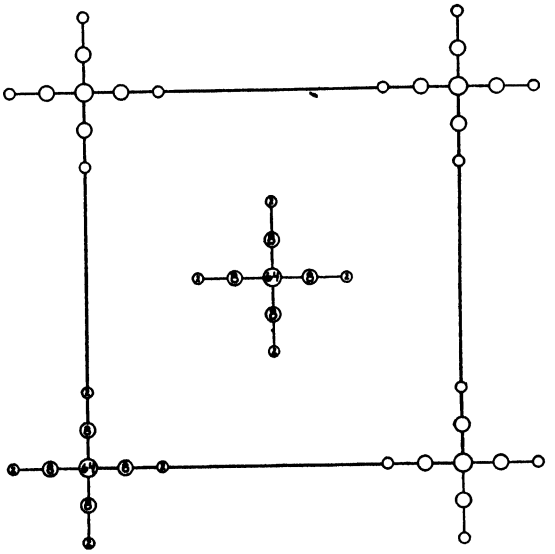
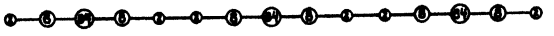


FIGURE 4. Relation between a hypothetical one-dimensional structure of point atoms (a) and its Patterson vector structure (e).



A. Patterson projection onto the XY (or XZ or YZ) plane (Equation 9).



B. Patterson projection onto the X (or Y or Z) axis (Equation 10).

FIGURE 5. Projections of the Patterson vector structure corresponding to the hypothetical structure of Figure 1A.

summation for a single unit cell. Similar summations for neighboring unit cells are also shown (with vertical displacement) in *d*. Adding these together gives *e*, representing the complete Patterson structure corresponding to *a*.

Since atoms are actually distributed over finite volumes, with continuously varying (time-average) electron densities, the straight addition of vectors just described is replaced by an integration and each vector has a weight equal to the product of the electron densities at its ends.

Two- and three-dimensional Patterson vector structures are related to the two- and three-dimensional electron structures in the same way as in the one-dimensional case, except that the vectors are no longer all oriented along a line. Comparison of Figures 3A and 3B with Figures 1A and 1B should make this relationship clear. An example of a Patterson-Harker section, obtained by means of Equation 7, is shown in Figure 9D.

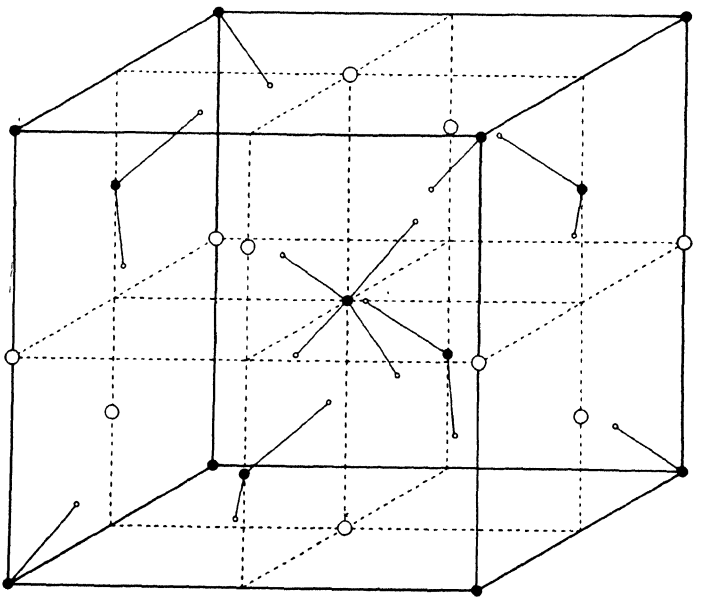


FIGURE 6A. Potassium dihydrogen phosphate (KH_2PO_4). Distribution of atomic centers in the unit cell. Large open circles represent potassium; filled circles, phosphorus; small open circles, oxygen.

An instructive comparison is shown in Figures 6A, 6B and 6C. 6A depicts the unit cell of the crystal structure of potassium dihydrogen phosphate, KH_2PO_4, deduced approximately by Hendricks,[13] using trial-and-error methods and, more accurately, by West [24] using Fourier summations. Figure 6B shows a contour map, obtained by West, representing the electron density projection onto the XY plane. This was made from the experimental F values, using Equation 4, the signs being determined for the most important reflections from the knowledge of the approximate atomic positions gained from the trial-and-error analysis and for the less important reflections by successive approximations. Figure 6C shows the corresponding Patterson projection.[20] Although this does not show what the structure would look like if we could see it, it does show certain facts about the relative positions of the atoms which (if the structure were not already known) would be of great help in working out further details—in checking hypotheses regarding the structure, etc.

summation is desired, or to the square of the structure factor, if a Patterson summation is to be made. The perforated plate is then bathed in a beam of monochromatic parallel light (from a point source at the focus of a lens in front of the plate) and the light which passes through the holes is focused by a second lens onto a photographic film. The interaction of the rays passing through the holes in the plate produces a photograph which reproduces, within certain limitations, the electron density or vector density distribution in the structure.

For electron density summations this method can be used only if the structure factors all have the same phase (*i.e.,* if the projection has a center of symmetry and the signs of the structure factors are all positive). Buerger [8] has proposed a scheme for making this method usable for any centrosymmetrical projection. The perforated plate is placed between crossed Polaroid films, and each hole in the plate is covered with a small piece of mica or strip of Polaroid film. These strips are oriented with their vibration directions parallel to a certain direction for positive structure factors and perpendicular to this direction for negative structure factors. Although theoretically possible, the practical difficulties are great—especially the necessity for keeping the optical thickness of each of the Polaroid sheets and of the whole set of Polaroid strips or sheets of mica uniform to within a small fraction of the wavelength of the light throughout the whole cross-section of the beam. It will be a tribute to Professor Buerger's ingenuity and technical skill if he is able to overcome these difficulties.

Buerger has also proposed that one use, instead of a brass plate in which tiny holes are bored, a print (on transparent film) from an x-ray diffraction photograph made by the de Jong and Bouman method. On such a photograph the spots are distributed as are the reciprocal lattice points, the blackening of each spot being produced by an x-ray intensity proportional to $|F_{hkl}|^2$ times certain other factors which he proposes to correct for by means of suitable rotating sectors. Then, by careful control, he should be able to obtain a print which will let light through a transparent spot at each lattice point, in an amount which is proportional to the value of F^2 for that point. This procedure should eliminate the necessity of measuring a large number of individual intensities before proceeding with a Patterson projection.

The fourth new type of method of more-automatic synthesis of Fourier projections and sections is also due to W. L. Bragg.[5, 6] Certain modifications of Bragg's experimental technique have been proposed by Hettich [14] and by Huggins.[15] According to Huggins' scheme, two masks are made, once for all, for each pair of indices (*e.g., hk*). One mask is used when the structure factor is positive, the other when it is negative. Each mask is crossed by light and dark bands (Figure 8), in such a way that the amount of light energy transmitted in a given time is proportional to $1 + \cos 2\pi(hx + ky)$ for a positive structure factor, or $1 - \cos 2\pi(hx + ky)$ for a negative structure factor. In making an *electron density* projection, successive exposures are made through those masks corresponding (in *h, k and sign*) to the structure factor data being used, the times of exposure being proportional to the magnitudes of the structure factors. If a *Patterson* projection is desired, only "positive" masks are used and the exposure times are proportional to the structure factor *squares.* For maps of the electron density or vector density distribution within a given plane in the structure, the exposure times and "signs" of the masks are determined by Equation 2c or Equation 7c or another similar equation (for a differently oriented plane).

The total intensity striking the film or paper on which the synthesis is being made is proportional (for an electron density projection) to

$$\sum |F_{hk0}| + \sum F_{hk0} \cos 2\pi (hx + ky),$$

summation is desired, or to the square of the structure factor, if a Patterson summation is to be made. The perforated plate is then bathed in a beam of monochromatic parallel light (from a point source at the focus of a lens in front of the plate) and the light which passes through the holes is focused by a second lens onto a photographic film. The interaction of the rays passing through the holes in the plate produces a photograph which reproduces, within certain limitations, the electron density or vector density distribution in the structure.

For electron density summations this method can be used only if the structure factors all have the same phase (*i.e.,* if the projection has a center of symmetry and the signs of the structure factors are all positive). Buerger [8] has proposed a scheme for making this method usable for any centrosymmetrical projection. The perforated plate is placed between crossed Polaroid films, and each hole in the plate is covered with a small piece of mica or strip of Polaroid film. These strips are oriented with their vibration directions parallel to a certain direction for positive structure factors and perpendicular to this direction for negative structure factors. Although theoretically possible, the practical difficulties are great—especially the necessity for keeping the optical thickness of each of the Polaroid sheets and of the whole set of Polaroid strips or sheets of mica uniform to within a small fraction of the wavelength of the light throughout the whole cross-section of the beam. It will be a tribute to Professor Buerger's ingenuity and technical skill if he is able to overcome these difficulties.

Buerger has also proposed that one use, instead of a brass plate in which tiny holes are bored, a print (on transparent film) from an x-ray diffraction photograph made by the de Jong and Bouman method. On such a photograph the spots are distributed as are the reciprocal lattice points, the blackening of each spot being produced by an x-ray intensity proportional to $|F_{hkl}|^2$ times certain other factors which he proposes to correct for by means of suitable rotating sectors. Then, by careful control, he should be able to obtain a print which will let light through a transparent spot at each lattice point, in an amount which is proportional to the value of F^2 for that point. This procedure should eliminate the necessity of measuring a large number of individual intensities before proceeding with a Patterson projection.

The fourth new type of method of more-automatic synthesis of Fourier projections and sections is also due to W. L. Bragg.[5, 6] Certain modifications of Bragg's experimental technique have been proposed by Hettich [14] and by Huggins.[15] According to Huggins' scheme, two masks are made, once for all, for each pair of indices (*e.g., hk*). One mask is used when the structure factor is positive, the other when it is negative. Each mask is crossed by light and dark bands (Figure 8), in such a way that the amount of light energy transmitted in a given time is proportional to $1 + \cos 2\pi(hx + ky)$ for a positive structure factor, or $1 - \cos 2\pi(hx + ky)$ for a negative structure factor. In making an *electron density* projection, successive exposures are made through those masks corresponding (in *h, k and sign*) to the structure factor data being used, the times of exposure being proportional to the magnitudes of the structure factors. If a *Patterson* projection is desired, only "positive" masks are used and the exposure times are proportional to the structure factor *squares.* For maps of the electron density or vector density distribution within a given plane in the structure, the exposure times and "signs" of the masks are determined by Equation 2c or Equation 7c or another similar equation (for a differently oriented plane).

The total intensity striking the film or paper on which the synthesis is being made is proportional (for an electron density projection) to

$$\sum |F_{hk0}| + \sum F_{hk0} \cos 2\pi (hx + ky),$$

the second summation being that desired and the first an unwanted background. A suitable choice of photographic film or paper enables one to eliminate this background almost completely.

A photograph, obtained in this way with the aid of a preliminary set of masks, is shown in Figure 9b. It represents the electron density in a portion of the unit cell

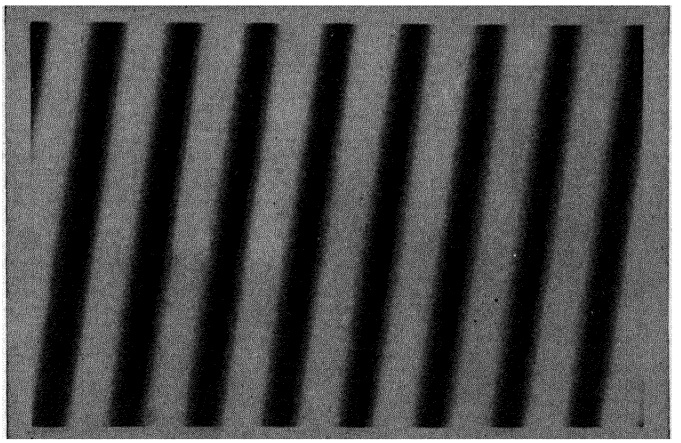


FIGURE 8. Mask [15] for optical Fourier synthesis of two-dimensional projections and sections (Equations 2, 4, 7, 9).

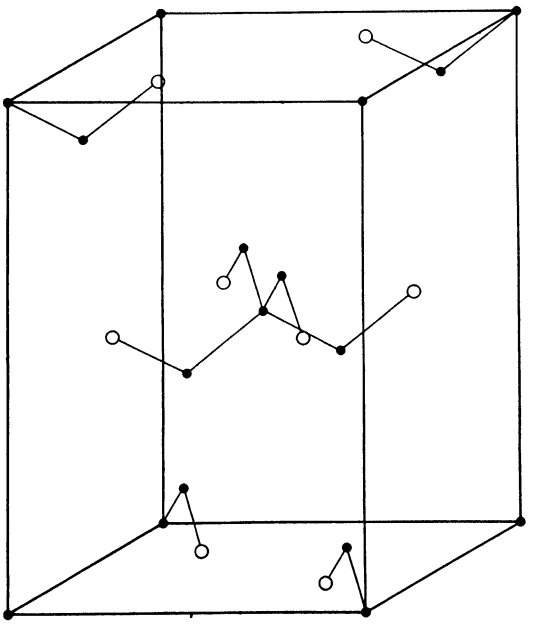


FIGURE 9A. Pentaerythritol, $C(CH_2OH)_4$. Distribution of atomic centers in the unit cell. Open circles denote oxygen; filled circles, carbon. Hydrogens are not shown.

the second summation being that desired and the first an unwanted background. A suitable choice of photographic film or paper enables one to eliminate this background almost completely.

A photograph, obtained in this way with the aid of a preliminary set of masks, is shown in Figure 9b. It represents the electron density in a portion of the unit cell

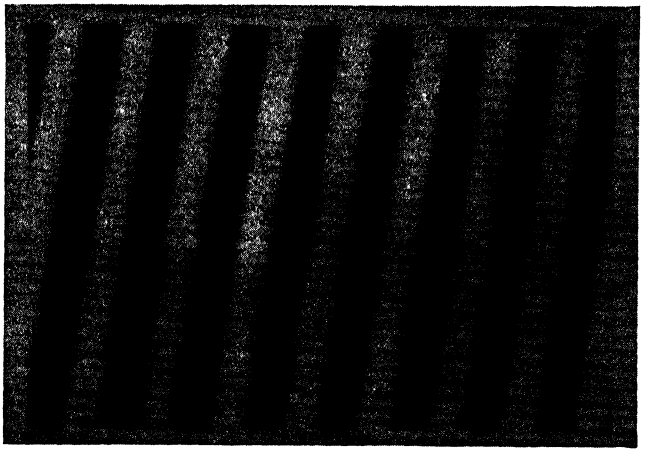


FIGURE 8. Mask [15] for optical Fourier synthesis of two-dimensional projections and sections (Equations 2, 4, 7, 9).

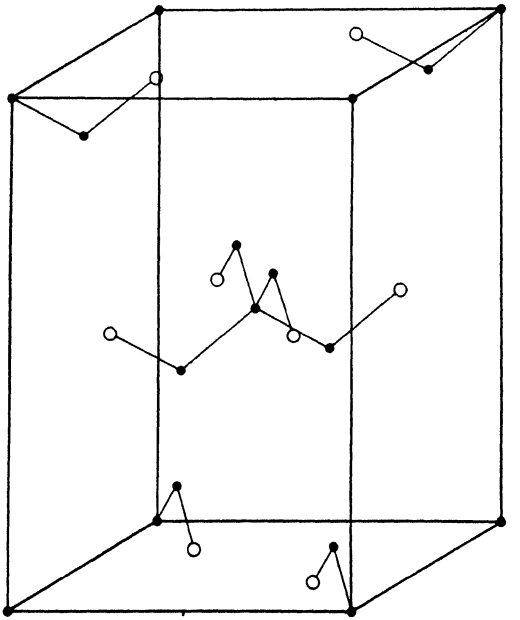


FIGURE 9A. Pentaerythritol, $C(CH_2OH)_4$. Distribution of atomic centers in the unit cell. Open circles denote oxygen; filled circles, carbon. Hydrogens are not shown.

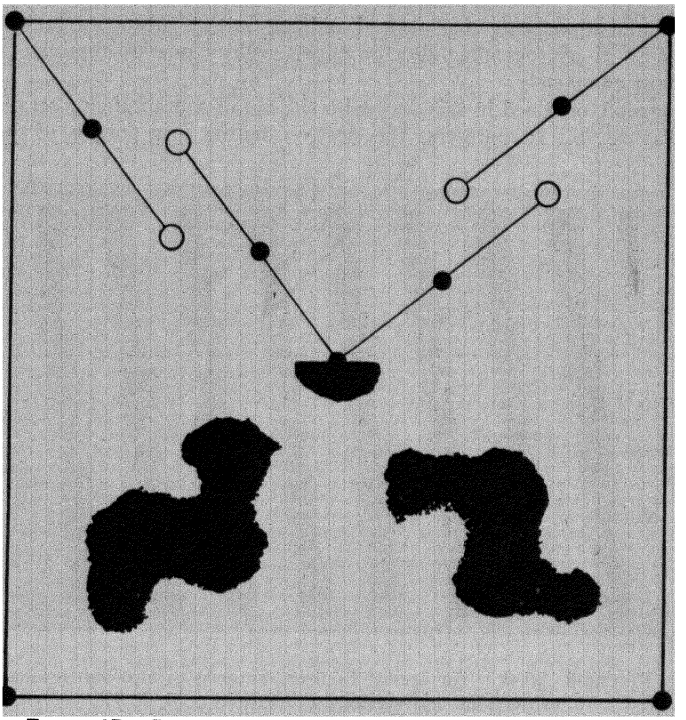


FIGURE 9B. Pentaerythritol. Projection onto the XY plane. The lower half of the figure contains a photographic representation of the electron density distribution, obtained by optical Fourier synthesis; the upper half shows only the relative positions of the atomic centers.

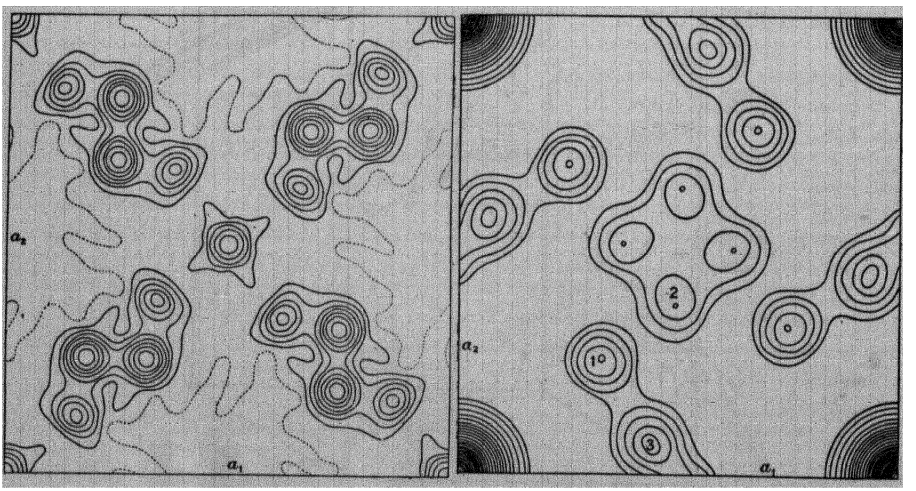


FIGURE 9C. Pentaerythritol. Contour map[17] representing the electron density distribution, projected onto the XY plane (Equation 4).

FIGURE 9D. Pentaerythritol. Contour map[17] representing the Patterson vector structure in the XY plane.

of pentaerythritol, $C(CH_2OH)_4$, projected onto the XY plane. The F_{hk0} values published by Llewellyn, Cox and Goodwin [17] were used. In the same figure are shown the positions of the atomic centers in the remainder of the unit of the projection. Figure 9A gives the atomic distribution within the whole unit cell. Figures 9C and 9D are reproductions of contour maps, published by the above-mentioned authors, representing the electron density distribution projected onto the XY plane (Equation 4) and the Patterson vector structure in this same plane (Equation 7), both obtained by straight numerical summation.

The method of optical summation is very simple and rapid (most summations requiring less than an hour), once the set of masks has been prepared. It shows promise of being especially useful during the preliminary stages of a structure analysis. Although the results are of course not as accurate as those obtained by direct summation methods, with or without punched-card machines, they are probably sufficiently accurate for most purposes.

Summary and Conclusion

The customary procedure for the analysis of complex structures with the aid of x-rays has been outlined, with special emphasis on the use of Fourier series to obtain the atomic distribution more or less automatically. The very real difficulties of obtaining sufficient accurate x-ray data of the right sort for the purpose and of determining, either by trial-and-error methods or from the results of vector summations, the phases of the structure factors—the signs of the coefficients of the terms in the series giving the electron density distribution—have been slurred over. These are difficulties which must be surmounted in ways which differ greatly from structure to structure and so cannot be adequately treated in the space available here.

It should be emphasized that, even with the newer methods, analyses of complex structures take considerable time and careful, intelligent labor. Partial analyses, giving information regarding the sizes, shapes, symmetries and relative orientations of the component molecules, can be obtained relatively easily; and often such partial analyses are sufficient, in combination with information regarding the structure from other lines of investigation, to solve the problem at hand.

References

1. Astbury, W. T., and Yardley, K., *Trans. Roy. Soc. London*, **A224**, 221 (1924).
2. Beevers, C. A., *Proc. Phys. Soc. London*, **51**, 660 (1939).
3. Beevers, C. A., and Lipson, H., *Phil. Mag.*, **17**, 855 (1934); *Nature*, **137**, 825 (1936); Lipson, H., and Beevers, C. A., *Proc. Phys. Soc. London*, **48**, 772 (1936).
4. Bragg, W. L., *Proc. Roy. Soc. London*, **A123**, 537 (1929).
5. Bragg, W. L., *Z. Krist.*, **A70**, 475 (1929).
6. Bragg, W. L., "The Crystalline State," New York, Macmillan, 1934.
7. Bragg, W. L., *Nature*, **143**, 678 (1939); **149**, 470 (1942).
8. Buerger, M. J., *Proc. Nat. Acad. Sci.*, **25**, 383 (1939); **27**, 117 (1941); "X-Ray Crystallography" New York, J. Wiley & Sons, 1942.
9. Compton, A. H., "X-Rays and Electrons," New York, Van Nostrand, 1926.
10. Duane, W., *Proc. Nat. Acad. Sci.*, **11**, 489 (1925).
11. Harker, D., *J. Chem. Phys.*, **4**, 381 (1936).
12. Havighurst, R. J., *Proc. Nat. Acad. Sci.*, **11**, 502 (1925); *Phys. Rev.*, **29**, 1 (1927).
13. Hendricks, S. B., *Am. J. Sci.*, **14**, 269 (1927).
14. Hettich, A., *Z. Krist.*, **A90**, 473 (1935).
15. Huggins, M. L., *J. Am. Chem. Soc.*, **63**, 66 (1941).
16. "Internationale Tabellen zur Bestimmung von Kristallstrukturen," Berlin, Gebrüder Borntraeger, 1935.
17. Llewellyn, F. J., Cox, E. G., and Goodwin, T. H., *J. Chem. Soc.*, **(1937)** 883.
18. Macewan, D., and Beevers, C. A., *J. Sci. Inst.*, **19**, 150 (1942).
19. Patterson, A. L., *Phil. Mag.*, **22**, 753 (1936).
20. Patterson, A. L., *Phys. Rev.*, **46**, 372 (1934); *Z. Krist.*, **A90**, 517 (1935).
21. Robertson, J. M., *Phil. Mag.*, **21**, 176 (1936).
22. Robertson, J. M., *Reports on Progress in Physics*, **4**, 332 (1938).
23. Shaffer, P. A., Jr., and Pauling, L., Unpublished work at the California Institute of Technology.
24. West, J., *Z. Krist.*, **A74**, 306 (1930).

Microradiography of Colloidal Materials

GEORGE L. CLARK

University of Illinois, Urbana, Ill.

Any method which involves the examination and delineation of small particles, especially where magnifications are required, is properly a technique of colloid chemistry. One of these, namely, the extension to very small objects of the branch of x-ray science known as radiography, so familiar in medical diagnosis and in industrial testing of gross soundness or homogeneity of materials, has been recognized since shortly after Roentgen's discovery of x-rays. But all attempts to develop a practical and dependable method of examining very small specimens of metals, colloidal particles or biological objects for gross structure (as contrasted with determination of fine structure by x-ray diffraction) were largely unsuccessful until very recently. Several workers saw how desirable it would be to use x-rays in the preparation of photographs which would delineate, by simple differential absorption of the rays, the various phases of different composition or regions of varying density in the same material. Such a technique would correspond to photomicrography with visible or ultraviolet light, except that x-rays penetrate through a finite thickness of a specimen, whereas the photomicrograph registers only a surface of an opaque body, such as a polished and etched metal, or at best only a microtomed section of a biological specimen.

The progress in the development of ordinary diagnostic radiography exceeded the hopes of the earlier workers. Numerous developments appeared to increase the contrast (usually by impregnation) and the clarity, or detail, of the images obtained. The logical tendency of such trends inevitably led to examination of more and more minute structures, and finally to actual enlargement of the images obtained on the photographic emulsion. Since, of course, there were no lenses for enlarging the x-ray image itself before registration, such methods of enlarging the radiograph for examination have become known as "microradiography," so named because of their relationship to the parent technical field of radiography.

Technique

Although a microradiographic method was employed by Heycock and Neville in 1898 [1] and by Goby in 1913,[2] the first work which might be considered to have biological interest was that by Dauvillier [3] in 1930, who examined radiographs of vegetable cells (elder pith, etc.). The biological use of the method was extended in 1936 by Lamarque and associates [4] in a study of plant and animal cells. All these researches, as well as those in non-biological fields, were carried out with low-voltage x-radiation using vacuum systems; for example, Lamarque used for his investigations 5000 volts on the x-ray tube with currents up to 100 milliamperes. These methods, employing soft general radiation, vacuum equipment, long exposure times, and photographic emulsions with silver halide particles too large to permit magnification of the image without loss of sharpness, were obviously subject to experimental limitations which restricted the possibilities of general use of the method. The need of a more usable method led the author and associates [5] to develop, in 1938, a technique employing ordinary x-ray diffraction equipment, and eliminating the

necessity for a vacuum. Another essential feature was the availability of photographic film with extremely fine-grain emulsions (Gevaert Lippmann emulsion) with which enlargements up to × 300 could be made without loss of detail. Characteristic radiations of suitable wave-length for most work are obtained in comparatively great intensity by working at 30,000 to 50,000 volts, using ordinary x-ray diffraction tubes with targets of molybdenum, copper, iron, cobalt, chromium, etc. The intensity of such radiations makes the use of vacuum equipment unnecessary, and permits short exposures. The radiation itself is essentially monochromatic—a factor of the utmost importance in interpretation. These methods are restricted to available target materials; it is of great advantage to have still longer wave-lengths available, but for many types of investigation this technique is certainly well adapted in its present form.

Generally, two possible methods are available for the microradiographic examination of biological materials: (a) using a technique of "absorption staining," that is, impregnating particular parts of a biological specimen with a material having higher absorption for the x-radiation employed than the normal tissue itself (like the use of barium sulfate in diagnostic radiographs of stomach and intestines); and (b) making use of the density variations in the specimen to obtain the necessary differential absorption.

To these two usual procedures may be added a third technique for highly heterogeneous materials, such as has been successfully applied by Clark and Gross [5] to complex alloys in which several phases (each of different composition and absorbing power) may be separately delineated. In this case microradiographs are made of the same specimen with two or more effective x-ray wave-lengths, selected so that there will be a maximum difference in the linear absorption coefficients of various pairs of phases. For example, for a bronze containing copper, lead and tin, radiation from a molybdenum target (0.71 Å) will produce light areas on the print for lead, dark for tin and copper; copper radiation (1.54 Å) for the same specimen produces both lead and tin light and copper dark. Solid solutions of tin or lead appear in various shades of gray.

In any case, the sample is placed in contact with the emulsion of a very fine-grain film (Lipmann emulsion has been found to be most satisfactory, and is available either in film form or on glass plates), and irradiated with characteristic x-radiation of the proper wave-length. The film and sample may be simply wrapped in black paper or various types of cassettes may be used. On the diffraction unit a disc of film about 0.5 inch in diameter may be placed at the end of a closed tube. Into this is inserted a closely fitting smaller tube, covered at one end with black paper to which the specimen is attached, until the specimen is against the bare film. The x-rays then pass through the tube and specimen and register the small image on the film. Subsequently the film is developed, and an enlargement made with a microscope and camera at the magnification desired. The limitation in enlargement is fixed by the grain size of the emulsion used for the radiograph. Ordinarily with Lipmann film the grain becomes more or less objectionable when 300 diameters is exceeded. It is very important that the object examined be in close contact with the emulsion of the film, since the detail recorded is very sensitive to sample-film distance. The x-ray beam employed need not be collimated and the exposures require but a few seconds when the commercial diffraction equipment is employed as radiation source.

Theory

The ordinary absorption laws hold for microradiography and may be expressed as

$$I = I_0 e^{\frac{-\mu}{\rho} \cdot \rho x}$$

where μ/ρ is the mass absorption coefficient as listed in the published tables in handbooks, x is the thickness of the sample in centimeters, and ρ is the density of the particular sample used. As may be observed from this expression, the density value ρ and the absorption coefficient fix the absorbing power of the material. The ratio of intensities for x-ray beams passing through two adjacent sections of a given sample (of the same thickness) may be written

$$I_a/I_b = e^{-\left(\frac{\mu_1}{\rho_1}\rho_1 - \frac{\mu_2}{\rho_2}\rho_2\right)x}$$

This expression is approximated below for the two most simple cases: (a) that in which density differentiation is important, the case which may be expected to be most common with biological materials where the absorption coefficient is often essentially identical, and (b) the case in which a variation in chemical composition appears, as for example, in metal alloys, bone or specimens in which some impregnating agent used for "absorption staining":

$$\text{(a)} \quad (I_b - I_a)/I_b = (\mu/\rho)x(d_b - d_a)$$

$$\text{(b)} \quad (I_b - I_a)/I_b = x(\mu_b - \mu_a)$$

The μ values given in formula (b) are linear absorption coefficients (obtained by multiplying the mass absorption coefficients by the density of the material indicated).

The first formula (a) requires that the sensitivity of the density differentiation on the radiograph be proportional to the differences in density; the second equation (b) that the sensitivity be proportional to the differences in the linear absorption coefficient.

Examples

That certain biological materials would be excellent subjects for microradiographic study is immediately evident. Cellulose, chitin and calcareous structures obviously are suitable because of extreme density variations in the samples or actual variations in chemical content which would correspond to low and high absorption of x-rays.

Figure 1 is a typical microradiograph of a very small insect, in this case a sweat-bee. The spots are identified as pollen caught on very fine hairs. Eyes, antennae, wings, digestional tracts and other anatomical details are readily revealed at magnifications usually of × 100. Microradiographs of wood sections, transverse, longitudinal tangential, and longitudinal radial, are very useful in wood technology. Here especially the addition of the third dimension, as the result of the penetration of x-rays through sections of considerable thickness, adds greatly to the information afforded by the optical microscope.

The high absorbing power of calcium for soft x-rays has been of great importance in ordinary radiography, and so it is to be expected that micrography would be quite successful with bone structures. Such has been the case, especially with extremely small specimens. Replacement of the calcium by other elements, such as lead in lead poisoning, is immediately detected; and as the result of disease variable darkening on the photograph serves to show variable density of the specimen under examination. The suitability of the bone material as subject matter for microradiographic investigations, and the importance of orientation as a function of previous history of the bone suggest that this method may be promising for future investigations in this field.

Other types of biological materials are less suitable for microradiography. Many tissues show good shadows but absence of microscopic structural detail. Moisture and the thickness of the samples used may combine to destroy any detail, since the photograph represents the superposition of effects from a finite thickness. With

such materials it is necessary to use sections thinner than the usual technique would permit for any hope of successful resolution, and longer wave-lengths would certainly be advantageous in the technique. Specimens of the muscle, brain, liver and nerve tissues of a frog and other microtome sections cast acceptable radiographic

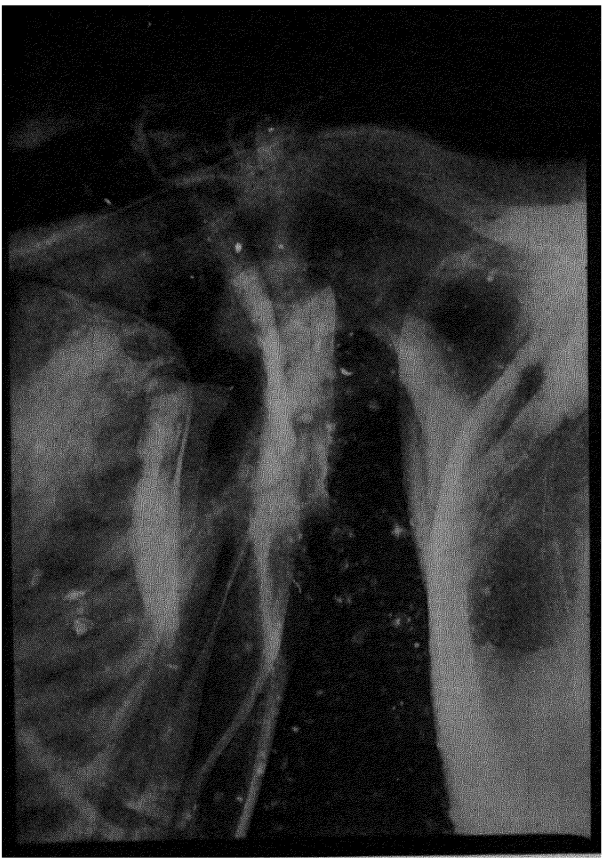


FIGURE 1. Microradiograph of body of sweat-bee (× 100), showing density variations of chitin and pollen particles caught on minute hairs.

shadows, but microscopic detail is difficult to observe when monochromatic rays shorter than 1.9 Å are used from the diffraction tubes usually available.

The method of "absorption staining" or impregnation extends applications of microradiography to many cases where differentiation is difficult. Radiographs were taken of embryonic beans soaked for different lengths of time in lead acetate solution. The progress of the solution into and through the beans was clearly indicated, and stereographic views were taken which enabled the absorption to be viewed in three dimensions. A more suitable demonstration of the possibilities of absorption staining is illustrated by the kidney of a thorotrast-impregnated frog. This microradiograph of a longitudinal section of the kidney shows as the light areas the glomeruli, and indicates that the thorium dioxide could not continue into the lumen of the kidney in the process of excretion, and accumulated at these points.

These are but a few examples of the possibilities of a simple extension of radiography, which though recognized and tried with little success for forty years, has just reached the stage of practical usefulness in a wide variety of biological and industrial fields.

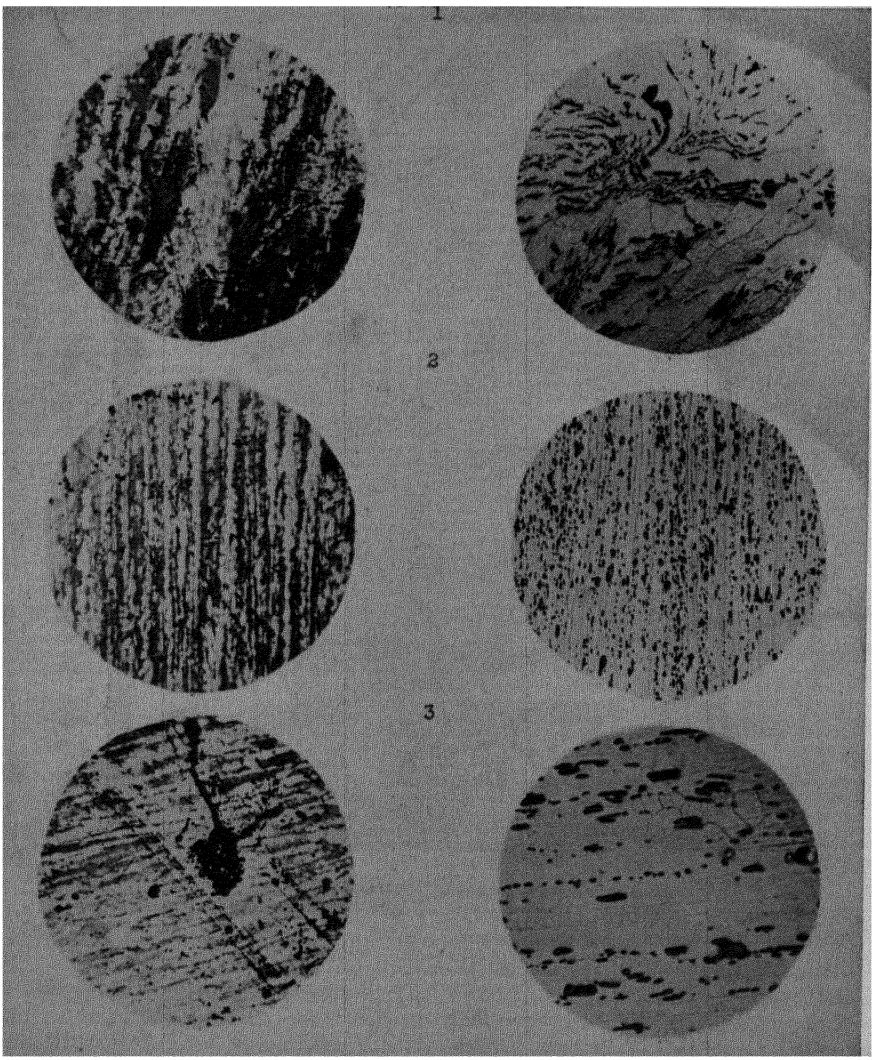


FIGURE 2. Comparison of photomicrographs (right) and microradiographs (left) of some specimens of aluminum-copper alloys used in airplane parts. (Aluminum dark, copper and alloy phases light.)

Thus far the most important contributions of the science of microradiography have been in the field of metals and alloys. Figure 2 presents comparative photomicrographs and microradiographs of the same areas of three samples of aluminum alloys used in aircraft construction. For the latter the light areas are for the constituents

heavier than aluminum. The control of production of superior airplane motor castings is now being accomplished with microradiographs. Papers by the writer and associates [5] give many examples of metallurgical applications.

As a more typical colloidal material, Figure 3 shows what may be accomplished with a study of the process of crystallization of inorganic salts, in this case cesium bismuth iodide. The microradiograph shows, at 300 diameters magnification, the very first stages of crystal organization on a skeleton framework, which subsequently fills in. It is evident that here is a valuable addition to the working tools of the colloid chemist, whatever may be his specific interests.

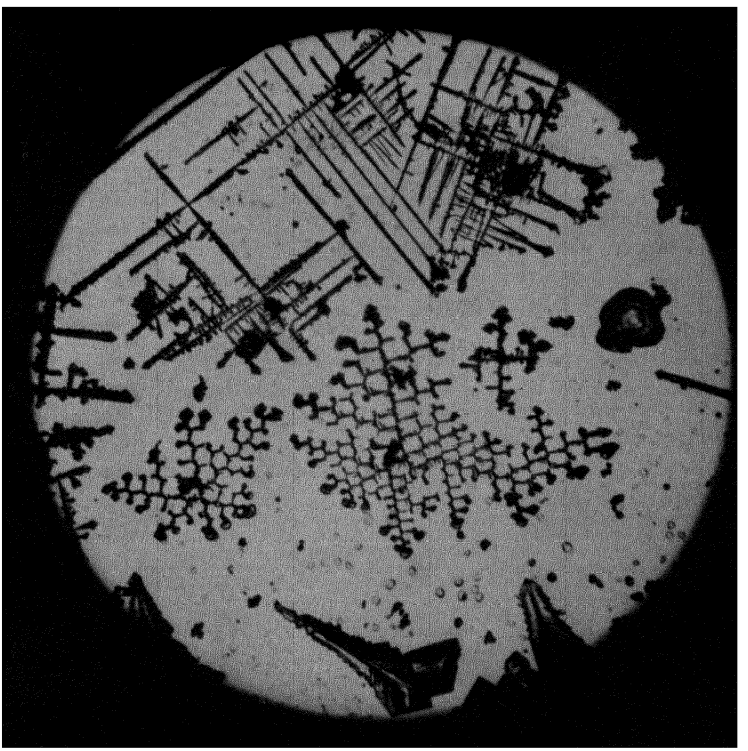


FIGURE 3. Microradiograph (× 300) of first stage in formation of crystals of cesium bismuth iodide.

References

1. Heycock and Neville, *Trans. Chem. Soc.*, **73**, 714 (1898); *Phil. Trans. Roy. Soc.*, **194a**, 201 (1900).
2. Goby, P., *Compt. rend.*, **156**, 686 (1913).
3. Dauvillier, A., *Compt. rend.*, **191**, 1287 (1930).
4. Lamarque, P., *Radiology*, **27**, 563 (1936).
5. Clark, G. L., *Photo Technique*, **1**, No. 12 (1939); Clark, G. L., and Shafer, W., *Trans. Am. Soc. Metals*, **13**, 732 (1941); Clark, G. L., and Gross, S. T., *Ind. Eng. Chem., Anal. Ed.*, **14**, 676 (1942); Gross, S. T., and Clark, G. L., *Industrial Radiography*, 1, No. 2, 21 (1942).

The Electron Microscope

ALBERT F. PREBUS

Ohio State University, Columbus, Ohio

I. INTRODUCTION

In the endeavor to correlate the observable properties of the substances encountered in all branches of the physical and biological sciences, structural units are sought which are common to a large number of similar materials and the characteristics of which determine with sufficient accuracy the macroscopic properties of the complex substances. The analysis and associated characterization of the significant constituents may be performed with techniques of varying degrees of directness and precision, this depending in a large measure upon the dimensional range in which the structural units are included.

A highly detailed resolution, in the atomic range of dimensions, is provided by a quantitative chemical analysis. This reveals the identity and relative proportions of the chemical elements. Together with certain spectroscopic, x-ray or electron diffraction studies, the existence of recurrent molecular or ionic groups may also be revealed. Interpretation of the data obtained by these means, however, is rapidly beset with difficulties in dealing with structural units which exceed a few millimicrons in size. The existence of such structures may be overlooked entirely if they occur in small numbers or in a random state of aggregation.

In many phases of microbiology, analyses in the atomic range of dimensions provide but meager information. The usefulness of such analyses to research workers engaged in the study of alloys, minerals, textiles, and other high polymers is equally limited. More significant results are obtained by studying the larger molecular aggregates, cells and organisms which can be observed directly with the aid of a microscope. There is, however, a lower limit in the dimensional range amenable to this method of study. It is fixed by the nature of light to between 100 and 200 millimicrons.

Between the upper dimensional limit of atomic analyses and the limit of resolution of the light microscope there exist highly important structural units in both the physical and biological sciences. These include the colloidal particles which bear such labels as crystallites, micelles, fibrils, viruses, and macro-molecules. Reactions in which these units maintain their identity are highly significant. A comprehensive understanding of the physical and chemical properties of these aggregates of submicroscopic dimensions, of their modes of generation and interaction, and of the mechanisms by which they are bound together to form larger structures, is of the utmost importance in the interpretation of the properties of many substances in the macroscopic form.

The study of matter in the colloidal state of aggregation has been pursued, until very recently, only with the aid of instruments and techniques which provide information of a *statistical character*. The fundamental requirement for the establishment of conclusive data from sedimentation, diffusion, and viscosity experiments and from

observations made with ultra- and polarization microscopes, is the existence in a specimen of large numbers of similar structures. The primary objective of nearly all these experiments is the determination of the physical dimensions of colloidal structures. From these data it is then possible to estimate the magnitude of specific surface properties, which determine in large measure the course of reactions observable on a macroscopic scale.

The electron microscope has been developed with the object of providing a means for the *direct observation of individual structural units* in this dimensional range. Electron micrographs of the colloidal particles occurring in the simpler suspensions or dispersions permit the shape of the particles to be inferred by inspection. Dimensions may be determined with a high degree of accuracy by direct linear measurements in the micrographs. With a practical resolving power below five millimicrons, the presence of isolated particles in the molecular range of dimensions may be detected. The results of shape and size determinations of the simpler colloids with the electron microscope have a more general significance, for they furnish data which permit surface and structure properties to be distinguished. They, therefore, provide an independent means of checking the formulations underlying the indirect methods, which must still be used in the study of many substances which cannot now be investigated successfully with the electron microscope for lack of suitable or reliable techniques of specimen preparation.

Since it is also possible to use the electron microscope within the working domain of the light microscope, an independent means is now available for scrutinizing the conclusions drawn from light microscope observations. This is particularly valuable in the range circumscribing the limit of resolution of the light microscope.

This review is devoted essentially to one of several types of electron optical instruments which are capable of forming magnified images. In the magnetic and electrostatic transmission microscopes, images are produced by means of high-speed electrons (accelerating potentials from 10,000 to 300,000 volts) which have been transmitted through specimens spread in the form of thin films or layers. The mechanism of image formation is in some respects analogous to the manner in which an image of a specimen is formed in the light microscope with either bright-field or dark-field illumination. This similarity, together with the system of lenses, *viz.*, the series *condenser, objective,* and *projector,* which these instruments have in common with the light microscope, has resulted in the designation of these instruments by the generic term *electron microscope.* Highest resolving power has been attained with magnetic objective lenses. Further development of the electrostatic objective lens undoubtedly will be pursued. A microscope employing the latter type of lens system is inherently more stable and less costly in construction.

Less than ten years have elapsed since the idea of developing an instrument which would surpass the resolving limit of the light microscope was recognized as a practical possibility. The resolving power of present-day instruments has been attained without the necessity of designing corrected lenses. Because of this good fortune, the development has been relatively easy, consisting in the elimination or avoidance of factors which prohibited the attainment of the highest resolving power of which the simple uncorrected lenses were capable. As a consequence, investigations of an intensive character are few. A thorough and comprehensive evaluation of the characteristics and limitations of the instrument cannot be carried out. Nevertheless, an extensive review of the fundamental concepts of the instrument, rather than of instrument accomplishments, was considered to be more valuable. Current reviews of this character do not exist in the English language. Several books, written in Germany, are difficult to obtain at the present time.

It has frequently been stated by electron-microscope workers that the interpretation of the electron microscope image is a matter of great simplicity in comparison with that of the light-microscope image. Such a comparison is not justified without

qualifying statements to specify the observable image data which is selected for consideration. Further, the relative ease of making conclusive observations from microscope images depends upon the dimensions of the observed structures relative to the resolving limits of the respective instruments. The research worker using the microscope as a tool in a specific investigation is in both cases confronted with the problem of interpreting the nature of his material in terms of the characteristics of the specimen observed in the microscope. The writer has endeavored to indicate the complexities of the problems involved in relating the photographic density variations in a micrograph with the phenomena of electron scattering in the specimen material. The most instructive theoretical and experimental contributions of many investigators have been taken into consideration. They provide, even to the most enthusiastic advocate of the electron microscope, only data of a qualitative character. Only the most elementary data provided in the electron microscope image can be utilized with reasonable assurance of accurate interpretation. The criticisms and suggestions of the writer have been made with the object of pointing out some of the pitfalls which may be encountered and of indicating the directions in which further intensive experimental work is urgently required. It is the nature of scientists to exploit their tools to the limits of their usefulness. In the range about the limit of resolution of present-day electron microscopes, the interpretation of the images is intimately related to the aberrations of electron lenses. This provides an additional reason for devoting considerable space to electron lenses in this paper.

II. Electron Lenses

The science of electron optics is an outgrowth of the fundamental discovery in 1926 by Busch [52] that the magnetic field in the neighborhood of the axis of symmetry of a circular solenoid possesses the properties of a lens with respect to radiation consisting of electrons or ions. Shortly thereafter, the lens characteristics of certain axially symmetrical electrostatic fields was recognized by Davisson and Calbick [55] and by Brüche and Johannson.[49] The early rapid development of magnetic electron lenses into useful electron optical instruments is due almost entirely to Ruska and associates,[95, 96, 97] then in the Technische Hochschule, Berlin. Intensive study of the electrostatic electron lenses, more versatile in technical applications, was taken up initially by Brüche and associates in the AEG Laboratories. In this same period, 1931 to 1936, the main theoretical aspects of electron optics were clarified, notably by Glaser and Scherzer. Descriptions and numerous references to this phase of the development and applications of electron lenses are included in the books of Busch and Brüche,[53] Myers,[125] Klemperer,[94] and Zworykin and Morton.[163] Further discussion here is limited to the main features of those lenses which have found application in electron transmission microscopes.

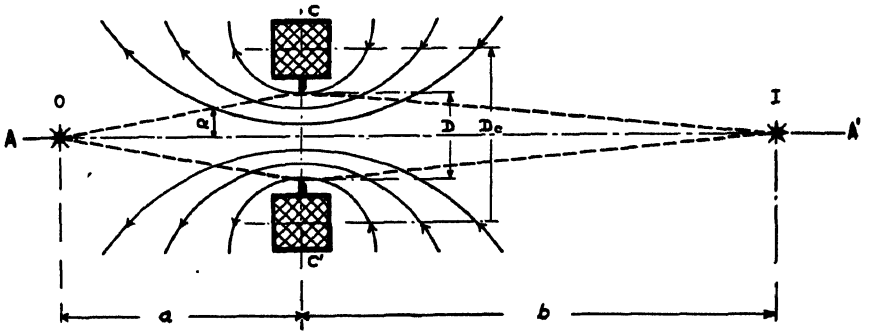


Figure 1a. Simple magnetic electron lens.

The Cardinal Points of Strong Magnetic Lenses. A simple magnetic electron lens is illustrated in Fig. 1 (a). CC′ represents the cross-section of a solenoid through the axis of symmetry AA′. The incompleted arcs illustrate the configuration of the lines of force of the magnetic field. If an object in a plane perpendicular to the axis AA′ and intersecting it in O, distant a from the center of the coil, emits or scatters electrons of uniform speed in a beam of small aperture α, an image is formed in the plane perpendicular to the axis and intersecting it in I, distant b from the coil center.

In the first approximation, the differential equations governing the trajectories of the electrons on their passage through the magnetic field may be reduced to the form [52]

$$\left.\begin{aligned} \frac{d^2r}{dz^2} + \frac{e}{8mc^2U}\cdot H^2(z)\cdot r = 0 \quad \text{.....(a)} \\ \frac{d\theta}{dz} = \sqrt{\frac{e}{8mc^2U}}\cdot H(z) \quad \text{.....(b)} \end{aligned}\right\} \tag{1}$$

where z is the distance along the axis AA′ with arbitrary origin; r is the radial distance of the electron from the symmetry axis; θ is the angular displacement of the trajectory in the cylindrical coördinate system r, θ, z; H (z) is the magnitude (emu) of the z-component of the magnetic field on the axis of symmetry; U is the accelerating potential (esu) of the electron beam; and e, (esu) m, c, have their customary significance.

Reduction of equation (1a) from the general Lorentz equation by Busch represents the discovery of the magnetic electron lens, for linearity of this equation in r implies that any axially symmetrical field free of current sources in the neighborhood of the axis may be regarded as a convergent symmetrical lens for electron beams of small aperture and homogeneous velocities. If the electron trajectory is considered to be confined in a plane containing the axis of symmetry, equation (1b) signifies that the plane of the trajectory rotates, the amount of rotation between any two axial points z_a and z_b being given by the integration of equation (1b) between these limits, *viz.*,

$$|\theta|_{z_a}^{z_b} = \sqrt{\frac{e}{8mc^2U}}\cdot\int_{z_a}^{z_b} H(z)dz \tag{2}$$

The magnetic lens does not in general provide an erect or inverted image. The image plane is rotated about the optic axis relative to the object plane through an angle determined by equation (2) with the limits of integration extended from the object to the image planes.

Under conditions in which the electron trajectories entering the field parallel to the optic axis are converged to a focal point at a large distance from the coil center, relative to the axial extent of the field, the lens is considered as thin and weak. Equation (1) may be integrated with sufficient accuracy by making the approximation that r remains sensibly constant within the field. In this case the focal length of the lens is determined by the relation

$$1/f = \frac{e}{8mc^2U}\int_{-\infty}^{+\infty} H^2(z)dz \tag{3}$$

This formula becomes increasingly inaccurate as the focal point approaches the region in which the field is appreciable and fails entirely to describe the optical properties of the lenses used in a transmission electron microscope.

The shortest focal length attainable with a simple coil is determined by the mean diameter D_c. For electrons with speeds in excess of 10,000 volts, required in a

transmission microscope, coils of large dimensions are required to converge the electrons, and focal lengths below several centimeters are not experimentally attainable. This limitation in the attainment of high-power lenses for fast electrons was removed by the introduction of the iron-encased solenoid by Knoll and Ruska in 1931.[95] By this means the magnetic field of a coil of arbitrary cross-section may be effectively concentrated into a small volume without destruction of image-forming properties. The diameter of the equivalent simple coil is determined by the internal radius of the terminals, or *pole pieces,* of the iron case between which the field is concentrated.[16, 134] Such a coil, in which the internal diameter of the pole pieces is denoted by D_p, is illustrated in Fig. 1 (b).

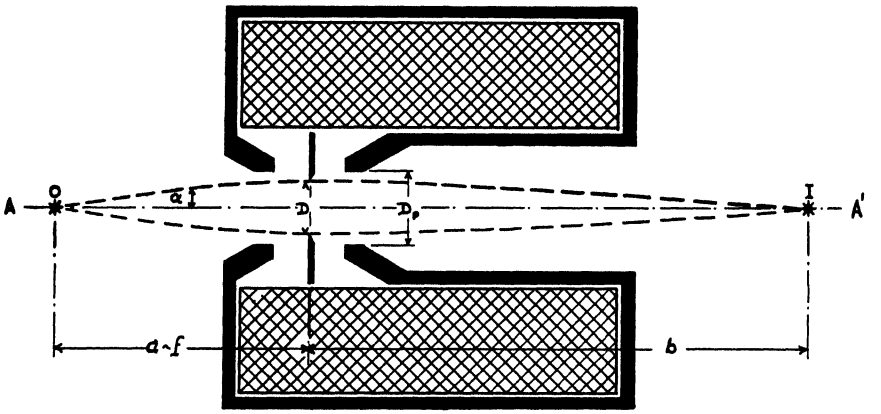


FIGURE 1b. Iron-encased short focal length magnetic electron lens. In the limiting case of thin and weak lenses, $1/a + 1/b = 1/f$ and $1/f = e\int_{-\infty}^{\infty} H^2(z)dz/8mc^2U$.

Empirical corrections have been applied[134] to the formula (3) to represent the focal lengths of the short focal length lenses obtained with the iron-encased solenoids. Even these, however, have been found inadequate to describe quantitatively the optical properties of strong lenses.

Although the formal optical theory of electron lenses has been developed to a satisfactory degree,[65, 144] general methods for the computation of the optical constants of strong magnetic lenses in terms of the geometrical and magnetic constants of the pole pieces have not yet been developed. Neither is it possible to determine the optical properties by exact analytical methods in terms of arbitrary theoretical or experimental magnetic fields. The construction of pole piece units to provide lenses with specified cardinal points and with certain aberrations reduced to minima, consequently has not yet been attempted.

The most comprehensive descriptions of the optical properties of the types of magnetic lenses used in transmission microscopes have been provided by Glaser[66] and Dosse.[56, 57] In the publication of Glaser, an analytic function is selected which represents with reasonable accuracy the field in a strong lens and which permits differential equation (1a) to be solved exactly. This procedure leads to a clear picture of the dependence of the cardinal points and of the important aberrations on the strength and width of the lenses.

With the function

$$H(z) = H_0/\left(1 + \frac{z^2}{a^2}\right) \qquad (z = 0 \text{ where } H(z) = H_0) \tag{4}$$

representing the magnetic field on the axis of symmetry, the "strength" of the lens is described by the parameter

$$k^2 = \frac{eH_0^2d^2}{8mc^2U} \tag{5}$$

and the "half-width" by the parameter d, equal to the axial distance from the maximum of the field to the point at which it has one-half of this value. In the case of the function (4), the half width d is equal to a.

By exact solution of equation (1), the focal points z_{f_0}, z_{f_1} of the lens can be expressed analytically by the simple relations

$$\left.\begin{aligned} z_{f_0} &= a\cot\frac{\pi}{\sqrt{1+k^2}} \quad \text{.....(a)} \\ z_{f_1} &= -a\cot\frac{\pi}{\sqrt{1+k^2}} \quad \text{....(b)} \end{aligned}\right\} \tag{6}$$

and the corresponding focal lengths f_0, f_1 by the relations

$$\frac{1}{f_0} = -\frac{1}{f_1} = \frac{1}{a}\sin\frac{\pi}{\sqrt{1+k^2}} \tag{7}$$

The positions of the principal points h_0, h_1 are determined by the relations

$$\left.\begin{aligned} z_{h_0} &= a\cot\frac{\pi}{2\sqrt{1+k^2}} \quad \text{.....(a)} \\ z_{h_1} &= -a\cot\frac{\pi}{2\sqrt{1+k^2}} \quad \text{.....(b)} \end{aligned}\right\} \tag{8}$$

From inspection of equation (7), it may be concluded that the shortest attainable focal length is equal to a, the half-width of the symmetrical field (4). This obtains when $k^2 = 3$. If k^2 is greater than 3, there are more than one pair of focal points in the lens, and multiple images occur. The impaired image definition resulting from multiple images has made it customary to avoid this range in electron microscope lenses. The following discussion is therefore restricted to the range of k^2 below the value 3.

If Z_o, Z_i, represent conjugated object and image distances measured from the focal points, *viz.*, $Z_o = z_o - z_{f_0}$, $Z_i = z_i - z_{f_1}$, they may be related by Newton's lens formula

$$Z_oZ_i = f_of_1 \tag{9}$$

The above analysis has demonstrated that the determination of conjugated object and image elements by the familiar Listing construction, based on a knowledge of only the cardinal points, is applicable in the case of strong magnetic lenses even with the object and image positions within the field. The positions of the cardinal points in a lens of this type, in terms of the parameter k^2, are illustrated in Fig. 2 (Fig. 6A [66]).

A comparison of these theoretical values with experimental measurements by Dosse [67] indicates that the function (4) is an inadequate representation of the magnetic fields in the range most frequently used in an objective lens. The measured fields are more sharply defined and become extended to fit equation (4) only when the saturation point in the material of the pole pieces is exceeded. In the useful range of field intensities, the fields have not been amenable to description with sufficient accuracy by any known function which makes equation (1) exactly soluble.

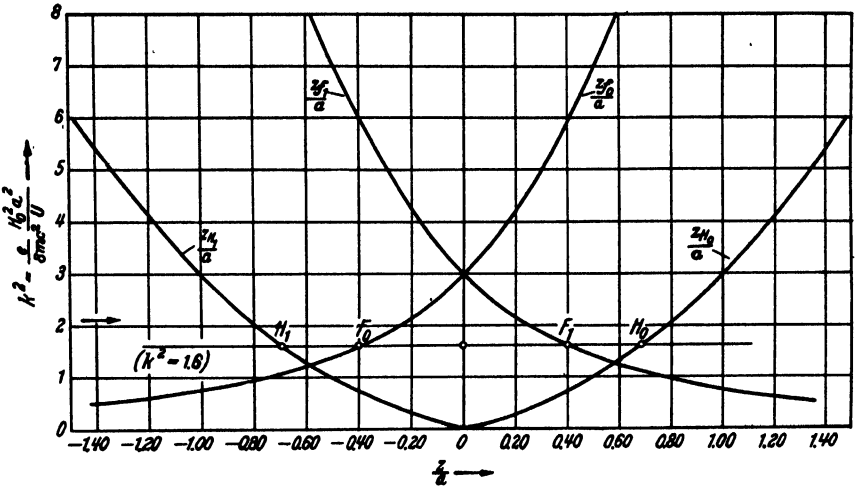


FIGURE 2. Principal points H_0, H_1 and focal points F_0, F_1 as functions of the parameter k^2 of strong magnetic lenses having the axial field distribution $H(z) = H_0/(1 + z^2/a^2)$. The focal lengths are given by the relations: $f_0 = z_{h_0} - z_{f_0}$; $f_1 = z_{h_1} - z_{f_1}$.[66]

Dosse,[57] therefore, carried out the calculation of the cardinal points for normal field distributions by numerical methods. In this investigation, the measured fields are described by the function

$$\xi_3 = \frac{H(z)}{H_0} = e^{-\frac{z^2}{a_3^2}}, \quad a_3 = 1.20d \tag{10}$$

the limiting case of the function

$$\frac{H(z)}{H_0} = \frac{1}{\left(1 + \frac{z^2}{a_3^2}\right)^{\mu}} \quad as\ \mu \to \infty \tag{11}$$

The accuracy with which the analytic fields (4) and (10) represent the measured fields in the two ranges (below and above saturation in the pole pieces) is illustrated in Fig. 3 (Fig. 1 [57]). In this chart, curve 1, representing equation (4), follows closely the measured field, curve 5, obtained beyond the saturation point. The limiting function (10), shown as curve 3, is a better approximation to the normal field, shown as curve 4.

In consequence of the variable permeability of the pole-piece material, it is evident that the shape of the $H(z)$ curve makes a gradual transition from the forms 4 to 5 as the current (ampere-turns) in the solenoid of the lens is increased. The comparison of the dependence of the focal length on k^2 in the two limits is therefore of interest. This is shown in Fig. 4 (Fig. 5 [57]), in which curve 1 demonstrates equation (7); curve 2 the numerically computed values on the basis of equation (10); and curve 3 the focal lengths in the ideal field $H(z) = H_0$ in the interval $-d$ to $+d$. For a given beam potential U and fixed pole-piece dimensions, the focal length decreases with increasing H_0 (determined by the current in the coil). In the range of small k^2, or large focal lengths, the variation may be expected to follow the trend of curve 2. As saturation is approached, the width of the field increases, and the focal length

decreases with increasing k^2 in a manner as yet undescribed until saturation is established. From this point onward, the focal length varies with k^2 in accordance with curve 1. The numerical range of k^2 in which this transition takes place depends upon

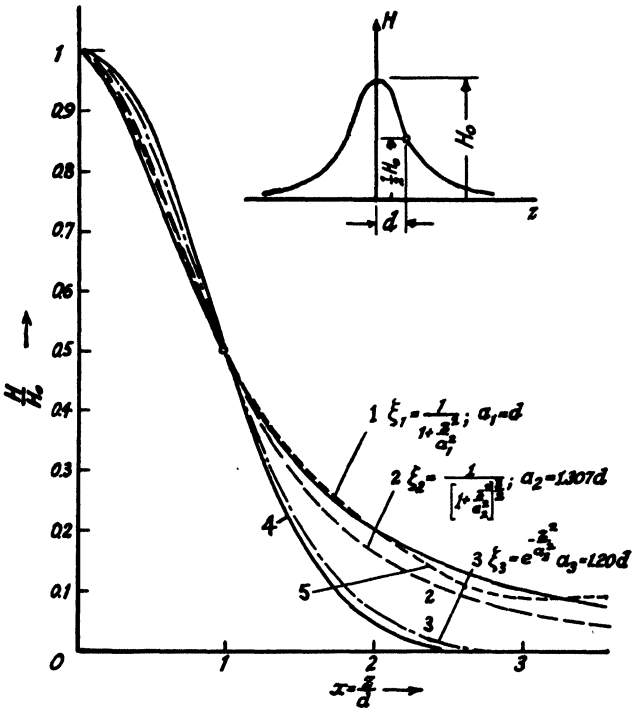


FIGURE 3. Theoretical and experimental axial fields, $H(z)$, of strong magnetic electron lenses. Curve 1: theoretical field $H(z) = H_0/(1 + z^2/a^2)$; curve 2: theoretical field of a simple coil with one turn of radius a_2; curve 3: theoretical field $H(z) = H_0 \exp(-z^2/a_3^2)$; curve 4: measured field in the normal case of unsaturated pole pieces; curve 5: measured field in the case of saturated pole pieces.[57]

U, decreasing with increasing U. A complete description of a magnetic lens (in terms of the general function (11)) would require the field to be expressed in the form of (11) with μ as a function of k^2, defined uniquely for fixed values of U.

An experimental comparison is illustrated in Fig. 5 (Fig. 8[57]). Although a numerical correlation of k^2 with the ampere turns (NI) in the coil and data of the pole pieces is lacking, it may be estimated by a comparison of these results with the earlier measurements of Ruska.[134] It is nevertheless interesting to observe the excellent agreement between the calculated value of the focal length, curve 5, on the basis of equation (10), with the experimental curves 1, 2, 3, 4, for small k^2. The agreement between the calculated values (marked x) on the basis of equation (4) with the focal lengths in the saturation range, further verifies the accuracy of these theoretical methods. In the electron microscope, the lenses are generally used in the range of k^2 preceding the establishment of complete saturation, with focal lengths somewhat above the minimum values.

(b) **Spherical and Chromatic Aberrations of Strong Magnetic Lenses.** The second approximation in the derivation of the lens equation from the general Lorentz equation reveals the existence and character of the third-order lens aberrations. These have been classified and labelled [65, 144] in terms of the known third-order aberrations

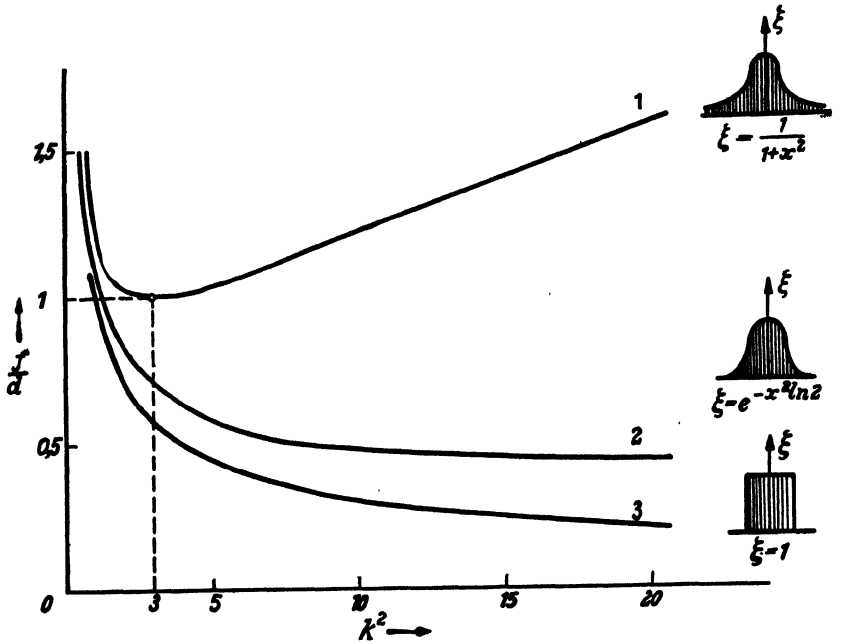


FIGURE 4. The focal length f (f_0 or f_1) of a magnetic lens as a function of the parameter k^2. Curve 1: with the field $H(z) = H_0/(1 + z^2/a^2)$; curve 2: with the more sharply defined field $H(z) = H_0 \exp(-z^2/a^2)$; curve 3: with the ideal most sharply defined field $H(z) = H_0$ $(-d < z < d)$, $H(z) = 0$ $(z < -d,\ z > +d)$.[57]

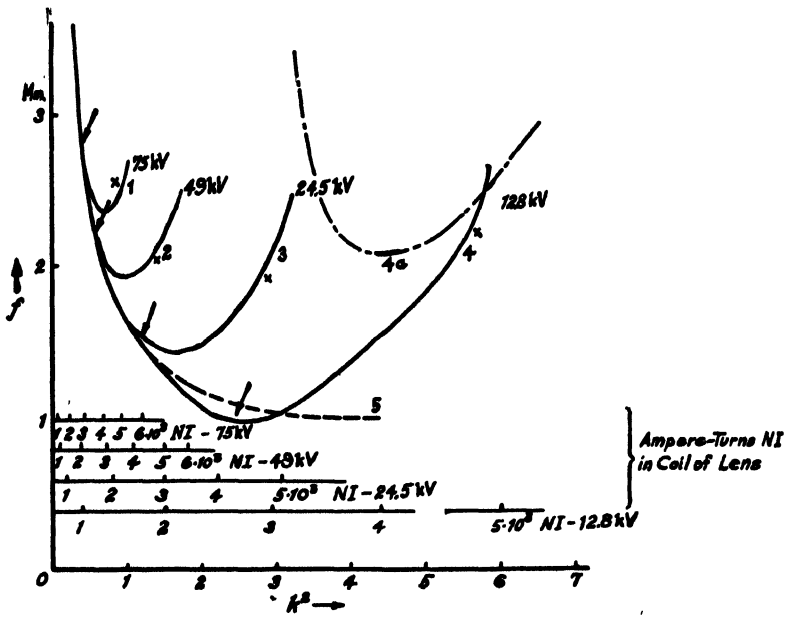


FIGURE 5. Experimental determination of the focal length f of a strong magnetic lens as a function of k^2. Points 1, 2, 3, 4: calculated focal lengths for the case of saturated pole pieces; curve 5: calculated focal lengths for the normal case of unsaturated pole pieces.[57]

of light lenses. Of particular interest in the study of microscope objective lenses is *spherical aberration,* the only aberration not reduced to zero as the object point approaches the optic axis of the lens. As a result of this aberration, the image of an axial object point appears as a disc of illumination, the radius of which, Δr, is proportional to the cube of the aperture α of the lens (see Fig. 1). This radius, divided by the image magnification, is used as a measure of the spherical aberration, denoted by the symbol δ_s. If V is the magnification, then

$$\delta_s = \frac{\Delta r}{V} = C_s \cdot \alpha^3 \tag{12a}$$

where C_s is a constant of the lens for fixed object and image positions.

In the case of the field represented by equation (4),

$$C_s/a = \left[\frac{\pi k^2}{4(1+k^2)^{\frac{1}{2}}} - \frac{1}{8} \frac{4k^2-3}{4k^2+3} \cdot \sin \frac{2\pi}{\sqrt{1+k^2}} \right] \frac{1}{\sin \frac{4\pi}{\sqrt{1+k^2}}} \tag{12b}$$

The analytical expression, equation (12b), for the case in which the object is at a focal point is shown by the dotted curve marked γ_s in Fig. 6.[57] It has been shown by Glaser that the minimum spherical aberration, insofar as it depends upon a selected object position, differs inappreciably from the values shown in the figure. The effect on the spherical aberration constant C_s of a dissymmetry in the field has been computed by Dosse.[56] In the practical range of k^2, the aberration increases rapidly to large values if the half width on the object side of the lens is less than the half width on the image side. The aberration is reduced in a comparatively small proportion if the larger

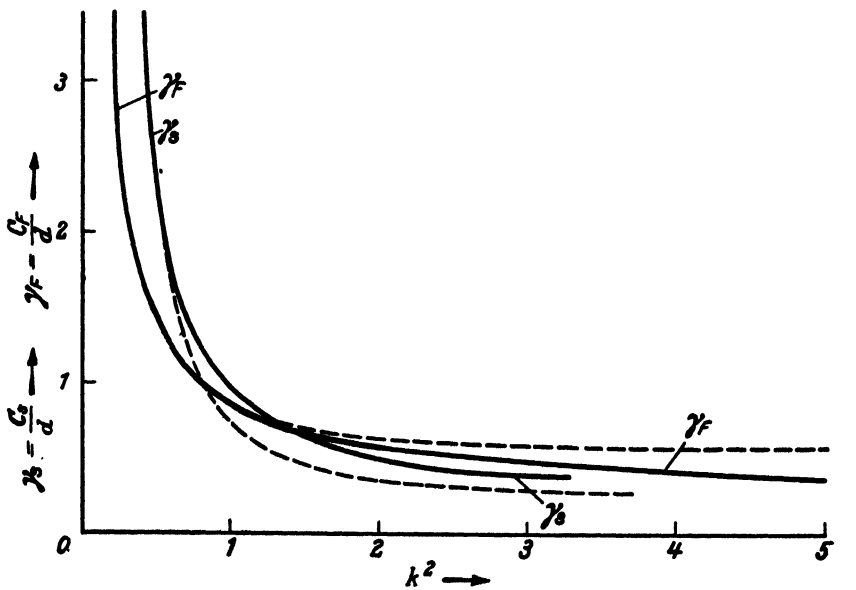


FIGURE 6. Spherical C_s and chromatic C_F aberration constants of magnetic lenses as a function of k^2. Curve γ_s (full) — calculated ratio C_s/d for unsaturated pole pieces; curve γ_s (dotted) — calculated ratio C_s/d for saturated pole pieces; curve γ_F (full) — calculated ratio C_F/d for unsaturated pole pieces; curve γ_F (dotted) — calculated ration C_F/d for saturated pole pieces.[57]

half width is on the object side. Some advantage, therefore, is gained by constructing the pole pieces with a larger internal diameter on the object side of the lens.

The numerically computed spherical aberration constants for the normal unsaturated fields, shown as the full γ_s curve in Fig. 6, agree very closely with the values which may be inferred from the analytical expression, equation (12b). For the purpose of estimating the effect of this aberration on the resolving limit of magnetic lenses [see Sec. III(c)], the analytical expression consequently appears to be adequate. The failure of this analytical method for the calculation of the cardinal points of strong microscope lenses, however, is well illustrated in Fig. 5.

The experimental measurement of the spherical aberration of objective lenses is difficult to carry out with high accuracy. The method of von Ardenne [11] requires the use of excessively large apertures to enhance the aberration to a sufficient extent to make it measurable. The experimental observations must subsequently be extrapolated several orders of magnitude to the range of lens apertures used in practice. A similar criticism applies to the method of von Borries and Ruska,[44] in which the spherical aberration is thought to be isolated by use of large lens apertures in a normal microscope image. The extrapolation is not justified because of the existence of errors of higher order. In the method of von Borries and Ruska, it is further doubtful if the assumed apertures obtain experimentally. Probably the most accurate measurements are those of Dosse,[57] using the method suggested by Boersch.[38] These experimental results agree very well with the theory reviewed above, thereby showing that the results of Ruska and von Ardenne are several orders of magnitude too high.

As the focal length of a magnetic lens is determined by the parameter k^2, containing the ratio H_0^2/U, it varies with independent variations of either H_0 or U. Independent variations of H_0 can be eliminated experimentally, for the field is determined by the current in the coil of the lens. It is therefore sufficient in this discussion to consider only the effect of variations in U. In consequence of a variation ΔU in the accelerating potential of the electron beam, the image of an axial object point appears in the original image plane as a disc of illumination. The radius of this disc, when divided by the magnification, is a measure of the chromatic aberration of the lens. It may be expressed in the form [66]

$$\delta_F = \cdot C_F \frac{\Delta U}{U} \alpha \cdot \tag{13}$$

where C_F is termed the chromatic aberration constant of the lens.

An analytical expression for C_F can be derived if the equation (1) is exactly soluble. On the basis of the theoretical field (4), Glaser obtained the expression

$$C_F/a = \frac{\pi k^2}{2(1+k^2)^{\frac{1}{2}}} \cdot \frac{1}{\sin^2 \frac{\pi}{\sqrt{1+k^2}}} \tag{14}$$

The chromatic aberration constant in this case is illustrated as a function of k^2 by the dotted curve marked γ_F in Fig. 6. Numerically calculated values, on the basis of the field (10), are illustrated by the full curve γ_F. The agreement in the range of small values of k^2 indicates that the chromatic aberration of magnetic lenses can be estimated with adequate accuracy by use of the analytical expression, equation (14).

With the exception of the test measurements of Dosse,[57] direct experimental measurements of the chromatic aberration of strong magnetic lenses have not been reported. The example given by Dosse demonstrates the agreement between the theoretical calculations and experimental measurements. For the case $U = 78$ kV, $\Delta U = 360$ volts, $f = 2.4$ mm, the measured value of the chromatic aberration leads to

a value of C_F between 1.35 and 1.65 mm. The corresponding theoretical value, deduced from Fig. 6 ($d = 1.7$ mm, $NI = 3900$, $k^2 = 0.64$) is 2.0 mm.

(c) **The Electrostatic Objective Lens.** The electrostatic electron lenses are more closely related to ordinary optical systems. Any axially symmetric electrostatic field, which, together with its derivatives, is continuous in the neighborhood of the axis of symmetry, may be regarded as an electron lens. The square root of the potential at any point is equivalent to the refractive index and the entire field to an inhomogeneous isotropic optical medium. In an axially symmetrical field the curvature of the equipotential surfaces is spherical in the neighborhood of the axis, and so the field may be regarded as a lens system composed of an infinite number of adjacent and centered spherical refracting surfaces. The differential equation governing the trajectory of a paraxial electron traversing an electrostatic field of this character may be written in the form [145]

$$V(z)\frac{d^2r}{dz^2} + \frac{1}{2}\cdot\frac{dV}{dz}\cdot\frac{dr}{dz} + \frac{1}{4}\cdot\frac{d^2V}{dz^2}\cdot r = 0 \tag{15}$$

where $V(z)$ is the electrostatic potential on the axis of symmetry and r is the radial displacement of the electron from the z axis. The image-forming properties of such a field are contained in the observation that the equation is linear and of the second

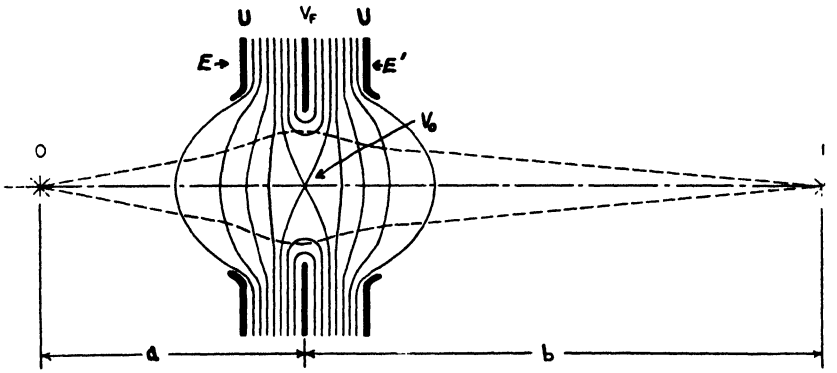


FIGURE 7. Electrostatic unipotential electron lens. The outer electrodes E E' are at a potential U, equal to the accelerating potential of the illuminating electron beam. The potential V_F of the focusing electrode determines the focal length f of the lens. In the limiting case of a thin and weak lens $1/a + 1/b = 1/f$; $1/f =$

$$\int_{-\infty}^{\infty}\left[\left(\frac{dV}{dz}\right)^2 \Big/ V^{\frac{3}{2}}\right]dz \Big/ 8U^{\frac{1}{2}}$$

order in r. Axial potential distributions with the necessary symmetry and continuity characteristics can be produced by applying potentials to a system of annular electrodes possessing rotational symmetry about a common axis. These electrodes may consist of any combination of cylinders, discs or cones.

Electrostatic lenses have been divided into two classes. In the *bipotential, immersion* or *accelerating* lenses, there is a potential difference between the object and image spaces. In the *unipotential* or *Einzel*-lenses, the potentials in the object and image spaces are equal. The bipotential lenses are particularly useful in the study of electron-emitting surfaces and in cathode-ray tubes in which it is desirable to combine the functions of accelerating the electrons and producing images in a single electrode system. The unipotential lens has been found more suitable for use in transmission electron microscopes.

This type of lens is illustrated in Fig. 7.[88] The outer electrodes EE' have potentials applied to them equal to the accelerating potential, U, of the illuminating beam.

The potential V_f of the inner or *focusing* electrode may be either positive or negative relative to the outer electrodes. In a transmission microscope objective lens it is convenient to select a focusing potential which lies between the cathode-anode potential of the electron beam. The value depends upon the dimensions of the electrodes and the desired focal length. The specimen is situated very close to the first outer electrode at O, and the image is formed at a considerably greater distance from the lens at I. In order to attain the short focal lengths which are required in a transmission microscope, the internal diameters of the electrodes and the spacing between them cannot exceed a few millimeters. An aperture-limiting diaphragm with internal diameter below $\frac{1}{10}$ millimeter is also desirable to enhance the contrast in the image and to reduce the spherical and chromatic aberrations of the lens. Since introduction of a diaphragm of arbitrary form would seriously distort the electrostatic field and its image-forming properties, it is necessary to provide an aperture diaphragm which is formed to replace a natural equipotential surface. The adopted procedure consists in forming one of the three lens electrodes with a sufficiently small internal diameter. In the instruments described by Mahl [105] and by Bachman and Ramo,[26] the outer electrode on the image side of the lens serves as an aperture. The diameter and shape of the focusing electrode is made so that a lens of selected focal length is obtained when this electrode is at cathode potential. This procedure obviates the necessity for a voltage divider in the high-voltage circuit and results in a lens of fixed focal length entirely independent of the accelerating potential of the electron beam. With the focusing electrode serving as an aperture diaphragm, a natural equipotential surface of a predetermined field distribution can be duplicated more precisely in principle. However, mechanical errors in forming the electrode are considerably more serious and a voltage divider is required.

In view of the fact that the electrode surfaces of the lenses are equipotentials, it is in principle an easy problem to compute the dimensions and positions of the electrodes which will produce a desired axial potential distribution. One procedure, demonstrated by Plass [120], consists in expanding the series solution of the Laplace equation,[50] *viz.*,

$$V(z,r) = \sum_{\nu=0}^{\infty} \frac{(-)^{\nu}}{(\nu!)^2} \cdot V(z)^{(2\nu)} \cdot \left(\frac{r}{2}\right)^{2\nu} \tag{16}$$

in terms of the axial function $V(z)$ to obtain the equipotential surfaces corresponding to each point on the axis. Electrodes shaped to conform to a small number of these equipotentials in regions removed from the axis may then be charged to the correct potential to provide the selected axial potential distribution. Electrode configurations for a few special axial functions may be computed by exact methods. Some of these are reviewed by Bachman and Ramo.[26a]

Although there are many methods available for the determination of the axial potential distributions of given electrode configurations, and subsequently the electron trajectories and aberrations,[34, 26a, 129, 163] relatively little data concerning the strong unipotential lens have been published.

Theoretical estimates of the magnitudes of the aberration constants of this type of objective lens have been reported by Dosse [57] and by Bachman and Ramo.[26a] In both investigations, numerical computations were carried out for special axial potential distributions which describe approximately the fields of real, strong lenses. The computations of Dosse indicate that the magnitudes of the spherical and chromatic aberration constants are from 5 to 10 times greater than in the equivalent magnetic lens.

The value of the spherical aberration constant estimated by Bachman and Ramo is somewhat smaller. In both investigations it is recognized that the chromatic aberration constant is greater in the electrostatic lens than in the magnetic lens. The latter result is not apparent if the lenses are considered as thin and weak.

The dependence of the focal lengths of both types of electron microscope lenses on the velocity distribution in the illuminating beam results in a reduction of the image definition in those areas of a specimen which are effective in destroying the velocity homogeneity of the illuminating beam. In the case of the magnetic lens, chromatic aberration may result from two additional sources which are independent of electron scattering in the specimen. These are variations in the accelerating potential of the illuminating beam and in the current of the lens.[3] An additional source of chromatic aberration in the electrostatic lens may also arise if the focusing potential is not maintained as a constant fraction of the cathode-anode potential of the illuminating beam. Departure from constancy may occur in a lens in which the focusing potential lies between the cathode-anode potential. Circuit reactances in the voltage divider must be adjusted to maintain a constant phase between the anode and focusing potentials relative to the cathode with an unregulated voltage supply if the effect is to be avoided.

The effect of the above aberrations on the resolving power exhibited in normal electron microscope images cannot be predicted quantitatively in terms of the above simple formulas. For a real specimen, an integration in the image plane must be performed in which both the distribution of the velocities of the scattered electrons and their angular distribution within the angle α are known. At the present time, only the qualitative data provided by the early scattering experiments are available. These data and the calculation of the magnitude of the chromatic aberration for various specimen thicknesses and beam voltages are presented in graphical form in publications of von Ardenne.[3, 5, 7]

(d) **Diffraction Aberration.** Although the effects of spherical and chromatic aberrations on the resolving power of a lens may be reduced without limit by reduction of the beam aperture α, there is a limit beyond which a third aberration becomes evident. This aberration is the electron-optical equivalent of *diffraction*, the limiting factor on the resolving power of the light microscope. It has been assumed that the kinematic wave theory of diffraction and the Abbe theory of microscope vision may be extended to describe the effects of diffraction on the resolving power of the electron microscope. It is of interest, therefore, to examine the conditions under which the validity of the extension might be verified experimentally. In the first case, an object is required in the form of an idealized line transmission grating of many elements. A beam of parallel light of wave-length λ incident on such a grating at an angle θ_K to its normal, but perpendicular to its elements, is scattered coherently. Sharply defined interference maxima occur at angles of inclination, θ, to the normal in accordance with the relation $n\lambda = \delta_d (\sin \theta_K + \sin \theta)$, where $n = 1, 2, 3, \ldots$ and δ_d is the grating interval. If the aperture of the objective lens is sufficiently large to receive both the zero order and the first order maxima which has the lesser inclination to the normal of the grating, the illumination in the image plane of the lens consists of a series of maxima and minima of intensity. The period of this pattern, when divided by the magnification, is equal to the period of the grating. The total number of diffraction maxima traversing the lens determines the degree of similarity of the image with the object. If not more than one order of the pattern enters the lens, the illumination in the image plane appears uniform and resolution of the grating is not obtained. Therefore, with oblique illumination and an objective lens of angular aperture α, the smallest grating space δ_d which can be resolved in the absence of other aberrations is determined by the relation $\delta_d = \lambda / (\sin \theta_K + \sin \alpha)$. If the specimen is mounted in a medium of refractive index μ, the objective lens aperture is circular,

and the illuminating beam is symmetrical about the optic axis with an aperture α, the above relation is modified to the form

$$\delta_d = \frac{1.22\lambda}{\mu(\sin\theta_K + \sin\alpha)} \tag{17}$$

This limiting formula therefore applies in the study of periodic structures of many elements and the resolution is determined essentially by the Fraunhofer diffraction arising in the object.

An alternative formula is applicable if the specimen consists of only two incoherent scattering centers or of two idealized independent point sources of illumination in the plane of the specimen. In this simple case, the illumination observed in the image plane consists of a pair of incoherent diffraction patterns resulting from Fresnel diffraction at the lens diaphragm. Under conditions which these two patterns may be just distinguished, the points are said to be resolved. The minimum distance of separation of the points in the object plane when they are resolved in accordance with the Rayleigh criterion is given by the formula

$$\delta_d = \frac{\lambda}{2\,\mu\sin\alpha} \tag{18}$$

This formula, directly applicable to an electron lens, has also been deduced from the uncertainty principle by Henriot. It is derived in a publication of Marton, Banca and Bender.[115]

Although von Ardenne[8] and von Borries and Ruska[42, 44] have adopted formula (17) and its implications to represent the limiting effects of diffraction on the resolving power of the electron microscope, the reasons for this choice are not apparent. Isolation of the diffraction aberration resulting from diffraction in the object plane would require a specimen consisting of a one- or two-dimensional grating with a period above the optimum resolving power of present-day objective lenses, at least 50 Å. With a modified criterion for resolution which takes into account atomic scattering amplitudes, the kinematic treatment is apparently justified.

It appears more logical to attempt the isolation of the diffraction aberration under conditions in which formula (18) is applicable. A suitable point source of illumination in the object plane might consist, for example, of a small opening (below 50 Å in diameter) in a heavy metallic foil illuminated by a beam of aperture equal to or greater than the aperture of the objective lens as determined by a metallic diaphragm. By this method, it is in principle possible to determine the distribution of the illumination within the objective aperture. Isolation of the Fresnel diffraction pattern originating in the lens diaphragm should be possible be reducing the aperture of the objective lens well below the value permitting attainment of optimum resolution. Validity of the extension of the kinematic wave theory to this case appears justified in view of the character of the straight-edge Fresnel diffraction pattern observed by Boersch.[89]

III. The Transmission Electron Microscope Image

(a) **Physical Interpretation of Micrograph Observables.** The mechanism of image formation in the transmission electron microscope is illustrated in Fig. 8. A simple specimen consisting of a small massive particle M supported on a thin film, is mounted immediately above the principal focus of the objective lens, in this illustration a magnetic lens, the pole pieces of which are denoted by N, S. The part B of the illuminating beam incident on the particle is scattered by it from a well-defined pencil of aperture δ (10^{-4} to 10^{-2} radian) into a diffused beam, the aperture of which is represented by α_0. The metal diaphragm D, in the magnetic field between the pole pieces, permits the passage of only a fraction of these scattered electrons through the

lens, *viz.*, those contained within the cone of aperture α. An illuminating pencil A, incident on an opening in the supporting film, is not scattered. With the aperture δ smaller than α, the entire incident pencil is transmitted through the lens. When the final image, formed by the pencils A', B', is observed on the fluorescent screen, the image of the opening appears bright, and the image of the particle dark.

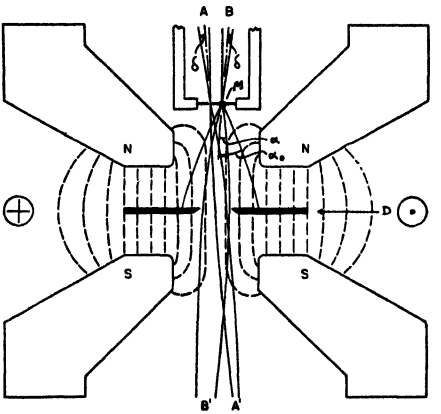


FIGURE 8. A convergent electron beam A incident upon an area free of material passes unscattered through the aperture diaphragm D between the pole pieces N, S of the objective lens. The image of such a clear area appears bright. An adjacent beam B striking a particle M in the specimen is scattered into a divergent beam. Only a fraction of the scattered electrons pass through the aperture of the lens to reach the image plane. The image of such a particle, therefore, appears dark. In the absence of a limiting diaphragm D, the image contrast is governed by the spherical aberration constant of the lens.

Well-defined images and adequate image contrast for the study of many materials may also be obtained without the use of an aperture-limiting diaphragm. This is thought to be primarily a consequence of the large spherical aberration constants of present-day electron lenses. The electron beam incident on an opening H (see Fig. 8) suffers no angular scattering and therefore passes through the objective lens with an aperture δ equal to that of the illuminating system. The boundary of such an opening is defined with high resolution by the electrons transmitted through the opening. An appreciable fraction of the electrons incident on any point in the supporting membrane are scattered over relatively large angles (of order 10^{-2} radian). Some of these appear in a diffraction pattern as diffuse rings, and as a consequence of spherical aberration, reach the image plane in areas far removed from the conjugated image points. A proportion of the electrons transmitted through the film thus arrives in the image plane within the area occupied by the Gaussian image of the opening. The total intensity within the latter area is the sum of the intensity transmitted through the opening and the scattered radiation from neighboring material. Only a small fraction of the electrons transmitted through a particle M in the specimen arrive in the image plane within the area representing its image. The remainder are scattered over the entire field of view, thereby enhancing the background intensity.

On the basis of the above qualitative explanation, it is apparent that the effective width of the intensity transitions in crossing the boundary between mass or thickness discontinuities is determined by the resultant of the aberrations in the area causing the *lesser* angular scattering. Although the resolving power within areas representing massive particles may be exceedingly low, the peripheries of such particles may

be sharply defined if the particles are suspended freely, or supported on a film of small density and thickness. In view of the fact that the peripheries of particles are defined with nearly equal sharpness either when they are supported freely or on a thin supporting film, it must be concluded that the effective angular aperture of the electrons scattered by such thin films is of the order of 1×10^{-3} radian. The electrons scattered over wide angles, such as those scattered coherently, have little effect upon image definition, but are responsible for the image contrast.

In general, both image contrast and definition are improved by use of an aperture-limiting diaphragm in the objective lens. An obvious exception occurs in the observation of the peripheries of particles suspended in the specimen without a supporting membrane. In the latter case, the definition is determined by the aperture of the illuminating beam alone.

The mechanism of image formation discussed above results in an interesting contrast between light and electron microscope images. Whereas the resolving power exhibited in a photomicrograph is essentially constant over the entire field of view, there are large variations in the resolving power over the field of view of an electron micrograph. In the bright background and about the peripheries of well-defined structures supported freely or by a thin supporting membrane, the optimum resolving power may always be attained. In the interior of a massive structure the resolving power is reduced, and with structure thicknesses above *ca.* 50 millimicrons, may be comparable with or less than the resolving power in a photomicrograph. This observation is of importance since it indicates that large structures commonly studied in the light microscope may also be studied with advantage in the electron microscope. Significant boundary details, obscured in a photomicrograph, are frequently revealed in an electron micrograph.

The immediate data provided by the image are two-fold. Providing there is sufficient image contrast, the *cross-section dimensions* of a structure obstructing the continuous rectilinear propagation of the electrons are immediately apparent. The *intensity differences* in the images of neighboring structures contain further information regarding the third dimensions of the structures and their composition.

The two dimensional image of a structure is not the image of a cross-section through any well defined plane, since the focal depth of the electron microscope is very large. With a resolving power δ_R, focal length f, and lens aperture diameter D. the focal depth is given by the relation

$$T = \frac{2 \cdot f \cdot \delta_R}{D} \tag{19}$$

In accordance with this simple formula, a focal depth of two microns is obtained with a resolving power of 3 millimicrons and a lens aperture $\alpha = 1.5 \times 10^{-3}$ radian.

The above formula may be regarded as reasonably accurate only if the objective lens aperture is defined by a mechanical diaphragm and is equal to or less than the aperture of the illuminating beam. In the absence of a mechanical diaphragm, the focal depth is a function of the angular scattering produced by the adjacent object elements used as test objects. It is determined essentially by the effective aperture of the scattered electrons emanating from the specimen area causing the lesser scattering. In the case of freely suspended specimen material, the focal depth, therefore, may be estimated by replacing $D/2f$ in formula (19) ,with the aperture of the illuminating beam. In consequence of this large focal depth, it is possible to change the inclination of the specimen relative to the optic axis and by superposition of the resulting micrographs to infer the magnitude of the third dimension (within the limits discussed in Section VI). Descriptions of several methods for obtaining stereoscopic micrographs and numerous illustrations are included in the publications of von Ardenne,[18] Mahl [104] and others.

Interpretation of the relative *intensities* in the electron microscope image is a much more complex and difficult problem. At the present time even qualitative conclusions regarding the significance of relative image intensities are not always beyond question. The earliest experiments performed with cathode rays by Lenard [101] and others were concerned with the establishment of empirical laws governing the changes in intensity, the average change in speed, the absorption and diffusion of fast electrons which resulted from their passage through relatively thick layers of solids. In all these experiments, however, solid films with a thickness below one micron were not investigated.

With the conception of the Rutherford model of the atom, many of these experiments were repeated with the object of determining the interaction of individual pairs of atomic particles. The critical experiments of this period involved the study of the scattering of electrons by individual atoms isolated in gases; and the study of collisions at large scattering angles by atoms or ions in solids. With the development of quantum mechanics and electron diffraction techniques the processes of small-angle elastic scattering in solids and inelastic scattering by gases were subjected to intensive investigation. However, there are at the present time insufficient quantitative data regarding the scattering of electrons in the range of small angles (10^{-4} to 10^{-2} radians) and small film thicknesses, *ca.* 10 to 500 Å, which are of greatest interest in the electron microscope problem, to permit a precise theoretical correlation of the observables in a micrograph with the optical constants of the electron microscope and the chemical composition of the specimens.

Marton [118] and von Ardenne [3, 5, 7] have carried out computations of the image intensities for various film thicknesses and beam voltages on the assumption that multiple scattering obtains throughout the range of specimen thicknesses of interest in the electron microscope problem. Comparison of image intensities predicted by these formulas with direct observations in the electron microscope shows that in the range of film thicknesses most frequently studied in the electron microscope (100 to 1,000 Å), the theoretical values are several orders of magnitude lower than the observed intensities. The reason for the discrepancy is evident when the multiple-scattering formula is compared, for example, with the experimental determination of angular scattering distributions by Kruppke.[99] The experimental curves show that multiple scattering is not fully developed until film thicknesses of the order of 10^{-4} cm are reached (for Al) and that for film thicknesses of the order of 300 Å, with electron velocities between 40 and 70 kv, the multiple-scattering formula fits the observations very poorly.

A second theory was postulated by von Borries and Ruska,[42] who described the mechanism of image formation solely in terms of the unimpeded rectilinear transmission of electrons through solid films and the scattering resulting from electron diffraction. Although this simplifying assumption provides an explanation of image contrasts which is in better agreement with observations, the theory is inadequate since a great proportion of the scattered electrons is excluded from consideration. The qualitative deductions regarding the dependence of image contrast on electron beam accelerating potentials, which may be inferred from these early explanations, have been demonstrated experimentally by von Borries and Ruska,[45] Müller and Ruska [124] and Zworykin, Hillier and Vance.[162]

A recent publication of Marton and Schiff [116] represents the first step of a quantitative character in the correlation of image intensities with established physical theory concerning the scattering of electrons. Although satisfactory agreement is obtained in three examples between the values of film thicknesses computed with the aid of the scattering formulas and the values deduced by more direct methods, the general validity of the application of this method is questionable. The addition of elastic scattering intensities from individual atoms and ions requires modification to take into consideration the pronounced interference effects occurring in crystalline

solids, for it has been shown by Hillier and Baker [82] and by Heidenreich [74] that reflections from crystal planes produce marked contrast anomalies. It is doubtful whether the laws of inelastic scattering of free electrons by individual atoms or ions may be extended to the case in which the same units are bound in a solid with the consequent obliteration of the discrete electronic energy levels.

Intimately associated with the problem of correlating the image intensities with the thickness of the specimen material are the effects of electron scattering on the resolving power of the objective lens. The magnitudes of both the spherical aberration and diffraction aberrations depend critically upon the angular distribution of the scattered electrons proceeding from the specimen through the objective lens. The chromatic aberration depends both upon the angular scattering and the velocity scattering. In consequence of these observations and also the possible presence of yet unrecognized aberrations, it has been difficult to set up a satisfactory criterion for the specification of the resolving power. The simple computations of von Ardenne [5, 7] illustrate qualitatively the variation of these effects under different conditions of instrument constants and specimen materials.

(b) **Image Contours.** Displayed prominently in nearly all transmission electron micrographs with high resolution is the presence of a distinct bright halo about the boundary of well-defined particles or film discontinuities. Proceeding normally in the image across the boundary from an area of greater to an area of lesser intensity, the intensity rises sharply to a maximum in the vicinity of the boundary, then drops to a minimum (not always observable), thereupon rising to the constant value of the darker area representing the image of the more massive material. The interval between the extremities of this boundary anomaly is generally contained within 150 Å. The phenomenon is of some interest since it obscures the location of the physical boundary of the mass discontinuity in the specimen, thereby making it difficult to determine accurately the dimensions of a very small particles in a specimen.

Various factors have been considered as the origin of these contours. The resulting explanations have been made on the basis of the study of a small number of selected micrographs obtained under normal operating conditions and are not consistent with observations. The explanation of Borries and Ruska [43] is based upon the assumption that the effective width of the angular distribution curve of scattered electrons increases appreciably with the thickness and density of the specimen. As a consequence, the proportion of the scattered electrons passing through the peripheral regions of the objective lens aperture increases with the mass of the material, this resulting in an increase in the spherical aberration. In crossing the boundary from a thick film to a thinner one, there is an abrupt change in the magnitude of the spherical aberration. The hypothesis of Hillier,[79] in which it is assumed that there is an observable discontinuity in the chromatic aberration on traversing the boundary, leads to a similar conclusion.

Inspection of the effective widths of the contours occurring in any micrograph in which there are particles or plates in a wide range of thicknesses leads to the conclusion that the essential character of the contours is unaffected by the ratio of thickness or density of neighboring specimen areas. A comparison of the contours in micrographs of similar specimens obtained with different objective lens aperture diameters indicates that the aberration constants are not altered sufficiently, relative to each other, to change the character of the contours. It is evident that the variations of the spherical and chromatic aberration constants are not the primary factors underlying the contour phenomenon. Neither are these hypotheses consistent with the observations described in the second and third paragraphs of Section III (a).

From the observation of micrographs obtained in the transition from bright field to oblique dark-field illumination, the writer has concluded that electrons grazing the edge of a solid are deflected toward it. The experimental conditions are illustrated in Fig. 9. To the left of the point A in the object plane of the lens provided with the

diaphragm D, the entire illuminating beam of aperture δ would be transmitted through the lens in the absence of scattering. This is the region of normal bright-field illumination. The entire beam to the left of B would be excluded, this being the region of oblique dark-field illumination. Without scattering material in the object plane, the intensity distribution in the image plane in proceeding from the bright field to the dark field is represented by curve 1. With a film having an opening H in the object plane, the incident electron beam is scattered and an image of the film is formed in

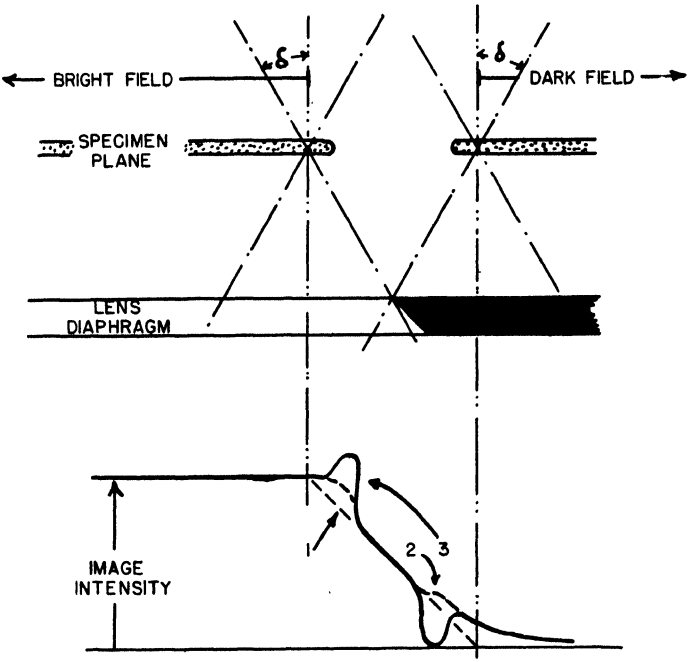


FIGURE 9. Image intensity across the diameter of a hole in a specimen supporting film in the transition region between bright field and oblique dark field illumination. Curve 1: intensity transition in the absence of a film in the object plane; curve 2: expected intensity transition in the image of the specimen in the absence of contour phenomena; curve 3: observed intensity transition across the diameter of a hole in a cellulose-nitrate film.

the dark-field region by scattered electrons alone. The intensity distribution which would be expected in the absence of boundary effects is represented by curve 2. The observed intensity distribution, however, represented in curve 3, shows a clear *maximum at the edge of the opening adjacent to the bright-field area* and a *minimum at the edge adjacent to the dark-field area.* A qualitative explanation of these image borders in the transition region may be based upon the existence of a potential drop of the order of 10 volts across the boundary of solids. The assumption that the specimens are electrically charged during observation is unsatisfactory in accounting for the sharpness of the contour maximum and the observations that the character of the contour is maintained when a metallic specimen is connected to a source of known and variable potential.

Further verification of the hypothesis that electrons grazing the edge of a solid are deflected toward the material is provided by the study of the image contours in "out-of-focus" micrographs. Fig. 10a is a representative series of micrographs of an

opening in a cellulose nitrate film in which the focal length of the lens was *decreased* in steps from the value for which the film is considered to be in best focus. In addition to the normally observed contour maximum, a new maximum and minimum are developed. The normal maximum and the out-of-focus minimum deviate *away from* the edge (radially inward in Fig. 10a), whereas the out-of-focus maximum suffers a larger *inward* (radially outward in Fig. 10a) deviation.

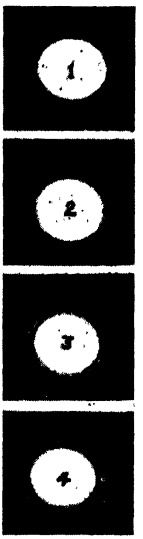


FIGURE 10a. Typical micrographs in "out-of-focus" series: 1: plane of specimen in focus; 2, 3, 4: successive planes between specimen and lens in focus.

Curve 1 in Fig. 10b illustrates the deviation of the normally observed contour maximum at one edge of the hole in Fig. 10a as a function of the displacement of the focal plane of the lens from the plane of the specimen. The center of the sharpest intensity transition in the micrograph in which the specimen is considered in best focus is used as a reference point. Curves 2 and 3 represent the deviations of the minimum and maximum which become clearly distinguishable with a focal displacement in excess of *ca.* 20 microns.

Assuming that the electron beam grazing the edge of the film suffers a real deflection, the true location of the physical boundary in the specimen may be inferred from this figure. Extrapolation of the out-of-focus maximum, curve 3, into the object plane determines the region effecting the deflection and which, by hypothesis, is the physical boundary. In this case, the physical boundary appears to lie approximately at the minimum of the normal image contour. As it is, however, difficult to specify the micrograph which is in best focus because of the large focal depth, the absolute values of the focal plane displacements and consequently the point of intersection of the curve 3 with the specimen plane, are obscured to some extent.

The first maximum of the Fresnel straight-edge diffraction pattern, in accordance with the demonstration of Boersch,[89] should be found along curve 5 in the absence of modifying effects. A modifying effect, *viz.*, spherical aberration of the lens, has been suggested by Hillier [80] as the reason for the displacement and distortion of the diffraction pattern. On this account, the first maximum of the Fresnel pattern should be expected along curve 4. Within the limits of experimental error in obtaining the curves 1, 2, and 3, it is possible to identify curve 1 with curve 4 and thus to explain the main features of the observed boundary anomalies.

(c) **Resolving Power of the Electron Microscope.** The review of the lens aber-

rations included in Section II (b), (c), (d) may be considered as incomplete for lack of a discussion of the resultant of the aberrations on the resolving power of the lenses. This conclusion was neglected because the experimentally estimated values of the resolving power depend not only upon the magnitude of the lens aberrations described in Fig. 6 and equations (17) and (18), but upon their magnitudes as determined by the scattering of the electrons by the specimens used as test objects. All

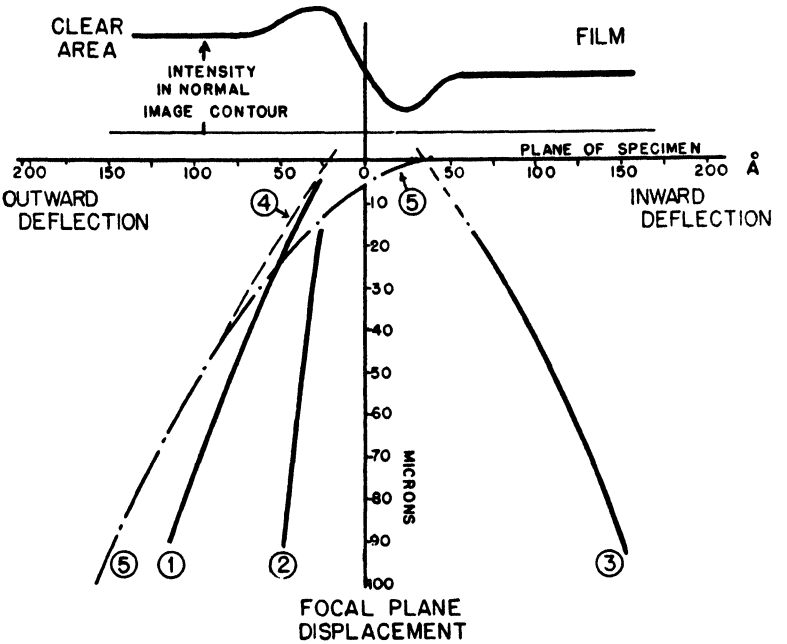


FIGURE 10b. Deflections of intensity extrema. Curve 1: normal image contour maximum (measured); curve 2: out-of-focus minimum (measured); curve 3—out-of-focus maximum (measured); curve 4: first maximum of Fresnel straight edge diffraction pattern displaced by spherical aberration; curve 5: theoretical deflection of first maximum of Fresnel diffraction pattern.

statements which have appeared in the literature with respect to the experimentally attained optimum resolving power refer essentially to the resolving power exhibited in the normal electron microscope image. Adequate knowledge of the instrument constants and electron scattering to permit these experimental values to be related with the magnitude of the lens aberrations is lacking.

The resolving-power measurements which have been reported are based upon observations in selected micrographs of normal specimens, *i.e.*, particles mounted in or upon a supporting film. Resolved "points" in the specimen consist of small, neighboring, scattering particles in the specimen. The images of such scattering centers appear as dark spots in a bright background. In general, no criterion is given by means of which the "points" are considered to be distinguishable, and a correlation of these dark point sources in a bright background with the bright point sources considered in the formulation of the aberrations is not made apparent. In addition to the unsatisfactory theoretical basis of these measurements, they are inadequate for the comparison of the relative merits of instruments developed or employed by different workers. The reported resolving power values are aptly described as "practical" [84] values, since pertinent data regarding the instrument constants and the in-

tensity distribution between the resolved points are not specified. Since the image contrast depends upon the effective apertures of the objective lenses as well as upon the relative densities and sizes of the "point" scattering centers, it might be assumed, for example, that the practical resolving power of 30 Å, reported by Hillier and Vance [84] from the observation of structural details revealed in a micrograph of Koroseal, would be lowered appreciably if they had obtained equivalent image definition in a micrograph of colloidal gold on a thin cellulose-nitrate film, the test specimen of von Ardenne.[19]

The metallic oxides proposed by Barnes and Burton [28] as test objects for the resolving power of electron microscopes may be useful as a qualitative comparison. From observations in micrographs of these materials it is difficult to formulate numerical values, even of the practical resolving power, which might be compared with the results of other workers. An examination of the numerous micrographs of magnesium oxide smoke which have been published indicates that the cubic crystal structure is maintained in the smallest particle sizes. The geometry of these crystals may be regarded as sufficiently simple to be used as test objects in the method of measuring practical resolving power suggested by the writer in the discussion of the results of von Kausche and von Borries [90] in Section VI.

Measurement of the effective widths of the intensity transitions in the image across the boundary of a dense and well-defined solid particle provides a measure of the resolving power if it is well above the range in which the image contours are observable. With the practical resolving power below *ca.* 100 Å, however, the intensity transitions are not monotonic, as described in Section III (b), and until a quantitative explanation of the contour phenomenon is available, the method is inapplicable.

The optimum resolving power is, of course, not attainable in the examination of the internal structures of thicker specimens (above *ca.* 2×10^{-6} gm/cm^2). It is assumed that both the spherical and chromatic aberrations are enhanced by the increased scattering. The additional effects of spatial scattering within the specimen reduce the attainable resolving power further. A qualitative demonstration of these effects on the contrast and resolving power in the image is included in the earlier publications of Marton [118] and von Ardenne.[3, 5, 7]

An optimum practical resolving power of approximately 50 Å was attained in selected micrographs by several laboratories in 1939, *viz.* (50 Å [33]; 50 Å [181]; 50 Å [51]). In the course of the last three years these limits have been extended in two respects. With considerable development in the power supplies, the stability characteristics of the instruments have been greatly improved. This has resulted in the attainment of an optimum practical resolving power of between 50 and 75 Å in the greater proportion of micrographs obtained under routine operating conditions. The practical limit in selected micrographs has also been reduced to some extent. The reported values, *viz.*, 25 Å [46]; 22 Å [20, 23]; 30 Å,[84] are difficult to compare with each other because of the somewhat arbitrary definition of resolving power. It is also difficult to compare the relative merits of different instruments in this respect or to estimate with great precision the improvement in the past three years.

On the basis of these achievements it might be inferred that the first phase in the development of the transmission electron microscope is completed. The initial objective of providing a practical instrument for investigations in the colloidal range of dimensions has been attained. The problems which have been overcome might be characterized as engineering difficulties, many of which have been solved in an empirical fashion. The second and perhaps more tedious phase of the development has already been initiated. One division of this development centers about the correction of the electron lenses. The immediate objectives may appear as lesser goals when compared with the initial success in surpassing the resolving limit of the light microscope by a factor of 20-50. The present problems are of a more fundamental character, and their complexity is well illustrated by the computations of the lens aberrations

by Glaser [66] and Dosse.[56, 57] A second and perhaps more extensive division of the problems which require attention is the interpretation of the electron microscope images in terms of the physical theories of electron scattering and interference phenomena. Initial steps in this direction have been taken and are exemplified by the computations of Marton and Schiff,[116] the observations of Hillier and Baker,[82] and the discussion of the image-contour anomalies in the previous section. The preparation of specimens and the interpretation of the micrographs in terms of the techniques and the possible artifacts introduced in the specimens under the influences of evacuation and electron bombardment are problems which must be solved by investigators using the instruments in systematic studies wherein the results may be integrated with those of older and established methods.

Of immediate interest is the question of the *ultimate* resolving limit of present-day types of electron lenses. To the approximation in which the magnetic fields may be described by the function (4), Dosse [57] has estimated the limit which could be obtained with the customary type of pole piece construction and the limitations imposed by modern magnetic materials. Assuming a beam potential of 70 kv and a maximum attainable field of 10,000 gauss ($H = H_0$), he concluded that the optimum field half-width, d, lies between 1.6 and 2.9 mm. Within this range, the focal length and the spherical and chromatic aberration constants are all very near minimum values. The numerical values C_s (min.) $= 0.92$ mm; C_c (min.) $= 1.57$ mm; and f (min.) $= 2.25$ mm are considered as possible of attainment. The resolving limit, determined by the resultant of spherical aberration and diffraction under these conditions, and calculated by a formula of Scherzer,[146] is given by the relation

$$\delta_R = \lambda \left(\frac{C_s}{\lambda}\right)^{\frac{1}{4}} = 0.57 \times 10^{-6} \text{ mm, or } 5.7\text{Å}$$

with an optimum lens aperture $\alpha = 6.7 \times 10^{-3}$ radian.

In view of the fact that the resolving limit depends upon the fourth root of the aberration constant C_s, it is evident that a reduction of the resolving limit below atomic dimensions is not likely to be attained by the present-day types of magnetic lenses, even if the maximum values of H_0 can be increased by the use of improved magnetic materials. Reduction of the spherical aberration constant by a factor of 16, for example, is effective in reducing the resolving limit only by a factor of 2. It is also pertinent to point out that the ultimate resolving limit of the electrostatic lens is not likely to fall far short of the magnetic lens in spite of the observation that the aberration constants are larger by a factor of between 5 and 7 than those of a corresponding magnetic lens.

The usefulness of a resolving power in the atomic range of dimensions has been discussed by Hillier [81] and Schiff.[147] On the basis of a quantum mechanical calculation, Schiff concludes that contrast considerations would limit the resolving power in the image, so that individual atoms would not be revealed in the normal bright field image with 60 kv electrons if the atomic number is less than *ca.* 7. The semi-classical considerations of von Ardenne [21d] indicate that both the Brownian motion and recoil from electron bombardment would be effective in obscuring the image of a freely suspended atom. However, with an atom embedded in a solid film, his computations indicate that the amplitude of the motion is below 1 Å. It may be concluded from these discussions that if the lenses can be adequately corrected, the method of dark-field illumination will provide decided advantages in the atomic range of dimensions in much the way it has proven valuable in colloidal investigations with the light microscope.

IV. Principles of Construction of an Electron Transmission Microscope

The first two-stage magnetic electron miscroscope was constructed in 1931 by Knoll and Ruska.[97] From a study of its characteristics, Ruska [133, 134] was able to

construct a second instrument in 1934, the resolving power of which later surpassed by more than an order of magnitude the resolving limit of the light miscroscope. On this continent, the development was initiated at the University of Toronto with the construction in 1938 of a microscope by Prebus and Hillier.[130, 51] These and other experimental developments [19, 111] have culminated in the construction on a commercial basis of electron microscopes for general laboratory use by the Siemens and Halske AG in Germany,[41] the RCA Manufacturing Company [84, 115] and the General Electric Company.[26]

Although they vary considerably in construction detail, the principal features of all magnetic electron miscroscopes in operation at the present time are sufficiently similar to be adequately illustrated by a description of the electron microscope constructed by the writer at the Ohio State University in 1940. At that time, it was common practice to set as an objective the attainment of a resolving power of 10 Å. The dimensions of an apparatus which would permit the observation and recording of images exhibiting this resolving power were estimated in terms of the resolving power of the human eye, of commercially available photographic emulsions, and of fluorescent screens prepared by customary techniques.

Assuming a resolving power of 0.1 mm at a viewing distance of 25 cm for the normal human eye, the necessary magnification of the electron microscope image exhibiting a resolving power of 10 Å, in order that all of its detail may be revealed, is given by the ratio of 0.1 millimeter to 1 millimicron, or 100,000 diameters. Direct observation of the electron microscope image at this magnification is, however, impractical. It is more convenient to secure a record of the image at a lower magnification and to enlarge the micrograph subsequently by customary photographic procedure to values for which the finest details can be seen by the unaided eye. With a photographic plate of fixed dimensions, the area of the specimen recorded in one exposure varies inversely as the square of the magnification. Of equal or greater importance is the fact that the necessary current density in the specimen for a given image intensity varies inversely as the square of the magnification. For these reasons, it is desirable to record the image at the smallest possible magnification. This requires the selection of an emulsion of high resolving power. Because of their relatively low sensitivity to light and their fine grain, lantern slide plates have been used extensively. These have an estimated resolving power of 0.01 mm. It has been demonstrated by von Ardenne [8] that the resolving power of an emulsion depends not only upon its grain size, but also upon its thickness, and upon the accelerating potential of the electron beam. It is, therefore, not permissible to assume that the resolving power of an emulsion is the same for electrons as it is for light. Speed and contrast characteristics also cannot be inferred from the manufacturers' characteristics. Some speed and contrast measurements with electrons have been reported by Marton [114] and by Baker, Ramberg and Hillier.[27] Assuming, however, that lantern slide plates with the above-indicated resolving power are used, it is apparent that an image exhibiting a resolving power of 10 Å would have to be magnified 10,000 diameters in order that all of its detail be recorded. This number represents the minimum necessary magnification for the implied, but as yet unattained, resolving power. With the highest resolving power now attained, *ca.* 25 Å, a magnification of 4,000 diameters is adequate for photographic recording.

The second factor to be considered in the choice of the minimum necessary magnification is the technique employed for focusing the images. A procedure which has been used extensively consists in focusing the image on a fluorescent screen with the unaided eye. Visual focusing is carried out by using an illuminating beam with an aperture which is greater than that required for optimum resolution. This results in the observation of an intense image in which the resolution is low, and which can, therefore, be focused accurately. Before making an exposure, the beam aperture is reduced. This reduces the intensity, but increases the resolving power. By proper

choice of apertures in the above two steps, it has been found possible to obtain an accurately focused micrograph exhibiting optimum resolution. Choice of the proper apertures is governed by the *relative focal depths* in the two cases.

If it is assumed, for simplicity, that the resolving power is the sum of spherical aberration and lens aperture diffraction as implied in the formula

$$\delta_R = C_s\delta^3 + \frac{\lambda}{2\delta} \tag{20}$$

the focal depth T is given by the relation

$$T = C_s\delta^2 + \frac{\lambda}{2\delta^2} \tag{21}$$

If, consistent with the discussion in the first four paragraphs of Section III (a), it is further assumed that the image definition is governed essentially by the aperture of the illuminating beam, optimum resolution is attained when

$$\delta = \left(\frac{\lambda}{6C_s}\right)^{\frac{1}{4}} \tag{22}$$

If the recorded image is to be accurately focused, the focal depth obtaining during the exposure must be *equal to* or *greater than* the focal depth obtaining when the image is being focused.

It is apparent from inspection of equations (20) and (21), that the value of δ for which T is a minimum is greater than the value of δ for which δ_R is a minimum. Hence, the focal depth under conditions of optimum resolution is greater than the minimum focal depth. Of greater significance is the result that there is a second and larger value of δ for which the focal depth is equal to the focal depth under exposure conditions. The latter value of δ is the limiting value which may be used during visual focusing if the recorded image is to be in best focus. On the basis of these simple formulas, numerical values may be estimated. Assuming, for example, that $\lambda = 5 \times 10^{-9}$ mm (50 kv electrons), and $C_s = 100$ mm, optimum resolution is obtained when $\delta = \delta_{opt} = 1.7 \times 10^{-3}$. The focal depth in this case is $T = 1.0 \times 10^{-3}$ mm. The computed limiting value of δ, *viz.*, $\delta_{11m} = 2.9 \times 10^{-3}$.

It is evident from the results of many workers that the "practical" resolving power, measured as the distance between "just distinguishable" small particles supported in the usual thin film, remains sensibly constant if the aperture of the illuminating beam is reduced by as much as a factor of order 10 below the above computed value for optimum resolution. This result indicates that the effective aperture of the electron beam scattered by the normal supporting film is of the order of 1×10^{-3} and consequently determines both the resolving power and focal depth. It is reasonable, therefore, to assume that the focal depth is essentially independent of the aperture of the illuminating beam in the range below δ_{opt}, and has a value close to that computed from formula (19) with $\delta = \delta_{opt}$. In the higher range of δ, where the resolution is determined primarily by spherical aberration, formula (19) may be regarded as reasonably accurate. It follows that the computed numerical values of δ_{opt} and δ_{11m} serve as a useful guide in selecting the proper aperture of the illuminating beam to assure sharply focused micrographs.

If the selected value of δ for visual focusing is in the neighborhood of 10^{-2} radian and an electron gun is used which produces a beam with discontinuities in the angular intensity distribution, multiple images may be observed in the micrographs. The occurrence of such multiple images is conclusive evidence that the selected aperture is excessively large. Since there are other factors, some of which are not clearly understood, which prevent the consistent attainment of the highest resolving power of

which a given lens is capable, it is difficult to determine the best value experimentally. It is, of course, desirable to use the largest permissible value in order to obtain greatest image intensity. In the course of preparing several hundred micrographs, it became apparent to the writer that the *average* of properly focused micrographs does depend upon the aperture chosen for visual focusing, and that the average is greater by adhering to apertures in the neighborhood of the computed limiting value.

Assuming the correctness of the numerical value 1.7 for the ratio $\delta_{opt}/\delta_{lim}$, the ratio of the resolving powers in the two cases is given approximately by the factor 1.7. Consequently, the minimum necessary magnification required for accurate visual focusing, with the unaided eye, of an image exhibiting a resolving power of 10 Å is 60,000 diameters. With the highest attained resolving power at the present time, *ca.* 25 Å, a magnification of only 23,500 diameters is necessary.

Because of the large difference between the magnification of 23,500 diameters required for accurate visual focusing and the magnification of 4,000 diameters required for photographic recording, the advantage of using an ocular to focus the image on the fluorescent screen is evident. The majority of the instruments now in use were designed for high electronic magnifications to avoid the use of an ocular. By the use of fine-grained fluorescent screens, however, it is possible to use a magnifying lens for visual focusing. The permissible magnification depends upon the type of screen. In the instrument constructed by the writer, the fluorescent screen may be viewed with an ocular magnifying it up to 5 diameters. Although the instrument was initially designed for higher magnifications, the images are focused upon the fluorescent screen at a magnification of approximately 7,500 diameters, and are viewed with aid of the ocular at a magnification of 37,500 diameters. Although these are above minimum necessary values, they provide relief from eye-strain. In the new instrument recently developed by the RCA Manufacturing Company, even smaller electronic magnifications are used. In both this instrument and the electrostatic electron microscope of the General Electric Company, the fluorescent screens may be magnified from 10 to 20 diameters without loss of detail in the images. Von Ardenne was perhaps the first to advocate the use of an ocular lens for visual focusing. He has used a grainless single-crystal fluorescent screen upon which the image could be observed, with full content detail, at the much greater magnifications provided by a large-aperture microscope objective lens.

There is a definite advantage in using the previously described two-step procedure in focusing micrographs. The image observed in the first step has much greater intensity than the image which is recorded. If this procedure is employed, the minimum necessary electronic magnification in an apparatus provided with a suitable fine-grained fluorescent screen and ocular may be estimated in terms of the resolving power of the selected photographic emulsion alone. This will result in an instrument of smallest dimensions.

Having selected a value for the necessary electronic magnification to permit the recording of the smallest image details, the length of the instrument from the objective lens to the final image plane depends upon the focal lengths of suitable lenses, and the number of these to be used. The selection of lenses is at present somewhat arbitrary. Objective lenses with focal lengths between 3 and 5 mm have been found satisfactory for the attainment of a resolving power below 50 Å. Although von Ardenne has reported the use of an objective lens with a focal length less than 1 mm, the construction of such a lens involves more serious mechanical difficulties. Such lenses are entirely satisfactory, however, for enlarging the images provided by an objective lens. In this application, the spherical aberration constant is of minor importance. The distortion of the lens requires more careful consideration.

An instrument consisting of a single lens would have to have a length of many meters to provide a sufficiently great electronic magnification. A reduction to an apparatus length of less than one meter is possible by constructing a two-stage instru-

ment in which the image formed by one lens, the *objective,* is enlarged by a second or *projector* lens. The apparatus length of a two-stage instrument may be easily computed (see also [26b]) in terms of the final magnification and the focal lengths of the lenses which are to be used. With an objective lens of focal length f_1; a projector lens of focal length f_2 distant d from the objective (both lenses regarded as thin); and an apparatus length L (distance from objective lens to final image plane), the final magnification M is given by the relation

$$M \backsimeq \frac{d(L+f_1) - d^2 - L(f_1+f_2)}{f_1 \cdot f_2}$$

With fixed focal-length lenses, maximum magnification is obtained when $d = (f_1 + L)/2$. With the lenses disposed in accordance with this condition, the magnification is determined by the relation

$$M \backsimeq \frac{L^2}{4f_1f_2}$$

Thus, assuming the use of a 4-mm objective lens and a 1-mm projector lens, magnifications of 4,000 and 10,000 diameters may be obtained with apparatus lengths of 25 cm and 40 cm, respectively.

In the instrument constructed by the writer, the projector lens is a doublet. This permits the attainment of the magnification of a two-stage instrument with the use of longer focal-length lenses. The advantage of longer focal-length lenses in the projector is the attainment of a larger field of view for images observed at lower magnifications. This advantage is offset, however, by the necessity of careful adjustment of the relative focal lengths of the lenses to overcome image distortion.

This instrument is shown in Figs. 11a and b. The source of the electrons is a heated tungsten hairpin filament F in the electron gun G situated at the upper end of the apparatus. The electrons are accelerated by application of a negative potential, variable from 10,000 to 90,000 volts (up to 300,000 volts in the high-voltage RCA microscope [162]) to the cathode of the gun at H. The divergent electron beam passing through the opening in the anode is converged by the magnetic field of the coil C serving as a *condenser* lens. An image of either the source or of the "cross-over" may be focused in the plane of the specimen, situated at the lower end of the cartridge L immediately above the *objective* lens O. Under these conditions, both the aperture and intensity of the illuminating beam are at a maximum. The aperture is determined by the ratio of the diameter of the condenser lens aperture diaphragm to the distance between the center of the condenser and the specimen. The aperture and image intensity may be decreased by either of two methods. The aperture diaphragm may be replaced by one of smaller size, or by a dark-field stop. Several diaphragms of different diameters are fastened in a movable slide A_o, controlled from the exterior. Alternatively, the beam aperture may be decreased by raising or lowering the focal plane of the "cross-over" image.[48]

The electron beam, after transmission through the specimen, is converged by the magnetic field between the pole pieces of the objective lens, thereby forming an image of the specimen at low magnification on a fluorescent screen I. After surveying this image, a selected area may be adjusted to fall over a small opening in the fluorescent screen. This area is then enlarged to a final magnification in the neighborhood of 7,500 diameters by the projector lenses P_1, P_2. The final image may be viewed on the fluorescent screen S through the window W. A photographic plate, inserted under the fluorescent screen from the compartment PM, is exposed by removing the fluorescent screen.

Since the entire optical system is in a vacuum, all mechanical motions of the

specimen, the lenses, electron source, and photographic plate must be transmitted from the exterior through vacuum-tight joints. Most of these joints are provided with flexible metallic bellows, one of which is shown at *B*.

Routine operation of this instrument is in many respects simpler and more rapid

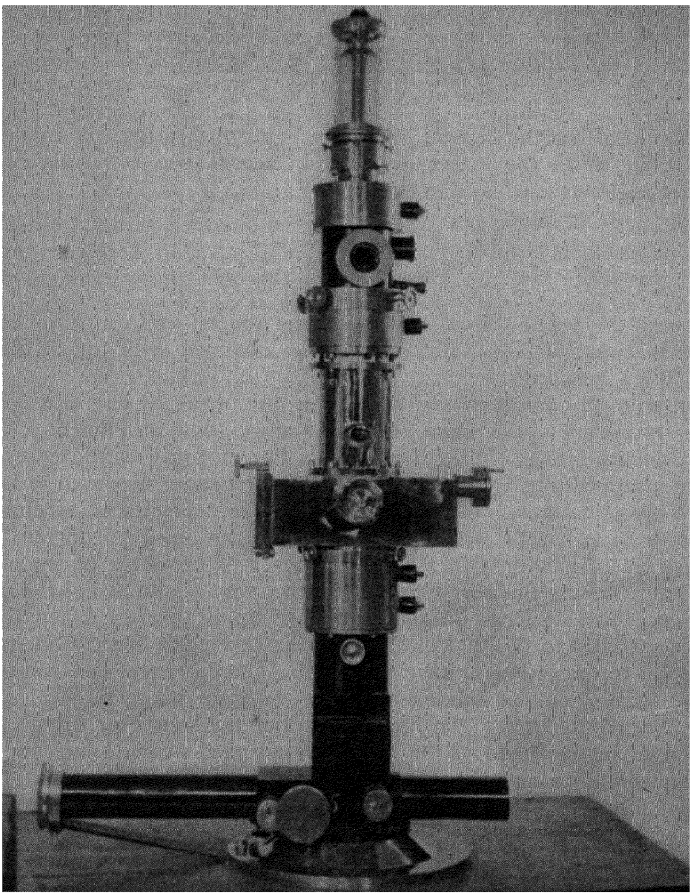


FIGURE 11a. Magnetic electron microscope constructed in the Radiation Laboratory of the Ohio State University in 1940.

than the operations of a high-power light microscope. Specimens may be inserted and ready for observation in an interval of from five to ten minutes. Photographic exposures rarely exceed 3 seconds. In this apparatus five two-inch square micrographs may be secured on one plate. An exposed plate may be replaced, and the system re-evacuated in less than four minutes. With improved air "locks," plates and specimens may be interchanged even more rapidly.[84]

In the photograph of the electron microscope illustrated in Fig. 11a there may be seen a second plate compartment, or camera, situated immediately over the projector lens system. This plate magazine has been installed for the purpose of recording the diffraction patterns of *selected microscopic areas* in the specimens normally studied

with the electron microscope. The principle of integrating the electron microscope with the diffraction camera may be considered as a development of the suggestion of Boersch.[86]

With the use of an electron gun which provides an intense beam of small initial cross-section and small aperture, it is possible in this instrument to restrict the illumi-

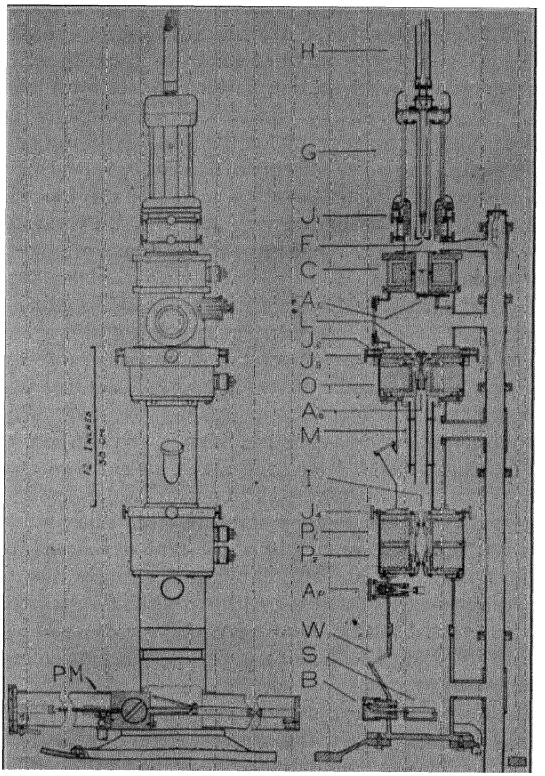


FIGURE 11b. Diagrammatic sketch of the electron microscope in Fig. 11a.

nated area of the specimen to a circle only a few microns in diameter. Even greater reduction of the illuminated area would be possible by the use of a specially designed condenser lens system such as described by von Ardenne.[22] Conversion of the optical system of the electron microscope to serve as a diffraction camera necessitates only the replacement of the normal condenser aperture diaphragm by one of small diameter. The new diaphragm serves as an electron source for the diffraction pattern and an image of it, together with the entire diffraction pattern, may be focused in the first image plane by adjustment of the focal length of the objective lens. Micrographs of the area in which the diffraction pattern is formed may be recorded either before or after recording the diffraction pattern without disturbing the specimen or the vacuum in the apparatus. It is also sometimes convenient to utilize the projector lens of the microscope to form highly magnified images of portions of the diffraction pattern. These may be recorded in the lower camera, thereby facilitating the measurement of intensity contours of the diffraction bands or spots.

In the apparatus employed by the writer at the present time, it is not possible to identify the area producing the diffraction pattern to within closer limits than

approximately 40 microns. In consequence, the electron gun is adjusted to provide an illuminated area of this diameter. A more precise identification of the specimen area studied by diffraction is considered possible in the design of a new instrument.

An electron diffraction attachment for both transmission and reflection studies has been developed and described by Hillier, Baker and Zworykin.[88] In this apparatus, the specimen holder may be taken out of its position above the objective lens and inserted between the lenses of a double projector. This permits the electron microscope to be used either as electron microscope or as a diffraction camera. In the latter case, the first projector provides a point source of electrons, and the second forms an image of this source together with the diffraction pattern on the final image plane.

Description of a magnetic electron microscope is incomplete without reference to the power supplies providing the current for the lenses and the accelerating potential for the illuminating beam. Since the chromatic aberration of this type of microscope is enhanced by fluctuations in the accelerating potential and lens currents, it is necessary to maintain these constant within very close limits.[84] Power supplies with special stabilizing devices have been developed for these instruments and may thus be considered as integral parts of them.

Storage batteries have been used successfully by many workers to serve as current sources for the lenses. They are, however, unsatisfactory in several respects. As a battery is essentially a constant-voltage source, long periods are required to establish temperature equilibrium in the coils of the lenses and control rheostats in order to ensure current constancy. Voltage drift of undesirable magnitude over periods exceeding one or two minutes are unavoidable without excessively expensive batteries. Both of these difficulties are overcome by the use of an electronic rectifier and current regulating system. Very high current stability (variations less than one part in 50,000) for short intervals and satisfactory stability over indefinitely long periods can be obtained from a stabilized power supply of the type developed by Vance.[157] Power supplies for the lenses may eventually be eliminated entirely through the use of magnetostatic or permanent-magnet lenses. Although relatively poor results were obtained with the experimental magnetostatic lens of von Borries, Ruska, Krumm and Müller,[47] a more careful design with the use of modern permanent magnet materials will result in useful lenses.

Stabilization of the high-voltage system presents a more difficult problem. In a high-voltage system consisting of a transformer and rectifier, there exist three sources of variations in the accelerating potential, *viz.*, variations in the voltage applied to the transformer primary, ripple, and variations in the output current. Fluctuations in "line" voltage can be reduced to negligibly small values by a number of methods, including the use of special generators, and special regulating transformers of the type manufactured by the Sola Electric Company. Ripple in the rectified voltage can similarly be reduced to small values by the use of capacity-resistance or -inductance filtering. The third source of voltage variation is more difficult to control. It is the result of electron current variations in the electron gun of the instrument which may be due to temperature variations affecting the specific emission from the filament and the interelectrode spacings in the cathode, and to changes in the emission constants of the filament.

The most elegant solution to this problem is the radio frequency power supply developed by Vance.[157] Since the necessary filter condenser capacity for a given ripple attenuation varies inversely as the frequency, the use of radio frequency effects a very large reduction in the capacity and consequently the physical size of the filter system. Of greater importance is the reduction of the time constant of the filter system. This makes it possible to detect and amplify very small variations in the potential at the cathode of the electron gun, and to compensate for any variations by application of a commensurate and opposed voltage variation in the primary of the high-voltage transformer.

A 60-cycle high-voltage system stabilized against the three sources of variations has been developed by the writer. A description of this circuit, illustrated in Fig. 12, may prove interesting to those unfamiliar with the high-frequency techniques required in the construction of a radio frequency system. The secondary of the high voltage transformer S and the low potential terminals of the filter condensers C_1, C_2 are connected together and insulated from ground. A high resistance R_3 (400 megohms) in series with the lower resistance R_4 (0.5 megohm) are connected in parallel with the electron microscope gun represented by M. By this means, a small

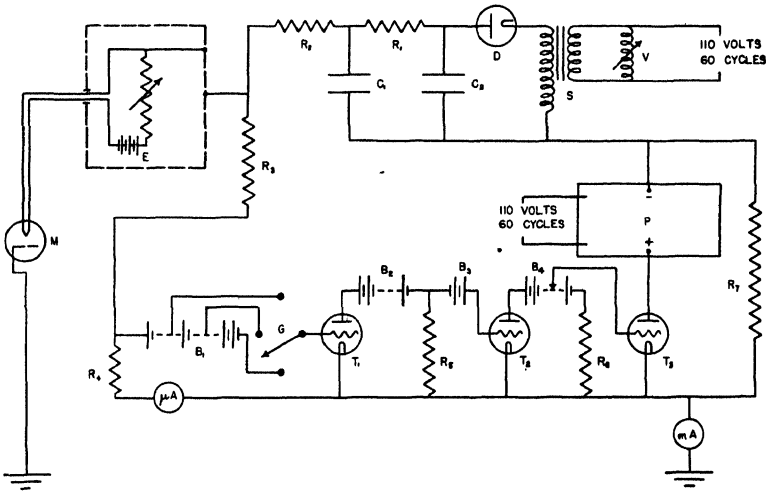


FIGURE 12. Circuit of the stabilized 60-cycle high-voltage power supply used with the Ohio State University electron microscope.

fraction of the total accelerating potential is impressed on the grid of the detector tube T_1 suitably biased with a battery B_1. Voltage variations detected by T_1 are amplified with high-voltage gain and fed to the grid of the power tube T_3. An independent, unregulated power pack P (2,000 volts d-c) in series with the output tube and resistance R_7 provides a compensating voltage (within a range of 1,500 volts) by raising or lowering the potential of the low-voltage terminal of the rectifier and filter system relative to ground. By use of this degenerative type of stabilizer (voltage gain 50,000) the time constant of the 60-cycle filter system is determined by the external surface of the system relative to ground. Although it has not been possible to effect a ripple reduction commensurate with the voltage gain of the amplifier on the assumption of a zero time constant, an adequate reduction, *ca.* 100 to 1.5 volts, as measured with a vacuum tube voltmeter, is provided. Reduction of the effect of slow drift is more effective since in this case the time constant plays a minor role. Over prolonged periods, however, there is an undesirable voltage drift, attributed to EMF drops in the batteries of the circuit.

A highly stabilized high-voltage system is entirely unnecessary with the electrostatic transmission electron microscope of the type described by Mahl [103, 105] and Bachman.[26] As the electrostatic lenses require no current, the additional power supply required for the lenses of the magnetic microscope is eliminated. For these reasons, the electrostatic microscope is much simpler and less costly in construction and has stability characteristics exceeding those of the magnetic electron microscope. The principles of the instrument constructed in the AEG Laboratories are illustrated in Fig. 13, reproduced from a publication of Mahl.[105] Electrons produced in

the cathode G are accelerated up to 50,000 volts to provide an illuminating beam emerging from the anode diaphragm B_1. A condenser lens has been avoided by suitable design of the cathode of the electron gun. The objective lens L_1, with the focusing electrode at the potential of the cathode of the gun, is provided with an aperture-limiting diaphragm in the form of a very small opening (1/10 millimeter in diameter) in the lower electrode which is at anode potential.

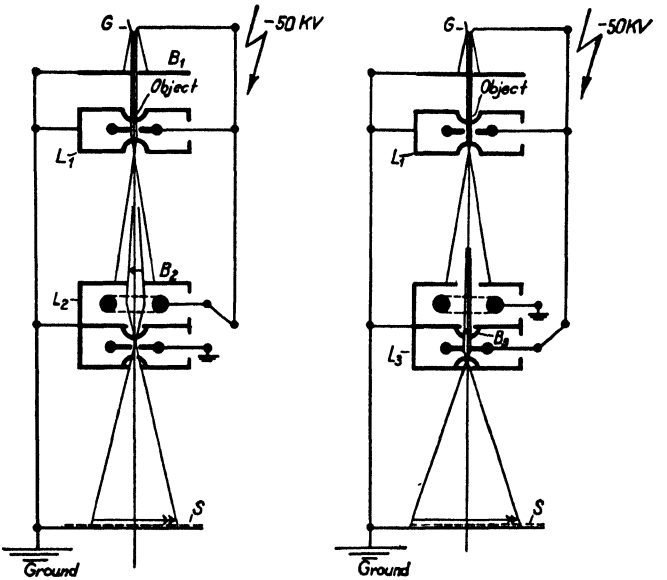


FIGURE 13. Schematic cross-section of the transmission electrostatic electron microscope described by Mahl [103,105]. (Left) Lens electrode connections for lowest magnification; (right) lens electrode connections for highest magnification. An intermediate magnification is obtained by connecting the focusing electrodes of both projector lenses to cathode potential.[105]

The double projector lens L_2, L_3, permits the specimen to be observed on the final fluorescent screen S at any one of three magnifications. These are fixed by the geometry of the instrument. A large area of the specimen may be observed at low magnification, *ca.* 1,000 diameters, by connecting the focusing electrode of the lens L_3 to the cathode and the focusing electrode of the lens L_2 to the anode as shown in Fig. 13 (left). A higher magnification, *ca.* 9,000 diameters, is obtained by interchanging the connections as shown in Fig. 13 (right). Intermediate magnification is obtained by connecting the focusing electrodes of both lenses to cathode potential. By use of the double projector lense, with these interchangeable connections, one of the main disadvantages of the electrostatic microscope with fixed focal length lenses is eliminated.

Independence of the chromatic aberration from fluctuations of the beam accelerating potential is demonstrated in the micrographs shown in Figure 14. The former was obtained with an unfiltered potential from a half wave rectifier, the latter with a filtered potential. It is evident from the equivalence of the image definition in these two micrographs that a total voltage fluctuation of 50,000 volts in each cycle has an imperceptible effect on the resolution.

Although the resolving power reported by Mahl, 80-100 Å, is somewhat less than the resolving power attained with the present-day magnetic electron microscopes,

there is no apparent reason for concluding that the limit of resolution of one is superior to the other in the same ratio. Indeed, the correction of the electrostatic lens may be a less difficult task, since the axial field is more intimately related to the geometry of the electrodes than is a magnetic field to pole piece geometry.

The limitation on the maximum attainable beam potential is more serious. This depends upon the greatest field which can be maintained between the closely spaced electrodes in the lenses. With the technique described by Mahl, the limit is about



FIGURE 14. Micrographs of magnesium-oxide smoke obtained with the electrostatic electron microscope: (left) with the electron beam potential varying over a range of 50,000 volts; (right) with a filtered beam potential. This comparison shows the insensitivity of the electrostatic electron microscope to fluctuations in the accelerating potential of the illuminating electron beam.[105]

60,000 volts. This is, of course, within the range of beam potentials commonly used at the present time with the magnetic electron microscope for which optimum contrast and resolution are obtained in the study of colloidal materials. This type of instrument could be used with equal success in many of the problems studied with the magnetic microscope. In view of its greater simplicity of construction and greater stability, it will probably be found more satisfactory in many applications, when developed further.

A commercial instrument of this type has been developed by the General Electric Company.[26] High resolving power has been sacrificed in this instrument in the interests of compactness, simple construction and stability of operation. In the many problems in which structural details requiring a resolving power below *ca.* 200 Å are of secondary interest, the application of this instrument is probably more economical than that of the magnetic electron microscopes now available. This instrument should prove particularly valuable to those research workers who are accustomed to using the light microscope near its limit of resolution.

V. THE PREPARATION OF SPECIMENS

Specimens to be studied in a transmission electron microscope differ in many respects from the type of specimen commonly studied with the transmission or ultramicroscopes. Since they are mounted in a vacuum chamber for observation, they must be free of volatile liquids. This requirement precludes the possibility of observing animation or reactions in liquid media. Whereas specimens of relatively

great thickness (several microns) are commonly prepared for observation in the light microscope and can be studied layer by layer with a high-power objective lens, the high resolving power of the electron microscope can only be exploited if the total thickness of the specimen together with its supporting slide and cover slip is below 50 millimicrons in thickness. As a consequence, the commercially available electron microscopes are of questionable value in the study of such specimens as cut sections. This limitation may be overcome by the use of other electron optical instruments less well known at the present time. It is evident that resolution surpassing the light microscope is possible by preparing wedge sections as described by von Ardenne [14] * and examining these in a high-voltage electron microscope of the type developed by Zworykin, Hillier and Vance [162] and by Müller and Ruska.[124] Even under these conditions, however, the resolving power does not approach the value obtained with the thinner specimens and lower-voltage instruments.

The conclusions of von Ardenne [10] regarding the possibility of observing living structures in the electron microscope are not considered as final. The killing effect of the electron radiation considered by him may not apply to small living organisms such as viruses which absorb relatively little energy from the illuminating beam compared to larger structures, such as bacteria and cells.

There are certain obvious ideal requirements to be satisfied by a material which is to serve as an electron microscope "slide." In the form of a thin film which is sufficiently transparent to the illuminating electron beam (thickness $\times$ density $=$ *ca.* 1×10^{-6} gm/cm^2) the molecular or crystal structure should be below the limit of resolution of the microscope. It should have sufficient tensile strength, flexibility and elasticity to withstand the mechanical strains resulting from the distortions and electrical charges of the specimen materials which may occur during preparation and observation. Resistance of the material to the action of the common chemicals and organic solvents generally employed in colloidal and biochemical preparations is highly desirable. From a practical standpoint, it is essential that the films can be prepared rapidly and under conditions permitting films of given properties to be easily reproduced.

Although many types of glass may fulfill the mechanical and chemical requirements to a satisfactory degree, no successful techniques have been developed for the preparation of sufficiently thin and uniform films. Oxide films of the lighter metals may come into use as the need for improved resistance to chemical action and high temperatures increases. The suitability of aluminum oxide films for this purpose has been demonstrated by Haas and Kehler.[70] Silica films, deposited on a structureless base by evaporation in a vacuum chamber, have been found suitable for specimen supports by Heidenreich.[74] These films will undoubtedly find extensive application in the future.

A small number of thermoplastic polymers, however, have been used almost exclusively in both electron diffraction and electron microscope studies. Thin structureless films of these materials are easily prepared by permitting the evaporation of the solvent from a dilute solution spread in a thin layer on the surface of either a liquid or solid. It has been difficult to select a polymer with optimum properties for this application on the basis of the known physical and chemical properties of these materials in bulk form. Negligible surface properties are highly significant if a material exists in the form of a film with a thickness less than the length of its molecules.

In the course of testing the suitability of a variety of polymers for this application, it has been possible to verify the application of the observations of Ott [126] to the case of these extremely thin films. The mechanical properties of films formed from cellulose nitrate samples of different viscosities improve perceptibly with increasing vis-

* See also paper by Richards, A. G., Anderson, T. F., and Hance, R. T.—Proc. Soc. Exp. Biol. and Med., 51, 148 (1942).

cosity up to approximately 20 seconds (ASTM). The most satisfactory films are obtained from a sample of uniform molecular weight. Films produced from mixtures of low and high viscosity samples have mechanical properties equivalent to the films produced from the low-viscosity component. As these observations probably apply to any polymer, they may in part account for the difficulty in finding suitable synthetic polymers for this purpose. The natural polymers have in general a more uniform distribution of molecular weights.

Although other liquids are not excluded, it is very convenient to use water as a base upon which to form films. This choice restricts the variety of solvents which may be used to those which are insoluble or only slightly soluble in water, spread readily on water, and have a lower specific gravity than water. The evaporation rate should be low enough to permit a drop of solution placed on the surface of the water to spread to a layer of uniform thickness. Cellulose nitrate dissolved in amyl acetate has been used most widely.

Although varying in details from the methods of other workers,[21a, 187] the following procedure for preparing plastic specimen supports has been found reasonably satisfactory in the laboratory of the writer. By means of a medicine dropper, a single drop of 2 per cent solution of cellulose nitrate of 20-second viscosity and containing a small proportion of plasticizer (dibutyl phthalate) is gently lowered to the surface of a shallow vessel filled to the brim with distilled water. The solution spreads rapidly to cover approximately half of the exposed water surface, whereupon the solvent evaporates, leaving a fairly uniform film of cellulose nitrate in the center of the vessel. The uniformity and cleanliness of the film depend upon the cleanliness of the water surface before the solution is deposited. By using a "Pyrex" pie plate with the flange ground flat and covered with paraffin, it is possible to sweep the surface clean with glass or chromium bars immediately before a film is to be deposited.

If the suspended area of the dried film has a diameter in excess of approximately 0.1 millimeter, it is likely to rupture when exposed to the electron beam. The film is therefore suspended over the surface of a fine screen before removal from the surface of the water. In order to reduce the contamination which is difficult to avoid in manipulating the small mesh discs normally mounted in the instrument, it is advantageous to remove the film from the water surface on a relatively large section of mesh (1 cm^2) which previously has been cleaned with dilute acid and rinsed thoroughly in distilled water. A rectangular section of mesh is immersed below the film with forceps and raised from beneath a selected film area. On removal from the vessel, excess water may be drained away with a folded edge of filter paper. When completely dry, the mesh supporting the film may be cut into smaller rectangular sections with shears. After mounting the material to be studied, suitable areas may be selected with the aid of a light microscope. Circular discs, containing the selected areas, may then be cut with a hollow ground punch without injury to the film.

With the technique developed by Schaeffer and Harker [148] it is possible to prepare satisfactory thin films of the synthetic plastic, polyvinyl formal, (Formvar, Shawinigan Products Corporation, New York City) by using mica or glass as a base from which to evaporate the solvent. This material apparently has somewhat better mechanical properties and is less susceptible to ageing. It is particularly well adapted to the preparation of replica specimens which will be described in Section VI.

As none of the plastics which have been used has adequate resistance to chemicals and organic solvents, it is preferable to suspend the material to be studied in water or to disperse it from a dried sample. Undesirable coagulation of colloidally dispersed specimen materials, which occurs as a result of the evaporation of the water in which they are initially suspended, is frequently difficult to avoid. It may be diminished by the addition of a wetting agent to the suspension. Its type and concentration must be selected with due consideration for the amount of solid material which re-

mains on the film after evaporation. It is also possible in some instances to prevent flocculation by addition of an electrolyte in an amount just sufficient to be completely absorbed on the particles. This has been found successful in the preparation of clay mineral specimens. A novel method of overcoming coarse flocculation has been suggested by von Ardenne.[15] The specimen holder is mounted on the end of a vibrating rod (tuning fork) during the drying period. Addition of a monolayer on the surface of the prepared film to make it hydrophyllic, as suggested by Schaeffer and Harker,[148] is probably a more satisfactory general procedure for the prevention of coagulation. The development of this procedure should also prove of value in the preparation of biological specimens which are suspended in normal saline solution.

Satisfactory dispersions of dry powdered materials are in general even more difficult to prepare. Several methods which have been used in the laboratory of the writer in the study of pigments yield specimens in which an appreciable proportion of the particles appear isolated. A thin plastic film is deposited on the surface of a glass slide which previously has been coated with gelatin and cooled in a refrigerator. Powdered material may then be dusted on the film and dispersed by means of a luminous discharge from a Tesla Coil. Although the film is ruptured by this discharge, satisfactory dispersions are obtained in selected areas. The method fails entirely for materials which do not adhere to the film. In the latter case, the powder is placed in a bottle which is saturated with the vapor of the plastic solvent. After the mouth of the bottle is covered with the coated side of the glass slide, the bottle is shaken vigorously. A thick deposit of powder is formed on the film, softened by the solvent vapor. On removal of the excess deposit by tapping the slide, the dried film retains a covering of partially embedded and dispersed particles.

After mounting the powder on the film by either of the above or yet other methods, the film is refloated on a warm water surface, specimen side upward, by teasing the gelatin from the glass slide. The gelatin is subsequently washed away and the film retrieved on a mesh in the usual manner.

Many materials may, of course, be dispersed in solvents containing a small proportion of cellulose nitrate or other film-forming material. Although this procedure simplifies the preparation of the specimen, the difficulty of coagulation is not in general avoided and the true dimensions of the particles may be obscured by the enveloping film.

VI. Some Applications of the Electron Microscope

(a) **In the Physical Sciences and their Technical Applications.** The optimum resolving power of the transmission electron microscope can be exploited to provide significant results from the determination of the dimensions of solid colloidal particles normally suspended in liquids or gases. From among the large number of publications in which the appearance of various types of the colloids have been illustrated, only a few examples can be selected for review. These demonstrate the usefulness of the instrument, its relationship to the established indirect methods of observation, as well as some of the difficulties which have been encountered in the interpretation of the micrographs. Perhaps the most fundamental problem in this field is the establishment of the three-dimensional geometry of small particles from the data provided in a two-dimensional image.

The determination of the geometrical form of particles which depart but little from spheres or from the circular shape in the plane of the specimen is discussed in a paper by Kausche and von Borries [90] concerning the sizes and shapes of the particles in some commercial colloidal gold preparations. The writers assume that crystallites in the colloidal range are bounded by plane faces intersecting in geometrical lines and that these would be revealed as such in the image provided by a microscope with infinite resolving power. On this basis they consider the effect of a finite resolving

power in obscuring the geometrical boundaries. Their conclusions are based on the approximation that the geometrical boundaries in the object appear in the image as boundaries of finite width, within which the image intensity falls continuously from the intensity of the background to the intensity of the particle. The effective width of the boundary is considered equal to twice the practical resolving power, δ_R, of the instrument and the physical boundary of the particle at a point midway between the particle and background intensities. In consequence of the finite resolving power, the image boundary at the intersection of two straight lines is rounded. It is assumed that the curvature of the rounded corner is circular, the arc being tangent to the intersecting edges at a distance δ_R from the point of intersection inferred by geometrical construction. In the case of a symmetrical particle (in the plane of the specimen), the rounded corners tend to obscure the crystalline shape. In accordance with the criterion of the writers, recognition of departure from circular shape is possible if the length of the straight portions of the edges in the image is equal to or greater than the length of the chord between the extremities of the arc. Fig. 15

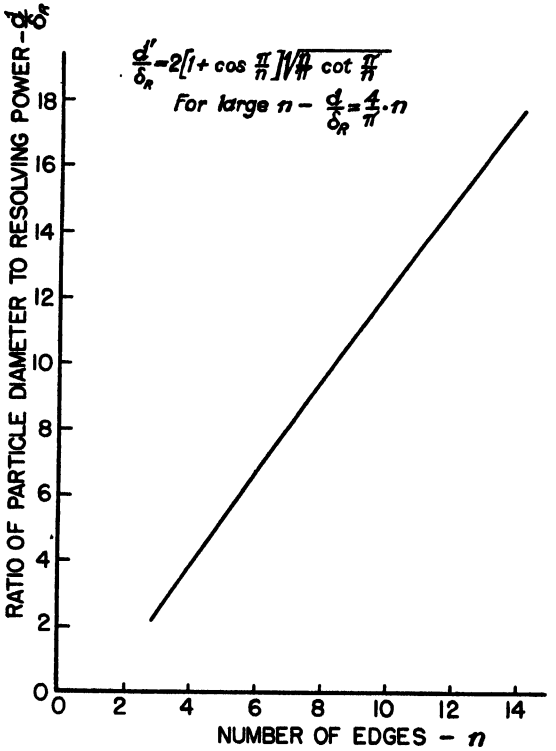


FIGURE 15. The image of a symmetrical figure with n corners or equal straight edges can be distinguished from the circular shape if the effective diameter d' of the figure is equal to or greater than the resolving power δ_R by a factor d'/δ_R.[90]

(Fig. 5[90]) illustrates this criterion for a particle with n equal edges in the plane of the specimen. The criterion is expressed as a function of the ratio d'/δ_R, where d' is the effective diameter of the particle (circle of area equal to that of the n-sided figure). From an inspection of one of their micrographs (Fig. 6[90]) it is shown that a hexagonal particle of diameter $d' = 40$ millimicrons can be recognized with a resolving power $\delta_R = 8$-10 millimicrons. The hexagonal structure is recognized with a

d'/δ_2 ratio between 4 and 5. The necessary value, according to the chart, is between 8 and 10, indicating that the criterion in this case is excessively stringent. The necessity for such a criterion is obvious. The above numerical comparison, however, indicates a further necessity for the specification of the conditions under which the particles are observed.

It is suggested by the writer that a measurement of the chords of the rounded corners of particles of known simple geometry might be used as a measure of the practical resolving power. Such linear measurements, independent of "just perceptible" contrast differences, in the micrographs of easily reproducible test specimens (*e.g.*, the freely supported magnesium oxide smoke particles [28]), would eliminate, further, the variables introduced by a supporting film and the dimensions and densities of neighboring "point" scattering centers which have been used as test objects in the past.

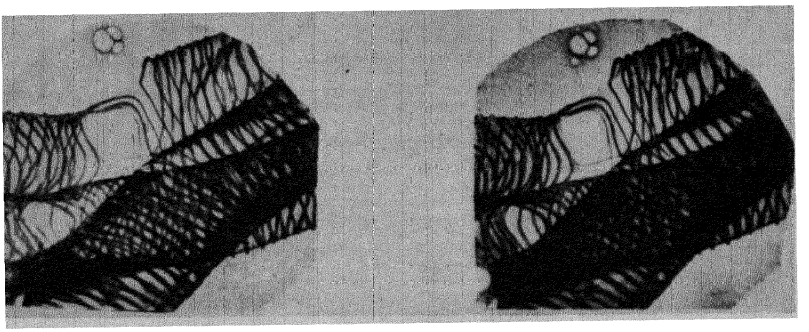


FIGURE 16a. Stereoscopic electron micrographs of a mosquito larva trachea. (Courtesy *Science News Letter*).[132]

Since the focal depth of the electron microscope is very large, it is possible to measure directly in a micrograph only the two dimensions of a structure which are parallel to the plane of the specimen. In the study of larger structures (dimensions in excess of *ca.* 1/10 micron), or of large aggregates of small particles, stereoscopic projections provide instructive pictures which reveal the three-dimensional geometry. The method is of great value in demonstrating the architecture of such structures as diatoms,[21b] and skeletal sections of insect morphology, examples of which are shown in Fig. 16.[1, 132] It also provides useful information regarding the state of aggregation of crystallites.

The value of this method for the numerical determination of the third dimensions of small colloidal particles, however, has not been proven. The success in obtaining stereoscopic projection depends critically upon the degree of accuracy with which the component micrographs can be superimposed. Slight variations in the relative disposition of the specimen particles resulting from altered conditions of illumination, as well as slight errors in the relative magnifications or orientation of the micrographs, have a serious effect on stereoscopic projection. The inherent inaccuracy of small measurements by parallax methods may be greatly enhanced by faulty projections.

One example of the inconclusive results which have been obtained by application of this method is provided in the publications of Schoon and Koch [149] and of Heering, Gizycki and Kirseck.[73] The former concluded that commercial colloidal carbon consisted of platelets of circular or elliptical shape. In the later publication of Heering *et al*, it is concluded that the shapes are spheroidal. The thickness determinations of thin crystal plates of clay minerals by this method, as reported by Eitel and Gott-

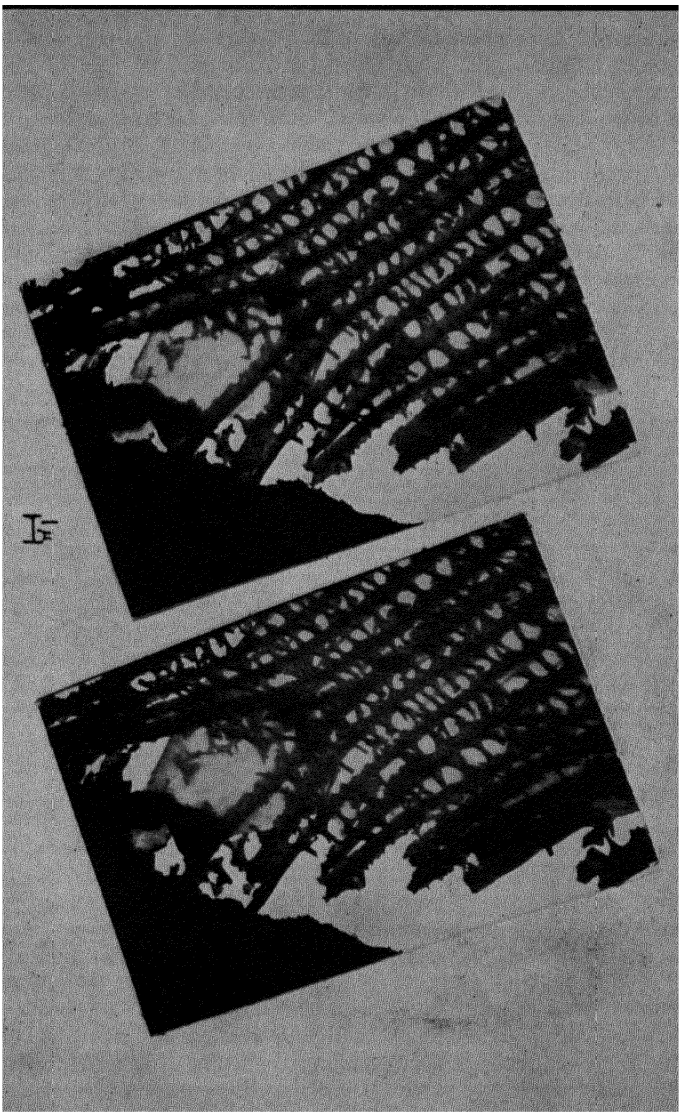


FIGURE 16b. Stereoscopic views of fractured non-iridescent wing scale of *Morpho Cypris* showing nature of supporting rods. 12° stereoscopic angle.[1]

hardt,[58] cannot be regarded as significant for yet an additional reason. The third dimension, or thickness, of the plates measured by these writers is evidently below the resolving power of the instrument. Consequently, even a direct measurement of the plate thickness would be impossible.

At the present time there appears to be only one possible direct method for the accurate determination of the third dimension. It is the simple procedure of rotating the specimen through large angles about an axis perpendicular to the optic axis of the instrument and procuring micrographs at angles of orientation for which selected particles reveal their *minimum* dimensions. Since particles tend to be oriented with

their flat sides parallel to a supporting film, this type of specimen support is unsuitable. The preferred orientation may be reduced through the use of a fine plastic screen in place of a film.

This direct method is inapplicable in the measurement of extended plates or films, the thickness of which is comparable with or below the resolving power of the instrument. In this range, the only apparent possible method at present is the indirect method of interpreting the image intensity, as described by Marton and Schiff.[116]

Studies of Clay Minerals.* Aside from the practical importance of clay minerals in soils, construction engineering, the ceramic, rubber, paper, petroleum, metal-founding and other industries, the complex structures of these hydrous aluminum silicates are of considerable interest from an academic viewpoint. Due to the efforts of research workers in many fields, a vast amount of information has been accumulated

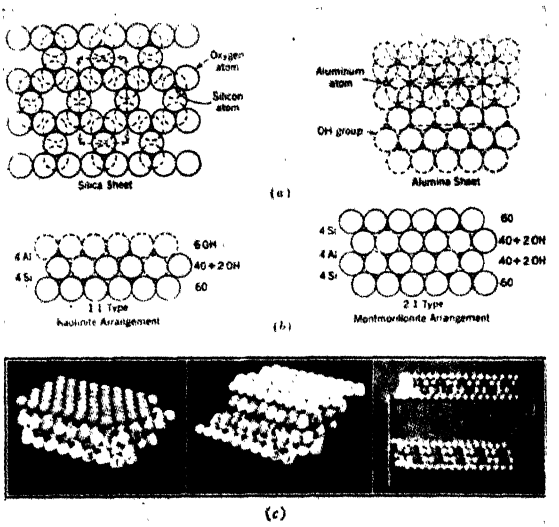


FIGURE 17. The crystal structure of the clay minerals. (a) Arrangement of atoms in the sheets; (b) arrangement of atoms in a single crystal plate: (c) in the models of kaolinite (left), montmorillonite (center) and the expanded lattice of montmorillonite (right).[81]

regarding the physical, chemical and crystalline properties of these materials. Grim [89] has thoroughly reviewed the significant findings regarding clay prior to the advent of the electron microscope, and for a full discussion, the reader is referred to his excellent paper.

According to Grim, "Two structural units are involved in the atomic lattices of the clay minerals. One is the alumina or aluminum hydroxide unit, which consists of two sheets of closely packed oxygens or hydroxyls between which aluminum atoms are embedded in such a position that they are equidistant from six oxygens or hydroxyls. Actually, only two-thirds of the possible aluminum positions are occupied in this unit, which is the gibbsite structure. The mineral brucite possesses a similar structure except that all possible aluminum positions are occupied by magnesium. The second unit consists of a sheet of tetrahedral silica (SiO_4) groups linked to form a hexagonal network of the composition Si_4O_{10} when repeated indefinitely. This unit

* A number of papers in earlier volumes of this series deal with various aspects of clays, *e.g.*, in Vol. III, by K. von Terzaghi, J. di Gleria, and Fr. Zuker, R. Bradfield. J. A.

may be viewed as a sheet of loosely packed oxygen atoms with each oxygen linked to two silicon atoms directly beneath. The silicon atoms are in tetrahedral positions, three valences being satisfied by linkage to three oxygens in the overlying sheet. The fourth silicon valency is satisfied below by an oxygen atom such that this silicon valency is analogous to the common hydroxyl group of gibbsite."

Marshall [109] classified the known clay minerals into two main lattice types. The 2:1 type is composed of an alumina unit bound between two silica units, while the 1:1 type consists of a combination of one alumina unit and one silica unit. The two lattice types are illustrated in Fig. 17. A single plate of kaolinite is shown at the left, and a single plate of montmorillonite at the right. In the third illustration of Fig. 17 two plates of montmorillonite are separated by an intervening space in which it is believed adsorbed cations are held. The spacing between plates varies with water content, as shown by Hofmann, Endell and Wilm.[85]

The properties of the 1:1 clay minerals—kaolinite, dickite, nacrite, and halloysite ($Al_2O_3 \cdot 2\ SiO_2 \cdot 2\ H_2O$)—are closely similar.[109] The unit plates of the 1:1 minerals are more firmly bound together than is the case with most 2:1 minerals, probably because of the attraction between O and (OH) layers which are adjacent.[69] These kaolin minerals exhibit only slight hydration and adsorptive properties. Base-exchange capacities are of the order of 3 to 15 milliequivalents per 100 g.

All 2:1 clay minerals are considered to have similar structural features, the 2:1 layer lattice of Hofmann, Endell and Wilm representing the type form. Marshall [109] and others treat the various minerals in this group as an isomorphous series. Certain replacements in the lattice are supposed to operate, *viz.*, aluminum replacing silicon in silica layers, as in the micas and feldspars, and magnesium replacing aluminum in alumina layers. Iron also replaces aluminum in the alumina layer. In the absence of any replacements, we have the alumino-silicate, pyrophyllite, and the analogous magnesium silicate, talc. These minerals do not exhibit variable spacings between unit plates and their base-exchange capacities are very low. In montmorillonite the chief replacement is magnesium for aluminum in octahedral coordination. This gives rise to a charged lattice, and exchangeable cations are necessary to balance the charge. In beidellite the chief replacement is aluminum for silicon in tetrahedral coordination. This replacement also requires exchangeable cations for balance. Ferric iron can be substituted for aluminum in octahedral coordination through the montmorillonite-beidellite series.[77] Possible end-members for these iron substitutions are nontronite and aluminum nontronite, which are essentially free of magnesium. Magnesium bentonite apparently owes its base-exchange capacity to a substitution of lithium for magnesium in octahedral coordination.[77]

Before the advent of the electron microscope, the shape and dimensions of clay particles had been inferred from indirect evidence. Baver,[81] in a discussion of shapes of clay particles in 1940 states, "This sheet-like structure of the crystal, which is similar to that of mica, would be expected to favor a plate-like shape of the entire particle, rather than a cubical, spherical or rod-like form." Grim [69] in 1941 stated, "At the present time there are few students of clays who would dispute the premises that clays are composed of flake-shaped particles." Hendricks' [77] estimates of particle dimensions are given in Table 1.

Table 1. Order of Particle Dimensions of the Silicate-Layer Minerals (in Centimeters)

Anauxite	10^{-1}	Antigorite	$10^{-5} - 10^{-6}$
Nacrite	10^{-1}	Micas	1.0
Dickite	$10^{-3} - 10^{-3}$	Hydrous micas	$10^{-4} - 10^{-5}$
Kaolinite	$10^{-4} - 10^{-5}$	Pyrophyllite	$1 \times 10^{-1} - 1 \times 10^{-3}$
Halloysite	10^{-5}	Talc	$10 - 10^{-3}$
Chrysotile	10×10^{-7}	Montmorillonites	$10^{-5} - 10^{-6}$

Since most of the physical and chemical properties of clays are closely related to particle shapes and the extent and nature of surfaces exposed, it is very natural that

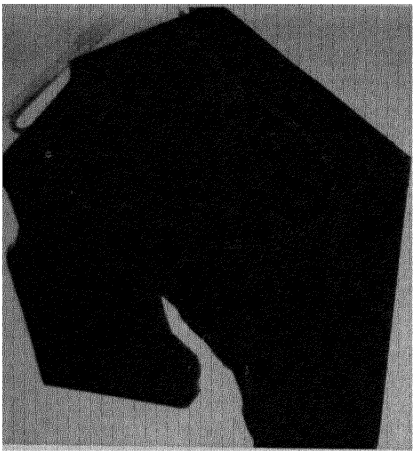


FIGURE 18a. Dickite crystal lying flat.

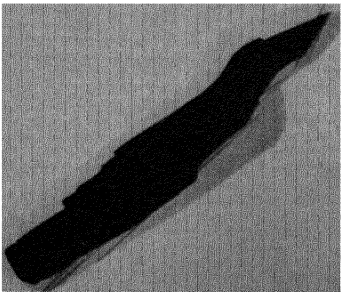


FIGURE 18b. Same dickite crystal standing approximately on edge.[87]

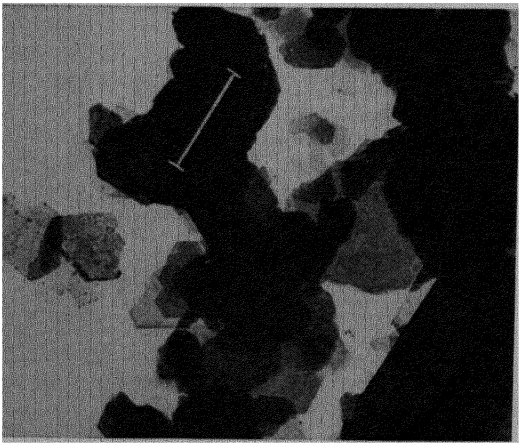


FIGURE 19a. Kaolinite.

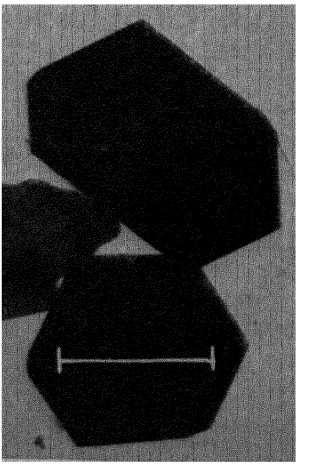


FIGURE 19b. Dickite.

clay scientists should be among the first to make practical use of the electron microscope. It was first used by the German investigators, Eitel and others [58, 59, 60, 61] and by von Ardenne, Endell and Hofmann [25] in 1939 and 1940. B. T. Shaw and his students at Ohio State University began their studies [86, 87, 98, 110, 151, 152] shortly afterward. The published work concerns two main objectives; (1) to characterize the various clay minerals by their appearance, and (2) to evaluate the extent and nature of the surface exposed by various size ranges of the several minerals.

The first *direct* evidence bearing on the plate-shaped nature of clay was provided by Humbert and Shaw [87] with successive exposures of the same crystal of dickite, first with the *c*-axis parallel to the optic axis of the instrument, Fig. 18a; and then with the *c*-axis approximately perpendicular to the optic axis, Fig. 18b. Since the crystal was vibrating during the second exposure it is not possible to give an accurate axial ratio. It is of the order of 10:1.

Three minerals of the 1:1 lattice type are shown in Fig. 19: (a) kaolinite,[151] (b) dickite [152] and (c) halloysite.[152] These electron micrographs are characteristic

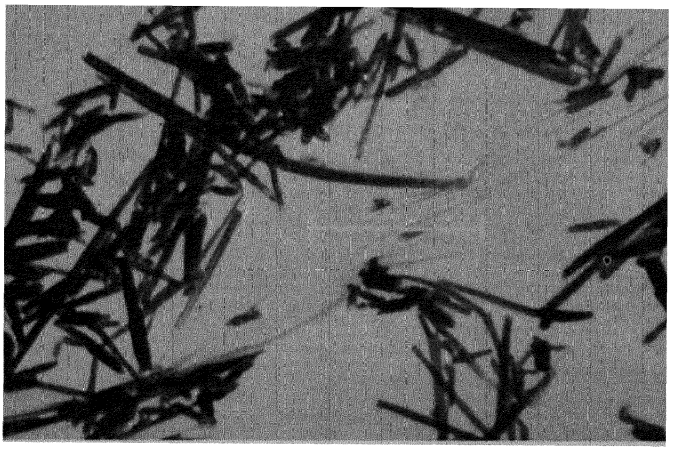


FIGURE 19c. Halloysite.[150,152]

of a large number of observations of minerals from various sources. Kaolinite particles are seen to be composed of an orderly arrangement of thin, well-defined crystals. The 120° interfacial angle is evident. Shaw estimates the thickness of the thinnest crystals to be less than 5 mμ. This would give axial ratios of the order of 100:1. The individual crystals of dickite show a much greater development along the *c*-axis than is the case with kaolinite. The axial ratios are of the order of 10:1 to 15:1.[98] The halloysite structure was quite surprising. Hendricks [77] stated in 1941, "There is no reason at the present time for considering halloysite ($Al_2O_3 \cdot 2\ SiO_2 \cdot 2\ H_2O$) to have types of layers different from kaolinite, nacrite and dickite. New evidence, however, might be found that would lead to a more complete understanding of its structure. The powder-diffraction pattern of halloysite differs somewhat from kaolinite, but in a manner that can be accounted for by its degree of organization." The marked differences in crystal habit of these three minerals is far greater than has been suspected on the basis of other evidence. They do have one feature in common. All show well-defined, sharp, straight-edged crystals.

Shaw [151] made some interesting observations regarding the exchange capacity of kaolinite. The basic material for his study was unground kaolinite (Fig. 20a) having an exchange capacity of 3.8 me/100 g. Samples of this material were ground in a ball mill in the dry state and also as a thin paste in water. The sample wet-ground

for 144 hours (Fig. 20b) had an exchange capacity of 16.6 me/100 g. The sample ground 144 hours in an air dry state (Fig. 20c) had an exchange capacity of 49.5 me/100 g. "It is seen that the wet grinding caused the particles to separate along cleavage planes. Also there has been some breaking of crystals as is evidenced by

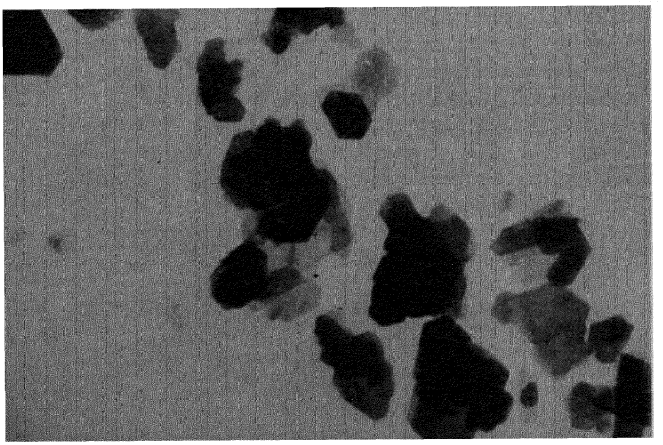


FIGURE 20a. $<0.5\mu$ Fraction of kaolinite (unground).

the absence of the well defined crystal edges shown in the unground material. There is little evidence of cleavage in the dry ground material. The relatively thick irregular fragments have obviously resulted from crystal fracture. The much greater increase in exchange capacity is fairly definite evidence that the exchange positions on kaolinite are located at plate edges."

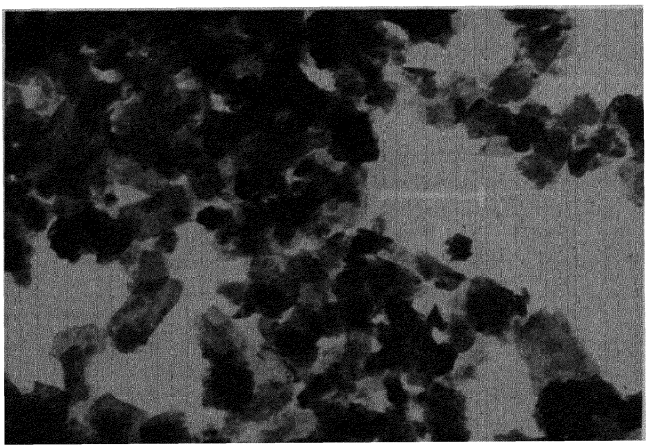


FIGURE 20b. $<0.25\mu$ Fraction of kaolinite (wet-ground).

The existence of a difference in the adsorptive properties of cleavage and fractured surfaces was suggested as the cause of anomalies in the study of clay gel viscosities by Ford, Loomis and Fidiam.[64] The observations of Shaw thus provide the first direct experimental evidence in this type of problem which is characteristic in the specification of the colloidal properties of a great many materials.

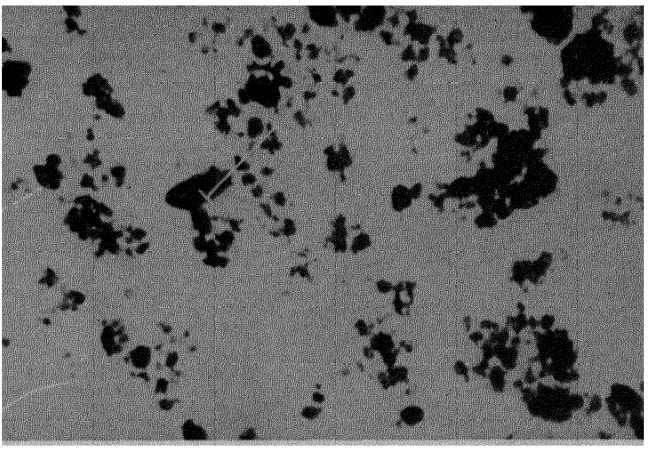


FIGURE 20c. <0.25μ Fraction of kaolinite (dry-ground).[151]

Four minerals of the 2:1 lattice type are shown in Fig. 21. These are: (a) montmorillonite from California;[87] (b) montmorillonite from Wyoming bentonite;[151] (c) beidellite from Putnam Clay (100-50 mμ, Equivalent Spherical Diameter, abbreviated to ESD in the following);[110] (d) beidellite from Putnam clay (< 20 mμ ESD);[110] (e) nontronite (200-50 mμ ESD);[110] (f) nontronite (< 50 mμ ESD);[110] (g) magnesium bentonite (500-200 mμ ESD);[110] (h) magnesium bentonite (< 50 mμ ESD).[110] If these four minerals belong to the same isomorphic series, as has been suggested, then the ismorphism is restricted to fine structures, indeed.

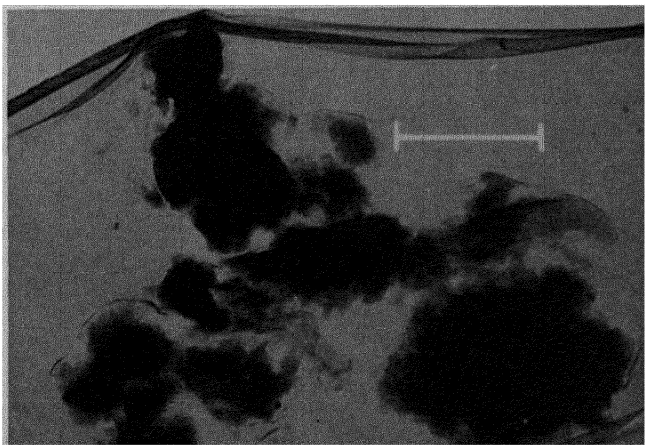


FIGURE 21a. California montmorillonite.

The crystal habit of montmorillonite is well shown in Fig. 21b. The larger particles are seen to be composed of extremely thin plates. Many of them exhibit a fair degree of orientation of plates. This material was fractionated by the double-layer technique of Marshall,[108] so that the sedimented particles should be in the size range from 2 to 1μ ESD. The fine plates evident in the micrograph were dispersed, in the gentle treatment of washing the material from the centrifuge tube, from the

larger particles characteristic of the settling velocity. This shows how loosely the unit plates of montmorillonite are bound together. The finest plates in the micrograph are very probably of unit thickness, (1 mμ) as will be discussed later. Ultimate dispersion would separate all particles into plates of unit thickness. Axial ratios for the unit plates are of the order of 250:1. Aggregates of overlapping

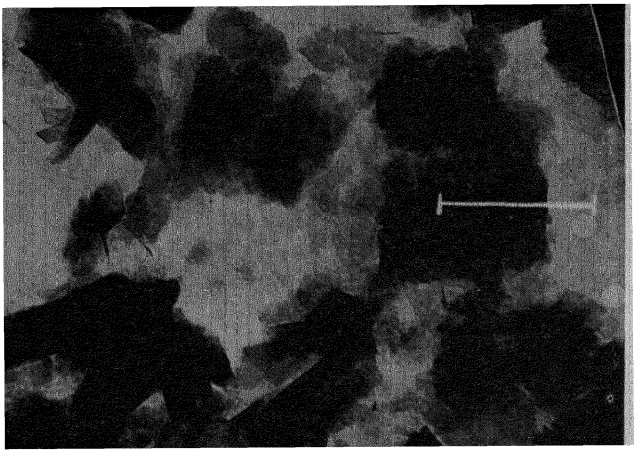


FIGURE 21b. 2 — 1μ Fraction of Wyoming bentonite.

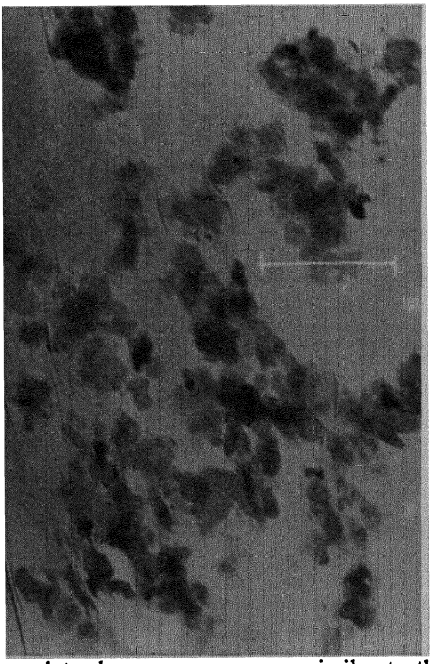


FIGURE 21c. 100 — 50 mμ Fraction of beidellite.

plates have an appearance similar to the particles of montmorillonite shown in Fig. 21a. This suggests that possibly the real difference between the two montmorillonites shown may be only one of plate area. The particles of the California montmorillonite appear to be relatively thick aggregates of small plates, whereas the particles

of Wyoming bentonite are composed of relatively large plates in parallel orientation. In either case the thickness of plates is probably of the same order of magnitude. The vast amount of exposed surface per unit weight of material is evident.

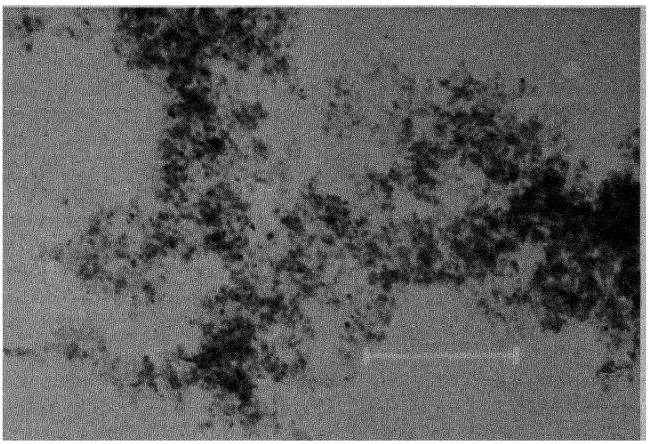


FIGURE 21d. <20 mμ Fraction of beidellite.

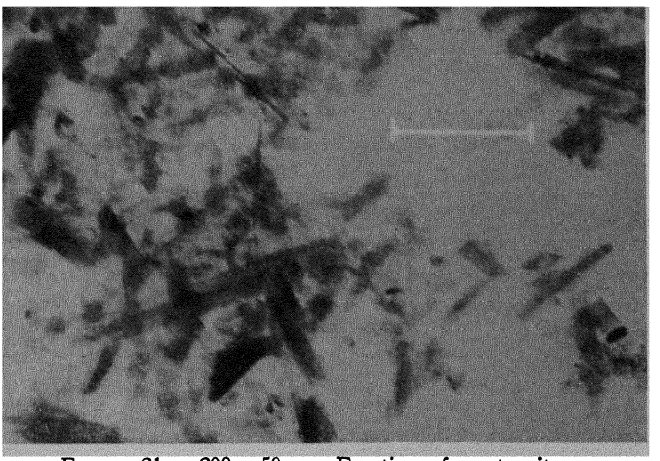


FIGURE 21e. 200 — 50 mμ Fraction of nontronite.

The beidellite shown in Figs. 21c and 21d, is also plate-shaped. In this particular clay there is no evidence that particles of smaller settling velocity are produced by micaceous cleavage alone from those of larger size. Neither do particles in the fraction 100-50 mμ show a composite character such as might be expected if they were built up from units in the 20-5 mμ range.[110] The substitution of aluminum for silicon in the crystal lattice has not created sufficient strain to inhibit the indefinite extension of the crystal in two directions so that beidellite is plate-shaped. The crystal plates are more firmly bound together in beidellite than they are in montmorillonite. This substantiates the views of Marshall.[109]

The nontronite shown in Figs. 21e and f has the shape of thin blades or laths. Evidently the substitution of ferric iron for aluminum distorts the lattice to limit

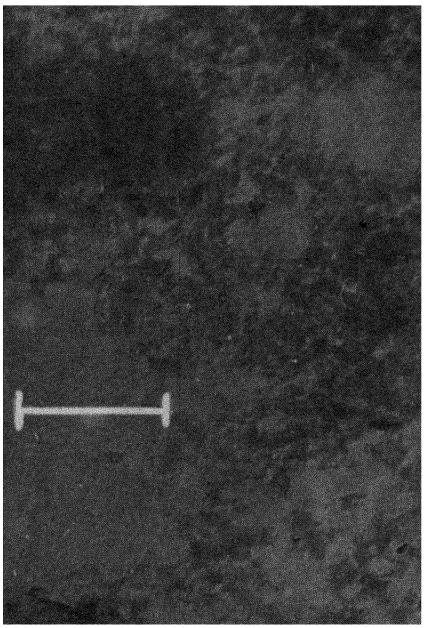


FIGURE 21f. <50 mμ Fraction of nontronite.

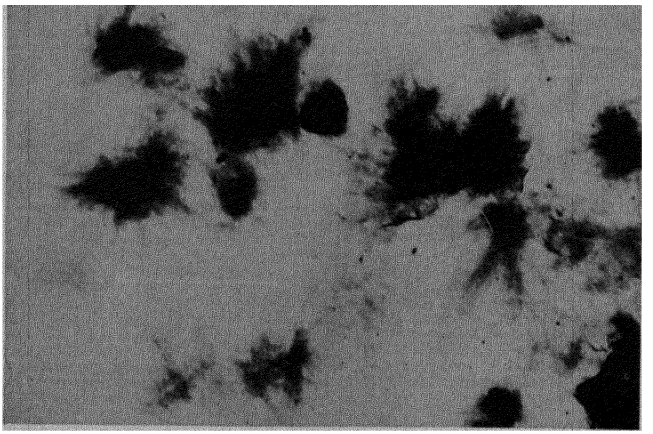


FIGURE 21g. 500 – 200 mμ Fraction of magnesium bentonite.

the width of crystals. This same phenomenon has been observed [77] when magnesium replaces aluminum in the kaolinite structure, thereby giving rise to chrysotile asbestos, Fig. 22. It is also shown when magnesium replaces aluminum in the montmorillonite structure, giving rise to magnesium bentonite, Figs. 21g and h. The magnesium substitution appears to give rise to a greater distortion than the iron substitution, since the crystal widths are much more limited in the case of magnesium.

The use of the electron microscope has thrown considerable light on the meaning of the mechanical analysis of clays. Kelley and Shaw [98] have studied sedimentation

analyses in relation to particle shapes. They found that a particle size distribution of many clays, notably montmorillonite, merely reflects the degree of dispersion. This is well shown in Fig. 23: (a) Wyoming bentonite (2-1μ ESD); (b) Wyoming bentonite (0.5-0.2μ ESD); (c) Wyoming bentonite (< 50 mμ ESD). These fractions were prepared from the same material as that shown in Fig. 21b, and in the same manner, except that the fractions shown in Fig. 23 were redispersed and resettled four times to assure sharp boundaries. A comparison of the four micrographs justifies the conclusion that the larger particles of this montmorillonite are aggregates of the thin crystal plates shown in Fig. 23c. By making use of the electron miscroscope it was possible to measure two of the dimensions of the particles. The third dimension could then be calculated from settling velocity considerations. In the case of the smallest fraction of montmorillonite, shown in Fig. 23c, the lower limit of settling velocity is not defined, so it is not possible to make mean thickness calculations. If one assigns the settling velocity of the upper limit to particles of greatest area, an estimate of maximum thickness of these particles can be made.

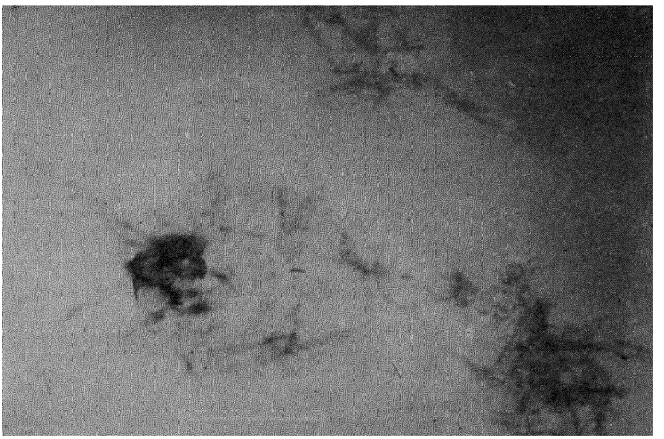


FIGURE 21h. <50 mμ Fraction of magnesium bentonite.[151, 152, 110]



FIGURE 22. Chrysotile asbestos.[110]

Such a calculation gives an average maximum thickness of 1.9×10^{-7} cm for ten of the larger particles.[151] A unit plate of montmorillonite is about 1.0×10^{-7} cm thick. It is highly probable that this fraction is composed largely of crystals of unit thickness. Shaw[151] explains many of the properties of montmorillonite that had been attributed to an expanding lattice as a consequence of the fact that an ultimate dispersion of this material results in plates of unit thickness.

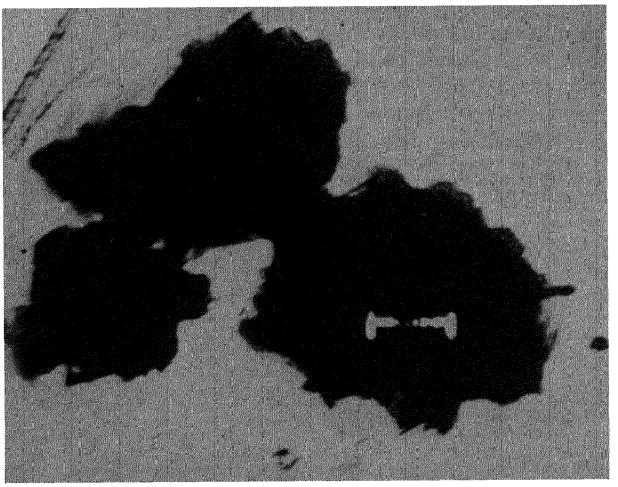


FIGURE 23a. $2 - 1\mu$ Fraction of Wyoming bentonite.

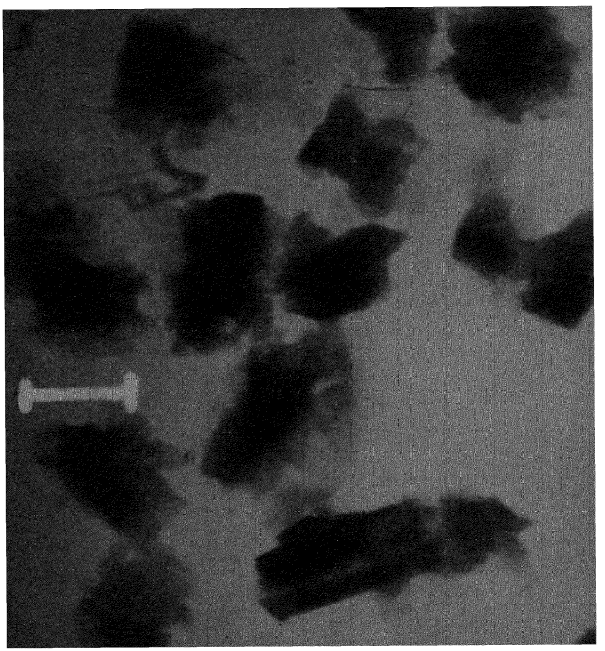


FIGURE 23b. $0.5 - 0.2\mu$ Fraction of Wyoming bentonite.

A summary of Kelley and Shaw's studies with plate-shaped particles is given in Table 2. Their studies of halloysite, a rod-shaped mineral shown in Fig 19c, led them to the conclusion that the fractionation of rod-shaped particles on the basis of settling velocity is not accomplished with the same degree of accuracy as can be obtained with plate-shaped particles. The fact that particles of the same size occur in all fractions is shown in Table 3.

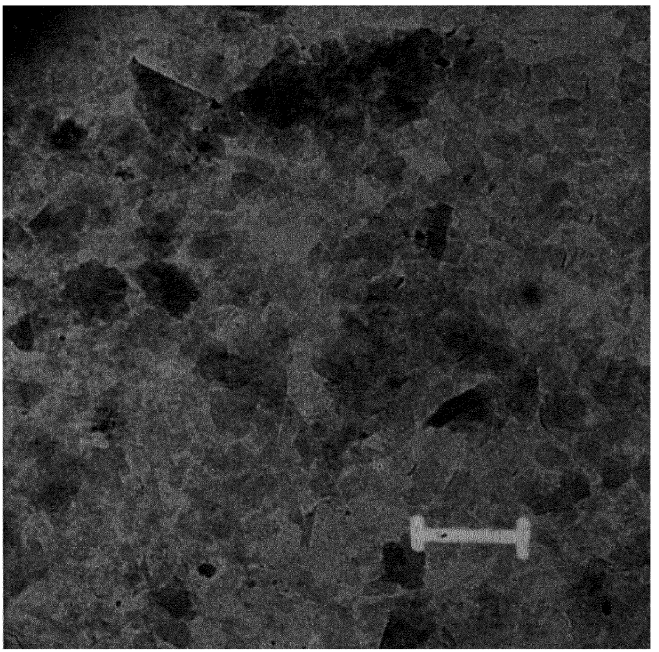


FIGURE 23c. <50 mμ Fraction of Wyoming bentonite.[151]

Table 2. The Mean Thickness, the Mean Axial Ratios and the Mean Area of Disk-Shaped Particles of Several Clay Minerals

Mineral	Equivalent Sph. Diameters (microns)	No. of Particles Measured	Mean Area $\times 10^{-8}$ cm²	Mean Thickness $\times 10^{-4}$ cm	Mean Axial Ratio a:b	Mean Settling Velocity $\times 10^{-4}$ cm/sec
Dickite	5 – 2	12	46.8	0.735	1:10	12.1
"	2 – 1	25	11.9	.264	1:14.8	2.22
Kaolin	2 – 1	26	5.58	.397	1: 6.7	2.22
Quartz	5 – 2	15	17.2	1.30	1: 3.5	12.4
"	2 – 1	21	2.79	.616	1: 3.1	2.28
"	0.5 – 0.2	38	.436	.076	1:10	0.124
Bentonite	5 – 2	28	20.7	1.18	1: 4.4	10.9
"	2 – 1	28	5.88	.387	1: 7	2.00
"	0.5 – 0.2	29	1.69	.0368	1:40	0.109
"	0.2 – 0.05	30	0.296	.0113	1:54	0.0139

Table 3. Mechanical Analysis of Halloysite

Equivalent Sph. Diameters (microns)	No. of Particles Measured	Average Width $\times 10^{-4}$ cm	Average Area $\times 10^{-8}$ cm²
2.2 – 1.1	26	0.271	1.52
1.1 – 0.56	19	0.273	1.20
0.56 – 0.22	21	0.205	1.12

The lack of fractionation of fibers on the basis of settling velocity is also shown by the work of Marshall and others[110] with the mineral attapulgite. The fraction 200-50 mμ ESD is shown in Fig. 24a and the fraction < 50 mμ ESD is shown in Fig. 24b. The cubical mineral grains are impurities in the attapulgite. It is evident that the impurity has been fractionated, whereas the dimensions of the attapulgite fibers are fairly constant for the two fractions.

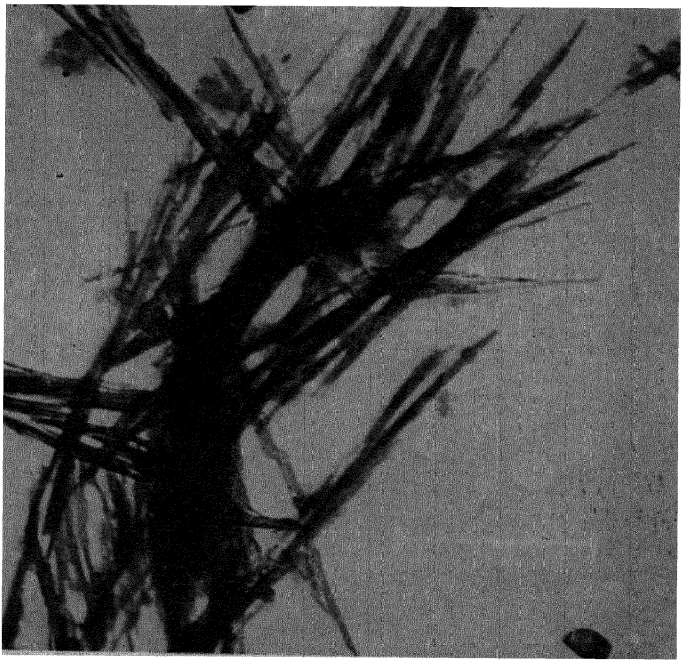


FIGURE 24a. 200 – 50 mμ Fraction of attapulgite.

The study of clays with the electron microscope has barely begun. Research in progress on the perplexing problem of aggregation in soils promises to be fully as enlightening as any studies so far completed. Considerable information is being gained regarding the changes that take place in clays subjected to such standard treatments as grinding, heating, electrodialysis, oxidation and reduction.

Studies of Colloidal Carbon. A comprehensive review of the significant properties of a material of great industrial importance, colloidal carbon, is contained in a publication of the Columbian Carbon Company Research Laboratories.[54] "Over two years ago, in Volume II of this series, the size and shape of the ultimate particles of carbon black (Micronex) were determined by the aid of the electron microscope. The surface area was shown to be more than double that previously assigned. Particle shape as an explanation for carbon differences was eliminated. This marked a major turning point in the theory and practice of reinforcement.

"With this new and refined instrument it is now possible to attempt the quantitative evaluation of surface area as a clue to rubber compounding and other chemical and physical properties. In the following studies the carbon spectrum is analyzed from this point of view. Surface area is shown to be the dominating influence, subject, however, to important and sometimes controlling anomalies. These are shown to be largely resolved on the assumption that carbon structure, in addition to surface area, plays an important role.

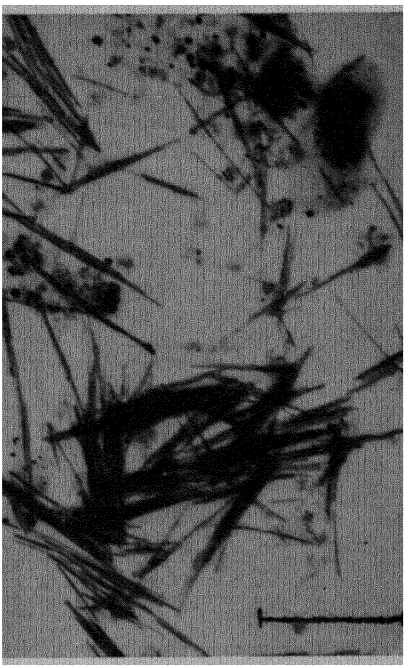


FIGURE 24b. <50 mμ Fraction of attapulgite.[110]

"During the past six months the above methods of analysis have been applied to the study of carbon reinforcement in synthetic rubbers. The role played by surface area has been found to be even more essential than in the case of natural rubber. The importance of carbon structure has likewise been confirmed.

"It is hoped that this new approach may serve a useful purpose, not only in the development of a general theory of reinforcement, but also in the most effective application of colloidal carbons to synthetic rubber compounding, now perhaps Research Problem No. 1 in this country." [54]

An interesting feature of the investigations reported in this brochure is the comparison of the particle-size determinations by several of the established indirect methods with the direct measurements in electron micrographs. This comparison is summarized in Table 4, in which the numerical values are the results of independent investigations extending over a period of several years. Exact references to these are included in the above publication.

The appearance of two of the more commonly used carbons is illustrated in Figs. 25 and 26 (Figs. 14, 16, and 8 [54]). Although the spheroidal shape of the particles cannot be established directly from the observations in these micrographs alone, the assumption is probably justified on the evidence that particles showing marked dissymetry are not to be found in any of numerous micrographs of specimens prepared by different procedures. If plate-shaped particles predominated, there should be an appreciable probability of finding some of these oriented otherwise than parallel to the specimen plane, especially about the aggregates evident in the micrographs.

In accordance with the criterion illustrated in Fig. 15, crystal habit of the carbon particles should be recognizable in the micrograph of Micronex, Fig. 25a, if the number of corners or straight edges in the images of the particles is less than 6 (mean particle diameter $d' = 28$ mμ, estimated $\delta_R = 5$ mμ, $d'/\delta_R = 5 - 6$). The experimental test of this criterion, discussed in Section VI(a), paragraph 2, indicates that crystal habit should be recognized if the number of straight edges is

less than 10 — 12. On this basis, it may be inferred that the greater proportion of the Micronex particles have more than 10 — 12 edges in the plane of the specimen, and therefore a poorly developed crystal habit. A similar and more definite conclusion may be drawn regarding the crystal development of the greater proportion of the particles of P-33, Fig. 26.

A second illustrative feature in these micrographs is the demonstration of the effect of specimen thickness on the resolving power exhibited in the micrographs [ref., paragraph 1 of Section V and Section III(c)]. With a particle thickness of 28 mμ and a beam potential of 44 kv, the existence of individual particles in clusters of small numbers is clearly revealed by the resolution of overlapping boundaries. In the micrograph of the larger particles, Fig. 26, taken with a beam potential of 54 kv, the complete contours of individual particles in clusters is entirely obscured.

Table 4. Particle Size Determinations of Several Industrial Carbons. Mean Diameters in Millimicrons

Carbon	Microscopic	Ultra-microscopic	Sedimentation	Adhesion Tension	X-Ray Diffraction	Nitrogen Adsorption	Electron Microscope
Thermax	1000	1000 1120		200-305	1120	430 490	240 (a) 374 (b) 274 (c)
Lampblack	100 300-400 300-600	1000		204-316		114	100-260 (a) 97 (c)
P-33	230	159		58-89	160	213 151	70 (a) 171 (b) 74 (c)
Acetylene		130				52 50	25-130-235 (a) 43 (c)
Micronex	80 100 100-200 150 80	124 106 83 15-200 50-60 50-200 50 61	25-35	45-68	64	31 29 30 32 28 23.6	30 (a) 39 (b) 28 (c)
Super Spectra	25	25		20-30	25	23 9.7 8 3.3	13 (c)

(a) Analysis by [54] of the distribution curves of Schoon and Koch [149] who assumed that the particles were plate-shaped.
(b) Measurement of the diameters of a few particles in the micrographs of Heering, *et al.*[73] by [54].
(c) Determined from distribution curves of large numbers of particles in each carbon by the Columbian Carbon Company Research Laboratories.[54]

The second micrograph of Micronex, Fig. 25b, is reproduced here for the purpose of illustrating the effect on image definition of the method of supporting the particles. The specimen of Fig. 25a was prepared by dispersing a sample over the surface of a uniform cellulose nitrate film. The specimen of Fig. 25b was dispersed in a cellulose nitrate solution which was subsequently cast into a film. The reduced contrast and resolving power may be ascribed to the enveloping film about the particles.

The relationship of carbon surface area to the macroscopic properties of suspensions is made evident in the charts, illustrated in this brochure. The electron microscope studies have provided the first definite evidence of a difference in the properties of colloidal carbons of equal surface areas but prepared under different conditions. Whether or not it will be possible to trace these differences by means of the electron

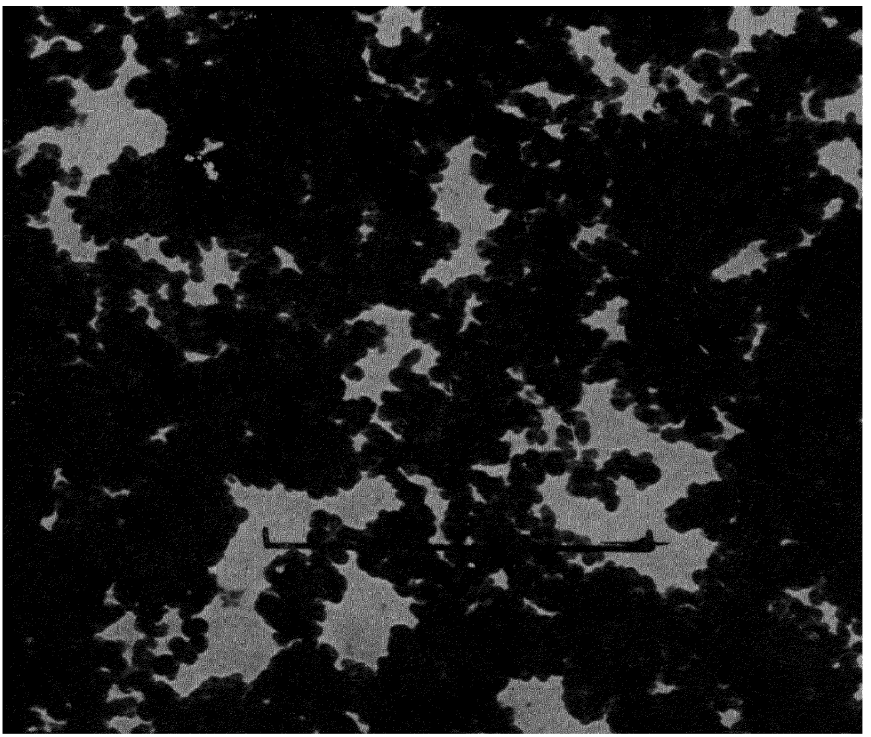


FIGURE 25a. Dry Micronex carbon dispersed on a film.

microscope cannot be predicted. That differentiation of surface properties in terms of crystal faces is possible is illustrated by the observations of Shaw.[151] The "structure" differences appearing in the carbons may be of two kinds, *viz.*, (a) the occurrence in suspensions of complexes composed of ultimate particle or particle-molecular aggregates; (b) individual carbon particles, composed of aggregates of small graphitic layers[86] may have states of organization wherein certain crystal faces are exposed in varying degree.

Photographic Emulsions. An interesting example which might be included in this section to illustrate the discovery of entirely unexpected particle shapes, is provided by the studies of photographic emulsions which have been carried out in the Research Laboratories of the Eastman Kodak Company. In a preliminary publication by Hall and Schoen,[72] it is shown that developed silver bromide crystals are not "coke-like" structures as previously had been inferred, but consist of a mass of tangled filaments. Lippman emulsions, containing the smallest photographic crystals, demonstrate this very clearly. Undeveloped silver bromide crystals are shown in Fig. 27a. When developed in common chemical developers, each small crystal grows into a single filament which is much longer and thinner than the original

crystal. The resulting worm-like structure of a developed Lippman emulsion is shown in Fig. 27b.

Cellulose Fibers. Electron microscope investigations which have been reported concerning the structure of textile fibers have had as their primary objective the exploration of the utility of the instrument in this field. In view of the several theories which have been postulated regarding the organization of the primary-valence molecular chains of cellulose into microscopic fibers, it is of some interest to review the published electron microscope observations.

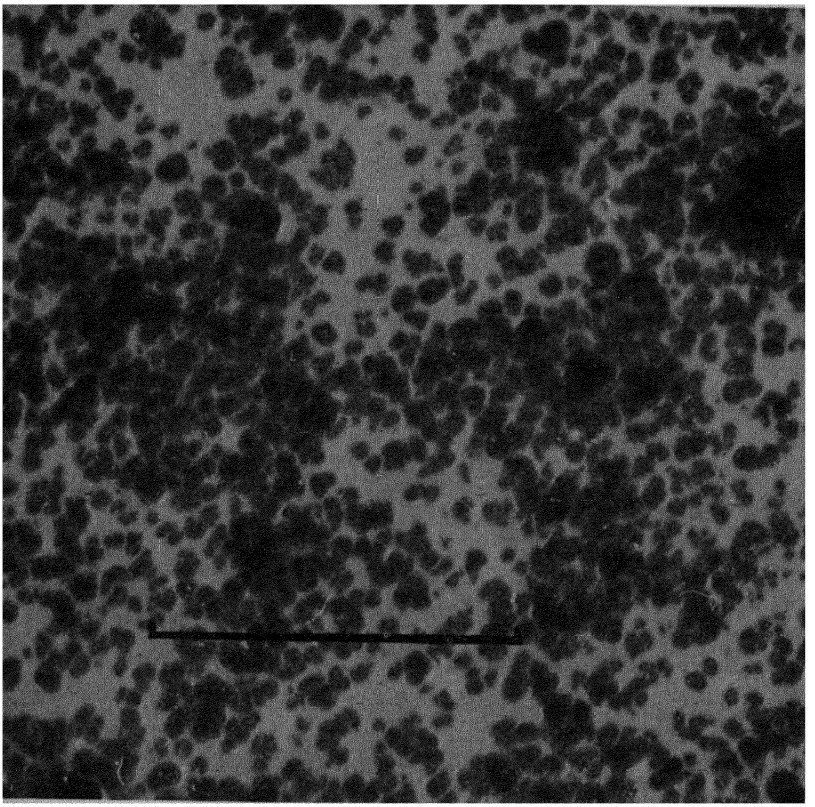


FIGURE 25b. Micronex carbon dispersed in cellulose nitrate solution. Mean particle diameter, 28 mμ. (Magnification 50,000 diameters).[84] (Courtesy Columbian Carbon Co.)

In the earliest publication, by Ruska,[188] micrographs of *cut sections* of a variety of textile fibers are illustrated. Ruska presupposed that crystalline micelles could be revealed in the micrographs of transverse sections, and would appear as a system of adjacent particles of circular cross-section with interspaces bounded by convex surfaces. His failure to observe such configurations in micrographs of specimens of *ca.* 1 micron thickness led him to unjustified conclusions. In his micrographs, the marrow of a cotton fiber appears as a network with a system of holes bounded by concave surfaces. Both the shapes of these holes and their dimensions (of the order of 1 micron in diameter) are regarded by Ruska as evidence to vitiate the existence of crystalline micelles as significant structural units. The conclusions drawn from these micrographs are not considered to be beyond question. Cut sec-

tions of organic materials of the thickness studied by Ruska are likely to suffer appreciable structural changes as a result of electron energy absorption. The observations of Barnes and Burton [29] verify this statement. This could completely obliterate the natural structure in the interior of such a specimen.

In a second publication, Ruska and Kretschmer [140] reported the observation of the products of degradation from cotton fibers treated with hydrochloric acid and cuprammonium. Electron microscope specimens were prepared by three procedures,

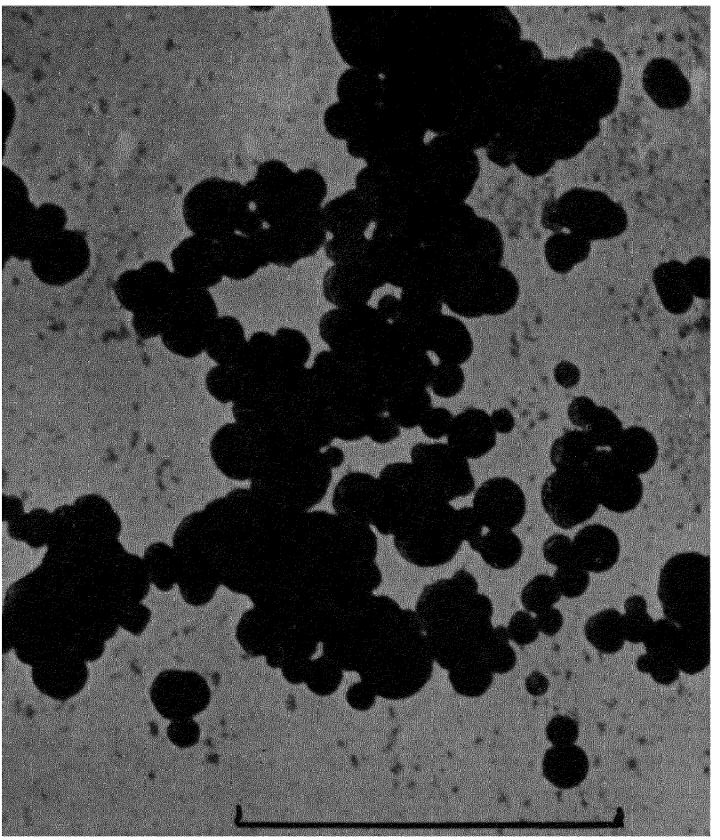


FIGURE 26. P-33 Carbon. Mean particle diameter, 74 mμ. (Magnification 50,000 diameters).[54] (Courtesy Columbian Carbon Co.)

vis., (a) centrifuging samples treated with HCl for various periods, dispersing the sediment in distilled water and drying the suspension on a specimen holder; (b) neutralizing the HCl, dialyzing the solution, and drying the products on a specimen holder; and (c) drying a dilute HCl suspension on a specimen holder. The specimens prepared by the procedures (b) and (c) exhibited a fine structure resembling a spongy film or the marrow of the untreated sectioned fibers. More clearly defined fibrous structures were observed in the specimens which had been centrifuged. In these specimens, the characteristic structures consisted of long fibers, each between 10 and 100 millimicrons in width and between 0.5 and 1 micron in length. Particularly interesting were the fibers revealed after a nineteen-hour acid treatment. These showed a width of *ca.* 5 millimicrons, corresponding to the x-ray measure-

ments of the cross-section of the cellulose crystallite. Ruska and Kretschmer express the opinion, however, that the source of these fibers is in the bark of the microscopic fibers.

The investigations of Sears and Kregel [150] on fine fractions of paper making fibers reveal the prominent existence of fine fibrous components. "The most interesting observation up to this time is the existence of fine fibrils in the fibers and fiber fragments of all stocks that have been examined." [150] Although the dimensions of these fibers, as reported by Sears and Kregel, do not indicate that fibers of the cross-sectional dimensions of crystallites are isolated, this investigation does contradict the conclusion of Ruska that the origin of the fine fibers is in the bark or skin substance alone.

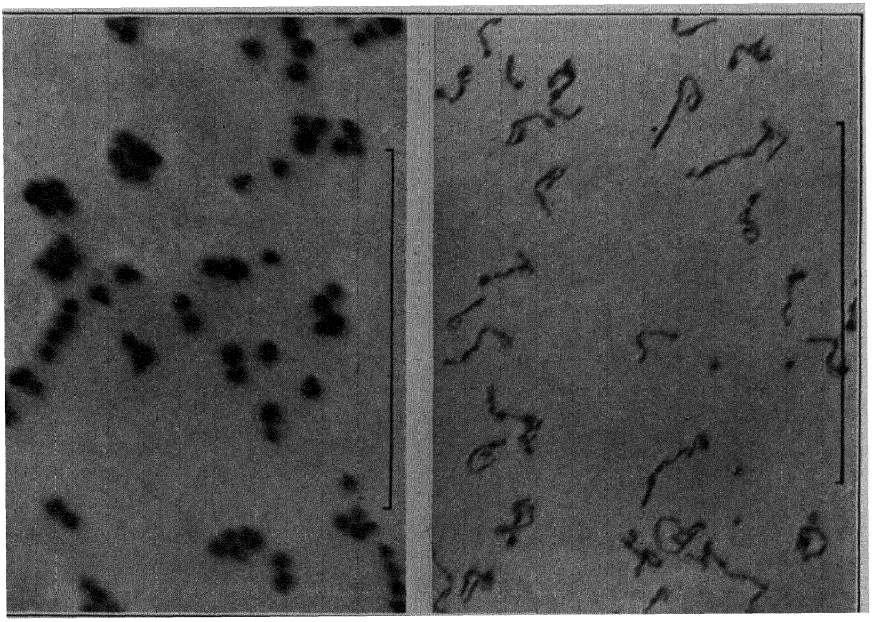


FIGURE 27. (a) (left) Undeveloped silver bromide crystals from Lippman emulsion; (b) (right) crystals developed in a common chemical developer.[72] (Courtesy Eastman Kodak Co., Rochester, N. Y.)

Barnes and Burton [80] examined specimens of both growing and mature cotton fibers disintegrated by mechanical shredding. Although two types of structures are revealed in their micrographs, *viz.*, very long fibers and a few rectangular-shaped particles, the greater proportion of the disintegrated material is fibrous in character. A micrograph of 15-day old cotton fibers is reproduced in Fig. 28. Structures in similar specimens of mature fibers are sufficiently similar to be indistinguishable from those observed in the specimen of growing fiber.

Inspection of the finer fibers revealed in all of the micrographs which have been published shows that the widths of the fibers vary continuously from below *ca.* 5 millimicrons up to several microns. Clear evidence of the existence of an ***ultimate unit*** fiber width is lacking. A number of fibers may be found in these micrographs which have a tapered shape, with the width of the marrow ends approaching the limit of resolution of the electron microscope. If the finer fibers consisted of crystallites, 60 millimicrons in length and cemented together end-to-end with a pectinous material, some periodic irregularity in the apparent width or density of the

fibers should be observable. Absence of observable density variations along the length of the smallest fibers, together with the continuous range of widths of the disintegrated fibers, might be considered as data compatible with the model of fiber structure postulated by Hengstenberg and Mark.[107]

The occurrence of a relatively small proportion of material in the form of rec-

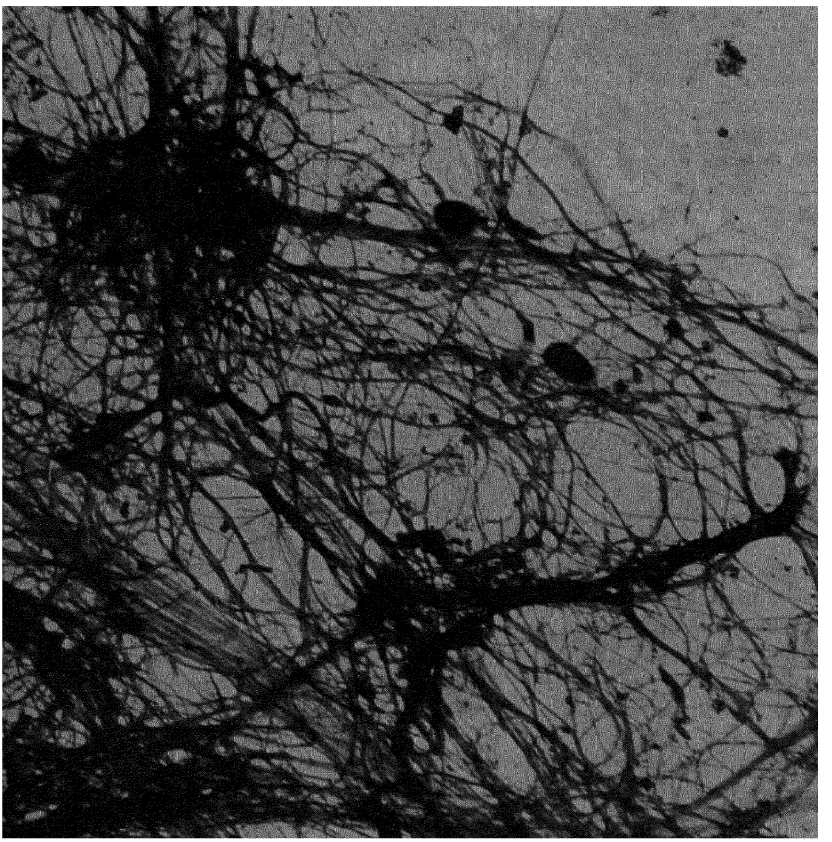


FIGURE 28. Fifteen-day cotton fiber (X 8000) partially disintegrated. Two phases can be seen—a very dense material which is almost completely opaque to electrons, and an extremely fibrous material.[30] (Courtesy American Cyanamid Co.)

tangular particles, such as those shown in Fig. 28 is of some interest. Although the dimensions of these particles are comparable with the structural units postulated by Farr and Eckerson,[65, 62] common identity does not appear justified because of their relatively small number. Other types of particles have been observed by Barnes and Burton, however, in the material leached out of the surface of 15-day cotton fibers. The dimensions of these particles are more nearly compatible with the elliptical units observed by Farr under the light microscope.

Surface Studies. The direct examination of the surface features of specimens by means of reflected electrons has met with little success. As a result of the large velocity distribution in a reflected beam, the resolving power is greatly reduced.

Ruska and Müller[185] investigated the possibilities of this method by mounting a

specimen with its plane tilted about 5° from the perpendicular to the optic axis of the instrument and illuminating it with a beam at 90° to this axis. The low resolving power and low image intensity obtained in this way limits the useful magnification of the image to that which can be obtained with a low resolving power light microscope. Other advantages of using the electron microscope for this purpose are not apparent.

Somewhat greater resolving power and considerably more image intensity are gained in the procedure of von Borries.[40] The surface plane of the specimen is nearly parallel to the optic axis and is illuminated with an electron beam at grazing incidence. The resulting image distortion, however, makes the method impractical.

"An image, no matter how produced, may be regarded as subdivided into individual image elements, each of which bears a certain definite relationship to an elementary area of a similarly subdivided object. If the geometrical arrangement of the corresponding elements in image and object is similar and if there is a regular relationship between certain characteristics of the image elements, for example, brightness or color, and the same or other characteristics of the object elements, such as transparency, reflectivity, or mass density, the image is regarded as faithful.

"The formation of an image may result either from the simultaneous imaging of all the elements, as in the projection of a picture on a screen, a photographic plate, or retina of the human eye by optical lenses, or by the successive recording of individual image elements, a process familiar today from the practice of television." [161]

von Ardenne[4, 6] was the first to attempt the exploitation of the principles in the above statement to overcome the limiting effects of chromatic aberration on the resolving power in the examination of thick specimens. In his scanning microscope, the specimen was placed in the principal focal plane of a strong magnetic lens which served as the second stage in the reduction of an electron source to a fine probe. Such a probe, incident on a specimen at any point, is scattered or absorbed in accordance with the discussion in Section III (a). By placing an aperture-limiting diaphragm between the specimen and the sensitive surface of a recording mechanism, the number of electrons reaching an image element is equal to the number which would pass through the objective lens of a transmission microscope as illustrated in Fig. 8. Thus the contrast in the images of this scanning microscope is determined by the factors governing the contrast in transmission microscope images obtained with an objective lens provided with an aperture-limiting diaphragm. Since the intensity is governed by the number of transmitted electrons, independent of their velocities, the effect of chromatic aberration in the final image is entirely eliminated. By placing a thin membrane between the final reducing lens and the specimen, which is sufficiently strong to serve as a vacuum seal, and yet transparent to an electron beam, von Ardenne considered it possible to attain a high resolving power in the examination of specimens mounted outside the vacuum chamber.

One successful method of recording the scanning microscope images provided by transmitted electrons involved placing a photographic film or paper immediately behind the specimen. The image of a small area of the specimen is recorded by displacing the photographic film in synchronism with the motion of the probe. In testing this apparatus with the type of specimen normally studied in the transmission microscope, von Ardenne concluded that a resolving power equal to the diameter of the scanning lines, *ca.*, 100 Å, was revealed in the direction of the scanning lines. In the perpendicular direction, the resolving power is limited by the distance between adjacent scanning lines. He did not succeed in demonstrating the usefulness of this apparatus for the study of the surface features of massive specimens. Although he recognized the possibility of using reflected or secondary electrons, the above described method of recording the images was inadequate.

The first successful microscope of this type, in which the resolving power of the metallurgical light microscope is exceeded, was described recently by Zworykin,

Hillier and Snyder.[161] The complex problems involved in the design of such an instrument, as well as the difficulties associated with the amplification of very small electron currents, are discussed in some detail in this paper. The principles of construction of this apparatus are illustrated in Fig. 29 (Fig. 9 [161]). The electron probe is produced by forming a greatly reduced image of a normal electron source with the aid of the first two electrostatic lenses. An image of this probe is subsequently focused in the plane of the specimen at approximately unit magnification by means

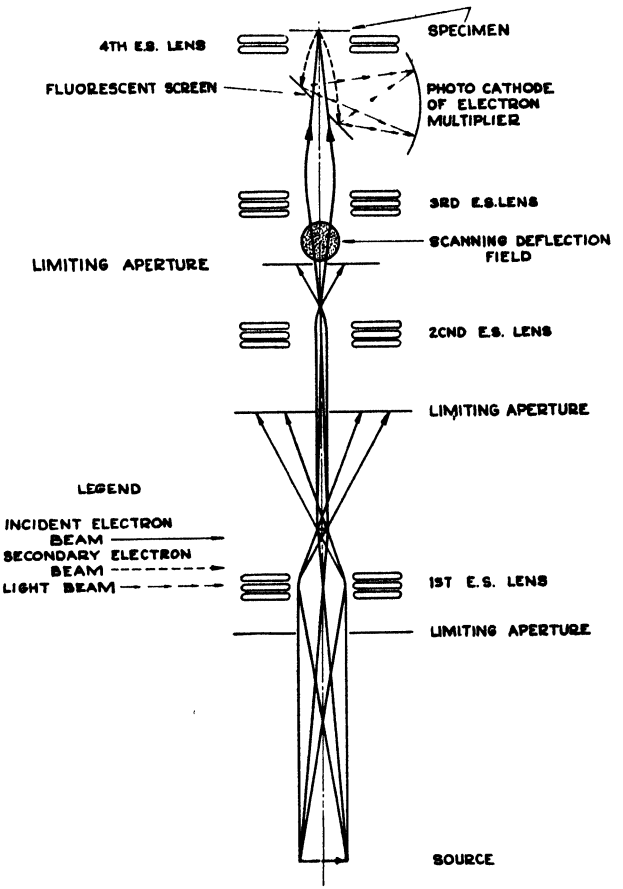


FIGURE 29. Schematic cross-section of the scanning microscope of Zworykin, Hillier and Snyder [161] developed for surface studies. (Courtesy RCA Laboratories)

of the third electrostatic lens. The use of this third relatively wide-aperture electron lens facilitates the insertion of a system of magnetic deflecting coils which control the motion of the electron probe over the surface of the specimen. The electron beam proceeding from the third lens passes through a weak decelerating lens to impinge upon the specimen. The decelerating lens serves the purpose of accelerating and guiding the slow secondary electrons emitted from the specimen to impinge upon a fluorescent screen without appreciably affecting the paths of the higher-speed electrons in the probe. Variations in secondary emission from different object elements become evident as variations in the light intensity emitted from the fluorescent screen. These light-intensity variations are subsequently amplified by means of a

photoelectric electron multiplier tube, ultimately to control the printing bar of a facsimile recorder providing the micrograph. The use of this photoelectric amplifying system provides a decided advantage in the detection of those secondary electron current variations which are of the same order of magnitude as the thermal noise in the input resistances of ordinary vacuum-tube amplifiers.

Fig. 30 (Fig. 13 [161]) illustrates the appearance of the micrographs which have been obtained with this apparatus. The specimen in this case is the surface of a

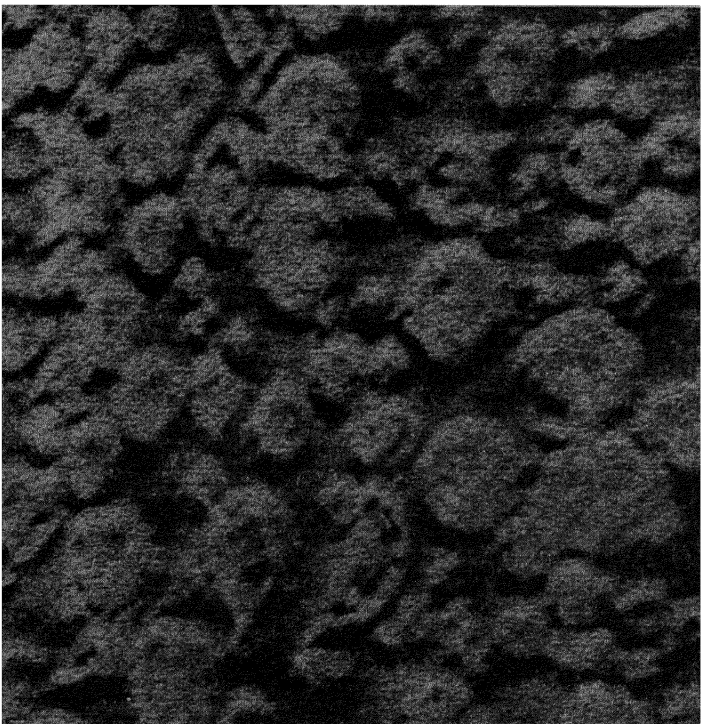


FIGURE 30. Micrograph of a ground and etched surface of a specimen of Bakelite containing carbonyl iron particles obtained with the scanning microscope of Zworykin, Hillier and Synder.[161] (Courtesy RCA Laboratories)

ground and etched surface of Bakelite containing carbonyl iron particles. The contrast in the image is due to variations in the secondary electron emission from different elements of the specimen. The contrast may be varied by using different electron probe accelerating potentials. Since the maximum sensitivity of secondary electron emission is in the neighborhood of several hundred volts, this scanning microscope is operated with lower beam accelerating potentials than are commonly used in transmission microscopes.

The intrinsic complexity of scanning apparatus relative to the simple lens system of the transmission microscope is a factor to be considered in evaluating the utility of the former type of apparatus in surface studies. It is apparent that results of greater immediate significance, particularly in metallographic problems, can be obtained through the study of *replicas* of surface features with the transmission microscope.

A number of methods for the preparation of replicas suitable for examination in

the transmission microscope have been described. The first, described by Mahl,[106] is applicable to the study of metals. It consists in preparing thin oxide films on the surface of metallographic specimens and removing these by chemical means. The success of this method depends upon the ease of oxidizing the metal and of subsequently removing the oxide film. Interpretation of the images of this type of replica is probably more difficult than in the case of the other types of replicas reviewed in the following.

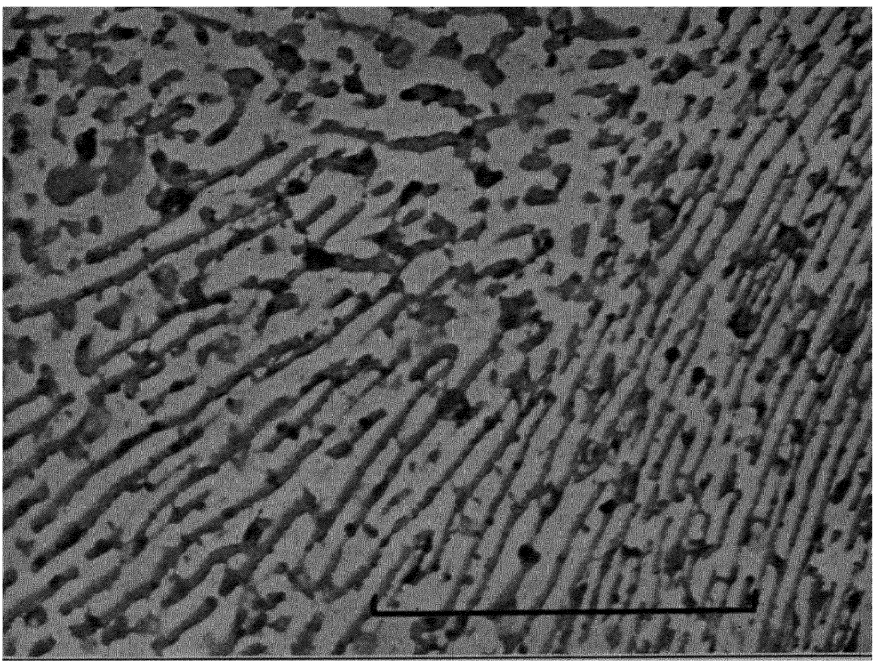


FIGURE 31. Replica of Pearlite structure prepared by double replica process of Zworykin and Ramberg.[104] Opaque areas represent elevations; transparent areas represent depressions in original sample. Some of the rectangular opaque areas are probably undissolved silver or salt crystals.

In the more versatile procedure described by Zworykin and Ramberg[104] a relatively thick film of silver, evaporated from a heating element in a vacuum chamber, is condensed on the surface of the specimen. After removal of the specimen from the vacuum chamber, the silver film is stripped mechanically from the specimen, thus providing a "negative" replica. The thickness, and consequently the tensile strength, of this film can be made sufficiently great to obviate the occurrence of large-scale distortions in the stripping process. A dilute solution of cellulose nitrate, or other plastic, is subsequently spread on the surface of the silver replica to a depth which will dry to a film of sufficient transparency for examination in the electron microscope. Upon removal of the silver film with dilute acid, the cellulose nitrate film serves as a "positive" replica. In the ideal replica, the thicker areas appearing as dark areas in a micrograph represent elevations; the lighter areas represent depressions in the original specimen.

The applicability of this method is limited to those materials which can be placed in a vacuum chamber without suffering disintegration. In metallographic work, the

most serious restriction is imposed by the alloying of the evaporated metal with the original specimen. This may be overcome in many cases by selection of a more suitable metal for the negative replica. There is also some difficulty in dissolving the negative replica from the plastic "positive" with assurance that the metal is completely dissolved and that the structure of the "positive" has not been altered. The latter difficulty does not appear to be insurmountable. Presumably it may be avoided or eliminated entirely by choice of more suitable materials than silver and cellulose nitrate.

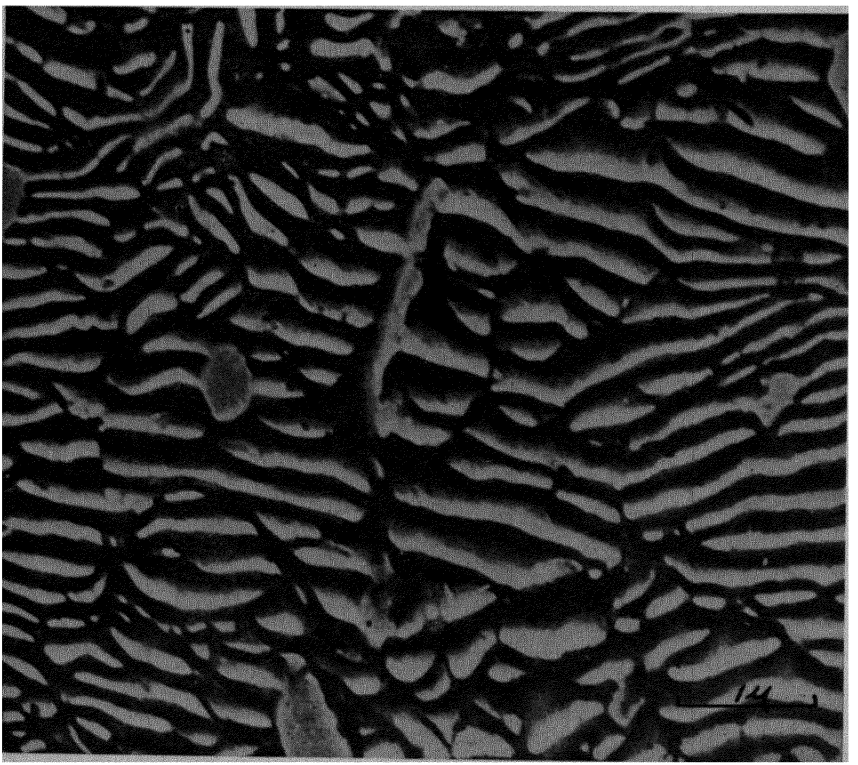


FIGURE 32. Replica of Pearlite structure prepared by initial single replica process of Schaeffer and Harker.[148] Opaque areas represent depressions; transparent areas represent elvations in original sample. The diffuse light lines are thought to represent strains in the replica. (Courtesy General Electric Co.)

The latter difficulty is, in fact, largely overcome in the double replica process developed by Heidenreich and Peck.[75, 76] In their technique, a negative replica is placed in a vacuum chamber and has deposited on it a thin film of evaporated silica. The high mobility of silica during condensation is effective in the production of a silica film, one surface of which follows the contours of the negative, the other of which is plane within limits of practical interest. Such a silica film is more resistant to chemical action than is a plastic film. Consequently, a more faithful replica, insofar as cleanliness and maintenance of structure in the silica film are concerned, may be expected.

In the first step of the Heidenreich and Peck process, polystyrene is moulded on the surface of the specimen with high pressure and elevated temperature. The

moulded negative replica is then separated from the specimen by mechanical shock or by dissolving the specimen. The method is not applicable to the study of soft materials or of those materials which may be altered at the temperatures required for moulding. In the methods of both Zworykin and Ramberg, and Heidenreich and Peck, a vacuum evaporator is required and the preparation of a replica requires a considerable amount of time.

Several techniques have been devised by Schaeffer,[141, 142, 143] for the preparation of negative replicas which may be examined directly in the electron microscope.

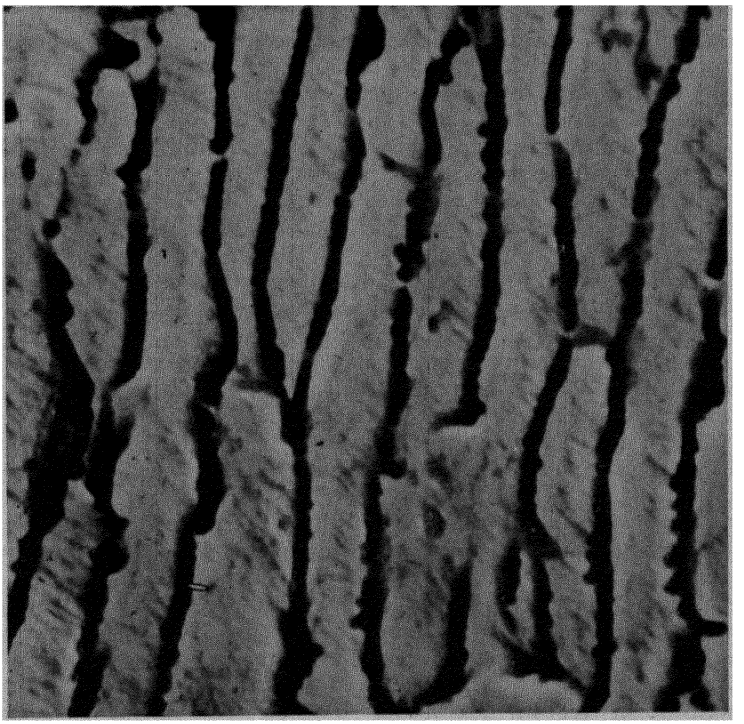


FIGURE 33. Replica of Pearlite structure prepared by double replica process of Heidenreich and Peck.[76] Opaque areas represent elevations; transparent areas represent depressions in original sample. (Courtesy Dow Chemical Co.)

The success in the preparation of sufficiently thin and strong replicas of this type depends in part upon the choice of replica material. Schaeffer and Harker discovered a satisfactory material, Formvar 15-95. Equally important to the success of the method is the observation of certain principles in the conditioning of the surface of the original specimen.

Schaeffer found that Formvar films of very small thickness, 100 Å or less, can be stripped easily from smooth surfaces, such as glass, if the glass is properly cleaned. "A single layer of molecules of a hydrophobic substance, such as grease or oil, will completely prevent carrying out this normal stripping process and it is necessary to increase the film thickness at least tenfold in order to add sufficient tensile strength to the film to enable its mechanical removal. Some evidence has been noted which indicates that conditioning metal surfaces with surface-active molecules such as tricresyl phosphate and oleic acid greatly facilitates the removal of replicas. Not more than a monolayer is needed for this purpose."

In order to obtain a coherent film from irregular surfaces, such as etched metallographic specimens, it is of course necessary to deposit films of average thickness considerable greater than 100 Å, for it is essential to have the thinnest areas sufficiently strong to permit stripping. From the results of experiment, Schaeffer and Harker estimate that an optimum compromise between adequate contrast and mechanical strength is obtained with average film thicknesses from 500 to 750 Å. Satisfactory replicas are obtained with the following procedure described by them.

"After the specimen has been given a good metallographic polish (we use MgO on polishing cloth known as kitten's ear for the final polishing of steel specimens), the surface is given an etch of such a depth that a fairly good photomicrograph can be obtained at 1000X in the light microscope. The etching solution is then removed and the surface quickly dried with a jet of clean filtered air. The specimen is then immediately lowered face down into a Petri dish containing a small amount of 0.5 per cent solution of polyvinvyl formal dissolved in dioxane. The specimen is removed with a smooth motion and swung into a position such that the plane of the surface is vertical with respect to the surface of the resin solution. The solvent is allowed to evaporate to completion while held in the vertical position. This procedure forms a wedge-shaped film on the metal surface thickest at the lower edge and thinnest at the top. With a specimen having a diameter of 2.5 cm the range in thickness will be from 500 to 750 Å, although at the very bottom there will be a small area much thicker than this where the draining liquid accumulates."

The replica film may be removed from the specimen by several methods. The simplest technique, which at the same time avoids contamination of both the original specimen and the replica, is described in a recent article by Schaeffer.[142] A standard specimen screen is placed over the region to be studied. "With the screen in place, a bit of moist air is applied to the coated surface and immediately a piece of scotch tape is pressed into light contact with the screen and the surrounding film. Holding the scotch tape against the specimen, on one side of the positioned screen, the tape is held rigidly and lifted away from the surface. When this is done the replica leaves the specimen and will be found in contact with the under surface of the screen ready for mounting in the specimen holders of the electron microscope. The scotch tape is easily removed from the mounting screen by placing the tape holding the screen in contact with a flat surface. The tape is then turned through 180°, and a finger placed on the sticky surface until the edge of the screen becomes visible. A razor blade or similar thin flat object held against the screen prevents it from becoming bent as the tape is peeled away from it. The entire process of replica formation, removal and mounting may be carried out in less than two minutes. Perhaps the most attractive feature of the process described is the fact that the specimen is not injured in any way. For example, from a specimen of 1.1 C Steel (used for developing the technique) at least fifty replicas have been obtained without injuring the surface in any way."

An evaluation of the relative merits of the above methods of preparing replicas in the study of materials for which any one of them may be applicable is difficult to carry out without making direct and detailed experimental comparisons. The most fundamental consideration in the value of a replica prepared by any of these methods is the faithfulness of the replica to the original surface. For a replica to be completely faithful, it should have one surface a precise counterpart of the original, while the other is plane. If both of these conditions are fulfilled, it is possible to use the thickness of the replica as a measure of the elevations or depressions in the original surface.

It is highly probable that the surface of an *evaporated* film adjacent to the original specimen is in intimate contact with it and consequently satisfies one requirement of an accurate replica. The degree to which it remains an accurate counterpart after removal depends upon the cohesive forces between the surface and the evaporated

film, and upon the magnitude of reentrant angles. It may be expected that in the stripping process, some portions of the negative remain embedded in the original, and in some cases, that some of the original features are pulled away with the negative.

A moulded replica may be in intimate contact with the original if it is sufficiently fluid during the moulding process. Accuracy of the replica in its finer details will also depend upon the relative magnitudes of surface tension in the plastic and of the cohesive forces between the plastic and the specimen surface. The amount of tearing of the plastic or of the original will depend upon the method used for separating them. Probably less rupturing occurs if the original is dissolved. The same factors must be considered if the negative replica is in the form of a solution when applied, as in the Schaeffer and Harker technique. In accordance with their test, it is apparent that an intimate contact is established between the resin film and the original, and it is thus permissible to regard the Formvar film as an accurate counterpart while it is in contact with the original.

On the basis of the above consideration, it is apparent that both the evaporated metal and the soluble plastic films are likely to be more accurate counterparts (in the finer details) than the moulded impressions. Of equal significance to the final result, however, is the effect on the replica of the method of separating the surfaces. The least alteration of the replica impression is likely to be incurred by dissolving the original specimen. Mechanical stripping will introduce distortions and ruptures to some extent. Large-scale distortions are likely to occur in stripping the thinner films of the Schaeffer and Harker technique. The existence of such distortions has been recognized by Schaeffer and Harker. Small-scale distortions due to cohesive forces, and reentrant angles are likely to occur in all the techniques. In the event of alloying, these effects are more serious in the Zworykin and Ramberg method.

The degree of flatness of the outer surface of the final replica is the second factor to be considered. In both the Formvar and silver collodion processes, it is assumed that the solutions spread on the irregular surface, will have a flat outer surface as a result of surface tension. The assumption has been partially verified by Schaeffer and Harker. It is evident, however, that the degree of flatness depends rather critically upon the depth of the film above the highest elevations in the irregular surface (in the original in the case of the single replica process; in the negative replica in the case of the double replica processes). A plane outer surface can only be assumed if the depth of the film is large in comparison to the depth of the highest projections (the case studied by Schaeffer and Harker). This probably is not the case if the films are made of a thickness to yield greatest contrast. In the polystyrene process, the high mobility of evaporated silica films was discovered by accident. Heidenreich and Peck express the opinion that the resultant film may be regarded as flat in the study of the finer details. Both surfaces of the silica film tend to follow the contours of the polystyrene surface. The outer surface, however, is considered to be smoother. The limits within which the outer surface may be regarded as flat have not been determined with any degree of accuracy.

It may be assumed that the film obtained by spreading a plastic solution over the irregular surface will result in a more faithful replica of the coarse features of the surface. In the single replica process of Schaeffer, the finer details are more likely to be revealed in the areas of the specimen representing elevations in the original surface. The films are thinnest in these areas and greatest resolution is obtained in the electron microscope images. On the other hand, the finer details in the depressions of the original surface are likely to be revealed in the double replica processes, since the thinnest film areas represent depressions. In those investigations in which finer details throughout the entire specimen are of primary interest, a silica film as a final replica will probably be most useful. The technique to be employed in any particular problem depends upon the nature of the material as well as upon the factors discussed above. Materials which are mechanically fragile, or sensitive to elevated

temperatures or abnormal pressures, are not capable of withstanding the treatment involved in the above double replica processes.

From inspection of the numerous micrographs which have been prepared through application of all of the above replica processes, it is evident that the highest resolving power exhibited in replica micrographs is lower than the optimum resolving power of the instruments. Although very sharply defined boundaries and particles frequently may be observed, it is not at all certain that these are true replica structures, for there is a strong probability that some of the finer structures of the original surface may be removed with a stripped replica. A reduced resolving power may be ascribed to one or both of two factors, *viz.*, insufficient image contrast to reveal the smaller thickness differences; failure of the replica to provide an accurate representation of the smaller structural units in the surface. It is apparent from the results obtained so far that crystallites of dimensions near the limit of resolution are not clearly revealed in the micrographs of replica specimens.

(b) **In Microbiology.** Exploitation of the electron microscope in the biological fields has been the objective of even the earliest workers engaged in the development of the instrument. The first micrographs of cell structures were obtained by Marton, in 1934,[112] who impregnated biological sections with osmium salts and so obtained images of the osmium skeletons which remained after the more sensitive cell materials had been destroyed by electron bombardment. Krause [98] later demonstrated that it was possible to study delicate cell structures without the use of dyeing or fixing agents. The results of these early investigations, however, were more significant in showing that the electron microscope was not particularly well adapted for the study of such massive structures as complete cells or cut biological sections. That the instrument was well suited for the study of the smaller biological structures represented by bacteria and viruses has been made clearly evident in numerous publications in which the appearance of many types of these organisms is illustrated.



FIGURE 34. Partially cytolyzed streptococcal cells under the action of sonic vibration.[122] (Courtesy RCA Laboratories)

Bacterial Morphology. A number of instructive investigations in bacterial morphology have been reported. In publications of Mudd and Lackman,[122] and of Mudd, Polevitsky, Anderson and Chambers,[128] evidence is presented to confirm the existence of a relatively rigid membrane enveloping the protoplasm of certain bacteria. The

most conclusive evidence is furnished by the micrograph Fig. 34 (Fig. 5 [122]) of streptococcal cells and Fig. 35 (Fig. 1 [123]) of *Bacillus megatharium* cells. In both cases, suspensions of the cells were partially cytolyzed under the action of sonic vibrations. Some of the cells among those in Fig. 34 have been cytolyzed and appear



FIGURE 35. Partially cytolyzed *Bacillus Megatharium* cells under the action of sonic vibration.[123]

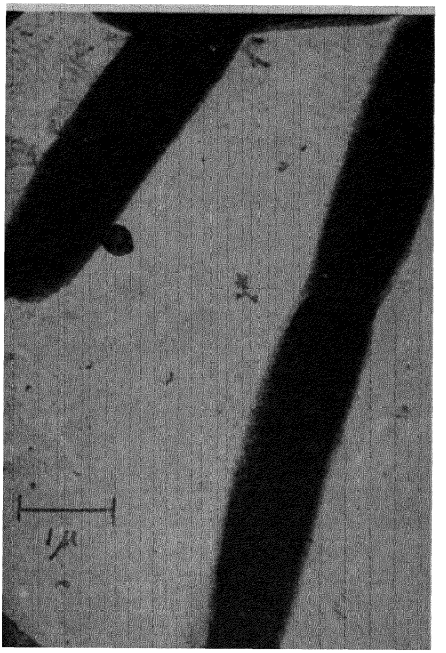


FIGURE 36. Intercellular plate of *Bacillus anthrisis*.[123]

(Courtesy RCA Laboratories)

as "ghost" cells. These do not differ significantly in their outlines from the intact cells, but since they are much more transparent to the electron beam, it is evident that the protoplasm has escaped. In Fig. 35, there appear fragments of cells produced by sonic vibration. The undistorted tubular fragments with ragged ruptured ends again indicate the rigidity of the cell wall. Of further interest in this publication is the demonstration of the existence of the intercellular plate shown in Fig. 36 (Fig. 11 [123]). This evidently consists of a thickened area of cell membrane.

An exploratory investigation of the value of staining bacterial preparations with the object of revealing protoplasmic structure has been reported by Mudd and Anderson.[121] Since the contrast in the electron microscope image is governed essentially by variations in the thickness-density of a specimen, characteristic internal structure is difficult to observe. Both the thickness and density of the organic compounds in

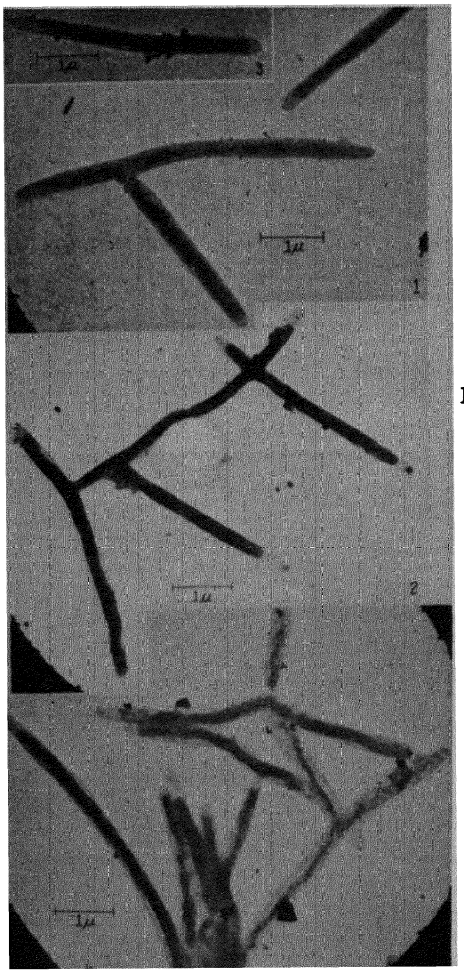


FIGURE 37. Enhancement of contrast by staining. (1) Unstained *Fusobacterium;* (2) stained with silver nitrate; (3) stained with mercury bichloride; (4) stained with lead acetate.[121]

the protoplasm are in general nearly uniform throughout. Some advantage might, therefore, be gained by the selective staining of cells with heavier elements, as initially suggested by Marton.[112] With these possibilities in mind, Mudd and Anderson prepared bacterial specimens stained with various heavy metals. It was found that the cell wall is unaffected by the stains, but that the stains are absorbed, as expected, by the protoplasm. Fig. 37 (Plate 11[121]) illustrates the absorption and increased contrast which results from the staining of specimens of *Fusobacterium* with silver nitrate, mercury bichloride and lead acetate. Although the selective action of these stains is not certain, the results do indicate that the development of this type of staining technique is a promising procedure in the study of protoplasmic structure.

The above demonstration, as well as the earlier investigations of von Borries, E.

Ruska and H. Ruska [186] concerning the identification of endo- and ektoplasm; the attempted demonstration of nucleic acid by Pickarski and Ruska; [128] and the study of tuberculosis bacilli by Lembke and Ruska [100] demonstrate the lack of sufficient "penetrating" power in this type of investigation with beam potentials below 100 kv. Greater progress in this phase of cell morphology will undoubtedly result through the use of higher-voltage instruments.

Several species of animal virus, in many respects closely related to bacteria, have also been demonstrated. The most instructive is the observation of morphological structure in vaccinia virus by Green, Anderson and Smadel.[68] It was found that the vaccinia virus particles have a rectangular shape in electron microscope specimens. An internal structure is revealed even with the use of a relatively low beam potential (78 kv). Elementary bodies of purified virus particles are shown in Fig. 38

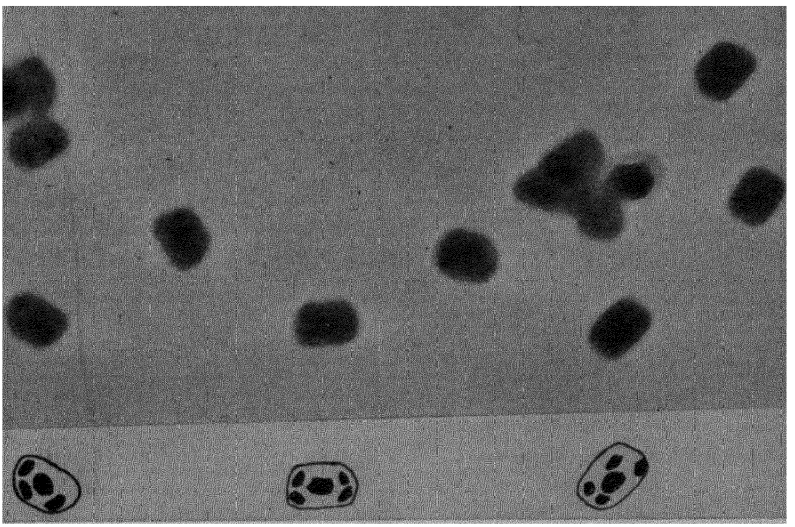


FIGURE 38. Elementary bodies of purified vaccinia virus. Rectangular-shaped virus particles include dark spots revealing internal structure.[68]

(Fig. 1a [68]). The dark areas within the rectangular boundaries represent the internal structure, which is not yet explained.

Bacterial Reactions. A variety of bacterial reactions have also been demonstrated with the electron microscope. An interesting example is the study carried out by Ruska [189] in which he recorded the process of lysis resulting from the addition of phage to a suspension of colon bacteria. This is illustrated in the series of micrographs, Fig. 39, reproduced from his paper. The initial rapid absorption of the phage on the bacterial membrane is shown in upper left. The subsequent disintegration of the membrane, escape of the protoplasm and final destruction of the bacillus is demonstrated in the remainder of the series. It is perhaps unnecessary to point out that four different bacilli are involved in the series of micrographs.

It is evident that the phage particles, or bacterial viruses, shown in the above series have suffered structural alteration from the effects of electron bombardment. This is clearly revealed in an analysis of the forms of different strains of bacteriophage by Luria and Anderson.[102] In the latter investigation it is shown, for example, that several species of colon bacteriophage consist of sperm-shaped particles.

A study of the alterations of the flagella and cell walls which result from specific

sensitization with anti-sera has been reported by Mudd and Anderson.[120] Figs. 40 and 41 (Figs. 3 and 4[120]) demonstrate their observations. The flagella and the sharply defined cell walls exposed by shrinkage of the protoplasm *in the control specimen* are shown in Fig. 40. Fig. 41 illustrates that the antibodies of the serum have been adsorbed on the flagella. The presence of adsorbed material on the cell walls is indicated by their greater opacity to the electron beam. The relatively uniform increase in thickness of the flagella and the decrease in sharpness of

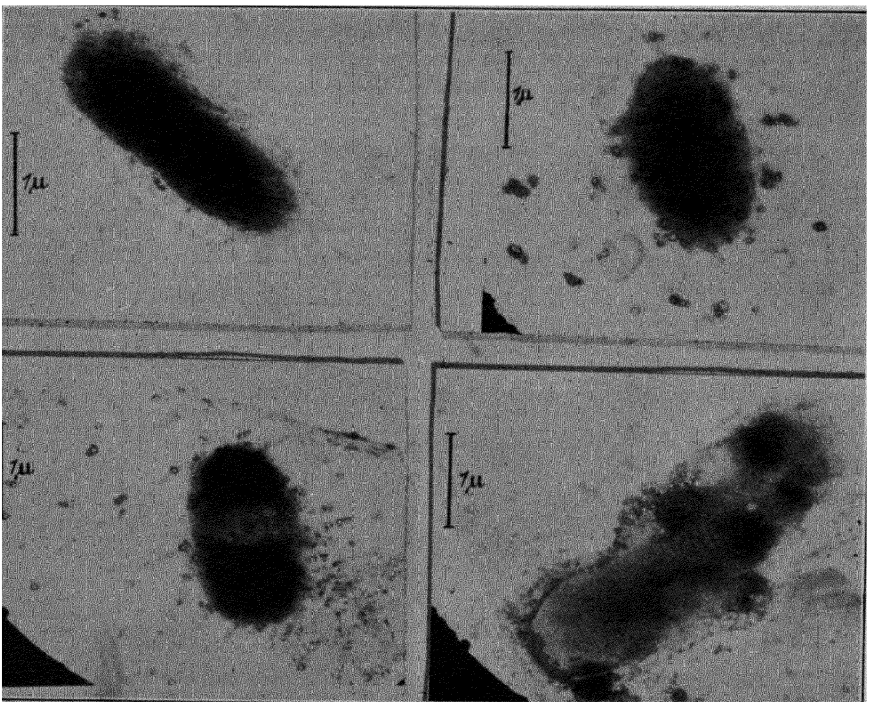


FIGURE 39. Lysis of colon bacteria by phage.[189] The structure of the phage particles, appearing as small annular particles, has been altered by electron bombardment.

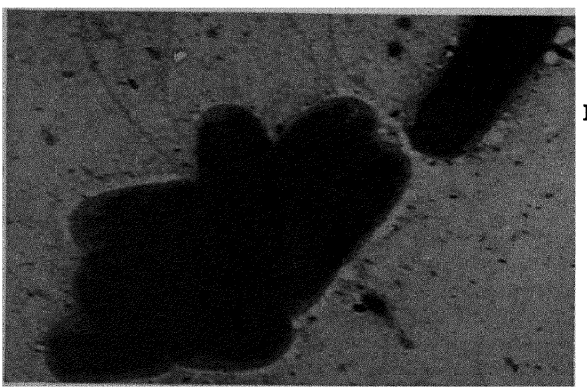


FIGURE 40. *Eberthella typhosa* with heterologous anti-streptococcal serum.

outline have been explained by Mudd and Anderson on the assumption that the antibodies are rod-shaped, and are adsorbed on the surface of the flagella end-on, in a thickness of 21 Å.

Biochemical reactions may be studied with equal success. An interesting aspect of this problem is the determination of the locality of the chemical reactions which are brought about by bacteria. In a study carried out by Morton and Anderson [119] diphtheria bacilli were grown in a chocolate agar containing potassium tellurite. In

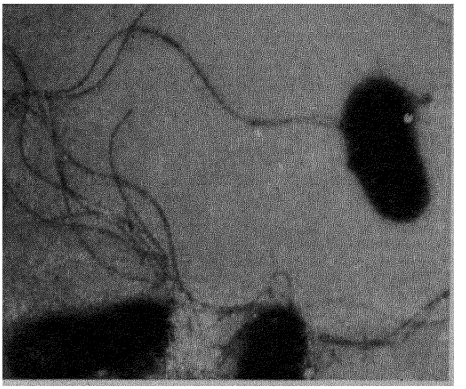


FIGURE 41. *Eberthella typhosa* with homologous anti-typhoid serum.[120]

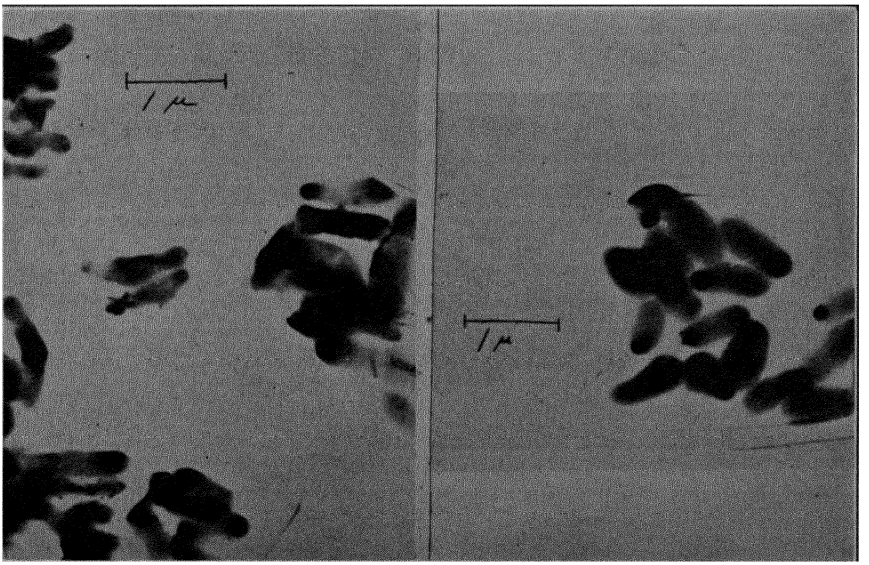


FIGURE 42. Reduction of potassium tellurite to metallic tellurium by diphtheria bacilli; (a) (left) bacilli grown in chocolate agar containing potassium tellurite; (b) (right) bacilli grown in control agar of horse blood.[119]

the presence of these bacteria, the tellurium salt is reduced to metallic tellurium. Electron micrographs have revealed that the reduction takes place in the interior of the organisms and may be regarded as part of the digestion process of the cells. This is demonstrated in Fig. 42. The long needle-shaped crystals visible within the cells

shown in (a), are the product of the reduction of the potassium tellurite. The typical polar granules of the normal cells are shown in the comparison micrograph (b).

Protein Molecules. Several publications have appeared in which the appearance of isolated globular and crystalline protein molecules have been demonstrated. It has been the purpose of these investigations to obtain a correlation between the dimensions of the molecules computed from sedimentation and diffusion experiments with the dimensions of the particles observed in electron microscope specimens. The first

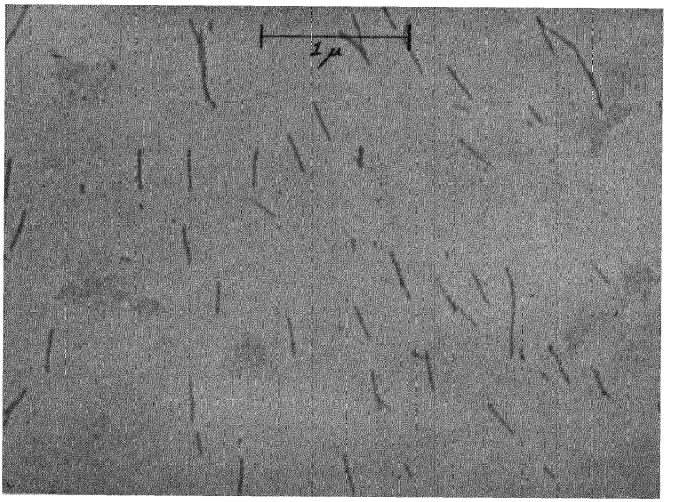


FIGURE 43. Single molecules of tobacco mosaic virus.[154] (Courtesy RCA Laboratories)

exploration of this kind was made by Kausche, Pfankuch, and Ruska,[91] who examined specimens of tobacco mosaic virus suspensions. They demonstrated the existence of rod-shaped particles in specimens of TM virus, the measured dimensions of which were in satisfactory agreement with the accepted values. A further study of the size and shapes of these molecules was carried out by Stanley and Anderson.[154] From direct measurements in the micrographs of TM virus suspensions, one of which is illustrated in Fig. 43, Stanley and Anderson deduced a mean width of 15 mμ and a mean length of 280 mμ for the dimensions of these molecules.

The precise agreement of the direct measurements of the smaller dimension with the cross-section dimensions determined from x-ray measurements is somewhat misleading. The investigators have measured the widths of the particles in the image between boundaries which are not clearly defined. In view of the image contour phenomena discussed in Section III(b), and because of the lack of a sufficiently detailed knowledge of the scattering of electrons by structures of such small dimensions, the writer regards the implied magnitude of the probable error in these measurements as too small.

In a later publication of Stanley and Anderson,[155] the dimensions of individual particles of several more proteins are correlated with indirect measurements. Three of these are illustrated in Fig. 44 (Figs. 4, 5, 6,[155]). In the upper micrograph, of *Busycon canaliculatum* hemocyanin, the average diameter of 75 particles was found to be *ca.* 22 mμ, the majority of the particles having diameters between 19 and 24 mμ. An analysis of the suspension from which the specimen was prepared in the analytical centrifuge revealed the existence of three components, in the proportions 40, 20,

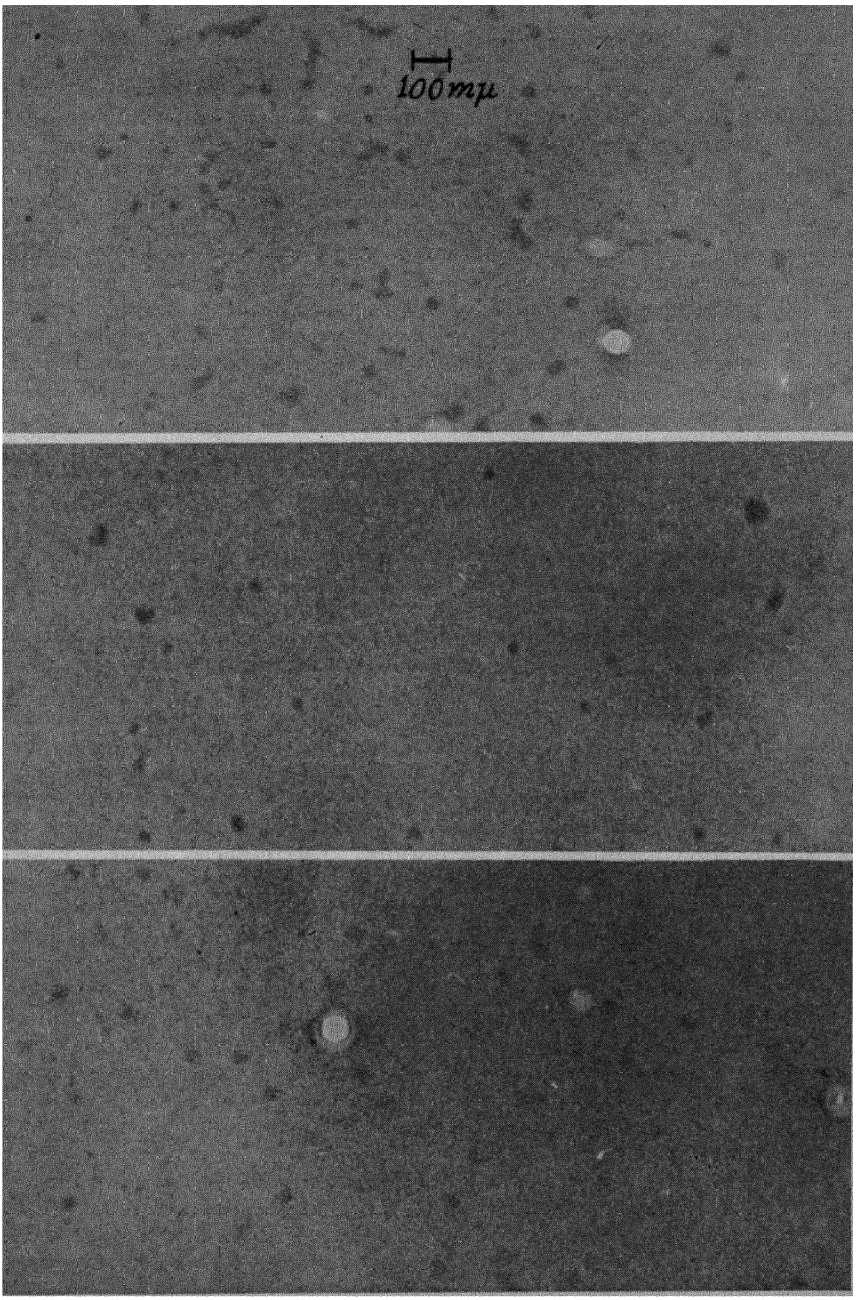


FIGURE 44. Protein molecules. Upper: *Busycon canaliculatum* hemocyanin with directly measured average particle diameter 22 mμ; center: *Limulus polyphemus* hemocyanin with directly measured average particle diameter 20 mμ; lower: *Viviparus maleatus* hemocyanin with directly measured average diameter 29 mμ.

and 40 per cent. Assuming that these molecules are spherical, the diameters computed from the sedimentation constants of these components are 10 mμ, 19 mμ, and 24 mμ respectively. It is evident from the value of the average of the directly measured diameters that the particles in the smallest size component are not revealed in the electron micrograph. Stanley and Anderson regard this as evidence to indicate that the molecules have the shape of prolate ellipsoids. This hypothesis is also indicated by the apparent thinness of the particles, judged by their relative transparency in the micrographs.

Somewhat similar conclusions were drawn from the electron micrographs of *Limulus polyphemus* hemocyanin, shown in the center of Fig. 44 (Fig. 5 [155]). The average diameter of 33 particles shown in this micrograph was found to be *ca.* 20 mμ. Analysis of the suspension from which the specimen was prepared indicated that there were two components. On the basis of spherical particles, the diameters of the particles comprising the greater proportion of the material, *i.e.* 90 per cent, have a computed value of 19 mμ. The greater opacity of the particles appearing in the micrograph, as well as the close agreement of the computed and directly measured diameters, led Stanley and Anderson to the conclusion that the particles in the specimen are spherical in shape.

In the case of the particles of *Viviparus maleatus* hemocyanin, shown in the lower micrograph of Fig. 44, about 32 particles have measured diameters of 29 mμ. Eighteen of these have diameters of *ca.* 36 mμ. The relative transparency of these particles indicates that they are probably disk-shaped. A centrifuge determination of the size of these molecules led to a diameter of 24 mμ, (on the assumption of spherical shape), considerably smaller than the average of the directly measured diameters.

"The results indicate that the electron microscope is very useful in the elucidation of the sizes and particularly the shapes of the larger protein molecules. In the case of molecules which are essentially spherical in shape, such as, for example, bushy stunt virus and probably also *Limulus* hemocyanin, the agreement between molecular sizes estimated by indirect methods and those estimated directly from electron micrographs is excellent. The fact that in one and perhaps two cases the molecular sizes estimated by means of the electron microscope are not in accord with those estimated by indirect methods based on the assumption of a spherical shape is a probable indication that the latter assumption is not justified. Discrepancies between molecular weights estimated by different indirect methods are usually regarded as being due to hydration, to asymmetry, or to both, and in most cases it is not possible to determine by the indirect methods which factor is the more important. For the larger protein molecules it is now possible to determine directly by means of the electron microscope the cases of molecular asymmetry due to a rod-like shape, such as occurs, for example, in tobacco mosaic virus, and it is possible to secure some measure of information regarding disk- or plate-like shapes, such as appear to occur in *Viviparus* hemocyanin and probably also in *Busycon* hemocyanin. This type of information, which can be provided by electron micrographs, should make it possible to evaluate better the relative importance of hydration and asymmetry in molecular weight estimations of the larger protein molecules."

In considering the application of the electron microscope in this field, the discussion of von Ardenne [17] is of some interest. It is concerned with the limiting dimensions of organic particles which may be detected and resolved with present instruments. The need for specimen-supporting films of smaller thickness-density than are commonly used is made evident, for in this discussion it is shown that the dimensions of the smallest observable particles depend critically upon the contrast in the image.

Plant Virus Reactions. Of the numerous biological and chemical reactions which have been demonstrated in the past three years, two typical examples will be reviewed to illustrate the data provided by electron micrographs. For a more extensive discussion, the reader is referred to the paper in this volume by Lauffer and Stanley.

As suggested by Kausche,[89] the Lange gold sol reaction may be used as a test to distinguish between the virus proteins of diseased potato and tobacco plants. Their identities may be established by the observation of the precipitation which follows the addition of a red-colored gold sol to suspensions of viruses obtained from diseased plants (see also [127]). On adding a red sol to a neutral suspension of potato-X virus, a voluminous red-colored sediment is produced. A precipitate is difficult to obtain by adding the red sol to a neutral suspension of tobacco mosaic virus, but if sufficient sol is added, a precipitate is formed which is blue in color. The red-colored precipitate characteristic of the potato-X virus reaction may be obtained, however, by add-

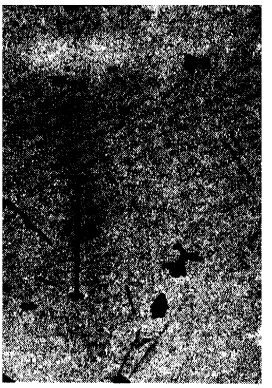


FIGURE 45a. Blue colored precipitate in neutral suspension of tobacco mosaic virus.

ing red sol to an acid suspension of the tobacco mosaic virus. In the latter case, it is difficult to separate the protein from the gold particles by centrifuging. This indicates that the gold particles are adsorbed on the protein molecules.

Kausche and Ruska [92] succeeded in demonstrating this precipitin reaction and thereby in explaining the observed results. The blue-colored precipitate of the sol in the neutral suspension of tobacco mosaic virus is illustrated in Fig. 45a (Fig. 3 [92]). The gold crystallites have aggregated into larger particles which settle out independently of the protein molecules. Since there is no evidence of a physical bond between the gold aggregates and the protein molecules, the ease of their separation by centrifuging may be explained. The red-colored precipitate which is obtained with the neutral potato-X virus and the acid tobacco mosaic virus suspensions is illustrated in Fig. 45b (Fig. 7 [92]). In this case, the adsorption of the gold particles on the protein molecules is clearly evident. Although the resultant aggregates are more massive than either the gold particles or protein molecules, the spacing between the gold particles is much larger than in the case in which they are aggregated by themselves, and the characteristic red color is maintained in the precipitate. Because of the strong attractive bond between the gold and protein particles, separation by centrifuging is difficult to carry out.

Of greater interest are the precipitin reactions indicating the inactivation of the virus proteins by antibodies which are generated in the blood stream of animals to combat disease. The first report of the investigations of Anderson and Stanley [2] is an excellent example of the work being carried on.

The absence of observable interaction between the blood serum of a normal rabbit with tobacco mosaic virus may be inferred from the appearance and dimensions of the virus molecules occurring in a suspension of tobacco mosaic virus to which had been added the normal serum. A specimen is illustrated in Fig. 46a (Fig. 3 [2]). The rod-shaped tobacco mosaic virus particles are identical with the normal virus particles shown in Fig. 43.

On adding a purified anti-serum obtained from a rabbit into which the virus previously had been injected, Anderson and Stanley were able to demonstrate the antibody-antigen reaction. The adsorption of the antibodies on the virus proteins may be inferred from the increased width and loss of sharp boundaries of the rod shaped

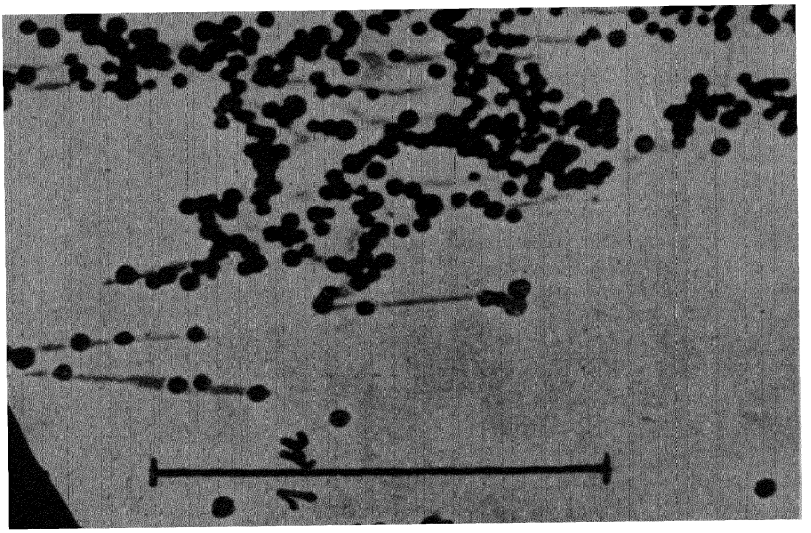


FIGURE 45b. Red colored precipitate in acid suspension of tobacco mosaic virus.[92]



FIGURE 46a. Normal serum and tobacco mosaic virus.[2]

particles shown in Fig. 46 (b) (Fig. 4[2]). The characteristic aggregation leading to precipitation is also demonstrated by the aggregates of enlarged molecules. The large black particle appearing in this micrograph is an impurity introduced in the specimen to facilitate focusing of the image. Without this, focusing is difficult because of the low contrast in the image between the specimen particles and the background consisting of the supporting film.

Collagen. The utility of the electron microscope for the investigation of organic structures composed of organized aggregates of long-chain molecules has been demonstrated recently by Schmitt, Hall and Jakus.[148] The structure of collagen fibers has been one of the most difficult to elucidate by means of the customary techniques employed in the study of large molecules. Attempts to describe the configuration of the polypeptide chains in terms of the keratin chain models have been unsatisfactory. The difficulty is partially attributed to the diversity of the R groups in the side chains

of the amino acid residues. This prevented the preparation of clear x-ray patterns from which the manner of folding or coiling of the chains could be interpreted. It is enhanced further by the abnormal extensibility and contractility of collagen fibers relative, for example, to wool. A distinguishing feature in the x-ray patterns of collagen, known for some years, is the existence of diffraction maxima indicating a periodicity of large spacing in the direction of the fiber axis. The latter periodic spacing, well above the resolving limit of the electron microscope, suggested the use of this instrument for the direct observation of collagen fibers. A detailed study of collagen, involving methods of preparation, determination of composition, and structure

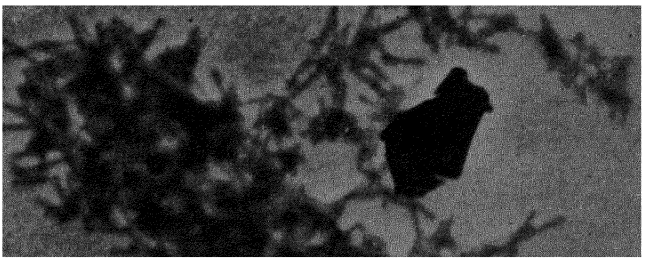


FIGURE 46b. Tobacco mosaic virus in presence of anti-serum.[2]

analysis by polarization, x-ray and electron-optical methods was therefore initiated by Schmitt.

In preliminary publications by Schmitt, Hall and Jakus,[71, 148] electron micrographs of collagen fibers from a variety of sources, including tail tendon, tendo Achilles, dermis of skin, interfibrillary connective tissue of nerve, cornea, spleen and ligamentum nuchae have been illustrated. Fibers for the specimens were obtained either by teasing small fibers in water or by dissolving the materials in acetic acid and reprecipitating the fibers by neutralization.

Typical structures observed in the fibers from specimens from all of the above sources is illustrated in Fig. 47 (Figs. 7, 8, 9[148]). The fibers are cross-striated with relatively opaque bands (A bands) and transparent bands (B bands) extending uniformly across them. The uniform optical density along the lengths of the A and B bands, as well as the appearance of folded fibers, indicate that the fibers have a sensibly uniform thickness and may be regarded as ribbon-shaped. The spacing between successive A or B bands was found to vary from about 400 Å to 1000 Å in fibers from different preparations. In unstretched individual fibers, however, the spacing is reasonably constant. On making a distribution curve of the fiber axis spacing from measurements in a single preparation of rat-tail tendon, the curve was found to have a sharp maximum at 640 Å. X-ray diffraction studies, by Bear,[32] of similarly prepared and air-dried specimens of collagenous materials revealed a fiber axis spacing in the range 638-648 Å, thus verifying the conclusion that the large fiber axis structural periodicity inferred from x-ray observations is identical with the striated structure revealed in the electron micrographs. It is also evident from these observations that the preparation of the specimens for electron-microscope observation causes appreciable contractions and extensions in the single isolated fibers.

The latter fact, together with the discovery, in several specimens, of greatly extended fibers, provide new data for the analysis of the molecular organization in collagen. One of the stretched fibers is shown at the right in Fig. 47 (Fig. 9[148]). At the upper end of this stretched fiber, the spacing between successive A bands has been increased to nearly 6,000 Å, nine times the average normal spacing of 640 Å. A study of the widths of the A and B bands shows that the A bands suffer a greater

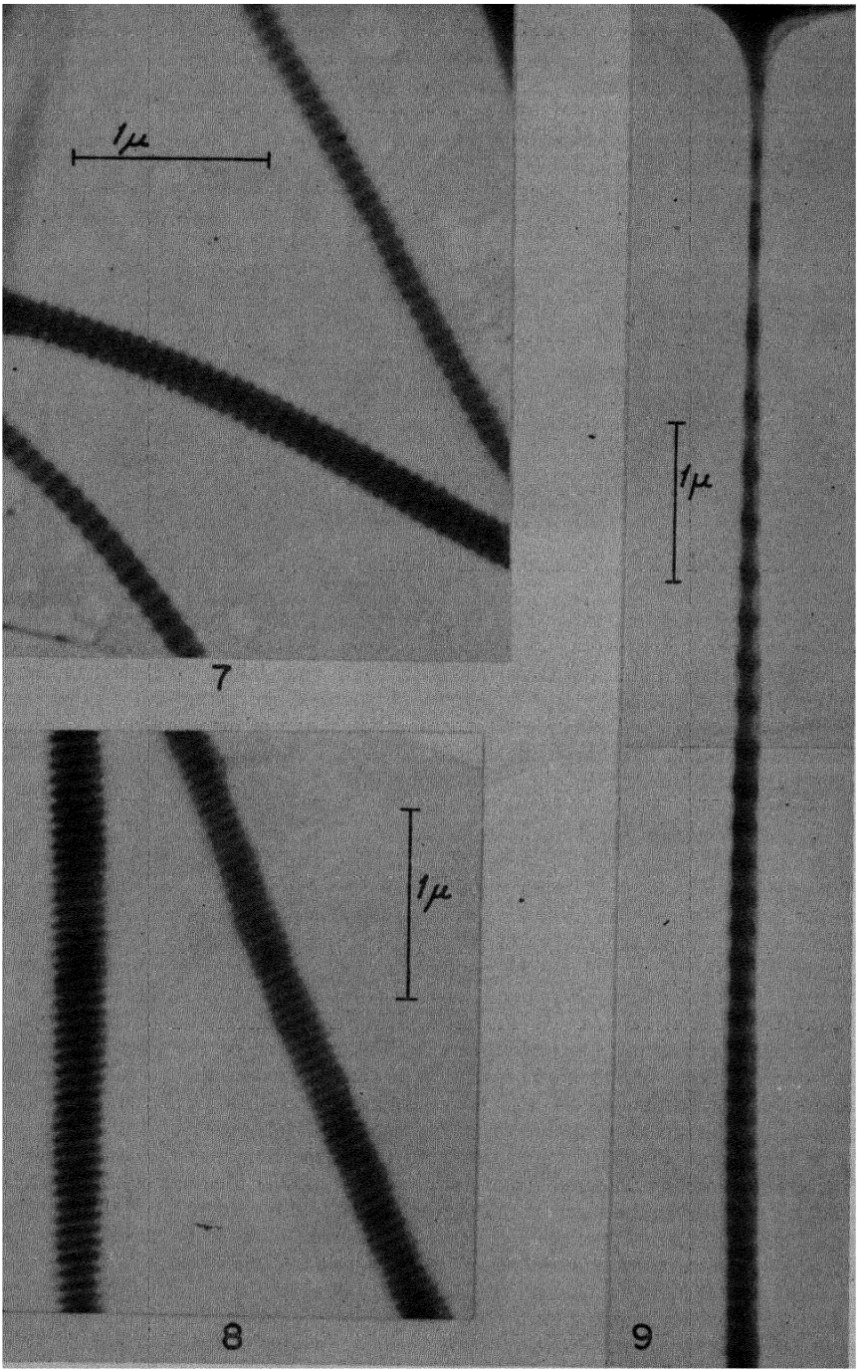


FIGURE 47. Collagen Fibrils. 7, fibrils from tail tendon of Kangaroo; 8, Beef Tendo Achilles fibrils; 9, fibril from rat tail tendon, enormously stretched by peeling back of collodion supporting film.[148]

extension than the B bands. This indicates that there is a greater degree of folding or coiling of the molecular chains in the A bands. The long-range extensibility of the fibers, revealed in the latter micrograph, is at variance with the extensibility exhibited by collagen fibers in intact tendons, in which an extension in excess of 10-15 per cent results in rupture.

"Any satisfactory theory of collagen structure must account for the above-mentioned facts. Most of the suggestions thus far offered are inadequate, since they are based on the assumption that there is little or no folding of the elementary columns. The "fringe" theory of Hermann, Gerngross and Abitz ('30, '32) pictures crystalline bundles of parallel chains which are joined together by the irregularly disposed chains into which the bundles fray. This theory is subject to certain objections such as those raised by Küntzel and Prakke ('33), and fails to account either for the regularity of the banding or for long range extensibility. Huggins ('42) proposed a structure consisting essentially of an assembly of parallel two-dimensional nets composed of spiral chains bonded to each other by NHO bridges. While this structure provides for some extensibility it offers no clues regarding the structural basis of the A and B bands.

"The experiments on very long range extensibility suggest that the folding, particularly in the A bands, may be much more extensive than that pictured by Astbury as characterizing the alpha state in keratin. It seems probable that the foldings characteristic of the A and B bands are imposed on the protofibrils by the steric requirements and interactions of the constituents amino and imino acid residues. Since the folding appears to be different in the A and B bands it is possible that the amino acid composition of the protofibrils may be different in these two regions."

"Wyckoff and Corey ('36) state that the x-ray long spacings obtained from collagen precipitated from acid solutions by addition of salt are identical with those of intact tendon. From this they conclude that collagen is not composed of bundles of indefinitely extended polypeptide chains, but rather of very large particles which are arranged in crystalline array in tendon and which retain their identity in acid solution. Addition of salt causes the particles to be recruited into a regular crystalline array characteristic of intact tendon. It was of interest, therefore, to determine whether the existence of these large particles in acid solution could be demonstrated by the electron microscope method.

"Thus far no such discrete particles have been observed in preparations made by drying acetic acid solutions of collagen (pH 3.2). From this little more can be concluded than that the protofibrils have a width smaller than the resolving power of the EM (*ca.* 50 A) or a thickness so small that the contrast is too low to make the units visible." [148]

In discussing the difference in extensibility between dried individual fibers and intact tendon, Schmitt *et al,* assume that the fibers in intact tendon have a finite length. When a tendon is ruptured, a shearing of fibers occurs before sufficient force can be exerted to demonstrate the long-range extensibility of the individual fibrils. Such a mechanism might be verified with the electron microscope by measurement of the dimensions of the frayed fibrils of a ruptured small tendon and obtaining force-extension curves for individual fibrils as demonstrated qualitatively in their paper.

References *

1. Anderson, T. F., and Richards, A. G., *J. App. Phys.*, **13**, 748 (1942).
2. —, and Stanley, W. M., *J. Biol. Chem.*, **139**, 339 (1941).
3. Ardenne, M. von, *Z. Physik*, **108**, 338 (1938).
4. —, *Ibid.*, **109**, 553 (1938).
5. —, *Ibid.*, **111**, 152 (1938).
6. —, *Z. techn. Physik*, **19**, 407 (1938).
7. —, *Z. Physik*, **113**, 257 (1939).
8. —, *Ibid.*, **114**, 379 (1939).
9. —, *Z. techn. Physik*, **20**, 235 (1939).
10. —, *Ibid.*, **20**, 239 (1939).

* An extensive bibliography has been compiled by Claire Marton and Samuel Sass, *J. App. Phys.*, **14**, 522, (1943).

11. —, *Ibid.*, **20**, 239 (1939).
12. —, *Naturwiss.*, **27**, 485 (1939).
13. —, *Ibid.*, **27**, 614 (1939).
14. —, *Z. wiss. Mikrosk.*, **50**, 8 (1939).
15. —, *Kolloid Z.*, **93**, 158 (1940).
16. —, *Naturwiss.*, **28**, 113 (1940).
17. —, *Z. phys. Chem. A*, **187**, 1 (1940).
18. —, *Naturwiss.*, **28**, 248 (1940).
19. —, *Z. Physik*, **115**, 339 (1940).
20. —, *Phys. Z.*, **42**, 72 (1941).
21. —, "Elektronen-Übermikroskopie," Julius Springer Berlin, 1940.
21a. —, *Ibid.*, p. 254 et seq.
21b. —, *Ibid.*, p. 282 et seq.
21c. —, *Ibid.*, p. 286
21d. —, *Ibid.*, p. 291
21e. —, *Ibid.*, p. 339
22. —, *Z. Physik*, **117**, 515 (1941).
23. —, *Physik. Z.*, **43**, 11 (1942).
24. —, and Beischer, D., *Z. phys. Chem. B*, **45**, 465 (1940).
25. —, Endell, K., and Hofmann, U., *Ber. deutsch. keram. Gesellsch.*, **21**, 209 (1940).
26a. Bachman, C. H., and Ramo, S., *J. App. Phys.*, **14**, 8 (1943).
26b. —, —, *Ibid.*, **14**, 69 (1943).
27a. Baker, R. F., Ramberg, E. G., and Hillier, J., *J. App. Phys.*, **13**, 450 (1942).
27b. —, —, —, *Ibid.*, **14**, 39 (1943).
28. Barnes, R. B., and Burton, C. J., *Chem. Eng. News (A.C.S.)*, **19**, 965 (1942).
29. —, —, *American Dyestuff Reporter*, 1, (May 25, 1942).
30. —, —, *Ind. Eng. Chem.*, **35**, 120 (1943).
31. Baver, L. D., "Soil Physics," p. 66, John Wiley and Sons, New York, 1940.
32. Bear, R. S., *J. Am. Chem. Soc.*, **64**, 727 (1942).
33. Beischer, D., and Krause, F., *Naturwiss.*, **25**, 825 (1937).
34. Bertram, S., *J. App. Phys.*, **13**, 496 (1942).
35. Biscoe, J., and Warren B. E., *Ibid.*, **13**, 364 (1942).
36. Boersch, H., *Ann. der Physik*, **26**, 631 (1936).
37. —, *Naturwiss.*, **27**, 418 (1939).
38. —, *Z. techn. Physik*, **20**, 346 (1939); **21**, 75 (1940).
39. —, *Naturwiss.*, **28**, 709 (1940).
40. Borries, B. von, *Z. Physik.*, **116**, 464 (1940).
41. —, and Ruska, E., *Wiss. Veroff. Siemens-Werken*, **17**, 99 (1938).
42. —, —, *Z. techn. Physik*, **19**, 402 (1938).
43. —, —, *Naturwiss.*, **27**, 281 (1939).
44. —, —, *Z. techn. Physik*, **20**, 225 (1939).
45. —, —, *Z. Physik*, **116**, 249 (1940).
46. —, —, *Erg. der exakt, Naturwiss.*, **19**, 237 (1940).
47. —, —, Krumm, J., and Muller, H. O., *Naturwiss.*, **28**, 350 (1940).
48. Brady, G. S., "Materials Handbook," p. 64, McGraw-Hill, New York, 1940.
49. Brüche, E., and Johannson, H., *Naturwiss.*, **20**, 49 (1932).
50. —, and Scherzer, O., "Geometrische Elektronenoptik," p. 66, J. Springer, Berlin, 1934.
51. Burton, E. F., Hillier, J., and Prebus, A., *Phys. Rev.*, **66**, 1171 (1940).
52. Busch, H., *Ann. der Physik*, **81**, 974 (1926); *Arch. f. Elektrot.*, **18**, 583 (1927).
53. —, and Brüche, E., "Beiträge zur Elektronenoptik," J. Barth, Leipzig, 1937.
54. Columbian Carbon Co. Research Laboratories, "The Surface Area of Colloidal Carbons," Mack Printing Company, Easton, Pa., 1942.
55. Davisson, C. J., and Calbick, C. J., *Phys. Rev.*, **38**, 585 (1931); **42**, 580 (1932).
56. Dosse, J., *Z. Physik*, **117**, 316 (1941).
57. —, *Ibid.*, **117**, 722 (1941).
58. Eitel, W., and Gotthardt, E., *Naturwiss.*, **28**, 367 (1940).
59. —, and Müller, H. O., *Ber. deutsch. keram. Gesellsch.*, **20**, 165 (1939).
60. —, and Radozewski, O. E., *Naturwiss.*, **28**, 397 (1940).
61. —, and Schusterius, C., *Naturwiss.*, **28**, 300 (1940).
62. Farr, W. K., *J. App. Phys.*, **8**, 228 (1937).
63. —, and Eckerson, S. H., *Contrib. Boyce Thompson Inst.*, **6**, 189 (1934).
64. Ford, T. F., Loomis, A. G., and Fidiam, J. F., *J. Phys. Chem.*, **44**, 1 (1940).
65. Glaser, W., p. 24 [53]
66. —, *Z. Physik*, **117**, 285 (1941).
67. Gray, F., *Bell Sys. Tech. J.*, **18**, 1 (1939).
68. Green, R. H., Anderson, T. F., and Smadel, J. E., *J. Exp. Med.*, **75**, 651 (1942).
69. Grim, R. E., "Modern Concepts of Clay Materials," *J. Geol.*, **50**, 225 (1942).
70. Hass, G., and Kehler, H., *Kolloid Z.*, **95**, 26 (1941).
71. Hall, C. E., Jakus, M. A., and Schmitt, F. O., *J. Am. Chem. Soc.*, **64**, 1234 (1942).
72. —, and Schoen, A. L., *J. Opt. Soc. Am.*, **31**, 281 (1941).
73. Heering, Gizycki, and Kirseck, *Kautschuk*, **17**, 56 (1941).
74. Heidenreich, R. D., *Phys. Rev.*, **62**, 291 (1942).
75. —, and Peck, V. G., *Ibid.*, **62**, 292 (1942).
76. —, and Peck, V. G., *J. App. Phys.*, **14**, 23 (1943).
77. Hendricks, S. B., *J. Geol.*, **50**, 276 (1942).
78. Hesz, K., and Rabinowitch, B., *Kolloid Z.*, **64**, 262 (1933).
79. Hillier, J., *Can. J. Res. A*, **17**, 64 (1939).
80. —, *Phys. Rev.*, **58**, 842 (1940).
81. —, *Ibid.*, **60**, 743 (1941).
82. —, and Baker, R. F., *Ibid.*, **61**, 722 (1942).
83. —, Baker, R. F., and Zworykin, V. K., *J. App. Phys.*, **13**, 571 (1942).
84. —, and Vance, A. W., *Proc. I. R. E.*, **29**, 167 (1941).
85. Hofmann, U., Endell, K., and Wilm, D., *Z. Krist.*, **86**, 340 (1933).
86. Humbert, R. P., *Bull. Am. Cer. Soc.*, **21**, 260 (1942).
87. —, and Shaw, B., *Soil Sci.*, **52**, 481 (1941).
88. Johannson, H., and Scherzer, O., *Z. Physik*, **80**, 183 (1933).
89. Kausche, G. A., *Naturwiss.*, **26**, 27 (1938); *Biol. Zbl.*, **54**, 194 (1939).
90. —, and Borries, B. von, *Kolloid Z.*, **90**, 132 (1940).
91. —, Pfankuch, E., and Ruska, H., *Naturwiss.*, **27**, 292 (1939).
92. —, and Ruska, H., *Kolloid Z.*, **89**, 21 (1939).

93. Kelley, O. J., and Shaw, B. T., *Soil Sci. Soc. Am. Proc.*, (In Press).
94. Klemperer, O., "Electron Optics," Macmillan, New York, 1939.
95. Knoll, M., and Ruska, E., *Z. techn. Physik*, **12**, 394 (1931).
96. —, and Ruska, E., *Ann. der Physik*, **12**, 607 (1931).
97. —, —, *Z. Physik*, **78**, 318 (1932).
98. Krause, F., p. 21.[53]
99. Kruppke, E., *Ann. der Physik*, **30**, 577 (1937).
100. Lembke, A., and Ruska, H., *Klin. Wochschrift*, **19**, 217 (1940).
101. Lenard, P., "Quantitäves über Kathodenstrahlen," Heidelberg, 1925.
102. Luria, S. E., and Anderson, T. F., *Proc. Nat. Acad. Sci.*, **28**, 127 (1942).
103. Mahl, H., *Z. techn. Physik*, **20**, 316 (1939).
104. —, *Naturwiss.*, **28**, 264 (1940).
105. —, *Kolloid Z.*, **91**, 105 (1940).
106a. —, *Z. techn. Phys.*, **22**, 33 (1941).
106b. —, *Metallwirtschaft*, **19**, 1082 (1940).
107. Mark, H., *J. Phys. Chem.*, **44**, 746 (1940).
108. Marshall, C. E., *J. Soc. Chem. Ind.*, **50**, 444 (1931).
109. —, *Soil Sci. Soc. Am. Proc.*, 1, 23 (1936).
110. —, Humbert, R. P., Shaw, B. T., and Caldwell, O. G., *Soil Sci.*, **54**, 149 (1942).
111. Marton, L., *Bull. Acad. Belg., Bruxelles*, **20**, 439 (1934); **21**, 553 (1935); **22**, 1336 (1936); *Phys. Rev.*, **58**, 57 (1940).
112. —, *Nature*, **133**, 911 (1934).
113. —, *Physica*, **9**, 959 (1936).
114. —, *Phys. Rev.*, **56**, 290 (1939).
115. —, Banca, M. C., and Bender, J. F., *RCA Review*, **5**, 232 (1940).
116. —, and Schiff, L. I., *J. App. Phys.*, **12**, 759 (1941).
117. Morton, G. A., *RCA Review*, **6**, 131 (1941).
118. —, and Ramberg, E. G., *Phys. Rev.*, **56**, 705 (1939).
119. —, and Anderson, T. F., *Proc. Soc. Exp. Biol. Med.*, **46**, 272 (1941).
120. Mudd, S., and Anderson, T. F., *J. Immunol.*, **42**, 251 (1941).
121. —, —, *J. Exp. Med.*, **76**, 103 (1942).
122. —, and Lackman, D. B., *J. Bact.*, **41**, 415 (1940).
123. —, Polevitzky, K., Anderson, T. F., and Chambers, L. A., *Ibid.*, **42**, 251 (1941).
124. Müller, H. O., and Ruska, E., *Kolloid Z.*, **95**, 21 (1941).
125. Myers, L. M., "Electron Optics," van Nostrand, New York, 1939.
126. Ott, E., *Ind. Eng. Chem.*, **32**, 1641 (1940).
127. Pfankuch, D., and Kausche, G. A., *Biochem. Z.*, **302**, 77 (1939).
128. Pickarski, G., and Ruska, H., *Arch. Mikrobiol.*, **10**, 302 (1939).
129a. Plass, G. N., *J. App. Phys.*, **13**, 49 (1942).
129b. —, *Ibid.*, **13**, 49 (1942). (bibliography)
130. Prebus, A., and Hillier, J., *Can. J. Res. A*, **17**, 49 (1939).
131. Radczewski, O. E., Muller, H. O., and Eitel, W., *Zemcnt*, **28**, 693 (1939).
132. Richards, A. G., and Anderson, T. F., *J. New York Ent. Soc.*, **50**, 147 (1942).
133. Ruska, E., *Z. Physik*, **87**, 580 (1934).
134. —, *Ibid.*, **89**, 90 (1934).
135. —, and Müller, O. H., *Ibid.*, **116**, 336 (1940).
136. —, and Ruska, H., *Klin. Wochschr.*, **17**, 921 (1938).
137. Ruska, H., *Naturwiss.*, **27**, 287 (1939).
138. —, *Kolloid Z.*, **92**, 276 (1940).
139. —, *Naturwiss.*, **28**, 45 (1940).
140. —, and Kretschmer, M., *Kolloid Z.*, **93**, 163 (1940).
141. Schaeffer, V. J., *Phys. Rev.*, **62**, 495 (1942).
142. —, *Science*, **25**, (1943) (In Press).
143. —, and Harker, D., *J. App. Phys.*, **13**, 427 (1942).
144. Scherzer, O., p. 33,[58]
145. —, p. 34 et seq.[58]
146. —, *Z. Physik*, **114**, 427 (1939).
147. Schiff, L. I., *Phys. Rev.*, **61**, 721 (1942).
148. Schmitt, F. O., Hall, C. E., and Jakus, M. A., *J. Cell. Comp. Physiology*, **20**, 11 (1942).
149. Schoon, and Koch, *Kautschuk*, **17**, 1 (1941).
150. Sears, G. S., and Kregel, E. A., *Paper Trade J.*, **114**, 43 (1942).
151. Shaw, B. T., *J. Phys. Chem.*, **46**, 1032 (1942).
152. —, and Humbert, R. P., *Soil Sci. Soc. Am. Proc.*, **6**, 146 (1941).
153. Stanley, W. M., *Science*, **81**, 664, (1935); *J. Biol. Chem.*, **115**, 673 (1936).
154. —, and Anderson, T. F., *J. Biol. Chem.*, **139**, 325 (1941).
155. —, —, *Ibid.*, **146**, 25 (1942).
156. Thomson, G. P., and Cochrane, W., "Theory and Practice of Electron Diffraction," p. 106 and p. 160, Macmillan & Company, London, 1939.
157. Vance, A. W., *RCA Review*, **5**, 293 (1941).
158. Wile, U. J., Picard, R. G., and Kearney, E. B., *J. Am. Med. Assoc.*, **119**, 880 (1942).
159. Wyckoff, R., and Corey, R. B., *J. Biol. Chem.*, **116**, 51 (1936).
160. Zworykin, V. K., *Electronics*, **16**, 64 (1943).
161. —, Hillier, J., and Snyder, R. L., *A.S.T.M. Bull.*, 15, (Aug. 15, 1942).
162. —, —, and Vance, A. W., *J. App. Phys.*, **12**, 738 (1941).
163. —, and Morton, G. A., "Television," Wiley, New York, 1940.
164. —, and Ramberg, E. G., *J. App. Phys.*, **12**, 692 (1941).

Rheological Properties of Simple and Colloidal Systems

HENRY EYRING AND RICHARD E. POWELL

Princeton University, Princeton, New Jersey

Introduction

Usually the diffusion coefficient, D, of colloidal particles in dilute solutions may be assumed to obey the Stokes-Einstein equation

$$D = \frac{kT}{6\pi r\eta} \tag{1}$$

or the appropriate modifications of Equation (1) which take into account the shapes of the particles. Here k, T, r and η are the Boltzmann constant, the absolute temperature, the radius of the particle and the viscosity of the medium, respectively. The viscosity, η, of a medium was shown by Einstein to be

$$\eta = \eta_0 (1 + 2.5\phi) \tag{2}$$

where ϕ is the volume fraction of large spherical particles in a medium of small molecules of viscosity η_0. To the approximation that these relationships hold, the theory of diffusion of large particles is reduced to the theory of the viscosity of the medium. We shall be concerned here with presenting some recent developments in the theory of viscosity of liquids, and of the diffusion of particles when the diffusing particles approach in size the molecules of the dispersing medium.

The Statistical Mechanical Theory of Reaction Rates

We only sketch this theory very briefly here, and refer the interested reader to a short exposition by Walter and Eyring,[1] to the book on the subject by Glasstone, Laidler and Eyring,[2] or to some standard text on statistical mechanics.[3]

A system of s atoms joined into a single molecule or into several molecules has a chance

$$P_i = a\, e^{-\epsilon_i/kT} \tag{3}$$

of being in a particular unique state of energy ϵ_i, provided this system is in equilibrium with its surroundings at a temperature T. The proportionality factor a is uniquely determined by the requirement that the sum of the probabilities, p_j, over all the possible states, initial and final, should add up to unity, *i.e.*, $1 = \sum_j a\, e^{-\epsilon_j/kT}$, or

$$a = \frac{1}{\sum_j e^{-\epsilon_j/kT}} \tag{4}$$

Thus for the equilibrium

$$m_1 A + m_2 B + \ldots = l_1 I + l_2 J + \ldots \tag{5}$$

we can write

$$\frac{\sum_f P_f}{\sum_i P_i} = \frac{n_I^{l_1} n_J^{l_2} \cdots}{n_A^{m_1} n_B^{m_2} \cdots} = \frac{\sum_f a e^{-\epsilon_f/kT}}{\sum_i a e^{-\epsilon_i/kT}} = \frac{\sum_f e^{-\epsilon_f/kT}}{\sum_i e^{-\epsilon_i/kT}} \tag{6}$$

or

$$K = \frac{\left(\frac{n_I}{V}\right)^{l_1}\left(\frac{n_J}{V}\right)^{l_2}\cdots}{\left(\frac{n_A}{V}\right)^{m_1}\left(\frac{n_B}{V}\right)^{m_2}\cdots} = \frac{\sum_f \frac{e^{-\epsilon_f/kT}}{V^{l_1+l_2+\cdots}}}{\sum_i \frac{e^{-\epsilon_i/kT}}{V^{m_1+m_2+\cdots}}} \equiv \frac{F_f}{F_i} \tag{7}$$

In the stoichiometric equation (5) we have indicated that m_1 molecules of type A react with m_2 molecules of type B, etc., to give l_1 molecules of type I, l_2 molecules of type J, etc. In equation (6), n_A indicates the actual number of molecules of type A in volume V for our system at equilibrium, the other n's having an analogous significance for the other types of molecules. The summation over subscript f indicates that the summation is to be over final states, while i has the same significance for initial states. This is perhaps made clearer by looking at the schematic diagram of Figure 1.

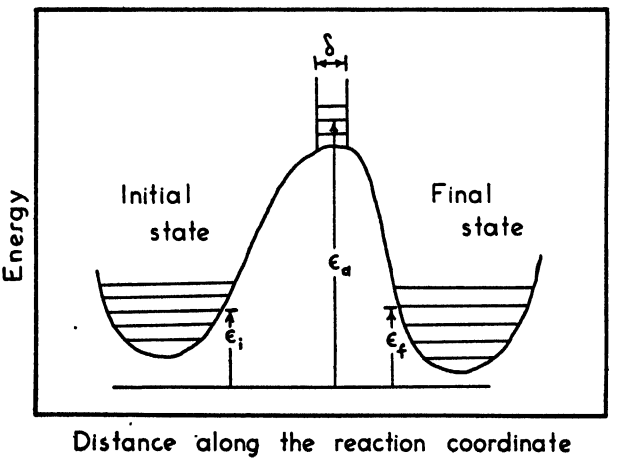


FIGURE 1. The energy relations of a reacting system.

In Equation (7), K is the equilibrium constant and $F_f = \frac{\sum_f e^{-\epsilon_f/kT}}{V^{l_1+l_2+\cdots}}$ is defined as the *partition function* at unit concentration for the final state, while F_i is the partition function for the initial state. Because ϵ_f can to a good approximation be written as the sum of $3s$ energies, one for each degree of freedom, the final partition function F_f can be written as a product of $3s$ well-known partition functions. The same is true of the initial partition function F_i.

Now this same sort of consideration enables us to write the expression for the rate of reaction. Thus if we describe an activated state by the requirement that the

system composed of s atoms have its reaction coördinate within the length δ at the top of the potential barrier, we have for the corresponding equilibrium constant

$$K_a = \frac{F_a}{F_i} \equiv \frac{F^\ddagger \frac{(2\pi m^\ddagger kT)^{\frac{1}{2}}\delta}{h}}{F_i} \equiv K^\ddagger \frac{(2\pi m^\ddagger kT)^{\frac{1}{2}}\delta}{h} \tag{8}$$

Thus the concentration of systems, c_a, lying in the length δ is

$$c_a = K_a (A)^{m_1} (B)^{m_2} \ldots \tag{9}$$

In Equation (8) the partition function for the activated state F_a can be factored into the partition function for the reaction coördinate $\frac{(2\pi m^\ddagger kT)^{\frac{1}{2}}\delta}{h}$, where h is Planck's constant and $m^\ddagger$ is the reduced mass for this coördinate, and into the partition function $F^\ddagger$ for the remaining $(3s-1)$ degrees of freedom. $(A) \equiv (\frac{n_A}{V})$ is the concentration of molecules of type A.

We now proceed exactly as though we were trying to calculate the number of automobiles passing over a hill. We multiply the concentration, c_a, of the activated systems by the mean velocity, $\sqrt{kT/2\pi m^\ddagger}$, then divide by the length δ, and finally multiply by a transmission coefficient κ which takes care of the chance that a system may turn back after passing the crest of the hill. We thus obtain the velocity of reaction

$$\text{rate} = \kappa C_a \sqrt{\frac{kT}{2\pi m^\ddagger}} \frac{1}{\delta} = \kappa K^\ddagger \frac{kT}{h} (A)^{m_1} (B)^{m_2} \ldots \tag{10}$$

Now since the specific reaction rate constant k' is defined by the equation

$$\text{rate} = k' (A)^{m_1} (B)^{m_2} \ldots \tag{10a}$$

we have by comparing Equation (10) and Equation (10a)

$$k' = \kappa \frac{kT}{h} K^\ddagger \tag{11}$$

We now apply this general expression for the rate of a reaction to the velocity of motion forward of a single molecule, or an aggregate of molecules, in a condensed phase when driven by some external potential. Since two molecules cannot simultaneously occupy the same space, it is natural to suppose that a prerequisite of a molecule's moving forward is that there be an empty space into which to move. While this is in fact true, as we shall see later, we shall develop a general formulation independent of any particular mechanism. Suppose that when no external potential is applied, a molecule on the average jumps k'' times per second, counting all possible directions, and that the average distance jumped is λ. Now if a potential is applied which does an amount of work, W, on a molecule (or aggregate) when it moves in the direction of the field through a distance corresponding to that lying between the normal and the activated states, we have for the average velocity, U, in the direction of the field

$$U = \frac{\int_0^\pi k''\lambda \cos\theta\, e^{-W\cos\theta/kT} \sin\theta\, d\theta}{\int_0^\pi \sin\theta\, d\theta} = k''\lambda \left\{ \left(\frac{kT}{W}\right) \cosh\frac{W}{kT} - \left(\frac{kT}{W}\right)^2 \sinh\frac{W}{kT} \right\} \tag{12}$$

Now when W/kT is very small we expand the exponential terms in Eq. (12) and obtain

$$U = k'' \lambda \frac{W}{3\,kT} \tag{13}$$

The condition that W/kT is small is easily satisfied for many rate processes, *e.g.*, viscous flow, diffusion, ionic conductance, and rotation of polar molecules in a fluid. A molecule in a simple liquid jumps to a new position some 10^9 times per second, and only one jump in several hundred is produced by the applied potential. It is apparent that under these conditions there will be no "saturation effects." On the other hand, such rate processes as solid friction and plastic flow have a very large value of W/kT; these processes are treated in a later section.

The result of Equation (13) has usually been obtained in a somewhat different way. Let us suppose that, when no force is acting, a molecule or aggregation of molecules jumps forward or backward k' times per second through the distance λ. The result is a zero net velocity. Now if a force is applied which does an amount of work, W, on the system when it moves forward from the normal to the activated state, it will do an amount of work $-W$ for a similar motion in the back direction. The net velocity then is

$$U = k' \lambda \left(e^{W/kT} - e^{-W/kT}\right) = k' \lambda\, 2 \sinh \frac{W}{kT} \tag{14}$$

When W/kT is much less than unity we obtain

$$U = k' \lambda \frac{2W}{kT} \tag{15}$$

This becomes identical with Equation (13) when we make the substitution $k' = k''/6$, which corresponds to saying that one-sixth of the molecules normally moving in all directions move forward. When W/kT is moderately large, Equation (12) and Equation (14) lead to slightly different results.

The Viscosity of Liquids

Viscous flow is the particular example of a transfer process in liquids where the driving potential is a shearing force, the work W being done by the force f acting upon an area $\lambda_2\lambda_3$ through a distance of $\lambda/2$.[4] The velocity equation is then

$$U = k' \lambda \frac{f\lambda_2\lambda_3\lambda}{kT} = \frac{f\lambda_2\lambda_3\lambda^2}{h} e^{-\Delta F\ddagger/kT} \tag{16}$$

Since viscosity is defined by the relation

$$\eta = \lambda_1/fU \tag{17}$$

the viscosity equation becomes

$$\eta = \frac{\lambda_1 h}{\lambda_2\lambda_3\lambda^2} e^{\Delta F\ddagger/kT} \tag{18}$$

The distances λ_1, λ_2, and λ_3 are lattice distances between molecules in the liquid. As the temperature rises, the molecular volume V/N of the liquid will increase. However, according to our present theory of liquid structure, this net expansion is due almost entirely to the presence of more and more "holes" in the liquid, the lattice expansion being comparatively small. The value of the product $\lambda_1\lambda_2\lambda_3$ may then be taken as constant and equal to the molecular volume V_s/N of the solid. The volume

V/N of the liquid is probably a somewhat poorer approximation, although the computed viscosity is quite insensitive to whether V or V_s is used.

The distance λ which a molecule jumps is certainly of the order of the other lattice distances, and if we make the approximation that $\lambda = \lambda_1$, Equation (18) takes the form

$$\eta = \frac{Nh}{V_s} e^{\Delta F^\ddagger/kT} \tag{19a}$$

or the equivalent forms

$$\eta = \frac{Nh}{V_s} e^{-\Delta S^\ddagger/k} e^{\Delta H^\ddagger/kT} \tag{19b}$$

$$\eta = \frac{Nh}{V_s} e^{-\Delta S^\ddagger/k} e^{\Delta E^\ddagger/kT} e^{P\Delta V^\ddagger/kT} \tag{19c}$$

The first requirement which Equation (19) satisfies is that it gives an exponential dependence on temperature, so that the logarithm of viscosity should plot as a straight line against $1/T$. This experimental result was found by Arrhenius,[5] de Guzman,[6] and Andrade.[7]

From the experimental data any of the thermodynamic properties of the activated state can readily be computed—for example, the free energy change of going from normal to activated states ($\Delta F^\ddagger$) is calculated from the viscosity itself, the heat change ($\Delta H^\ddagger$) from the temperature coefficient, the volume change ($\Delta V^\ddagger$) from the pressure coefficient, the entropy change ($\Delta S^\ddagger = (\Delta H^\ddagger - \Delta F^\ddagger)/T$) and the energy change ($\Delta E^\ddagger = (\Delta H^\ddagger - P\Delta V^\ddagger)$ by difference, the change in heat capacity by $d(\Delta H^\ddagger)/dT$, the change in compressibility by $d(\Delta V^\ddagger)/dP$), and so on. We may emphasize that even though the activated (flowing) molecular state has the short life-time of about 10^{-13} second, it is a real state with perfectly definite properties of its own.

It is instructive to compute the properties of a molecule in the activated state and correlate them with other properties of the liquid. The results of several such studies on viscous flow will be here summarized.[8,9,10] To facilitate comparison with other thermodynamic properties, the thermodynamic properties of activation ($\Delta F^\ddagger$, $\Delta H^\ddagger$, etc.) have been computed in calories per mole instead of ergs per molecule.

The volume change of activation $\Delta V^\ddagger$ is about ⅐ or ⅛ the volume of a molecule, for a number of simple liquids. This has been interpreted as the volume of the hole which must be provided for a molecule to flow. For molten metals the volume of the hole is about ⅛ the volume of the metal ion, in agreement with other evidence that in molten metals the flowing individual is the ion instead of the atom.

The energy change of activation $\Delta E^\ddagger$ is equal to the heat change of activation $\Delta H^\ddagger$ at normal pressures, since the pressure effect is negligible. It is about ⅛ or ¼ the heat of vaporization for a number of non-associated organic liquids. The energy required to form a hole may well be expected to be some fraction of the energy of vaporization, since the same type of bonds within the liquid are broken in both processes. If the hole were the full size of a molecule, the full energy of vaporization would be required to make a hole: it is thus apparent that a hole necessary for viscous flow is smaller than a molecule. Typical heats of activation are for ether 1.6 kilocalories, and for hexane 1.8 kilocalories.

If the liquid molecules fit extremely well in the normal state, it will be more expensive in energy for flow to occur, even after a hole has been provided. High heats of activation are then to be expected (1) in cycloparaffins and in polycyclic aromatic hydrocarbons, where the molecules are flat and therefore pack together well in the normal state, and (2) in hydrogen-bonded liquids, where these bonds must be broken before a molecule can flow. It is well known that the viscosity of lubricating oils

containing a high proportion of naphthenes and aromatics is more temperature-dependent than that of paraffin-base oils. The heat of activation for naphthalene is 6.3 kilocalories. As typical hydrogen-bonded liquids we may mention isobutyl alcohol ($\Delta H^{\ddagger} = 5.5$ kcal) and glycerol ($\Delta H^{\ddagger} = 14.8$ kcal).

When the viscosities of successively longer and longer straight-chain hydrocarbons are measured, the heat of activation does not increase without limit, but instead approaches asymptotically the value of about 8 kcal.[11] Even more striking is the viscosity behavior of linear polyesters, where the heat of activation remains sensibly constant at 8.4 kcal over a range of molecular weight of 500 to 10,000.[12] These phenomena have been interpreted as demonstrating that long molecules flow by segments of length about 20 to 25 chain atoms.[11] This segment model has also received considerable support from investigations on properties other than viscosity, such as annealing of polyamides,[13] rates of polyesterification,[14] thermodynamic properties like melting point and osmotic pressure,[15] and dielectric dispersion of polymer solutions.[16]

The entropy change of activation $\Delta S^{\ddagger}$ might be expected to be near zero for a simple liquid, since the flowing molecule is substantially the same as the normal molecule, all its degrees of freedom being unchanged except that one vibration has become a short translation. The experimental fact is that the entropy of activation for many substances is small, of the order of 0 to -3 entropy units. Those liquids which show large heats of activation (plate-like and hydrogen-bonded) also show large positive entropies of activation: naphthalene, $\Delta S^{\ddagger} = 6.7$ E.U.: glycerol, $\Delta S^{\ddagger} = 26$ E.U. These liquids are in the normal state, probably held relatively fixed by their van der Waals or hydrogen bonds. Although a flowing molecule must expend considerable energy to break such bonds, it gains a substantial amount of rotational entropy in its flowing state. In fact, the heat and the entropy changes largely compensate for each other.

The long straight-chain molecules which move by segments have an additional negative entropy term which probably arises as follows: Many of the jumps made by a segment will be unprofitable, for the molecule will progress only when all segments by chance jump in the same direction. The longer a molecule is, the less is the chance of all its segments jumping in the same direction, and this effect introduces a term $\Delta S^{\ddagger} = -$ const. $Z^{1/2}$, where Z is the chain length. The effect of chain length on the viscosity of polyesters was first found experimentally,[12] and was later deduced from theoretical considerations, even the value of the constant being estimated.[11]

As a result of the tendency for heat and entropy changes to compensate, the *free energy change of activation* $\Delta F^{\ddagger}$ shows much less irregularity than either the heat or the entropy change. It is determined almost entirely by the energy necessary to create a hole in the liquid, and may be estimated fairly closely for a wide variety of liquids by taking it as 1/2.45th of the energy of vaporization.[10]

As an illustrative example, all the thermodynamic properties of activation for viscous flow will be computed for a Pennsylvania, an Oklahoma, and a California lubricating oil. The data are those of R. B. Dow.[17] While the oils are of course mixtures, they are well enough defined to give consistent and easily interpretable results.

The logarithm of viscosity was plotted against pressure, giving good straight lines (Figure 2) from whose slope the value of $\Delta V^{\ddagger}$ was obtained (Table 1).

Table 1

	Pennsylvania	Oklahoma	California
38° C.....	50.9 cc	59.9 cc	79.0 cc
54	46.1	51.5	61.5
99	36.6	41.0	54.1

Since lubricating oils approximate a C_{35} hydrocarbon, of density 0.86, the molal volume is about 430 cc. The value of $\Delta V^{\ddagger}$ for the Pennsylvania oil at 38° C is thus about 1/8.5th the volume of a molecule, in agreement with expectation. It is interesting to mention that according to the segment theory of long molecules, a paraffin, however long, will move in segments of about 25 atoms in length. Thus the activation volume would not be expected to increase indefinitely with increasing chain length, but to approach a limiting value of approximately 50 cc. Experimental data

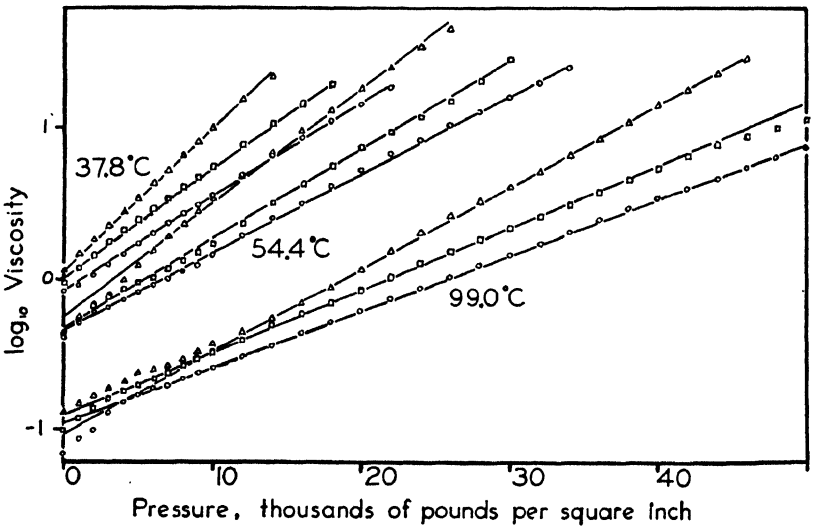


FIGURE 2. Effect of pressure on the viscosity of lubricating oils.

Circles	Pennsylvania oil
Squares	Oklahoma oil
Triangles	California oil

are not at present available to test this prediction. The value of $\Delta V^{\ddagger}$ for the California oil is particularly large, probably because of the larger activation volumes of the naphthenes (cyclic hydrocarbons) which California oil is known to contain.

It is to be noted that $\Delta V^{\ddagger}$ decreases regularly with increasing temperature. At higher temperatures there are more holes in the normal liquid, so that less have to be introduced to go to the activated state. The value of $\Delta V^{\ddagger}$ must in fact decrease steadily with rising temperature, until it becomes zero at (or somewhat above) the critical temperature, for it is well known that the viscosity of a perfect gas is independent of pressure.

The logarithm of viscosity (extrapolated to zero pressure) was plotted against $1/T$, and from the slope of the line $\Delta E^{\ddagger}$ was obtained. These values of $\Delta E^{\ddagger}$, together with the corresponding values of $\Delta F^{\ddagger}$ and $\Delta S^{\ddagger}$, are assembled in Table 2.

Table 2

	Pennsylvania	Oklahoma	California
$\Delta E^{\ddagger}$ (kcal).....	7.66	7.78	9.32
$\Delta F^{\ddagger}$ (kcal).....	7.05	7.13	8.22
$\Delta S^{\ddagger}$ (E.U.).....	2.0	2.1	3.5

Note that while the Pennsylvania and Oklahoma oils have $\Delta E^{\ddagger}$ near the limiting value for an infinitely long straight hydrocarbon, the California oil is considerably

higher than the limit—again because of its naphthenic hydrocarbons. Note also that the increased $\Delta E^{\ddagger}$ is partially compensated by an increase in $\Delta S^{\ddagger}$ for the California oil.

The foregoing studies on viscosity have led to a definite model for the process of viscous flow. The molecules in a solid are analogous to a crowd of people packed closely together: each individual must stand in his place, and can only breathe. When the solid melts to a liquid, a finite number of holes are introduced, giving an increase in volume—a man in the crowd will now occasionally have an empty position next to him, and can step into the new place. An observer watching from above could see that any one individual will move to and fro many times, in random directions, while the crowd as a whole does not move. However, if some force is applied so that the individuals move north slightly more often than south, the crowd will on the average move northward.

From this qualitative picture of liquid structure, the viscosity data have provided a quantitative measure of the size of holes, the energy necessary to create them, and the effect of such phenomena as van der Waals and hydrogen bonding. In addition, the same model of liquid structure can be applied with success to a quantitative calculation of all the thermodynamics of liquids. The application to liquid neon, argon, krypton, xenon, nitrogen, sodium, mercury and benzene has been made, and many more liquids are at present being studied. The detailed results are published elsewhere:[18,19] it is here sufficient to say that by using only the physical properties of the solid, it has been possible to calculate the melting constants, critical constants, vapor pressure, density, and heat capacity of the liquid with considerable accuracy.

Viscosity of Surface Films

The viscous flow of molecules within a unimolecular film is completely analogous to the viscous flow of molecules in bulk. The derivation of the viscosity equation proceeds exactly as in the foregoing section, so the details will not be repeated here (see [20]). One obtains the result

$$\eta = \frac{h}{a} e^{\Delta F^{\ddagger}/kT} \tag{20}$$

Equation (20) is identical with the usual viscosity equation, Equation (19), except that the area a of a molecule has now replaced the volume V_s/N of a molecule. This is of course because surface forces are defined and measured in only two dimensions.

The viscosity data for surface films are less plentiful than for liquids in bulk. The free energy of activation, $\Delta F^{\ddagger}$, has been computed for several condensed liquid films, and the values are assembled in Table 3.[20]

Table 3

Myristic acid	C_{14}	$\Delta F^{\ddagger} = 10.36$ kcal
Palmitic acid	C_{16}	10.90
Stearic acid	C_{18}	11.03
Oleic acid	C_{18}	10.36
Cetyl alcohol	C_{16}	15.3

These free energies are approximately twice as high as those for the same liquids in bulk.

The heat of activation for simple films would be expected to be very nearly the same as the free energy of activation. The heat of activation for films of arachidic acid (C_{20}) is about 15 kcal, as estimated from preliminary measurements.[21]

The surface change of activation, $\Delta a^{\ddagger}$, is the analog in films of the volume change of activation in liquids, and it is as before evaluated by plotting the logarithm of

viscosity against pressure. Its value is of the order of the surface of one molecule, or about 22 $Å^2$; the experimental values vary from 31 $Å^2$ for arachidic acid (C_{20}) to 56 $Å^2$ for palmitic acid (C_{16}).[21] These high values are to be contrasted with the low values of the volume change of activation for liquids in bulk (see page 240). We shall offer an explanation below for these and certain other anomalies.

When the pressure vs. area of certain films is investigated, an additional phase is found between the "condensed liquid" and the "solid" phases. This phase is less compressible than the liquid and more compressible than the solid, and has been called the "plastic solid."[22] The viscosity-vs.-pressure curves also exhibit marked changes at approximately the same places at which the area vs. pressure curves have breaks. From such data a phase diagram (temperature vs. pressure) has here been sketched for a film of octadecanol (Figure 3).*

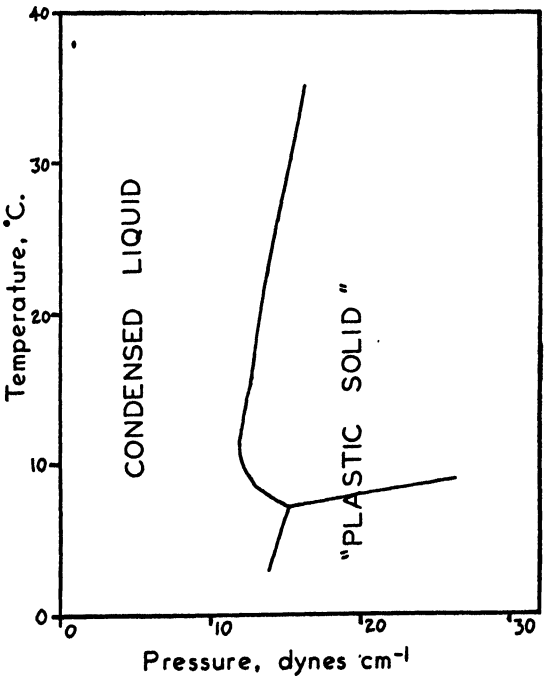


FIGURE 3. Phase diagram of an octadecanol film.

A region somewhere at the extreme right of the diagram represents the solid phase. The region at the left of the diagram represents the condensed liquid phase. The viscosity of this condensed liquid film is closely parallel to that of a normal liquid; it increases with increasing pressure and decreases with increasing temperature. The intermediate portion of the diagram represents the plastic solid phase. In the low-temperature region, this also behaves like a normal liquid, the viscosity increasing with increasing pressure and decreasing with increasing temperature. However, in the high-temperature region the viscosity behavior of the plastic solid is quite different: (1) the viscosity is sensibly independent of pressure; (2) it increases with increasing temperature; (3) the viscosity of the plastic solid (high-temperature form) may be either higher or lower than that of the condensed liquid. If it is lower, the viscosity-vs.-pressure curve has a negative slope in the region between condensed liquid and plastic solid; if it is higher, the viscosity-vs.-pressure curve is abnormally

* We are indebted to Prof. W. D. Harkins and Dr. Edward Boyd for the viscosity and pressure-area data from which Figure 3 was constructed.

steep. Similarly, if the viscosity of the plastic solid (high-temperature form) is less than that of the plastic solid (low-temperature form), the viscosity-vs.-pressure curve in the region between them is abnormally steep. An explanation for these phenomena is now to be sought.

Viscosity anomalies for liquids in bulk are known to be associated with highly organized structure. Thus the viscosity of most liquids increases with an increase in pressure; water, which is known to be hydrogen-bonded, behaves in the opposite way within a certain range of pressures: increase in pressure decreases the viscosity. One expects more of these anomalies in surfaces, with their greater structural complexities.

We shall now develop one example of such a possible complexity. Suppose that on the surface the mole fraction N_1 of the molecules are oriented in some one way, and the remaining N_2 are oriented in another way. Let ΔF be the difference in free energy between these two states. Then we have the relations:

$$N_1 + N_2 = 1 \tag{21a}$$

$$N_2/N_1 = e^{-\Delta F/kT} \tag{21b}$$

whence

$$N_1 = \frac{1}{1 + e^{-\Delta F/kT}} \tag{21c}$$

$$N_2 = \frac{1}{1 + e^{\Delta F/kT}} \tag{21d}$$

Now the two forms may differ in viscosity. If the mixture law for viscosity

$$\eta = \frac{h}{a} e^{(N_1 \Delta F_1^{\ddagger} + N_2 \Delta F_2^{\ddagger})/kT} \tag{22}$$

applies (as it does very accurately for liquids in bulk,[10] and cannot in any event be far wrong) we obtain, with a little algebraic calculation, the relations

$$RT\left(\frac{\partial \ln \eta}{\partial P}\right)_T = N_1 \Delta a_1^{\ddagger} + N_2 \Delta a_2^{\ddagger} + \frac{\Delta a}{2\left(1 + \cosh \frac{\Delta F}{kT}\right)} \frac{\Delta F_1^{\ddagger} - \Delta F_2^{\ddagger}}{kT} \tag{23}$$

$$R\left(\frac{\partial \ln \eta}{\partial 1/T}\right)_P = N_1 \Delta H_1^{\ddagger} + N_2 \Delta H_2^{\ddagger} + \frac{\Delta H}{2\left(1 + \cosh \frac{\Delta F}{kT}\right)} \frac{\Delta F_1^{\ddagger} - \Delta F_2^{\ddagger}}{kT} \tag{24}$$

Where the symbols with the superscript $\ddagger$ indicate, as usual, the change in going from the normal to the activated state, and the symbols without superscript indicate the change between species 2 and species 1. As an illustration, the logarithm of viscosity has been plotted against pressure for a hypothetical liquid with $\Delta F_1^{\ddagger} = 4000 + 0.33P$ calories, $\Delta F_2^{\ddagger} = 2000 + 0.33P$ calories, $\Delta A = 1000$ calories and $\Delta V = 100 \times 10^{-24}$ cc, in Figure 4. (The numerical values are of course arbitrary, and Figure 4 must be regarded as schematic.) It is seen that even though the viscosities of both species 1 and 2 are increased with pressure, the viscosity of the mixture may decrease with rising pressure because of the effect of pressure in shifting the equilibrium in the direction of the more fluid molecules. This produces the negative slope of Figure 4. If we had assumed species 2 to be more viscous than species 1, instead of less so, we would have obtained an abnormally rapid rise of viscosity with pressure in the region of shifting equilibrium. Entirely analogous results will be obtained for the variation of viscosity with temperature, as Equation

(24) shows. Thus we believe that the rapid variations with pressure or temperature exhibited by octadecanol films is to be explained in terms of such shifts in equilibrium between two types of surface molecules. The same explanation may be offered for the unusually high values of $\Delta a^{\ddagger}$ for the long-chain acids.

The "plastic solid" octadecanol surfaces at high temperature and pressure have a viscosity which is independent of pressure, increases with temperature, and is

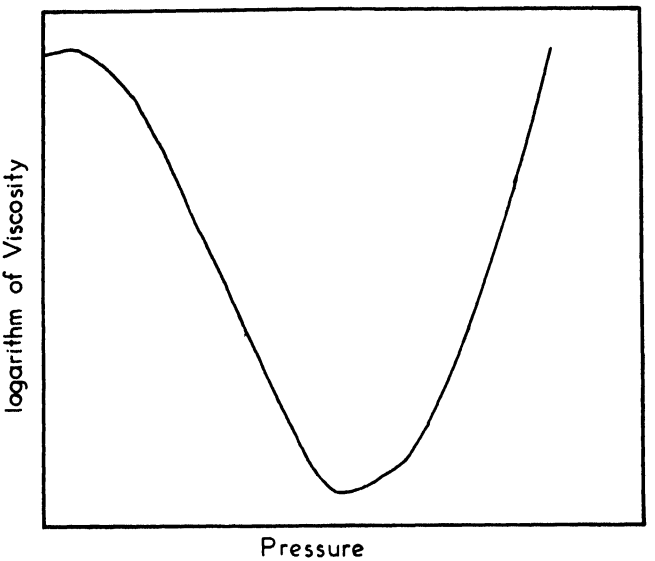


FIGURE 4. Effect of a shifting equilibrium on viscosity. Schematic.

furthermore thixotropic, *i.e.*, the viscosity decreases with increased velocity of motion of the viscometer disc over the surface. This is the type of behavior expected if at these high pressures the slippage changes from (mechanism one) that of one surface molecule past another to (mechanism two) a slippage between the surface atoms and the disc. In this case compressing the surface atoms would tend to inhibit mechanism one and bring about a shift to mechanism two, which latter would show no effect of lateral pressure on the viscosity, since the motion of film molecules will no longer be involved in the shearing process. The loss of momentum from the shear plane is now presumed to take place by momentum transfer normal to the surface, much as in the gas phase, and this loss in momentum would be expected to increase with temperature as for a gas, also to become less at high velocities since increased velocity lessens the time in which momentum can be dissipated.

The foregoing explanations, even if they should prove to be in part inapplicable to this particular example, indicate some things which must be taken into consideration in understanding the problems of surface viscosity.

Film Lubrication

Lubrication by thick films of liquids has been extensively considered by the methods of hydrodynamics, and the effectiveness of lubrication expressed in terms of the dimensionless quantity $N\eta/W$, where N is the speed, η the viscosity, and W the load.[28] A desirable lubricant for thick film lubrication is therefore one of high viscosity.

Simple liquids usually have viscosities of about 2 centipoises at their melting points. This is because in the process of melting there is an increase in entropy due

to the introduction of about 0.8 mole of new equilibrium positions (or holes) per mole of matter, giving all simple liquids about the same concentration of holes. As we have seen previously, the free energy of activation for viscous flow arises almost entirely from the necessity of having holes; hence to a good approximation substances with an equal number of holes will have an equal fluidity. As the temperature is raised there will be more holes in the liquid and the liquid will always become more fluid. Thus a suitable lubricant, having a viscosity of the order of a hundred times that of simple liquids, even at temperatures well above its melting point, is not obtainable from simple liquids by temperature changes. Likewise, viscosities greater than those of a simple liquid at the melting point are not obtainable by compression, for when the increase in pressure has depleted the number of holes to around 0.8 mole of holes per mole of molecules, the liquid will solidify no matter what the temperature. These considerations are consistent with Batschinski's rule that fluidity is proportional to volume.

Since in the liquid range the viscosities of simple liquids are less than 0.02 poise, while in the solid range the resistance to shear corresponds to enormously high viscosities, it is clear that intermediate viscosities are associated with the unstable intermediate volumes between liquid and solid.

(A) One interesting way to get high viscosity is to bring a substance into the liquid state without adding the number of holes necessary to melt it. This is done by adding to substance A a substance B which is soluble in the liquid but not in the solid A: this gives an entropy advantage to liquid A, so that fewer holes need be supplied to melt A, and it will therefore melt below its regular melting point. Only the few holes present will then contribute to the fluidity. This is one reason why the oils occurring in nature are useful lubricants: they are mixtures. It is to be pointed out that when two liquids mix with a net decrease in volume one expects —and finds—an increase in viscosity. This is because such a decrease occurs at the expense of the holes present.

(B) If the molecules of B are very large they may increase the viscosity of A also by reducing the fraction of space in which flow takes place [See Equation (2)].

(C) Long-chain molecules, by moving in segments, may introduce large entropies of activation to increase the viscosity. The corresponding viscosities will be relatively independent of temperature, which is an advantage in a lubricant.

(D) Hydrogen bonds, as in glycerol and certain other alcohols, produce high viscosities because these bonds must be broken in flow. However, the viscosity is strongly dependent on temperature.

(E) In certain colloidal solutions where the system approaches the gel state, flow may involve the shearing of secondary structures extending through the liquid.

Thus by taking sufficiently complicated systems one can obtain a viscosity in almost any range, and with some care the temperature coefficient can be kept reasonably low.

Boundary Friction and Boundary Lubrication

When two unlubricated surfaces are moved past each other, the ratio of the force required to maintain motion to the force pressing the surfaces together is a constant, called the coefficient of friction, or $\phi/Wg = \mu$. This law was discovered by Amontons in 1699.[24]

The rate of motion may be treated by the method of absolute reaction rates in the following way:[19]

The surfaces of even optically smooth metals have irregular hills and valleys some 10^{-5} cm—500 to 1000 atom diameters—deep. In addition, the surface has smaller periodic valleys, one between each atom of the metal.[25] A schematic diagram of such a surface is given in Figure 5.

A second surface is pulled over the first with a velocity of the order of 1 cm sec^{-1}. Since the distance between equilibrium positions is about 10^{-8} cm, the time spent in each potential minimum is about 10^{-8} sec. The vibration frequency of an atom in a solid is about 10^{13} sec^{-1}; therefore the atom vibrates some 10^5 times in each minimum and is able to dissipate as heat all of its activation energy. Thus the activation energy must be furnished anew for each distance λ the surfaces move relative to each other. (We shall consider later the case where the surfaces move extremely rapidly.)



FIGURE 5. The surface of a solid.

We can calculate the work $\Delta F^{\ddagger}$ which must be supplied for the upper surface to move to the top of a potential hill, a distance forward of $\lambda/2$. Let the upper block press down with a force Wg. The block rises an amount s due to the atomic irregularities, and the work done is Wgs. The upper block also rises an amount l due to the gross surface irregularities, and the work done is Wgl; however, this work will be on the average positive and negative equally often, so its net effect will be small. In the third place, the moving block must overcome an energy of attraction, part of which is proportional to the weight—those bonds which are produced by plastic deformation of the surface: the work done may be written Wgr. Finally, there are work terms independent of the weight—bonds which exist when the weight is zero, non-linear terms for the elastic bonds, the work necessary to tear off surface atoms; such terms are collected into $\Delta F_0^{\ddagger}$. The total work is now

$$\Delta F^{\ddagger} = \Delta F^{\ddagger}_0 + Wg\,(s + l + r).$$

The general rate expression is

$$U = \lambda \frac{kT}{h} e^{-\Delta F^{\ddagger}/kT} \left(e^{\phi\lambda/2kT} - e^{-\phi\lambda/2kT}\right) \tag{25}$$

$\phi\lambda/2kT$ is an extremely large number (about 10^{16}) so the second term in parentheses is negligible. If the velocity is 1 cm sec^{-1}, λ is about 10^{-8} cm, kT/h is 6×10^{12} at room temperature; then

$$\frac{\phi\lambda}{2kT} = \frac{\Delta F^{\ddagger}}{kT} - 11.5 \tag{26}$$

It is seen that the term 11.5 is also negligible. The coefficient of friction is then independent of temperature:

$$\mu = \frac{\phi}{Wg} = \frac{2\,\Delta F^{\ddagger}}{\lambda\, Wg} = \frac{2}{\lambda}\left(\frac{\Delta F_0^{\ddagger}}{Wg} + s + l + r\right) \tag{27}$$

For two surfaces not actually welded together, $\Delta F_0^{\ddagger}$ may well be negligible. Then

$$\mu = \frac{2(s+l+r)}{\lambda} \tag{28}$$

Equation (28) shows that the important term in the coefficient of friction is $2s/\lambda$, or the ratio of the rise to the horizontal distance moved in going from the normal to the activated state. One might well surmise the actual result that μ is not very different from unity.

We now return to the problem of extremely rapid slippage. If the upper block were moved rapidly enough it could make only a few vibrations in each minimum, and so could not dissipate much of its activation energy. The system would then need to be brought only to the top of its first barrier, after which it could coast over all successive barriers. The critical speed is of the order of the speed of sound in solids, or about 5000 meters per second. A high explosive which produces a detonation wave with such a speed will demolish any material, because after the first unlocking along a shear plane the solid is torn apart extremely efficiently. The detonation velocities of several high explosives are listed in Table 4.[26]

Table 4

Mercury fulminate	3920 meters per sec.
Nitromannite	6900-7700
Nitrocellulose	3700-6200
Picric acid	7700-8200
Trinitrotoluene	7200
Gelatin dynamite	7000
Ammonium nitrate explosives	2700-3700

In contrast, if explosives are to be useful propellants, their velocity must be of the same magnitude as that of the projectile. For example, the velocity of explosion of black powder is 300 meters per second: the flame velocity in an internal combustion engine is of the same magnitude as the piston velocity (some tens of meters per second), a too rapid explosion producing "knocking," as is clearly shown in photographs of knocking and non-knocking flames in a gasoline engine.[27]

When the two surfaces are lubricated with a monomolecular layer of primary alcohols, paraffins, carboxylic acids, or secondary alcohols the coefficient of friction is [28]

$$\mu = \mu_0 - d - c(n-1) \tag{29}$$

where μ_0 is the coefficient for unlubricated surfaces, d a decrease in coefficient which is characteristic of the lubricant and independent of chain length, and $c(n-1)$ gives a decrease characteristic of the lubricant and proportional to chain length; c is independent and d almost independent of the kind of surface material. Table 5 presents some values for these parameters compiled by Adam.[29]

Table 5

Surface	Hydrocarbon			Primary Alcohol		Secondary Alcohol		Acid	
	μ_0	c	d	c	d	c	d	c	d
Glass	.94	.021	.15	.023	.26	.027	.39	.059	.19
Mild steel	.79	.021	.23	.023	.33	.027	.37	.060	.18
Alloy steel	.88	.022	.24	.023	.35			.060	.22
Med. carbon steel	.83	.022	.23	.023	.33			.060	.19
Nickel chrome steel	.93	.022	.23	.023	.33			.060	.23
Phosphor bronze	.94	.022	.22	.023	.34			.060	.24

A lubricant may reduce the cofficient of friction by reducing $\Delta F_0^{\ddagger}$, s, l, r, or by increasing λ. The reduction independent of length, d, probably arises from a decrease in $\Delta F_0^{\ddagger}$, for hydrocarbons make very poor van der Waals bonds with each other—much poorer than metal or glass. The reduction proportional to chain length probably arises as follows: In the unlubricated surfaces a large number of surface atoms "mesh," which causes s, the mean vertical displacement of unlocking, to be large. A tightly adsorbed layer of a monatomic substance would not improve the situation. However, a longer molecule with one end tightly anchored to the surface can still orient the rest of the molecule randomly, will prevent meshing and so decrease s. Furthermore, if each additional chain atom has as much chance of hindering the meshing as did the previous chain atom, the observed linear decrease of μ with chain length would be understandable. When the hydrocarbon becomes long enough to move in segments (longer than 20-25 atoms) the coefficient will be reduced to that for the liquid, *i.e.*, approximately zero. The decrease per chain atom for short chains should also be about $\frac{1}{20}$ or $\frac{1}{25}$ of the original coefficient. The observed value is 0.023 for hydrocarbons and 0.06 for acids.

Plasticity

It is frequently stated that plastic deformation sets in at some definite value of the applied stress. While practically speaking this is about right, it is often misunderstood. It is exactly analogous to the statements that compounds start decomposing at a certain temperature or that at a definite overvoltage current starts flowing. What should be said in each of the three cases is that at a certain value of the stress, the temperature, or the applied potential the phenomenon becomes noticeable with the particular instrument one has chosen for observing the change. Progressively more accurate observations would reveal that the change sets in at progressively lower values of the disturbing influence. Thus for the velocity of a shear displacement we have already seen in Equation (14) that

$$\mu = \lambda \frac{kT}{h} e^{-\Delta F^{\ddagger}/kT} \left(e^{-\phi\lambda/2kT} - e^{-\phi\lambda/2kT}\right) \tag{30}$$

Now if the free energy of activation for shear is large the applied force ϕ must be large before U will have a noticeable value. This insures that the second term in parentheses is negligible. But now we see that the velocity of displacment depends exponentially on the applied force ϕ, so that observable motion will set in fairly suddenly at some critical value of the applied force, ϕ_c. Because the first observable velocity U_c is only a few orders of magnitude smaller than $\lambda kT/h$, the quantity $(\Delta F^{\ddagger} - \phi_c\lambda/2)$ must be quite small. This insures that

$$\phi_c = \frac{2\,kT}{\lambda} ln\left(\frac{U_c h}{\lambda\, kT}\right) + \frac{2\Delta F^{\ddagger}}{\lambda}$$

will depend very little (calculated on a percentage basis) upon the temperature.

The same general theory and type of treatment as that we have applied to solid friction and to plasticity lead to analogous relations for overvoltage or critical decomposition potential.

Diffusion

As was pointed out in the beginning of this article, the theory of diffusion of large particles reduces to a theory of the viscosity of the surrounding medium: so that by treating viscosity we have completed at least one phase of the diffusion problem.

The question as to what happens when the diffusing particles are of the same size as the particles of the surrounding medium still requires consideration. The

mechanism of flow within the liquid will in general be the same whether the flow neutralizes a shearing stress, a concentration gradient, or an electrical potential gradient. We have already seen that the velocity of forward motion of a particle is

$$U = \lambda k' \frac{2W}{kT} \tag{15}$$

The definition of the diffusion coefficient is $-D =$ the net number of molecules passing through a square centimeter of surface per second, divided by the concentration gradient normal to this surface, *i.e.*,

$$D = -\frac{Un_1}{dn_1/dx} \tag{31}$$

where n_1 is the number of particles, of the kind diffusing, per cc. The partial molal free energy $\overline{F_1}$ (or chemical potential) of a particle can be written as $\overline{F_1} = -kT \ln a_1$, where a_1 is the activity. Then the work done in going from the normal to the activated state is

$$W = \frac{\lambda}{2}\frac{d}{dx}(-kT \ln a_1) = -\frac{\lambda}{2} kT \frac{d \ln a_1}{d \ln n_1} \frac{d \ln n_1}{dx} \tag{32}$$

Substituting Equation (32) and Equation (15) into Equation (31),

$$D = \lambda^2 k' \frac{d \ln a_1}{d \ln n_1} = \lambda^2 \frac{kT}{h} e^{-\Delta F^\ddagger/kT} \frac{d \ln a_1}{d \ln n_1} \tag{33}$$

For a perfect solution, $d \ln a_1/d \ln n_1$ is of course unity. Remembering that the equation for viscosity is

$$\eta = \frac{\lambda_1 h}{\lambda_2 \lambda_3 \lambda^2} e^{\Delta F^\ddagger/kT} \tag{18}$$

we have

$$D\eta = \frac{\lambda_1 kT}{\lambda_2 \lambda_3} \frac{d \ln a_1}{d \ln n_1} \tag{34}$$

This expression, for diffusing particles of the same size as the surrounding solvent, is to be compared with the appropriate Stokes-Einstein relation for large particles,

$$D\eta = \frac{kT}{6\pi r_2} \frac{d \ln a_1}{d \ln n_1} \tag{35}$$

Equation (34) and Equation (35) differ only by a numerical factor, which has to do with the difference in the lines of flow around small and large particles. The intermediate size range requires further investigation.

The applicability of these equations has been considered in detail elsewhere,[10] so we give here only one typical example of the application of Equation (34) to the mixture of chloroform and acetone. Figure 6 illustrates the expected simplification in the curve for $D\eta$ against composition when one takes into account the imperfection of the solution by dividing by $d \ln a_1/d \ln n_1$.

If a particle is diffusing under an electrical potential as well as a chemical potential, the extra work done by the electrical field is included in the expression for W to be substituted into Equation (15). In this way one obtains all the usual equations for ionic conductance.[2]

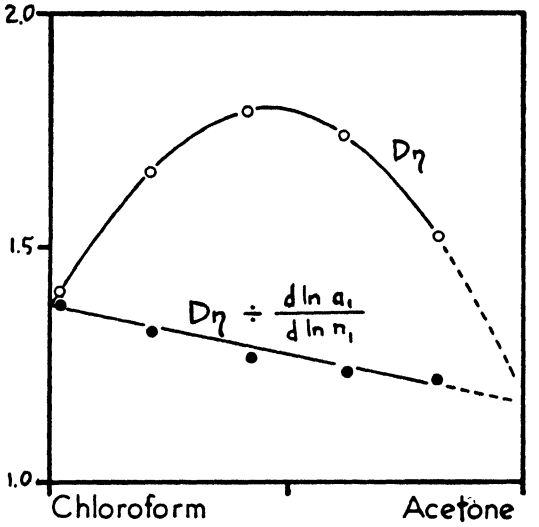


FIGURE 6. Effect of activity on diffusion.

References

1. Walter, J., and Eyring, H., *J. Chem. Ed.*, **18**, 73 (1941).
2. Glasstone, S., Laidler, K., and Eyring, H., "The Theory of Rate Processes," McGraw-Hill Book Co., 1941.
3. Fowler, R. H., "Statistical Mechanics," Macmillan, 1936.
4. Eyring, H., *J. Chem. Phys.*, **4**, 283 (1936).
5. Arrhenius, S., *Meddel. Vetensk. Nobelinst.*, **3**, 20 (1916).
6. Guzmán, J. de, *Anales soc. espan. fis. quim.*, **11**, 353 (1913).
7. Andrade, E. N. da C., *Phil. Mag.*, **17**, 497, 698 (1934).
8. Ewell, R., and Eyring, H., *J. Chem. Phys.*, **5**, 726 (1937).
9. Frisch, D., Eyring, H., and Kincaid, J., *J. App. Phys.*, **11**, 75 (1940).
10. Roseveare, W. E., Powell, R. E., and Eyring, H., *J. App. Phys.*, **12**, 669 (1941); *Ind. Eng. Chem.*, **33**, 430 (1941).
11. Kauzmann, W., and Eyring, H., *J. Am. Chem. Soc.*, **62**, 3113 (1940).
12. Flory, P. J., *J. Am. Chem. Soc.*, **62**, 1057 (1940).
13. Fuller, C. S., Baker, W. O., and Pape, N. R., *J. Am. Chem. Soc.*, **62**, 3275 (1940).
14. Baker, W. O., Fuller, C. S., and Heiss, J. H., *J. Am. Chem. Soc.*, **63**, 2142 (1941).
15. Powell, R. E., Clark, C. R., and Eyring, H., *J. Chem. Phys.*, **9**, 268 (1941).
16. Bridgman, W. B., *J. Am. Chem. Soc.*, **60**, 530 (1938).
17. Dow, R. B., *J. App. Phys.*, **8**, 367 (1937).
18. Walter, J., and Eyring H., *J. Chem. Phys.*, **9**, 393 (1941).
19. Eyring, H., Walter, J., and Stearn, A. E., *Science* (1942) (in publication).
20. Moore, W., and Eyring, H., *J. Chem. Phys.*, **6**, 391 (1938).
21. Harkins W. D., and Boyd, E., *J. Am. Chem. Soc.*, **61**, 1188 (1939).
22. Harkins, W. D., and Boyd, E., *J. Chem. Phys.*, **7**, 203 (1939).
23. Hersey, M. D., "Theory of Lubrication," John Wiley and Sons, 1936.
24. Amontons, *Mem. Acad. Roy. Sci.* (1699) 206.
25. Derjaguin, B., *Z. Physik.* **88**, 661 (1934).
26. Brunswig, H., "Explosives," p. 98, John Wiley and Sons, 1912.
27. Withrow, L., and Boyd, T. A., *Ind. Eng. Chem.*, **23**, 539 (1931).
28. Hardy, W. B., and co-workers. See ref. 29, p. 227. Also Vol. I, this series, p. 288-305.
29. Adam, N. K., "The Physics and Chemistry of Surfaces," Oxford Press, 1938.

High-Vacuum Distillation *

K. C. D. HICKMAN

Distillation Products, Inc., Rochester, N. Y.†

Distillation can be done at successively lower temperatures as the pressure is reduced until a limit is reached below which continued reduction makes no alteration. It is then found that if the apparatus is redesigned to dispense with the usual links between evaporator and condenser, a further lowering of temperature can be secured. Finally, a new limit is reached beyond which no device yet known will reduce the temperature of distillation. This limit is realized in apparatus called a "molecular" or "unobstructed path," high-vacuum still. Such stills, which operate best with pressures of residual gas of less than $\frac{1}{1000}$ of a millimeter, require a specialized technique of vacuum pumping. The combination of short-path still and high-vacuum pumps provides a medium for distilling labile substances with less decomposition than previously thought possible. Indeed, many substances which had been considered as fixed liquids or fixed solids have now become readily distillable. The natural fats, oils and waxes of vegetable and animal origin, many of the vitamins, the sterols, hormones, some of the simpler dyes and a large range of pharmaceutical compounds are examples.

There are so many steps in the progression from the ordinary high-vacuum still to the most efficient molecular still that the description will include some of the more useful pieces of intermediate apparatus. The steps are pictured in Figure 1 in reverse order. Here *A* represents a drop of liquid suspended freely in the center of a glass flask evacuated through the side tube *Z*. The vacuum is assumed to be so high that a molecule leaving the central drop will reach the wall without experiencing any collisions on the way. In fact, the molecule will not "know" that there are any other molecules present; its flight will be free and unobstructed. In such hypothetical apparatus, distillation will occur entirely unhindered and, therefore, at the lowest possible temperature. A practical interpretation is the small pot still *B* where the liquid rests on the bottom of a glass alembic. The bottom is warmed; part of the vapor, as shown by the cone of dotted lines, reaches the cooled ceiling where it condenses and runs out of the tube *X*. The vacuum is applied at *Y*. Such a still is only partially unobstructed since the walls condense and return to the *distilland* about half of the evaporated molecules.

Figure 1C shows a simple near approach to the unobstructed still that can be made with the minimum of glass blowing skill from an ordinary wide-necked flask. The neck should be extended considerably and at the joint a small bulb is blown which is then pushed in to form the alembic *V*. The apparatus is evacuated by side arm *W* and distillate is collected through the tube *T*. A more ambitious piece of apparatus, but one working at considerably higher pressure, is the high-vacuum refluxing still *D*. Here a large flask *S* is attached to the long neck *R* which is bent in

* "Molecular distillation," as it has been termed, gives a method for isolation of many kinds of molecules of high molecular weight, some of which approach or enter the field of colloidal dimensions, and are not easy to recover otherwise. EDITOR

† Communication No. 40 from the laboratories of Distillation Products, Inc.

two places, r^1 and r^2. Experience has shown that bends placed at these positions prevent spray reaching the condenser and at the same time accommodate wells for thermometers. The condensate is collected in the air-cooled portion *O* and withdrawn by the tube *P*.

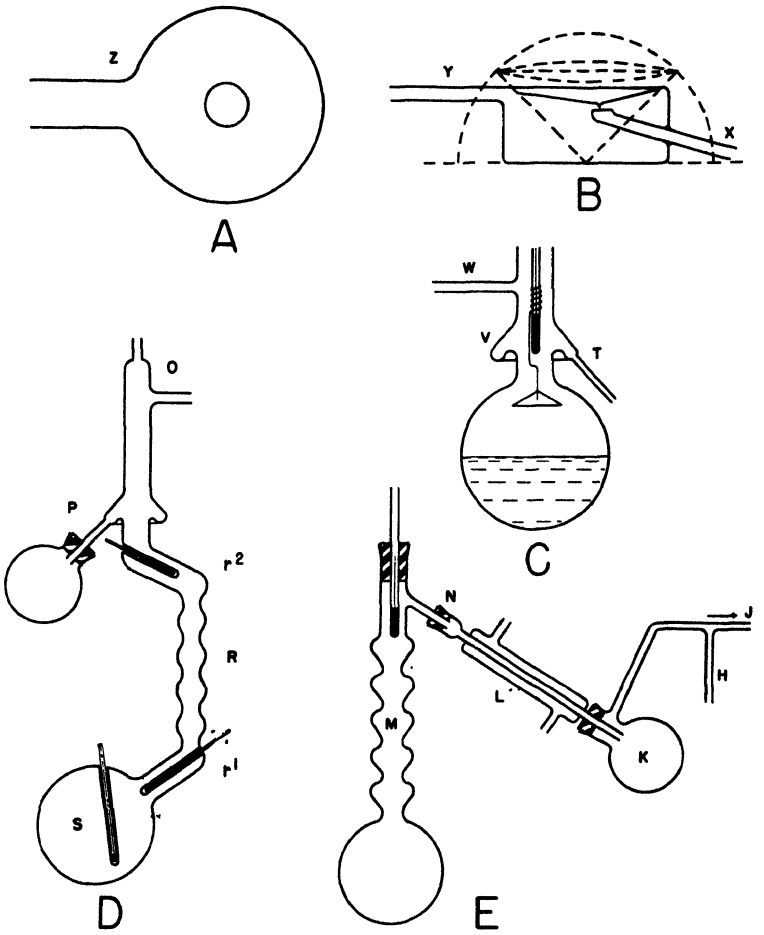


FIGURE 1
A—A drop of fluid evaporating freely in all directions in high vacuum
B—Pot still with fluid evaporating freely only in directions shown by cone
C—High-vacuum pot still of non-molecular type
D—High-vacuum fractionating still
E—Ordinary laboratory vacuum fractionating still

From still *D* it is a short retrograde step to still *E*, which is of the type in common use in organic laboratories. A flask and fractionating column are attached by side arm *N* to condenser *L* and receiver *K*, manometer *H* and vacuum pump *J*. The probable pressures existing in each region during operation are as follows: *M*, 4 mm; *N*, 2mm; *L*, 1 mm, and *K*, 0.05 mm. The progressive increase from the pump to the flask shows the resistance to the outgoing vapors which is imposed by the construction of the vessel, and it is an illustration of the advantage of using apparatus ap-

proximating the simpler diagram *A* when the lowest temperature distillation is required. However, it should be realized that still *E* gives considerable fractionation and therefore has a separating power of more than one theoretical plate. Under ordinary conditions, when it is operated at perhaps 10-millimeter boiler pressure, a separation equal to 3 or 4 theoretical plates may be secured. Even with complicated distilling columns, separation comparable with stills of the Podbielniak type is seldom reached. All the stills *D, C,* and *B,* provide fractionation in lesser degree, even still *B* returning about half the distillate for redistillation. This demonstrates that short-path distillation is undertaken at the expense of sharpness of separation and there is, therefore, every reason *not* to use a short-path still except when the perishability of the substance under treatment makes it necessary.

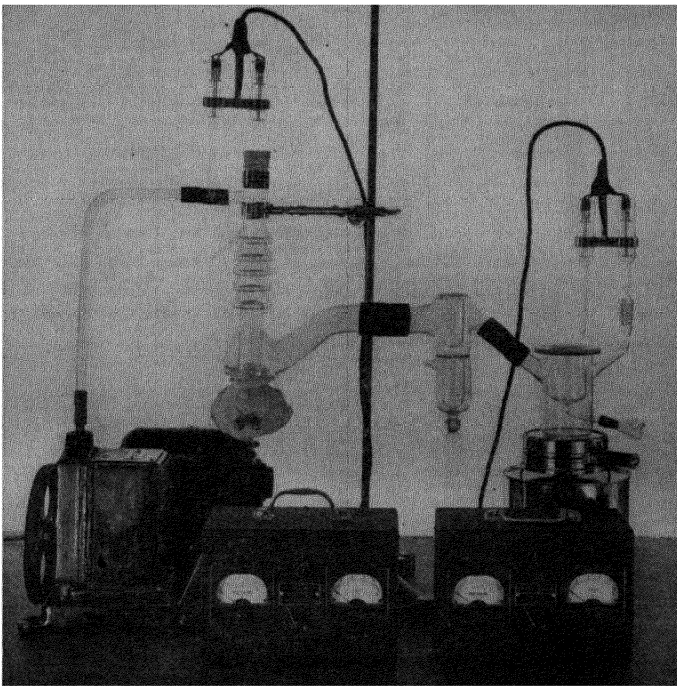


FIGURE 2. Molecular pot still with pumps, trap, and high-vacuum gauges.

Before proceeding with a detailed development of the subject, it is considered best to familiarize the reader with the general appearance and technique of the unobstructed-path distillation assembly. In Figure 2 the pot still shown on the right is fitted with a detachable re-entrant cover and connections to the receiver for distillate and to a high-vacuum gauge. This is conveniently a Pirani hot-wire instrument, the indications of which are shown in the left-hand meter of the right-hand box. Material to be distilled is placed in the flask and warmed by a small electric heater. It is convenient if a shallow beaker of castor oil is interposed between the flask and heater to keep the temperature uniform and the walls warm. Between the flask and the high-vacuum pump is placed a solvent trap which may have a vacuum-jacketed base to retain without serious re-evaporation any liquid which condenses at the temperature of the Dry Ice or other refrigerant used. The condensation pump

need not be an elaborate piece of apparatus, a single-stage pump filled with butyl phthalate or Octoil being adequate if the tubes between the still, trap and pump are sufficiently wide. The condensation pump in turn is attached by a wide tube to a mechanical pump which should have an effective operating capacity of 100-500 cc per second at 0.08 mm of mercury (the pressure is measured by the gauge on the left). It is more important that the mechanical pump shall have adequate volumetric capacity than reach very low pressure. Accordingly, if the choice is to be made between a single- and a double-stage oil-immersed pump of approximately the

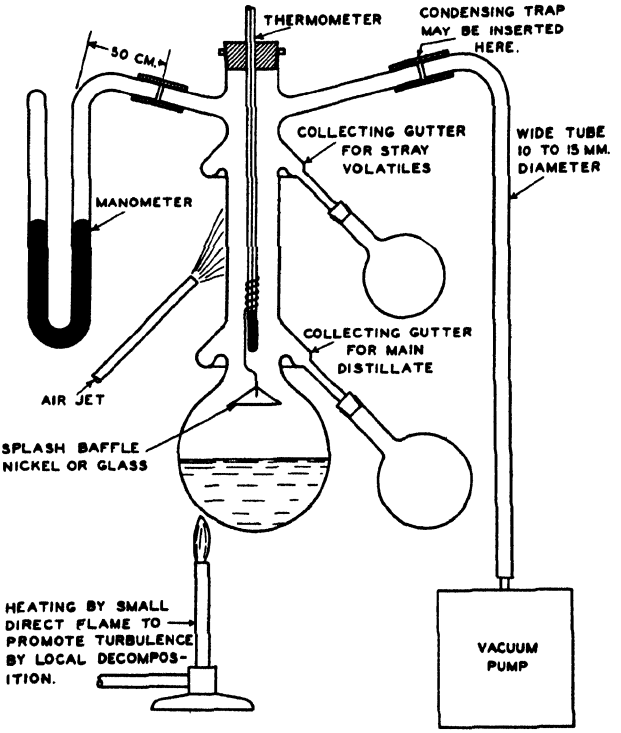


FIGURE 3. High-vacuum pot still assembly.

same cost, the former should be chosen, but should be supplied frequently with fresh oil. The higher volumetric capacity will enable the apparatus to be brought under vacuum quickly.

To operate the apparatus, the material is placed in the alembic or still and is warmed gently to melt or reduce its viscosity. Vacuum is applied cautiously by operating the mechanical pump intermittently, and in this way the material can be carried through the eruptive part of the degassing treatment without seriously splashing the lid of the alembic. Should such splashing be troublesome, air may be admitted momentarily, the detachable lid removed and cleaned, and the apparatus again brought under vacuum. The second degassing will not be as violent as the first if the change is made rapidly.

A useful compromise, according to the scheme C in Figure 1, is shown in greater detail in Figure 3. Here the material to be distilled is held in a wide-necked flask with a low alembic on the neck. Another alembic much higher up is used for catching stray volatile materials. The whole of the neck of the flask is cooled by air

blast. A clean mercury manometer is attached directly to one side of the apparatus, while the other side is connected through a Dry Ice trap to the vacuum pump. Stirring of the contents is induced by a small pin point flame. Such apparatus will give very satisfactory separation of many materials which otherwise could only be distilled in the molecular pot still.

Historical

The researches that have led up to the modern highly specialized industrial or research-type molecular stills are buried in obscurity, but Detwiler [1] has discovered some early references.[2] Perhaps the first scientific analysis of short-path distillation conditions was made by Brönsted and von Hevesy [3] during their work on the separation of the isotopes of mercury. In the early 1920's Waterman [4] and collaborators in Holland made many experimental low-temperature distillations using semi-unobstructed-path apparatus. In the same decade, C. R. Burch built an unobstructed-path tray [5] still which he attached to the newly invented Langmuir condensation pump.[6] He was able to separate out from "undistillable" Vaseline a number of comparatively fluid oils of extremely low volatility. These materials had many interesting properties and could be used, for instance, to replace mercury in the condensation pump where they produced in one step extremely low pressures of the order of 10^{-5}, 10^{-6} mm without the interposition of a cooled trap. Burch's laboratory experiments were transmuted into a self-contained industrial technique which was set forth in British [7] and other patents.[8]

Experiments similar to Burch's and Waterman's were started in other places, notably the Bureau of Standards at Washington, D. C., by Washburn [9] *et al.*, and at the Kodak Research Laboratories [10] in Rochester.

Improvements in the Molecular Still

It was soon apparent that there are two interrelated factors influencing thermal decomposition of a labile substance: temperature and time. There is an indirect factor—viscosity—which has a marked effect, since with substances of high molecular weight such as are generally treated in the molecular still there is very poor stirring of the distilland, and the distilling surface becomes impoverished and no longer represents the composition of the main body of the liquid. These factors—temperature, time and viscosity—make it desirable to operate with very thin films of distilland. Burch and Bancroft [8] devised tray stills in which the distilland wandered from one tray to another, thus providing a thin layer under moderate turbulence. A more important development and one which has been universally adopted is the falling-film still.[11, 12] Here the distilland is distributed around the sides of a vertical cylinder down which it flows in a fairly uniform stream, the uniformity depending on the exactness with which the liquid is distributed on the top of the column, on the cleanness of the column, its orientation and on surface tension changes occurring during distillation. The falling-film column is maintained at high temperature and is housed within a cooled cylindrical condenser; the space between the column and condenser is evacuated so highly that the evaporating molecules meet no obstructing particles on their journey to the condenser.

A diagram of a laboratory falling-film still is shown in Figure 4. The distilland is held in the reservoir *A* and is admitted through a drip feed or sight glass *B* through an adjustable stopcock to the preheating tube *C*. It emerges into the "explosion" bulb *D*, where much of the dissolved gases and vapors are given off to be drawn away by a small auxiliary vacuum pump. The partially degassed liquid runs down a wide spiral tube *F* where the temperature is increased and more gas is drawn away to the main vacuum pump. The liquid then passes on to the hot falling-film column *H*, which may have a plain surface or an embossed pattern, or may be given a ground-glass finish as suggested by Markley [13] and Detwiler. The column is

housed within the condenser *K*, evacuated by means of a condensation pump *G*. The distillate collects in the small flask to the right, and residue passes to reservoir *L*. The column *H* may be warmed by electrical winding or by the circulation of hot fluid. In the present case it is warmed by vapor supplied by the flask at the base of the picture. This apparatus shows no vacuum meter other than the McLeod gauge attached to the trap. Various laboratory falling-film stills have been described by Waterman,[4] Markley,[18] and others. The control board and one of the distillation units in a larger industrial falling-film still [16] is shown in Figure 5. This happens to

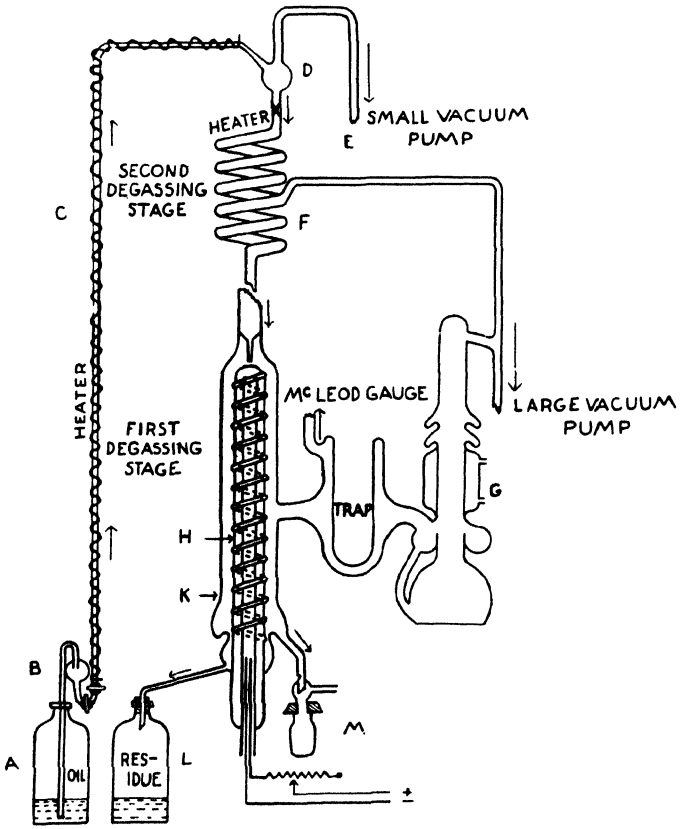


FIGURE 4. Laboratory falling-film still.

be the fifth and last unit in a series of stills each unit of which accommodates 200 gallons of oil per day. The metal column has a finely embossed pattern for distributing the distilland and is covered by a bell jar. The vacuum is applied at the base which also serves as a support. Very wide ducts and condensing channels are provided so that there shall be little obstruction to the residual gas which emerges.

Centrifugal Still

The duration of distillation, which was 1-2 hours in the pot still, was reduced to about 10 minutes and 30 seconds in the industrial and laboratory falling-film stills, respectively. There remained much room for improvement both in evenness of distribution and time of exposure of the distilland. Both factors have been improved

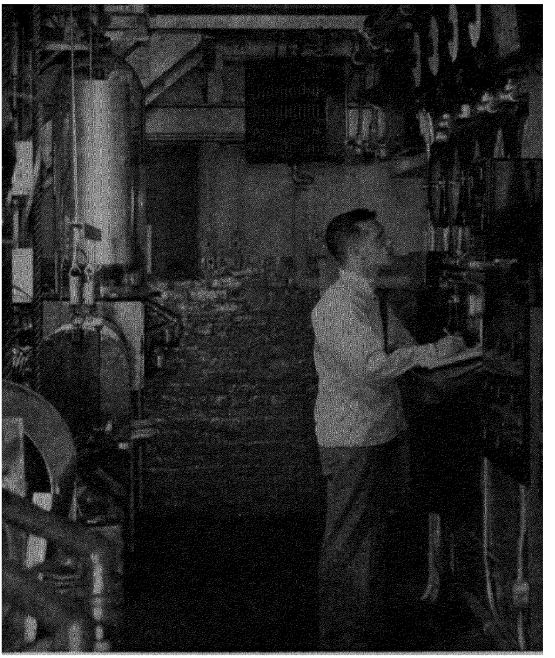


FIGURE 5. Fifth unit and controls of a commercial molecular still installation. The falling-film still is seen under the bell jar to the left, and the thermometers and heat controls to the right.

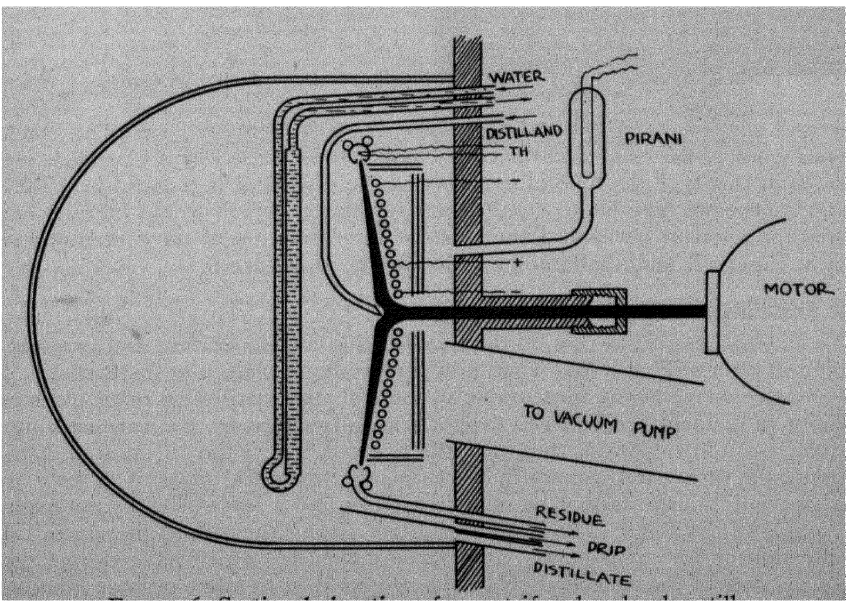


FIGURE 6. Sectional elevation of a centrifugal molecular still.

in a still fitted with a rotating distilling surface.[16] A diagram of the centrifugal still is shown in Figure 6, which presents a section in elevation. The apparatus is mounted on a metal base plate, shown as the vertical crosshatched portion which carries a bearing and stuffing box through which the shaft of a conical rotor projects. The shaft and rotor are driven by a motor shown to the right. The distilland is applied to the center of the rotor whence it is flung outward to the water-cooled collecting gutter and passes away from the still through the residue pipe. A radiant

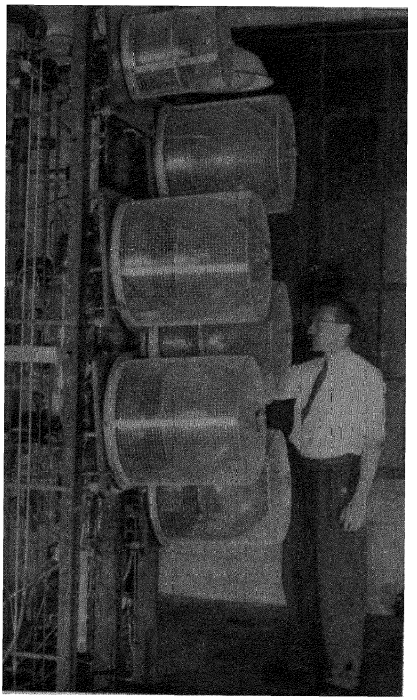


FIGURE 7. A five-plate centrifugal molecular still used in commercial production of vitamin E.

heater and succession of reflectors are situated behind the rotor. The condenser stands in front of the rotor and the entire assembly is housed by a glass bell jar. In apparatus of this type the time of exposure of the distilland is reduced to 0.2-0.1 second and it becomes possible to distill many substances which would literally explode if heated for a longer period. The industrial interpretation of the centrifugal still is shown in Figure 7, and a laboratory cylic variety in Figure 8.

Cyclic Distillation; The Elimination Curve

The chemist trained to rely on the interrelation of temperature and pressure may find himself somewhat at a loss when first performing a molecular distillation. There is no pressure conveniently measurable in the still, and distillation takes place at any difference of temperature between evaporator and condenser, the rate varying with the conditions. The operator thus has little to guide him in the selection of distillates.

Again, the separation, generally less than one theoretical plate, is so poor judged by his previous standards that he is distressed to find a little of every component in each sample of distillate. The distillates differ among themselves merely in relative concentration of the components. It is these difficulties which have caused the valuable properties of the molecular still to be neglected in so many organic laboratories.

A system of molecular distillation is required which will provide a substitute for the ordinary boiling-point measurement. Such a system has been provided by the technique of the "elimination curve" introduced by Embree [17] and Hickman.[18] Cumbersome and laborious, the method nevertheless gives the required information and provides a systematic way of comparing molecular distillations done in different apparatus under varying pressures of residual gas.

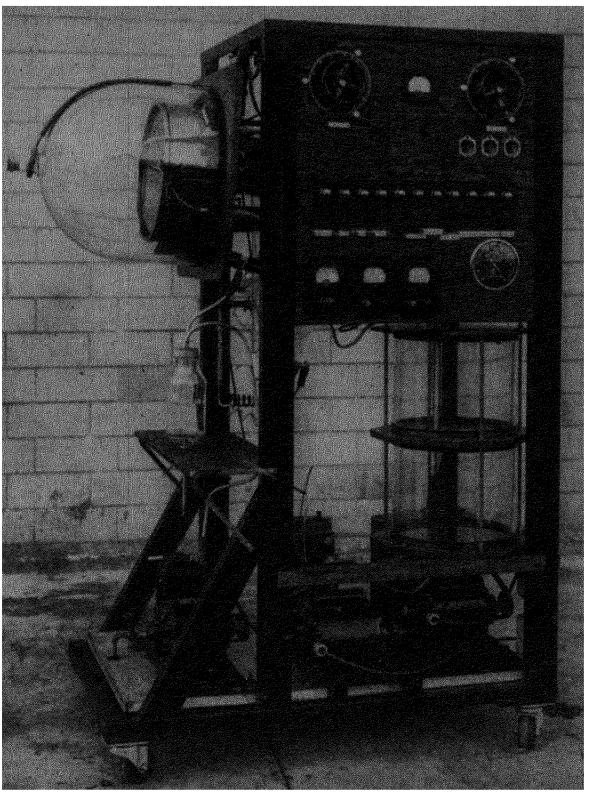


FIGURE 8. A cyclic centrifugal single-plate molecular still of semi-commercial proportions.

The method requires the withdrawal of a large number of distillate fractions over equal periods of time, the temperature of the distilland being raised a constant amount whenever the receiver is changed. The exact technique for this procedure has been detailed elsewhere.[15] In brief, the material to be distilled is dissolved in a specially prepared mixture of "constant yield" and "residue" oils. These can be made from glyceride or petroleum raw materials. Where biological substances are under investigation the glycerides are preferable because of their digestibility and the ease with which they can be removed by saponification. A typical distillation required 200 cc of such a mixture which is cycled for 5- or 10-minute intervals with temperature increments of 5 or 10°. Some 15 to 30 fractions will be withdrawn between the limits of 80 and 250°; these are now examined for all the constituents of interest. It will be found that the concentration of a simple substance will always vary from one fraction to another according to the elimination curve of Figure 9. The reason for

this has been exhaustively presented by Embree.[17] The curve has the fundamental quality of being dependent upon arithmetic and not upon the chemical nature of the substance. However, the curve can be distorted by chemical changes occurring during distillation so that the shape of the curve is a very sensitive indicator of the state of the molecule.

The curve has a maximum which can be used conveniently as a boiling point. In Figure 9 the main curve has a maximum at 125°. Every time the material is distilled under these conditions this substance will collect in maximum yield at 125°.

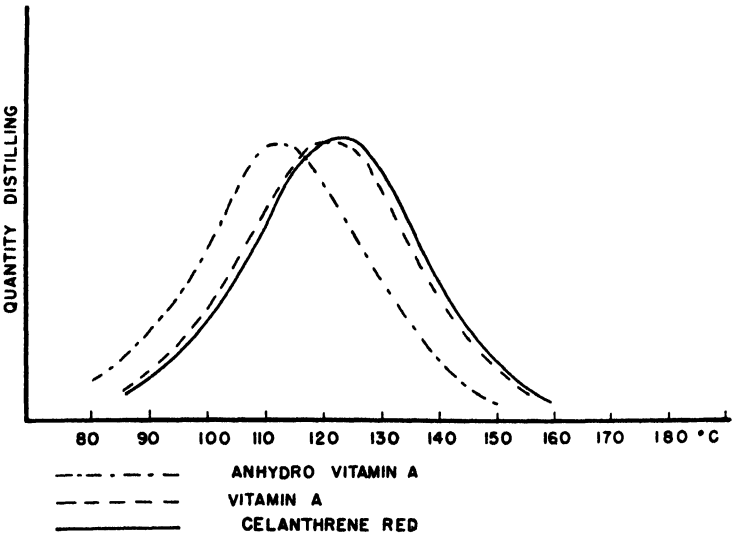


FIGURE 9. Typical elimination curves.

Unfortunately, if the distillation operation is quickened or slowed or performed in apparatus of different design, the curve will shift along the temperature axis. The maximum is, therefore, not an absolute analog of boiling point. It has been found that a series of substances, distilling from the same kind of constant yield and residue oil, generally provide maxima in the same order and it is, therefore, only necessary to include an easily recognizable pilot substance, such as a brightly colored dye, to be able to fix the maximum of the experimental substance. In this way a fairly reliable series of relative evaporation temperatures can be assembled; and it is possible to follow the concentration, changes of properties and general chemistry of many substances by molecular distillation with the same order of exactness that can be derived from other well-known physicochemical measurements.

The technique has been applied as a routine matter in the laboratories of Distillation Products, Inc. during the past five years. The elimination curve is, of course, only a glorified distillation curve and such curves have been known since the beginnings of chemistry. Its importance in the present connection is that it provides a conventionalized method of picturing the behavior of substances in the molecular still. The operator who has been provided with the elimination curves of a mixture of substances can separate them by distillation with the least trouble.

Applications of the method outside of this laboratory have been slow to appear in the literature. A notable exception is the elimination curve of tocopherol presented by Steenbock, Quackenbush, *et al.*[19] Illustrations of the method must thus be taken from our own work dealing with the oil-soluble vitamins.

We shall consider the constituents of a fish-liver oil. As a first step, the oil is saponified and mixed with constant-yield and residue oils, adding a few parts to the million of the pilot dye celanthrene red. The mixture is placed in the cyclic falling-film still and separated into about 20 fractions. The relative concentration of celanthrene red in the fractions is determined by colorimeter and the concentration of vitamin A and cyclized vitamin A by the ultraviolet spectrograph. The extinction coefficients are multiplied by the volume of the fractions, and the total relative yield per fraction is plotted against temperature, giving the three curves of Figure 9 which differ in height and position on the temperature axis but are approximately similar in shape. The curves correspond to vitamin A-alcohol, anhydro vitamin A and celanthrene red. Their relative elimination maxima or boiling points are as follows:

Substance	Relative Elimination Maxima (° C)
Vitamin A-alcohol	126
Anhydro Vitamin A	108
Celanthrene Red	127

If the distillation had been done on a fish-liver oil without the intervening saponification, a different set of curves would have been obtained. Thus if the fractions had been examined for vitamins A and D, the former by ultraviolet or antimony trichloride tests, the latter by animal assay, very small quantities of the vitamins would have been found in the low-boiling fractions, while large quantities would have appeared at higher temperatures. But in no case would a pure elimination curve have been obtained. The reason for this is that the vitamin A occurs naturally esterified with fatty acids selected from the fat of the animal from which the vitamin A is recovered. The same applies to vitamin D, but as there is more than one sub-

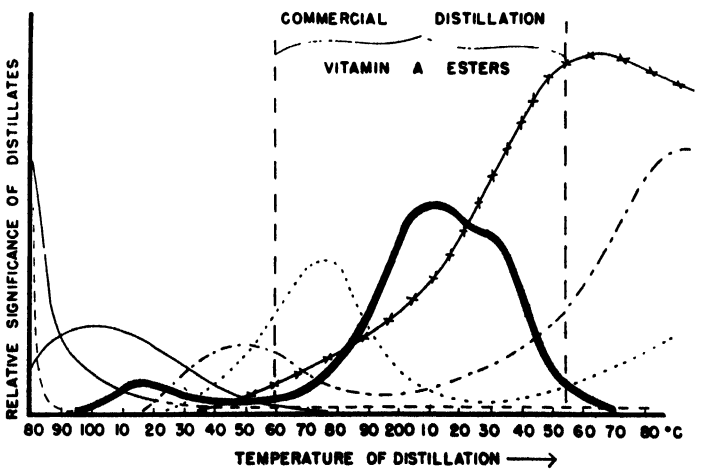


FIGURE 10. Elimination diagram of important constituents of a fish-liver oil.

– – – – – – Residual Protein Odor
———— – – – Rancidity and Reversion Odor
———————— Free Fatty Acids
– – – – – – – – Sterols, Vitamin D, Glyceride Ethers; and Their Esters
· · · · · · · · · · Preservatives, Tocopherols; and Their Esters
———————— Vitamin A and Esters
-x—x—x—x— Glyceride Fat

stance with vitamin D activity, a great many different esters occur. In the common fish-liver oils the C_{16}, C_{18} and C_{22} esters predominate.

The more important constituents of an oil, as revealed by the elimination technique, are shown in Figure 10. The relative heights of the curves are weighted arbitrarily by psychological standards. The vitamin A and glyceride fat curves are accorded first importance, followed by the preservative-tocopherol curve and the odor and fatty acid curves. While the quantity of material contributing to odor is small its importance can not be overestimated in practical nutrition. Figure 10 is more than an illustrative map since it indicates the way in which commercial vitamin A distillate is made from a fish oil and shows why this method of manufacture has come into general use. The oil requires little but a simple passage through the molecular still to strip out the unwanted volatile materials and then to separate the vitamins unharmed and in their natural ester forms. The remaining oil passes unaltered from the stills and becomes available for other edible or industrial purposes. The vitamin A esters together with the tocopherols provide a relatively stable mixture suitable for medicine and food fortification.

It is not the purpose of this paper to lay too much stress on the elimination method. This is a *tour de force* which provides more information than it does practical separation of constituents. The information must always be applied afterward in some other way, either by guiding distillations done without constant yield of residue oils, or indicating some other method of purification.

Applications of the Molecular Still

Soon after Burch's inventions in high-vacuum distillation, it became apparent to chemists that a revolution was impending in fat technology. This revolution, however, did not come about because the method was inexact, and it was extravagant in apparatus, vacuum requirements and the energy consumed in distillation. The hot evaporator is held opposite and near to the cold condenser so that 5 to 10 times as much heat is used in prewarming the oil and keeping it hot as is actually employed in distillation.

Industrial stills at this time of writing are used almost exclusively for the production of oil-soluble vitamins. Vitamin A is concentrated from low-grade, low-potency fish-liver oils which are otherwise unsuitable for human consumption. Vitamin E similarly is distilled from crude or inedible vegetable products.

Some Theoretical Considerations

This review has stressed the practical nature of molecular distillation, it being assumed that the reader who wishes to study the subject more intimately will refer to the original literature. For this reason the technique of high-vacuum pumps and peculiarities of measuring gauges have not been related. It is desirable, however, to include some quantitative figures on the kind [20] and rate [21] of distillation to be expected in the molecular still.

Under molecular conditions and direct transfer to the condenser, molecules do not re-enter the distilland, and there is no equilibrium between liquid and vapor. The quantity distilling is proportional to $P\sqrt{T/M}$. The relative quantities of two or more constituents distilling at a given temperature are:

$$\frac{p_1}{\sqrt{m_1}}, \frac{p_2}{\sqrt{m_2}}, \ldots \frac{p_n}{\sqrt{m_n}}$$

In ordinary distillation where the rate of escape of vapor is less than the rate of generation, molecules re-enter the surface and produce approximate equilibrium between vapor and liquid phases. The quantities of constituents distilling are then proportional to their partial pressures $p_1, p_2 \ldots, p_n$. Molecular distillation will

thus effect separation between substances of equal vapor pressure, providing that molecular weights are different. On the other hand, substances having $p_1/m_1 = p_2/m_2$ which are inseparable molecularly can be separated if the molecular still is abandoned and one of the partially obstructed-path alembics is used instead.

The actual mass, W, in grams of a constituent evaporating per second, per centimeter squared, of substances of molecular weight, M, at absolute temperature, T, at evaporating pressure, P millimeters or p microns of mercury is:

$$W = 5.83 \times 10^{-2} \times P \times \sqrt{M/T}$$

and

$$p = \frac{17{,}200 \times W}{\text{cm}^2} \times \sqrt{T/M}$$

Now it so happens that the factor $\sqrt{T/M}$ approaches unity; it also happens that the molecular weight rises hand in hand with temperature of distillation, so that the substances in the range accommodated by the molecular still have an almost constant rate of distillation—a factor of about 33 per cent, accounting for all likely variations. We may thus assemble the following list of useful distillation rates:

One square meter of evaporating surface will distill per second, at a saturation pressure of 1μ: 0.5 gram of stearic or palmitic acid, or; 0.55 gram of sterols, or; 0.75 gram of a typical glyceride fat.

Bearing this simple relation in mind, the rate of distillation can be adjusted in a small still so that it will occur under approximately molecular conditions.

References and Acknowledgements

The author wishes to acknowledge a debt of gratitude to his colleagues at Distillation Products who have placed their work so fully at his disposal. Much of the material quoted from the general literature has been traced through "review articles," important among which are the two bibliographies of Detwiler,[1, 22] the symposium articles by Burch,[20] Fawcett,[21] Burrows,[23] Phipps and Meade [24] and Waterman.[4] A recent compilation from the pen of D. D. Howat has appeared in the Chemical Age.[25] The reader is referred to these and original papers for further details. Not mentioned in this short review are the important researches of Knudsen on gaseous flow through tubes and orifices. The reader is referred to a small monograph by Knudsen [26] and to the standard works on high-vacuum technique such as those by Dushman,[27] Dunnoyer [28] and Strong.[29]

References

1. Detwiler, S. B., Jr., *Oil and Soap*, **16**, 2 (1939); **17**, 241 (1940).
2. Krafft, F., *et al.*, *Ber.*, **29**, 1316, 2240 (1896); **42**, 202, 210 (1909). Caldwell, K. S. and Hurtley, W. H., *J. Chem. Soc.*, **95**, 853 (1909).
3. Brönsted, J. N. and von Hevesy, G., *Phil. Mag.*, **43**, 31 (1922).
4. Waterman, H. I. and van Vlodrop, C., *Rev. Chim. Ind.*, **48**, 314 (1939).
5. Burch, C. R., *Proc. Roy. Soc. London*, **A123**, 271 (1929).
6. Langmuir, I., *J. Franklin Inst.*, **182**, 719 (1916).
7. Burch, C. R., *et al.*, British Patent 303,078 (December 21, 1928). Reissued as U. S. Patent 1,955,321 (April 17, 1934).
8. Burch, C. R., *et al.*, U. S. Patent 1,955,321 (April 17, 1934). Reissued from British Patent 303,078 (December 21, 1928).
9. Washburn, E. W., *et al.*, *J. Research Nat. Bur. Stand.*, **2**, 467 (1929).
10. Hickman, K. and Sanford, C. R., *J. Phys. Chem.*, **34**, 637 (1930). Hickman, K., *J. Franklin Inst.*, **213**, 119 (1932).
11. Hickman, K., U. S. Patent Re20,705 (April 26, 1938). Reissued from U. S. Patent 1,925,559 (September 5, 1933).
12. Waterman, H. I., and Oosterhof, D., *Rec. trav. chim.*, **52**, 895 (1933).
13. Detwiler, S. B., Jr. and Markley, K. S., *Ind. Eng. Chem., Anal. Ed.*, **12**, 348 (1940).
15. Hickman, K., *Ind. Eng. Chem.*, **29**, 968 (1937).
16. Kodak Ltd., British Patent 482,883 (April 6, 1938). Hickman, K., U. S. Patent 2,180,050 (November 14, 1939).
17. Embree, N. D., *Ind. Eng. Chem.*, **29**, 975 (1937).
18. Hickman, K., *Nature*, **138**, 881 (1936).

19. Quackenbush, F. W., Gotlieb, H. L. and Steenbock, H., *Ind. Eng. Chem.*, **33**, 1276 (1941).
20. Burch, C. R. and van Dijck, W. J. D., *J. Soc. Chem. Ind.*, **58**, 39 (1939).
21. Fawcett, E. W., *J. Soc. Chem. Ind.*, **58**, 43 (1939).
22. Detwiler, S. B., Jr., Abstracts of Articles and Patents on Molecular or Short-Path Distillation. Urbana, Ill., U. S. Regional Soybean Industrial Products Laboratory, 1941.
23. Burrows, G., *J. Soc. Chem. Ind.*, **58**, 50 (1939).
24. Jewell, W., Mead, T. H. and Phipps, J. W., *J. Soc. Chem. Ind.*, **58**, 56 (1939).
25. Howat, D. D., *Chem. Age.*, **45**, 309, 323 (1941); **46**, 3, 41, 53 (1942).
26. Knudsen, M., (See) "The Kinetic Theory of Gases," pp. 21-26, "Methuen's Monographs on Physical Subjects," London, 1934.
27. Dushman, S., "High Vacuum," Schenectady, General Electric Review Press, 1922.
28. Dunnoyer, L., "Vacuum Practice," New York, D. Van Nostrand Company, 1926.
29. Strong, J., "Procedures in Experimental Physics," New York, Prentice-Hall, Inc., 1939.

States of Aggregation of Some Paraffin Chain Compounds

ROBERT D. VOLD
Assistant Professor of Chemistry
and
MARJORIE J. VOLD

Research Associate in Chemistry, University of Southern California

Paraffin chain compounds are remarkable for the number and variety of crystalline and mesomorphic forms they exhibit. Thus the *n*-paraffin hydrocarbons (I) furnish examples of three different crystal forms and one probably mesomorphic modification. For the *n*-fatty acids (II) there are three and perhaps four crystal forms; for the alcohols (IV), two; for the esters (III) one or more crystal forms and also a mesomorphic form; and for the simple triglycerides (III) one crystal form, and two probably mesomorphic forms. Sodium salts of the *n*-fatty acids (VI) display probably three alternative crystal forms, as well as a sequence of at least as many as five stable mesomorphic forms.* The purpose of this paper is to correlate the known experimental facts concerning these states with various hypotheses that have been advanced concerning their structure and stability.

It is desirable at the outset to distinguish between real differences in polymorphic form (*i.e.*, discrete phases) and other variations in the condition of matter in which degree of dispersion, protective colloids or other special circumstances have profoundly modified the physical properties of a substance without any fundamental alteration in the packing together or manner of arrangement of its molecules.

It is not at present possible to specify how extensive an alteration in molecular arrangement is required to produce a phase change. For example, a liquid is thought of as containing a high degree of "local order," the distribution over a few molecular diameters about many single molecules being, at any instant, similar to that found in the crystal. Although at high temperatures the number and persistence of such groups are much less than at temperatures just above the freezing point of the substance, the corresponding variation in measurable properties is continuous. However, a definite discontinuity in primary physical properties, such as volume or heat content, accompanies the transformation: nematic liquid crystal—isotropic liquid; this transformation is still regarded as consisting only in a variation in the number and persistence of groups of oriented molecules, although here a genuine phase change has occurred.

Similarly, glasses are sometimes thought of as a separate state, and sometimes as

* The total number of forms that can be realized for sodium stearate now totals probably ten. Of these, two are certainly liquid crystalline and three crystalline. However, positive classification of the remaining forms into one category or the other requires further experimental work or may even remain indefinitely a matter of definition.

undercooled liquids, there being no satisfactory theoretical criterion for their classification. The unstable γ form of simple triglycerides, obtainable by rapidly chilling the liquid, is possibly an example of an anisotropic glass.[25] Parallelism of the paraffin chains seems to be achieved, but not regular arrangement of their ends in planes by the time the thermal energy becomes too low to allow the molecules sufficient mobility to complete the process of crystallization.

The best criterion for characterizing a given variant on the spatial distribution of molecules as a separate form or phase is the experimental one of the detection of transitions. Most of the discussion in this paper is concerned with modifications of paraffin chain compounds which are separate phases in this sense. The problem is to try to account for the very large number of forms, and to relate changes in behavior to the internal structure of the phase and to the physical constitution (*i.e.*, anisometry, heteropolar character, etc.) of the molecules.

Crystal Forms of Paraffin Chain Compounds

Crystal forms of paraffin chain compounds fall into two main categories, rhombic and monoclinic. In the rhombic form the axis of the zig-zag chains lies perpendicular to the planes containing the terminal groups. As many as four monoclinic forms of one compound have been reported. These differ from one another principally in the angle of inclination of the chain axis to the terminal planes.

Not all paraffin chain compounds have been obtained in the rhombic form. Examples are the fatty acids and the methyl esters of an odd number of carbon atoms. The α form of the ethyl esters, although the chains are perpendicular to the terminal planes, is possibly not fully crystalline. Potassium acid soaps [18] and lead soaps [35] ordinarily crystallize in the rhombic form. The non-appearance of a rhombic form is associated with strong attractive forces between the terminal groups, leading to their closer packing. Thus for the hydrocarbons,[6] when the monoclinic lattice is strained by addition of homologous impurities, whose different chain length prohibits efficient packing of the heads, transformation occurs to the rhombic form.

Figure 1 shows the c spacing (separation of the planes containing the terminal groups) for the rhombic form of several series of paraffin chain molecules. It is at once apparent that there are two series of such compounds. In one series successive molecules are oppositely oriented, so that the lattice unit is a linear double molecule of a total length similar to that of the true compounds, such as the lead soaps or potassium acid soaps. This condition does not occur in the hydrocarbons (since the molecules are symmetrical), nor in ketones other than methyl ketones, nor in ethyl esters. Methyl esters of an even number of carbon atoms behave like ethyl esters. The odd ones crystallize as double molecules, but not in the rhombic system.

For both series the increment in c spacing is very nearly 1.27 Å per CH_2 group, as is required for a plane zig-zag carbon chain with tetrahedral bond angles and the normal C-C bond distance.

Schoon [18] has given an excellent discussion of the concept that the various monoclinic forms may be considered as derivable from the rhombic lattice by means of a sort of shear along some definite plane. The amount of the displacement is "quantized," being some multiple of 1.27 Å in order to maintain an efficient packing of the chains. Moreover, in all forms the short spacings must be such that each molecule has about the same apparent cross-section perpendicular to its axis, since the van der Waal's forces between the chains have not been altered in the shift. Thus, for a shear parallel to the b-c plane ([100]), one has the two relations,

$$\tan\ (90 - \beta) = 1.27\ n/a_0$$

$$a_n = a_0/\sin\ \beta$$

where n is an integer giving the relative displacement of the chains in number of methylene groups, β is the monoclinic angle, and a_n one of the principal axes of the new crystal. These relations can be seen from Figure 2, taken from Schoon's paper.

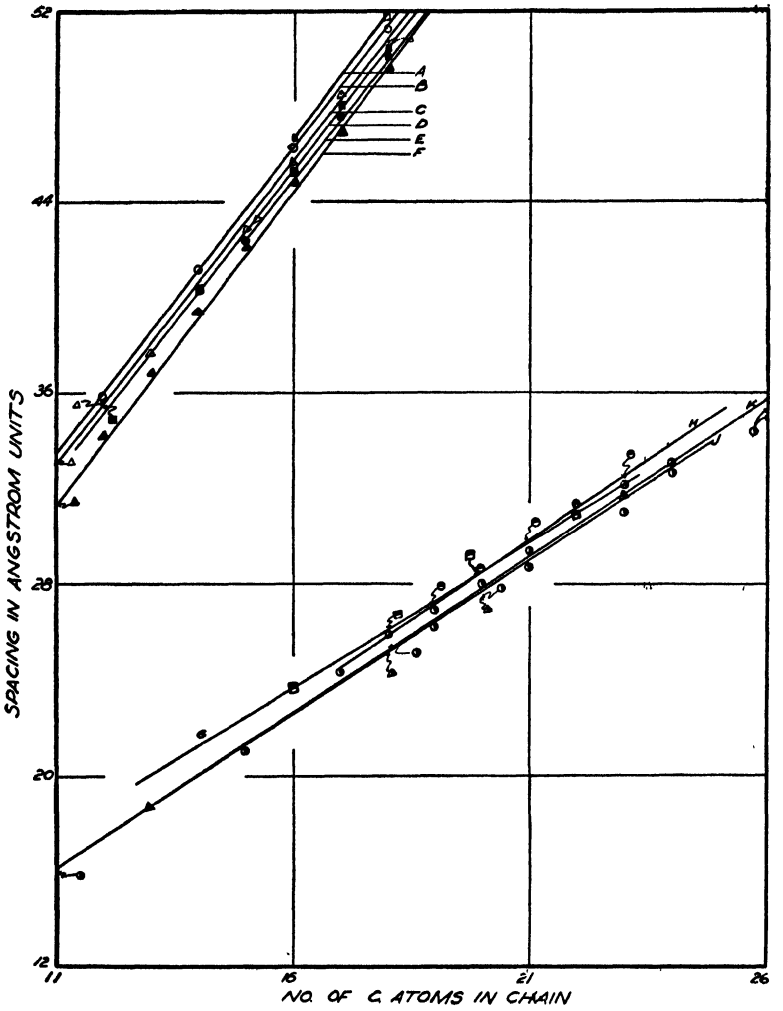


Fig. 1 C Spacings of the Rhombic Form of Several Series of Paraffin Chain Compounds

FIGURE 1. c Spacings of the Rhombic Form of Several Series of Paraffin Chain Compounds. A, sodium soaps; B, lead soaps; C, simple triglycerides; D, potassium acid soaps; E, methyl ketones; F, aliphatic alcohols; G, methyl esters (Even no. of C atoms); H, ethyl esters; J, ketones other than methyl ketones; K, hydrocarbons.

The b spacing remains constant. The c spacing is $c_0 \sin \beta$, owing to the tilt of the chains, and the a spacing varies to keep $ab \sin \beta$ constant. A second possible series of monoclinic forms results from a shear along the a-c planes ([010]), a and b spacings and the values of β for these hypothetical structures are shown in Table 1.

The monoclinic forms derivable by this means differ from one another in the magnitude of the monoclinic angle β between the chain axis and the terminal planes, and also in the values of the a spacing. One series has $b = 7.47$ Å and the other has $b = 4.97$ Å. When the c spacings have been determined for a homologous series, β can be determined from the increment per CH_2 group, which is 1.27 sin β. This

fact assists in a tentative classification of the observed results according to Schoon's scheme even where the a and b spacings have not been determined.

Where there are several polymorphic crystal forms for a series of paraffin chain compounds it has become conventional to designate the rhombic form by the symbol A or α, and the monoclinic forms by the symbols B (β), C, D, etc., in the order of decreasing c spacing. Occasionally, as for the fatty acids,[13] the symbol A has been

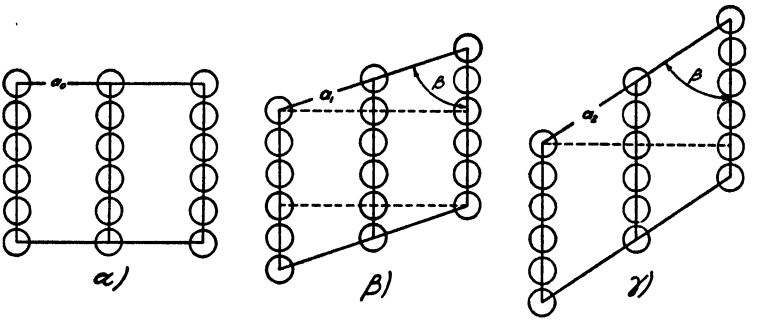


FIGURE 2. Alternative Methods of Packing of Hydrocarbon Chains: $a = a_0/\sin\beta$; $b = b_0$; $c = c_0 \sin\beta$. [After Schoon, T., *Z. physik. Chem.*, **B39**, 385 (1938).]

used for the monoclinic form of longest c spacing when a rhombic form does not exist. The situation is confused by the use of the same symbols to designate forms with similar temperature regions of stability without regard to structure, as in the case of the α form of the ethyl and methyl esters of n-fatty acids. Schoon's system, given in Table 1, seems more rational, though it is admittedly cumbersome and, in addition, presupposes that all crystal forms of these compounds can be so classified.

Table 1. Characteristics of Monoclinic forms of Paraffin Chain Compounds derived by Distortion of the Rhombic Lattice.

symbol	a	b	β
	1. Displacement parallel to [100]		
β_1	5.135	7.47	75° 15′
γ_1	5.575	7.47	62° 58′
δ_1	6.26	7.47	52° 30′
ϵ_1	7.105	7.47	44° 20′
	2. Displacement parallel to [010]		
β_2	7.58	4.97	80° 19′
γ_2	7.88	4.97	71° 13′
δ_2	8.38	4.97	62° 57′
ϵ_2	9.04	4.97	55° 50′

Although only one of several crystal forms of a given compound can be stable at a given temperature, different ones can be obtained in a metastable condition by suitable experimental procedures, and persist indefinitely. For example, sodium stearate at room temperature can be either rhombic or monoclinic.[40] Piper [8] finds that films of palmitic acid melted onto glass show the C form, while pressed films give a mixture of A and C forms. Similar results obtain for other fatty acids.

Differences in c spacing of up to 0.7 Å may occur within the same crystal form prepared in different ways, because of distortion of the lattice.

Progress in the elucidation of the structure of the crystal forms of paraffin compounds was given a great impetus by the painstaking preparation of highly purified materials. This is particularly significant for x-ray work, where conclusions

are based on the quantitative as well as the qualitative features of the diffraction pattern. Sometimes [15, 16] homologous impurities form solid solutions with the material under investigation and give rise to changes in the lattice constants roughly proportional to the amount present. Sometimes, as for solutions of hydrocarbons C_{29} or C_{26} in C_{28}, the qualitative behavior is altered. In this instance [6] C_{28} normally occurs in the C monoclinic form, but with additions of 10 per cent of C_{26} or only 2.5 per cent of C_{30}, transformation to the rhombic (A) form occurs.

In the following summary of the reported crystal forms of various paraffin chain compounds, an effort has been made at classification according to the Schoon system. When a and b spacings and the value of β are available, they show immediately to which of Schoon's classes, if any, the substance belongs. When only a value of β calculated from the increment in c spacing is available, the assignment is not so definite, γ_1, and δ_2, for example, being indistinguishable.

Hydrocarbons of less than 26 carbon atoms apparently crystallize in form γ_1, (B) while those of more than 26 carbon atoms give form δ_1, (C) judging from the increments per methylene group in the c spacing. The alcohols commonly crystallize in the rhombic form (α) but have also a monoclinic type (β) which appears to correspond to Schoon's category ε_2. The methyl esters (β) belong to class γ_1. Here there are two series with slightly different c spacing, one for the odd and one for the even members. Alternation of this kind does not occur for the rhombic forms, and not universally for monoclinic forms. The β form of the ethyl esters is also γ_1, but the lattice unit is a single molecule rather than an oppositely oriented pair, as is the case for the methyl series. The β form of the simple triglycerides has the same monoclinic angle as the ethyl esters, but its crystal structure is probably more complex than that of the esters of monohydric alcohols. The behavior of the fatty acids is very complex. Plots of c spacing against the number of carbon atoms in the acid [18] show seven straight lines, 3 for the even series and four for the odd series. The slopes for the even series correspond closely to that expected for forms γ_2(A), γ_1(B) and ε_2(C) respectively. The B and C forms for the odd series have nearly the same increment in c spacing per methylene group, and therefore do not fit into the scheme of Table 2. The a and b spacings for β-margaric acid have been determined.[18] Apparently its structure differs from that being here assumed, since the cross-section, $ab \sin \beta$, is larger (39.6Å^2) than the fairly constant value of 37.0 for other paraffin chain compounds; also the b spacing is very small (4.63Å, compared to 4.97). Potassium soaps [34] have a quite small increment in c per CH_2 group, putting them possibly into class ε_2 or δ_1. The monoclinic form of sodium soaps studied by Thiessen and Stauff [40] does not seem to fit the Schoon system.*

Although the apparent cross-section area of the paraffin chain remains nearly constant from crystal to crystal ($ab \sin \beta = 37$Å), the field of force about the axis of the chain cannot be considered to have cylindrical symmetry. Thus Muller [5] finds that the thermal expansion along the a axis for the rhombic form of the hydrocarbons is about 7 per cent between liquid-air temperatures and the melting point, while that along the b axis is only about 2 per cent. (Expansion along the c axis is negligible). As the temperature is raised, oscillations about the chain axis increase in amplitude despite the lack of symmetry of the field of force, until at a sufficiently high temperature free rotation occurs, as in the case of ethyl stearate.[30] Ethyl undecylate, on the other hand, melts at a temperature too low for the molecules to have sufficient rotational energy to overcome the crystal forces opposing rotation.

* According to J. W. McBain and A. de Bretteville, Thiessen and Stauff's paper contains many numerical errors which probably vitiate any conclusions drawn from it concerning the structure of the monoclinic form.

Liquid Structure of Paraffin Chain Compounds

The transformation of a paraffin chain compound from the crystalline to the liquid state consists, as for other crystals, in the disappearance of long-range order. The overall distribution of molecules relative to one another is random, but there still exist transient groupings of small numbers of molecules with the same arrangement as in the crystal.

To this concept must be added, for paraffin chain compounds, the consideration that the chains can become coiled in the liquid state. Thus, Kauzmann and Eyring [48] find from viscosity measurements that not more than an average of 14 methylene groups can maintain themselves in an extended plane zig-zag arrangement for any appreciable time. Meyer and Lühdemann,[49] on the basis of osmotic pressure measurements in fairly concentrated solutions, suggest a kinetically active unit of 17 carbon atoms. Similarly, Huggins,[50] in calculating the entropy of liquid paraffin hydrocarbons (which he finds to be a linear function of the number of carbon atoms at any one temperature), considers that the chains are kinked in a completely random manner, or perhaps even sufficiently coiled up to give essentially spherical molecules. Hildebrand [51] has ascribed the nearly ideal behavior of solutions of paraffins, even of quite different chain length, to the fact that the molecules lie parallel, two short ones end to end not acting very differently from one longer one. The occurrence of nearly ideal solutions, however, does not necessitate a parallel arrangement of extended chains if the coiled structure is regarded as changing continuously with thermal agitation, rather than being rigid. X-ray evidence [47, 52, 58, 59, 60] shows, on the other hand, that adjacent molecules are on the average roughly parallel and, further, that for molecules of up to ten carbon atoms, the chains are essentially straight.

The Melting Process and Mesomorphic Forms

The transformation of a substance from crystal to liquid is often thought of as occurring sharply at a single temperature which has come to be known as the melting "point." Though justified in many cases by laboratory experience, the concept of a melting point has proved difficult to incorporate into statistical theories of the solid and liquid states. Molecules are held in position on a lattice by mutual attractions and repulsions. As the temperature is raised they oscillate about these positions with ever-increasing amplitude and anharmonicity; finally the "vibration" is replaced by translatory movement, all semblance of gross crystallinity (long-distance order) disappears, and the substance is said to have melted. A definite amount of energy is required for a given molecule to break loose from the forces binding it to its equilibrium position, and the proportion of molecules having sufficient energy for this scission increases with increasing temperature. What is required of theory is an explanation of why the proportion of free molecules increases so tremendously over a temperature interval sufficiently narrow to give rise to melting phenomena at all (cf. ref. 53, p. 103).

In presenting their ingenious work on the melting behavior of paraffin chain ketones and their mixtures, Oldham and Ubbelohde [33] discuss the hypothesis of "coöperative lattice flaws." According to this theory, when one imperfection develops in a lattice, the energy required to loosen adjacent molecules diminishes; when two imperfections develop close together they exert much more than a double effect in diminishing the energy required to loosen intervening molecules; finally a network of such coupled lattice defects reduces the remaining energy requirement for complete catastrophic breakdown of the crystal to a negligible value, so that melting occurs.

In a heteropolar molecule, such as, for example, palmitic acid, the constraints operating to maintain the terminal—COOH groups in place must be quite different, in

both degree and kind, from those operating on centrally located methylene groups. The existence of a sharp melting point must here be due to a further coupling, whereby a network of defects or flaws in the hydrocarbon part of the lattice reduces the energy required to destroy the more rigid framework of the carboxyl groups so that complete disruption can occur at one temperature.

However, there are cases where interaction between groups of different polar character is not very strong, so that breakdown of the crystal lattice does occur in steps. Bernal and Crowfoot [54] have summarized the structural requirements for the occurrence of liquid crystalline or mesomorphic phases: namely, (1) rod-like or plate-like geometric form, (2) heteropolar character, but with the polar group not too strongly polar. Classic examples of mesomorphism are the nematic liquid crystalline phase of *p*-azoxyanisole and the smectic liquid crystalline form of ethyl *p*-azoxybenzoate. These, and others of the same type, have been described by many authors.[53] Sodium salts of the *n*-fatty acids constitute the most outstanding example of this kind of behavior yet reported, there being at least five, and perhaps more, different intermediate forms of the same substance, each with its own definite temperature range of stability.

The step-wise breakdown of some paraffin chain lattices offers the opportunity to achieve a detailed understanding of the mechanism of the melting process, when the structure of each of the intermediate forms has been determined. At present, however, only scattered fragments of information are available.

The development of free rotation about the chain axis could be considered as a step in the melting process. Baker and Smyth [55] advance the idea that many compounds in which the entropy of fusion is abnormally small can be considered to have developed degrees of freedom, over a range of temperature preceding the melting point, which are usually acquired only during melting. The idea is that when the substance can increase its entropy through such means with only a very small change in free energy, actual melting can be economically postponed to a higher temperature. Such compounds have therefore abnormally high melting points, or alternatively, abnormally short temperature ranges of stability of the liquid phase. This phenomenon is found most strikingly for molecules of nearly spherical shape where three-dimensional rotation is possible; thus it is pronounced for tertiary butyl bromide, less so for isopropyl bromide, and absent for *n*-amyl bromide.

The normal paraffin chain hydrocarbons [5] undergo a transformation near their melting points which can be regarded as a step toward melting involving even less rearrangement than the development of rotational freedom. It follows from the plane zig-zag structure of the hydrocarbon molecule that the plane through the methylene groups of a single molecule can assume various orientations, *i.e.*, the molecules are lath-shaped rather than rod-shaped. On the rhombic crystal form the chains are oriented so that these planes are perpendicular to each other along the diagonal of the *ab* crystal face, as shown in Figure 3a. Just below the melting point transformation occurs to the layer structure shown in Figure 3b. This is revealed experimentally by the disappearance of lines in the x-ray diffraction pattern coming from the [200] planes.

Methyl and ethyl esters of fatty acids separate a transparent form (α) on cooling the liquid. For esters of odd acids this form undergoes a reversible transition to a monoclinic structure (β); for the even acids this transition is monotropic. Except for methyl esters of odd acids, the α form has the chain axis perpendicular to the planes of the terminal groups. There is only a single lateral spacing indicating a mean distance of closest approach of the parallel paraffin chains of about 4.2 Å. Perhaps the planes of the individual molecules are here randomly oriented. This structure can be considered to belong to Bernal and Crowfoot's [54] ordered smectic category of liquid crystals. If the randomness is due only to the circumstance that thermal motion is insufficient to permit re-orientation of these planes into a regular

crystal pattern rather than being the thermodynamically stable condition, this structure should be regarded as an anisotropic glass.

The α form of the simple triglycerides [25] is probably also an example of the ordered smectic state, since the terminal groups lie in planes, and there is a fairly constant separation of the paraffin chains, which, however, are said to be frozen in

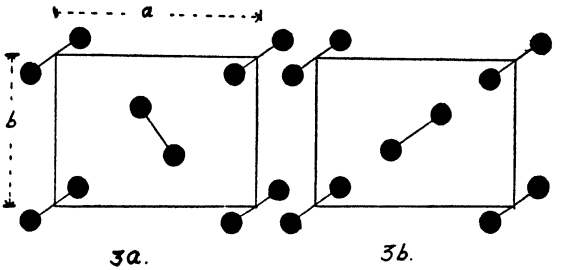


FIGURE 3. Relative Orientation of the Molecules in (a) the Rhombic Crystal, (b) Layer Structure of Paraffin Hydrocarbons. •—• represents the projection of a hydrocarbon molecule on the $a-b$ crystal face. [After A. Muller, *Proc. Roy. Soc. London*, **A127**, 417 (1930).]

random orientation with respect to each other. The γ form of the triglycerides, obtained by quenching the liquid form, is, on the other hand, apparently more closely related to a nematic liquid crystal, since the terminal groups do not lie in planes (no long spacings appear), and the only regularity seems to be a parallel alignment of the chains. Because of its rigidity and also its metastability, this form has also been referred to as an anisotropic glass.

Mesomorphic Forms of the Sodium Soaps

As has been already pointed out, the sodium soaps * afford the most striking illustration of a stepwise melting process yet discovered. The x-ray studies necessary to complete understanding of the structure of each phase have not been carried out, although they have been begun in several laboratories.† The principal effort has thus far been directed toward finding the number and temperature ranges of existence of the phases for different soaps, and the relations between the phases of one soap and those of another. Recent calorimetric studies provide the experimental basis for initial speculations concerning the nature of some of the transitions.[44]

Sodium palmitate has six definitely established transitions [41, 42] in addition to the monotropic transition from the rhombic crystal to a monoclinic form [40] and the genotypic point.[38] At the high temperature extreme it is an ordinary isotropic liquid. At room temperature the substance normally exists as a monoclinic crystal. When sodium palmitate is prepared by evaporation of the solvent from an aqueous solution, its physical form is fibrous, individual fibers being exceedingly thin, sometimes as thin as two molecular lengths (electron microscope [45]). This circumstance has resulted in frequent reference to the crystalline form of the soaps as "curd fibers" or "curd fiber phase." A rhombic crystal form can be prepared at room temperature by crystallization from ethyl alcohol.[40] It is possible [44, p. 2020] that more than one monoclinic form can exist.

The naming of the known phases intermediate between crystalline sodium palmitate and liquid sodium palmitate reflects the history of their discovery and investigation. When one mesomorphic form was first recognized it was given the

* Recent unpublished experiments of the authors show that other alkali metal soaps also exhibit sequences of mesomorphic forms.

† F. G. Chesley, *J. Chem. Phys.*, 8, 643 (1940). See also ref. 38, 39, 40. J. W. McBain, personal communication.

name "neat," because it was thought to be continuous in a phase rule sense with the anisotropic liquid of that name present in the commercial soap kettle.[46] Then a second intermediate form was discovered and called "waxy" because of its resemblance (in appearance) to paraffin. The names subwaxy, superwaxy, and subneat were assigned to subsequently discovered phases to indicate their position in the sequence. (subwaxy—waxy—superwaxy—subneat—neat.)

The neat soap phase of sodium palmitate is a translucent, viscous, doubly refracting fluid. Trapped bubbles are elongated because of the anisotropic character of the resistance offered by the liquid to expansion of the air. The phase separates from isotropic liquid either as globules (on which a uniaxial cross is visible through crossed Nicols with a polarizing microscope) or as batonnets. In very thin layers on glass neat soap becomes rapidly oriented, with the optic axis perpendicular to the glass so as to simulate isotropic behavior. Curiously, the orientation is more rapid near the center of the temperature range of stability of the phase than at higher temperatures.

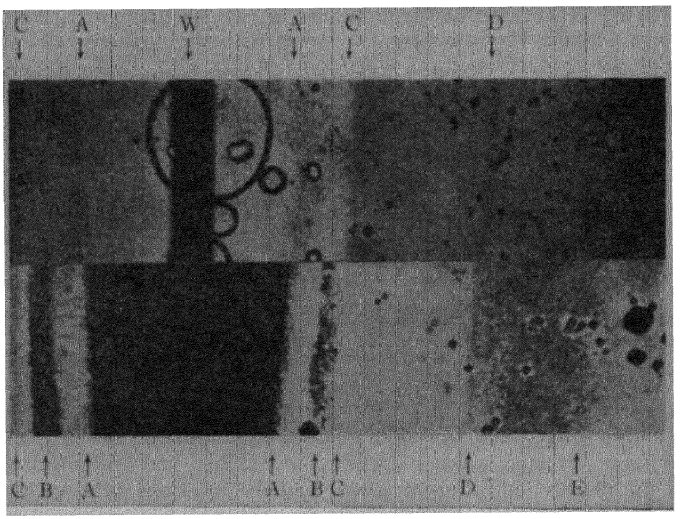


FIGURE 4. Photomicrographs of sodium stearate by the hot-wire technique. Five of the six phases of sodium stearate appear successively as the temperature falls from about 350° near the wire (W) to about 140° at the right-hand edge of the photograph. A, boundary between isotropic liquid and neat soap; C, boundary between neat soap and subneat soap; D, boundary between subneat soap and superwaxy soap; E, boundary between superwaxy soap and waxy soap. The upper photograph was taken with the Micropolychromar, the lower, showing an adjacent portion of the same field of view, with crossed polaroids. Magnification is about 20 fold. At B the neat soap has begun to become oriented, simulating the appearance of isotropic liquid. [From Vold, M. J., Macomber, M., and Vold, R. D., *J. Am. Chem. Soc.*, 63, 168 (1941).]

The subneat phase of sodium palmitate is also a translucent, plastic, optically anisotropic material. It is more rigid than neat soap. The size of uniformly oriented aggregates is smaller, and extensive orientation, as for neat soap on a glass surface, has not been observed.

Viewed in the microscope with white light through crossed polaroids, both neat soap and subneat soap appear as a mosaic of different colors which arise from the different thicknesses of uniformly oriented patches of liquid crystal. For neat soap

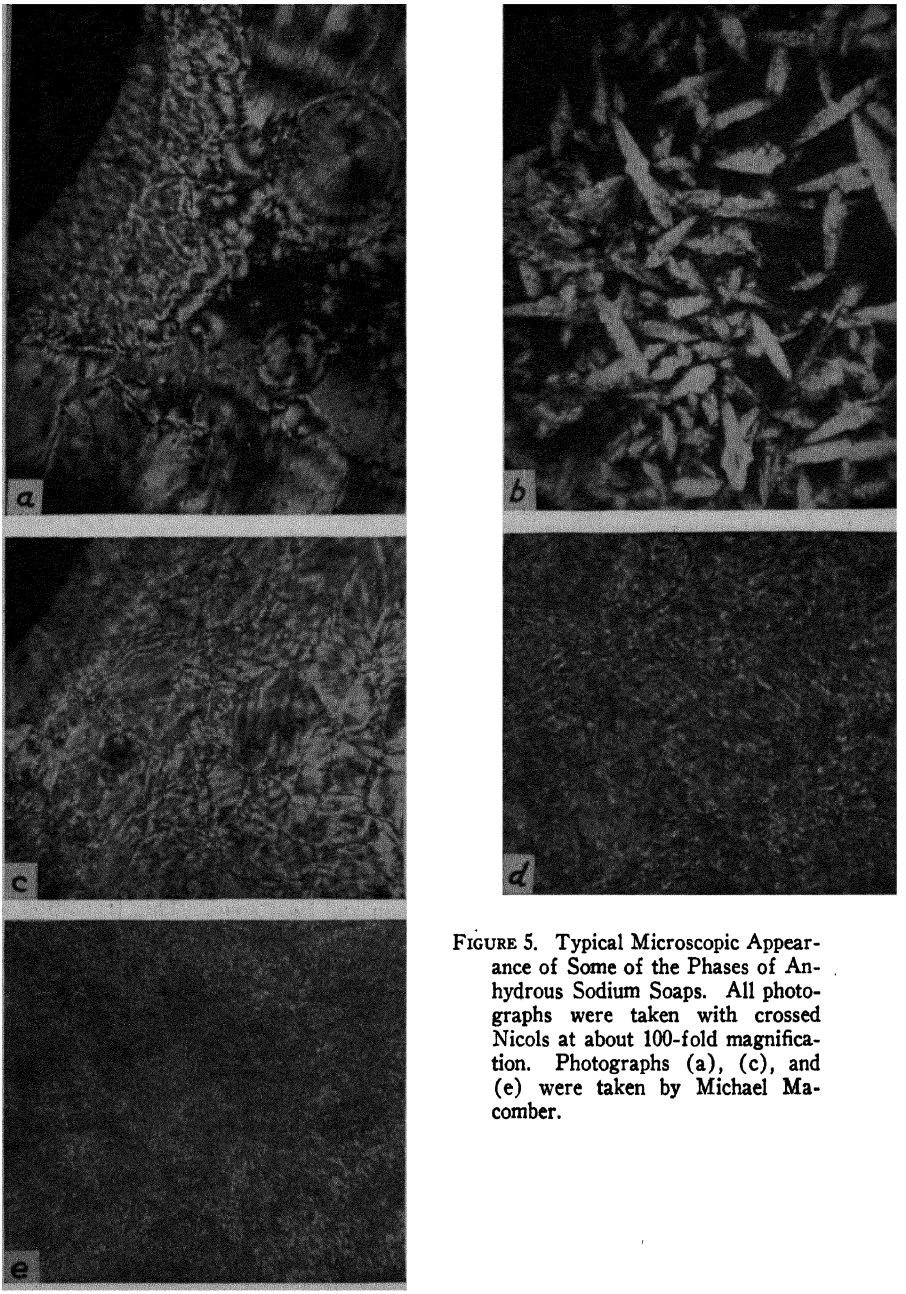


FIGURE 5. Typical Microscopic Appearance of Some of the Phases of Anhydrous Sodium Soaps. All photographs were taken with crossed Nicols at about 100-fold magnification. Photographs (a), (c), and (e) were taken by Michael Macomber.

(a) Sodium Laurate at 336° C. Neat Soap in Equilibrium with Isotropic Liquid.
(b) Sodium Palmitate at 295° C. Neat soap in Equilibrium with Isotropic Liquid. (Batonnet Structures).
(c) Sodium Laurate Neat Soap below 336° C.
(d) Sodium Palmitate at 220° C. Subneat Soap.
(e) Sodium Stearate at 140° C. Waxy Soap.

uniformly colored areas are larger, and conic structures of the type described by Friedel,[58] Lawrence,[56] and others are frequently visible.

The complete series of phases can be viewed together when a thin film of material is subjected to a temperature gradient produced by a hot wire placed at the edge of the film.[42] The films are too thin to exhibit polarization colors, but characteristic differences nevertheless appear, leading to sharply defined boundaries between the different phases. Figure 4 shows a part of the sequence for sodium stearate. Different methods of illuminating the sample were employed in order to render one or another boundary more conspicuous.

The "waxy" phases (subwaxy, waxy, and superwaxy) are very similar in appearance, at least in the case of sodium palmitate and sodium stearate. When prepared by heating or cooling a film of anhydrous material, they all display a granular stippled appearance without any obvious structural features or optical patterns. They differ most in translucency, superwaxy being the most nearly transparent, and subwaxy the least. This difference is largely responsible for the conspicuous boundaries seen in Figure 4. Figure 5 shows the characteristic microscopic appearance of several of these phases.

Even for the most thoroughly studied soaps, sodium palmitate and sodium stearate, the entire gamut of experimental behavior is not fully understood. Among the unsolved problems are the following.

First is the interrelation between the various crystal forms. Sodium stearate usually crystallizes from solvents in a rhombic form if the crystals form first below 50° C. This form undergoes an irreversible transition at 52° to a monoclinic form. However, the preparations always contained about 1% moisture, removal of which by room temperature dehydration destroys the capacity of the soap to exhibit any transition at 52°. Both in this dehydrated material, and in the slightly moist monoclinic form produced by the transition at 52°, the genotypic transition discovered by Thiessen and co-workers[37, 38] is present, and appears to be reversible. However, this genotypic transition is absent from all preparations that have been heated hot enough to form subwaxy soap, and from some heated only hot enough to form "supercurd," even though both these transitions themselves seem reversible. Thus the number of crystal forms possible, and their relative stability are still undetermined.

Second is the relation between subneat soap and neat soap. This transition appears to occur (microscopic observations) at 253° C for sodium palmitate; a small but definite and reproducible reversible change in volume also occurs at this temperature. However, there is no observed change in heat content at 253°, while there is a *reversible* heat effect of 1540 cal/mole at 237° unaccompanied by visual or dilatometric evidence of a transition. Further, M. Heldman has found that an enormous increase in electrical conductivity occurs continuously in the entire range of temperature from 237° to 253°. Is this behavior subject to interpretation in terms of the structure of subneat soap, or must we admit the existence of yet another intermediate form?

Also there is the problem as to why orientation of neat soap on glass should proceed more rapidly at 270°, near the center of its temperature range of stability, than at the melting point (295° C), where it is actually in contact with mobile isotropic liquid.* (See Figure 4 for an illustration of this same effect for sodium stearate.)

* Since the above summary was written (December, 1941) further research, particularly on sodium stearate, has partially cleared some of the reported anomalies. Thus M. J. Buerger, L. B. Smith, A. De Bretteville and F. V. Ryer [*Proc. Nat. Acad. Sci.*, **28**, 526 (1942)] find two of the reported crystal modifications of sodium stearate to be hydrates of very low water content. The transition of "sodium stearate" occurring at 53° appears to be the decomposition temperature of the hemi-hydrate. For discussion of recent x-ray work, see paper by R. S. Vold, *J. Am. Chem. Soc.*, 1943-4.

Similar sequences of mesomorphic forms have been investigated for the sodium salts of fatty acids from C_6 to C_{20}[42, 43, 44]. This study provides more information about the individual transitions from the variation of transition temperature, heat of transition, volume increment of transition, etc., with the number of carbon atoms in the molecule. It appears that the sequence of phases is not always the same from soap to soap, so that the problem is more complex than originally supposed.

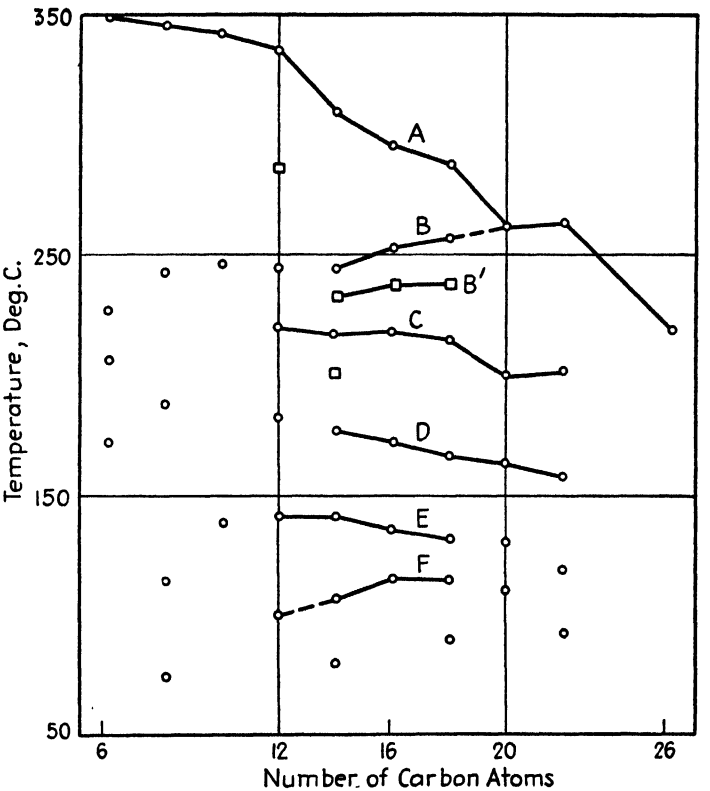


FIGURE 6. Transition Temperatures of Sodium Soaps. ○, visual and dilatometric values; □ calorimetric results significantly different from the former. Lines connect transitions presumed to involve similar changes in structure.

Figure 6 summarizes the results that have been obtained. Wherever there is any evidence that two transitions are of the same kind, *i.e.*, at least one of the two phases involved has the same structure for both soaps, the corresponding points have been connected by lines. The nature of the evidence is quite varied, similarities in appearance, volume increment, change in heat content, mutual solubility, and regular variation of temperature with number of carbon atoms all having been considered.

The transitions for sodium stearate are similar to those for sodium palmitate, except for the existence in the stearate of an additional change occurring prior to the formation of subwaxy soap. Though absent for sodium palmitate, this transition appears again for sodium myristate. It is absent for sodium laurate and apparently also for sodium arachidate. McBain has proposed to call this transition the formation of "supercurd phase."

Reference has already been made to the apparent occurrence of the neat-subneat

transition for C_{14}, C_{16} and C_{18} at a temperature lower according to calorimetric evidence than according to visual and dilatometric results. One or two other unexplained discrepancies of this character appear in Figure 6.

Aside from the additional transitions noted above, the sequence of phases appears to be similar for sodium myristate, palmitate and stearate, but not for sodium laurate. The formation of subwaxy soap from the crystal form stable at room temperature involves a change in heat content of the same order of magnitude for C_{14}, C_{16}, and C_{18}. For C_{12} the value is considerably smaller. This may reflect some difference between the crystal form of C_{12} and the others, since the transition from subwaxy soap to waxy soap is of the same character for all four. At the next transition, waxy-superwaxy, laurate behaves differently from C_{14}, C_{16}, and C_{18} in that there is for laurate a change in heat content of about 600 cal/mole. At the superwaxy-subneat transition all four behave similarly, while at the subneat-neat transition C_{12} again deviates, in this instance having no detected change in heat content as compared to a value of about 100 cal/mole extrapolated from the results for C_{14}, C_{16}, and C_{18}. That differences in behavior between C_{12} and the others should appear only at alternate transitions is at present quite inexplicable.

The curve of melting point as a function of the number of carbon atoms in the molecule is far from smooth. The break at C_{20} (sodium arachidate) can be ascribed to the fact that transformation to isotropic liquid takes place from subneat soap at and beyond C_{20}, but from neat soap below C_{20}. A further break between C_{12} and C_{14} has not been explained. It is, however, possibly to be attributed to differences in the nature of the liquid phase. For myristate, palmitate and stearate, the liquid is very viscous and contains microscopically visible patches of faintly anistropic material, while for laurate the liquid is clear and mobile. The residual anisotropy diminishes with increasing temperature. In fact, Vorländer once reported 316° instead of 295° as the true melting point of sodium palmitate.[43]

As the number of carbon atoms in the molecule increases, the temperature of final melting to isotropic liquid decreases, while that of the formation of a mesomorphic phase from the crystal increases; that is, the temperature interval in which mesomorphic forms exist is narrowed down. Presumably a soap of sufficiently high molecular weight would exhibit a single melting point, similar to the limiting value of about 117° C for other paraffin chain compounds.[61] Consequently, as chain length increases, the temperatures of the several transitions are crowded together. At C_{20} the curve for the subneat-neat transition joins that for the melting points. Therefore, at C_{20} melting involves the same change in structure in a single step that takes place in two discrete steps for soaps of shorter chain length.

Transformation from crystal to liquid should require the same increase in heat content, regardless of the number of steps or separate transitions by which the process occurs. Values found for the sum of the heat effects of all the transitions are of the same order of magnitude as experimental results for the heats of fusion of the corresponding fatty acids (12,570 cal/mole for sodium palmitate, 13,030 cal/mole for palmitic acid [62]).

Transitions which involve loosening of the hydrocarbon chains from one another can be expected to show considerable variation in heat effect with chain length. This is the case for the transition from crystal to subwaxy soap and from subwaxy soap to waxy soap. About two-thirds of the total heat of fusion is associated with these two transitions. Heats of transition for the remaining four changes are smaller, and vary but little with chain length (for the three substances sodium myristate, palmitate and stearate). Presumably the structural rearrangements are here largely concerned with the heads of the molecules.

Neat soap, the liquid crystalline form stable at temperatures just below the melting point, probably has a smectic structure. According to Hermann,[36] a liquid crystalline form of thallium palmitate (relation to the neat soap of sodium palmitate un-

known) has the terminal groups lying in definite planes, but randomly rather than regularly arranged. The conic structures (see for example Figure 5a) are further evidence of this structure.[63]

Assignment of similarly detailed structures, even tentatively, to the remaining mesomorphic forms cannot be made without further experimental evidence. One direction of thought is that there are here a number of different kinds of constraints holding the crystal lattice together, and that these can be released one by one at different temperatures. At least six, and perhaps more, discontinuous steps must be thought of as occurring in the transformation from a crystal, containing pairs of oppositely oriented molecules disposed in a perfectly regular manner, to a liquid in which the molecules are, on the average, randomly arranged, with the hydrocarbon chains possibly not fully extended. Moreover, since the sequence of phases is not identical for all the soaps, alternative routes of stepwise lattice breakdown may have to be provided.

Another hypothesis * is that in some of the mesomorphic forms the fundamental unit is larger than the molecule; *i.e.*, two forms may have the same disposition of one molecule relative to its neighbors, and differ only in the character of the secondary structure, or the arrangement of such primary groupings of molecules. Whether or not *two discrete phases* can differ *only* in size or nature of such secondary aggregates can be answered only by further investigation.

In all the mesomorphic forms, microscopic or submicroscopic aggregates certainly do exist. This is evident from their appearance between crossed Nicols, the fact that they scatter light, etc. The existence of such aggregates might be thought of as similar to the mosaic structure of imperfect crystals.[64] The crystal form of the sodium soaps frequently has a fibrous structure in which individual fibers are very long and narrow (40-50 Å).[45] Whether or not there is a characteristic size of aggregate in any of the mesomorphic forms has not been determined.

References

I. Papers on n-hydrocarbons

1. Muller, A., and Saville, W. B., *J. Chem. Soc.*, **127**, 599 (1925).
2. Muller, A., *Proc. Roy. Soc. London*, **A154**, 628 (1936).
3. *Ibid.*, **A120**, 437 (1928).
4. *Ibid.*, **A124**, 319 (1929).
5. *Ibid.*, **A127**, 417 (1930).
6. Piper, S. H., Chibnall, A. C., Hopkins, S. J., Pollard, A., Smith, J. A. B., and Williams, E. F., *Biochem. J.*, **25**, 2074 (1931).
7. Ubbelohde, A. R., *Trans. Far. Soc.*, **34**, 292 (1938).
8. Piper, S. H., *J. Chem. Soc.*, **234** (1929).

II. Papers on n-fatty acids

9. Muller, A., *J. Chem. Soc.*, **123**, 2043 (1923).
10. Muller, A., and Shearer, G., *J. Chem. Soc.*, **123**, 3156 (1923).
11. Gibbs, R. E., *Trans. Chem. Soc.*, **125**, 2622 (1924).
12. Muller, A., *Proc. Roy. Soc. London*, **114A**, 542 (1927).
13. Francis, F., Piper, S. H., and Malkin, T., *Proc. Roy. Soc. London*, **A128**, 514 (1930).
14. La Tour, F. Dupré, "Actualités Scientifiques et Industrielles #424. The Polymorphism of the Fatty Acids." Hermann et Cie, Paris, 1936.
15. Ott, E., and Slagle, F. B., *J. Phys. Chem.*, **37**, 257 (1933).
16. Ott, E., and Wilson, D. A., *Science*, **78**, 16 (1933).
17. Thiessen, P. A., and Stuber, C., *Berichte*, **71**, 2103 (1938).
18. Schoon, T., *Z. physik Chem.*, **339**, 385 (1938).
19. Trillat, J. J., *Ann. de phys.*, **6**, 5 (1926).
20. Piper, S. H., *J. Chem. Soc.*, **1929**, 348.

III. Papers on fats and esters

21. Malkin, T., *J. Chem. Soc.*, **1931**, 2796.
22. Malkin, T., *Proc. Roy. Soc. London*, **A128**, 224 (1930).
23. Smith, J. C., *J. Chem. Soc.*, **1931**, 802.
24. Phillips, J. W. C., and Mumford, S. A., *J. Chem. Soc.*, **1931**, 1732.
25. Ferguson, R. H., and Lutton, E. S., *Chem. Rev.*, **29**, 355 (1941) (excellent bibliography).
26. Kohlhaas, R., *Feit. Krist.*, **A98**, 418 (1938).
27. Shearer, G., *J. Chem. Soc.*, **123**, 3152 (1923).
28. Baker, W. O., and Smyth, C. P., *J. Am. Chem. Soc.*, **60**, 1229 (1938).

IV. Papers on n-aliphatic alcohols

29. Wilson, D. A., and Ott, E., *J. Chem. Phys.*, **2**, 231, 239 (1933).
30. Baker, W. O., and Smyth, C. P., *J. Am. Chem. Soc.*, **60**, 1229 (1938).

V. Papers on n-paraffin chain ketones

31. Saville, W. B., and Shearer, G., *J. Chem. Soc.*, **127**, 591 (1925).

* Suggested to the authors by J. Alexander.

32. Ubbelohde, A. R., and Oldham, J. W. H., *Nature*, **142**, 74 (1938).
33. Oldham, J. W. H., and Ubbelohde, A. R., *Proc. Roy. Soc. London*, **A176**, 50 (1940).
VI. Papers on salts of n-fatty acids (soaps)
34. Piper, S. H., *J. Chem. Soc.*, **1929**, 234.
35. Trillat, J. J., *Ann. de phys.*, **6**, 5 (1926).
36. Hermann, K., *Trans. Far. Soc.*, **29**, 972 (1933).
37. Thiessen, P. A., and Klenck, J. V., *Z. physik. Chem.*, **A174**, 335 (1935).
38. Thiessen, P. A., and Ehrlich, E., *Z. physik. Chem.*, **B19**, 299 (1932).
39. Thiessen, P. A., and Spychalski, R., *Z. physik. Chem.*, **A156**, 435 (1931).
40. Thiessen, P. A., and Stauff, J., *Z. physik. Chem.*, **A176**, 397 (1936).
41. Vold, R. D., and Vold, M. J., *J. Am. Chem. Soc.*, **61**, 808 (1939).
42. Vold, M. J., *J. Am. Chem. Soc.*, **63**, 160 (1941)
43. Vold, M. J., Macomber, M. and Vold, R. D., *J. Am. Chem. Soc.*, **63**, 169 (1941).
44. Vold, R. D., *J. Am. Chem. Soc.*, **63**, 2915 (1941).
45. Marton, L., McBain, J. W., and Vold, R. D., *J. Am. Chem. Soc.*, **63**, 1990 (1941).
46. McBain, J. W., "Colloid Chemistry," ed. J. Alexander, Vol. I, p. 137, New York, Chem. Catalog Co. (Reinhold Publishing Corp.), 1926.
VII. Liquid structure in paraffin chain compounds
47. Warren, B. E., *Phys. Rev.*, **44**, 969 (1933).
48. Kauzmann, W., and Eyring, H., *J. Am. Chem. Soc.*, **62**, 3113 (1940).
49. Meyer, K. H., and Lühedemann, R., *Helv. Chim. Acta*, **18**, 307 (1935).
50. Huggins, M. L., *J. Phys. Chem.*, **43**, 1083 (1939).
51. Hildebrand, J. H., *J. Am. Chem. Soc.*, **59**, 794 (1937).
52. Stewart, G. W., *Phys. Rev.*, **31**, 174 (1928).
58. Stewart, G. W., and Morrow, R., *Phys. Rev.*, **30**, 232 (1927).
59. Stewart, G. W., *Proc. Nat. Acad. Sci.*, **13**, 787 (1927).
60. Morrow, R., *Phys. Rev.*, **31**, 10 (1928).
VIII. Mesomorphic forms and the melting process
53. Friedel, G., "Colloid Chemistry," ed. J. Alexander, Vol. I, p. 102, New York, Chem. Catalog Co. (Reinhold Publishing Corp.), 1926.
54. Bernal, J., and Crowfoot, D., *Trans. Far. Soc.*, **29**, 1032 (1933).
55. Baker, W. O., and Smyth, C. P., *J. Am. Chem. Soc.*, **61**, 1695 (1939).
56. Lawrence, A. S. C., *J. Roy. Microscop. Soc.*, **58**, 30 (1938); see also ref. 33, 5.
57. Spaght, M., Thomas, S. B., and Parks, G. S., *J. Phys. Chem.*, **36**, 882 (1932).
61. Timmermans, J., *Bull. Soc. chim. belg.*, **28**, 392 (1919).
62. Garner, W. E., Madden, F. C., and Rushbrooke, J. E., *J. Chem. Soc.*, **1926**, 2491.
63. Bragg, W. L., *Trans. Far. Soc.*, **21**, 1056 (1933).
64. Zwicky, F., *Proc. Nat. Acad. Sci.*, **15**, 253 (1929).

Polymerization

H. MARK
Polytechnic Institute, Brooklyn, N. Y.

Definitions and General Survey

Polymerization processes are chemical reactions which combine small molecules by chemical valences to form very large ones. They occur with both organic and inorganic compounds, and usually result in solid materials of crystalline or vitreous character having outstanding thermal, mechanical, electrical, and optical properties. According to W. H. Carothers,[2] there are two main types of polymerization reactions: (a) A-polymerizations and (b) C-polymerizations.

The first group represents association polymerization which at present also is, for short, called polymerization. The reaction consists in a successive addition of molecules containing double or triple bonds, until a giant molecule is formed, in which the different parts are linked together by main valence bonds. Examples: Liquid sulfur polymerizes to long-chain molecules which represent elastic sulfur; ethane polymerizes to a long-chain substance called polythene.

The molecule which undergoes polymerization is called the *monomer*. The number of monomers linked together in one molecule is termed "degree of polymerization" or "polymerization degree" (P). If P is a number above 100, one speaks of high polymers; the lowest numbers of such a series of polymers are usually termed by the Greek prefix for their degree of polymerization, such as: dimer, trimer, tetramer, etc.

A series of molecules built up from the same monomer, but having different degrees of polymerization, belongs to one *poly-homologous series*. Many such series

are known; in some cases, polymerization degrees as high as 10,000 have been measured experimentally.

Such large molecules are called *macromolecules,* or *giant molecules;* if they have the form of a chain, they are also known as *main valence chains.* It has been observed that under certain conditions, two or more monomers are capable of entering the same macromolecule; this is called *copolymerization,* or *interpolymerization.*

Condensation (C) polymerizations take place if two or more bi- or multi-functional molecules combine, with the elimination of water. Such processes are also called *polycondensations.* Examples: NaOH and orthosilicic acid form polysodium silicate; glycol and adipic acid give polyglycol adipate.

The molecules (two or more) which enter the reaction are called *monomers.* The reaction products are again termed dimers, trimers, high polymers and eupolymers according to their degree of polymerization. The resulting molecules again are macromolecules or giant molecules.

If only two-functional monomers participate, one gets only *chain polymers,* which are also called covalent chains. If one or more of the monomers are tri- (or higher) functional, two- or three-dimensional networks are obtained, and the degree of polymerization exceeds any measurable limit. In such a case, one observes *gelation.* Examples: hexamethylene diamine and adipic acid build up linear macromolecules belonging to one homologous series. Substances of this type are called *nylon.* Glycerin and phthalic acid build up macromolecules of a non-linear type having complicated three-dimensional structures with very high degrees of polymerization. Such substances are called *glyptals.*

Table 1 shows a few atomic groups which undergo A-polymerization with comparative ease and are therefore used to a large extent in the commercial production of synthetic elastomers and plastics. The first column contains the characteristic group of atoms which makes the material polymerizable; the second gives the name of one of the more important substances containing this group; and the third column lists a few trade names of products being built up by the systems under consideration.

Table 1.

Atomic Groups	Name of Monomer	Trade Names of Polymers
$CH_2 = CH_2$	ethylene	Polythene
$CH_2 = CHCl$	vinylchloride	Koroseal
$CH_2 = CH - CH = CH_2$	butadiene	Buna
$CH_2 = CCl - CH = CH_2$	chlorobutadiene	Neoprene
$CH_2 = CHOOC \cdot CH_3$	vinylacetate	Vinylite
$CH_2 = CH\ (C_6H_5)$	styrene	Victrone
$CH_2 = CH - CO \cdot OCH_3$	acrylic ester	Lucite

In Table 2 is the corresponding information about C-polymerizations. The first column shows pairs of two- or three-functional molecules, which are capable of condensing; the second column gives the names of a few products.

Table 2.

Pair of Polyfunctional Monomers	Trade Names of Polymers
Glycerol + phthalic anhydride	Glyptal
Formaldehyde + phenol	Bakelite
Formaldehyde + urea	Pollopas
Adipic acid + hexamethylene diamine	Nylon

Characterization of Polymers

One of the most important problems is the adequate characterization of macromolecules. Their empirical chemical formula is given by the formulas of the monomers and can, in most cases, be obtained by elementary analysis. Much more difficult is the determination of the size and shape of the macromolecules.

In practically all cases, there is no definite molecular weight, but the sample has to be described by a molecular size or molecular weight distribution curve, which gives the number of molecules having a polymerization degree between P and $P + dP$ as a function of P. It is called the number distribution curve of the molecular weight. More frequently, the weight distribution curve is employed; this gives the total weight of the material contained in molecules of a polymerization degree between P and $P + dP$ plotted against P. Fig. 1 shows such a weight distribution curve of a polystyrene sample prepared at 130° C. It can be seen that this material is rather heterogeneous, inasmuch as it contains a considerable amount of substance in the range as low as $P = 200 - 300$, and also a fraction as high as $P = 3000$. All natural and synthetic polymers are more or less non-uniform, and it seems to be an important question to what extent this heterogeneity affects their technical properties.

Sometimes one is satisfied to use *average values* instead of the whole distribution curve, and distinguish between the number average molecular weight, M_n, and the weight average molecular weight M_w. Both values can be directly derived from the corresponding distribution curves.[4, 9, 11, 16]

The following experimental methods are used to determine these average values:

(a). The measurement of the osmotic pressure in very dilute solution (weight concentration c_w below 0.2 per cent) allows the computation of the number average molecular weight with the aid of the Van't Hoff equation:[10, 14]

$$M_n = \frac{RT}{p} \cdot c \tag{1}$$

where

R = gas constant 7.2×10^7 erg/gram
T = temperature (°C)
p = osmotic pressure in dynes per sq cm
c = concentration in g per cc

The usual way to obtain p is direct measurement in an osmometer with a semipermeable membrane.

(b). The determination of the number of end groups of a chain polymer by chemical means also permits the calculation of the number average molecular weight of the material. The end groups usually are either double bonds or aldehyde groups which give certain characteristic chemical reactions, and can thereby be determined analytically.[1, 3, 5, 6, 7, 8, 17]

Aldehydic end-groups of, *e.g.*, cellulose chains can be determined by their reducing power against Fehling's solution; double bonds as end-groups in polystyrene, by their ability to add bromine. Acidic or alkaline end-groups of poly-condensation products, such as nylon, can be titrated with monoacidic bases or monobasic acids.

(c). The specific viscosity of a very dilute solution of a high polymer (weight concentration below 0.1 per cent) allows the computation of the weight average molecular weight with the aid of the Staudinger equation:[17]

$$\eta_{sp} = K_m \cdot M_w \cdot c_m \tag{2}$$

η_{sp} = specific viscosity (relative viscosity increase due to the dissolved material).
K_m = characteristic viscosity constant
c_m = concentration in moles of the monomer per 1000 cc

Equation (2) is an empirical law, which holds with a certain approximation and is sometimes useful to figure out an approximate value for M_w. Staudinger has determind the K_m-constants for certain important high polymers in various solvents.[17]

Instead of the specific viscosity, one can use to greater advantage the intrinsic viscosity, according to E. O. Kraemer.[9, 10]

(d). The ultracentrifuge permits determination of the weight average molecular

weight by measuring the rate of sedimentation. Still another average value can be obtained by the sedimentation equilibrium method. Also, in these cases it is essential that the solutions investigated should be very dilute (weight concentration below 1.0 per cent).

Table 3 shows a series of values of weight average molecular weights for certain natural and synthetic high polymers. It can be seen that these values cover a fairly wide range and represent materials of considerable technical interest.

Table 3.

Substance	Polymerization degree	Molecular Weight
Rubber	2000-4000	140,000-280,000
Cellulose	2000-3000	300,000-450,000
Polyisobutylene	1200-2400	60,000-130,000
Polystyrene	1000-2500	100,000-250,000
Polyvinyl acetate	400-1200	30,000- 90,000
Polyglycerol phthalate	100- 500	30,000-150,000

The distribution curves can be determined either in the ultracentrifuge, or by fractionate solution or precipitation of a given sample by the use of appropriate solvents and precipitants.

Another characterization is possible by distinguishing between polymers which are built up of chain molecules and those which present more complicated systems netted and cross-linked in two or three dimensions. It is not yet possible to give a completely clear distinction between chain- and net- or space-polymers; but generally it can be said that chain polymers are usually soluble in appropriate liquids, while the latter substances merely swell. A certain amount of the material goes into solution, but the main mass simply absorbs the swelling agent without losing its own coherence.

Such three-dimensional polymers have exceedingly high molecular weights, and with a certain degree of justification the whole sample may be considered one single giant molecule. As they are insoluble, it is as yet impossible to carry out any of the above-mentioned methods for determination of the degree of polymerization, and a suitable method for their characterization has not yet been developed.

The Formation of Macromolecules

The mechanism of the formation of macromolecules has not yet been elucidated in all its details, but it is possible to give at least a general description of the most important elementary steps.

In the case of C-polymerizations, the situation seems to be comparatively simple. According to Carothers [2] and Flory,[4] these reactions consist of a very large number of elementary steps of esterification, anhydrization, etherification or amidification. All consecutive steps are identical with the foregoing, and the whole chain or space-polymer is held together by covalent chemical bonds of identical character.

It seems that A-polymers are produced by a considerably more complicated mechanism. Its essential elementary steps seem to be (a) the activation or start reaction; (b) the propagation or growth reaction; (c) the termination or break reaction.

In certain cases, still other steps, such as branching, ring closure, chain transfer, and so on, may occur and interfere with the steps mentioned above.

Activation processes. It seems that the first step of a polymerization reaction is the activation of certain molecules of the monomer. This activation can be effected by the absorption of a light quantum by simple thermal collision or by the aid of a catalyst. Table 4 shows the activation energies for the starting reaction for a series of polymerization processes. It can be seen that the order of magnitude corresponds to that of a chemical process during which covalent bonds are opened and closed.

We may therefore formulate the activation step as the formation of a radical or biradical, according to the equations:

$$CH_2 = CH_2 + CH_2 = CH_2 \rightarrow -CH_2 - CH_2 - CH_2 - CH_2 -$$
$$CH_2 = CH_2 + h\nu \rightarrow -CH_2 - CH_2 -$$
$$CH_2 = CH_2 + \text{catalyst} \rightarrow \text{catalyst} - CH_2 - CH_2 -$$

Table 4.

Reaction	Activation Energy
Polymerization of	
acetylene	40,500
ethylene	35,000
styrene	26,000
vinylacetate	25,000
indene	26,000
butadiene	24,700

In some cases, however, it seems to be more appropriate to consider the activated state as being the double bond in one of its excited electronic states, without actually having the double bond opened:

$$CH_2 = CH_2 + CH_2 = CH_2 \rightarrow CH_2 = CH_2 + CH_2 = {}^*CH_2$$
$$CH_2 = CH_2 + h\nu \rightarrow CH_2 = {}^*CH_2$$
$$CH_2 = CH_2 + \text{catalyst} \rightarrow CH_2 = {}^*CH_2$$

In the first case, the starting reaction is very slow as compared with all other elementary steps, and each nucleus will grow very quickly in a long-chain molecule as soon as it has been formed. The average lifetime of the activated centers (growing chains) is presumably below $1/100$ of a second. This free-radical mechanism seems to work during the heat polymerization of styrene, vinylchloride, vinylacetate, acrylic esters, etc.

The activated double bond mechanism, on the other hand, has been observed with the photochemical polymerization of methacrylate and the sodium-catalyzed polymerization of butadiene.

The formation of nuclei with successive growth of these nuclei is a process which shows a close analogy to the phenomenon of crystallization. G. Tammann in his book "Crystallization and Fusion" has developed the idea that the presence of certain centers or nuclei is essential for crystallization and that the slow formation of those nuclei is responsible for the phenomenon of supercooling.

The propagation reaction. If a monomeric molecule collides with an activated center, a process of growth takes place according to the equation:

$$\text{Chain} - CH_2 - CH_2 - + CH_2 = CH_2 \rightarrow \text{Chain} - CH_2 - CH_2 - CH_2 - CH_2 -$$
$$\text{Chain} - CH = {}^*CH_2 + CH_2 = CH_2 \rightarrow \text{Chain} - CH_2 - CH_2 - CH = {}^*CH_2$$

In the case of the radical chain mechanism the activation energy of propagation [11, 16] is very low (around 4000 or 5000 cal per mol monomer), and this accounts for the fast propagation and the short average lifetime of the chains.

In the case of the activated double bond mechanism, the activation energy for the propagation seems to be somewhat higher (6000-10000 cal per mol of the monomer), which would account for the higher stability of the intermediate products.[11, 18]

The experimental measurement of those activation energies is accomplished by creating a certain number of active molecules in the monomer and then determining the rate of polymerization at different temperatures.

It seems that one and the same monomeric substance can undergo both types of polymerization according to the experimental conditions, and that even under given conditions both types of growth can take place simultaneously.

The termination reaction. Experience shows that chains do not grow indefinitely, but are terminated by some additional chemical process. Radical chains are very sensitive, and can be broken in many different ways:

(1) A collision of two free radical chain ends leads to a mutual saturation and to the disappearance of two active centers. This process can be expressed by the formula:

$$R - CH_2 - + - CH_2 - R' \rightarrow R - CH_2 - CH_2 - R'$$

(2) A collision of a free radical chain end with an unactivated monomer usually leads to a step of propagation, but sometimes a so-called chain transfer takes place according to the equation:

$$\text{Chain} - CH_2 - CH_2 - + CH_2 = CH_2 \rightarrow \text{Chain} - CH_2 - CH_3 + CH_2 = CH - \text{ or}$$
$$\text{Chain} - CH_2 - CH_2 - + CH_2 = CH_2 \rightarrow \text{Chain} - CH = CH_2 + CH_3 - CH -$$

(3) The two free ends of the chain can saturate each other by the formation of a poly-membered ring.

(4) Many substances—deliberately added or present as impurities—react with the free valencies at the chain ends and therefore stop the further growth of the molecule. They are usually called inhibitors or stabilizers.

(5) The migration of a hydrogen atom from one end of the chain to the other also leads to a removal of both free valencies according to the equation:

$$- CH_2 - CH_2 - \text{Chain} - CH_2 - CH_2 - \rightarrow CH_2 = CH - \text{Chain} - CH_2 - CH_3$$

and hence to a termination of growth of the molecule under consideration.

It has been possible[1, 4, 10, 11, 12, 13, 15, 16] to work out general formulas for the interference of the different elementary steps listed above, and it can be said that in several cases, a satisfactory agreement with the observed data has been obtained. It seems, therefore, that the above scheme covers the majority of the observed polymerization reactions and provides a possibility of dividing them into their different elementary steps.

References

1. Burk, R. E., Thomson, H. E., Weith, A. J., and Williams, I., "Polymerization," Reinhold Publishing Corp., New York, 1937.
2. Carothers, W. H., "Collected Papers," Interscience Publishing Co., New York, 1940 .
3. Ellis, C., "Chemistry of Synthetic Resins," Reinhold Publishing Corp., New York, 1935.
4. Flory, P. J., *J. Am. Chem. Soc.*, **58**, 1877 (1936); **61**, 3334 (1939); **62**, 1057, 1064, 1561, 2255, 2261 (1940).
5. Freudenberg, K., "Tannin, Cellulose, Lignin," Berlin, 1933.
6. Haworth, W. N., "The Constitution of Sugars," London, 1929.
7. Heuser, E., "Cellulose Chemistry," New York, 1942.
8. Hibbert, H., *J. Am. Chem. Soc.*, **61**, 1905, 1910, 1912 (1939).
9. Kraemer, E. O., and Lansing, W. D., *Ibid.*, **55**, 4319 (1933).
10. Mark, H., "Physical Chemistry of High Polymers," Interscience Publishing Co., New York, 1941.
11. Mark, H., and Raff, R., "High Polymeric Reactions," Interscience Publishing Co., New York, 1941.
12. Marvel, C. S., *J. Am. Chem. Soc.*, **61**, 3156, 3241, 3244 (1939); **62**, 45, 2666, 3499 (1940).
13. Melville, W. H., *Proc. Roy. Soc. London (A)*, **163**, 511 (1937); **175**, 392 (1940).
14. Meyer, K. H., "High Polymeric Systems," Interscience Publishing Co., New York, 1942.
15. Norrish, G. R. W., and Brockman, E. F., *Proc. Roy. Soc. London (A)*, **171**, 147 (1939).
16. Schulz, G. V., *Z. Phys. Chem. (B)*, **32**, 27 (1936); **34**, 187 (1936); **43**, 25, 385 (1939); **44**, 227 (1939); **47**, 155 (1940).
17. Staudinger, H., "Hochpolymere Organische Verbindungen," J. Springer, Berlin, 1932.

Fundamental Aspects of the Elasticity of Rubber and of Rubberlike Materials

EUGENE GUTH
Department of Physics, University of Notre Dame, Notre Dame, Indiana

Introduction

Rubber is one of the most important colloidal materials. In addition to its widespread use in daily life, it may be considered as a simple prototype of important biological substances, like the proteins. Rubber is much more unstable than a metal or a glass. This very instability permits an amazing versatility of change in its structure and properties. It can be made electrically conducting or brought into a glassy state. It belongs to a class of substances which nature uses as a framework for all living things. Its study, though developed by the necessities of daily life, is deeply interconnected with all those sciences which make a fundamental inquiry into organic life: biochemistry; biophysics; physiology and pathology, dealing with nerve and muscle; vitamins and proteins.

Rubber chemistry and technology may be considered as an important link in the monumental attempt of human mind to unveil the mystery of organic life. The development of rubber technology starts with the discovery of the latex of *Hevea brasiliensis,* for in this tree nature is a manufacturer of rubber. After discovering the latex and its transformation into solid crude rubber by the typical colloidal process of coagulation, the next important step was the discovery of vulcanization by Goodyear in 1839. Though unknown to Goodyear, vulcanization is also an imitation of an artifice of nature to preserve an organic material from decay and make it more enduring. Wool protein, or keratin, contains sulfur as a protective agent. Hair consists of such "vulcanized" protein, and may last a long, long time after the "crude" (*i.e.* unprotected) proteins of an animal body have disintegrated. The synthesis of rubberlike materials, though (like vulcanization) developed mainly through needs of daily life, is a successful attempt of mankind to replace nature as a manufacturer of rubber or to find substitutes for this substance.

The aim of this paper is to present a simple picture of the interrelations of the elasticity and structure of rubber and of rubberlike materials. The theoretical work by H. M. James and the author[1] and the experimental work by the author's associates[2] at Notre Dame University will principally be considered. Unfortunately there is neither time nor space to permit a critical appraisal of the excellent work of other investigators.

A series of interesting articles appeared on rubber in Vol. III and Vol. IV of this series * which may very well form a background to the present article. The most complete book on all aspects of rubber is "Chemistry and Technology of Rubber" edited by C. C. Davis and J. T. Blake.†

The most characteristic properties of rubber are, undoubtedly, its long-range reversible elasticity and its anomalous thermoelastic behavior. These properties are so characteristic that all materials having them may be called "rubber," or elastomers, regardless of their chemical composition. When, in the following, we speak of rubber, we mean both natural and synthetic rubber possessing the aforementioned characteristic properties to a marked extent. Most practical uses of rubber are connected with its high reversible elasticity.

A substance possessing such an unusual elasticity must obviously differ greatly in structure from other materials, and conclusions regarding its structure may be drawn from its observed elasticity. To anticipate one of our main points: the elas-

* The articles by P. P. von Weimarn in Vol. III, and by E. B. Spear, Vol. IV; also in the same volume the articles by A. van Rossem, R. P. Dinsmore and N. A. Shepard.

† Reinhold Pub. Corp., N. Y., 1937.

ticity of rubber may be explained quantitatively on the basis of a very general model. Without a detailed knowledge (such as we do not now possess) of the structure of unstretched rubber, comparison of theory and experiment permits conclusions about some details of the model.

Fundamental Experiments on Elasticity

The manifold uses of rubber in daily life are based mainly on its remarkable physical properties. In order to obtain compounds with desirable properties for specific applications, compounds of varying composition are made, and then a series of physical tests is applied to find out the compositions needed for a particular application. The compounding and curing, *i.e.*, the chemical part of this process though important, is more incidental. Compounds of widely different composition may give the same essential physical properties. The physical tests applied may be characterized here merely by their name as, for instance, rebound, plastometer, flexometer, stress-strain curve, tensile strength, and so on. However, in these physical tests rubber undergoes a rather complicated deformation. This makes it difficult to correlate directly the results of the usual physical tests with the structure of rubber. Moreover, some of these tests may be closely related so that the results of the one test can be predicted from those of another.* From the practical point of view it is desirable to reduce the number of tests to a minimum, using only those giving independent results. From the theoretical point of view it is desirable to design a set of simple, fundamental experiments on elasticity of rubber and the results of such experiments should show a simple correlation to the structure of rubber. In turn these fundamental experiments may indicate an evaluation of the physical tests in terms of rubber structure. So we have the sequence: rubber structure → fundamental experiments → physical tests.

Fundamentally, rubber shows three thermoelastic effects; (1) it generates heat on fast stretching; (2) if stretched by a constant load (isotonic stretch) it contracts visibly when heated; (3) the stress in rubber kept at constant extension (isometric stretch) increases with rising temperature. Another fundamental fact is the S-shape of the stress-strain curve of rubber and its continuation for compression.

Unaware of earlier observations of Gough (1805) and Page (1847), Joule (about 1855) discovered independently the first thermoelastic effect of rubber, namely that it generates heat on fast stretching. William Thomson (Lord Kelvin) then pointed out to Joule that, according to thermodynamics, the second thermoelastic effect must occur, namely that rubber stretched by a constant load will contract on heating. As a matter of fact, this prediction of Lord Kelvin constituted one of the earliest applications of the second law of thermodynamics and Joule promptly verified the prediction experimentally.

Fig. 1 shows the change in temperature which Joule observed on fast stretching of a strip of vulcanized rubber.[3] It is seen that a small cooling was observed for the lower elongations. However, at a critical extension, defining a thermoelastic inversion point, the change in temperature becomes zero. At extensions above this critical extension there is a rise in temperature. Normal solids, like metals, glasses and wood, show a cooling effect on fast stretching. The elastic behavior of rubber is normal only at small extensions, but it is anomalous above the thermoelastic inversion point.

Fig. 2 shows the contraction which Joule observed on heating a sample of vulcanized rubber. Again a thermoelastic inversion point appears.

The third thermoelastic effect, namely the rise of stress in isometrically (*i.e.*, at

* To mention an example from the related field of metals: Hardness measurements constitute one of the simplest non-destructive types of tests for metals. Now, for steel there is a correlation between hardness and tensile strength. Thus in many cases it will not be necessary to carry out the (destructive) tensile strength test; its result may be predicted from the measured hardness.

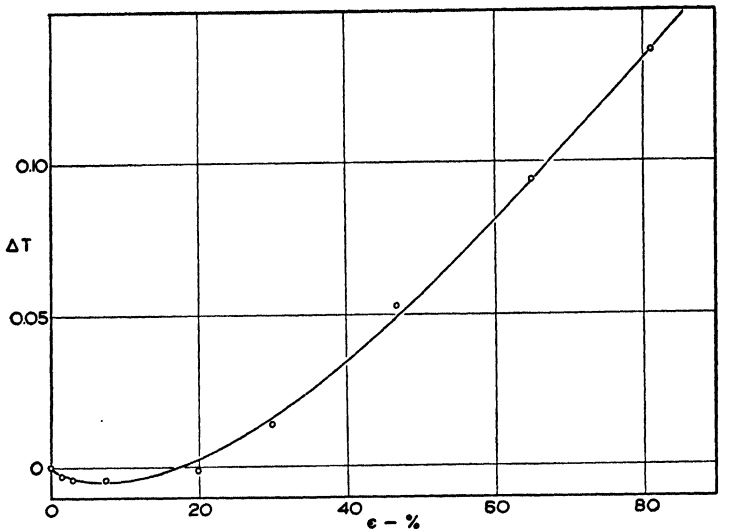


FIGURE 1. Change of temperature ΔT on fast stretching as a function of the strain ϵ according to Joule.

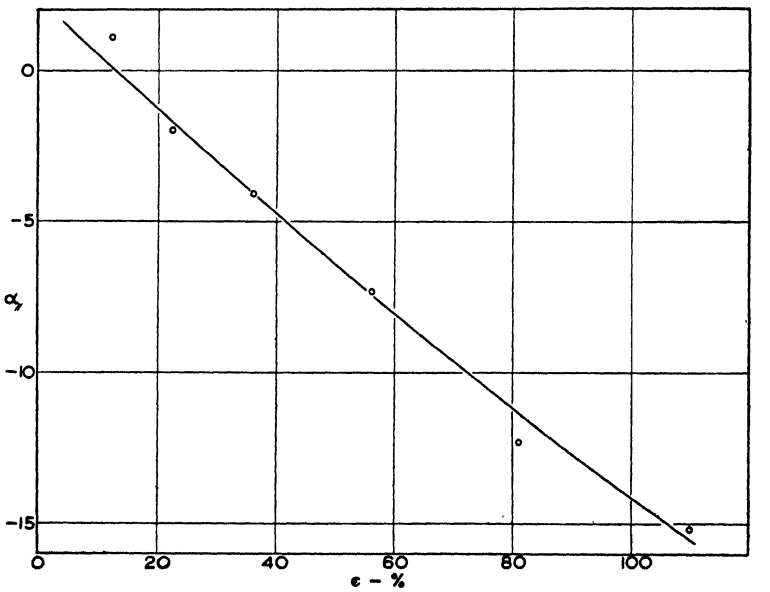


FIGURE 2. Linear thermal expansion coefficient $\alpha_{,,}$ parallel to the direction of the stress as a function of the strain ϵ according to Joule.

constant length) stretched rubber with increasing temperature, was not investigated until recently.

Figs. 3a and 3b show isometrics obtained by the author's associates at Notre Dame. Again the existence of a thermoelastic inversion point is clearly evident.

The first stress-strain curves on rubber were taken by Villari. Fig. 4 shows a stress-strain curve obtained by Villari for rubber cord. This curve exhibits nicely the characteristic S-shape of the rubber stress-strain curves. The dashed compression branch is taken from data of Sheppard and Clapson.[4] The rubber compound they investigated differed of course from that used by Villari.

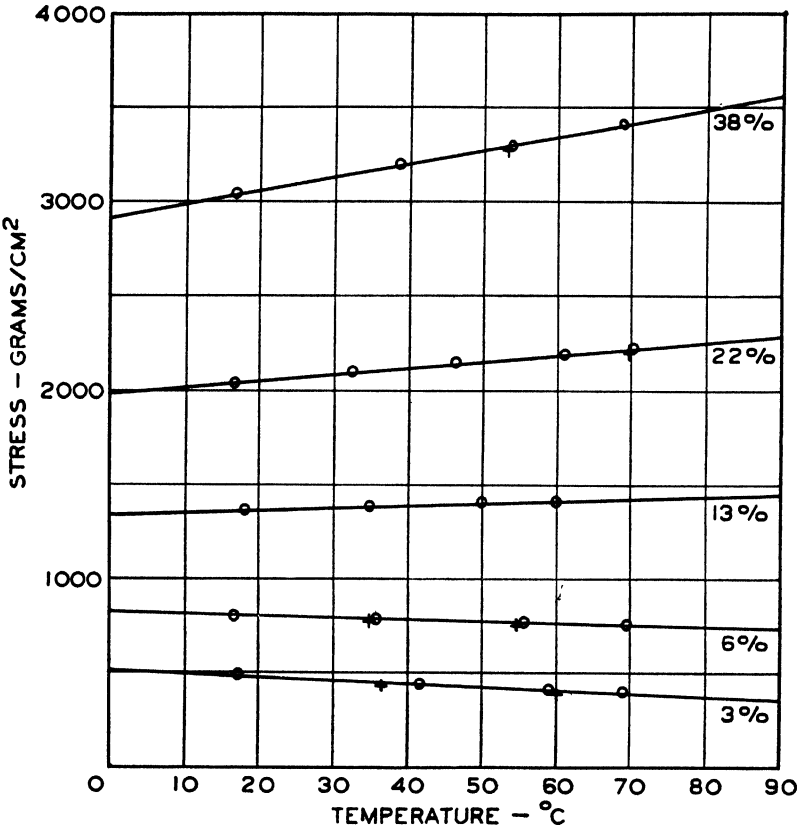


FIGURE 3a. Stress-temperature curves (isometrics) at various extensions for an unaccelerated pure gum stock.

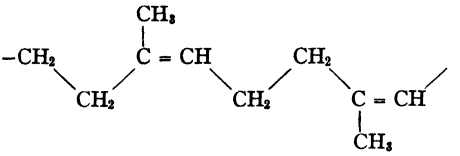
Any theory of rubber elasticity must explain the three thermoelastic effects and the S-shape of the stress-strain and compression curve quantitatively. Such a theory was developed by H. M. James and the author [1] and will be described briefly in the later sections of this article. The next sections contain a short discussion of the molecular structure of rubber.

Structure of Rubber

The Rubber Molecule. There are five main questions connected with the structure of rubber. The first concerns its elementary formula. This may be deduced from purely chemical facts and turns out to be $(C_5H_8)_n$. The second question pertains to the structure within the (C_5H_8) groups. Chemistry shows that this structure is:

$$-CH_2 - CH = C(CH_3) - CH_2-$$

containing one double bond and three single bonds (isoprene). The third question concerns the relation of the isoprene units to one another, *i.e.*, whether rubber is in the *cis*- or the *trans*-configuration; the fourth question relates to the degree of polymerization *n*. Physicochemical data, especially x-ray analysis, show that rubber is in the *cis*-configuration and that *n* is very large. The rubber "molecule" consists of a long chain of isoprene units:



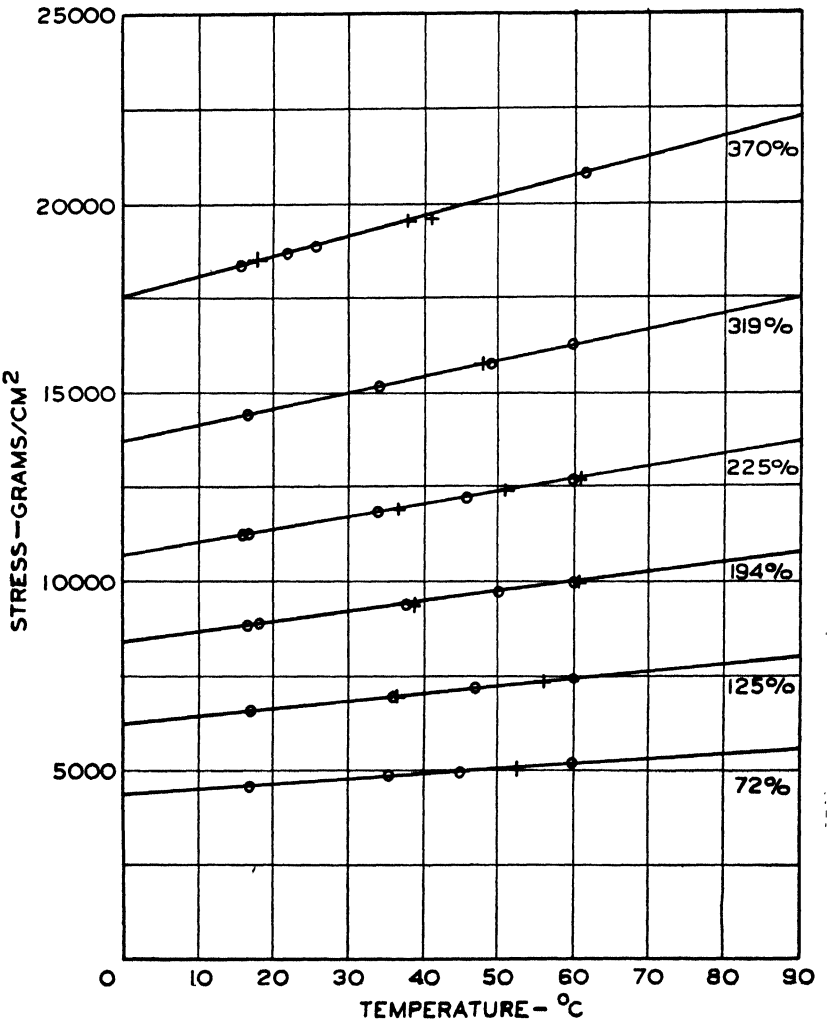


FIGURE 3b. Same as 3a for larger extensions.

These four questions are the only ones which arise in more or less dilute solutions of rubber.

Since rubber is a solid there must be more or less perfect connections, or linkages, between the hydrocarbon chains; otherwise rubber would behave as a liquid and would flow if deformed. The fifth question we now raise pertains to the structure of bulk rubber.*

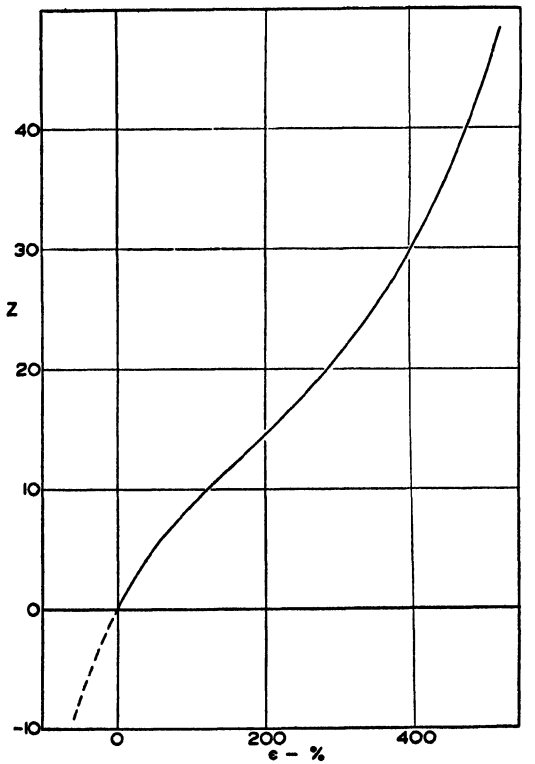


FIGURE 4. Stress-strain curve for a rubber cord according to Villari. The dashed curve represents compression data by Sheppard and Clapson.

For solid bulk rubber the concept of a rubber "molecule" will not in general be applicable. The situation is similar to that in the case of diamond, for instance. One may consider a whole macroscopic piece of diamond as a single giant molecule, but there is little physical sense in considering "molecules" of diamond, since diamond forms a regular network, or lattice.

The Theory of Gelation. Until recently the term "gel" was only vaguely defined and any theories of gelation were necessarily qualitative in nature. Recently, however, definite progress leading to a statistical treatment of "gelation" was made mostly by Flory,[7] and followed by Stockmayer.[7'] We shall follow closely in this section Flory's excellent review. The term "gel" will be applied to insoluble, infusible three-dimensional polymers. A polymerization process leads to gelation, if an intermolecular reaction incorporates tri- or higher functional molecules. These molecules join together at random, to form network aggregates. Flory derived the critical conditions for the formation of infinite networks, finite clusters of molecules and for the coexistence of finite clusters with infinite networks.

* Spear, *l.c.*, mentions that the relation of the rubber molecules to one another in bulk rubber "has not received all the attention it deserves."

Flory also treats the more difficult problem of the composition of three-dimensional polymers. He obtained solutions for two cases: (a) bi- and tri-functional units joined at random and (b) chains of uniform length cross-linked at random at various points.

The weight fraction of molecules composed of a finite number of chains is given by the Poisson distribution:

$$W_n = \frac{n^{n-1}}{(\gamma n)!}(\gamma e^{-\gamma})^n \tag{1}$$

where γ, the "cross-linking index" is defined as the number of cross-linked units in the entire polymeric mixture divided by the total number of chains. The degree of cross-linking, *i.e.*, average number of cross-links per chain, is given by $\gamma/2$, two cross-linked units being involved in each cross-linkage.

When $\gamma \leqq 1$, the summation ΣW_n over the weight fractions of all finite species is less than unity and equal to $W_s = \gamma'/\gamma$ where $\gamma' e^{-\gamma'} = \gamma e^{-\gamma}$ defines γ', W_s represents the weight fraction of the sol (finite clusters of chains), and γ' measures the cross-linking index of this sol fraction.

$$W_g = 1 - \gamma'/\gamma \tag{2}$$

represents then the weight fraction of the gel (infinite network). The cross-linking index γ'' for the gel fraction is given by $\gamma'' = \gamma + \gamma'$. $\gamma = 1$ represents the critical point for incipient formation of infinite gelation networks. There $\gamma' = \gamma = 1$ and $\gamma'' = 2$. [For limitations of (2) cf. Stockmayer [7] l. c. pp. 49-50]

The treatment of non-uniform chains is more complicated and has not yet been carried out. It can be seen, however, that non-uniformity of the chains will reduce the critical value of γ. [cf. however Stockmayer [7]]

Flory points out that gelation-network polymers should be characterized not by an average molecular weight (since a portion of the material possesses essentially infinite molecular weight). Rather he recommends two fundamental quantities: (a) the number of branch units or of cross-linkages per chain, (b) the number average chain length or chain molecular weight. Determination of the weight fraction of the gel will give (a), using formula (1). Therefore formula (2) then yields γ', the degree of cross-linking in the sol fraction. γ' combined with the so-called number average molecular weight $\overline{M}_n$* (= weight of the sample divided by the total number of molecules) determines the average chain length.

Flory discusses applications of his theory of gelation to rubber. Kemp and Peters have shown that the sol fraction of rubber films carefully prepared from latex is ten per cent or less, *i.e.*, $W_s = 0.1$ and $W_g = 0.9$. The presence of a gel fraction in raw rubber is probably due to occasional cross-linkages between the polyisoprene chains. According to Flory's theory, the division of rubber into two separate fractions of (practically) the same chemical composition is a natural consequence of the random cross-linkage of the long rubber chains.

Assuming, for orientation, the polyisoprene chains to be of uniform length, $W_g = 0.9$ leads to $\gamma = 2.5$ for the latex rubber films of Kemp and Peters; thus there are 1.25 cross-linkages per chain.

From $\overline{M}_n = 300{,}000$ (obtained from osmotic pressure measurements) a number average chain molecular weight of about 200,000 follows, assuming about one cross-linkage per two or three chains. Possible degradation and the non-uniformity of the chains would tend to increase this value, perhaps up to 1,000,000, *i.e.*, to a degree

* $\overline{M}_n$ may be obtained by end-group determinations or measurements of the colligative properties of polymer solutions.

of polymerization of about 10,000 isoprene units per chain. This estimate leads to the very low value of about one cross-linkage per 8000 structural units.

Free Rotation in the Rubber Chains. In an isoprene group three out of the four C-C bonds are single and only one is double. Both quantum mechanics and experiment show that there can be free rotation around a single C-C bond. On the other hand, theory and experiment agree that a double bond is rigid and defines a certain plane which is characteristic of the group containing a double bond.

Free rotation, as used here, means only that there are different configurations (stable arrangements) of equal relative probability. The configurations might be separated by energy barriers, so that a free rotation in the literal sense does not take place. Nevertheless, if the barriers are not too high, a relative equilibrium of the different relative configurations may occur in a short time.

Measurement of the heat capacity of ethane indicates that the energy barrier hindering the free rotation of the two CH_3 groups relative to each other is of the order of 3kcal/mole. However, for isoprene the presence of the double bond may very well lower the barrier by a factor of 3 or more. In fact, for dimethylacetylene ($CH_3 - C \equiv C - CH_3$) the presence of the triple bond lowers the barrier so much that the two CH_3 groups rotate almost freely relative to each other. These points seem to have been overlooked by Bresler and Frenkel, in their attempt to calculate the influence of the potential barrier on the contracting force in single long hydrocarbon chains.*

Structure of Bulk Rubber. In the section on gelation it was shown that raw rubber consists of long chains interlinked by a few cross-bonds to a three dimensional gelation network. In this infinite network as a matrix there may be embedded finite clusters of chains, only loosely connected with the network, corresponding to a sol component in the gel matrix.

The Bundle Model and Its Statistical Treatment

In the first semi-quantitative treatment of rubber elasticity by Guth [9] and coworkers in 1934 a simplified model was used for bulk rubber: a bundle of parallel chains or strings running through the whole piece of rubber, let us say a unit cube with its axes parallel to three coordinate axes x, y, z. If M designates the number of chains per unit cross-section, then a unilateral stress, Z, acting on two sides of the cube, say in the Z-direction, will be carried equally by the M strings. Therefore, it is sufficient to consider the behavior of one chain under the action of a stress Z/M and derive an equation of state for this chain.

This problem is closely related to the derivation of an equation of state for a gas and may be solved in an entirely similar fashion.

A string of rubber molecules may be considered as analogous to a gas whose molecules are tied together by flexible strings. The string must be flexible because of the free rotation. If there were no strings the heat motion would cause an expansion of the gas. The presence of the string will cause a coiling up of the resulting chain in a random fashion. To mention a simple illustration: an actual piece of flexible string thrown into the air will come down in a coiled, rather than in a straight form because of the greater probability of coiled configurations in contrast to the less probable straight configuration.

The quantitative treatment of our rubber chain is again very similar to the case of a gas. For simplicity we may consider an ideal gas. A configuration of gas molecules is the more probable the greater the volume the gas molecules occupy. Therefore, it is plausible to put this configurational (or thermodynamic) probability, C, proportional to the volume, V:

$$C(V) = A.V; A = \text{constant} \tag{3}$$

* Recently a detailed criticism of the treatment of Bresler and Frenkel was given by Flory and Rehner [8].

In order to derive an equation of state we have to connect the configurational probability $C(V)$ with the entropy $S(V,T)$ of the gas. Obviously, nature will prefer more probable configurations to less probable ones. This tendency of nature means that the entropy will change in conformity with the configurational probability: a statement qualitatively equivalent to the second law of thermodynamics. Quantitatively, the connection between $S(V,T)$ and $C(V)$ is given per molecule by Boltzmann's principle:

$$s(V,T) = k \log C(V) = k \log (A.V), \tag{4}$$

and per mole with N molecules introducing the gas constant $R = Nk$

$$S = Ns = R \log (A.V) \tag{5}$$

The equation of state $p = p(V,T)$ follows from the entropy by purely thermodynamical considerations:

$$p = T\left(\frac{\partial s}{\partial V}\right)_T = \frac{RT}{V} \tag{6}$$

For a rubber chain containing N freely rotatable links, the configurational probability $C(L)dL$ will be a measure of the probability that the ends of the chain will be apart by a distance between L and $L + dL$. It is natural to expect that $C(L)$ will have the form of the error law, *i.e.*, of the Gaussian distribution:

Here:
$$\left.\begin{aligned} C(L)dL &= Be^{-\beta L^2}dL;\ B = \text{constant} \\ \beta &= \frac{3}{2}\frac{1}{Nl_e^2};\ l_e = l\tan\frac{\delta}{2} \end{aligned}\right\} \tag{7}$$

with l equal to 1.54Å, the distance of the C — C bond and δ equal to 109°, the valence angle. Rigorous analysis, which we omit here, verifies this expectation.

Equation (7) corresponds to equation (3). To the volume, V, of a gas there corresponds the length L between the ends of the rubber chain.

Again Boltzmann's principle will connect the configurational probability $C(L)$ with the entropy $s(L,T)$ of the rubber chain:

$$s(L,T) = k \log C(L) = k\log B - k\beta L^2 \tag{8}$$

and for a *bundle* of M parallel chains

$$s(L,T) = M \cdot s(L,T) \tag{9}$$

This equation of state $Z = Z(L,T)$ where Z designates the force (normalized to unit original cross-section) stretching the rubber follows from the entropy by the analogue of equation (6):

$$Z = -T\left(\frac{\partial S}{\partial L}\right)_T = M \cdot 2\beta \cdot kT \cdot L = KTL \tag{10}$$

Applied to bulk rubber L may be interpreted as relative length, *i.e.*, extended length/original length.

This equation does explain *qualitatively* the three thermoelastic effects exhibited by rubber and holds *quantitatively* fairly well for medium extensions.

Thermoelastic Effects. 1. For the rise in temperature ΔT on fast stretching thermodynamics yields the formula:

$$\Delta T = \frac{T}{C_L}\int_1^L \left(\frac{\partial Z}{\partial T}\right)_L dL \tag{11}$$

Inserting the value of $(\partial Z/\partial T)_L$ using (10) one obtains:

$$\Delta T = \frac{T}{C_L}\frac{K}{2}(L^2 - 1) \tag{12}$$

This formula explains the experimental results given in Fig. 1, above the thermoelastic inversion point but it fails both for lower and higher extension.

2. The isotonic stretch may be characterized by the linear thermal expansion coefficient parallel to the direction of the stress,

$$\alpha_{\parallel} = \frac{1}{L}\left(\frac{\partial L}{\partial T}\right)_z \tag{13}$$

$\alpha_{\parallel}$ may be computed from (10) using the thermodynamic relation:

$$\left(\frac{\partial L}{\partial T}\right)_z = -\left(\frac{\partial L}{\partial Z}\right)_T \left(\frac{\partial Z}{\partial T}\right)_L ; \tag{14}$$

one obtains:

$$\alpha_{\parallel} = -\frac{1}{T} \tag{15}$$

i.e., the linear thermal expansion coefficient of stretched rubber has exactly the same magnitude (in the approximation of equation (10)) as the cubic expansion coefficient of an ideal gas but with a negative sign. Thus, equation (15) explains why rubber stretched isotonically contracts visibly when heated; $\alpha_{\parallel}$ (in the approximation of equation (10) is quite unrelated to the linear expansion coefficient of *unstretched* rubber, which is of the order: $2 \cdot 10^{-4}$. Again equation (15) only holds well above the thermoelastic inversion point.

3. For isometric stretch, equation (10) shows the (proportional) increase of stress with temperature.

Larger Extensions. The upward curvature of the stress-strain curve as shown in Fig. 4, may be explained on the basis of the bundle model. For it can be proved that the Gaussian distribution (7) holds only if the "length" L between the ends of the chain is small compared to the maximum length, which the chain has if it is completely straightened out by stress. For a simple one-dimensional model described in the Appendix of this paper, the maximum length is

$$L_{max} = l \cdot N \tag{16}$$

where N is the number of links. The condition for the validity of the Gaussian distribution reads:

$$L \ll L_{max} \text{ or } t \ll 1; t = L/L_{max} = L/Nl \tag{17}$$

The generalized distribution derived in the Appendix is

$$C(L,N) = \frac{2}{(2\pi N)^{1/2}} \cdot \exp\left(-\frac{N}{2}\int \log\frac{1+t}{1-t}dt\right) \tag{18}$$

since

$$log\frac{1+t}{1-t} = 2(t + \frac{t^3}{3} + ...) \tag{19}$$

(18) reduces to the Gaussian distribution if only the first term in the parenthesis is retained. Inserting $C(L, N)$ from (18) into equations (8), (9) and (10) we obtain:

$$Z = M \cdot \frac{kT}{l}\frac{1}{2}\log\frac{1+t}{1-t} \tag{20}$$

Introducing the quantity

$$f = \frac{Zl}{kT}\frac{1}{M} \tag{21}$$

we can transform (20) into the form

$$Z = M \cdot \frac{kT}{} \cdot f(t) \cdot \text{with} f(t) = \tanh^{-1} t = \tfrac{1}{2} \log \frac{1+t}{1-t} \tag{22}$$

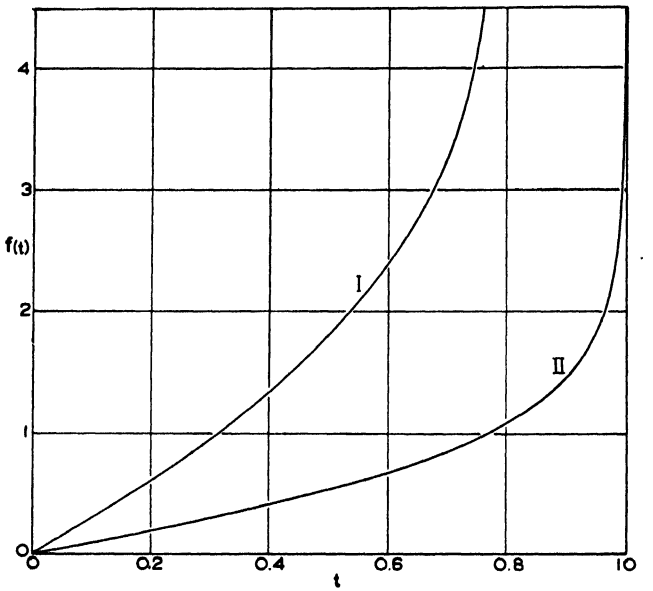


FIGURE 5. Inverse Langevin function (I) and inverse tanh function (II).

The function $t = \tanh f$ occurs in Ehrenfest's theory of magnetism. The function $f(t)$ is shown as curve II in Fig. 5. This curve explains the upward curvature of the rubber stress-strain curves simply as due to deviations from the Gaussian law as the chains are straightened out more and more by the stress.

A three-dimensional model corresponding to the one-dimensional model just treated, yields

$$f(t) = \mathcal{L}^{-1}(t); \quad = \mathcal{L}(f) = \text{cotn} f - \frac{1}{f} \tag{23}$$

where $\mathcal{L}$ (f) is the Langevin function known from Langevin's theory of paramagnetism and Debye's theory of polar molecules. The physical reason for the appearance of the Langevin function and of the function $t = \tanh f$ in the theory of rubber elasticity is elaborated in the Appendix of this paper.

The inverse Langevin function is shown as curve I in Fig. 5. The curve 3 $\tanh^{-1} t$ deviates from $\mathcal{L}^{-1}$ (t) only when t approaches 1.

Thus the bundle model does explain the anomalous thermoelastic properties of rubber. It also explains the shape of the stress-strain curve at medium and larger extensions.

Limitations of the Bundle Model. The bundle model has, however, the following shortcomings. (a) It is an essentially one-dimensional model, stretchable only in the direction of the axis of the bundle. $\alpha_\perp(L)$, the linear thermal-expansion coefficient perpendicular to the direction of the stress cannot be computed. (b) It does not hold for small extensions and for compression. (Hooke's law holds in the limiting case of very small extensions). It gives $Z > 0$ for $L = 1$, and $L = 0$ for $Z = 0$

i.e., a volume zero for the unstretched state. (c) It does not explain the thermoelastic inversion point.

The New Model: Irregular Network and Internal Pressure

We know from the theory of gelation that even in raw rubber we do not have a bundle of isolated long-chain molecules. Rather, the long-chain molecules form a coherent mass by some sort of cross-linking. There must exist strong forces between the molecules to prevent their slipping past one another. Otherwise raw rubber could not be elastic, but would show as much plasticity as beeswax and chewing gum, which substances, when stretched, do not retract to any extent.

In the vulcanization process the cross-links already present in raw rubber are reinforced and new ones formed. Cross-links may be primary and secondary bonds. This general picture of the network structure of rubber is sufficient for the theory developed here.

For soft rubber, even after vulcanization, there are only a few cross-links per chain. Although a translation of the chains with respect to one another is suppressed, there is enough freedom of motion of the chains to make possible a free rotation around single bonds in the sense explained before.

We take as our model a region crossed by a complicated network of chains, very irregular in detail, but homogenous on the average. The chains may run between perpendicular plane boundaries of the volume of the rubber over which their ends will be uniformly distributed. The flexible chain molecules are bonded into the network by steric forces.

One action of the network will be a pull, the development of a contracting force due to the tendency of the chains of the network to coil up because of free rotation. This contracting force would lead to a collapse of the network if there were not present another action of the network exactly balancing the contracting force.

The second action of the network is due to the effects of molecular bombardments on the bounding surfaces. The chains jerk at the boundaries of the bulk material. In addition there may be inert clusters of chains (a sol component) present but not connected with the network. Both molecular bombardments on the bounding surfaces and sol component will lead to the production of a pressure as in an ordinary liquid.

In our model we separate rather sharply the two actions of the network: the contracting pull of the network due to the tendency of the chains to coil up, and the hydrostatic pressure, due to the sidewise pushes of the chains. We compute the first action, the pull of such a network directly, but represent its second action, the production of internal pressure, simply by filling the model with a fictitious * liquid, the "rubber liquid," which in a good approximation one may take as incompressible. Any surface of the model must, of course, be in equilibrium under all the forces which act on it—the pull of the network, the push of the hydrostatic pressure, and any external forces.

Equation of State for Incompressible Rubber Consisting of Gaussian Chains

Let us consider the irregular cross-bonded network as described above. It can be proved [2] that this network can be replaced by a regular network of cubic symmetry. The independent chains of this regular network may run parallel to the three coordinate axes. For this network it is easy to obtain the equation of state for stretched rubber.

We make two fundamental assumptions:

I. The Gaussian distribution is valid for $C(L)$, the configurational probability; *i.e.*, the rubber chains are extended only to a length small compared to their maximum length;

* This "liquid" is fictitious as far as it simulates effects of steric forces and molecular bombardments but is more real as far as it represents a largely non-extractable sol component i.e. branching.

II. Rubber is incompressible, *i.e.*, it does not change its volume when stretched.

Assumption II holds experimentally for soft gum compounds up to about 300 per cent extension, according to Holt and McPherson.[10]

Let us consider a unit cube of unstretched rubber. After a force Z is applied in the z-direction, the dimensions of the resulting parallelepiped will be L_x, L_y, L_z. For unilateral deformations we have the relation

$$L_x = L_y \tag{24}$$

According to our assumption II and (24)

$$1 = V = L_x L_y L_z = L_y^2 L_z \tag{25}$$

As described earlier, there will be three forces acting in our model:

Z_1, the inward pull due to the contracting force;

Z_2, the outward push due to the internal pressure of the rubber liquid; and

Z_3, the outward pull due to the external force.

Because of the equivalence of the irregular and of the regular cubic network, we may use for the contracting force the expression (10) derived for an isolated chain. Then

$$Z_1 = KT \cdot L_z \tag{26}$$

The internal pressure P may be calculated from the condition of the equilibrium of the forces applied to one of the sides (L_y L_z) of the cube on which no external force acts. The outward push of this pressure is balanced by the pull of M molecular chains, each extended to length L_z and thus exerting an average force of $Z_z = KT \cdot L_z$. Then,

$$PL_y L_z = KT \cdot L_x; \tag{27}$$

using (24) we obtain

$$P = KT \cdot 1/L_z \tag{28}$$

Thus the internal pressure of rubber decreases for tension and increases for compression in accordance with physical expectations.

Considering the forces acting on the end (L_z L_y) faces, we have

$$\begin{aligned} Z_1 &= KT \cdot L_z \\ Z_2 &= PL_x L_y = KT \cdot 1/L_z^2 \end{aligned} \tag{29}$$

For equilibrium between these three forces, the sum of the outward forces must be equal to the inward pull, or

$$Z_2 + Z_3 = Z_1 \tag{30}$$

Inserting the value of Z_1 and Z_2, we obtain

$$Z = Z_3 = Z_1 - Z_2 = KT[L_z - 1/L_z^2]. \tag{31}$$

This is the stress-strain relation we were looking for. Physically, it states that the external force balances the difference between the tension due to the internal Brownian motion and that due to the internal pressure.

The stress-strain relation (31) holds both for extension ($L_z < 1$) and for compression ($L_z > 1$), thus removing an outstanding difficulty of the bundle model. It holds also, of course, for small extensions. At large L (31) has an asymptote intersecting the Z-axis at $L = 0$; at small L (31) has the Z-axis as its asymptote.

Fig. 6 curve B is an observed stress-strain curve obtained at Notre Dame. The circles are theoretical values according to equation (31). The knee of the observed curve is explained well by the theory.

Young's modulus, E, may be defined as the slope of the stress-strain curve given by equation (31):

$$E(L_s) = \frac{dZ}{dL} = KT\cdot\left[1 + \frac{2}{L_s^3}\right] \tag{32}$$

E has for $L_s = 1$ the value

$$E(1) = 3KT, \tag{33}$$

which is threefold the value

$$E(L_s \gg 1) = KT. \tag{34}$$

To see the range of validity of Hooke's law we introduce instead of the relative length L the extension or strain $\epsilon = L - 1$. (31) becomes thus

$$Z = KT\left[(1+\epsilon) - \frac{1}{(1+\epsilon)^2}\right] = 3KT\cdot\epsilon\left[1 - \epsilon + \frac{4}{3}\epsilon^2 - \ldots\right] \tag{35}$$

From (35) it follows that for extensions as small as 5 per cent there is a 5 per cent deviation from Hooke's law. Young's modulus is, of course, the same for both variables L and ϵ as reference.

Nominal Versus True Stress-Strain Curve

Equation (31) represents a *nominal* stress-strain curve, *i.e.*, the stress is the force referred to unit original cross-section. The *true* stress-strain curve for which the stress is referred to the actual cross-section may be simply obtained from (31) using (25). Designating the true stress

$$\zeta = Z\cdot L \tag{36}$$

we have

$$\zeta = KT\left[L^2 - \frac{1}{L}\right] \tag{37}$$

Young's modulus is now defined:

$$E_\zeta(L) = \frac{d\zeta}{dL} = KT\left[2L + \frac{1}{L^2}\right], \tag{38}$$

and has the values

$$\begin{aligned} E_\zeta(1) &= 3KT \\ E_\zeta(L \gg 1) &= 2KT\cdot L \end{aligned} \tag{39}$$

for $L = 1$ and large L respectively.

In terms of the strain ϵ the equation (37) reads:

$$\zeta = KT\left[(1+\epsilon)^2 - \frac{1}{1+\epsilon}\right] = 3KT\epsilon\left[1 + \frac{\epsilon^2}{3} - \frac{\epsilon^3}{3} + \ldots\right] \tag{40}$$

Thus, for the true stress-strain curve Hooke's law holds over a much wider range of strains than for the nominal stress-strain curve.

For metals in particular a true strain

$$\lambda = \int_1^L \frac{dL}{L} = \log L = \log(1+\epsilon);\ L = e^\lambda;\ \epsilon = e^\lambda - 1 \tag{41}$$

is introduced usually instead of the nominal strain ε. In terms of ζ and λ as dependent and independent variables (40) takes on the form:

$$\zeta = KT[e^{2\lambda} - e^{-\lambda}] \quad . \tag{42}$$

Young's modulus is now in terms of λ:

$$E_\zeta(\lambda) = \frac{d\zeta}{d\lambda} = KT\left[2e^{2\lambda} + e^{-\lambda}\right] \tag{43}$$

Therefrom

$$E_\zeta(\lambda \gg 0) = 2KT \cdot e^{2\lambda} \tag{44}$$

Young's modulus for small extensions has the same value for the three pairs of variables as it should; for large values of the strain, however, the moduli differ from one another.

Instead of λ one uses sometimes the logarithm of the reduction in area to measure the strain, because for metals in the stage of necking the strain is not directly observable. For rubber, because of its incompressibility, the logarithm of the reduction in area is simply the reciprocal of the true strain.

We introduced here true stress and true strain because it is instructive to compare analytically the expressions for stress and strain in the three different pairs of variables. It is obvious that in the range of stress and strain in which the volume of rubber does not change with extension, there is no advantage in using a true stress-strain curve instead of a nominal one.* The situation is different when rubber crystallizes, particularly when necking takes place before breaking. Then the use of the true stress-strain, or rather true stress vs. reduction in area curves probably will offer the same advantages over the nominal stress-strain curves as in the case of metals.

The Thermoelastic Effects for the New Model

For the rise in temperature on fast stretching insertion of (31) in (11) yields:

$$\Delta T = \frac{KT}{2C_L}\left[L_x^2 + L_x - 2\right]\left[\frac{L_x - 1}{L_x}\right] \tag{45}$$

The linear thermal expansion coefficient parallel to the direction of the stress $\alpha_\parallel(L_x)$ becomes:

$$\alpha_\parallel(L_x) = \frac{1}{L_x}\left(\frac{\partial L_x}{\partial T}\right)_Z = -\frac{1}{T}\frac{L_x^3 - 1}{L_x^3 + 2} \tag{46}$$

This expression does not differ significantly from (15) to which it reduces for $L_x \gg 1$. A novel feature, however, is the possibility of computing the linear thermal expansion coefficient perpendicular to the direction of the stress:

$$\alpha_\perp(L_x) = \frac{1}{L_y}\left(\frac{\partial L_y}{\partial T}\right)_Z = \frac{1}{2T}\frac{L_x^3 - 1}{L_x^3 + 2} \tag{47}$$

In deriving (47) the constancy of the volume of rubber during stretching, *i.e.*, the relation (25), was used. The possibility of introducing $\alpha_\perp(L_x)$ rests on the three-dimensional character of the new network model in contrast to the essentially one-dimensional bundle model.

The formulas (46) and (47) still do not contain the cubic expansion coefficient of unstretched rubber and consequently do not show a thermoelastic inversion point.

* In addition the area under the nominal stress-strain curve has a simple thermodynamical meaning: it represents the work done by the stretching force.

Generalization of the Stress-Strain Relation to Larger Extensions

The generalization of our model to larger extensions involves three parts:

(A) The contracting force for a single chain has to be generalized to hold beyond the range of validity of the Gaussian distribution.

(B) An irregular network has to be considered and reduced to a form in which the expression for the contracting force for a single chain may be utilized for the contracting force of the whole network.

(C) The contracting force of the network has to be balanced by the internal pressure of the rubber liquid and any external forces.

The result for part (A), the contracting force for a single chain, was recorded earlier. Part (B) could not be treated in the same rigorous fashion as was possible for the Gaussian chains. Instead of carrying out the reduction from an irregular to a regular network, we simply assume that in the unstretched state of the material each independent chain is taken to have an extension equal to κ times its maximum length. The maximum stretch is thus given by a factor $1/\kappa$, a quantity which will depend upon the constitution of the particular rubber considered. κ is an adjustable parameter in our treatment.

Considering again the equilibrium of forces and using for the contracting force the three-dimensional result (23)

$$Z = K'T\left[\mathcal{L}^{-1}(L_s\kappa) - \frac{3\kappa}{L_s^2}\right] \tag{48}$$

This relation involves essentially two parameters, the scale factor K' and the new maximum stretch parameter κ. When the extension is small compared to the maximum

$$\mathcal{L}^{-1}(t) \cong 3t \tag{49}$$

$$Z \cong 3\kappa K'\left[L_s - \frac{1}{L_s^2}\right] \tag{50}$$

i.e., (48) reduces to the Gaussian expression (31), $3\kappa K'$ taking the place of the old parameter K.

Equation (48) yields an S-shaped stress-strain curve. The occurrence of the inflexion point, *i.e.*, of the S-shape, is explained as due to the superposition of the effects of the contracting force and those of the internal pressure.

Referred to the actual cross-section the nominal stress-strain relation (48) assumes the form:

$$Z = K'T\left[L_s\mathcal{L}^{-1}(L_s\kappa) - \frac{3\kappa}{L_s}\right] \tag{51}$$

This true stress-strain curve does *not* show the S-shape of the nominal stress-strain curve. Fig. 6 curve A shows the true stress versus nominal relative length curve corresponding to the nominal curve B. The S-shape of curve B is well explained by equation (48), and the shape of curve A by equation (51).

Influence of the Thermal Expansion on the Stress-Strain Relation

With the successive refinements introduced we were able to explain the peculiar thermoelastic properties of rubber, except the thermoelastic inversion point. To explain this phenomenon we have to take into account the cubic thermal expansion of unstretched rubber, characterized by a coefficient of thermal expansion α defined as

$$V(T) = V(T_0)[1 + \alpha(T - T_0)] \tag{52}$$

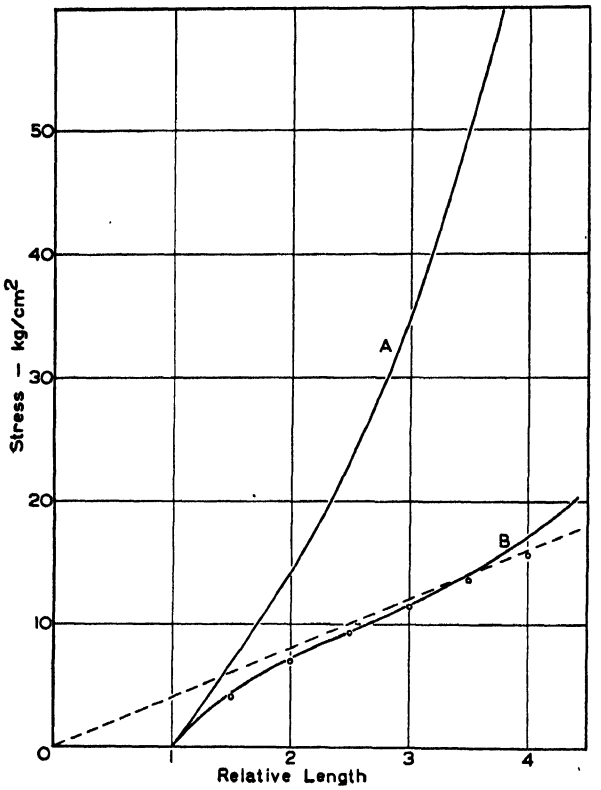


FIGURE 6. Curve A. True stress vs. nominal relative length for an accelerated pure gum stock. Curve B. Nominal stress vs. nominal relative length corresponding to Curve A. The circles are theoretical values according to Equation (31). The dashed line represents the asymptote of the theoretical curve.

The previous stress-strain relation (31) yields zero stress for $L = 1$, the unstretched state. When thermal expansion is considered, the unstretched state is characterized by $L = 1 + \frac{\alpha}{3}(T - T_0)$

The previous equation (31) should be changed to

$$Z = KT\left[L - \frac{1 + \alpha(T - T_0)}{L^2}\right] \tag{53}$$

which gives $Z = 0$ at $L = 1 + \frac{\alpha}{3}(T - T_0)$. The thermal expansion enters only into the internal pressure term, and the contracting force (the first term in the bracket) is independent of α. We emphasize that L by definition is independent of α, the original length being taken at $T = T_0$.

Equation (53) predicts a thermoelastic inversion point as the solution of the equation

$$\left(\frac{\partial Z}{\partial T}\right)_L = K\left[L - \frac{1 + \alpha(2T - T_0)}{L^2}\right] = 0. \tag{54}$$

Thus, the critical extension at which the stress-temperature curves, the isometrics, change their slope, is given by the isometric thermoelastic inversion point

$$L^3_{iso} = 1 + \alpha(2T - T_0) = 1 + \alpha T_0; \text{for } T = T_0, \tag{55}$$

Because of the thermodynamic relation

$$\left(\frac{\partial L}{\partial T}\right)_Z = -\left(\frac{\partial L}{\partial Z}\right)_T \cdot \left(\frac{\partial Z}{\partial T}\right)_L \tag{56}$$

the isotonic thermoelastic inversion point is identical with the isometric one.

From (52) and (56) the linear thermal expansion coefficients in the direction of the stress and perpendicular to it: $\alpha_{\parallel}$ and $\alpha_{\perp}$ may be computed. One obtains for $T = T_0$

$$\alpha_{\parallel}(L_x) = \frac{1}{L_x}\left(\frac{\partial L_x}{\partial T}\right)_Z = -\frac{1}{T}\frac{L_x^3 - 1}{L_x^3 + 2} + \frac{\alpha}{L_x^3 + 2}, \tag{57}$$

and

$$\alpha_{\perp}(L_x) = \frac{1}{L_y}\left(\frac{\partial L_y}{\partial T}\right)_Z = +\frac{1}{2T}\frac{L_x^3 - 1}{L_x^3 + 2} + \frac{\alpha}{2}\frac{L_x^3 + 1}{L_x^3 + 2}. \tag{58}$$

In deriving (58) the constancy of the volume of rubber during stretching was used:

$$V = L_y^2 L_x = 1 + \alpha(T - T_0) \tag{59}$$

In the unstretched state $L_x = 1$ we have

$$\alpha_{\parallel}(1) = \alpha_{\perp}(1) = \alpha/3, \tag{60}$$

as for any isotropic material. $\alpha_{\parallel}$ changes from positive to negative values with increasing L_x, whereas $\alpha_{\perp}$ always stays positive.

The formula corresponding to equations (12) and (45) becomes for $\alpha \neq 0$:

$$\Delta T = \frac{KT}{2C_L}\left[L_x^2 + L_x - 2(1 + \alpha T_0)\right]\left[\frac{L_x - 1}{L_x}\right] \tag{61}$$

According to this formula ΔT starts with the value zero at $L_x = 1$, reaches a minimum for $L_x = L_{1x0}$ as given by (55), then becomes zero for the adiabatic thermoelastic inversion point, determined by

$$L_{ad}^2 + L_{ad} - 2(1 + \alpha T_0) = 0 \tag{62}$$

and takes on positive values for $L_x > L_{ad}$. (57) and (61) explain Joule's data in Figs. 2 and 1.

The introduction of the cubic thermal expansion of rubber into the stress-strain relation (51) is simple, because only the internal pressure is modified. One obtains:

$$Z = K'T\left[\mathcal{L}^{-1}(L_{x\kappa}) - \frac{3\kappa[1 + \alpha(T - T_0)]}{L_x^2}\right] \tag{63}$$

Comparison of the Theoretical Stress-Strain Relations with Experiment

In comparing the theory with experiment, a resolution of the stress into its additive parts due to entropy and internal energy has to be carried out both for the experimental results and for the theory.

In van der Waals' equation of a real gas

$$p = -\frac{a}{V^2} + \frac{R}{V - b}T \tag{64}$$

the first term $p_{(U)} = -a/V^2$ is due to the internal energy U and the second term $p_{(S)} = RT\ (V - b)$ is due to entropy S. According to thermodynamics:

$$p_{(U)} = -\left(\frac{\partial U}{\partial V}\right)_T, \quad p_{(S)} = T\left(\frac{\partial S}{\partial V}\right)_T = T\left(\frac{\partial p}{\partial T}\right)_V \tag{65}$$

so that

$$p = p_{(U)} + p_{(S)} = -\left(\frac{\partial U}{\partial V}\right)_T + T\left(\frac{\partial p}{\partial T}\right)_V \tag{66}$$

For a unilateral stress Z instead of a uniform pressure p we have to replace p by $-Z$ (since pressure and stress act in opposite directions) and the volume V by the relative length L. Then we obtain

$$Z = Z_{(U)} + Z_{(S)} = \left(\frac{\partial U}{\partial L}\right)_T + T\left(\frac{\partial Z}{\partial T}\right)_L \tag{67}$$

Using equation (67) we may deduce experimental $Z_{(U)}$ and $Z_{(S)}$ curves plotted versus L from any experimental stress-strain curve. The theoretical $Z_{(U)}$ and $Z_{(S)}$ vs. L curves follow from (53) and (63) by the help of (67):

$$Z_{(U)} = \alpha KT^2 L_a^{-2};\ Z_{(S)} = KT\left\{L_a - [1 + \alpha(2T - T_0)L_a^{-2}\right\} \tag{68}$$

$$Z_{(U)} = 3K'\kappa\alpha T^2 L_a^{-2};\ Z_{(S)} = K'T\left\{\mathcal{L}^{-1}(L_a\kappa) - 3\kappa[1 + \alpha(2T - T_0)]L_a^{-2}\right\} \tag{69}$$

(69) we compare with experimental results obtained by Anthony, Caston and Guth. Fig. 7 is taken from a paper by James and Guth which appeared in the *Journal of Chemical Physics* in Oct. 1943. The over-all agreement between theory and experiment is rather good, except at the lowest extensions where van der Waal's forces, neglected in the theory, may play a role. For further comparisons between theory and experiment we refer to Section 7 of the paper by James and Guth and to the papers by the author and his associates.[2]

Concerning the molecular interpretation of the two constants κ and K' of the theory we must refer to the paper by James and Guth. There the new concept of the fraction of the chains actively involved in the network is introduced in contrast to the old concept of the molecular weight.

Limitations of the Theory

The theory of James and Guth does not consider crystallization phenomena which are not essential for the occurrence of rubber-like elasticity. In fact it is known that Buna S (GR-S) type synthetic rubber, though it does not crystallize, still exhibits the characteristic S-shape in its stress-strain curve. The unaccelerated pure gum stock of Fig. 7 also does not show crystallization up to about 400 per cent extension, while revealing an S-shape.

It is not possible to develop a theory of the crystallization of rubber-like materials as general as our theory of rubber elasticity. The magnitude and geometrical distribution of the intermolecular forces regulating crystallization will depend greatly upon the chemical composition of the elastomer.

One may consider how the average molecular weight or chain length and the chain length distribution curve of milled rubber in solutions will influence the elasticity of the cured product. In our theory this effect will be evident in the constants κ and K'. The amount of branching or cross-linking may similarly be treated.

It should be pointed out that the theory of James and Guth provides a basis for the explanation of many other physical properties of stretched rubber.

General Occurrence of Rubber-like Materials

According to the principles advanced in our theory, any long-chain compound with free rotation will exhibit rubber-like elasticity above its melting point. Conversely, any compound exhibiting rubber-like elasticity will be a long chain compound possessing free rotation. Long-chain compounds have a melting point or melting range above which there is free rotation. Whether this melting point is be-

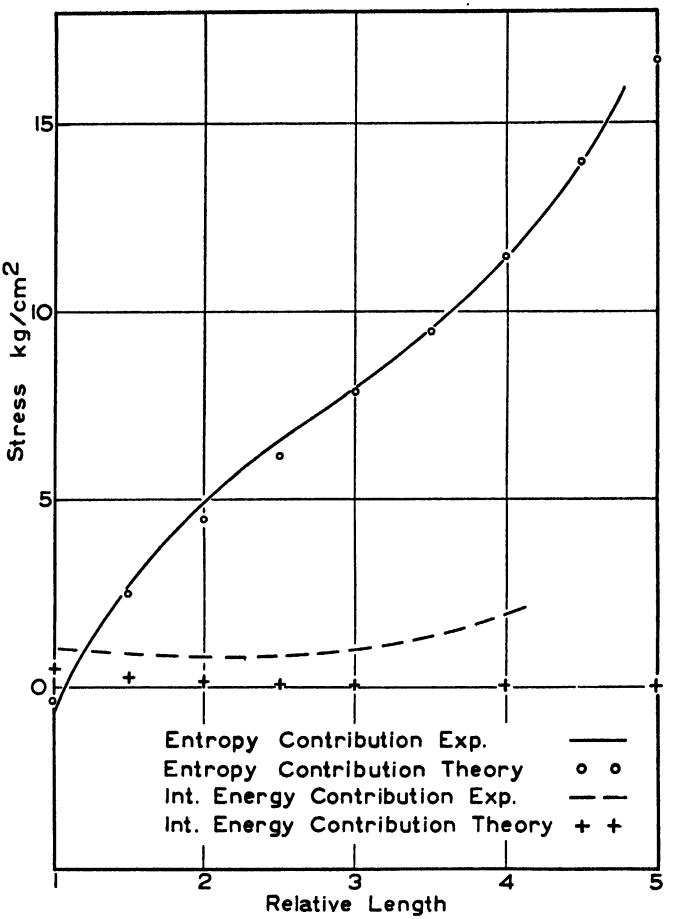


FIGURE 7. Internal energy and entropy contributions to the stress for an unaccelerated pure gum compound. The circles are theoretical values according to the equation (63).

low or above room temperature depends on the competition between free rotation and molecular forces. The free rotation may be started either by raising the temperature or by introducing a solvent either by plasticizing or swelling. A survey of rubber-like materials was given by the author in another paper.[12] Here we restrict ourselves to a few examples:

(a) gutta-percha, the trans-isomer of rubber, shows rubber-elasticity above 70° C;

(b) the inorganic substances sulfur, selenium, phosphorus chloronitride, and sulfur trioxide possess elastic modifications at higher temperatures and under special circumstances;

(c) synthetic elastomers, like neoprene, butyl rubber, thiokol and the butadiene types; synthetic linear polymers [18];

(d) polystyrene, polyindene and allied substances;

(e) cellulose derivatives;

(f) nylon;

(g) biological materials, like gelatin, glue, wool, silk, and the muscles.

Summary

Fundamental facts on the elasticity of rubber are the thermoelastic effects and the S-shape of the stress-strain curve with its continuation for compression. There are three thermoelastic phenomena for elastomers:

(1) They exhibit a rise in temperature on fast stretching.
(2) If stretched by a constant load (*i.e.*, isotonically) they contract visibly when heated.
(3) The stress in an elastomer kept at constant extension (*i.e.*, isometric stretch) increases with rising temperature.

These three effects are correlated by thermodynamics.

The theory of James and the author explains quantitatively all these fundamental facts on rubber elasticity. It explains observed finer details of thermoelasticity. The change in temperature on fast stretching is negative (cooling) for low extensions, passes through a minimum, becomes zero (adiabatic thermoelastic inversion point), and finally assumes positive values (heating) for increasing extensions. It explains the observed thermal anisotropy of isotonically stretched rubber. The linear thermal expansion coefficient is negative in the direction of the stress but positive in the direction perpendicular to the stress. Both for isotonic and isometric stretch there exists a thermoelastic inversion point which is about half of the value of the critical extension for the adiabatic thermoelastic inversion point. The visible contraction of isotonically stretched rubber when heated is due to the fact that the negative linear thermal expansion coefficient of stretched rubber is of the magnitude of the (cubic) thermal expansion coefficient for a gas. The thermoelastic inversion points are determined in a simple manner by the cubic thermal expansion coefficient of unstretched rubber. The S-shape of the stress-strain (compression) curve is also explained by the theory.

The thermoelastic effects (except the inversion phenomena) may be explained on the basis of the bundle model introduced earlier by the author. In this model bulk rubber is assumed to consist of a bundle of parallel strings built up from flexible (free rotation) long chain molecules. However, only the new network model by James and the author can explain the stress-strain curve at low extensions and its continuation for compression. This model is also needed to explain the thermal anisotropy of stretched rubber. The inversion phenomena, finally, may be explained by taking into account the cubic thermal expansion of unstretched rubber.

In the discussion of the structure of bulk rubber, a brief account is given of the theory of gelation by Flory. The question of nominal versus true stress-strain curves is discussed. In the range of strains for which the volume of stretched rubber remains constant, no advantage is gained by using the true stress-strain curve. However, if the volume changes and particularly in the stage of necking before breaking, the true stress versus reduction in area curve may offer similar advantages as in the corresponding case of metals.

Appendix

Finite Flexible Chains in One and Three Dimensions

We consider a chain of N links, each of which has the probability of 1/2 of contributing $+l$ or $-l$ to the net length L of the chain. If N_+ represents the number of links contributing a length $+1$, to the net length and N_- the number of links contributing a length -1, the probability that we have N_+ links in one direction and N_- links in the other is

$$C(N_+,N_-) = \left(\frac{1}{2}\right)^N \frac{N!}{N_+!N_-!} \qquad (I)$$

It is obvious that $N_+ + N_- = N$ and that the net length of the chain will be $L = l\,(N_+ - N_-)$. Solving for N_+ and N_- in terms of L we obtain as the probability distribution for a given length L

$$C(L,N) = \left(\frac{1}{2}\right)^N \cdot \frac{N!}{\left(\frac{N + L/l}{2}\right)!\left(\frac{N - L/l}{2}\right)!} \tag{II}$$

Using Stirling's approximation, we have

$$N! = N^N e^{-N}\sqrt{2\pi N} \tag{III}$$

and equation (II) becomes, writing $t = L/Nl$,

$$C(L,N) = \frac{2}{(2\pi N)^{\frac{1}{2}}}(1+t)^{-\frac{N}{2}(1+t)}(1-t)^{-\frac{N}{2}(1-t)}(1-t^2)^{-\frac{1}{2}} \tag{IV}$$

This reduces to equation (18) of the text if we neglect the slowly varying factor $(1 - t^2)^{-\frac{1}{2}}$

The occurrence of the function

$$t = \tanh f \tag{V}$$

in equation (22) of the text may be made plausible by observing the analogy of our problem with Ehrenfest's problem of the orientation of a magnetic particle capable only of two (parallel and antiparallel) positions with regard to an external magnetic field. According to the Maxwell-Boltzmann relation the average length $\bar{l}$ of a link l on which a stress Z acts is then, for our one-dimensional chain,

$$\bar{l} = l\frac{e^f - e^{-f}}{e^f + e^{-f}} = l\tanh f; f = \frac{Zl}{kT} \tag{VI}$$

Noting that $\bar{l}/l = t$, equation (VI) reduces to (V).

The occurrence of the Langevin function for the three-dimensional flexible chain may be explained similarly. For this chain the projections of each link may take on continuously all values between $-l$ and $+l$. The Maxwell-Boltzmann relation then yields.

$$\bar{l} = \frac{\int_0^\pi \exp\left[\frac{lZ\cos\theta}{kT}\right] l\cos\theta\sin\theta\,d\theta}{\int_0^\pi \exp\left[\frac{lZ\cos\theta}{kT}\right]\sin\theta\,d\theta} \tag{VII}$$

where θ is the angle between the directions of l and Z. Equation (VII) yields

$$\bar{l} = l\left(\coth f - \frac{1}{f}\right) = l\mathcal{L}(f) \tag{VIII}$$

This three-dimensional problem is completely analogous to the orientation of polar molecules in an electric or magnetic field. The thermal agitation of the chain (characterized by kT) tends to disorient the links and make the average component in the direction of the stress vanish. The tension Z in the chain tends to orient the links.

References

1. Guth and James, *Ind. Eng. Chem.*, 33, 624 (1941); Guth and James, papers read before A.C.S. Rubber and Colloid Divisions, Detroit, 1940; James and Guth, *Phys. Rev.*, 59, 111 (1941); James and Guth, *Ind. Eng. Chem.*, 34, 1365 (1942); James and Guth, *J. Chem. Phys.*, 11, 455 (1943). This last paper gives a detailed account of the new theory with many applications not reported in the present article; the earlier work of other authors is referred to in *Ind. Eng. Chem.* (1941), and in *J. Chem. Phys.* (1943). The simple stress-strain relation (31) is a special case of a relation ob-

tained by Mooney, *J. Appl. Phys.*, **11**, 582 (1940), who correlated stress-strain and shear by the theory of elasticity. Such a consideration does not yield, however, the temperature dependency of the stress, which is essential for the explanation of the thermoelastic phenomena; it intends neither to give a physical picture associated with the two terms L and $-1/L^2$ of the relation (31). Such a picture may be obtained only by using statistical mechanics, as in the theory of James and the author.

2. Anthony, Caston and Guth, *J. Phys. Chem.*, **41**, 826 (1942); Caston, Anthony and Guth, *Phys. Rev.*, **60**, 542 (1942); Peterson, Anthony and Guth, *Ind. Eng. Chem.*, **34**, 1349 (1942); Dart, Anthony and Guth, *Ind. Eng. Chem.*, **34**, 1340 (1942).
3. An excellent account of Joule's and other older work is given by Whitby, "Plantation Rubber and the Testing of Rubber," Chapter XX, pp. 453 ff. Longmans, Green and Co., London, 1920; for more recent experiments, cf. Dart, Anthony and Guth.[2]
4. Anthony, Caston and Guth.[2]
5. cf. Whitby,[3] pp. 225-226.
6. Sheppard and Clapson, *Ind. Eng. Chem.*, **24**, 782 (1932).
7. Flory, *J. Am. Chem. Soc.*, **63**, 3083, 3091, 3096 (1941); *J. Phys. Chem.*, **46**, 132 (1942); Stockmayer, *J. Chem. Phys.*, **11**, 45 (1943). Stockmayer (personal communication) has solved recently the problem of cross-linking of non-uniform chains. The critical value of γ becomes equal to the ratio of the number average to the weight average length of the chains.
8. Flory and Rehner, New York Academy of Science, preprints for Conference on High Polymers (Jan. 1943).
9. Guth and Mark, *Monatsh. f. Chem.*, **65**, 93 (1934); Guth, *Kautschuk*, **13**, 201 (1937).
10. Holt and McPherson, *J. Res. Natl. Bur. Standards*, **17**, 657 (1936).
11. The resolution of experimental stress-strain curves into entropy and internal energy components was first attempted by Wiegand and Snyder, *Trans. Inst. Rubber Ind.*, **10**, 234 (1934). It was carried out in a more complete fashion first by Anthony, Caston and Guth [2] by Caston, Anthony and Guth,[2] and by Peterson, Anthony and Guth.[2] On the thermodynamics of stretched rubber cf. also Gerke, *Ind. Eng. Chem.*, **22**, 73 (1930) and article by W. W. Vogt in "Chemistry and Technology of Rubber," ed. by Davis and Blake, Reinhold Publishing Corp., 1936.
12. Guth, paper read at Surface Chemistry Symposium, 50th Anniv. Univ. of Chicago, 1941; A.A.A.S. pub. No. 21 (1943).
13. Linear synthetic polymers were studied extensively by Fuller and his associates; cf. Fuller, *Chem. Rev.* **26**, 143 (1940); Fuller and Baker, *J. Chem. Ed.*, **20**, 3 (1943).

The Vitreous State

MAURICE L. HUGGINS AND KUAN-HAN SUN
Eastman Kodak Co., Rochester, N. Y.
and
ALEXANDER SILVERMAN
Department of Chemistry, University of Pittsburgh, Pittsburgh, Pa.

General Considerations

If a substance possesses a degree of rigidity comparable to that possessed by typical crystalline substances, yet lacks the dependence on direction (characteristic of the crystalline state) of such properties as hardness, rate of growth and solution, and (sometimes) refractive index, it is said to be in the vitreous * state.†

From the structural point of view, the vitreous state of matter may be defined, alternatively, as that state in which the arrangement of the component atoms, ions, or molecules possesses a *permanence* similar to that in the crystalline state and a *randomness* similar to that characteristic of liquids.

The terms *vitreous substance, vitreous solid, amorphous solid, supercooled liquid, glass,* and *plastic* may be considered to be practically synonymous. The term used by a given author at a given time is largely a matter of preference, depending in part on the type of characteristic property (*e.g.*, rigidity, lack of directional proper-

* Vitreous is a derivative of the Latin "vitreus," itself derived from "vitrum," which means "glass." [Incidentally, because of their glassy appearance the sulfates of zinc, copper and iron were called respectively white, blue and green vitriol. The acid distilled from green vitriol was called oil of vitriol, or commonly "vitriol." See "The Natural History of C. Plinius Secundus," translation by K. C. Bailey, p. 180, London, 1932. J. A.]

† Certain properties (*e.g.*, the refractive index) of crystals belonging to the cubic (isometric) system are independent of direction. Macroscopic properties of substances which are microcrystalline, such as many metals as ordinarily produced, may also apparently be the same for all directions, the measured properties being averages or summations of the properties of the individual randomly oriented "crystallites."

ties, method of production) which it is desired to emphasize. For example, industrial usage distinguishes between glasses and solid organic plastics, on the basis of characteristic differences in hardness.

A rigorous definition of a vitreous substance is impossible, without the introduction of purely arbitrary quantitative limiting values for the rigidity (or viscosity) and the dependence of various properties on direction or for the degree of permanence and lack of order in the atomic distribution. Rather than to introduce such arbitrariness, it seems better to consider the term *vitreous* merely as a convenient descriptive adjective, realizing that there are many borderline cases to which it may be applied or not, as convenient.

On raising the temperature, most vitreous solids are transformed to liquids; the transformation, however, is a gradual one, whereas change from the crystalline state to the liquid state is sharp.* A vitreous substance in the transition range has properties intermediate between those characteristic of an amorphous solid and of a liquid. It may be described either as a very viscous liquid or as a plastic solid, as suits one's fancy. Tars are typical examples of vitreous substances whose transition range is at ordinary temperatures. The transition ranges for thermoplastic resins, for example, are somewhat above, and for common silicate glasses very much above, ordinary room temperatures.

Gels, consisting of a more or less rigid, high-molecular framework with interspersed small molecules, each in much the same state as in a liquid, are also intermediate between the vitreous and liquid states, though in a somewhat different sense.

As examples of substances intermediate between vitreous and crystalline solids, one may mention glasses which are inhomogeneous and show a tendency toward crystallization, and also organic long-chain polymers which have been given directional properties by stretching or cold-drawing.

Stretched rubber or other elastic substances may be considered to be intermediate between the vitreous, crystalline, and liquid states. As regards its rigidity, it is intermediate between a liquid and a vitreous solid. It has, to a limited extent, the directional properties characteristic of the crystalline state.

Looking at the subject from an atomic point of view, the reasons for the existence of many borderline cases are obvious. Like atoms (or ions or groups of atoms) tend to be surrounded by other atoms (or ions or atomic groups) in a like manner. In any condensed system—crystalline solid, vitreous solid, or liquid—there tends to be a regularity in the local atomic distribution immediately surrounding each atom (or ion or group of atoms). Thus, in all silicate crystals of known structure and presumably also in silicate glasses, each silicon atom is tetrahedrally surrounded by four oxygen atoms at a distance of about 1.6 Å, and each oxygen has adjacent to it either two silicon atoms or one silicon and one or more atoms of another element or elements. If, in addition to the local regularity, there is also a long-distance regularity, the substance is crystalline; if not, it is a vitreous solid or a liquid, depending on the degree of permanence of the long-distance structure—a function of temperature.

A crystalline arrangement of atoms or molecules has (nearly always, at least) a lower energy than a non-crystalline arrangement, although the difference is usually not large. If the interatomic (or interionic or intermolecular) forces tending to maintain a momentarily existing long-distance structure are small enough, the thermal vibrations of the atoms will cause the structure to change rapidly to a different one: the substance is a liquid. If the substance is then cooled rapidly to a temperature such that the average kinetic energy tending to change the structure is considerably smaller than the energy humps tending to prevent rearrangement to a

* Except in substances which show mesomorphic states. See paper by G. Friedel in Vol, I of this series. J. A.

crystalline structure, a vitreous solid results. If, on the other hand, the cooling takes place sufficiently slowly, and if the energy of a structure having long-distance regularity is enough lower than the energy of any other structure, a crystalline substance is produced. The existence of vitreous substances is thus closely related to the strengths of the interatomic, interionic, and intermolecular forces and to the rate of cooling from the liquid state.

In general, a substance tends to be a solid if interatomic bonds, strong enough successfully to resist breaking by thermal collisions, form a three-dimensional network structure extending throughout the mass. If a *regular* network structure has a considerably lower energy than any *irregular* structure, the solid tends to be crystalline. If, however, different arrangements have practically equal energy, it tends to be amorphous. For example, long-chain polymers having occasional crosslinks (or strong interchain forces) between the chains tend to be amorphous, because different orientations around each of the single bonds in the chains have nearly the same energy and so are almost equally probable. Moreover, if the distribution of crosslinks is a random one, as is usually the case, regularity of structure is impossible.

Similar considerations apply to silicate glasses. The strongest interatomic forces in these are usually between adjacent silicon and oxygen atoms. Because of these forces, as already noted, each silicon is bonded tetrahedrally to four oxygens, and each oxygen either to *two* silicons (the angle between the bonds being roughly the "tetrahedral" angle) or to *one* (usually in addition to one or more other metallic atoms). In most ordinary glasses, the ensemble of Si-O bonds forms a network structure. The rings, of which this network may be considered to be composed, are of various sizes—partly because of the effects of the other atoms, of various sizes and coordination numbers, and partly because the energy of the network depends only very slightly on the ring-size distribution (See Fig. 1). The fact that different

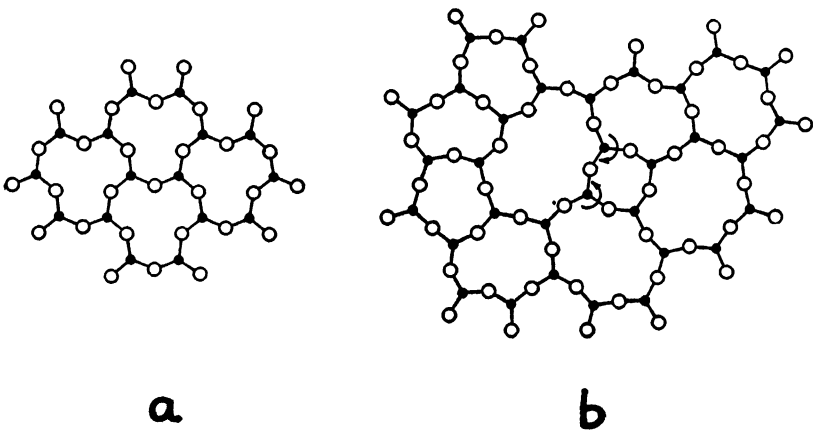


FIGURE 1. Two-dimensional representations of crystalline (a) and vitreous (b) SiO_2 structures (after Zachariasen[47] and Warren[41]).

relative orientations of practically equal stability are possible for the two Si-O bonds extending from many of the oxygen atoms contributes both to the irregularity of distribution of ring sizes and to the irregularity in shape of the rings of the same size.

A polyhydroxy small molecule compound, such as glucose, tends to be vitreous, when solid, both because the intermolecular hydrogen bridges are of nearly equal strength, regardless of which oxygen atoms are connected by them and regardless

of the O—H · · · O angle (within limits), and because different orientations of each O—H (or O · · · H) bond, relative to the adjacent C—O bond, give nearly equal stability.

All of the substances which we consider amorphous, both liquids and solids, possess some degree of order—at least the already-mentioned approximate regularity of arrangement around each kind of atom and approximate uniformity of corresponding interatomic distances, bond angles, etc. Order over somewhat larger distances also exists, in many cases at least, although macroscopic properties give no direct evidence of this on account of the randomness of orientation of the small regions (usually of colloidal dimensions) possessing this order. The permanence or evanescence of short-range order and intermediate-range order depends largely on the strengths of the interatomic forces involved and on the temperature.

Crystalline substances often possess some degree of disorder; to the extent that this is the case, they are not perfectly crystalline. Foreign atoms (impurities) and vacant atomic sites may be distributed irregularly throughout the structure. "Crystallographically equivalent" positions may be occupied irregularly by atoms of two or more kinds [3, 4, 5, 40] (in many "solid solutions"). Twinning may occur, producing differences of crystal orientation in adjacent portions of the structure. The orientations of certain atomic groups (*e.g.*, in organic molecules) in a crystal may be more or less random, even though the distribution and orientation of the molecules as a whole is regular.[32] Atomic groups and even whole molecules may, in fact, be continuously rotating in the crystal.[33, 38] Strictly speaking, also, there is a randomness in the instantaneous distribution of atoms in any crystal, resulting from their temperature vibrations and rotations.

Some Semiquantitative Considerations

Most of the temperature motions of the atoms in any solid or liquid can be considered as oscillations (not mutually independent) about equilibrium positions. In addition to the changes of relative atomic positions resulting from these oscillations, larger changes may occur, in which the energy (strictly, the *free* energy[25]) of the system first increases and then decreases; an energy hump is passed over. In many instances the change taking place can be (approximately) described very simply, *e.g.*, as a shift of position of a single atom or a shift of position or of orientation of a group of atoms, even though smaller displacements of other neighboring atoms must necessarily occur concurrently.

Each rearrangement of this sort is characterized by two important energy quantities: (1) the difference (ΔE) between the energy of the system before the change and that after the change, and (2) the magnitude of the energy hump passed over, *i.e.*, the difference between the energy of the system before the change and its maximum energy at any time during the change (See Fig. 2).

The latter quantity is called the "activation energy" for the change; we shall designate it as E_{a1} for a change from a lower energy state (1) to a higher state (2) and as E_{a2} for a change from a higher energy state to a lower. Obviously,

$$\Delta E = \pm(E_{a1} - E_{a2}), \tag{1}$$

the plus sign applying to the change from state 1 to state 2, the minus sign to the reverse change.

Generally speaking, amorphous substances are those for which many internal changes of this sort are possible, with ΔE small in magnitude. This statement will now be amplified and put on a semiquantitative basis.

If, for a given atomic shift or rearrangement, states 1 and 2 are in thermodynamic equilibrium with each other (but not with any additional states), it can be shown that the probability that the system is in state 1 is

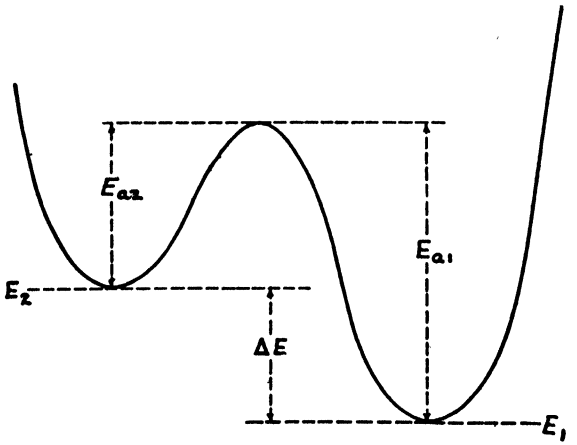


FIGURE 2. Illustrating certain energy relationships pertaining to a structural rearrangement in a solid or liquid.

$$P_1 = \frac{e^{\Delta\epsilon}}{1 + e^{\Delta\epsilon}}, \tag{2}$$

where

$$\Delta\epsilon = \frac{|\Delta E|}{k'T}, \tag{3}$$

T being the absolute temperature and k' being a constant depending on the Boltzmann constant (k) and also on the number of atoms making the shift and the tightness with which these atoms are bound to each other and to neighboring atoms, *i.e.*, the independence of their oscillational motion. For the purpose of this discussion, k' can be considered to have a uniform, constant value, approximately equal to k.

The probability that the system is in state 2 is

$$P_2 = 1 - P_1 = \frac{1}{1 + e^{\Delta\epsilon}} \tag{4}$$

Figures 3 and 4 show the variation of P_1 and P_2 with $\Delta\epsilon$ and with T (for a constant $\Delta E/k'$). The greater the energy difference, at a given temperature, the

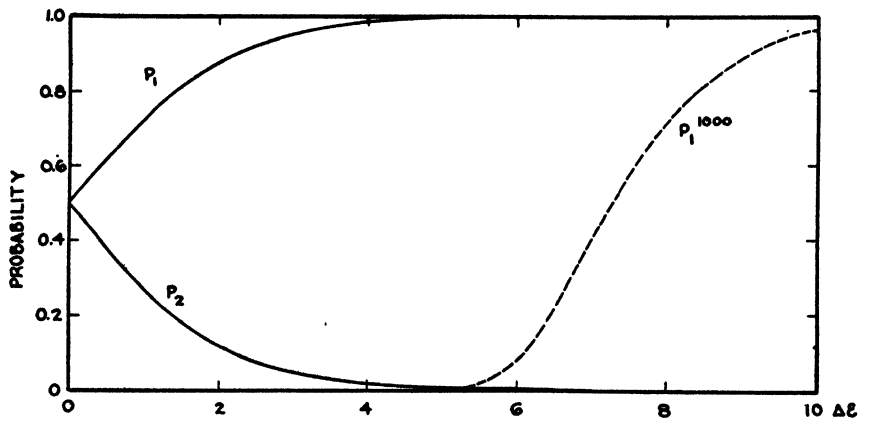


FIGURE 3. The probabilities of two states, in equilibrium with each other, as functions of $\Delta\epsilon$. The probability of the simultaneous existence of 1000 independent subsystems (all having the same value of $\Delta\epsilon$) is also shown (dotted line).

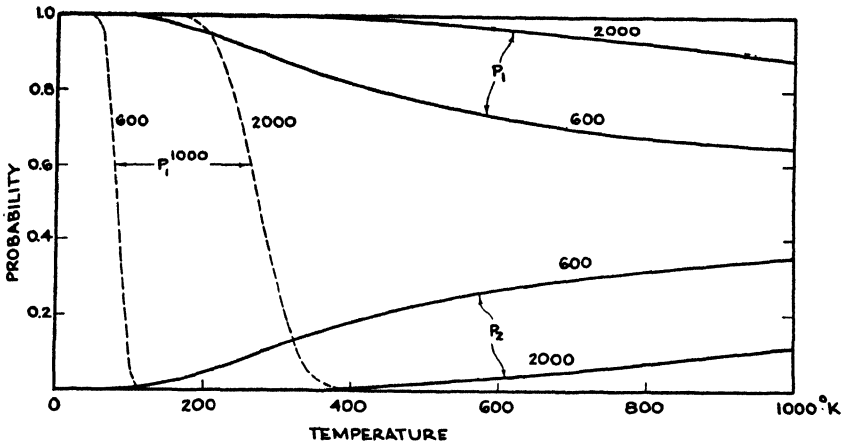


FIGURE 4. The probabilities of two states, in equilibrium with each other, as functions of the absolute temperature, assuming $\Delta E/k'$ constant (equal to 600 or 2000, as indicated). The probabilities of the simultaneous existence in state 1 of 1000 independent subsystems are also shown (dotted lines).

greater is the probability that the system is in the lower energy state. The higher the temperature, for a given value of $\Delta E/k'$, the more nearly equal are the probabilities for the two states.

If a given sample of the substance under consideration contains a large number of small subsystems, each of which can have two or more atomic distributions differing in energy by the same amount (or, more precisely, by the same value of $\Delta E/k'$, the probabilities just discussed and represented in Figures 3 and 4 are equal to the fractions of these subsystems which are in the respective states at a given instant, provided there is thermodynamic equilibrium.

A substance of this sort should be considered crystalline only if (practically) all of the subsystems throughout relatively large regions, containing many thousands of atoms, have the same atomic distribution—normally that (state 1) having the lowest energy. The probability that a mole of the substance is crystalline is (practically) P_1^n, n being the number of subsystems, of the sort just discussed, in the mole. The dotted curves in Figs. 3 and 4 show how P_1^n varies with $\Delta\epsilon$ and with T, for $n = 1000$. It is obvious that with very large values of n, a crystalline structure is very improbable, unless $\Delta\epsilon$ is large.

The change of entropy (per mole) accompanying crystallization, for the sort of substance being discussed, is

$$\begin{aligned}\Delta S_c = S_c - S_v &\approx k \ln P_1^n - k \ln (1 - P_1^n) \\ &\approx kn \ln P_1. \end{aligned} \tag{5}$$

The molal change in heat content—equal to the change in energy if there is no volume change—accompanying crystallization is

$$\Delta H_c = \Delta E_c = E_c - E_v = -nP_2 \Delta E. \tag{6}$$

The free energy change associated with crystallization is therefore

$$\begin{aligned}\Delta F_c = F_c - F_v &= \Delta H_c - T\Delta S_c \\ &\approx -nP_2 \Delta E - nkT \ln P_1. \end{aligned} \tag{7}$$

Fig. 5 shows the manner in which the free energy of crystallization varies with the temperature, according to this equation and Equations (2), (3), and (4).

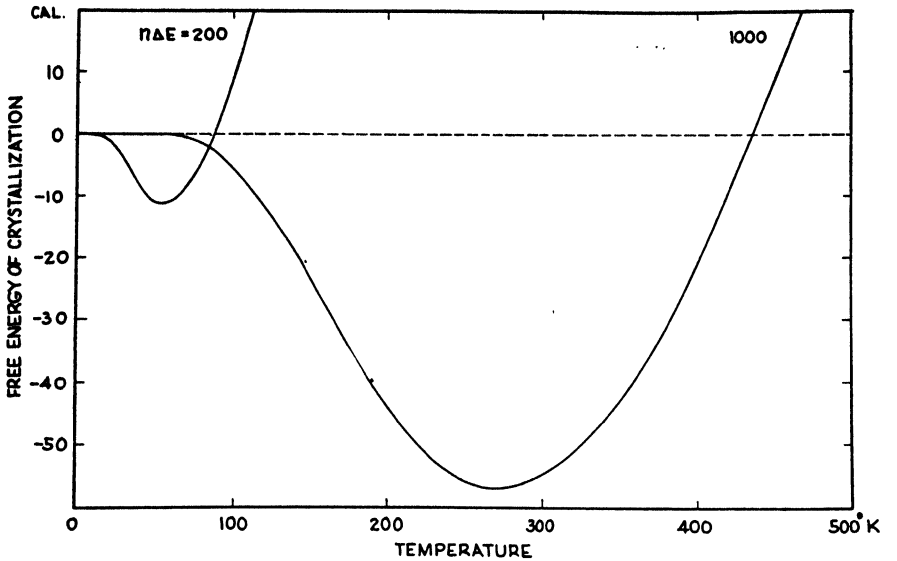


FIGURE 5. Dependence of the free energy of crystallization on the absolute temperature, according to Eq. (7), assuming $n\,\Delta E = 200$ cal and 1000 cal, with $k' = k$ and $n =$ Avogadro's number in both cases.

The temperature at which $\Delta F_c = 0$ is that below which the crystalline state is more stable than the amorphous, and above which a noncrystalline distribution is the more stable. This temperature is the melting point of the crystalline material, if the viscosity of the noncrystalline substance at that temperature is sufficiently low for it to be liquid.

It should be emphasized that the foregoing discussion applies only to a substance in a state of thermodynamic equilibrium. The rate at which the equilibrium state is approached and the viscosity of the amorphous material both depend primarily, not on the energy difference (ΔE) between more or less stable states, but on activation energies (E_{a1} and E_{a2}) which may be much larger than the ΔE values.

Neglecting certain minor considerations, the rate of change from state 1 to state 2 (or from 2 to 1) is proportional to the number of subsystems in state 1 (or state 2) and to $e^{-\epsilon_1}$ (or $e^{-\epsilon_2}$), where ϵ_1 and ϵ_2 are defined by the equations

$$\epsilon_1 = \frac{E_{a1}}{k'T} \tag{8}$$

and

$$\epsilon_2 = \frac{E_{a2}}{k'T} \tag{9}$$

If we denote by f_2 the instantaneous fraction of subsystems in state 2 at time t, by $f_{2,0}$ the fraction in this state at $t = 0$, and by $f_{2,\infty}$ that fraction at $t = \infty$ (i.e., at equilibrium), the net rate of change of f_2 can be shown to be

$$\begin{aligned}\frac{df_2}{dt} &= k''(f_{2,\infty} - f_{2,0})(e^{-\epsilon_2} + e^{-\epsilon_1})\exp[-k''(e^{-\epsilon_2} + e^{-\epsilon_1})t] \\ &= k''(f_{2,\infty} - f_{2,0})(1 + e^{-\Delta\epsilon})\exp[-\epsilon_2 - k''e^{-\epsilon_2}(1 + e^{-\Delta\epsilon})t]\end{aligned} \tag{10}$$

A similar relationship holds for df_1/dt. For our present purpose, k'' can be considered merely as another constant.

If $\Delta\epsilon$ is not too small, $e^{-\Delta\epsilon}$ can be neglected relative to unity; then

$$\frac{df_2}{dt} = k''(f_{2,\infty} - f_{2,0})\exp[-\epsilon_2 - k''e^{-\epsilon_2 t}]. \tag{11}$$

For the actual fraction, f_2, at time t, the relation (also for $\Delta\epsilon$ not too small) is

$$f_2 = f_{2,\infty} + (f_{2,0} - f_{2,\infty})\exp[-k''e^{-\epsilon_2 t}]. \tag{12}$$

In place of Equation (11), we may therefore write

$$\frac{df_2}{dt} = k''(f_{2,\infty} - f_2)\exp[^{-\epsilon_2}]. \tag{13}$$

Figure 6 shows the dependence of the net rate of change from state 2 to state 1

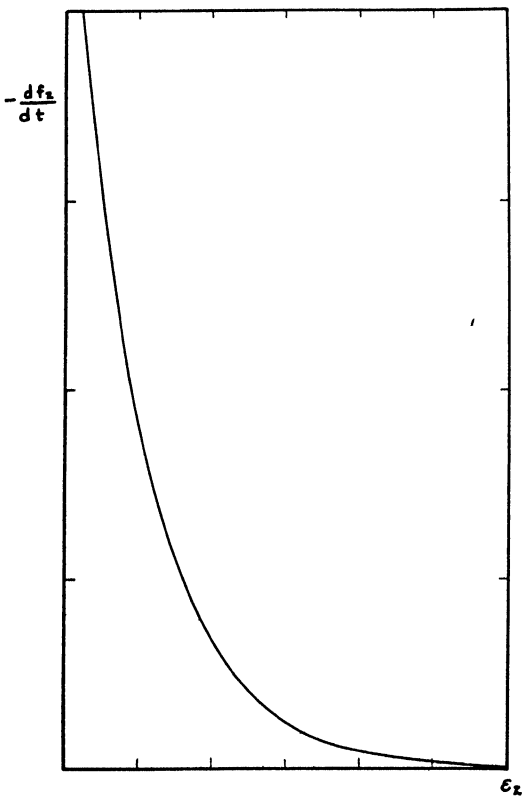


FIGURE 6. Dependence of the net rate of change from state 2 to state 1 on the magnitude of ϵ_2, according to Eq. (13). (The units are arbitrary.)

on the magnitude of ϵ_2, according to Equation (13). The scale is arbitrary, depending on the magnitude of $k''(f_{2,\infty} - f_2)$ at the time being considered. At constant temperature, the rate is seen to vary greatly with the magnitude of the activation energy E_{a2} (or, better, with E_{a2}/k').

Fig. 7 shows the very great dependence of the net rate of change from state 2 to state 1 on the temperature, for a constant value of E_{a2}/k'.

Fig. 8 shows the manner in which the fraction of subsystems in the upper energy state varies with time, for various values of $k''\epsilon_2$, according to Equation (12). The

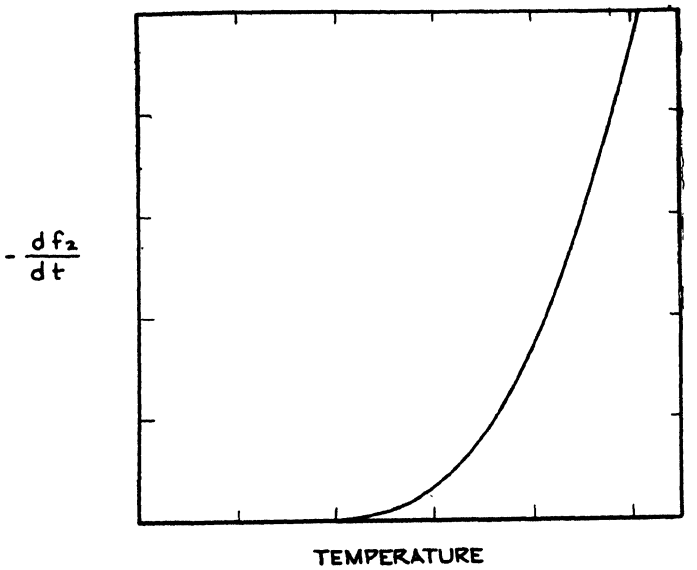


FIGURE 7. Dependence of the net rate of change from state 2 to state 1 on the absolute temperature, for a constant value of E_{a2}/k', according to Eq. 13. (The units are arbitrary.)

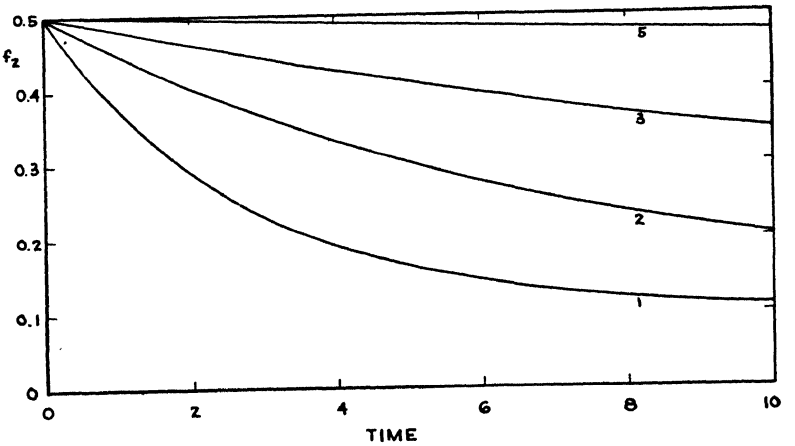


FIGURE 8. Variation with time (in arbitrary units) of the fraction of subsystems in the upper energy state (2), for various values of k'' e^2 (indicated by the numbers alongside the curves), according to Eq. (12).

different curves again illustrate the great dependence on the activation energy and the temperature of the rate at which the subsystems reach true stability (state 1).

In actual substances, there will, as a rule, be more than one type of structure shift, each characterized by two independent constants,

$$\frac{\Delta E}{k'} \text{ and } \frac{E_{a1}}{k'} \left(\text{or } \frac{E_{a1}}{k'} \text{ and } \frac{E_{a1}}{k'} \text{ or } \frac{\Delta E}{k'} \text{ and } \frac{E_{a1}}{k'}\right).$$

These constants determine the relative fractions of the appropriate subsystems in the two energy states at equilibrium and the rate at which equilibrium is approached at any given temperature. A substance will be amorphous at a given temperature if there are considerable numbers of subsystems in energy states other than the lowest —which will be the case (at equilibrium) if $\Delta\epsilon$ is small for any (not too rare) type of structure shift. The substance will also be amorphous, even though equilibrium at the existing temperature would require practically all of the subsystems to be in their lowest energy states, if too rapid cooling from a higher temperature (at which many of the subsystems were in higher energy states) has so slowed the rates of the structure shifts as to prevent the attainment of equilibrium at the lower temperature.

Considerations similar to those above apply also to the theoretical treatment of viscosity. Without going into details, it may be mentioned that the viscosity is determined largely by the magnitudes of the energy humps (or, better, of the ratios $E_a/k'T$) which must be passed over as each small region of the substance shifts, in a given direction, relative to neighboring regions.

Let us now consider the state of affairs in a silicate glass of ordinary composition. The strongest interatomic forces are undoubtedly those between silicon and oxygen atoms. At ordinary temperatures, the ensemble of silicon and oxygen atoms is an irregular giant network (negatively charged). At sufficiently high temperatures, such as those at which the glass was made, the viscosity is that of an ordinary liquid; the network is continually being broken and re-formed, with a change of configuration. This necessitates the breaking and re-formation of Si-O bonds. Both the activation energy and the absolute temperature are large; the rate of the break-and-make process is determined chiefly by the ratio of the former quantity to the latter.

As the temperature is lowered, the rate of breaking of Si-O bonds rapidly becomes smaller and the viscosity greatly increases. Within the temperature range in which the viscosity is rapidly changing (or at any higher temperature), the structure existing at any instant is far from regular—because the differences in energy between different configurations (different sizes of rings, etc.) for each small region of the structure are small relative to the thermal energy available.

On further cooling, practically no Si-O bonds are broken; in this respect the structure remains fixed. Other changes, having smaller activation energies, continue to occur at temperatures somewhat below the range of rapid viscosity increase. These include the breaking and making of other metal-oxygen bonds and changes of orientation of Si-O bonds in the silicate net, insofar as possible without breaking any of these bonds or changing Si-O-Si bond angles too greatly from their most stable values. Those changes are preferred, over changes in the opposite direction, which result in a decrease in (free) energy.

If the cooling takes place very slowly through each temperature range in which the rate of shift, for a given type of structure change, is rapidly decreasing, approximate equilibrium, with respect to each of these types of changes, can be maintained. In those types of subsystems in which one of two alternative distributions is considerably more stable than the other, practically all will have the more stable distribution, when the temperature is reached at which the rate of shift becomes negligible. Slow cooling over all of these critical regions therefore produces a glass at room temperature which approaches the lowest possible energy—maximum stability. Rapid cooling through any of the critical regions results in the "freezing-in" of some relatively unstable atomic distributions. This results in strains and inhomogeneities. The tendency to relieve these strains, to give a thermodynamically more stable structure, produces slow internal changes persisting over a long period of time.

In general, a solid is much more readily fractured if its atoms are in a crystalline arrangement than if they are in an amorphous arrangement. The crystalline or

other atomic arrangement at and near the surface is especially important, since fracture nearly always starts from surface cracks. (Etching a glass surface usually increases the mechanical strength.) The "tempering" or "case-hardening" of vitreous products—sudden, uniform surface chilling from a temperature only slightly below the softening temperature—thus produces a uniformly amorphous surface region in which cracks are not readily started. The stability and other properties of the resultant product are of course very dependent on the temperature just before chilling, the rate and uniformity of chilling, etc.

Very slow cooling through temperatures only slightly below those at which the glass is fluid favors the production and growth of crystalline regions in the mass, the importance of this tendency being very greatly dependent on the composition, however. Such crystalline regions—especially if the crystal structure is quite anisotropic, with strength and other properties varying greatly with direction—may be very deleterious, producing inhomogeneity and other undesirable properties in the final product. On the other hand, the presence of many small crystalline regions in an amorphous matrix may make a glass artistically desirable, as in Aventurine glass (*e.g.*, "Goldstone").

Vitreous Sulfur and Selenium

In this and the following sections some typical substances and types of substances which form vitreous solids will be discussed, with emphasis on the structural relationships.

Among the elements known in the vitreous state are sulfur and selenium. According to the Lewis theory of valence and atomic structure,[26] a neutral atom of one of these elements has six valence electrons. It can complete a tetrahedral, four electron-pair valence shell by forming two single bonds—sharing two electron-pairs—with other atoms. (Double bonds are apparently unstable.) One would therefore expect these elements to form either rings or very long chains, each atom in either case (except for the few terminal atoms of the long chains) being bonded to two others, with a bond angle at least roughly approximating the tetrahedral angle ($\sim$110°). This expectation has been verified. The ordinary crystal forms of sulfur [8, 44] and (probably) two of the crystal forms of selenium [24] contain 8-atom rings. Another crystal form of selenium [6, 36] contains long helical chain molecules, extending completely through the crystal.

Amorphous sulfur and selenium can be obtained by rapidly cooling the liquids, after heating at a temperature sufficient to break the rings. These amorphous substances contain long chains of atoms [27] and probably also rings of various sizes, the chains and rings having many different configurations, because of the nearly equal stability of various orientations of each single bond relative to each of the two adjacent bonds. This picture is consistent with the rubberlike elasticity of amorphous sulfur at ordinary temperatures and of amorphous selenium at somewhat higher temperatures—a property now known to be due to the entropy decrease occurring when randomly kinked chains become more or less straightened out and mutually aligned, on stretching the material.[9, 23, 28] The elasticity disappears as the temperature is lowered (as in the case of rubber) because the thermal motions of the atoms become insufficient to produce the changes of orientation at a sufficiently rapid rate.

Vitreous Silica

The structure of vitreous silica, SiO_2, will next be discussed. Reasoning again from the Lewis theory of valence, it was concluded [20, 26] that in silicon dioxide each silicon atom should be tetrahedrally bonded to four oxygen atoms and each oxygen should be bonded, by bonds making (approximately) tetrahedral angles with each other, to two silicons. These conclusions have been verified by the results of x-ray studies of quartz [7, 18, 20, 45] and vitreous silica.[42, 43] (In the structures first deduced for

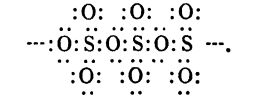
tridymite [14] and cristobalite,[35, 46] the crystalline forms of silica which are stable at higher temperatures, the two bonds at each oxygen are assumed to be collinear. This may be the case at temperatures somewhat above room temperature, but, as later work has shown,[2] in cristobalite the Si-O-Si bond angles are somewhat less than 180°. This is probably also true of tridymite).

Because of the possibility, without much energy change, of different orientations of each Si-O bond relative to the other Si-O bond leading from the same oxygen atom, many different atomic arrangements satisfying the bond requirements just mentioned have nearly the same stability. The most stable of these, at room temperature, is that of quartz, but at much higher temperatures, many less regular structures must exist, in equilibrium with each other (See Fig. 1). A change from one of these structures to another involves (in most instances) the breaking of relatively strong Si-O bonds. The activation energy for this is large—so large that the rate must be negligibly slow at ordinary temperatures. Sudden cooling of the melt, through the temperature range within which this rate is not negligible, therefore "freezes-in" the irregular arrangements, producing silica glass.

Other Oxide Glasses

Other pure oxides which have yielded glasses on rapid cooling from the molten state include the following: B_2O_3,[15, 30, 47] GeO_2,[15, 30, 47] ZrO_2,[10] P_2O_5,[15, 30, 47] As_2O_3,[15, 30, 47] As_2O_5,[15, 47] Sb_2O_3 [47] and Bi_2O_3.[34] Possibly P_2O_3,[47] V_2O_5,[15, 30, 47] Sb_2O_5 [47] and SO_3 [12] should also be included in this list. In all of these substances one would expect, theoretically,—in both the crystalline and vitreous states—rather strong metal-oxygen electron-pair bonds. Except in ZrO_2, one would expect two such bonds extending approximately at tetrahedral angles from each oxygen atom (or, in P_2O_5, As_2O_5, Sb_2O_5, and V_2O_5, from most of the oxygen atoms). The crystal structures of most of these compounds have been determined and have been found to agree with these theoretical expectations. According to both the theory and the crystal structure results, three or more of the oxygen atoms surrounding each atom of the more metallic element are bonded also to other "metal" atoms, except in the case of SO_3. This permits a three-dimensional network structure, and such a structure doubtless exists in these glasses. The randomness of the atomic arrangements in the glasses results from the possibility of different relative orientations of the two bonds connecting each oxygen to other atoms and, in those instances in which all of the oxygens surrounding each "metal" atom are not also bonded to other atoms, from a randomness in the distribution of the bridging and nonbridging oxygens.

An amorphous, elastic form of sulfur trioxide can be produced [12] as a skin on the surface of a solution of SO_3 in SO_2. This probably contains long, *kinked* chains, corresponding to the dot formula (pairs of dots representing valence electron-pairs)



These chains may perhaps have side chains; rings may also be present.

Silicate Glasses

In most silicates, both crystalline and vitreous, the Si-O bonds are much stronger than any others present. (B-O and perhaps Al-O bonds are comparable in strength. For simplicity we shall neglect these here.) In agreement with theoretical expectations, it has been found that in silicate *crystals* each silicon is tetrahedrally surrounded by four oxygens and that each oxygen is bonded to either one or two silicons. Such evidence as there is regarding the structure of silicate *glasses* favors the assumption that a similar state of affairs exists in these also.

The SiO_4 tetrahedra may be classified as types 0, 1, 2, 3, and 4, on the basis of the number of oxygens shared with other silicons. At least in crystals and glasses in which there are no other bonds comparable in strength with the Si-O bonds, it seems reasonable to assume that for greatest stability only types designated by adjacent numbers (0 and 1 or 1 and 2 or 2 and 3 or 3 and 4, but not 0 and 2 or 1 and 3, etc.) can exist together in a given sample. In other words, two SiO_4 tetrahedra of types 1 and 3 (having one and three oxygens shared), for example, would tend to react together in such a way as to produce two tetrahedra of type 2. The authors know of no contradictions to this generalization among silicate crystals. Among glasses we might expect some deviations, especially if the composition is such that nearly all of the tetrahedra are of a single type. For instance, a glass having nearly all of type 2 might well have some of type 1 and some of type 3. Nevertheless, some conclusions based on this generalization are of interest and may be considered to be approximately correct.

For this purpose, we make use of a quantity N_{Si}, defined as the ratio of the number of silicon atoms to the number of oxygen atoms in the glass (or crystal). It may be computed [17] from the weight fractions (f_M) of the different oxide components (having formulas M_mO_n and molecular weights W_M) by means of the relation

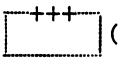
$$N_{Si} = \frac{f_{Si}/W_{Si}}{\Sigma n_M f_M/W_M} = \frac{f_{Si}/60.06}{\Sigma n_M f_M/W_M}. \tag{14}$$

N_{Si} and s, the average number of oxygen atoms of each SiO_4 tetrahedron which are shared with other silicons, are related by the equation

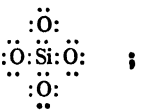
$$s = 8 - \frac{2}{N_{Si}}. \tag{15}$$

Table 1 gives both the values of s and of f_0, f_1 . . . f_4, the fractions of the SiO_4 tetrahedra which are of types 0, 1 . . . 4, respectively, for various compositions (N_{Si} values) and composition ranges, assuming the generalization stated above to be strictly correct.

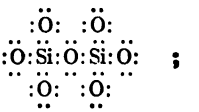
Table 1. Hypothetical Dependence of Structures on Composition, for Silicate Glasses

N_{Si}	s	f_0	f_1	f_2	f_3	f_4	Diagrammatic Representation
0.250	0	1	0	0	0	0	+
.250-.286	0-1	$\frac{2}{N_{Si}} - 7$	$8 - \frac{2}{N_{Si}}$	0	0	0	+ and ++
.286	1	0	1	0	0	0	++
.286-.333	1-2	0	$\frac{2}{N_{Si}} - 6$	$7 - \frac{2}{N_{Si}}$	0	0	++···+ (chains), with or without  (rings)
.333	2	0	0	1	0	0	Rings and/or infinitely long chains
.333-.400	2-3	0	0	$\frac{2}{N_{Si}} - 5$	$6 - \frac{2}{N_{Si}}$	0	Network, composed of ····+···· and ····+····, with or without rings
.400	3	0	0	0	1	0	Network, composed of ····+····
.400-.500	3-4	0	0	0	$\frac{2}{N_{Si}} - 4$	$5 - \frac{2}{N_{Si}}$	Network, composed of ····+···· and ····+····
.500	4	0	0	0	0	1	Network, composed of ····+····

In the right-hand column, the silicate structures present are symbolized diagrammatically. The crosses represent SiO_4 tetrahedra: + denotes SiO_4^{-4} ions, having the "dot formula"



++ denotes $Si_2O_7^{-6}$ ions, having the dot formula



etc.

In the composition range $0.286 < N_{Si} < 0.333$, the average number of silicon atoms per chain is given by the equation

$$\bar{n}_{Si} = \frac{f_c N_{Si}}{1 - 3N_{Si}}, \tag{16}$$

where f_c is the fraction of the total number of silicon atoms which are in chains, rather than in rings.

Except for pure SiO_2 ($N_{Si} = 0.50$), each aggregation of silicon and oxygen atoms, joined together through Si-O bonds, is negatively charged, the magnitude of the charge being given by the formula $Si_xO_y^{4x-2y}$. This charge is balanced by positively charged ions, such as Na^+, Ca^{++}, Al^{+++}, also present in the structure. The distribution of the positive ions, although of course important in determining the properties of the glass, will not be discussed at this time.

It must be emphasized that the distributions of silicon and oxygen atoms indicated in Table 1 are, at best, only approximate, limiting distributions. The energy change for the reaction between a type 1 tetrahedron and a type 3 tetrahedron to give two type 2 tetrahedra, for instance, is probably not large. At any temperature at which the rate of this reaction would not be negligible, it is probable that the equilibrium distribution would contain a considerable fraction of types 1 and 3 coexisting in the same sample. Glasses having $N_{Si} \leqq 0.333$ would therefore be expected to have some *branched* chains and those having $N_{Si} \geqq 0.333$ would be expected to have some chains with loose ends. The relative numbers of tetrahedra of the different types are certainly affected, also, by the kinds and numbers of other atoms or ions present, especially by those (B, Al) having the strongest attractions for the oxygens.

In spite of these limitations, it is felt by the authors that the dependence of structure on composition indicated in Table 1 is approximately correct and will prove useful in future discussions of the dependence of the properties of silicate glasses on their compositions. Considerations such as these have, in fact, already proved helpful in leading to reasonably satisfactory treatments of the dependence of the densities,[17, 21] refractive indices,[18, 21] and dispersions [19, 21, 22] of silicate glasses on their compositions.

The conclusions regarding the structure of silicate glasses containing no atoms, other than silicon, which form strong electron-pair bonds to oxygen, are readily generalized, on the basis of similar assumptions, to other mixed oxide glasses—both those containing silicon, plus other elements such as boron, phosphorus, germanium, etc., and those containing no silicon.

For instance, Table 1 is applicable *in toto* to glasses containing as components only SiO_2, GeO_2, P_2O_5, and As_2O_5 (in all of which each atom of the more metallic element is bonded to four oxygen atoms) and oxides of more electropositive elements, provided N_{Si} is replaced by $\sum_A N_M$, the summation being over the "Class A" elements: Si, Ge, P^V, As^V. Boron can probably also be included among the Class A elements, provided $\sum_A N_M$ is less than one-half; otherwise some boron atoms will be bonded to only three oxygen atoms and the relationships are more complicated.

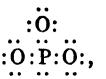
The assumption that types 1 and 3 (or 0 and 2 or 2 and 4, etc.) do not coexist is a more radical assumption if two or more Class *A* elements are present than if only one is present. Two or even three or four of the oxygens around many of the silicon atoms in a structure containing both silicon and phosphorus, for instance, may perhaps be shared, even though some of the PO_4 tetrahedra share none of their oxygens with other elements. If this sort of situation occurs, the generalization of Table 1, just described, is incorrect.

If the glass contains oxides of "Class B" elements, such as P^{III}, As^{III} and perhaps Sb^{III} (in which each of these atoms is bonded to only *three* oxygens), and oxides of more electropositive elements—but no oxides of Class A elements—Table 2 replaces Table 1. For simplicity, only those compositions corresponding to integral *s* values

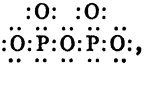
Table 2. Hypothetical Dependence of Structure on Composition for Glasses Containing Oxides of Class B Elements

$\sum_A N_M$	*s*	Diagrammatic Representation
0.333	0	⊥
.400	1	⊥ ⊥
.500	2	···⊥··· (rings and/or infinite chains)
.667	3	···⊥··· (network)

(or zero) are listed. The symbols ⊥ and ⊥ ⊥, in the last column, represent (if M = phosphorus) PO_4^{-3},



and $P_2O_5^{-4}$,



respectively.

This table also is based on the assumption that types 0 and 2, 0 and 3, and 1 and 3, will not coexist; it may well be incorrect in many cases, especially if more than one kind of B atom is present.

If both Class A and Class B elements are present, Table 3 applies—again provided the corresponding assumption, regarding the mixing of different types, holds.

Further experimental and theoretical studies are required before one can proceed much further with the correlation of structure with composition, in glasses containing more than one type of element of Class A or Class B, or both.

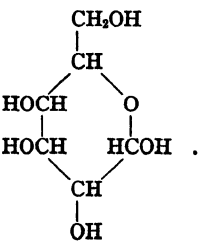
If elements forming more than four strong bonds with oxygen are present, still further complications arise. These will not be considered here.

Table 3. Hypothetical Dependence of Structure on Composition for Glasses Containing Oxides of both Class A and Class B Elements

Composition	s	Diagrammatic Representation
$4\sum_A N_M + 3\sum_B N_M = 1$	0	+, ⊥
$\frac{7}{2}\sum_A N_M + \frac{5}{2}\sum_B N_M = 1$	1	++, ⊥⊥, +⊥
$3\sum_A N_M + 2\sum_B N_M = 1$	2	····+····, ···· ⊥ ···· (rings and/or infinite chains)
$\frac{5}{2}\sum_A N_M + \frac{3}{2}\sum_B N_M = 1$	3	····+····, ···· ⊥ ···· (network)
$\frac{5}{2}\sum_A N_M + \frac{3}{2}\sum_B N_M = 1$	3-4	····+····, ····⊥···· (network)

Glasses Containing Hydrogen Bridges

Glasses in which molecular (or ionic) units are held together by OHO hydrogen bridges will now be discussed. Most organic hydroxy compounds form vitreous masses of this sort. A typical example is glucose, which has the molecular formula

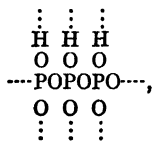


.

The possibilities of different orientations of the OHO bridges relative to the adjacent C—O bonds and of considerable variation in bond angles at (and within) these bridges, as well as other factors (in many instances), help to produce an amorphous rather than a crystalline structure. The forces tending to prevent rearrangement of a momentarily existing structure are much less than for silicate glasses; the substances are therefore fluid at much lower temperatures. This greatly simplifies the investigation of the variation of various properties with temperature, through the transition region between the liquid and the vitreous conditions.

Curves showing the variation with temperature of the specific heat, coefficient of expansion, and dielectric constant of glucose are shown in Fig. 9. It is evident that, although there is no actual discontinuity, as in the case of a crystal-liquid transition, rather large changes in these properties take place within a small temperature interval. As would be expected, the dielectric constant—the magnitude of which depends chiefly on the ability of the polar molecules and atomic groups to orient in the electric field within the time of a half-cycle—attains its maximum value only at temperatures at which the substance is quite fluid.

Two examples of vitreous substances composed (primarily, at least) of long chains, held together by means of OHO hydrogen bridges, may be mentioned. One is glacial phosphoric acid, one of the forms of metaphosphoric acid, HPO_3. This substance, the authors believe, consists of long chains,



kinked in a more or less random manner and connected to each other by means of the hydrogens. Some closed rings may also be present.

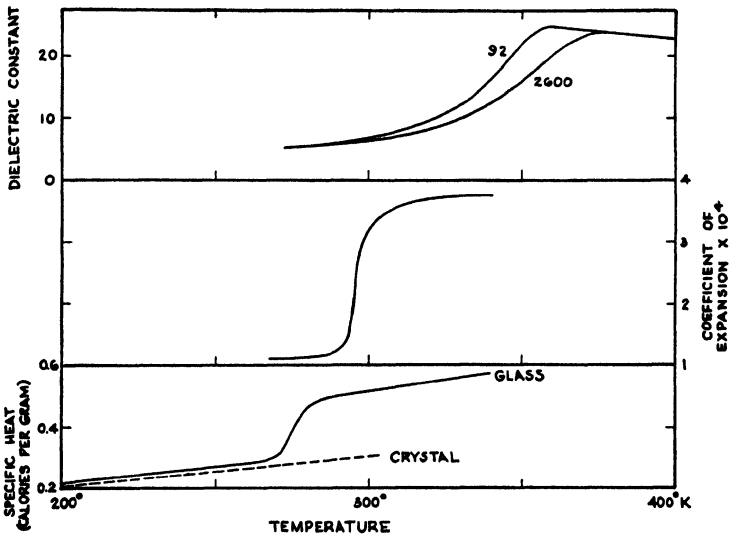
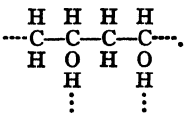


FIGURE 9. Variation with temperature of the specific heat at constant pressure,[81] coefficient of expansion,[81] and dielectric constant[89] (at 92 and 2600 kilocycles per second) of glucose. Data by Parks and co-workers.

The other example is polyvinyl alcohol, which also consists chiefly of long chains,



On warming and stretching, the chains straighten out to a zigzag form and line up so that the axes are parallel, with different chains, in pairs, bonded together by hydrogen bridges, as shown in Fig. 10. In the unstretched (vitreous) state, each chain is kinked and connected by hydrogen bridges to many others, rather than to one only.

Organic Polymers

Polyvinyl alcohol is but one of many organic polymers containing long chains which are kinked in a random manner—provided they have not been extended and oriented by stretching of the material. Such substances are rigid solids if the activation energies for rearrangement of the atomic positions are large relative to kT. They are soft and non-rigid if these activation energies are roughly equal to, or smaller than, kT. They show rubberlike elasticity if (1) most of the atoms have considerable freedom of motion, being restricted only by small energy humps, yet (2) long-distance displacements are prevented by the existence of a three-dimensional

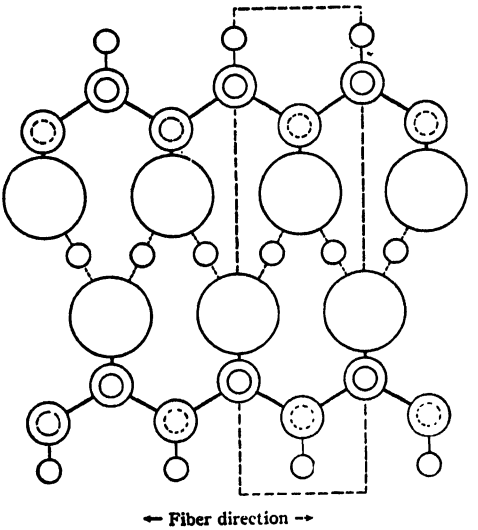


FIGURE 10. Projection of a pair of molecular chains in crystalline polyvinyl alcohol, according to Mooney.[29] The circles, in descending order of size, represent oxygen, carbon, and hydrogen. The dotted and solid circles distinguish between hydrogens on alternate sides of the chains.

network of strong bonds, capable of being broken only if large energy humps are passed over. Without such a network, these substances exhibit plastic flow, rather than long-range elasticity.

Many organic high polymers, although amorphous, are rigid, rather than elastic or plastic, at ordinary temperatures. Structural rearrangements are practically impossible because of the sizes of the energy humps which would have to be passed over. *Long-chain* polymers, as a rule, are *thermoplastic*—softening as the temperature rises, that is, as the thermal energy becomes great enough to pass over more and more of the energy humps. Many resins, however, contain three-dimensional networks of strong bonds, produced either by reactions together of small molecules—each becoming connected to more than two others, on the average—or by extensive crosslinking between already formed chain molecules. If these reactions, producing network formation, are induced by heat, the resin is said to be produced by a *thermosetting* process. A three-dimensional network polymer is elastic or rigid, depending on the tightness with which the atoms in the network are held in fixed positions.

Low-molecular Organic Compounds without Hydrogen Bridges

Most pure organic substances composed of small molecules crystallize on slow cooling. This is an indication that a single definite distribution of molecular centers, with a single definite orientation for each molecule, gives somewhat greater stability than any more random arrangement. If, however, there are no localized intermolecular forces favoring a particular orientation of each molecule relative to its neighbors—or if such orienting forces are very weak—the substance is likely to produce a vitreous mass on cooling. Vitreous formation is especially likely if the molecules are irregular in shape and if there are two or more *intra*molecular distributions of atoms having practically the same energy and practically the same *inter*molecular attractions.

To illustrate these statements, it may be mentioned that many unsymmetrical *iso*-paraffins, unlike the ***normal*** compounds and the symmetrical *iso* compounds, form

vitreous masses, rather than crystals, on cooling from the liquid state.[11, 16, 87] Glasses are also formed (at ∽ −160° C) by isoamyl and isobutyl bromides. The dielectric constants of these substances, as functions of temperature, frequency and previous treatment of the samples, have been the subject of an interesting study by Baker and Smyth.[1]

The presence of more than one molecular species in the samples being cooled—especially if the different species are very similar as regards their intermolecular attractions—also favors the formation of a vitreous, non-crystalline material. Mixtures of isoparaffins, for example, usually become vitreous on solidification.[37] As all organic chemists know, the crystallization of organic compounds from impure melts can often be accomplished only with great difficulty, if at all.

Summary

In this chapter the difficulty of defining the vitreous state in a rigorous manner, the existence of many borderline cases, and the theoretical requirements which must be met if a substance is to be vitreous have been discussed. A start has been made toward the development of a quantitative treatment of the subject. Certain typical classes of vitreous substances have been treated in some detail, with especial regard to the relationships between their structures and vitreous properties.

References

1. Baker, W. O., and Smyth, C. P., *J. Am. Chem. Soc.*, **61**, 2063 (1939).
2. Barth, T. F. W., *Am. J. Sci.*, **23**, 350 (1932); **24**, 97 (1932).
3. Barth, T. F. W., *Forts. Min. Krist. Petr.*, **70**, 25 (1932).
4. Barth, T. F. W., *J. Wash. Acad. Sci.*, **21**, 255 (1931).
5. Barth, T. F. W., and Lunde, G., *Z. physik. Chem.*, **122**, 293 (1926).
6. Bradley, A. J., *Phil. Mag.*, **48**, 477 (1924).
7. Bragg, W. H., and Gibbs, R. E., *Proc. Roy. Soc. London*, **A109**, 405 (1925).
8. Burwell, J. T., *Z. Krist.*, **A97**, 123 (1937).
9. Busse, W. F., *J. Phys. Chem.*, **36**, 2862 (1932).
10. Cohn, W. M., *Ber. deut. keram. Ges.*, **12**, 118 (1931) *Trans. Electrochem. Soc.*, **68**, 65 (1935).
11. Edgar, G., Calingaert, G., and Marker, R. E., *J. Am. Chem. Soc.*, **51**, 1483 (1929).
12. Gerding, H., *Naturwiss.*, **25**, 251 (1937).
13. Gibbs, R. E., *Proc. Roy. Soc. London*, **A107**, 561 (1925); **A110**, 443 (1926).
14. Gibbs, R. E., *Proc. Roy. Soc. London*, **A113**, 351 (1926).
15. Hagg, G., *J. Chem. Phys.*, **3**, 42 (1935).
16. Huffman, H. M., Parks, G. S., and Thomas, S. B., *J. Am. Chem. Soc.*, **52**, 3241 (1930).
17. Huggins, M. L., *J. Opt. Soc. Am.*, **30**, 420 (1940).
18. Huggins, M. L., *J. Opt. Soc. Am.*, **30**, 495 (1940).
19. Huggins, M. L., *J. Opt. Soc. Am.*, **30**, 514 (1940).
20. Huggins, M. L., *Phys. Rev.*, **19**, 363 (1922).
21. Huggins, M. L., and Sun, K.-H., *J. Am. Ceram. Soc.*, **26**, 4 (1943).
22. Huggins, M. L., Sun, H.-K., and Davis D. O., *J. Opt. Soc. Am.*, **32**, 635 (1942).
23. Karrer, E., *Pys. Rev.*, **39**, 857 (1932).
24. Klug, H. P., *Z. Krist.*, **A88**, 128 (1934).
25. LaMer, V. K., *J. Chem. Phys.*, **1**, 289 (1933).
26. Lewis, G. N., "Valence and the Structure of Atoms and Molecules," Reinhold Publ. Corp., New York, 1923; pp. 94-95.
27. Meyer, K. H., and Go, Y., *Helv. Chim. Acta*, **17**, 1081 (1934).
28. Meyer, K. H., von Susich, G., and Valkó, E., *Koll.-Z.*, **59**, 208 (1932).
29. Mooney, R. C. L., *J. Am. Chem. Soc.*, **63**, 2828 (1941).
30. Morey, G. W., "The Properties of Glass," Reinhold Publ. Corp., New York, (1938), pp. 74-75.
31. Parks, G. S., Huffman, H. M., and Cattoir, F. R., *J. Phys. Chem.*, **32**, 1366 (1928).
32. Pauling, L., *J. Am. Chem. Soc.*, **57**, 2680 (1935).
33. Pauling, L., *Phys. Rev.*, **36**, 430 (1930).
34. Randall, J. T., and Rooksby, H. P., *J. Soc. Glass. Tech.*, **17**, 287 (1933).
35. Seljakow, N., Strutinski, L., and Krasnikow, A., *Z. Physik*, **33**, 53 (1925).
36. Slattery, M. K., *Phys. Rev.*, **21**, 378 (1923); **25**, 333 (1925).
37. Smittenberg, J., Hoog, H., and Henkes, R. A., *J. Am. Chem. Soc.*, **60**, 17 (1938).
38. Smyth, C. P., and Hitchcock, C. S., *J. Am. Chem. Soc.*, **54**, 4631 (1932).
39. Thomas, S. B., *J. Phys. Chem.*, **35**, 2103 (1931).
40. Vegard, L., and Schjelderup, H., *Phys. Zeit.*, **18**, 93 (1917).
41. Warren, B. E., *J. Applied Phys.*, **13**, 602 (1942).
42. Warren, B. E., *Z. Krist.*, **A86**, 349 (1933).
43. Warren, B. E., and Biscoe, J., *J. Am. Ceram. Soc.*, **21**, 49 (1938).
44. Warren, B. E., and Burwell, J. T., *J. Chem. Phys.*, **3**, 6 (1935).
45. Wyckoff, R. W. G., *Am. J. Sci.*, **11**, 101 (1926); *Z. Krist.*, **A63**, 507 (1926).
46. Wyckoff, R. W. G., *Z. Krist*, **A62**, 189 (1925); *Am. J. Sci.*, **9**, 448 (1925).
47. Zachariasen, W. H., *J. Am. Chem. Soc.*, **54**, 3841 (1932).
See also "Glasses, Organic and Inorganic" by Harry Moore, *Chemistry and Industry*, **58**, 1027-1037 (1939). J. A.

Some Irreversible Effects of High Mechanical Stress *

P. W. BRIDGMAN

Research Laboratory of Physics, Harvard University

Some more or less arbitrary selection of the irreversible effects to be considered is necessary because of the vastness of the subject if all irreversible phenomena are admitted. Thus every case of rupture, or of plastic flow in a solid, or of flow of a viscous liquid is properly an example of an irreversible effect produced by mechanical stress. The overwhelming majority of the processes of daily life, which are sidestepped by thermodynamics on the plea that they are "irreversible," are of this kind. It is possible to distinguish various possible meanings for "irreversible." I shall use it with the connotation of thermodynamics as implying a dissipation of energy or increase of entropy, and rule out such purely mechanical cases of irreversibility as the winding of a watch with a ratchet. All that I can attempt in this enormous field is to select some of the more interesting and unusual cases, and the selection will usually be from instances which have occurred within my own experience in the laboratory. The "high" mechanical stress of the title is also in general fixed by my own laboratory experience, which means stresses of the order of thousands of kilograms per square centimeter.

Within this field it is possible to set up a formal scheme for systematic treatment. The mechanical stress may be distinguished as being either a pure hydrostatic pressure, or something more complicated, with a shearing component. The sorts of effect produced by the stress may be distinguished as consisting of: (1) rearrangements or distortions of the initial atomic arrangement which occur without change of type of space lattice, as in recrystallizations or various accommodation effects, or (2) a rearrangement of the lattice into a new type, thus constituting a change of phase, or finally (3) chemical changes, including both decompositions and syntheses. The combination of the two different stress factors with the three sorts of change gives a formal classification of six kinds of irreversible change to be considered.

EFFECTS OF HYDROSTATIC PRESSURE

Effects Not Involving Phase Change or Chemical Reaction

Under this heading come the ordinary mechanical effects. Every body, so far as recorded human experience goes, is completely elastic if the stress within it is a pure hydrostatic pressure; rupture or permanent change of dimensions never occurs. No one knows whether hydrostatic pressures of very much greater intensity than are available under terrestrial conditions, such as the pressures in the interiors of the stars, would squeeze matter permanently to the density of the companion of Sirius or not. Although cases have been observed in terrestrial laboratories of permanent increases of density under hydrostatic pressure, cavities have always been initially

* Many physico-chemical effects involve the close approach of particulate units. It is therefore of great interest to consider some of the results of high pressures, which have both theoretical and technological aspects. J. A.

present. Of course, if such cavities are present, the stress in the immediately neighboring walls is not a hydrostatic pressure, even if the whole body is immersed in a fluid under pressure. Our generalization thus stands even in such apparently exceptional cases. There are other situations in which the stress within the interior of a solid body is not a hydrostatic pressure even when the body as a whole is immersed in a fluid under pressure. The compressibility of non-cubic crystals is in general different in different directions, so that the geometrical shape of such a crystal is altered when the crystal is placed in a fluid under hydrostatic pressure. It follows that the stress in a haphazard aggregate of non-cubic crystals is not hydrostatic when immersed in a fluid carrying hydrostatic pressure, but is a complicated function of the relative orientation of the different crystal grains. If the hydrostatic pressure in the fluid is not too high, the adaptations demanded in the neighboring grains will be within the elastic limit, and the effects are reversible. When pressure rises too high, however, local elastic limits are exceeded, and the effect becomes irreversible. Any permanent change of dimensions produced in this way is, under the laboratory conditions of my own experience, in many cases too small to be important, although detectable, but there are other effects which may be larger. In particular, the electrical resistance of a polycrystalline wire of non-cubic crystals is often permanently altered after an application of hydrostatic pressure. After a number of applications of pressure, such a wire settles down to a steady state with no further alteration of zero resistance as long as the original pressure range is not exceeded. The more complex the structure, the more complex and unusual the effects. Such effects are particularly likely to be found in rocks or minerals. "Pipestone," or catlinite, is a very fine-grained and perfectly homogeneous-appearing mineral to the eye. But if its linear compressibility is measured,[1] it will be found that the change of length is not a linear or even single-valued function of pressure, but describes a loop, closed at both ends, which can be repeated indefinitely. Furthermore, this loop is not like the ordinary hysteresis loop, in which the strain lags behind the pressure: the strain is in advance—a highly paradoxical effect.

Most rocks are either initially porous, or else are made porous by hydrostatic pressure because of the differential compressibility of the grains. Dr. Birch [2] has encountered many examples of this sort of irreversibility in his measurements on rocks. In general, it is necessary to specify whether the measurements are made with the rock directly immersed in the pressure fluid, or "jacketed" to prevent entrance of the liquid into the pores.

If the material is not initially perfectly free from internal strain, in some cases the application of hydrostatic pressure will produce irreversible changes of geometrical form. This effect is particularly prominent in glass, and it is not unusual for a badly annealed glass vessel to be broken by the application of hydrostatic pressure. It is almost impossible to get accurate values for compressions at pressures of the order of several thousand kilograms in glass piezometers, because of the hysteresis of the glass. One case of disorganization of internal structure by pure hydrostatic pressure has been reported. Meyer, Hopf, and Mark [3] report that starch which has been exposed to 20,000 atmospheres gives a disordered x-ray pattern; it is natural to suspect the presence of a shearing component in the stress.

Effects Involving Phase Changes

We may distinguish less and more drastic effects produced by pressure. In the less drastic sort of change, the new phase which is induced to appear by the action of pressure is the thermodynamically stable phase, which might have appeared reversibly at a slightly different pressure or temperature; the action of pressure in inducing the appearance of the phase is simply that of an agitator, suppressing or diminishing the effect of internal friction. In the more drastic sort of change, pressure causes a new phase to appear, which is not formed reversibly from the initial

phase under any known conditions. The less drastic sort is by far the most common. Practically all liquids subcool to a certain extent before freezing, and practically all transitions between solids have to be carried an appreciable distance beyond the point of thermodynamic equilibrium before the reaction will start. Any phase change which thus runs displaced from the point of thermodynamic equilibrium runs irreversibly. This very common phenomenon may, however, be manifested in various ways.

To visualize the possibilities, consider the phase diagram of a substance, in which temperature and hydrostatic pressure are coördinates, and the regions of thermodynamic stability of the different phases are separated by the various transition lines on which the adjoining phases are in thermodynamic equilibrium. Consider now a phase initially at some temperature and pressure within its region of stability, and then carry it, by variation of temperature or pressure, across one of the bounding transition lines into the region of thermodynamic stability of another phase. The reaction from the phase which has now become unstable to the new phase involves two different sorts of phenomena: a nucleus of the new phase must be formed, and after the nucleus has been formed, the surface of separation between the two phases advances by a molecular rearrangement at the surface. Since one can arrive at a given point in the diagram either by a variation of pressure along the isothermal passing through the point or by a variation of temperature along the isobar passing through the point, it is arbitrary whether one chooses to describe the irreversible transition which is occurring at the point as brought about by a change of temperature or a change of pressure (or both).

Nucleus formation is the necessary precursor of the transition. The rapidity with which new nuclei are formed is a function of the coödinates of the point in the phase diagram. At first, as the point recedes from the transition line, the speed of formation of new nuclei will in general increase. A precise study of the variation of speed of nucleus formation as a function of the coördinates all over the pressure temperature-pressure plane is evidently necessary for a full understanding of the possibilities, but only very partial studies have been made. Tammann,[4] fifty years ago, studied the speed at atmospheric pressure as temperature is depressed. His important result was that in many cases the speed of nucleus formation passes through a maximum as temperature is lowered below the point of thermodynamic equilibrium, and beyond the maximum may drop to practically zero with further drop of temperature. This indicates what we are to expect in the full temperature-pressure field. As we travel away from the line of thermodynamic equilibrium, we are to expect that in general the speed of nucleus formation will at first increase, pass through a maximum, and then decrease. Incidentally, there is no reason to think that the speed becomes zero on the line of thermodynamic equilibrium, but the possibility should be recognized that there is a finite speed of formation of the nucleus of phase B within the region of thermodynamic stability of the phase A. As long as the rate of growth of the surface of separation of phase B into phase A is negative or zero in this region, there is no trouble. One of the reasons why we should expect a fall in speed of nucleus formation with increasing pressure is obviously the increased viscosity produced by increasing pressure; this for certain organic substances [5] may be as high as 10^7 fold for a pressure increase of 10,000 kg/cm^2. This point of view makes it doubtful whether a maximum is to be expected if the displacement from the transition line is in the direction of decreasing pressure; so far as I know the question has never been investigated by direct experimental attack, and there is no indirect evidence in all my experience that such a maximum ever occurs with decreasing pressure. There are, however, effects which become understandable if there is a maximum with increasing pressure. For example, the reversible melting pressure of ethyl acetate at room temperature is about 12,000 kg/cm^2. The solid is very reluctant to appear, however, and pressure may be

carried from 12,000 to 50,000 without freezing.[6] But on release of pressure, when the melting pressure is approached within approximately 10,000 kg/cm², nuclei of the solid are formed and the liquid freezes irreversibly. This freezing has a most paradoxical appearance, for it is a phase change with decreasing volume brought about by a decrease of pressure. Such a transition is thermodynamically impossible if it is reversible. Another, somewhat similar, example[6] is afforded by CBr_4, in which a transition runs against the direction possible for thermodynamically reversible transitions when the region of appreciable velocity of nucleus formation is entered from above. In the latter case, the phenomena have additional complications, for which the original paper must be consulted.

The speed of advance of the surface of one phase into the other is the second factor determining how a transition actually runs. For common substances this speed is zero at the thermodynamic equilibrium line, and increases at distances from the line. Whether it also passes through a maximum as pressure increases at constant temperature has not been established by direct means, but the presumption is that it does. The increase of viscosity brought about by pressure would lead one to expect it, and I have recently found[6] on measuring melting curves over a pressure range reaching up to 40,000 kg/cm² that the tendency is for freezing to become much more sluggish at the high pressure end of the curve. For instance, the solid and liquid forms of methyl cyclohexane were maintained in contact for a couple of hours at pressures varying from 12,000 to 50,000 in the region of stability of the solid, during which less than one-tenth of a volume of 0.2 cm³ froze. Under normal conditions the freezing is not notably sluggish.

There are substances, moreover, for which the speed of advance of the phase vanishes, at least to a different order of small quantities, within a finite region on both sides of the thermodynamic transition line, so that within this region the two phases may exist indefinitely in contact, in spite of the fact that thermodynamically they are out of equilibrium. A transition between two such phases must always run irreversibly. Many examples of such have been found.[7]

In view of all these general considerations, one would expect to find cases of substances which at atmospheric pressure are not far from a line of thermodynamic transition to another phase, but which can be forced to assume this other phase by the exercise of sufficient pressure to bring it into the region of finite velocity of nucleus formation, and which when then brought back to atmospheric pressure retain the new phase because the reverse transition velocity is too low. Such cases do exist, but they are not as common as perhaps might be expected. Thus selenium exists in both a crystalline form and a glassy modification. By application of 50,000 kg/cm² to the glassy modification some of it is permanently and irreversibly transformed, presumably to a crystalline form.[8] As another example, complicated and not yet elucidated internal changes are produced in arsenic[9] by hydrostatic pressures of the order of 10,000 kg/cm².

If one avails oneself of full coöperation between changes of temperature and of pressure one may find many examples of the realization of phases under atmospheric conditions which are normally unstable. Thus one may first raise the temperature to a region of appreciable transition velocity; then raise pressure until the domain of stability of a new phase has been entered and the new phase appears; then lower temperature under pressure so as to maintain the new phase; and then remove pressure at room temperature, where the transition velocity has become sensibly zero, so that the unstable phase is maintained. Any substance which has phases with finite reaction velocities at high temperatures and zero velocities at low temperatures should show this phenomenon; a number of examples will be found in my papers. One of the most spectacular is afforded by ice. It was first shown by Tammann[10] that the unstable ices II and III could be realized at atmospheric pressure at liquid air temperatures. The ices were formed by the action of pressure

at temperatures of perhaps −40° where the reaction velocity is considerable; and then the apparatus was cooled in liquid air, pressure released, and the apparatus opened. Dr. McFarlan [11] has made x-ray determinations of the structure of these ices by utilizing this effect, and has found that they are apparently indefinitely permanent at atmospheric pressure at liquid air temperature.

If the system is not chemically homogeneous, irreversible effects are easily produced by the application of hydrostatic pressure. Thus when an impure liquid is forced to freeze by the application of hydrostatic pressure, the impurities may separate on freezing and segregate in the bottom of the receptacle. When pressure is released, because of the finite time of solution and the time required for equilibrium to be set up by the action of convection currents, the system will not retrace its original configuration, so that the phenomenon is irreversible. Kerosene is such a substance.

Recently, increasing attention has been given to transitions other than the simple transitions of the "first kind" characterized by sharp discontinuities in volume and energy content. For example, the new types of transition occur in conjunction with the anomalies of the specific heat and thermal expansion of some substances at low temperatures. These transitions do not run with sharp discontinuities, and show hysteresis between increasing and decreasing temperature. Such hysteresis is always evidence of irreversibility. Since the transition temperature is in general a function of pressure, hysteresis loops should also be producible for such substances by a change of pressure at constant temperature. I have found [12] several examples of such loops at room temperature among the alloys; these loops are obtained when length is measured as a function of pressure. An especially large effect of this kind is shown by a non-metallic substance,[13] Ag_2O. Other physical properties similarly would be expected to show hysteresis, electrical resistance, in particular. Wilson [14] has found such electrical effects for a number of alloys which are known from other considerations to be capable of the order-disorder transition.

Hydrostatic pressure is known to affect the speed of certain irreversible reactions. Thus van Wert [15] found that the speed of segregation of the insoluble component of several age-hardening alloys is lowered by the application of pressure.

In all the examples hitherto considered, the phase which has been induced to appear irreversibly by the action of pressure is one which somewhere has a region of thermodynamic stability, so that the reaction may be made to run reversibly under proper conditions. But there are other phase changes produced by pressure which are not known ever to occur under reversible conditions. These are the more drastic reactions already referred to. Not many examples of these are as yet known. An example is HgS. This is known to the chemist in two forms: black and red. The black form is unstable and may be transformed into the red by an increase of temperature at atmospheric pressure. In nature the red form is common, but there is mention of a natural black form in textbooks of mineralogy. Professor Palache, however, states that he has never found any of these supposedly black forms in nature which did not turn out on examination to be really red; he believes that the black form never appears in nature. This is consistent with the observation [16] that the unstable black form made artificially in the laboratory is irreversibly transformed into the red by the application of hydrostatic pressure at room temperature.

Another example is black phosphorus. If ordinary yellow phosphorus is heated to temperatures above 200° and subjected to pressures of more than 12,000 kg/cm^2, it is irreversibly transformed [17] into a black form with large decrease of volume and drastic change in its physical properties; the new form superficially resembles graphite. The phenomena have recently been examined in much detail by Jacobs,[18] who finds that there are at least two forms of black phosphorus and that the relations are highly complex. I have found that black phosphorus may also be formed from yellow at room temperature [19] by pressures above 35,000 kg/cm^2.

Recently it has been found [20] that when CS_2 is subjected to pressures in excess of 40,000 kg/cm^2 and temperatures above 175° it is irreversibly and permanently transformed to a black solid, which must be a new form of CS_2. It cannot be a mechanical mixture of the constituent carbon and sulfur, because the density is 4 per cent greater than that of such a mixture, and it does not melt or soften at the melting point of sulfur, but decomposes slowly when the temperature is raised to the neighborhood of 200°.

There seems to be no theoretical method at present of anticipating the possibility of such substances as black phosphorus or solid carbon disulfide; if these had not been found experimentally after exposure to unusual conditions their existence would never have been suspected. It is fascinating to speculate whether it may not be possible that many other of the substances of ordinary experience can be pushed over a hill of potential into a new, permanent, and totally different form, if only sufficient pressure is applied.

Chemical Changes Produced by Hydrostatic Pressure

Among the complicated compounds of biology a comparatively large number are known which may be changed permanently by the application of hydrostatic pressure. Early in the century it was found at the West Virginia Agricultural Experiment Station [21] that microörganisms of various kinds are killed by the application of hydrostatic pressures of 5,000 or 6,000 kg/cm^2, and the application of such pressures was proposed as a commercial method for the sterilization of milk and canned products. Similar experiments were made later in my laboratory by Dr. Bovie and Stuart Ballantine, but were never published. I had previously found [22] that egg white and the proteins of meat are coagulated by pressures of the same order. This coagulation gives an explanation of the killing of microörganisms. Somewhat later Dr. Alice R. Davis made unpublished experiments in my laboratory on the coagulation of blood. Dr. Davies [23] published a thesis in the Department of Biology of Harvard University on experiments with my apparatus in which he established that the germination of certain seeds is much facilitated by an exposure to pressure. The optimum pressure was around 2,000 kg/cm^2; the seeds were killed by pressures much higher. This effect has received no sufficient explanation. It is not grossly mechanical, because no cracks can be discovered under the microscope in the coatings of the seeds. Germination is known to require some preliminary physicochemical change within the seed after ripening on the plant; this change may possibly be hastened by pressure, analogous to coagulation. I have found in unpublished experiments that the effect is not the same on all species of plants. Recently Dow [24] and his pupils have made similar experiments. Going a step beyond microörganisms, it has been found by Basset [25] and his collaborators that ferments, enzymes, and viruses may be killed by high pressures; the pressures required to kill these substances may rise as high as 15,000 or 20,000 kg/cm^2.

Among less complex organic compounds Conant and collaborators [26] have found a number of examples of polymerization. For example, isoprene is polymerized to rubber. This is probably not a straight pressure effect because of the smaller volume of the polymer. This might at first appear to be the natural explanation, because the volume of isoprene, for example, is under normal conditions materially larger than that of its polymer. The speed of the polymerization increases very rapidly with increasing pressure, however, whereas the volumes of the two substances become more nearly equal with increasing pressure. The highest rate of polymerization of all was found at 20,000 kg/cm^2, and at this pressure there is practically no change of volume when the reaction runs. Conant came to the conclusion that the largest part of the effect is the catalytic action of slight impurities, contact between catalyst and isoprene being furthered by the pressure.

Among the simpler inorganic compounds the examples of chemical changes

produced by pressure are less numerous. HgTe appears to be a case of a true pressure effect.[27] This compound is formed from the elements with considerable increase of volume, unlike the vast majority of compounds, so that decomposition to the elements would be facilitated by hydrostatic pressure. This does in fact slowly occur at pressures higher than 10,000 kg/cm^2. Another effect, not uncommon, is for the water of crystallization to be squeezed out of a salt by hydrostatic pressure.[28] Although this process may be thermodynamically reversible on a microscopic scale, as it actually occurs it is irreversible because of the mechanical segregation of the constituents. Similarly, certain double salts have been decomposed by pressures of 10,000 kg/cm^2 or more; $(NH_4)_2KPO_4$ is an example.[29] I have observed one example of an explosion produced by pure hydrostatic pressure.[30] A mixture of kerosene and $NaClO_3$ may be made to explode at room temperature by the application of somewhat over 2,000 kg/cm^2. It is curious that the chemically less stable $KClO_3$ cannot be made to explode by similar treatment.

There is an effect of pressure which is the reverse of decomposition. Many organic compounds decompose when temperature is raised. In a number of cases I have found that this decomposition is inhibited by the action of hydrostatic pressure. Thus CBr_4 has been heated to 175° without decomposition under pressure,[31] and urea to 200° at pressures of 10,000 kg/cm^2 or more.[32] There may be interesting possibilities here of bringing organic substances into reaction at temperatures where normally reaction would be impossible because of decomposition.

Effects of Stress with a Component of Shear

Effects Involving neither Phase Change nor Chemical Change

As already suggested, all cases of rupture properly belong here. There is space to mention only two rather unusual sorts of rupture which occur only at high stresses. Under normal conditions any piece of glass is covered with a surface layer of water many molecules deep so tightly sorbed as to be virtually in solution. If the glass is immersed in water and exposed to a pressure of several thousand kilograms, the water is driven into solution in the glass to a greater depth. When pressure is released the glass may be ruptured because of mechanical strains set up by the water trying to come out of solution.* Rupture may be avoided if pressure is released with extreme slowness. This effect was established by Poulter.[33]

The opinion has often been expressed that any two bodies ought to adhere if they are only pressed together hard enough. The idea, of course, is that if they are pushed so close together that the parts are brought within the range of molecular forces, there is nothing to distinguish the body from one initially in one piece, so that welding must occur. Ordinarily these simple considerations are complicated by the fact that any body in the air is covered by a film of adsorbed air molecules, so that molecular contact cannot be forced, and the expected welding normally does not take place. If, however, a body is fractured deep in its interior, so that an adsorbed gas layer never has a chance to form, and if the body is under a high mean hydrostatic pressure so as to push the fractured surfaces immediately into contact again, it is to be expected that any fracture will immediately heal itself.

I have arranged conditions of this sort in some recent experiments.[34] Small quantities of various materials have been subjected to mean pressures of 50,000 kg/cm^2, and while under this pressure have been subjected in addition to shearing stresses of sufficient magnitude to produce plastic flow. Many bodies will not permit indefinite plastic flow under such conditions, but when flow has proceeded to a critical extent, the substance breaks with a snap, and the shearing stress drops to

* According to Zsigmondy, the explosive eruption of Mt. Pelée (Martinique) involved a pressure-solution of water in volcanic rock or glass, and its fine pulverization and dispersion when the pressure was suddenly released. Bentonites seem to have been laid down similarly. J. A.

zero. Because of the prevailing hydrostatic pressure, the material immediately self-welds, and the whole process may be repeated indefinitely. A simple description of the phenomenon is that the body "lets go to get a fresh hold." It must be that fractures occur under such conditions deep in the interior of the earth; the application to geophysics and in particular to the explanation of deepseated earthquakes need not be elaborated.

Not only does self-welding occur under such conditions, but if one material is sheared across the surface of another under sufficient normal stress and for a sufficient distance to rub off any initially adhering film of gas, welding may be expected to occur. In my shearing experiments many examples of such welding have been found. Thus borax glass may be welded in the cold to steel just as tightly as if it had been fused.

An interesting example of the efficacy of the adsorbed surface film is afforded by the behavior of mercury and iron. Normally mercury does not wet iron, nor can it be made to amalgamate with it. If, however, a piece of iron is fractured under the surface of mercury, thus exposing a perfectly clean surface, the iron is completely wet and amalgamated. Amalgamation of iron may be produced by pressure. A hollow cylinder filled with mercury will support an internal pressure of only a few thousand atmospheres;[85] at higher pressures the mercury is driven by the pressure into the expanded pores of the iron, where it produces amalgamation because the iron is clean. Since the amalgamated region is mechanically much weaker than the untouched iron, the system is mechanically unstable, the amalgamated region rapidly eats its way through, and the cylinder bursts.

Recrystallization is a well known phenomenon brought about by shearing stress. The classic experiments of Carpenter and Elam[86] come to mind in which large single crystals of aluminum are grown in sheets that have been stretched beyond the elastic limit and then suitably annealed so that recrystallization takes place. A converse effect is the observation by Kapitza[87] that the success of producing single crystals of bismuth by solidification from the melt is very sensitive to slight mechanical stress. During the recrystallization that occurs when a polymorphic transition is brought about by a change of pressure, the growth of the new form may take place is a strongly preferred direction,[88] and may exert non-hydrostatic stresses sufficient to rupture heavy steel containers.

All sorts of minor rearrangements may be brought about by shearing stresses, ranging up to almost complete destruction of the lattice and amorphization, as when the surface film of a metal is heavily flowed by polishing. Elastic after-effects, and phenomena of creep, accommodation, and hysteresis are to be found here. Under conditions of high stress such effects may sometimes assume paradoxical forms. Thus I have found[89] that if a heavy-walled hollow cylinder is strongly collapsed by a high external pressure so that the internal bore is reduced to a fraction of its initial value while still retaining its circular figure, and if the external pressure is then released, such a complicated state of strain may be produced in the walls that the initial effect of a reapplication of pressure to the outside is to *increase* the internal bore. Or, it is not unusual for such heavily collapsed cylinders to fracture in extension across axial planes on *release* of external pressure.[40] Heavy glass cylinders exposed to external pressure may at first show no obvious effect on release of pressure, but that nevertheless some flow has been produced is shown by the fact that they may spontaneously fracture on standing for some hours.

It at first appears paradoxical that the density of metals is usually decreased by the heavy stresses that produce plastic flow; the reason is that the regular space lattice of the virgin material is a closer-packed arrangement than the more amorphous aggregate to which the metal is reduced by flow.

If plastic flow is sufficiently prolonged, a steady state must eventually be reached. This steady state probably consists of a highly distorted lattice, parts of which may

be so highly distorted as to be essentially a glass. A. B. Greninger at the Harvard Metallurgical Laboratory has made many x-ray analyses of metals sheared under high pressure; the structure in nearly every case is very much broken up. The lines are broadened, and the size of the elementary grains is very much reduced. One extreme case has been recently found,[41] sulfur, where there are no x-ray lines at all in the final product; but in general complete reduction to a glass does not occur. Microscopic examination by Larsen [42] seems to establish, however, that a portion of the distorted material at least sometimes is in the condition of a true glass. There is a converse effect; if a true glass is sheared at high stresses some lattice structure is produced. Thus when quartz glass is sheared, faint x-ray lines of crystalline quartz appear.

Geology is full of examples of recrystallization in orientations definitely related to the direction of flow. This effect has been studied by Sander.[43] It has been found possible to imitate a number of these orientation effects in the laboratory.[42]

Phase Changes Accompanying Shearing Stress

The effect of non-hydrostatic stresses on phase-changes has been very little investigated. Formally, the possibility must be recognized that a substance may exist in thermodynamic stability under high enough shearing stress in a form which may never be stable under any hydrostatic pressure whatever. The conditions of reversible transition under such conditions should be capable of formulation, but so far as I am aware no attempt has ever been made. Doubtless the reason is the absence of experimental evidence for the existence of such phenomena. There is one case which I have observed,[44] however, which may belong in this category; in my experiments with lithium under shearing stress at pressures of 50,000 kg/cm^2, I found a sharp break in the curve of a sort which in all other cases corresponds to a polymorphic transition which can be realized at some hydrostatic pressure. In the case of lithium, however, there is no such transition under any known hydrostatic pressure, even when pressure is extended to 100,000 kg/cm^2, as I have done recently.[45]

Shearing stress does exert a very great effect, however, on many ordinary polymorphic transitions. In general, the effect is that of an agitator or reducer of internal friction, so that the transition will run at pressures closer to the line of thermodynamic equilibrium, or at temperatures so low that ordinarily it would not run at all. Transitions may thus be made to run at room temperature by the exercise of a shearing force great enough to produce plastic flow, which under pure hydrostatic conditions have ceased to run perceptibly at temperatures 100° or perhaps 150° higher. Or, a transition may be made to occur under shearing stress that is not known to occur at all without it; thus *violet* phosphorus may be converted to black by high shearing stresses at room temperature combined with a high mean hydrostatic pressure.[46] This transition is, of course, irreversible, as are all the changes induced only by shearing stress, with the possible exception of lithium.

Wurtzite is completely converted to sphalerite by shearing under pressure.[47] Chalcocite is similarly changed to the ordinary cubic modification of Cu_2S. Complicated changes are produced among the various modifications of SiO_2, with evidence for a new crystalline form under some conditions.[42]

There are cases in which phase changes originally ascribed to hydrostatic pressure turn out to be actually produced by shearing stress. Thus Jeffries [48] has shown that many instances of permanent alteration of density ascribed by Kahlbaum to high hydrostatic pressure are to be explained by the plastic flow which was produced in the metal by the congealing of the oil with which pressure was transmitted.

Chemical Changes Brought About by Shearing Stress

Perhaps the first attempt at a systematic investigation of this subject was made in 1893 by Carey Lea,[49] who found many examples of chemical reactions under the

shearing stresses produced by vigorous rubbing of a pestle in a mortar. It is obviously difficult under such conditions to make any good estimate of the intensity of the stresses involved. This investigation of Lea attracted little attention, and apparently has been forgotten. I did not discover it myself until making a literature search in connection with my shearing experiments.[84] During these experiments I have observed many cases of irreversible chemical change produced by shearing stress; in fact, irreversible permanent changes are the only ones that can be established, there being no way of observing the product during the application of stress. Except for a specification of the magnitude of the stresses involved, however, these experiments of mine have been mostly qualitative in character. The reason is that the quantities of transformed material are so minute that in most cases an analysis of the product is not feasible. The reasons why the quantities are so minute are first, that the original quantities have to be small, and secondly that the altered material is in most cases confined to a thin surface layer, since the reaction is usually self-limiting. This self-limiting quality is explained by the fact that when a thin surface film of the new product is once formed, stress is thereby prevented from building up to the transformation stress in the rest of the original material. The fact of some sort of transformation can be established in many cases by alterations in the surface color. My detailed paper must be consulted for many special cases. Thus a brown discoloration frequently appears on shearing the compounds of cesium. There are, however, some cases in which the transformation occurs throughout the entire mass, or at any rate through a mass large enough to be susceptible to x-ray analysis. There are numerous examples of reduction of a salt to one of different valence, or of complete decomposition to the metal. Thus SnO_2 is reduced to SnO; Bi_2O_3 is reduced to metallic bismuth. Decomposition to the pure metal occurs especially among the salts of bismuth, mercury, and lead. Syntheses may be effected; thus copper and sulfur sheared together are converted to CuS. The thermite reaction is initiated by shearing under pressure, with explosive effects.

Transformations among the organic compounds are not as frequent as among the inorganic, perhaps because of a greater effect of increased viscosity and molecular interlocking. Anthracene,[50] however, affords an example of complete and irreversible conversion to some black substance.

In an address presented to the Congrès Internationale de Physique in Paris, 1900, Walthère Spring mentions the work of Carey Lea and much other early work (see "Oeuvres Completes de Walthère Spring," pp. 1746-1814, Brussels, 1923). Incidentally, in another paper (p. 1850) Spring points out that he had used the ultramicroscopic method prior to Siedentopf and Zsigmondy, and had sent one of them a reprint of his paper on request. Spring did much work showing that at high pressures chemical and physical unions take place.

The following item of interest appears in Technical Manual TM 9-2900 of the War Department, and is here given by permission (p. 118): "When loaded into commercial detonators, mercury fulminate is usually compressed at pressures of about 3,000 pounds per square inch. In this condition its explosive properties are not appreciably different from those of loosely pressed material. At greater densities obtained by higher pressures, there is a gradual reduction in sensitivity, until at such extreme pressures as 25,000 to 30,000 pounds per square inch fulminate entirely loses its property of detonating when ignited and will only burn. In this condition it is referred to as 'dead pressed.' If, however, such highly pressed fulminate is initiated by loose fulminate or other initial detonating agent, it will detonate at even higher rates than are obtainable at low densities." [Note added by the editor with the approval of the author.]

References

1. Bridgman, P. W., *Am. J. Sci.*, 7, 92 (1924).
2. Birch, F., and Bancroft, D., *Jour. Geol.*, 46, 59 and 113, (1938).

3. Meyer, Hopf, and Mark, *Berichte*, **62B**, 1103 (1929).
4. Tammann, G., "Kristallisieren und Schmelzen," p. 131 ff. Barth, Leipzig, (1903).
5. Bridgman, P. W., *Proc. Am. Acad.*, **61**, 57-99 (1926).
6. Bridgman, P. W., unpublished, probably to appear shortly in *Proc. Am. Acad.*
7. Bridgman, P. W., *Proc. Am. Acad.*, **52**, 57-88 (1916).
8. Bridgman, P. W., *Phys. Rev.*, **60**, 351 (1941).
9. Bridgman, P. W., *Proc. Am. Acad.*, **68**, 39 (1933).
10. Tammann, G., *Z. Anorg. Chem.*, **63**, 285 (1909).
11. McFarlan, R. L., *J. Chem. Phys.*, **4**, 253 (1936).
12. Bridgman, P. W., *Proc. Am. Acad.*, **68**, 27-93 (1933).
13. Bridgman, P. W., *Rec. Trav. Chim.*, **5**, 627-632 (1932).
14. Wilson, T. C., *Phys. Rev.*, **56**, 598 (1939); *J. Chem. Phys.*, **8**, 13 (1940).
15. Wert, L. R. van, Paper presented at the Sixteenth Annual Convention of the American Society for Metals, New York City, Oct. 1934.
16. Bridgman, P. W., *Proc. Am. Acad.*, **74**, 43 (1940).
17. Bridgman, P. W., *J. Am. Chem. Soc.*, **36**, 1344 (1914); **38**, 609 (1916).
18. Jacobs, R. B., *J. Chem. Phys.*, **5**, 945 (1937).
19. Bridgman, P. W., *Phys. Rev.*, **45**, 844 (1934).
20. Bridgman, P. W., *J. Applied Phys.*, **12**, 466 (1941).
21. Hite, B. H., Giddings, N. J., and Weakley, Jr., Chas. E., West Virginia University Agricultural Experiment Station, Bull. **146**, 1914.
22. Bridgman, P. W., *J. Biol. Chem.*, **18**, 511 (1914).
23. Davies, P. A., *Am. J. Bot.*, **15**, 149 (1928).
24. Dow, R. B., and Matthews, J. E., *Phil. Mag.*, **27**, 637 (1939).
25. Basset, J., et Macheboeuf, M. A., *Compt. Rend.*, **195**, 1431 (1932); Basset, J., Lisbonne, M., et Macheboeuf, M. A., *Compt. Rend.*, **196**, 1540 (1933).
26. Conant, J. B., and Tongberg, C. O., *J. Am. Chem. Soc.*, **52**, 1659 (1930); Conant, J. B., and Peterson, W. R., *ibid.*, **54**, 628 (1932).
27. Bridgman, P. W., *Proc. Am. Acad.*, **74**, 44 (1940).
28. Bridgman, P. W., *Proc. Am. Acad.*, **52**, 159 (1916).
29. Bridgman, P. W., *Ibid.*, **52**, 160 (1916).
30. Bridgman, P. W., *Proc. Am. Acad.*, **51**, 78 (1915).
31. Bridgman, P. W., *Proc. Am. Acad.*, **72**, 229 (1938).
32. Bridgman, P. W., *Proc. Am. Acad.*, **72**, 233 (1938).
33. Poulter, T. C., and Wilson, R. O., *Phys. Rev.*, **40**, 877 (1932).
34. Bridgman, P. W., *Proc. Am. Acad.*, **71**, 387 (1937).
35. Bridgman, P. W., *Proc. Am. Acad.*, **46**, 325 (1911).
36. Carpenter, H. C. H., and Elam, C. F., *Proc. Roy. Soc. London*, **107**, 171 (1925).
37. Kapitza, P., *Proc. Roy. Soc. London*, **119**, 358 (1928).
38. Bridgman, P. W., *Proc. Am. Acad.*, **52**, 106 (1916).
39. Bridgman, P. W., *Phys. Rev.*, **34**, 21 (1912).
40. Bridgman, P. W., *Phys. Rev.*, **34**, 16 (1912).
41. Bridgman, P. W., *J. Applied Phys.*, **12**, 467 (1941).
42. Larsen, Esper S., and Bridgman, P. W., *Am. J. Sci.*, **36**, 81 (1938).
43. Sander, B., *Z. Angew. Min.*, **1**, 20 (1939).
44. Bridgman, P. W., *Phys. Rev.*, **48**, 836 (1935).
45. Bridgman, P. W., *Phys. Rev.*, **60**, 351 (1941).
46. Bridgman, P. W., *Phys. Rev.*, **48**, 828 (1935).
47. Bridgman, P. W., *Am. J. Sci.*, **237**, 18 (1939).
48. Jeffries, Z., *J. Inst. Met.*, **18**, 243 (1917).
49. Lea, C., *Am. J. Sci.*, **46**, 241, 413 (1893).
50. Bridgman, P. W., *Proc. Am. Acad.*, **72**, 265 (1938).

Sonic and Ultrasonic Waves in Colloid Chemistry

KARL SOLLNER
University of Minnesota, Minneapolis, Minnesota

Introduction

High-frequency inaudible sound waves, "ultrasonics," of great intensity are able to produce a great diversity of phenomena pertaining to the realm of colloid chemistry, as first shown in 1927 by R. W. Wood and A. L. Loomis.[256] Later it was found that substantially the same phenomena may be likewise obtained with low-frequency sound waves, provided their intensity is sufficiently high.

It must be emphasized that there is no fundamental physical difference between ultrasonics and audible sound. The distinction is based on a virtually accidental physiological difference—the ability of the human ear to respond to sound frequencies between about 16 cycles per second and about 15,000 to 20,000 cycles per second. Higher and lower frequencies are inaudible. Sound frequencies above the audible range are commonly referred to as "ultrasonics" or "suprasonics."

Sound waves of high intensity may bring about all kinds of phenomena of dis-

persion, such as peptization, the mutual emulsification of two bulk liquid phases, the disintegration of certain solids, the depolymerization of high-molecular substances, the formation of fogs, etc. On the other hand they may also cause accumulation, aggregation and coagulation in liquid as well as in gaseous systems, and the orientation of anisometric particles, etc.

High-intensity sound waves may cause some additional phenomena of interest to the colloid chemist; they favor the formation of new phases, *i.e.*, start and accelerate the crystallization from many supersaturated solutions and affect the crystalline structure of metals and alloys when applied to the melts during the period of cooling and crystallization.

Though there are no fundamental physical differences between sound waves of low, medium and high frequencies, great differences exist in the ease with which high intensities may be produced and the ease with which the sound energy may be transferred to the system under investigation. For most colloid work sound intensities are used which are much greater than those to which we are accustomed. For example, to produce emulsification an intensity of 10 watt/cm^2 is necessary, compared to the 10^{-9} watt/cm^2 of a comfortably tuned radio.

Originally all the work in colloids was done with ultrasonics of several hundred thousand cycles per second; lately frequencies down into the audible range have been more and more used.

The fact that the wave length decreases as the frequency increases makes itself felt in all experimental work; with higher frequencies the experimental system may easily be made many times larger than the wave length. Another important fact is that the absorption of high-frequency waves is much stronger than that of lower frequencies. (See p. 355). At very high frequencies the energy losses due to absorption become excessive. This seems to be one of the reasons why hardly any colloid-chemical work has been done with frequencies of more than a million cycles per second. On the side of the lower frequencies another difficulty arises. The amplitudes necessary to transmit a given quantity of energy at low frequency are much greater than those at higher frequencies. At medium and lower audible sound frequencies the necessary amplitudes are actually so large that the mechanical strength of the vibrators is the factor limiting the sound transmission; unless the amplitudes stay below a certain limit, the vibrators are destroyed.

To obtain an idea of the wave length, λ, of sound waves of different frequencies, n, we must remember the simple equation $V = n\lambda$, V being the sound velocity. Table 1 is calculated on the basis of this equation. Readers interested in accurate information on this and other points concerning the pure physics of ultrasonics must be referred to the literature, *e.g.*, the monograph of Bergmann,[6, 7] and particularly to the excellent book of Hiedemann.[106]

Table 1. Wave Length at Different Sound Frequencies in Air, Liquids and Solids

Frequency n (cycles per second)	Wave Length (λ)		
	Air (V = 330m/sec)	Average Liquid (V = 1200 m/sec)	Average Solid (V = 4000 m/sec)
16 (audible)	20.6 m	75 m	250 m
435 (")	76.0 cm	2.75 m	9.2 m
15,000 (")	2.2 cm	8.0 cm	26.6 cm
25,000 (ultrasonic)	1.3 cm	4.3 cm	16.0 cm
200,000 (")	1.6 mm	6.0 mm	2.0 cm
1,000,000 (")	0.3 mm	1.2 mm	4 mm

Sonic and Ultrasonic Generators

The literature on sound generators useful for colloid work has never been discussed critically, or even collected in one place. However, the success of further

work in this field will necessarily depend upon the correct choice of experimental equipment.

It is beyond the scope of this article to set forth in detail the physical principles on which the constructions of the different sound-generating devices are based. However for the foregoing reasons it seems imperative to discuss critically those types of sonic and ultrasonic generators which have been used successfully for colloid work. It may be added that the patent literature contains a great number of grossly exaggerated and misleading claims; it is therefore disregarded in this review. Four main groups of sound generators will be discussed, and their usefulness for special purposes will be indicated.

(a) **Piezoelectric Sound Generators**, advantageously used in the frequency ranges from 100,000 to 1,000,000 cycles a second. The upper limit is determined by the usefulness for colloid-chemical work and not by the possible range of such generators (which may be built for frequencies up to 10^9 cycles per second).

(b) **Magnetostrictive Sound Generators**, successfully used in the frequency range between several thousand and about 50,000 cycles a second, *i.e.*, between higher audible sound and lower ultrasonic frequencies.

(c) **Electromagnetic Sound Generators**, useful mainly in the range of lower and medium audible sound frequencies.

(d) **The Gas Current Vibration Generator**, a device useful only in gaseous systems. Its range extends from very low (infra-acoustic) up to high ultra-acoustic frequencies.

A discussion of the engineering aspects involved in the construction of such generators must here be omitted completely. Readers interested in constructing such generators will find below some of the pertinent references. Some generators are now commercially available, but the author has no personal experience with them.

Piezoelectric Generators. The piezoelectric sound generator is a device for transforming high-frequency electric oscillations into mechanical oscillations by making use of the piezoelectric effect. Piezoelectric generators were the first used for colloid work and are still most commonly used.

If pressure is applied to a crystal belonging to a crystal system without center of symmetry, certain crystal faces become electrically charged; conversely, if a potential difference is applied to certain faces of the crystal, it contracts or expands according to the direction of the applied voltage. These phenomena are called piezoelectric effects. For technical and economic reasons, only discs and plates cut in a suitable manner from large quartz crystals have been used as piezoelectric oscillators for work on colloids.

If an alternating electrical field is applied to such a quartz plate it contracts and expands periodically—it vibrates. The frequency of these vibrations is, of course, that of the electric field applied. At a given field strength the vibrations become much stronger if the natural mechanical frequency of the disc coincides with the frequency of the applied electric field, *i.e.*, if the quartz vibrates in resonance. Under these conditions electrical energy can be transformed efficiently into mechanical oscillations. As the frequency of the natural mechanical oscillations of a solid body depends upon its shape and size, it is easily understood that the dimensions of the quartz disc determine the frequency of the electric field which should be applied to achieve maximum effects through resonance. Harmonic oscillations of higher frequency can be stimulated; they are, however, not nearly as powerful as those at the natural fundamental frequency. Circular quartz crystal discs 60 to 80 mm in diameter have mostly been used. Their thickness was usually between 7 and 14 mm, corresponding to frequencies from 150,000 to 300,000 cycles per second. Unfortunately, large quartz discs are rather difficult to obtain and are expensive. For detailed information on the physics and technology of piezo-quartzes the reader may consult the monographs of Scheibe,[230] and Vigoureux.[258]

If an oscillator vibrates in air the energy transmission from solid to gas is very poor. Since one cannot cause a quartz oscillator to vibrate really strongly—for it would crack under the stress associated with greater amplitudes—the use of the other devices, listed below, for work on gaseous systems is indicated. If, however, a quartz crystal vibrates in a liquid, *i.e.*, in a medium of a specific gravity and sound velocity similar to the quartz oscillator itself, then the energy output may become rather high even with small amplitudes. For most work on colloids one makes use of this and immerses the quartz oscillators several centimeters deep in a bath of transformer oil. Vibrations transmitted into the oil are easily conducted to any other liquid or solid system dipped into the vibrating oil. The use of oil has the additional advantage that rather high voltages can be applied to the quartz oscillator, without the danger of sparking, etc. If properly handled, oscillators seem to last indefinitely, provided they were carefully selected (free from twinning, impurities, etc.). The author and his collaborators have used the same quartz plate without serious difficulties over a period of several years for several thousand hours.

The high-frequency electrical oscillations to activate the crystal are generated by a suitable electronic oscillator with an energy output of several hundred watts or more. The sender is coupled to the quartz plate by means of a Tesla coil; the plate is fixed between suitable electrodes in a bath of transformer oil (Fig. 1).

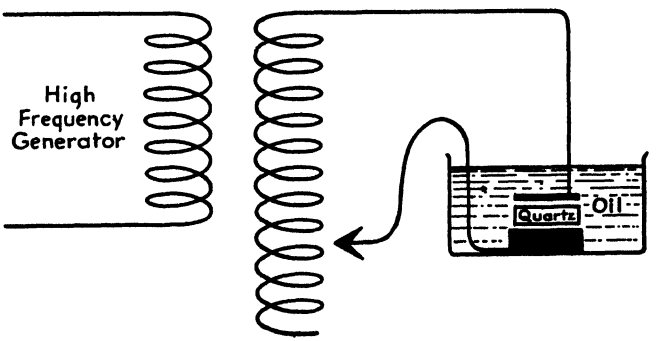


FIGURE 1. Piezoelectric sound generator.

Descriptions of 2-KW transmitters have been given by Wood and Loomis;[256] Freundlich, Rogowski and Sollner;[76] and others. (See, *e.g.*,[162] compare, *e.g.*,[6, 7, 50, 91, 106, 191])

The secondary high-frequency voltage which is applied to the quartz oscillator is varied according to the desired effect. For fast emulsification Freundlich, Rogowski and Sollner [76] applied 30,000 to 40,000 volts. The energy consumed under these conditions by the quartz oscillator, the electrode system, and the oil bath—as measured calorimetrically—was 250-300 watts, for stronger effects up to 500 watts. How big a fraction of the calorimetrically measured energy is available as sound energy cannot be determined accurately; it depends on the frequency, the electrode system, etc. The difficulty of these measurements (see, *e.g.*,[185]) is one among several reasons why hardly any information exists concerning the influence of the frequency under comparable energy conditions.[57, 71]

Small ultrasonic quartz generators for colloid work have been described in several instances. (See, *e.g.*,[179]). The writer considers them to be unsatisfactory for investigative or preparative purposes, for the energy output is so small that their usefulness is much limited. The time required to obtain the effects which may be had within seconds with more powerful generators is inordinately long. It also seems doubtful whether such effects as the formation of fogs, or the dispersion of solids, can be achieved with such outfits at all. However, it seems probable that

fairly small high-frequency generators may become really useful if used with more efficient quartz vibrators than those used so far for colloid work. (See below). With the conventional quartz crystals and quartz mountings, generators with less than 300 watts high-frequency output and quartzes of less than 60 mm diameter are not recommended.

The mounting of quartz plates for colloid work was described by Wood and Loomis;[256] in detail by Freundlich, Rogowski and Sollner,[76] Claus,[53, 57] and others.[4, 6, 7, 106] One such electrode system is shown in Fig. 2. The quartz crystal is

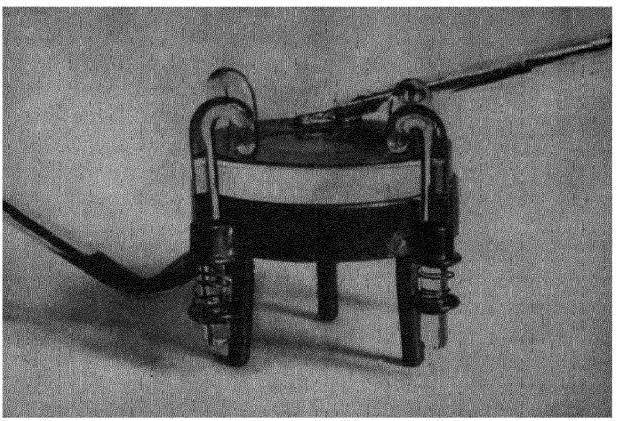


FIGURE 2. Quartz oscillator and electrode system, about one-half actual size.

best laid upon a thick piece of lead, the surface of which is ground plane; this serves as the lower electrode. The upper electrode is formed by a brass ring, the best dimensions of which are determined empirically. It is held in position by glass hooks provided with springs. The electrodes are connected to the secondary side of the Tesla coil. Some investigators prefer the use of quartz vibrators the circular faces of which are metallized, mounted in similar electrode systems.

If a crystal, radiating upward, vibrates at low energies in the oil bath, a slight tremor of the surface of the oil is seen; at greater energies a mound of oil is raised above the surface over the crystal; its height varies from bare visibility to several centimeters. At still higher energies (those useful for fast emulsification) the oil mound breaks up forming a fountain of oil 6, 8, or more centimeters high. This oil fountain is due less to "radiation pressure," than to a peculiar pumping action of the vibrating oscillators.[143]

It is practically never possible to perform experiments on colloids directly in the oil bath, chiefly because of the danger of sparking, which could easily damage the crystal. One usually dips a test tube containing the system under investigation into the oil fountain, and thus exposes it to the action of the sound oscillations. The containers used for this purpose should be of strong and fairly thin glass. The wall thickness is not nearly as critical a factor as one might expect *a priori*, for the sound waves are not transmitted as longitudinal waves through the glass (if this were the case its thickness would be very critical), but transverse vibrations set up in the walls of the container. These in turn are transmitted to the liquid inside.

Gruetzmacher [84] and others have constructed electrode systems in which only one of the circular faces of the quartz vibrator is in contact with a medium of very low acoustic impedance (air). This results in an almost complete reflection of the sound waves at the air side of the quartz; they are in phase with those radiated from the other side of the vibrator and the useful energy ouput is thus almost doubled.

Gruetzmacher [88] also describes a piezoelectric quartz, shaped like a spherical concave lens, which brings the ultrasonic waves proceeding from the vibrating quartz surface to a focus. Gruetzmacher was able in this manner to concentrate the energy at the focus to more than 100 times the value near the vibrating quartz itself. So far use has not been made of such improvements for work on colloids. Upon their successful application will depend how far piezoelectric sound generators will ultimately be used for laboratory or industrial work.

Magnetostriction Generators. The magnetostriction generator is a device to transform alternating electric currents into sound vibrations, making use of the magnetostriction effect as first described by Pierce.[170, 171, 172, 178]

If a rod or tube of ferromagnetic material is brought into a magnetic field parallel to its axis, its length is changed slightly. This change of length is independent of the sign of the field, and may be either a decrease or increase. This depends on the nature of the material, its previous treatment, the degree to which it was previously magnetized, the applied field strength, etc. This effect is known as "magnetostriction." The changes of length so produced are of the order of magnitude of a few parts in a hundred thousand with easily produced field strengths. Besides pure nickel, various nickel alloys (Invar, Monel metal, etc.) and occasionally cobalt are used.

The basic principle of all magnetostriction oscillators is this: if a nickel rod, for example, is brought into an alternating magnetic field it is shortened periodically by magnetization. It is easy to see that if the rod is not previously magnetized it will vibrate with double the frequency of the alternating field. If, however, it is suitably premagnetized, the mechanical change in length will be in step with the alternating frequency. If there is resonance between the natural elastic period of the nickel rod and the frequency of oscillations caused by the magnetic field, the amplitude of the oscillations will be at a maximum. The alternating magnetic field used in the magnetostriction sound generators is, of course, produced by an alternating electric current of the same frequency.

The fundamental frequency, ν, of a rod is given by

$$\nu = \frac{V}{2L}$$

V being the velocity of sound in the rod, and L its length. For colloid work harmonics are rarely used. With a sound velocity in nickel of about 5,000 m/sec, the length of the rod or tube is about 50 cm for a frequency of about 5,000 cycles per second, 12.5 cm for about 20,000 cycles a second, and so forth.

To permit such a rod to vibrate freely at its fundamental frequency it is held at its middle, which leaves the ends free to vibrate. The exciting magnetic field, or in the case of the more frequently used generator with pregmagnetization, the two magnetic fields are applied by the use of suitable coils, or magnetic yokes, or a combination of the two.

For practical purposes, instead of nickel rods, nickel tubes are used which are closed at the upper end and cooled from below with a stream of water. They are supported in the middle by a rubber ring carrying the reaction vessel. Many investigators split the lower half of the nickel tube to reduce eddy currents. Suitable arrangements have been described by Gaines;[77] Chambers and Gaines;[46] Chambers and Flosdorf;[42] Freundlich and Gillings;[71] Clair;[48, 49] and many others.[190] Fig. 3 shows the construction of a laboratory model described by Chambers and Flosdorf.

The magnetostriction sound generators are used not only in liquid but also in gaseous systems. In the latter case the vibrating nickel tube carries a heavy, piston-like end-plate of large diameter to facilitate the transmission of energy into the gaseous phase.

The relatively great length of the sound waves produced by the magnetostriction sound generators is one of their main disadvantages for work in liquid systems, as it is of the same order of magnitude as the dimensions of convenient containers. Consequently the action of the sound waves, which is strongest in the nodes and loops of stationary waves (see below), is more or less confined to a few planes or points many centimeters apart; such points may not properly occur at all in small systems.[71]

An advantage of the magnetostriction oscillators is their relatively simple and cheap construction and the absence of the costly quartz crystals. They have the

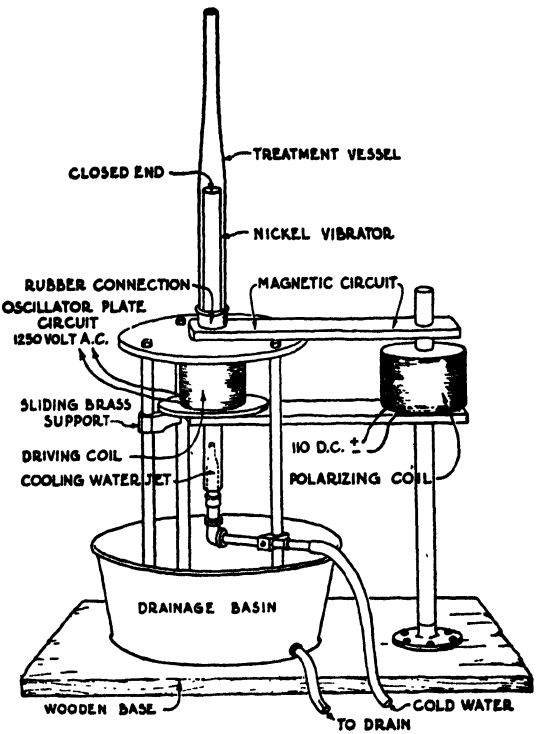


FIGURE 3. Magnetostriction sound generator.

additional advantage that the oscillator itself—the vibrating nickel tube—may in most cases be brought in direct contact with the substances to be treated without the interposition of another medium. Thus the unavoidable but great losses of vibrational energy suffered by dissipation and reflection in the oil bath of the piezoelectric sound generators is avoided here.

It seems likely that magnetostriction generators will at least partially displace piezoelectric oscillators for colloid work. Their efficiency can be increased greatly by the use of some sound concentrating devices, such as described by Chambers and Gaines;[46] one also could try the use of hollow-ground end-plates of the nickel vibrator similar to Gruetzmacher's[88] quartz crystal lenses.

Electromagnetic Generators. The electromagnetic sound generators have so far been used in only a few cases for colloid work. Since their frequency is entirely in the audible range, rather large amplitudes are necessary to produce the same quantity of sound energy. Their effectiveness reported for liquid systems, at least in some cases, is not due to typical sound action.[71]

Chambers[87] describes two types of oscillators, one of conventional electromagnetic design, and one of the Fessenden type which he used for the treatment of milk.

(See also [226]). Their description by Chambers [37] is quoted here. "1. Electromagnetic oscillators available were similar to those used in submarine communication and echo depth sounding consisting essentially of a heavy, loaded, steel membrane actuated by alternation in an electromagnetic circuit." (See Fig. 4a) "Three oscillators of this type were available with resonant frequencies of approximately 1200, 2160, and 3000 cycles per second respectively. Power was supplied from appropriate motor driven generators at one-half the oscillator frequency since the diaphragms were non-polarized. Input of electrical energy in each case was from 400 to 600 watts, depending somewhat on the loading conditions. Actual output of acoustic

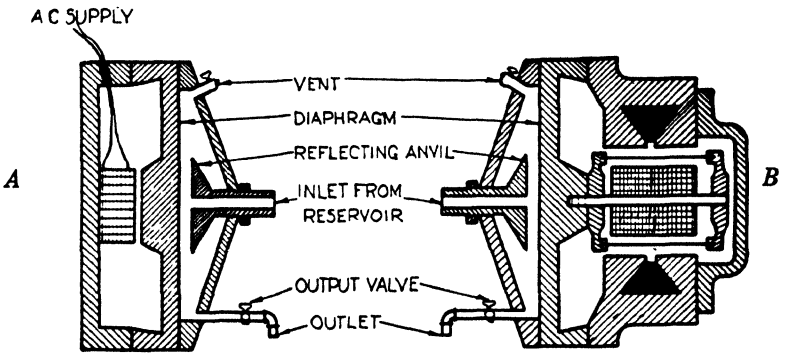


FIGURE 4. Two types of electromagnetic sound generators.

energy at 1200 cycles was approximately 175 watts while the electrical efficiency of the other oscillators was considerably lower. In fact the acoustic output of the 3000 cycle apparatus was so low that no satisfactory tests could be made at that frequency.

"2. Oscillators of the Fessenden type were also used, in which the vibrating membrane is fixed to a copper tube, the whole being polarized by an inciting current of about 2.75 amperes. Movement of the diaphragm depends on reversal of eddy currents in the copper sleeve induced by current oscillations in the activating coil." (Figure 4b). "Consequently the frequency of vibration is equal to the frequency of the activating current and not twice that value as in the electromagnetic oscillators. Two oscillators of the Fessenden type were used, one with a resonant period of about 1090 cycles, the other with resonant periods at 360 and about 610 cycles. Power input was variable up to 2200 watts. The electrical efficiency was about 45 per cent except at the 610 cycles resonance point where the power ratio was somewhat lower. . . . Accessory equipment . . . consisted of a heavy bell metal cover, tin-dipped, bolted to the margins of the steel diaphragm. Two marginal openings were provided, one for the ingress or egress" of the liquid to be irradiated and the other opening to "enable elimination of an air pocket from the chamber formed between cover and diaphragm. In addition a central opening was provided as an outlet or inlet for" the liquid. ". . . This opening passed centrally through a stainless steel 'anvil' or reflecting plate capable of micrometer adjustment in such a way that the layer of 'liquid' between anvil and diaphragm could be reduced to any desired value up to one-half inch in steps of .001 inch. . . ."

Clair [48, 49] describes briefly an electromagnetic generator, developed for a study of the coagulation of smoke; it is claimed to be very efficient. So far Clair's generator has only been tested in gaseous, not in liquid systems. Clair's [49] description follows: "The vibrating element consists of a solid cylinder of Duralumin supported by an annular web at its midsection and free to vibrate longitudinally as a free-free bar. On the upper end is a driving ring which projects into the radial gap of a pot magnet. It is excited by a high frequency current which flows through a coil wound

around the center pole of the magnet adjacent to the driving ring. This current induces in the driving ring a much greater current which interacts with the magnetic field to cause an alternating force to be applied to the vibrator, just as the current in the voice coil drives the diaphragm of a dynamic loudspeaker. The vibrator is made self-exciting by means of a built-in electrostatic pick-up unit which causes the vibrator to initiate the electrical oscillations which drive it." ". . . The vibrator executes simple harmonic motion even at the maximum amplitude. . . ." ". . . Sound energy is radiated from the lower face of the vibrator into the cylindrical chamber" containing the gaseous system to be irradiated.

Freundlich and Gillings [71] tried with little success the vibrating mechanism of an automobile horn with a maximum input of 100 watts.

Since only a few investigators have used electromagnetic sound generators it is difficult at present to appraise their merits justly. One has the impression that they are very promising in the treatment of gaseous systems. The effects obtainable in liquid systems seem to be moderate at best. One is not sure that the effects are entirely due to typical sound actions, and not, at least partially, the result of immediate mechanical action. The amplitudes which are necessary at low frequencies in order to produce an intense sound field are so large that various types of effects due to gross stirring may occur, as indicated by the results of Freundlich and Gillings.[71] If it should be found possible to operate electromagnetic sound generators really efficiently in liquid systems, it seems likely that this method could find wide industrial application.

Gas-current Vibration Generators. For attaining considerably greater energies of vibration in air than those obtainable with conventional or Galton whistles, Hartmann [85, 86, 87, 88, 89] constructed the gas-current vibration generator. It is based on an effect described first by Mach and Salcher.[131] If a current of air is allowed to issue from a nozzle at a speed greater than sound velocity, *i.e.*, under a pressure of 0.9 atm or more, a periodic structure is formed in the air stream. The pressure, *P*, varies periodically at different distances from the mouth of the nozzle, *D*, as indi-

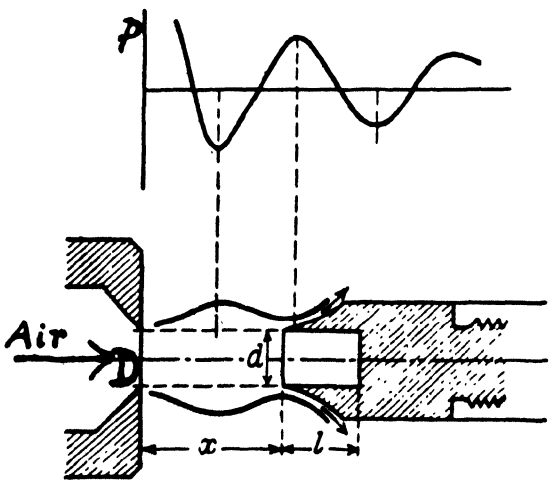


FIGURE 5. Gas-current sound generator.

cated in the upper half of Fig. 5. The regions of rise of pressure are intervals of instability in the jet. They make possible the production of sound waves, when a hollow body, serving as an oscillator, is brought into these regions of instability. (Fig. 5, lower half). The hollow oscillator is periodically filled with an overpressure of air and in the intervals this air is forced out again, whereby an oscillation is pro-

duced. The frequency of these sound waves is determined by the dimensions l and d of the hollow oscillator, and its distance x from the nozzle D. For example, if the diameter of D is the same, d, as that of oscillator, and if $d = l = 4$ mm, the frequency is 11,000 to 17,000 cycles per second, according to its distance x from D. (See Fig. 5). The frequencies obtainable with Hartmann's sound generator range from the infra-acoustic range, through the whole acoustic range, up into high (500,000 cycles per second) ultrasonic frequencies. Hartmann [89] estimates that 5 per cent of the power input is transformed by this gas-current vibrator into sound energy. Sound outputs of 100 watts and more have been claimed for the apparatus.

On account of its simplicity and the lack of expensive auxiliary equipment—all that is necessary is a source of compressed air—the gas-current vibrator generator seems to be destined to play an increasingly important role in colloid work in gaseous systems, particularly for industrial purposes. Certain difficulties have been pointed out by Clair.[49]

The Main Effects Caused by Sound Waves in Physicochemical Systems and the Mechanisms which Bring Them About

We have mentioned before that the most characteristic actions of intense sound waves are the disintegrating and ponderomotive effects: emulsification, formation of fogs, dispersion of solids, liquefaction of gels, etc. on the one side, and accumulation, coagulation, and orientation on the other. Besides these main phenomena there are several more effects which must be discussed, at least briefly: the degassing of liquids, thermal effects, the acceleration of crystallization, and finally the influence of sound waves of high intensity upon chemical reactions. Most of the phenomena mentioned were first observed by Wood and Loomis [256] and described in their pioneering paper.

Emulsification and Emulsions. If a test tube containing water and mercury, or water and an immiscible organic liquid, is dipped into the oil fountain of a piezoelectric sound generator, or if the same two phase systems come in contact with the vibrating nickel rod of a magnetostrictive sound generator, emulsification occurs at once. Grey clouds of very fine mercury droplets are thrown into the water from the water-mercury interface; in oil-water systems white clouds of dispersed water or organic liquid, respectively, are produced where the two liquids meet; soon more or less concentrated emulsions are obtained. In non-protected systems 6 g of mercury per liter and 50 to 60 g of benzene or similar substances may be dispersed in one or two minutes, and these concentrations do not increase on further irradiation. In the presence of suitable emulsifiers very high concentrations can be obtained.

In 1929 Richards [186] emphasized the differences in the nature of the emulsification of mercury and in the dispersion of organic liquids in water. For liquid and molten metals the mechanism of dispersion in water or organic liquids, according to Richards [186 see also 217] is as follows: when the vessel containing the metal and the water is brought into the oil fountain, violent transverse vibrations are set up in the walls of the vessel; these vibrations pump small quantities of water into the liquid metal; the water droplets rising in its interior reach the interface metal-water covered with a thin film of metal. When this film breaks, a cloud of minute metal droplets is thrown into the water. This process, as described for the macro-interface metal-water, obviously occurs also between the water droplets in the interior of the liquid metal, thus accentuating the effect. Sollner [215] irradiated a low-melting alloy-water system and cooled it during the irradiation. The metal solidified under these conditions is sponge-like, as is apparent from Fig. 6, showing its cross-section and its surface, which is covered with little blisters. For a reason which will be understood later, it must be stressed that this dispersion of metals occurs equally well *in vacuo* or under high external pressure.[217] The mechanism of the formation of Hg-sols by sound waves is basically the same as by normal mechanical dispersion,

which was investigated by Nordlund;[154] Kremnew[119] has elaborated on this point.[see also 192]

Now let us turn our attention to the mechanism of emulsification in oil-water systems.[216] The facts concerning emulsification in non-metallic systems, as described first by Newton Harvey,[91] are as follows. Emulsification occurs neither *in vacuo* nor when sufficiently high outside pressure is applied, the liquids in the latter case being gas-free or saturated with gas at a lower pressure only. This rule applies equally well in the presence of even the best emulsifiers, where mere shaking



FIGURE 6. Emulsification of a metal by ultrasonics. A low-melting alloy solidified during irradiation. About twice actual size.

may easily produce stable and fine-grained emulsions.[216] When the system is in equilibrium with a gas phase, emulsification always occurs on irradiation, provided this pressure is neither too high nor too low. The lower limit is much more sensitive; in toluene-water systems, for example, an absolute gas pressure of at least 100 mm of mercury is necessary. How striking the effect of the gas pressure is may be seen in Fig. 7; the test tube at the left contains air at atmospheric pressure, the test tube at the right is thoroughly degassed.[215]

Next we must turn to the question of how the sound waves cause what at first sight appears to be such complicated phenomena, the effects being substantially the same whether performed at a few thousand or at a million cycles a second. According to Chambers and Gaines,[46] and Bondy and Sollner[216] the explanation is as follows. A sound wave travelling through a liquid compresses and stretches it periodically. If the stretch is moderate and the irradiated liquid is free of gas, nothing spectacular occurs; but if the liquid is saturated with gas, gas bubbles appear, as was shown in detail by Boyle[16] and Newton Harvey[91] and their co-workers (see below, p. 363). What happens if the liquid is stretched unduly was described nearly seventy years ago as a curiosity by Kundt and Lehmann[124] (working at low fre-

quencies). They said: "While the whole system was vibrating violently the water close to the end of the vibrating rod turned turbid during the vibration. Since it was entirely free of air, these small bubbles causing the turbidity could only be due to the disruption of the water under the influence of the intense vibrations." When irradiating with sound waves of sonic [46] or ultrasonic [218] frequencies carefully de-

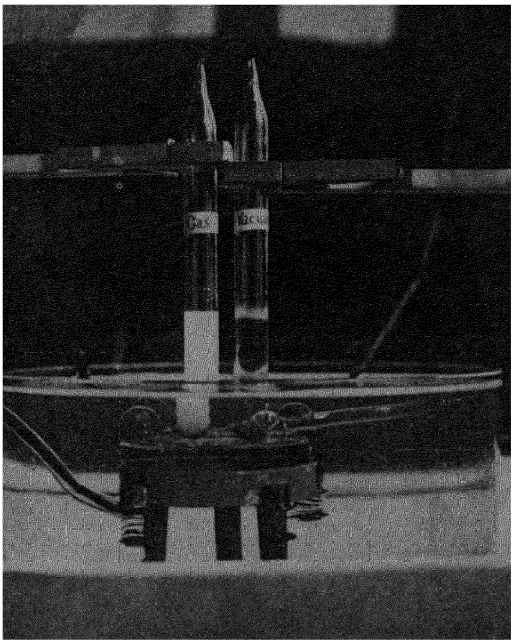


FIGURE 7. Emulsification by ultrasonics. The influence of the presence of a foreign gas upon emulsification. About one-third actual size.

gassed liquids, such as benzene, toluene, or water, zones of a slight and somewhat glittering opacity were always observed, although no bubbles rise to the surface; a hissing noise is heard under these conditions.

Thus it is clearly indicated that the liquid disrupts under the stretch of sound waves; a high hydrostatic pressure prevents this disruption and thus the subsequent phenomena. If the same experiment is performed in a vacuum even at low energies, the liquid bubbles and boils but no hissing noise is heard. Under neither of these conditions does emulsification occur. Strong mechanical action by ultrasonics is always accompanied by a hissing noise, its loudness being so characteristic that it was used by the author and his collaborators as a very convenient and rather accurate indicator of their efficiency. It is apparent that, although the formation of cavities does not give rise to destructive action, their disappearance, always connected with the hissing sound, does.

Lord Rayleigh [188] calculated the pressures which may occur when a vapor bubble collapses in a liquid and found that many thousands of atmospheres may be obtained locally in this way. These enormous local stresses, of course, can cause all kinds of intense mechanical effects. In engineering the entire phenomenon involving the formation of cavities and their violent collapse due to outside pressure is called "cavitation." This term will be used in this sense.

It can easily be demonstrated that the collapse of steam bubbles produces emulsifi-

cation when occurring at the interface of two liquids; this collapse of the steam bubbles is always accompanied by a rattling sound,[216] the same as in the case of emulsification caused by cavitation due to sound waves.

Thus, we may safely conclude that the emulsification by ultrasonics is due substantially to "cavitation," as defined above.[215, 216] This explanation accounts readily for the fact that no emulsification occurs in a vacuum or when a high outside pressure is applied. In a vacuum cavities may be formed, but lack of outside pressure prevents their violent collapse. On the other hand, a sufficiently high pressure prevents the formation of cavities. (The prevention of cavitation by the application of outside pressure is well known in engineering.)

Disruption due to cavitation occurs in an irradiated system preferentially at weak spots, where the energy of disruption is relatively low. The disruption is favored at oil-water interfaces,[213, 216] a fact which probably contributes to the remarkable speed of emulsification; thus cavitation is favored where emulsification may occur. Similarly dissolved gases may act as "weak spots"; [216] they are driven out by irradiation, forming small bubbles, thus favoring cavitation.

It seems permissible to visualize cavitation which leads to emulsification (and other destructive processes) as a very violent hammering of the collapsing cavities, occurring at the phase boundary.

It is always observed that considerable, often turbulent stirring occurs if the ultrasonic irradiation is strong enough to cause rapid emulsification. There is first the collapse of cavities, which causes violent stirring near the phase boundaries liquid-liquid and liquid-gas; next we have to consider the irregular reflection of sound energy at the same places; and finally we must remember that the walls of the container holding the irradiated system vibrate in an irregular manner; thus they cause gross stirring which is frequently made conspicuous by the formation of vortices.

According to Smith,[207] the oscillation of gas bubbles in a sound field is the cause of many of their mechanical actions in liquids, such as emulsification. There seems little doubt that the ideas of Smith deserve very careful consideration; if properly coördinated with the cavitation theory of emulsification, they will considerably deepen our insight into the mechanism of emulsification. The ideas of Marinesco [183, 184, 185, 187, 188, 189] concerning the mechanism of emulsification are rather detached from any observed facts.

The influence of pressure and vacuum is not observed when dispersing liquid metals, since, as already mentioned, their dispersion is not due to cavitation.

Emulsions prepared by sound waves do not show any distinguishing features. The concentrations and particle size vary greatly according to the materials used, the intensity of the irradiation, its duration (unless a stationary state is reached), and the size and shape of the reaction vessel.

The best-investigated case among the metal dispersions is that of mercury.[54, 218, 220] The concentrations obtained by Bondy and Sollner,[218] in samples of a few ml and using high sound intensities, were about 6 gm/l for water and aqueous solutions of stabilizing electrolytes, and about 60 gm/l in aqueous systems containing such stabilizing agents as lysalbinic acid. The stationary concentration which does not increase on longer irradiation was reached in about one minute. The particle size varied from the ultramicroscopic range to diameters of 5-8 microns, larger particles settling out rapidly. Quantitative particle size distribution measurements were given by Bondy and Sollner.[218]

With organic solvents (cyclohexane, petroleum ether, benzene, toluene, xylene, ether, acetone, chlorobenzene, carbon disulfide) at room temperature only extremely dilute, rather unstable emulsions of mercury were obtained. At low temperatures, near the melting point of mercury, more concentrated emulsions may be obtained. Their stability is also greater, if they are kept at these lower temperatures.

With the alcohols the mercury concentration rises from 2 per cent with methyl alcohol to 18 per cent with isoamyl alcohol, etc. Generally in viscous liquids, such as glycol, glycerin, and benzyl alcohol, rather concentrated emulsions may be obtained (10-20 per cent). Particle size distributions were measured in many liquids by Bondy and Sollner.[218] In the presence of stabilizers, such as rubber (in benzene), highly concentrated and very stable emulsions may be obtained.

A very interesting and still unexplained influence of gases on the stability of mercury emulsions was found. It is possibly of fundamental interest. In the absence of gases the amount of mercury dispersed in water by ultrasonic waves is very small, about 0.5 gm/l; this is less than 10 per cent of the concentration which is obtained under otherwise identical conditions in the presence of air. The gas-free emulsions are very unstable; after a short time the water is quite clear again, all the disperse phase having settled down and united with the bulk mass, which then shows its normal reflecting surface. When gases are present (air, N_2, O_2, H_2), ultrasonics rapidly produce fairly concentrated and much more stable emulsions. The droplets, after having settled down, do not coalesce, but form a gray sediment which coalesces with the bulk of the mercury so slowly that it does not entirely disappear even after many days. This sediment may be redispersed merely by shaking. Since this effect is found whether either hydrogen or oxygen is present, chemical action may be excluded. Feebly adsorbable electrolytes do not influence this effect of gases. More strongly adsorbable electrolytes render the gas influence less distinct. In the presence of true protective agents, such as soaps, gelatin, or lysalbinic acid, the effect of gas disappears entirely. In liquids, such as butyl alcohol, amyl alcohol, nitrobenzene, aniline, etc., the effect was especially marked. Although the emulsions in these liquids show only a small difference as to their concentration shortly after irradiation, the stability in the gas-free systems is much less (Table 2).

Table 2. Change of Concentration of Hg-Emulsions in Course of Time[218]

Time after Emulsification	Disperse Phase in gm/l			
	In Nitrobenzene		In *n*-Butyl Alcohol	
	In Air	*In Vacuo*	In Air	*In Vacuo*
1 minute	8.4	6.4	7.6	6.0
30 minutes	6.4	1.6	3.5	1.2
3 hours	5.0	1.3	2.4	1.5
16 hours	4.1	0.9	2.0	Nearly clear
48 hours	3.0	Nearly clear	1.6	clear

Kremnev's [119] assumption that gases favor the process of dispersion cannot explain this difference in stability. The stabilizing effect of gases is presumably general. [192, 217, 218] Cone, Tartar and Taylor [55] have found that Zsigmondy gold sols change their color to a brighter red and have a higher precipitation value after being saturated with gas; hydrogen and oxygen exert their action in the same direction. (For further references see [217].) A more detailed investigation of the stabilizing effect of gases seems very desirable.

Other metals besides mercury may easily be dispersed if their melting point is low enough. They are dispersed only in the liquid state; solid metals cannot be dispersed. The alkali metals form concentrated emulsions in organic liquids, such as paraffin oil and xylene.[54, 218] Those in paraffin oil have very intense colors: reddish violet in the case of sodium, bluish green in that of potassium.[218] Concentrated emulsions in aqueous and non-aqueous liquids may be prepared with Wood's metal.[218, 256] Very finely grained emulsions of this metal are obtained in bromonaphthalene and tetrachloroethane. When the coarser particles have settled a brown, stable, truly colloidal solution remains.

Oil in water and water in oil emulsions were produced with many pairs of incompletely miscible liquids.[218] The limiting concentrations in the water phase are 4-5 per cent with such substances as benzene, toluene, xylene, aniline, nitrobenzene, etc. The water-in-oil emulsions are in most cases very dilute. In properly protected systems, *e.g.*, in the presence of soaps, 75 per cent emulsions of oil in water are readily obtained. Particle size distribution measurements have been published by Bondy and Sollner.[218] (Further references on emulsification and emulsions: [38, 87, 120, 147, 149, 178, 189, etc.]) Chambers [37, etc.] has suggested the homogenization of milk to induce soft curd character by intense sonic treatment. This seems to be the first attempt to use the emulsifying action of sound waves on an industrial scale.

Formation of Fogs. The formation of fogs, *i.e.*, the "atomization" of nonmetallic liquids by ultrasonics, was first described by Wood and Loomis.[256] With fairly volatile liquids, such as benzene, toluene, water, etc., fog formation is best shown by irradiating a beaker containing a little of one of these liquids with sound waves of fairly high intensity. The beaker fills rapidly with a white cloud; this fog is in a continuous turbulent motion.

To produce fogs of more viscous liquids, *e.g.*, of transformer oil, another arrangement must be used. A test tube about 2.5 cm in diameter is drawn out in the middle to form a thin-walled constriction of about 7 mm diameter. When it is dipped into the oil bath of an ultrasonic generator, the oscillations of the glass at the constriction are very vigorous. If a little oil is now applied to the outside of the tube above the constriction, it spreads over the surface and seems to be thrown out in jets of spray, and a dense cloud gathers about the tube. Close inspection shows that the oil on the collector gathers in more or less regular rings which form dots here and there, and the jets of spray seem to be shot out from these dots.[256]

The strong heat effects (see pages 367 to 369) always connected with intense irradiation might be supposed to cause the formation of fogs by a mechanism of evaporation and immediate recondensation. But, as shown by the author,[212] this is not the case, for if solutions containing a non-volatile compound, such as a dye, are used, colored fogs are readily formed and hardly any residue of the non-volatile substance is left after the whole liquid has been dispersed into the air. Thus, "temporary" evaporation is excluded as the cause of formation of fogs.

Another possible explanation might lie in the fact that, during irradiation, all liquids form mounds or fountains which throw drops of the liquid into the air.[256] However it was demonstrated that this phenomenon is in no way directly related to the formation of fogs.[212]

The movement of the fog away from the collector, as if coming from jets, is due to air currents [143] set up in the neighborhood of the vibrating collector; these air currents are strong enough to blow the flame of a micro-burner horizontally sideways when it is brought close to the collector.[212] The air currents, however, do not explain the formation of the fog droplets.

An indication as to the true mechanism of the fog formation may be seen in the fact that, if the intensity of irradiation is increased, the surface of a liquid always becomes rippled as if it were being hit from above, just before fog formation sets in. It was previously mentioned that phase boundaries are frequently a preferred place of cavitation; for the case of the air/liquid interface, this was demonstrated in emulsification experiments by Sollner and Bondy.[219] It, therefore, seems very likely that fog formation is also due to cavitation. This would be proved if the effect disappears when cavitation is rendered ineffective. Obviously, the pressure method cannot be applied since the surface, at which the fog is formed, is in contact with the gas by means of which pressure can be applied, and saturation in the liquid film with which we are concerned occurs much too rapidly to allow conclusive experiments.[212] The vacuum method, however, can be used. (Liquids having a low vapor pressure are best suited for this, because they exclude minor complications

caused by the vapor pressure of the irradiated and consequently heated liquid.) No fogs are formed if the liquid was first carefully degassed and sealed in test tubes *in vacuo;* this holds true even with such strong irradiation that a considerable fountain is formed in the liquid and large drops abundantly thrown out from its surface. This surface *in vacuo* also lacks the characteristic rippled appearance always associated with fog formation. One may therefore safely conclude that the formation of fogs is due to the same mechanism as that producing all other disruptive and destructive effects in non-metallic systems, *i.e.,* cavitation.[212]

In the manner described by Wood and Loomis many organic liquids have been successfully "atomized": benzene, toluene, tetralin, decalin, paraffin oil, and many others.

It has not been possible to produce fogs which are stable for more than a very few minutes, since the particles are too large. For this reason they have not been investigated extensively. Ultrasonics actually do not at present provide a really satisfactory method for the preparation of fogs: other methods yield them much more readily and with much smaller particles. The converse phenomenon, the coagulation of fogs by sound waves, which was very extensively studied, will be discussed below.

Dispersion of Sediments, Precipitates and Gels. The peptizing action of ultrasonic waves on gels, gel-like substances, precipitates and sediments has been extensively investigated. All such systems contain pre-formed particles of colloidal or semi-colloidal size, and it has been found that these particles may be separated from one another readily by the action of audible and inaudible sound waves. The preparations made by sound action do not show any special features; the sound waves act merely as a mechanical dispersing device. However, the sound method of dispersion and peptization of sediments, precipitates and gels undoubtedly offers many advantages to the investigator in the field of colloids on account of the cleanliness and good reproducibility of this method and the small quantities of materials necessary.

Freundlich and Sollner [78] have demonstrated that the peptizing action of sound waves is due to cavitation. Neither under a high pressure nor *in vacuo* is peptization ever observed.

Mention may also be made of two further points of importance in the peptization of gels: the vigorous movement to be seen in two-phase systems when irradiated, and the heat (see below) which is developed in many disperse systems. The gross stirring and the increase in temperature, of course, in many cases greatly increase the rate of peptization. That the same is true for the rate of reaction in heterogeneous systems in general [see *e.g.,* 144] hardly needs to be mentioned.

Some of the reported cases of dispersion of precipitates, sediments and gels are: dyestuffs and pigments; [203, 247, 248, 249] metal hydroxides; [73, 74, 76, 198, 200, 201] starch; [158, 159, 160] carbon in aqueous alcohol; [194] reversible liquefaction of thixotropic systems, mostly of metal hydroxides; [73, 74, 76, 198, 199] thixotropic silica gel; [70, etc.] the liquefaction of thixotropic gels gradually blends in the reduction of structural viscosity of similar but "liquid" systems. This problem will be dealt with in a separate section below, after the discussion of the dispersion of coherent solids.

The preparation of extremely fine-grained photographic emulsions of great sensitivity was reported by Claus,[50, 52, 53] and Dangers,[56] who irradiated silver bromide emulsions during their preparation. Claus reported in detail on the properties of these photographic preparations. The mechanism of the action of the sound waves in this case is not definitely known. One would be inclined to attribute it to the general destructive and disruptive action of sound waves caused by cavitation; however, other suggestions have been made; the thermal effects (to be discussed below) associated with strong sound action have been strongly emphasized. Chemical activation (see below) may also come into play. (See, *e.g.,*[140, 180, 181, 182].)

At the end of this section we must mention some very important work by Claus. As will be seen in the next section, a satisfactory dispersion by sound waves of solids of great cohesion, *e.g.*, of metals, is at present a practical impossibility. To circumvent this difficulty Claus,[51, 53] and Claus and Schmidt,[54, see also 252] submit systems in which metallic precipitates are formed by electrolysis or chemical reaction during these processes to the action of strong ultrasonic waves. As cathode in the electrolysis process a noncorrosive metal, *e.g.*, stainless steel, is used, to which the products of the electrolysis adhere poorly. The freshly formed metal is torn off the electrode (undoubtedly by cavitation) and thus dispersed in the liquid. The authors cited report that they produced in this manner fine-grained dispersions of Pt, Au, Ag, Hg, Cu, Cd, Pb, Bi, Ni, Cr, Fe, Al. (It seems doubtful whether the dispersions of the less noble metals should be considered as being of a purely "metallic" character.) The Hg emulsions produced by the method of Claus were much finer-grained than those produced by dispersion of the bulk phase.[54]

Dispersion of Solids in Liquids. The first to try the dispersion of coherent solid bodies by ultrasonic waves was Richards;[186] he reports unsuccessful attempts to disperse solids (glass, quartz) in liquids. Rschevkin and Ostrowsky[189] claimed to have dispersed tin, sulfur, bismuth, lead, copper and even silver. The "emulsions" of sulfur, tin, bismuth and lead are said to be true colloidal solutions with a particle size of about 10^{-6} cm.[see also 223] Sollner,[214] in criticizing their work, pointed out that one can consider a solid body to be truly dispersed by sound action only if one starts out with clean surfaces, carefully freed of all adhering dust or foreign substance. This is conveniently done by irradiating the sample under investigation repeatedly in water for several minutes. When initially irradiated, nearly all solid substances show dispersion. The difference between those substances which really are dispersed and those in which only adhering material is torn off becomes apparent after three or four irradiations. The author's criterion of a true dispersion was that the seventh to tenth irradiation (of two minutes each) invariably yielded the same positive result.[214] With these proper precautions one finds that silver, tin, glass, quartz and marble are totally unaffected, even by prolonged, very intense irradiation. Fibers of glass wool are broken, asbestos is split into fibers, and even thin foils of metals may be torn apart; but a true dispersion, yielding colloidal or semi-colloidal suspensions, is never observed in these and similar cases. Examples of substances comparatively easily dispersed in water, yielding colloidal or semi-colloidal solutions, are (crystals of) mica, gypsum, steatite, hematite, sulfur, and graphite.[214] They are all substances of only moderate cohesion, which are easily split and broken.

The suspensions so obtained contain particles of different sizes, their diameter being from several microns down to colloidal size. In the cases of mica, steatite, hematite and graphite, for example, the suspensions were particularly rich in particles of truly colloidal and semi-colloidal dimensions. (For the dispersion of camphor and iodine see [252].)

The concentration of these suspensions is generally rather low, unless the irradiation is continued for a considerable period; *e.g.*, a piece of steatite with an apparent surface of about 5 cm^2 yielded 8 mg of dispersed matter in 5 cc of water after two minutes of irradiation at as high an energy as could be applied. Hematite, with a surface of about 8 cm^2 gave 1 mg under the same conditions, the particles obtained being non-spherical and particularly small.[214] The amount dispersed depends upon the character of the irradiated specimen; well developed crystals are less liable to destruction.

The solid substance to be dispersed need not necessarily be of macroscopic size; suspended particles have been successfully dispersed. Even the particles of truly colloidal solutions may be broken down further by irradiation. The particles of a V_2O_5 sol prior to irradiation were truly colloidal in two dimensions, their length

being between 10-15 μ. After irradiation their length was less than 3 μ. In this and similar cases the effect is more pronounced, and is obtained in a shorter time, if the concentration of the irradiated sol is decreased. Substantially the same results are obtained with colloidal graphite suspensions. The number of particles visible with bright field illumination decreases on irradiation, while the number of particles visible in the dark field increases greatly. The increase in particle number continues, with further irradiation, after all particles visible in the bright field have vanished, thus indicating that particles truly colloidal in all three dimensions are further broken down.[214]

The dispersion of solids in liquids was shown to be due to cavitation.[214]

The status of the problem from the point of view of the colloid chemist can be summarized as follows. Sound action is capable of dispersing only solid bodies of moderate cohesion. Thin metal foils may readily be broken and torn into small fragments, but no true dispersion ensues.[51] From the engineering literature we know that cavitation causes the corrosion of the propellers of steamships (for references see [216]), though we do not know yet whether this is due entirely to direct mechanical action; chemical factors (activation of oxygen, see below) may also come into play. Claus and Schmidt [54] have reported a corrosion of the edge of steel blades on prolonged irradiation with ultrasonics of high intensity; the brittleness of steel and the peculiar geometrical configuration of the cutting edge of a steel blade make it doubtful whether this corrosion would proceed very far. (See also the careful investigation of Schumb, Peters and Milligan [244].) The limits of what can be achieved experimentally depend quite obviously on the intensity of the irradiation. The best chance to extend the sonic dispersion of solids probably lies at present in the use of sound concentrating devices, such as the quartz crystal lenses of Gruetzmacher.[88]

Changes in the Anomalous Viscosity of Colloidal Solutions, Splitting of Macromolecules and Related Phenomena. The liquefaction of thixotropic gels [73, 74, 76] was discussed in a previous section. This effect can be considered as an extreme case of change in anomalous ("structural") viscosity. "Fluid" systems the viscosity of which is decreased by sound waves were described by Szent-Györgi,[229] and Szalay [227, 228] (gelatin, starch, gum arabic); Kimura [113] (gelatin, gum acacia, Na-stearate); Heymann [103] (methyl cellulose); Sibata [205, 206] (rubber); Freundlich and Gillings[69] (gum tragacanth, gum arabic, gelatin, agar, cotton yellow, benzo-purpurin, Na-stearate, and Na-oleate); Schmid and collaborators [233, 234, 240, 241] (polystyrene, polyvinyl acetate, polyacrylic acid esters, nitrocellulose, rubber, etc.) [See also 61, 195, 196, 246] It may be worth while to note that the true viscosity of liquids and solutions showing normal viscosity, such as glycerin or its aqueous solutions, is not affected by sound waves.[69]

Freundlich and Gillings have investigated the general problem carefully on a fairly broad basis.[69] Of all the classes of substances investigated by Freundlich and Gillings, the gums show the simplest behavior. Dilute solutions of gums, *e.g.*, gum tragacanth, are very viscous and show a small degree of structural viscosity. The quantitative relationship can easily be established with a Couette viscosimeter. By treatment with sound waves the viscosity and its anomalies are reduced; the gum solutions become less viscous and show smaller deviations from Poiseuille's law. To bring about these effects, short (½ to 3 minutes) irradiations suffice, of about the same intensity as those used for emulsification. The viscosity of the irradiated solutions rises slowly on standing, but does not reach the original value. By applying pressure it can be shown that this effect is due to cavitation; no change in viscosity occurs when a sufficiently high outside pressure is applied during irradiation.

The behavior of gelatin solutions is not so simple; 0.5 per cent solutions show a distinct structural viscosity. Irradiation for 1 to 3 minutes with ultrasonics of different intensity results in all cases in a marked decrease in viscosity, but of a fairly

complex nature. Freundlich and Gillings [69] find that weak irradiation, sufficient to reduce the viscosity at low shear rates, may have no great effect at the higher rates, whereas more intense irradiation effects an overall reduction of viscosity at all shear rates. The solutions recover their structural properties rapidly on standing. There appear to be two kinds of irradiation effects which seem to cause the anomalous viscosity of gelatin solutions. In the region of lower shear rates the resistance of the solutions to flow is probably bound up with gel formation; at the higher rates we have a closer approach to a viscous resistance, the gel structure is destroyed, and the behavior is similar to that of a non-gelating solution, like a gum. Experiments on the mechanism of the attack of the ultrasonics confirm this view. If cavitation is rendered ineffective *in vacuo,* or if cavitation is prevented by pressure, the anomalous viscosity of gelatin solutions is much less affected by ultrasonic irradiation, as described in detail by Freundlich and Gillings.[69] However some reduction in viscosity cannot be prevented in this manner.

The behavior of agar solutions is similar to that of gelatin. Schmid and collaborators [233, 234, 240, 241] found a similar complexity of behavior in several other cases, cavitation being responsible for part of the effect. Schmid, who irradiated his samples for long periods—up to 3 hours—thinks that frictional forces resulting from the relative motions of a rigid network of macromolecules and the vibrating solvent may cause the breakdown of macromolecules.[234] Some other valuable ideas concerning this problem have been published by Szalay.[228] High polymers seem to be a fruitful field for further investigations. The problem of recovery should be studied further; it blends gradually with that of influencing polymerization reactions.

Different from that of the preceding groups is the behavior of certain dyestuffs. Solutions of cotton yellow, *e.g.,* have characteristic colloid properties, especially as regards viscosity and optical behavior. Dilute solutions show distinct anomalous viscosity and strong streaming double refraction. Both effects are certainly connected with the presence of rod-shaped particles which are visible in the ultramicroscope. The structural properties of the dye were found to be very sensitive to ultrasonic irradiation; a short irradiation of low intensity was sufficient to effect a complete reduction of anomalous viscosity. The optical properties of the solutions are changed parallel with the reduction of structural viscosity. Ultramicroscopic examination showed a uniform dispersion of needle-shaped particles in the original solution; after irradiation the long particles had disappeared, and a very fine dispersion of apparently isotropic particles remained; correspondingly, the strong stream double refraction of the original solution had disappeared after irradiation. There was but little recovery on standing. In this case the dispersing action was due to the collapse of cavities. Irradiation under pressure produced no change in viscosity and none in optical properties.

Benzopurpurin sols showed a decrease in particle size after fairly strong irradiation, but without change in viscosity. Freundlich and Gillings also investigated V_2O_5 sols and found no change in particle size or in viscosity. However the intensity of irradiation used by these authors was not great. (Compare below, page 356 and [214]).

Sodium stearate solutions have also been carefully investigated. Such solutions show considerable deviation from Poiseuille's law and an unusually great increase of viscosity with concentration. Freundlich and Gillings [69] found that the viscosity of these solutions is reduced rapidly by ultrasonics, provided the intensity exceeds a certain minimum value which rises with increasing concentration. The viscosity changes in this case are accompanied by a most marked change in appearance: whereas the original solution has a curdy, opaque, somewhat gelated, but generally homogeneous appearance, the irradiated solution is a limpid suspension of silky, needle-shaped, crystalline particles, showing streaks on stirring.

Ultrasonics seems in this case to accelerate a normally slow process somewhat

akin to crystallization. The collapse of cavities is essential to viscosity reduction.

Sound waves may, in the long run, be found to be a valuable tool for the investigation of the structure of many colloidal systems. An additional indication in this direction may be seen in the work of Brohult,[82] who reports that the hemocyanin molecule is split into fragments of one-half and one-eighth the original value, if irradiated with ultrasonics. Confirmation and expansion of this highly interesting observation seems most desirable.

The denaturation of proteins has been reported many times.[40, 41, 48, 128] In many cases it seems likely that there is a kind of surface denaturation. Further investigations would be desirable.

Accumulation and Orientation in Gaseous Systems. The best known ponderomotive effect of sound waves is the accumulation of fine solid particles in the nodes of a field of stationary sound waves in gaseous systems—Kundt's dust figures.[122, 123]

With audible sound these figures are customarily demonstrated in the following manner (Fig. 8). One end (in Fig. 8, the left) of a horizontal glass tube is closed



FIGURE 8. Kundt's dust figures in gases.

by a movable air-tight piston. A glass rod, clamped at the middle, projects into the other end of the glass tube. A light-weight disc fitting the tube loosely it attached to the end of the rod within the tube. The rod is made to vibrate longitudinally by rubbing it. Its ends become antinodes; the disc at the end of the rod vibrates like a piston, thus imparting a regular succession of impulses to the air column enclosed in the tube. The piston on the other end of the tube is adjusted until at a certain position the node of the glass rod is considerably amplified. Now the air column is in resonance, its length being a simple multiple of the half wave length of sound in the gas, and stationary waves are set up. If some light, powdered material, *e.g.*, cork dust, is present in the tube, it collects in the nodes.

In the sound field generated by an ultrasonic vibrator the phenomenon is much the same; fine dust brought into the sound field collects in the nodes (see, *e.g.*, [223]). With high intensities the particles may even be kept suspended for a considerable time in the sound field, between the vibrator and a reflector.[34] (The coagulation resulting from this accumulation of particles will be discussed below.) The movements of the particles of an aerosol in a sound field are too complicated to be discussed quantitatively here. They have been considered by König;[117] Brandt and Freund;[25, 28, 29, 30] Brandt, Freund and Hiedemann;[107, 109, 110] Lewis and Farris;[127] Wagenschein;[254] and others.[125, 132] We shall confine ourselves to a qualitative outline of this phase of sound physics.

The observed accumulation of the particles of an aerosol is the result of the slightly asymmetrical oscillations of the particles in the sound field. The individual particles of an aerosol follow the oscillations of the gaseous medium the more closely, (1) the lower the frequency, (2) the smaller the mass of the particle, (3) the lower its density, and (4) the higher the viscosity of the gaseous medium. Small, light particles follow the movements of the medium almost completely at low frequencies; at higher frequencies this movement decreases as the frequency increases. Large particles oscillate but little even at low frequencies, their inertia being great compared with the forces acting upon them; they are practically stationary at high frequencies. Brandt, Freund and Hiedemann[110] have published photographs of particles of different diameters oscillating in a sound field. The images of the small particles are drawn out into streaks in the direction of the propagation of the sound waves, large particles appearing as stationary circles.

In the diagram of Fig. 9 are plotted the ratios of the amplitudes of particle and gaseous carrier, $\frac{Ap}{Ag}$, as a function of the particle radius, r, for four different frequencies, as calculated by Brandt, Freund and Hiedemann.[110] These calculations refer to air of atmospheric pressure; the specific gravity of the particles is assumed to be 1. The curves of Fig. 9 are a quantitative expression of the above-mentioned general considerations for some special cases; they seem to be in satisfactory agreement with the observed facts. It is evident that with smoke and fog particles of the sizes that usually occur in natural and artificial smokes and fogs (1 to 5 μ), the strongest effects can be expected in the region of sonic and lower ultrasonic frequencies; with higher ultrasonic frequencies only the very smallest particles are still appreciably affected.

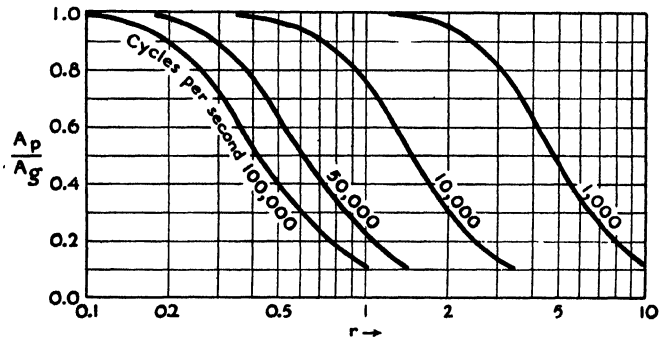


FIGURE 9. The ratio of the amplitudes $\frac{Ap}{Ag}$ in its dependence of particle radius for different frequencies.

It is easy to visualize qualitatively that the impulses received by the oscillating particles toward a node and away from it are not strictly symmetrical. The impulses from the side of the loops are slightly larger, since the loops are the places of the maximum movement of the oscillating gas. The net result of this asymmetry of oscillation is that the particles are slowly shifted toward the nodes, where they accumulate. This accumulation leads to coagulation, and is therefore of interest to the colloid chemist.

Brandt, Freund and Hiedemann [110] have also reported that non-spherical aerosol particles are oriented in a sound field. A detailed discussion of this problem is lacking.

Accumulation in Liquid Systems. As in gaseous systems, "dust figures" occur in liquids. If a sound wave travels through a liquid which contains dissolved gas, a part of the gas is driven out (see below) and unites in small bubbles; the gas bubbles accumulate in the nodal planes of stationary sound waves. This effect, first observed by Dörsing,[64] has been investigated carefully by Boyle and his collaborators [15, 17, 18, 19, etc.] In progressing (not stationary) waves the bubbles are driven away from the source of sound. Similar effects can be observed with dispersed solids and liquids, as described by Boyle *et al.*[16, 20] Particles settling in a liquid in which a field of stationary sound waves is set up accumulate in planes of half wave length and may, as they settle, be collected in beautiful stationary wave patterns on a horizontal plate.

Accumulation can be readily observed in all emulsions and suspensions.[62, 188] Sollner and Bondy [219] have investigated accumulation more closely from the colloid chemist's point of view. Stationary wave patterns produced by ultrasonics in a liquid

were investigated in this manner: the emulsion (or suspension) is put into a thick-walled capillary slightly U-shaped in the middle; this part is dipped into the oil fountain of an ultrasonic generator, the two ends protruding horizontally sideways (Fig. 10). This arrangement avoids the very disturbing changes in concentration due to creaming up or settling down close to the meniscus. Stationary waves cause the formation of Kundt's dust figures (Figs. 10, 11a and b). With increasing time of irradiation the zones of accumulation become much sharper. Under favorable conditions thin discs with a well-defined boundary seem to stand upright in the liquid. The distance between two bands is $\lambda/2$, λ being the wave length in the liquid. With aqueous emulsions of benzene, toluene, etc., the first band is found at a distance of $\lambda/4$ from the meniscus (cf. Fig. 11a), *i.e.*, at a node, the meniscus always being the locus of an antinode. If exactly the same experiment is done with aqueous emulsions of nitrobenzene or mercury, or with a suspension of quartz, the first zone

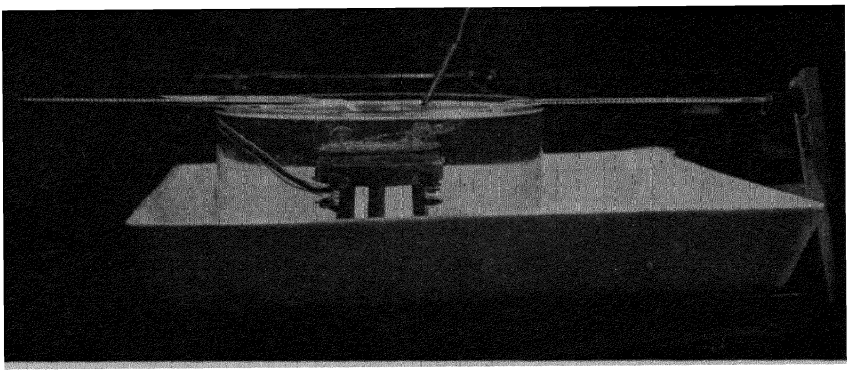


FIGURE 10. Arrangement for the study of stationary wave patterns, about one-fourth actual size.

of accumulation lies at the meniscus itself, the next a distance of $\lambda/2$ from the meniscus, *i.e.*, at the antinodes (cf. Fig. 11b). So far as the experimental evidence goes at present, accumulation occurs at the nodes if the dispersed substance is less dense than the medium of dispersion, and at the antinodes if it is denser. However the decisive factor may not be simply the difference in specific gravity between the medium of dispersion and the disperse phase; it may be some closely related function which takes into account the different sound velocities. In an aqueous system containing particles both denser and less dense than water, ultrasonics may separate the two kinds of particles from each other. This is shown in Fig. 11c for a mixture of an emulsion of toluene and a suspension of quartz; this figure is evidently produced by a superposition of the effects shown in Fig. 11a and 11b. The zones of accumulation of toluene and of quartz alternate with each other, the distance between a band of the one and the other being $\lambda/4$. Toluene and quartz are readily distinguished by the fact that the former creams, whereas the latter settles. (The photographs 11a, b and c were taken from the side.)

The rate of accumulation of dispersed substances at the nodes or antinodes depends on the size of the particles, as is shown by the following experiments: three capillaries having the same dimensions were filled with suspensions of quartz (concentration about 3 per cent). The quartz powder had been roughly fractionated. The capillaries contained particles with diameters between 4 and 10 μ, 1 and 4 μ, and below 1 μ. On irradiation with medium energy, the wave patterns are formed immediately with quartz particles of 4 to 10 μ, the liquid between them being free

from particles. With the particles of 1 to 4 μ it takes a few seconds for the space between the striations to become practically free of particles, many seconds for the striations to become somewhat narrow, and a still longer time until the pattern is similar to that of the larger particles. With a suspension of particles of 1 μ and smaller, zones of diminished concentration appear after several seconds, and zones of a visible accumulation appear after a minute; but even after thirty minutes the whole pattern is blurred. In suspensions containing particles less than 0.5 μ the effect is still less distinct. In truly colloidal solutions accumulation has not been obtained at all by Sollner and Bondy with a frequency of 214,000 cycles per second.

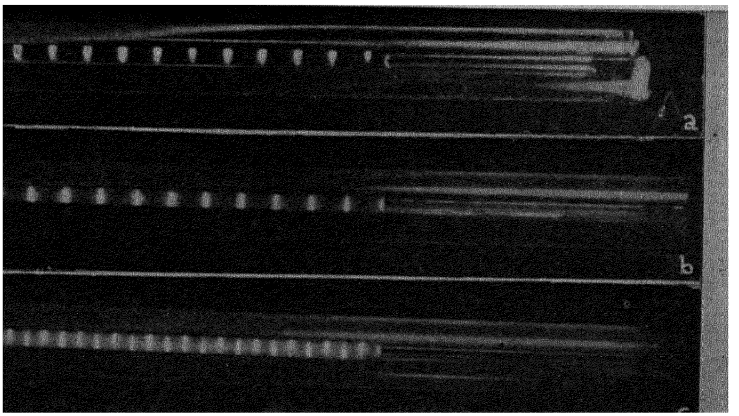


FIGURE 11. Stationary wave patterns. (a) Toluene in water; (b) quartz in water; (c) toluene and quartz in water. About 1.5 times actual size.

The ponderomotive effects of sound waves were discussed by Boyle (see above) and King.[114, 115] According to King the pressure on a sphere with a radius R in a sonic field of an energy E watt/cm^2 and a wave length λ is given by the expression

$$P = \pi R^2 \frac{(2\pi R)^4}{(\lambda)} E$$

The radiation pressure increases with the sixth power of the particle radius. Numerical evaluation[101] shows that, in agreement with the observations of Sollner and Bondy,[219] no accumulation can be expected in truly colloidal solutions, whereas strong effects are predicted for emulsions and suspensions with microscopic particles.

The electrical effects which accompany the oscillations of colloidal particles in a sound field have been considered by Hermans[98, 99] on the basis of earlier work on electrolytes by Debye.

A further quantitative study of the accumulation effect and its correlation to the mechanical properties of medium of dispersion and disperse phase, particle size and charge, etc., would be highly desirable; it is bound to yield much basic information from the point of view of the physicist as well as of the colloid chemist.

Orientation in Liquid Systems. On the basis of simple hydrodynamic considerations, Burger and Sollner[221] looked for and found a sonic orientation effect; anisometric particles are oriented in a sound field with their long axes perpendicular to the flux of energy. The phenomenon is best observed in a field of stationary waves. If a long test tube containing a dilute suspension of mica, mosaic gold, cer-

tain clays, or similar material is dipped into an oil bath containing a vibrating quartz oscillator, silky, brilliantly glittering zones at more or less regular half wave distance appear immediately, a sure sign of orientation; the term "sonic double refraction" has been suggested for this phenomenon. In suspensions with larger particles the orientation can readily be detected by close visual inspection. If the irradiation stops, orientation disappears immediately, except with rather large particles, in which case the orientation may persist for several seconds. The energy necessary to bring about orientation is very much smaller than that required for accumulation. Thus the two phenomena are very readily separated experimentally. This is easily understood, because in order to obtain any appreciable accumulation, particles must be moved several tenths of a millimeter, whereas orientation is effected by simply turning the particles into the favorable position, the maximum movement necessary being only of an order comparable with the length of the particles. Of course, the forces which bring about orientation are the same as those which (with higher energies) cause accumulation. Orientation is more pronounced in the antinodes than in the nodes, and larger particles are oriented much more readily than smaller ones. For example, in the experiments of Burger and Sollner three tubes containing quartz powder suspensions with particles from 4 to 10 μ diameter, 1 to 4 μ, and smaller than 1 μ respectively, were irradiated with medium energy. The 4 to 10 μ suspensions at once showed distinct glittering; subsequently the 1 to 4 μ suspensions followed, showing a weaker effect; finally came the finest suspension, the effect here being still weaker and not increasing, even after many seconds. It became distinct only when the energy was increased. Orientation (contrary to accumulation) can be observed with truly colloidal solutions, as was demonstrated with aged V_2O_5 and Fe_2O_3 sols; [221] in concentrated sols the orientation can be detected in reflected light; in dilute sols transmitted polarized light is necessary. In very viscous, gel-like sols several minutes may be necessary to produce orientation; conversely the orientation may persist for many days after irradiation has ceased. Hermans,[100] Oka,[155, 156] and Petralia [169] have investigated the orientation phenomenon further, and have tried to develop the physical theory of the effect, mainly on the basis of theoretical considerations of Pohlman [176] and King.[115]

Coagulation in Gaseous Systems. The coagulation of aerosols, *i.e.*, of fogs and smokes, is without doubt the most intensively investigated action of sound waves in colloidal or semi-colloidal systems. Sonic coagulation in aerosols has been discussed by Patterson and Cawood;[165] Brandt and Freund;[26, 27] Pearson;[168] Andrade and Parker;[2, 3, 163, 164] Gottschalk and Clair;[82] Clair;[48, 49] Dean;[60] Tatum;[245] Gorbatschew;[80, 81] Gies;[78] and others. It has been thoroughly investigated, particularly by Brandt, Freund and Hiedemann;[24, 104, 105, 107, 108, 109, 110, 111] and reviewed jointly [107 109, 110] and separately by these authors.[23, 104, 106]

It is well known that fogs and smokes coagulate spontaneously; practically all collisions between particles are non-elastic and lead to the formation of aggregates. The coagulation of aerosols can be greatly accelerated if they are exposed to sound waves. Photomicrographs of the sediments from fogs and smokes obtained with and without sonic treatment have been published by Brandt, Freund and Hiedemann.[106, 108, 110] The untreated aerosols settled down as fine dust, mainly as primary particles; the sound-treated aerosols yielded sediments composed substantially of large aggregates.

The importance of the frequency of sound waves acting upon an aerosol was pointed out in a preceding section (see also Fig. 9); the strongest effects are obtained with higher audible and the lower inaudible frequencies. Most investigators, therefore, use magnetostrictive sound generators for work in aerosols in preference to piezoelectric generators. The gas-current vibration generator of Hartmann is also used frequently; Clair [48, 49] has used an electromagnetic sound generator. (See pages 344 and 345.)

The usual experimental set-up consists substantially of a cylindrical gas chamber, either filled with the aerosol, or, more advantageously, through which the aerosol is allowed to flow continuously. For use with electrostriction sound generators, one end of this chamber is made of a pistol-like end-plate carried by the magnetostriction rod; the position of the other end-plate is adjustable, so that its distance from the vibrator plate is an integral number of half wave lengths. In this manner a stationary wave field may be set up, the intensity of which is many times that obtainable without this precaution. With air-jet sound generators, the Hartmann whistle is placed in the middle of a cylindrical gas chamber, provisions being made to adjust for resonance by means of adjustable end-plates.

Brandt and Freund [27] have followed the process of aggregation by means of microscopic measurements of the rate of sedimentation of the particles of aerosols. The results of such experiments are shown in Table 3. The aerosol contained in a cylindrical chamber was exposed to the sound action for five seconds. The frequency used was 10,000 cycles per second, produced by means of a magnetostriction oscillator. As the measure of the sound intensity the amplitude of the oscillations performed by the end-plate of the vibrator is given in the first column. In the second column is given the average observable rate of sedimentation of the particles; in the third, the calculated radius of the particles; and in the fourth, the ratio of the particle radii before and after irradiation. In the last column is the ratio of the masses of the particles before and after irradiation, as calculated from the rate of sedimentation.

Table 3. Coagulation of an Aerosol with Sound Waves [27]

Amplitude of the Vibrator (μ)	Rate of Sedimentation (mm/sec)	Average Radius of Particles (μ)	Ratio of Radii r/r_0	Ratio of Masses m/m_0
0	0.5	1.8	...	...
9	0.77	2.3	1.28	2.1
18	1.8	3.7	2.06	8.7
36	2.5	4.3	2.4	13.6
54	16.7	10.5	5.85	200

In some cases aggregations to a mass several thousand times that of the primary particles have been observed. Brandt [24] has also studied these processes continuously by measurement of the light absorption. Fig. 12 shows the rate of aggregation expressed as m/m_0 for three different sound intensities, as before, in terms of the amplitude of the sonic vibrator. If the sonic treatment is carried out for any length of time, the aggregation leads to particles which are too big to be influenced by the sound field and which settle rapidly.

A quantitative theory of sonic coagulation is confronted with extraordinary difficulties, though the behavior of individual particles in a sound field is fairly well understood (see page 346 *et seq.*). In the process of coagulation, aerosols become very heterogeneous and the aggregates so formed deviate widely from a spherical shape. The worst difficulty is that the mechanism of coagulation is not simple, a number of not entirely independent factors working simultaneously to produce aggregation. The importance of these factors varies greatly with the particle size. At present one cannot do more than try to recognize the different possible causes for coagulation and to try to estimate their relative contribution to the observed effects. It is known that in aerosols nearly all collisions between particles are non-elastic and lead to coagulation. Therefore, all processes which increase the number of collisions between particles increase the rate of coagulation. The two main and primary factors are hydrodynamic attraction between the oscillating particles and the increased

probability of kinetic collisions between the particles oscillating with different amplitudes; the latter factor, as pointed out by Sollner,[211] can be considered as a special case of orthokinetic coagulation (Wiegner;[255] Müller;[148] and Tuorila[250]). In addition, there are the changes in concentration and the spontaneous sedimentation of the aggregated particles which leads to orthokinetic coagulation, as originally conceived by Wiegner and collaborators for liquid systems.

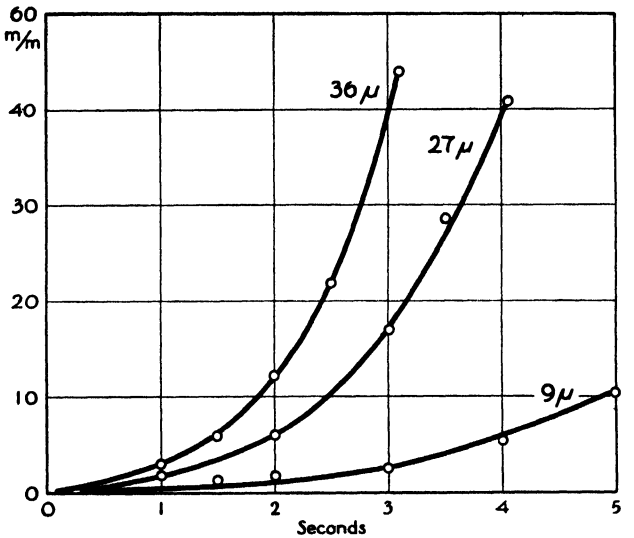


FIGURE 12. The rate of aggregation in an aerosol for three amplitudes of the sonic vibrator.

Let us first consider orthokinetic coagulation. In their work on suspensions Wiegner *et al*[148, 250, 255] have distinguished between a "perikinetic" and an "orthokinetic" coagulation. Perikinetic coagulation is due to disoriented Brownian movement. In dilute aerosols it proceeds at a fairly low rate, much more slowly than sonic coagulation; it therefore can be neglected here. Orthokinetic coagulation, according to Wiegner and collaborators, occurs if a force, *e.g.*, gravity, acting upon the particles of the disperse phase makes them move with differing velocities according to their sizes. This causes collisions between particles of different size (in excess of those caused by Brownian movement) which lead to "orthokinetic" coagulation. Since aerosols are never homodisperse but always contain particles of various sizes, a sound field imparts different velocities to different particles (see Fig. 9); this leads to additional ("orthokinetic") collisions and in turn to orthokinetic coagulation. Brandt, Freund and Hiedemann[104, 109, 110] have tried to estimate the orthokinetic fraction of the observed coagulation in a sound field. They based their calculation on the assumption that there are two kinds of particles in the aerosols: large ones which do not oscillate at all and small ones which oscillate with the same amplitude as that of the medium. Thus, each large particle can be considered to be surrounded by a zone of aggregation, the magnitude of which depends on the oscillations of the small particles. On the basis of these considerations the above-mentioned authors conclude that with suitable frequencies and in fairly concentrated aerosols, orthokinetic coagulation is a sizable fraction of the observed effect.

The problem of sonic coagulation becomes still more complex if we turn to the forces of attraction which originate between two oscillating particles. The most important among these forces is hydrodynamic attraction, as discussed by Bjerknes[11] and König.[117, 118] Further, there are vortex formation, the pulsation of liquid par-

ticles, etc., which seem to play only minor roles however. König has calculated the force of attraction between two oscillating particles. It is believed to be proportional to the third power of the particle radii and inversely proportional to the fourth power of the distance between the two particles. Brandt, Freund and Hiedemann [109, 110, see also 48, 49, etc.] have tried to use these calculations to estimate the rate at which the two spheres approach each other. However, it seems a practical impossibility to apply such calculations with any degree of accuracy to aerosols, for it would be necessary to consider not two, but numerous particles influencing one another over different distances, which vary with time. The whole problem is further complicated by the fact that particles which approach each other through the forces which cause orthokinetic coagulation are brought into positions which are favorable to hydrodynamic attraction; thus, the two different processes are linked together and accelerate the coagulation. Brandt, Freund and Hiedemann, in view of the tentative nature of all these considerations, have confined themselves to rather qualitative estimates, and conclude that, in the range of sonic and lower ultrasonic frequencies, hydrodynamic attraction cannot account for the observed rate of coagulation without assuming orthokinetic coagulation to occur as well. In spite of much thorough work this is about the present state of our knowledge of sonic coagulation—a rather unsatisfactory one from the physicist's point of view.

Numerous attempts have been made to apply sonic coagulation methods to the precipitation of industrial fogs and smokes. Since coagulation occurs readily in streaming aerosols it is possible to handle large volumes; however, no definite statement can be made at present as to whether or not sonic coagulation may be developed into an industrial method; the chances that this may be achieved seem to be rather good.

Coagulation in Liquid Systems. The coagulating action of sound waves in emulsions and suspensions containing macroscopic particles on the one hand, and truly colloidal solutions on the other, seems to be due to entirely different mechanisms; the two cases must therefore be treated separately.

The coagulation of systems containing visible particles has been reported repeatedly by Dörsing,[64] by Boyle and collaborators,[16, etc.] by Wood and Loomis,[256] and others. These authors attribute the phenomenon to radiation pressure. Coagulation in liquid systems was investigated later by Sollner and Bondy [219] on a qualitative basis from the colloid chemist's point of view. A quantitative study accompanied by a thorough theoretical treatment of the problem, though highly desirable, is still missing.

The coagulation of systems with microscopic particles can be demonstrated very strikingly by irradiating an unstable suspension, *e.g.*, quartz powder in an organic liquid. If a test tube containing such a suspension is irradiated even with low sound intensities, the particles accumulate and aggregate; the aggregates so formed soon reach a size which favors rapid sedimentation. A few seconds of weak irradiation are sufficient to clear suspended particles from the liquid almost completely. The coagulation and coalescence of gas bubbles developed in a saturated, or better, in a supersaturated solution of gas in a liquid is also striking. If such a supersaturated solution of gas is irradiated with sound waves for a few seconds, it turns milky white a few seconds later; innumerable small gas bubbles are formed (see page 349). A further brief irradiation causes the gas bubbles to coalesce and to accumulate in more or less regular stationary wave patterns. Soon the gas bubbles become so large that they can no longer be held at these points by the sound waves and rise to the surface.

In a less spectacular manner the same result is also obtained with more stable systems. The concentration of a mercury emulsion (in the absence of a bulk mercury phase), for example, was reduced from about 6 to less than 1 g per liter by 30 seconds of fairly strong irradiation. The absence of a bulk mercury phase is neces-

sary to study the phenomenon of coagulation without interference by simultaneous emulsification. As pointed out in a previous section, in the presence of a mercury macro-phase a stationary concentration of disperse phase is soon reached on irradiation: emulsification and coagulation balance each other.

In the case of the coagulation of mercury emulsions one actually observes not only the aggregation of the original particles, but also their coalescence into larger mercury drops.

Coagulation quite visibly occurs in the zones of accumulation in the liquid. These zones become more pronounced and more regular if higher columns of the emulsion are used in long test tubes. This explains why the concentration of dis-

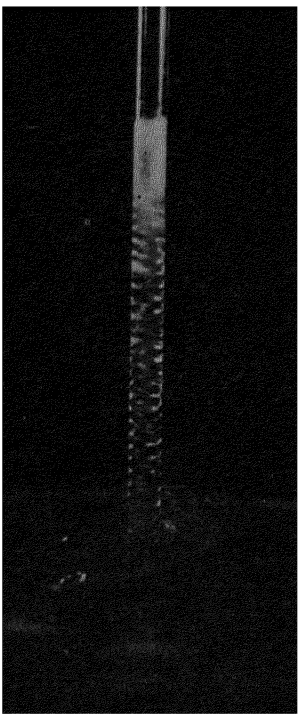


FIGURE 13. Coagulation of an emulsion by ultrasonics, about one-third actual size.

persed mercury, which can be obtained by sound treatment, depends greatly on the volume of the aqueous phase; for, in a given test tube and with a given quantity of bulk mercury the amount of mercury dispersed per second at a given intensity of irradiation is independent of the size of the water phase; on the other hand, the rate of coagulation in the water phase as a whole increases with its volume.

With oil emulsions the situation is very similar. Cavitation, however, becomes a disturbing factor unless very low sound intensities are used or unless it is prevented by pressure. A toluene-in-water emulsion after a few minutes of moderately strong irradiation (without the prevention of cavitation) is shown in Fig. 13. Accumulation, coagulation and coalescence occur rapidly in the middle and lower part of the test tube; the three phenomena may be observed readily under a microscope.[219] Near the surface, where cavitation and thus emulsification is favored, the milky appearance of the emulsion persists; a stationary state between emulsification and coagulation, which is different for different layers (see Fig. 13), is soon reached.

Thus with proper precautions emulsions may be broken readily. It is obvious that the shape and size of the vessel in which emulsification and coagulation occur simultaneously must necessarily play an important role in determining the ultimate concentration of the emulsion obtainable. By prolonged application of sound intensities too weak to cause emulsification, emulsions may be broken completely. Under favorable conditions this is true also for many stabilized emulsions.

With highly protected emulsions the particles aggregate, but little coalescence occurs; the aggregates may be dispersed easily by little stirring after irradiation stops. The same, of course, is true for highly charged, stable suspensions.

Qualitatively, Sollner and Bondy[219] explain the coagulating action of sound waves in systems with microscopic particles in this manner: first, the particles are accumulated and the rate of spontaneous coagulation increases rapidly as the concentration increases; secondly, particles of different size migrate with different velocities toward the zones of accumulation, thus being liable to additional collisions, a kind of "orthokinetic" coagulation;[148, 250, 255] thirdly, the particles do not migrate only in one direction; their macroscopic movements is merely the result of the asymmetry of the oscillations which they perform according to their (different) size; this too causes a kind of orthokinetic coagulation. Whether the factors enumerated are a sufficient and complete explanation of the phenomenon is still open to question; it is not impossible that hydrodynamic attraction between oscillating bodies (*e.g.*,[161]) in the zones of accumulation come into play; dipole forces which are believed[101] to be the most important single factor with truly colloidal solutions should possibly be considered also in the case of macroscopic particles.

The coalescence of the particles of poorly stabilized emulsions, which is observed as soon as coagulation sets in, does not need further discussion; this process, occurring at a slow rate in the absence of sound waves, is greatly enhanced by the ponderomotive action which brings the particles into close contact. It seems not improbable that sonic coagulation of suspensions and emulsions (crude oil) may assume technical importance in the future.

In truly colloidal solutions coagulation may also be brought about by sound action. The mechanisms involved, however, must of necessity be of a different nature, since the ponderomotive effects responsible for the coagulation in systems with microscopic particles are practically absent, as pointed out on page 359. Boutaric and Bouchard[18] observed that the rate of coagulation of unstable sols of gum mastic and As_2S_3 may be increased by ultrasonic irradiation, if the latter causes a gross stirring. If, however, the system was arranged to prevent gross stirring and only a field of stationary waves was set up, no acceleration of the rate of coagulation was observed. Thus no specific sound action on the rate of coagulation was observed by Boutaric and Bouchard; the effect of gross stirring is not specific, being well known under the name of "mechanical coagulation" (see also[209]). Hermans has investigated the problem further theoretically,[98, 99, 102] as well as experimentally,[101] with dilute ($0.3 - 3 \cdot 10^{-3}$ mol/l) AgI sols containing a 50 per cent excess of KI. The resulting bluish-green sols were reasonably stable, and in the lower range of the concentrations used were perfectly clear. Their coagulation rates were studied tyndallometrically.

The very slow rate of spontaneous coagulation in the original sols is hardly influenced by sound treatment. In the presence of electrolytes NH_4Cl, $NaNO_3$, KNO_3 ($0.6 - 0.24$ mol/l) and $BaCl_2$ ($0.001 - 0.02$ mol/l) the rate of coagulation is increased in a sound field of moderate or high energy. This influence is very pronounced at low electrolyte concentrations, that is in the range of "slow coagulation," while it becomes almost negligible at high electrolyte concentrations, where the coagulation is rapid. In the case of $BaCl_2$, the effect is less pronounced than with the uni-univalent electrolytes. In a sol protected with starch no influence of a sound field was observed. Hermans thinks that the turbulent motions and convection

currents observed in his systems upon irradiation are of no consequence. He goes on to show that cavitation does not cause the increase in the rate of coagulation; nor do heat effects, the kinetic energy of oscillations, orthokinetic coagulation, accumulation, or hydrodynamic interaction between the particles, according to Hermans, offer a satisfactory explanation. The most promising explanation at present is seen in the dipolar interaction of the oscillating particles ("migrating dipoles") which results from the distortion of the electric double layer. The basic consideration in this concept is this: as soon as a particle acquires a velocity with respect to the solvent, its electric double layer loses its spherical symmetry, and a dipole is set up.[98, 99] These dipoles are all lying in the same direction, namely, that in which the sound is propagated. The dipole forces between two particles which follow each other in the same line are attractive, and result in an increased rate of coagulation.

The problem of sound coagulation in truly colloidal solutions still seems to be far from a definite solution; much more experimental work with a variety of solutions is needed.

Influence of Sound Waves in Rheopectic and Dilatant Systems. *Rheopexy.* The time of solidification of certain types of thixotropic systems is shortened by submitting them to a more or less regular movement, such as tapping the test tube containing the sol or slowly rolling it to and fro between the palms of the hands. This phenomenon was called rheopexy.[72] Rheopectic solidification seems to have a relationship to coagulation produced by stirring. Rheopexy is found with any suitable disperse phase only in narrow ranges of concentration and electrolyte content. Rheopexy in sound fields was found by Juliusburger and Pirquet[112] and studied by Burger and Sollner.[221]

A kaolin with particles mostly below 1 μ gave a thixotropic and rheopectic suspension if mixed with 1.6 ml *2N* NaCl per g kaolin. The spontaneous (thixotropic) solidification time was 17 minutes. Upon tapping, this suspension solidified after 30 seconds; in an ultrasonic field of moderate energy only 15 seconds were required for solidification. When it was irradiated for a longer time, zones of accumulation were gradually formed in the paste. Similar results were obtained with gypsum which, like kaolin, has anisometric particles which may be oriented by sound waves.

The particles of Solnhofen slate are approximately cubical; they do not show an orientation effect. However, rapid solidification can be obtained by weak ultrasonic treatment. Therefore, in this case at least, rheopexy is possible without orientation, the accumulation of the discharged particles probably being the only cause of their becoming a paste.

Plastic clay, as used for molding purposes, shows rheopexy when in suitable concentrations, upon tapping the test tube containing the sample or rolling it to and fro. Its particles are anisometric, for dilute suspensions show the orientation phenomenon. It was, however, impossible to detect a shortening of the time of solidification in a sound field under the various conditions of concentration and electrolyte present. The tendency to form stiff gels—as distinct from the rather soft pastes of kaolin and Solnhofen slate—is apparently also the reason that rheopectic solidification with ultrasonics cannot be observed. The individual particles are bound so firmly to one another that the sound waves are unable to move and rearrange them. Zones of accumulation have never been observed in concentrated systems of this kind.

Rheopectic solidification due to ultrasonics has so far not been observed in V_2O_5 sols (containing suitable electrolytes), which otherwise show pronounced rheopexy. The interaction between the particles of these very viscous gel-like V_2O_5 systems is obviously so strong that even orientation by sound waves is rendered impossible.

Generally speaking, a solidifying influence of ultrasonic radiation is to be expected in those rheopectic systems which are more paste- than gel-like, and whose

particles are big enough to be influenced—by being either oriented or moved—individually by the sound waves.[221]

Dilatancy. If the supernatant liquid is poured off the sediments of suspensions, such as quartz powder or sea-sand, the sediments are somewhat moist and soft, when only a slight pressure is applied. However, as soon as a stronger, one-sided, pressure is applied, a sediment of this kind becomes dry and hard and offers considerable resistance to a penetrating instrument; as soon as the pressure ceases, the whole mass usually becomes moist again. This phenomenon, called dilatancy, was explained by Osborne Reynolds [184] in the following way: in moist sand the particles tend to become very closely packed. If an external force displaces the particles, the packing becomes looser and consequently water is sucked in, thus leaving the sediment apparently dry and hard. While a system normally exhibiting dilatancy is exposed to ultrasonics it is possible to drive an instrument through the sand to the bottom of the vessel without much resistance. Other experiments show that the sand under the influence of the ultrasonic radiation is more closely packed than after being left undisturbed for quite a long time, or after being tapped frequently to promote settling.

It seems to hold true generally that dilatant and, therefore, non-plastic systems can be rendered somewhat plastic for the duration of intense sound treatment.

It can readily be shown that cavitation is not connected with the influence of ultrasonics on dilatancy; quite obviously the effect of sound waves on dilatancy is due to the ponderomotive action of the sound waves. The particles when oscillating and moving under the influence of sound waves do not stick to each other, but may move relatively freely over one another as an instrument penetrates the sediment.

Degassing of Liquids. If a sound wave travels through a liquid which is saturated with gas, a part of the gas is driven out and unites in small bubbles, as first described by Dörsing.[84] What happens to these gas bubbles once they are formed was discussed above (page 348). The degassing effect has been studied carefully by Boyle and collaborators.[14, 15, 16, 17, 18, 19, 21, 22] Boyle is inclined to see the explanation of the phenomenon in the coalescence of minute gas bubbles mainly adhering to dust particles which are supposed to be always present in a liquid. These gas bubbles coalesce on sonic irradiation and thus become visible. The bubbles originally present or enlarged by coalescence obviously could grow when the liquid is stretched locally by a sound wave travelling through it. Boyle does not think that this stretching of the liquid alone could account for the appearance of gas bubbles. It seems to the author that the true explanation should emphasize these considerations. If a sound wave passes through a liquid which is saturated with gas, the liquid is stretched periodically, and in this period it is supersaturated with respect to the gas. The assumption of the universal presence of minute (and therefore unstable) gas bubbles seems most unlikely and is not supported by other evidence. This is another phenomenon which needs careful reinvestigation.[166, 224] In many experiments, particularly those involving stationary wave patterns and related phenomena, the occurrence of degassing is a rather disturbing factor; this difficulty can be overcome by working with partially degassed liquids, or by the application of a small outside pressure on the liquid.

The degassing of liquid and fused masses like melted metals and glass was suggested by Krüger [121] and others.[116, etc.] In these cases, however, the degassing may easily be more an accumulation and coagulation of preformed gas bubbles than a primary true degassing. That this may occur even in very viscous systems was shown early by Freundlich, Rogowski and Sollner [76] in the case of very viscous technical viscose solutions. The degassing of glass, etc., may easily assume practical importance in the future.

Thermal Effects of Sound. Intense sound treatment in liquid and solid systems is always accompanied by considerable thermal effects which are very greatly increased

in the presence of extended phase boundaries, as they occur in micro-heterogeneous systems.

In pure liquids and in homogeneous solutions the thermal effect is due to the absorption of sound energy which is intimately connected to the viscosity of the liquid. Wood and Loomis [256] point out that the kinematic coefficient of viscosity increases as the square of the frequency. Thus sound absorption is very pronounced at medium-high ultrasonic frequencies (about 200,000 cycles per second); temperature increases in irradiated liquid samples of up to 1° C in three seconds have been reported by Wood and Loomis; 5 to 10° C temperature rises are common and usually unavoidable when studying emulsification, etc. Frequencies of 500,000, and more so of 1,000,000 cycles per second, are so strongly absorbed that their propagation through the irradiated systems is very limited. This is one of the main reasons why these higher frequencies have not been used to any extent for colloid work, for which high intensities are ordinarily required. A good discussion of the problem of heat evolution was given by Richards. [187] Measurements with different substances have been reported, *e.g.*, by Dognon and Biancani.[62]

It is well known that the propagation of sound is much poorer in a liquid containing gas bubbles than in their absence; vibrational energy is dissipated and transformed into heat. Emulsions and suspensions heat up rapidly if irradiated, many times faster than do the corresponding homogeneous macro-phases. The importance of interfaces in the heating effect has been very clearly demonstrated for the case of solids by Wood and Loomis. A block of newly formed ice (distilled water frozen in a beaker by ice and salt) was subjected for two minutes to the action of intense sound waves in a beaker of ice water containing numerous small fragments of ice which kept the temperature of the water at 0° C. At the end of the sonic treatment, the block of ice, on being squeezed between the fingers, broke up into small fragments, showing that liquefaction had taken place throughout the mass of the ice. With natural ice (pond ice) this experiment could not be duplicated. Wood and Loomis ascribed this to the fact that in the latter case one was dealing with a single crystal, whereas in the case of artificial ice we have a mass of interlocking crystals, the heating taking place at the crystal interfaces.

Some early observations on heterogeneous systems were reported by Freundlich, Rogowski and Sollner,[76] who found with macroscopic objects (glass beads, lead shot) in water an increasing heat effect with decrease in size. In heterogeneous systems the heat effect is due to a number of factors, the relative importance of which depends on the system under consideration. There is first the absorption of sound in the medium of dispersion and the disperse phase; gross stirring, the irregular reflection and the scattering of sound on the interfaces medium of dispersion—disperse phase; the friction between the particles and the oscillating dispersion medium. From the discussion on pages 357 to 359, it is easy to see how the last-mentioned factors must depend very strongly on the particle size; systems with very small particles, as true solutions, scatter but little energy, the solution acts as a homogeneous medium. With macroscopic particles only a limited area of interfaces can be accommodated in a given volume. One may assume that maximum heat effects with medium-high ultrasonic frequencies are obtained in systems with microscopic particles.

There is one additional factor not mentioned so far, which under the conditions of colloid work frequently comes into play: cavitation. The formation of cavities consumes considerable energy which finally appears as heat, and the cavities, as long as they exist, prevent the smooth flow of sound energy by reflection and scattering. A detailed investigation of the heat effect in heterogeneous systems has never been carried out, for such a study does not seem to promise results of great or general interest which could not be obtained more easily in another way. (Compare also [188].)

The generation of heat is actually a disturbing factor in many investigations with sound waves, particularly in the range of ultrasonic frequencies. Cooling is only a partial remedy; by cooling, the average temperature of an irradiated system may be kept within reasonable limits; the local effects at the phase boundaries, however, cannot be suppressed. There is little doubt that this local increase in temperature accelerates many dispersing actions of sound waves. How intense this local heating may be was demonstrated early by Freundlich, Rogowski and Sollner,[76] who irradiated crystals of gallium in water with strong ultrasonic waves; a concentrated suspension of gallium was obtained, though the crystals were not liquefied by the heating of the system, but kept their shape. Close inspection showed that the crystals were melted at their surfaces; this indicates that the gallium was actually dispersed in the liquid state, being first locally liquefied by the action of the sound waves. Many similar observations can be found in the literature.

The fact that considerable heat development cannot be prevented in many phases of colloid work is one of the drawbacks of the application of sound methods in this field.

It has been suggested [75, etc.] that one might utilize ultrasonic radiation for medical purposes, such as the heating of joints, the marrow inside intact bones, etc.

Chemical Effects, Crystallization, etc. A discussion of the chemical action of sound waves is, strictly speaking, outside the framework of this review and can be referred to only briefly. The chemical effects can be classified arbitrarily into: (a) oxidation reactions which are usually accompanied by some light effects; (b) depolymerization which was discussed on pages 354 to 366; (c) the destruction of labile states, the influencing of crystallization, the acceleration of reactions; (d) electrochemical actions, such as the neutralization of passivity.

Oxidation. Schmitt, Johnson and Olson [242] first showed that strong ultrasonic irradiation of water containing oxygen leads to the formation of H_2O_2. This, as established later, is due to the activation of oxygen by the collapse of cavities. In the presence of suitable reagents the activated oxygen, of course, may bring about many varied oxidation reactions. The activation of oxygen is accompanied by the emission of light (sonoluminescence). The problem of sonic oxidation and sonoluminescence was reviewed by Newton Harvey,[92] to which review the reader is referred. (For further references see: [10, 31, 36, 39, 45, 66, 67, 68, 79, 126, 129, 130, 157, 167, 177, 222, 242].) The influence of ultrasonic irradiation of the photographic plate (increase of blackening, occurrence of stationary wave patterns, etc.) is probably due mainly to oxidation.[184, 187, 188, 141, etc.] The oxidation effect may also play some role in the case of the improved photographic emulsions reported by Claus and collaborators (see page 352). Whether or not oxidation reactions play a role in the depolymerization of high polymers is not known. (For the acceleration of hydrolysis etc. see [197, 248, 251].)

Labile states may frequently be changed to more stable ones. Such effects as the sudden boiling of superheated liquids [256] or the detonation of highly sensitive explosives when subjected to sound waves [12, 186, 256] do not require further comment.

Of greater interest here is the spontaneous crystallization from supersaturated solutions which frequently occurs; its mechanism is unknown. (See [1, 5, 8, 58, 59, 204, 210, 256].) This effect has considerable metallurgical interest.[35, 65, 202, 208, 237, 238, 239] The crystalline structure of metals and alloys is changed by sonic treatment during solidifaction; the grain size is generally reduced and the growth of dendrites sometimes favored; however, the results of different authors vary widely, probably due to the difference in the systems investigated and the different sound frequencies used.

Masing and Ritzau,[142] and Schmid and Ehert [237] have prepared dispersions of one immiscible metal in another (Pb in Al, Ca in Si, etc.), by the application of intense sound waves. The mechanism causing these dispersions is unknown.

It seems possible that sonic methods may soon find practical application in attempts to improve the properties of metals and alloys.

Electrochemical effect. A number of electrochemical actions of sound waves have been reported, foremost among them the neutralization of passivity by sonic treatment, and changes in electrode potentials. The neutralization of passivity finds its obvious explanation in the destruction of protecting layers, undoubtedly due to cavitation. If the protecting layers are sufficiently damaged, the electrode potential changes rather abruptly to the reversible value. For further observations and discussion we must refer to the literature.[145, 146, 150, 151, 152, 153, 174, 175, 281, 282, 285, 286, 287, 257]

Biological Effects. A discussion of the biological effects of sound is still further from the central theme of this review than are the chemical effects sketched in the preceding section; a few general remarks must suffice here. If one considers the different disruptive and destructive, ponderomotive, and chemical actions of intense sound waves discussed, it is small wonder that living systems too may be affected greatly by intense sound treatment. Intense irradiation may lead to a destruction of the organized cell structure and even to complete disintegration, as first shown by Wood and Loomis.[256] A very great number of observations have been reported; at present, however, an attempt to coördinate this material may be premature. Somewhat arbitrarily the reader is referred to the work of Harvey and collaborators,[90, 91, 94, 95, 96, 97] and Chambers and collaborators[38, 42, 44, 47, 93] etc. The reviews of Dognon and Biancani,[62, 63] Harvey[91] and the fairly recent one by Hiedemann[106] give a useful picture of the application of high-intensity sound waves to biology.

Acknowledgments

The author wishes to thank Dr. E. S. Fetcher, Jr. and Dr. M. B. Visscher for reading the manuscript; the editor of the *Transactions of the Faraday Society* for permission to use freely his papers published therein and to reproduce a number of photographs; the editor of the *Journal of Physical Chemistry* for reproducing some figures; the editor of the *Journal of Dairy Science* for permission to quote verbatim an article by Dr. L. A. Chambers and to reproduce Fig. 4; and the editors of the *Proceedings of the Society for Experimental Biology and Medicine* and Dr. L. A. Chambers for the permission to reproduce Fig. 3. Acknowledgment is also made for a verbatim quotation from a paper by H. W. St. Clair, presented at the dedication of Bureau of Mines Experiment Station, Salt Lake City, Utah.

References

1. Altberg, W. J., *Priwda (U.S.S.R.)*, **25**, 64 (1936).
2. Andrade, E. N. da C., *Phil. Trans. Roy. Soc. London (A)*, **230**, 413 (1932).
3. —, *Trans. Farad. Soc.*, **32**, 1111 (1936).
4. Ardenne, E. v., *Funktechn. Montashefte*, 285 (1935).
5. Belynski, S. W., *J. Exptl. Theoret. Phys. (U.S.S.R.)*, **6**, 1176 (1936).
6. Bergmann, L., "Ultrasonics and their scientific and technical Applications," (translated by H. Stafford Hatfield), London, Bell & Sons, Ltd., 1938.
7. —, "Der Ultraschall und seine Anwendung in Wissenschaft und Technik," 2nd. Ed., Berlin, VDI-Verlag, 1939.
8. Berlaga, R. Y., *J. Exptl. Theoret. Phys. (U.S.S.R.)*, **9**, 1397 (1939).
9. Beuthe, H., *Z. phys. Chem.*, **A163**, 161 (1933).
10. —, Furbach, E., and Sörensen, C., *Akust, Z.*, **4**, 209 (1939).
11. Bjerknes, V., "Vorlesungen über hydrodynamische Fernkräfte nach C. A. Bjerknes' Theorie," Leipzig, J. A. Barth, 1900/02.
12. Bobolev, V., and Khariton, Y., *Acta Physicochim. (U.S.S.R.)*, **7**, 146 (1937).
13. Boutaric, A., and Bouchard, J., *Bull. soc. chim.*, **51**, 543 (1932).
14. Boyle, R. W., *Trans. Roy. Soc. Canada (3)*, **16**, 157 (1922).
15. —, *Nature*, **120**, 476 (1927).
16. —, *Science Progress*, **23**, 75 (1928).
17. —, and Field, G. S., *Canadian J. Res.*, **6**, 102 (1932).
18. —, —, and Froman, D. K., *Nature*, **129**, 693 (1932).
19. —, and Froman, D. K., *Canadian J. Res.*, **1**, 405 (1929).
20. —, and Lehmann, J. F., *Trans. Roy. Soc. Canada*, **19**, 159 (1925).
21. —, and Taylor, G. B., *Phys. Rev. (2)*, **27**, 518 (1926).
22. —, —, and Froman, D. K., *Trans. Roy. Soc. Canada*, **23**, 187 (1929).
23. Brandt, O., *Z. physik. chem. Unterr.*, **50**, 1 (1936).
24. —, *Kolloid-Z.*, **76**, 272 (1936).

25. —, *Ibid.*, **81**, 2 (1937).
26. —, and Freund, H., *Z. Phys.*, **92**, 385 (1934).
27. —, —, *Ibid.*, **94**, 348 (1935).
28. —, —, *Bl. Unters. u. Forsch. Instr.*, **9**, 57 (1935).
29. —, —, *Z. physik. chem. Unterr.*, **48**, 259 (1935).
30. —, —, *Z. Phys.*, **95**, 415 (1935).
31. Bresler, S. E., *Acta Physicochim.* (*U.S.S.R.*), **12**, 323 (1940).
32. Brohult, S., *Nature*, **140**, 805 (1937).
33. Brown, E. P., *Milk Plant Monthly*, **30**, 52 (1941).
34. Bücks, K., and Müller, H., *Z. Phys.*, **84**, 75 (1933).
35. Cerovska, J., *Strojnicky Obzor*, **19**, 145, 185 (1939).
36. Chambers, L. A., *Phys. Rev.*, **49**, 881 (1936).
37. —, *J. Dairy Sci.*, **19**, 29 (1936).
38. —, *J. Biol. Chem.*, **117**, 639 (1937).
39. —, *J. Chem. Phys.*, **5**, 290 (1937).
40. —, and Flosdorf, E. W., *J. Am. Chem. Soc.*, **55**, 3051 (1933).
41. —, —, *J. Bact.*, **31**, 570 (1936).
42. —, —, *Proc. Soc. Exp. Biol. and Med.*, **34**, 631 (1936).
43. —, —, *J. Biol. Chem.*, **114**, 75 (1936).
44. —, —, *Science*, (Feb. 21, 1941).
45. —, —, and Malisoff, W. M., *J. Am. Soc.*, **58**, 1069 (1936).
46. —, and Gaines, N., *J. Cell. Comp. Physiol.*, **1**, 451 (1932).
47. —, and Weil, A. J., *Proc. Soc. Exp. Biol. and Med.*, **38**, 924 (1938).
48. Clair, H. W. St., *U. S. Bur. Mines, Rept. of Investigations 3400*, 51 (1938).
49. —, *Conf. Met. Res. Met. Div., Bur. Mines* (Salt Lake City) 99 (May, 1940).
50. Claus, B., *Z. techn. Phys.*, **15**, 74 (1934).
51. —, *Ibid.*, **16**, 80 (1935).
52. —, *Ibid.*, **16**, 109 (1935).
53. —, *Ibid.*, **16**, 202 (1935).
54. —, and Schmidt, E., *Kolloid-Beih.*, **45**, 41 (1936).
55. Cone, W. M. H., Tartar, H. V., and Taylor, T. I., *J. Phys. Chem.* **36**, 735 (1932).
56. Dangers, H. W., *Z. Phys.*, **97**, 34 (1935).
57. Daniewski, W., *Acta Phys. Polonica*, **2**, 45 (1933).
58. Danilov, V. I., and Teverovskii, B. M., *J. Exptl. Theoret. Phys.* (*U.S.S.R.*), **10**, 1305 (1940).
59. —, Pluzhnik, E. E., and Teverovskii, B. M., *Ibid.*, **9**, 66 (1939).
60. Dean, R. S., *U. S. Bur. Mines, Repts. of Investigations, 3480* (1940).
61. Demann, W., and Asbach, H. R., *Tech. Mitt. Krupp Forschungsber.*, **3**, 12 (1940).
62. Dognon, A., Biancani, E. and H., "Ultra-Sons et Biologie," Gauthier-Villars, Paris, 1937.
63. —, —, *Radiologica*, **3**, 40 (1938).
64. Dörsing, K., *Ann. Phys.* (*4*), **25**, 227 (1908).
65. Förster, F., and Scheil, E., *Z. Metallkunde*, **28**, 245 (1936).
66. Frenkel, Y., *Acta Physicochim.* (*U.S.S.R.*), **12**, 317 (1940).
67. —, and Bresler, S. E., *Rec. Chem. Prog.*, **2**, 12 (1941).
68. Frenzel, H., and Schultes, H., *Z. phys. Chem.*, **27**, 421 (1934).
69. Freundlich, H., and Gillings, D. W., *Trans. Farad. Soc.*, **34**, 649 (1938).
70. —, —, *J. Chem. Soc.*, 546 (1938).
71. —, —, *Trans. Farad. Soc.* **35**, 319 (1939).
72. —, and Juliusburger, F., *Ibid.*, **31**, 920 (1935).
73. —, and Sollner, K., *Ibid.*, **32**, 966 (1936).
74. —, Rogowski, F., and Sollner, K., *Z. phys. Chem.*, (*A*), **160**, 469 (1932).
75. —, —, —, *Klinische Wochenschrift*, **11**, 1512 (1932).
76. —, —, —, *Kolloid-Beih.*, **37**, 223 (1933).
77. Gaines, N., *Physics*, **3**, 209 (1932).
78. Gies, J. R., *Z. Ver. deut. Ing.*, **6**, 177 (1938).
79. Gohr, H., and Schultes, H., *Angew. Chem.*, **49**, 420 (1936).
80. Gorbatschew, S. W., *Bull. Acad. Sci.* (*U.S.S.R.*), 843 (1936).
81. —, and Severny, A. B., *Kolloid-Z.*, **73**, 146 (1935).
82. Gottschalk, V. H., and Clair, H. W. St., *Mining and Met.*, **18**, 224 (1937).
83. Gruetzmacher, J., *Z. Phys.*, **96**, 342 (1935).
84. —, *Z. techn. Phys.*, **17**, 166 (1936).
85. Hartmann, *J. Phys. Rev.* (*2*), **20**, 114, 719 (1922).
86. —, "Modus operandi of the air-jet pulsator," (Vorträge auf d. Geb. d. Aerodynamik u. verw. Geb. Aachen 1929), Berlin, J. Springer.
87. —, *Verh. 3. Internat. Kongr. f. techn. Mech. Stockholm*, 3, 329 (1931).
88. —, *Phil. Mag.* (*7*), **11**, 926 (1931).
89. —, *J. Phys. Radium* (*7*), **2**, 49 (1936).
90. Harvey, E. N., *Am. J. Physiol.*, **91**, 284 (1929).
91. —, *Biol. Bull.*, **59**, 306 (1930).
92. —, *J. Am. Chem. Soc.*, **61**, 2392 (1939).
93. —, and Chambers, L. A., *J. Morph. Physiol.*, **52**, 155 (1931).
94. —, and Loomis, A. L., *Nature*, **121**, 622 (1928).
95. —, —, *J. Bact.*, **17**, 373 (1929).
96. —, —, *J. Gen. Physiol.*, **15**, 147 (1931).
97. —, Harvey, E. N., and Loomis, A. L., *Biol. Bull.*, **55**, 459 (1928).
98. Hermans J. J., *Phil. Mag.*, **25**, 426 (1938).
99. —, *Ibid.*, **26**, 674 (1938).
100. —, *Rec. trav. chim.*, **57**, 1349 (1938).
101. —, *Ibid.*, **58**, 139 (1939).
102. —, *Ibid.*, **58**, 164 (1939).
103. Heymann, E., *Trans. Farad. Soc.*, **31**, 846 (1935).
104. Hiedemann, E., *Kolloid-Z.*, **77**, 168 (1936).
105. —, *Trans. Farad. Soc.*, **32**, 1120 (1936).
106. —, "Grundlagen und Ergebnisse der Ultraschallforschung," Berlin, W. de Gruyter and Co., 1939.
107. —, and Brandt, O., *Kolloid-Z.*, **75**, 129 (1936).
108. —, —, *Trans. Farad. Soc.*, **32**, 1101 (1936).
109. —, —, and Freund, H., *Kolloid-Z.*, **77**, 103 (1936).
110. —, —, —, *Z. Phys.*, **104**, 511 (1937).
111. —, —, —, *Verh. dtsch. phys. Ges.* (*3*), **18**, 28 (1937).
112. Juliusburger, F., and Pirquet, A., *Trans. Farad. Soc.*, **32**, 445 (1936).
113. Kimura, O., *J. Chem. Soc. Japan*, **56**, 842 (1935).

114. King, L. V., *Proc. Roy. Soc. London* (*A*), **147**, 212 (1934).
115. —, *Ibid.*, **153**, 1 (1935).
116. Klein, V., *Glastech. Ber.*, **16**, 232 (1938).
117. König, W., *Wiedemann's Ann. Phys.*, **42**, 353, 549 (1891).
118. —, *Z. phys. chem. Unterr.*, **8**, 191 (1895).
119. Kremnev, L., *Kolloid-Z.*, **67**, 171 (1934).
120. —, *Ibid.*, **68**, 16 (1934).
121. Krüger, F., *Glastechn. Ber.*, **16**, 233 (1938).
122. Kundt, A., *Pogg. Ann. Phys.*, **127**, 497 (1866).
123. —, *Ibid.*, **128**, 337, 496 (1866). 6
124. —, and Lehmann, O., *Ibid.*, **153**, 1 (1874).
125. Laidler, T. J., and Richardson, E. G., *J. acoust. Soc. Amer.*, **9**, 217 (1938).
126. Levshin, V. I., and Rzhevkin, S. N., *Compt. rend. acad. sci.*, (*U.S.S.R.*), **16**, 399 (1937).
127. Lewis, E. P., and Farris, L. P., *Phys. Rev.*, **6**, 491 (1915).
128. Liu, S., and Wu, H., *Proc. Soc. Exp. Biol. and Med.*, **28**, 782 (1930).
129. —, —, *J. Am. Chem. Soc.*, **54**, 791 (1932).
130. —, —, *Ibid.*, **56**, 1005 (1934).
131. Mach, E., and Salcher, P., *Wiedemann's Ann. Phys.*, **41**, 144 (1890).
132. Mache, H., *Meteorol. Z.*, **50**, 393 (1933).
133. Marinesco, N., *Compt. rend.* (*Paris*), **194**, 1824 (1932).
134. —, *Ibid.*, **196**, 346 (1933).
135. —, *Bull. soc. roumaine phys.*, **36**, 181 (1934).
136. —, *Compt. rend.* (*Paris*), **201**, 1187 (1935).
137. —, *J. Chim. Phys.*, **33**, 99.
138. —, *Compt. rend.* (*Paris*), **202**, 757 (1936).
139. —, "Propriétés piézo-chimiques, physiques et biophysiques des ultrasons. Actualités sciéntiques et industrielles No. 522. Technique des ondes élastiques des haute fréquence No. 523. Destruction des microorganismes. Préparation des colloides a basses températures. Réactions explosives, réactions photochimiques," Paris, Hermann & Cie., 1937.
140. —, and Reggiani, M., *Compt. rend.*, **220**, 548 (1935).
141. —, and Trillat, J. J., *Ibid.*, **196**, 858 (1933).
142. Masing, G., and Ritzau, G., *Z. Metallk.*, **28**, 293 (1936).
143. Meissner, A., *Phys. Z.*, 28, 621 (1927); *Naturwiss*, **17**, 25 (1929).
144. Moriguchi, N., *J. Chem. Soc. Japan*, **54**, 949 (1933).
145. —, *Ibid.*, **54**, 1047 (1933).
146. —, *Ibid.*, **55**, 749 (1934).
147. Morozov, A. A., *Byull. Ivanovo Nauch. Issledovatel. Inst.*, **4-5**, 105 (1938).
148. Müller, H., *Kolloid-Beihefte*, **27**, 223 (1928).
149. Myers, R. J., and Blumberg, H., *Proc. Soc. Exp. Biol. and Med.*, **35**, 79 (1936).
150. Nikitin, L. V., *Compt. rend.* (*Leningrad*), **4**, 309 (1934).
151. —, *Chim. Shurn.* (*A*) *Shurn. obtschtschei Chim.*, **6**, 1401 (1936).
152. —, *Compt. rend.* (*Moscow*), **2**, 67 (1936).
153. —, *J. Gen. Chem.* (*U.S.S.R.*), **6**, 1393 (1936).
154. Nordlund, I., *Kolloid-Z.*, **26**, 121 (1918); Dissertation, Upsala, 1918.
155. Oka, S., *Kolloid-Z.*, **87**, 37 (1939).
156. —, *Z. Phys.*, **116**, 632 (1940).
157. Olson, A. R., and Garden, N. B., *J. Am. Chem. Soc.*, **54**, 3617 (1932).
158. Ono, S., *Rev. Phys. Chem. Japan*, **14**, 25 (1940).
159. —, *Ibid.*, **14**, 101 (1940).
160. —, *J. Chem. Soc. Japan*, **61**, 997 (1940).
161. Oseen, C. W., "Hydrodynamik," Leipzig, 1927.
162. Ostrovskii, E. P., *Compt. rend. Acad. Sci.* (*U.S.S.R.*), **14**, 491 (1937).
163. Parker, R. C., *Trans. Farad. Soc.*, **32**, 1115 (1936).
164. —, *Proc. phys. Soc.*, **49**, 95 (1937).
165. Patterson, H. S., and Cawood, W., *Nature*, **127**, 667 (1931).
166. Paunow, P. G., *Ann. univ. Sofia, II, Faculté phys.-math.*, **35**, 179 (1939).
167. —, *Compt. rend.*, **209**, 33 (1939).
168. Pearson, E. B., *Proc. phys. Soc.*, **47**, 136 (1935).
169. Petralia, S., *Nuovo cimento*, **17**, 498 (1940).
170. Pierce, G. W., Brit. Pat. No. 283, 116 (1928).
171. —, *Proc. Am. Acad. Arts Sci.* (*Boston*), **63**, 1 (1928).
172. —, *Proc. Inst. Radio Eng.*, **17**, 42 (1929).
173. —, *J. acoust. Soc. Am.*, **9**, 185 (1938).
174. Piontelli, R., *Atti Accad. Lincei*, **27**, 357 (1938).
175. —, *Ibid., Classe sci. fis. mat. nat.*, **27**, 581 (1938).
176. Pohlman, R., *Z. Phys.*, **107**, 497 (1937).
177. Polotskii, I. G., *J. Gen. Chem.* (*U.S.S.R.*), **8**, 1691 (1938).
178. Popow, K., *Massloboino-shirowoje Djelo*, **12**, 397 (1936).
179. Porter, B. H., *Ind. Eng. Chem., Anal. Ed.*, **12**, 748 (1940).
180. Protas, I. R., *Kinofotokhim. Prom.*, **1**, 38 (1940).
181. —, *Kinofotokhim. Prom.*, **5**, 43 (1940).
182. Randolph, A., *Z. Phys.*, **103**, 414 (1936).
183. Rayleigh, Lord, *Phil. Mag.* (*6*), **34**, 94 (1917).
184. Reynolds, O., *Phil. Mag.* (*5*), **20**, 469 (1885); *Nature*, **33**, 429 (1886).
185. Richards, W. T., *Proc. Nat. Acad. Sci.*, **15**, 310 (1929).
186. —, *J. Am. Chem. Soc.*, **51**, 1724 (1929).
187. —, *Proc. Nat. Acad. Sci.*, **17**, 611 (1931).
188. Rytov, S. M., Vladimirskii, V. V., and Galanin, M. D., *J. Exptl. Theoret. Phys.* (*U.S.S.R.*), **8**, 614 (1938).
189. Rzhevkin, S. N., and Ostrovskii, E. P., *Act. phys.* (*U.S.S.R.*), **1**, 740 (1935).
190. Salisbury, W. W., and Porter, C. W., *Rev. Sci. Instr.*, **10**, 142 (1939).
191. —, —, *Ibid.*, **10**, 269 (1939).
192. Sata, N., *Kolloid-Z.*, **71**, 48 (1935).
193. —, *Bull. Chem. Soc. Japan*, **12**, 536 (1937).
194. —, *Kolloid-Z.*, **81**, 182 (1937).
195. —, *Ibid.*, **87**, 186 (1939).
196. —, *Ibid.*, **88**, 182 (1939).
197. —, and Harisaki, Y., *Bull. Chem. Soc. Japan*, **15**, 180 (1940).
198. —, and Naruse, N., *Kolloid-Z.*, **86**, 102 (1939).
199. —, —, *Ibid.*, **89**, 341 (1939).

200. —, and Watanabe, S., *Ibid.*, **73**, 50 (1935).
201. —, —, *Ibid.*, **78**, 277 (1937).
202. Seemann, H. J., *Metallwirtschaft*, **15**, 1067 (1936).
203. Sesjulinski, B. M., and Tumanski, S. S., *J. techn. Phys. (U.S.S.R.)*, **7**, 1922 (1937).
204. Shablykin, P. N., *Mineral. Syr'e*, **12**, 28 (1937); *Khim. Referat. Zhur.*, **1**, 23 (1938).
205. Sibata, K., *J. Soc. Rubber Ind. Japan*, **11**, 232 (1938).
206. —, *Repts. Tokyo Imp. Ind. Res. Inst. Lab.*, **35**, 1 (1940).
207. Smith, F. D., *Phil. Mag.* (7), **19**, 1147 (1935).
208. Sokoloff, S. J., *Acta physicochim. (U.S.S.R.)*, **3**, 939 (1935).
209. —, *Techn. Phys. (U.S.S.R.)*, **3**, 176 (1936).
210. —, *Ibid.*, **5**, 619 (1938).
211. Sollner, K., *Trans. Farad. Soc.*, **32**, 1119 (1936).
212. —, *Ibid.*, **32**, 1532 (1936).
213. —, *Ibid.*, **32**, 1537 (1936).
214. —, *Ibid.*, **34**, 1170 (1938).
215. —, *J. Phys. Chem.*, **42**, 1071 (1938).
216. —, and Bondy, C., *Trans. Farad. Soc.*, **31**, 835 (1935).
217. —, —, *Ibid.*, **31**, 843 (1935).
218. —, —, *Ibid.*, **32**, 556 (1936).
219. —, —, *Ibid.*, **32**, 616 (1936).
220. —, and Bull, H. B., *Kolloid-Z.*, **60**, 263 (1932).
221. —, and Burger, F. J., *Trans. Farad. Soc.*, **32**, 1598 (1936).
222. Solov'eva, L. R., *J. Phys. Chem. (U.S.S.R.)*, **9**, 77 (1937).
223. —, *Colloid J. (U.S.S.R.)*, **5**, 289 (1939).
224. Sörensen, C., *Ann. Phys.*, **26**, 121 (1936).
225. Straubel, H., *Phys. Z.*, **32**, 379 (1931).
226. Submarine Signal Company, Boston, "The Submarine Signal Sonic Oscillator for the Production of Sonic (Soft Curd) Milk."
227. Szalay, A., *Z. Phys. Chem.*, **A164**, 234 (1933).
228. —, *Phys. Z.*, **35**, 293 (1934).
229. Szent-Györgi, A., *Nature*, **131**, 278 (1933).
230. Scheibe, A., "Piezoelektrizität des Quarzes," Dresden, Verl. Steinkopff, 1938.
231. Schmid, G., *Angew. Chem.*, **49**, 117 (1936).
232. —, *Z. Elektrochem.*, **44**, 728 (1938).
233. —, *Z. phys. Chem.*, **A186**, 113 (1940).
234. —, *Phys. Z.*, **41**, 326 (1940).
235. —, and Ehret, L., *Z. Elektrochem.*, **43**, 408 (1937).
236. —, —, *Ibid.*, **43**, 597 (1937).
237. —, —, *Ibid.*, **43**, 869 (1937).
238. —, and Roll, A., *Ibid.*, **45**, 769 (1939).
239. —, —, *Ibid.*, **46**, 653 (1940).
240. —, and Rommel, O., *Z. Phys. Chem.*, **A185**, 97 (1939).
241. —, —, *Z. Elektrochem.*, **45**, 659 (1939).
242. Schmitt, F. O., Johnson, C. H., and Olson, R. A., *J. Am. Chem. Soc.*, **51**, 370 (1929).
243. Schumb, W. C., and Rittner, E. S., *J. Am. Chem. Soc.*, **62**, 3416 (1940).
244. —, Peters, H., and Milligan, L. H., *Metals and Alloys*, **3**, (May, 1937).
245. Tatum, G. R., *J. Acoust. Soc. Am.*, **8**, 210 (1937).
246. Thieme, E., *Phys. Z.*, **39**, 384 (1938).
247. Tumanskii, S. S., *Colloid J. (U.S.S.R.)*, **6**, 603 (1940).
248. —, and Maksimova, Ch., *Ibid.*, **5**, 517 (1939).
249. —, and Zezyulinskii, V. M., *J. Phys. Chem. (U.S.S.R.)*, **11**, 801 (1938).
250. Tuorila, P., *Kolloidchem. Beih.*, **24**, 1 (1927).
251. Urazovskii, S. S., and Polotskii, I. G., *J. Gen Chem. (U.S.S.R.)*, **10**, 812 (1940).
252. —, —, *Colloid J. (U.S.S.R.)*, **6**, 779 (1940).
253. Vigoureux, P., "Quartz Resonators and Oscillators," London, H. M. Stationary Office, 1931.
254. Wagenschein, M., *Ann. Phys.* (4), **65**, 461 (1921).
255. Wiegner, G., *Z. Pflanzenernährung, Düngung und Bodenkunde* (*A*), **11**, 185 (1928); *Trans. J. Soc. Chem. Ind.*, **50**, 55 (1931); *Kolloid-Z.*, **58**, 157 (1932).
256. Wood, R. W., and Loomis, A. L., *Phil. Mag.* (7), **4**, 417 (1927).
257. Young, W. T., and Kersten, H. *J. Chem. Phys.*, **4**, 426 (1936).

The Cyclotron and Some of Its Uses *

JEROME ALEXANDER

Many important advances in science and technology radiate from the discovery of new facts, new apparatus and new methods which, though at first laboratory curiosities, may later enter wide use in the home, hospital or factory. It is only a

* This paper was assembled by the Editor mainly from publications of Professors Ernest O. Lawrence, John H. Lawrence, and Joseph G. Hamilton, of the Crocker Radiation Laboratory, University of California, Berkeley, Cal., who have also kindly consented to check over the manuscript. Figure 1 is reproduced from *Electrical Engineering*, July, 1942, and the remaining illustrations are from cuts kindly supplied by Rutgers University Press from "Molecular Films, The Cyclotron and the New Biology" (1942) by Hugh Stott Taylor, Ernest O. Lawrence and Irving Langmuir. The Editor expresses sincere thanks for this coöperation.

few years now since mechanical refrigeration, founded on the experiments of physicists, supplanted the iceman in so many homes; and out of the "Edison effect" of leaking electrons there developed the radio tube.

The cyclotron, which already means much in physical, biological and medical research, is a magnetic resonance accelerator in which ionized particles are given successive impulses of acceleration by charged electrodes, while being constrained to follow a circular path in a powerful magnetic field.

According to a theorem originally due to Larmor, the angular velocity of a charged particle in a magnetic field is independent of its linear velocity. Figure 1

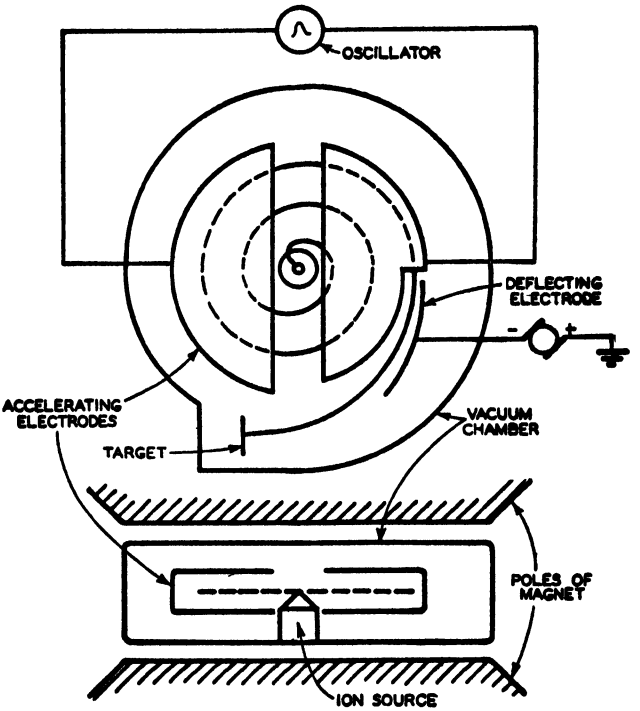


FIGURE 1. Schematic diagram of the cyclotron.

gives a schematic diagram of the cyclotron, above in section, below in elevation. The D-shaped hollow accelerating electrodes (called "dees") are maintained in a hydrogen vacuum (pressure of about 10^{-4} mm of mercury), and are activated by an alternating voltage applied across the electrodes by an oscillator at a frequency of the order of 10 megacycles. Consider the behavior of a single positive ion, liberated, for example, by a heated tungsten filament between the dees. The ion will be attracted toward whichever dee happens to be negatively charged at the moment, and entering the hollow of the electrode (where it is free from electrostatic attractions) follows a semi-circular path in the magnetic field until it reaches the gap between the dees. There it encounters the attractive force of the opposite dee (for the oscillation speed is suitably synchronized) and enters the hollow of this dee with a further increase in velocity. With each passage across the electrode gap, the ion is thus given a new increment of acceleration and follows an ever-widening circle, until near the periphery of one of the dees, where there is a window in a jog, it passes out of the hollow electrode under the pull of a deflecting electrode and strikes a target with tremendous velocity.

Assuming an alternating current of 50 kv across the dees, after 50 revolutions (100 gap crossings) an ion of unitary charge (a proton $_1H^1$, or a deuteron, $_1H^2$) would have an energy of five million electron-volts. Under like conditions an alpha particle (helium nucleus, $_2He^4$) because of its double charge, would acquire twice this energy. The 60-inch cyclotron at the University of California (Berkeley, Cal.), the largest now in operation, gives a deuterium ("heavy hydrogen") beam with an energy of 16 million electron-volts, which produces a lavender luminosity in air for a distance of 4.5 feet after emerging from a special window or a vacuum gate. The 100 million electron-volt cyclotron, now under construction at Berkeley, is expected to give a beam 140 feet long in air.

"The beam in air looks rather pretty, but its appearance hardly suggests its latent powers. However, some conception of the energy in the beam is gained when a steel plate is placed in the path of the beam, for it is immediately melted and cut through. . . . A much more subtle danger, moreover, lurks in it because, as the swiftly moving particles lose their energy, they make nuclear collisions giving rise to penetrating nuclear radiations—the gamma rays and neutron rays, which like x-rays produce harmful and even lethal effects in excessive doses. It is for this reason that the cyclotron is so carefully surrounded by large masses of absorbing material to protect the operators." [17] (Figure 2).

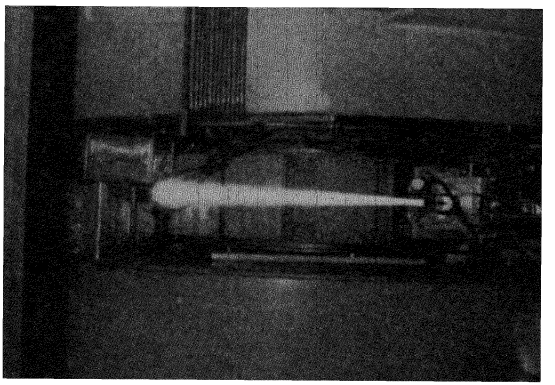


FIGURE 2. The 16 million electron-volt deuteron beam emerging from the cyclotron chamber into the air. An indication of its size may be gained from noting the meter stick placed below the beam.

In 1919 Ernest Rutherford (later Lord Rutherford) demonstrated the first laboratory transmutation by showing that particles (helium nuclei) emitted with energies of 5 to 10 mv by naturally radioactive elements (radium "C," polonium) could, upon collision with the nucleus of a nitrogen atom, cause the ejection of a proton. About five years later P. M. S. Blackett was able to take photographs which indicated that the helium nucleus ($_2He^4$) was captured by the nucleus of the main nitrogen isotope ($_7N^{14}$), forming an unstable grouping which immediately split into two particles; one a proton, the other having a mass of 17. Assuming the now well recognized conservation of charge this change may be thus indicated:

$$_7N^{14} + {}_2He^4 = {}_8O^{17} + {}_1H^1$$

It was later discovered that while 99.76 per cent of oxygen atoms have a mass number (M) of 18, there are two oxygen isotopes, 0.04 per cent with $M = 17$, and 0.20 per cent with $M = 18$.

The "natural" particle projectiles were able to produce transmutations in many among the first score of elements in the periodic table, but the heavier elements resisted change. Scientific advance demanded projectiles of increased energy with which to bombard the resistant atomic nuclei, and Cockcroft and Walton (Cam-

bridge) developed a circuit whereby a pulsating voltage of about 700 kv could be applied to a discharge tube so that the tube could deliver protons having this maximum energy. Lauritsen, of the California Institute of Technology, produced particles having a million volts by using cascaded transformers. Van de Graaff (Massachusetts Institute of Technology) constructed very high voltage electrostatic generators. The tremendous difficulties of handling these high voltages and of constructing tubes to withstand them were cleverly avoided by the cyclotron of Lawrence and Livingston (University of California[1, 8, 12]) which accelerates a particle not once but many times, to produce projectiles having energy sufficient to crash into the nucleus of any atom.

The present uses of the cyclotron stem from the fact that its potent beam is able to produce nuclear changes in all the 92 elements, resulting in transmutation of the atoms into new elemental or isotopic forms. In most cases there are produced unstable nuclei having various periods of half-life, which are radioactive and, on their course to stability, emit subnuclear particles (*e.g.*, neutrons, electrons, positive electrons) as well as intense γ radiation.

If the target of the cyclotron contains beryllium atoms and the projectiles are deuterons (heavy hydrogen nuclei) a very unstable radioactive isotope of boron is formed, which immediately breaks down, liberating *neutrons* with an energy of 21 million electron-volts. Since neutrons are without any net charge, they possess profound penetrating ability, like x-rays, whereas charged particles, *e.g.*, protons, are stopped by a few millimeters of solid material. Energetic neutrons are being widely applied in the medical and biological fields, to treat cancer * and to produce chromosomal changes. Figure 3 shows schematically the liberation of high-energy neutrons.

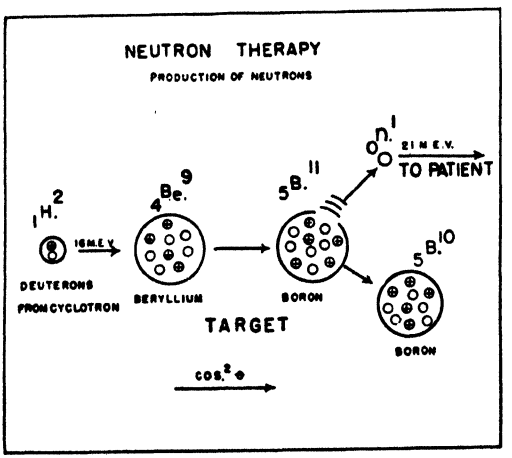


FIGURE 3. The nuclear reaction which is graphically represented here indicates the mode in which the bombardment of beryllium atoms by deuterons results in the release of energetic neutrons.

The potent local effect produced by a neutron beam is shown in Figure 4, which illustrates the intense ionization produced in a Wilson cloud chamber containing air, hydrogen and water vapor. The thick-appearing, dense tracks were produced by recoil protons resulting from collisions of neutrons with hydrogen nuclei; the thin tracks arise from secondary electrons liberated by gamma-rays. Incidentally, although these tracks appear more or less continuous, they really consist of colloidal or near-colloidal water droplets which form about ionized particles when the moving piston causes sudden expansion and chilling of the moist air within the chamber.

The pioneer experiments of G. Hevesy,[9] who employed radium D (a radioactive

* See Podolsky, "The War on Cancer," Reinhold Publishing Corp., 1943.

isotope of lead) to investigate lead metabolism in plants, were abortive because lead as well as bismuth, which was later tried (radium E is radio-bismuth), are not normal plant constituents. Hevesy was able to measure quantitatively the uptake and distribution of the lead by determining the radioactivity of various parts of the plant; for this method is over a million times more sensitive than ordinary chemical or physical analysis, and the traces of metal used were far below the threshold of toxicity.

The discovery of artificial radioactivity by I. Curie and F. Joliot[5] made available radio-nitrogen, radio-silicon, and radio-magnesium; but the discovery of the cyclotron by E. O. Lawrence quickly led to the production of a great number (to date, about 350) of radio-isotopes and radio-elements,* some of which are important

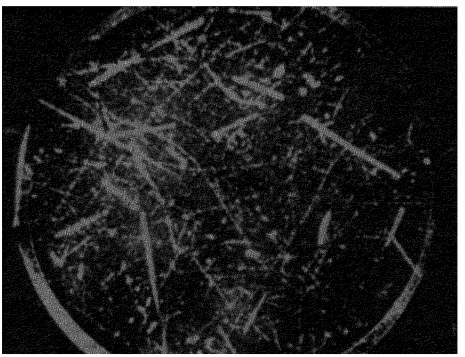


FIGURE 4. Ionization produced in a Wilson cloud chamber filled with a mixture of air, hydrogen and water vapor after bombardment with neutron rays, and gamma-rays from the cyclotron. The thin tracks of ions were produced by secondary electrons liberated by the gamma-rays, while the thick and very dense tracks were produced by the recoil protons resulting from collisions of neutrons with the hydrogen atomic nuclei. This picture demonstrates the more localized and intense ionization produced in tissues by neutrons when compared with X-rays or gamma-rays.

normal constituents of plants and animals, can be produced in "quantity," and have a sufficiently long "life" to make them experimentally valuable, *e.g.*, phosphorus, sodium, iodine, calcium, carbon, and nitrogen. The cyclotron can also produce usable quantities of the rare "triple hydrogen," *tritium*, which has a half-life of about 10-15 years, and can be readily stored and handled by being combined with lithium as the "hydride," lithium tritide (LiT).

Three techniques have been developed to apply the radio-isotopes to medical and biological problems:

(1) Administration of a suitable chemical compound, followed by measurement, with an electroscope or a Geiger counter, of the radiation from each individual tissue.

(2) Autoradiography, first devised by Lacassgne and Lattes,[11] which involves placing portions of the biological material, for suitable periods, within effective range of photographic plates, which on development indicate the loci where radio-elements are fixed.[8, 14, 15]

(3) The *in vivo* technique, developed mainly by J. G. Hamilton, involving the study of the absorption, distribution, and deposition of the various radio-elements in the intact organism, by placing the Geiger counter over the part to be studied, *e.g.*, over the thyroid with radio-iodine.

Investigations with radio-phosphorus, which emits beta rays having a penetration of less than a centimeter in tissue, showed that it was localized mainly in bone marrow and bone. This led J. H. Lawrence to experiment on mice suffering from leukemia. He found that wherever there is leukemic infiltration there is a high phosphorus uptake.[13] Leukemic tumor (lymphoma), a lymphosarcoma, and a mammary gland cancer also showed a higher turnover or uptake of phosphorus than

* See table of elements and their isotopes by Prof. Robley D. Evans, published in this volume.

any of the soft tissues.[10] Cautious exploration showed that in man, leukemia white cells take up more phosphorus than do normal white cells, and that intravenous administration results in a greater effect than oral. Encouraging therapeutic results have followed radio-phosphorus therapy in cases of chronic mylenogenous and chronic lymphatic leukemia. The doses given are small, equivalent to 2 to 4 roentgens of whole body irradiation per day, and are controlled by lethal studies on animals and previous knowledge of x-rays. There are no radiation reactions and the simplicity of dosage avoids frequent trips to the x-ray department. In no sense, however, can radio-phosphorus be considered a cure for leukemia.

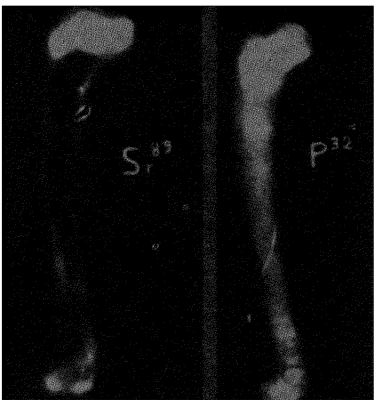


FIGURE 5. A comparison of the accumulation of radio-phosphorus and radio-strontium in the femurs of two different rabbits. The radio-autograph of the left indicates that radiostronium was predominantly accumulated in the marrow.

Polycythemia vera (excess of red blood cells) can be readily controlled by small doses of radio-phosphorus, 5 to 7 millicuries intravenously in the form of sodium radiophosphate, followed three weeks later by a similar dose. Six to eight weeks after the first dose the erythrocyte count begins to fall, and may remain normal for a year without further treatments. Encouraging results are also recorded in generalized lymphosarcomatosis, involving chiefly the lymph nodes, and in multiple myeloma, involving the bone marrow.

The late Charles Pecher, a young Belgian physician, found in preliminary experiments[16] that radio-calcium was selectively adsorbed by the bone rather than by the soft tissues. Since radio-strontium behaved like radio-calcium and was more readily prepared in the cyclotron, Pecher gave rabbits intravenous injections of sodium radio-phosphate and of radio-strontium lactate. As may be seen from Figure 6, radio-phosphorus is indicated in a disease like leukemia, which involves both bone marrow and the soft tissues, whereas radio-strontium would logically be used in diseases involving the skeleton. Radio-strontium lactate is also being used in an attempt to mitigate post-operative cancerous metastases, *e.g.*, in osteoclastic (bone-destroying) invasion of the skeleton by breast cancer, and in osteoplastic (bone-forming) metastases in cancer of the prostate. Since 50 to 1,000 times as much radio-strontium is taken up by the bone as by any of the soft tissues, theoretically it becomes possible to irradiate the skeleton and bone marrow selectively with controlled doses of beta rays. In the prostatic case, treatment was followed by a drop in the high serum phosphatase content to nearly normal, which indicates a certain degree of cessation of troublesome and painful bone formation.[14]

The normal human thyroid gland, weighing about twenty grams, contains from 5 to 20 milligrams of iodine, which is about 5,000 times the average content of the other body tissues. Thyroxine, a potent hormone exerting control over the rate of metabolism, is produced in and liberated by the thyroid, which has remarkable

selective power of adsorbing iodine and radio-iodine. Hamilton and Soley[6] administrated orally to a group of normal persons and to patients with various thyroid disorders, sodium iodide containing 14 milligrams of iodine with a definite quantity (10 — 50 μ C) of radio-iodine; cancerous thyroid tissue took up only traces of radio-iodine.[8]

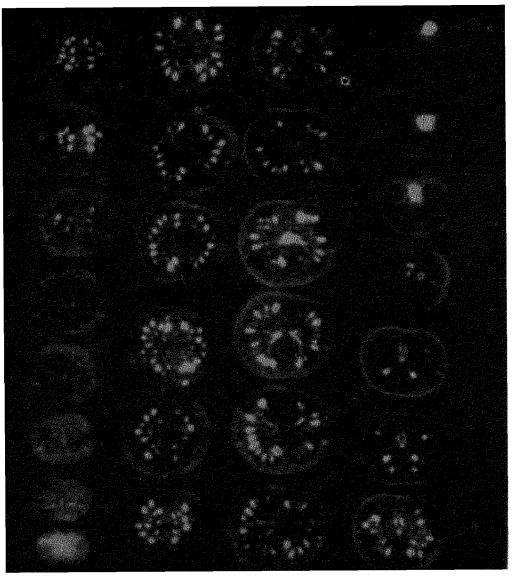


FIGURE 6. The uptake of labelled zinc in the fruit of the tomato plant is demonstrated by the technique of radio-autography. The light areas indicate the regions of maximum deposition of the labelled zinc in slices of the tomato fruit. It can be seen that the seeds apparently accumulate most of the absorbed radiozinc.

The anatomical unit of the thyroid gland is the acinus, with a spherical envelope made of a single layer of specific cells, which encloses a clear, non-cellular material termed "colloid." The grape-like acini, from 20 to 50 microns in diameter, are bound together by numerous strands of connective tissue and are well supplied with blood. Comparing photomicrographs of normal and diseased thyroid sections taken after radio-iodine administration, it was found that normally both the colloid and the thyroid cells take up iodine, while in hyperplastic thyroid tissue it is the colloid that appears to concentrate the recently assimilated iodine to the greater degree. In a case of non-toxic goiter, the labelled iodine appeared to be concentrated in the cells to a greater degree than in the colloid, but in cancerous thyroid the uninvaded tissue took up the iodine, and the cancerous tissue failed to take up any appreciable quantity of the labelled iodine.

Eka-iodine, the newly discovered[4, 7] radioactive halogen element 85 with a half-life of 7.5 hours, behaves similarly to iodine. It is prepared by bombarding metallic bismuth with alpha particles having an energy of 32 mv, and can be separated from the bismuth by vacuum distillation.

It is not possible to do justice, in this brief outline, to such a novel, potent, and rapidly developing field. However, enough has been said to awaken scientists in all fields to the general nature and importance of the cyclotron, and to some of its direct and indirect uses.

References

1. Aebersold, P. C., "The Cyclotron: A Nuclear Transformer," *Radiology*, **39**, 5, 513-540 (1942).
2. Aebersold, P. C., and Lawrence, J. H., "The Pathological Effects of Neutron Rays," *Am. Rev. Phys.*, **4**, 25-48 (1941).
3. Brobecker, W. M., "The Cyclotron," *Elect. Eng.* (July, 1942).
4. Corson, D. R., MacKenzie, K. R., and Segrè, E., *Phys. Rev.*, **58**, 672 (1940).

5. Curie, I., and Joliet, F., *Nature*, **133**, 201 (1934).
6. Hamilton, J. H., and Soley, M. H., *Am. J. Physiol.*, **127**, 557-572 (1939).
7. *Ibid., Proc. Nat. Acad. Sci.*, **26**, 483 (1940).
8. Hamilton, J. G., "The Use of Radioactive Tracers in Biology and Medicine," *J. Appl. Phys.*, **12**, 6, 440-460 (1941).
9. Hevesy, G., *Biochem. J.*, **17**, 439 (1923).
10. Jones, H. B., Chaikoff, I. L., and Lawrence, J. H., *Am. J. Cancer*, **40**, 2 (1940).
11. Lacassgne and Lattes, *Compt. rend. Soc. biol.*, **90**, 352 (1924).
12. Lawrence, E. O., "The Biological Action of Neutron Rays," *Radiology*, **29**, 313-322 (1934); "The Medical Cyclotron of the Wm. H. Crocker Laboratory," *Science*, **90**, 407 (1939); Atoms, New and Old, "Science in Progress," I, Yale Univ. Press, 1939; "The New Frontiers in the Atom," Rept. Smithsonian Inst. for 1941, 163-73.
13. Lawrence, J. H., and Gardner, Wm. U., *Am. J. Cancer*, **33**, 1 (1938).
14. Lawrence, J. H., *Proc. Inst. Med., Chicago*, **12**, 30-49 (1942).
15. Lawrence, J. H., "The New Nuclear Physics and Medicine," *Am. J. Roent. Rad. Therapy*, **48**, 3, 283-301 (1942).
16. Pecher, C., *Proc. Soc. Exptl. Biol. Med.*, **46**, 86 (1941).
17. Taylor, H. S., Lawrence, E. O., and Langmuir, I., "Molecular Films, The Cyclotron and the New Biology," Rutgers University Press, 1942.

The Betatron: A Generator for High-Energy Electrons and X-rays

PROFESSOR D. W. KERST

University of Illinois, Urbana, Ill.

With the advent of intense interest in nuclear physics which followed Cockcroft and Walton's successful attempt to produce atomic disintegrations by artificially accelerated positive ions, many physicists turned their efforts toward the production of generators for high-speed particles. Not all of those devised could be used for the acceleration of both positive ions and electrons. The electrostatic generator of Van de Graaff can produce either positive ion or electron beams in its linear accelerating tube, depending on the type of source which is installed in the high-voltage electrode. Also the voltage multiplier apparatus used in the early work by Cockcroft and Walton could be used for electrons as well as for positive ions. The energy produced by these types of apparatus to date has reached about 4.5 million volts. To increase this voltage, difficulties in maintaining an electrode and operating a vacuum tube at such an extremely high potential must be overcome.

On the other hand, the cyclotron has been able to accelerate the positive ion of heavy hydrogen to an energy of about 16 million volts, and it has been a very successful instrument for the artificial production of radioactive substances and for the production of neutrons by positive ion bombardment. But the cyclotron cannot give its enormous energy to electrons. Consequently electron experiments requiring large currents had been confined to the energies obtained with the other types of accelerators. The cyclotron principle operates successfully only with particles which have a small velocity compared with that of light, and consequently heavy particles are used to obtain high energy. Since electrons of an energy as small as 0.5 million volts have a velocity which is almost 0.9 the velocity of light, they cannot be accelerated to very high energies by the cyclotron method. However, the betatron principle applies to the acceleration of particles at both classical and relativistic energies; and hence an electron can be started at a very low velocity and acquire a velocity which is within a fraction of a per cent of that of light. In the present betatron 20 million volts are reached.

This method of acceleration, which was successfully accomplished for the first time at the University of Illinois,[1, 2, 3] shows promise not only of having practical value for x-ray and electron therapy and many other purposes to which x-rays can

be put, but also of opening new fields of research in nuclear physics and in an energy range now found only in cosmic radiation. Some of these applications will be discussed later.

While the general idea of accelerating electrons by magnetic induction seems to be very old, the attempts which have been made to use it were unsuccessful.[4,5] Before the construction of the betatron, the details of an electron's motion in the betatron were carefully calculated and characteristics of the motion were discovered which made possible the design of a successful accelerator.

The energy is given to an electron stream by the accelerating effect of a time-varying magnetic field. Electrons circulate in essentially one plane between the poles of the betatron magnet which looks somewhat like a small cyclotron. The principle of acceleration is, however, entirely different from that used in the cyclotron, and the betatron's magnet operates on alternating current, whereas the cyclotron's magnet uses direct current.

An electron gun injects electrons into a doughnut-shaped vacuum tube between the poles of the betatron at a time when the magnetic field intensity is small. These electrons circulate between the poles of the magnet while the magnetic field is increasing, and the time rate of change of flux within the orbit produces an electromotive force directed circumferentially along the orbit. The voltage gain per complete trip of the electron around the pole is equal to the instantaneous voltage which would be read on a voltmeter connected to one turn of wire at the orbit which the electron describes. Since the electrons travelling in the vacuum tube can make an enormous number of revolutions while the flux linkage is increasing sinusoidally from zero to its peak value, the energy in electron-volts which is finally obtained is approximately equal to the voltage which would be generated on a secondary coil of the same number of turns placed in the betatron at the position of the orbit. The orbit of the electron is thus similar to the secondary of a transformer, and the great advantage of the betatron is that it is unnecessary to produce the full voltage on a secondary coil and then to apply that voltage to a high-vacuum x-ray tube for the acceleration of the electron. Instead the energy is continually and smoothly supplied to the electron stream.

The chamber or doughnut in which the electrons are accelerated is shown in Figure 1. In the original betatron the path was approximately 80 miles long. This path length is determined by the frequency at which the magnet is excited. Electrons are shot out from the injector at the time A in the figure, and after they have traveled their long path in the doughnut they strike the injector where they produce x-rays and scatter out of the doughnut. The x-rays are produced when the field is large, as at C. Without the disturbing action of the orbit expansion coils, which are shown placed above and below the doughnut, the electron stream would remain in the circle called the equilibrium orbit; however, a sudden pulse of current through these coils spirals the beam outward to the target. In each cycle the injection and expansion are repeated in succession, with a period of three-fourths of a cycle when no electrons are in the doughnut.

While the electron gains energy because of the increasing flux linkage, the magnetic field at the orbit of the electron grows stronger in such a way that the electron stream remains at a fixed radius within the vacuum chamber. Were the field at the orbit not increasing, the electrons would soon spiral toward the outer wall of the doughnut. It is necessary to increase the magnetic field, H, in proportion to the momentum, mv, which is produced by the increase in flux linkage. This arrangement of the magnetic field, then, causes the electron orbit to remain at a fixed radius.

The rate of change of momentum of the electron is, by Newton's second law, the force $d(mv)/dt = f$; and this force is equal to the energy gained per centimeter of

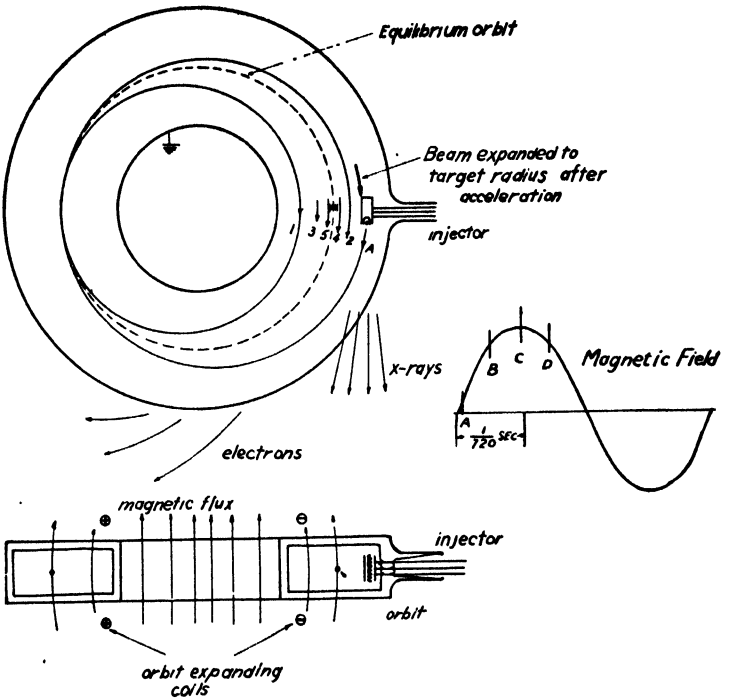


FIGURE 1. The doughnut-shaped vacuum tube in which electrons are accelerated. Electrons are injected at a time, A, in the magnetic cycle, and the orbit is expanded to the target at the time, C. If it is expanded at B or D the energy is less than the maximum which is available.

path. Since the voltage induced in a circuit is $10^{-8}d\phi/dt$, where ϕ is the flux linking the circuit, the voltage gain per centimeter by an electron travelling in a circle about a changing flux is $(10^{-8}/2\pi r)d\phi/dt$, or in ergs per centimeter,

$$(e/2\pi rc)d\phi/dt = f, \tag{1}$$

e being the electronic charge and c the velocity of light.

Integrating (1) we get

$$mv = \int_0^t f dt = (e/2\pi rc)(\phi - \phi_0) \tag{2}$$

but for the balance of centripetal and magnetic force

$$\begin{aligned} mv^2/r &= (e/c)Hv, \text{ or} \\ mv &= (e/c)Hr \end{aligned} \tag{3}$$

Combining (3) with (2) we find that

$$\phi - \phi_0 = 2\pi r^2 H \tag{4}$$

must hold if the electron is travelling in a circle.

For the case of a field and flux produced in an air gap of the same electromagnet, $\phi_0 = 0$ when $H = 0$ and ϕ is always proportional to H. The flux condition (4) is thus automatically fulfilled in the simplest kind of magnetic circuit.

We have assumed that the electron travels in a circle of radius r and have proved that if this is so the flux must fulfill the condition (4). It was necessary to prove the converse which is: given an arrangement of magnetic field and flux satisfying (4), the equilibrium orbit will be a fixed circle. This was done before the original betatron was constructed; and in the development of the complete theory, characteristics of the motion of electrons were discovered which play a vital part in the successful operation of the induction acceleration scheme.

Since the electron path is of the order of 100 miles, it must be possible to form a very stable beam. It is also necessary to find a way to introduce electrons so that they do not strike the injector or fly out of the acceleration region on one of their first few revolutions.

The theoretical treatment [8] solved these problems. It was found that stray electrons which had been deflected from the equilibrium orbit by scattering from residual gas in the vacuum tube could be brought back to the orbit by their executing a damped oscillation about this orbit. This characteristic of the motion also plays an important part in the injection of the electrons, as will be seen later.

Figure 2 shows the shape of the laminated pole pieces for the betatron, which

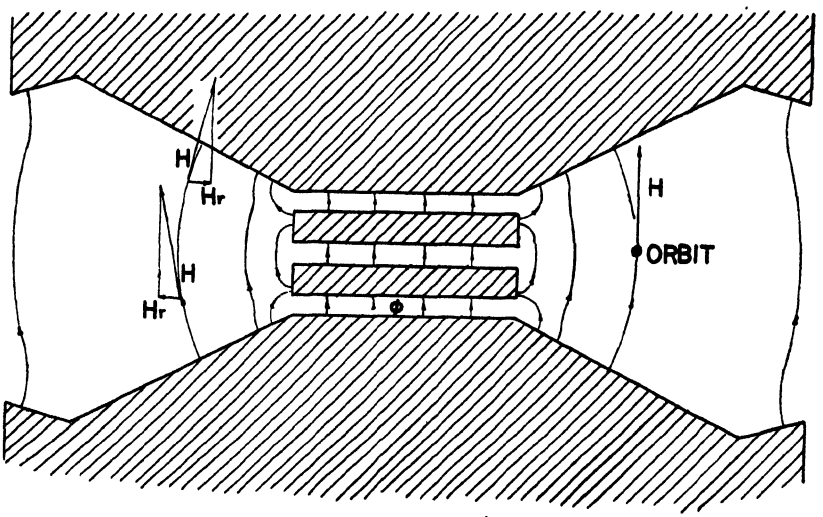


FIGURE 2. Approximate shape of the poles in the betatron. The curvature which they produce in the magnetic field results in the radial component H_r being oppositely directed above and below the plane of the orbit. H_r always forces straying electrons back toward this plane.

succeed in providing the required amount of flux within the orbit to maintain it at a fixed radius and which will produce the damped axial and radial oscillations about the equilibrium orbit. The iron discs in the center of the gap must be of exactly the correct thickness so that (4) is satisfied. The fringing field from between these discs and the curved field resulting from conically shaped pole surfaces have a radial component, H_r, at all points in the gap except in the plane of the orbit. This slight radial field influences straying electrons so that they are always forced back toward the plane of the orbit. The electron thus can oscillate axially with the decreasing amplitude which brings it back to the equilibrium orbit.

The conical shape of the poles means that the field intensity decreases with increas-

ing radius. Provided this decrease is not too rapid, radial oscillation across the equilibrium orbit can also occur. Figure 3 shows the centripetal force, F_c, necessary to hold an electron in a circle of radius r. This curve is hyperbolic, since $F_c = mv^2/r$ and v, the velocity, changes so slowly during several oscillations that it can be considered constant for the present. The force supplied by the magnetic field is $F_m = (e/c)Hv$. If the field H, which is proportional to F_m, is so shaped that F_m is less than F_c when r is less than the radius r_0 of the equilibrium orbit, then the magnetic force is not sufficient to hold the electron in at such a small radius. The electron will thus move

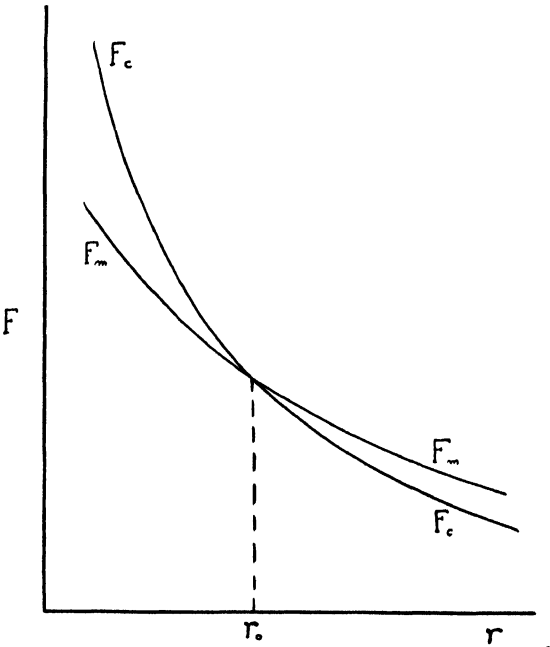


FIGURE 3. Forces acting in the radial focussing process. F_m equals (e/c) Hv for the magnetic force which is supplied to the electron. $F_c = mv^2/r$ is the centripetal force required to hold the electron at radius r. Oscillation occurs about r_0.

out past r_0 where, if F_m is now greater than F_c, the magnetic force is greater than the required centripetal force and the electron is curved inward toward r_0.

The straying electron will thus oscillate radially as well as axially about the equilibrium orbit. The damping of these oscillations is due to the slight rise in field strength with time. The effect of the increasing magnetic field is somewhat analogous to that of a stiffening spring which supports an oscillating mass. Theory shows that the amplitude of these oscillations is proportional to $H^{-\frac{1}{2}}$ and hence decreases with increasing field strength.

It is because the oscillation of an electron is damped with a relatively large damping when H is small, that electrons can be injected into the vacuum doughnut from a point close to r_0 so that they do not hit the injector on one of the first few revolutions. This makes possible the process which is used to get the electrons started.

Figure 4 shows the original 2.3-million volt betatron beside the 20-million volt betatron in the laboratory at the University of Illinois. A 600-cycle per second magnetic field was used in the original machine, and because of this fairly high frequency the voltage gain per revolution was large. This facilitated the injection

process and provided a high rate of repetition of charging and discharging the acceleration region with electrons, which is desirable for a high output current. The x-ray intensity produced by the small accelerator is equivalent to the gamma-radiation from a gram of radium, and the electron current which is accelerated is estimated at about 0.03 microampere.

The 20-million volt accelerator was constructed during a leave of absence from the University at the General Electric Company, and it was an intermediate step toward the construction of a betatron following the University's 100-million volt de

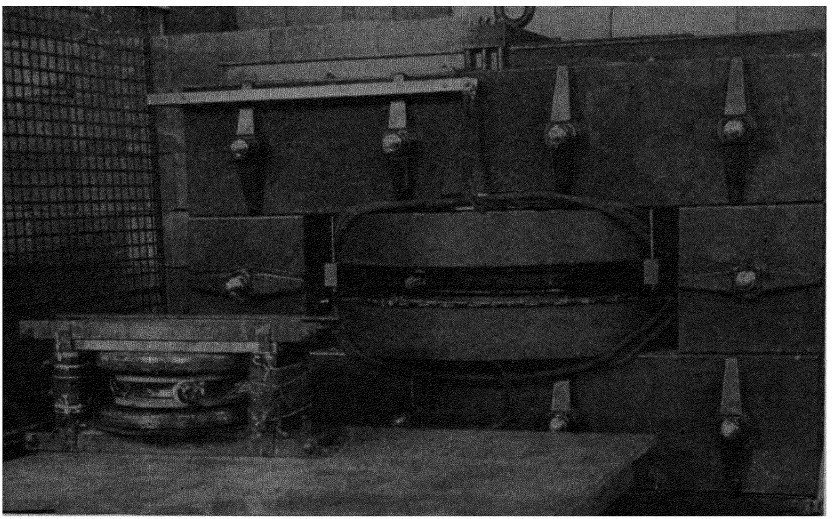


FIGURE 4. The original 2.3-million volt betatron is on the table in front of the 20-million volt betatron in the university betatron laboratory.

sign. The 20-million volt betatron operates at 180 cycles per second and produces a current at the target which can be as high as one microampere. When measured in a thick-walled ionization chamber placed 70 centimeters from the target, the x-ray intensity is about 50*r* per minute. This intensity is as great as that now commonly used in x-ray therapy; and therapy is one of the possible practical applications of the betatron.

Since the recoil electrons produced when 20-million volt x-rays pass into tissue-like material can penetrate as far as 10 centimeters, a distribution of ionization deep below the surface which is very different from that produced by low voltage x-rays is expected. With 400-kilovolt x-rays the most energetic secondary Compton electron can penetrate only about 1.3 millimeters beyond the point where the x-ray was absorbed. Since it is the secondary electron which produces the ionization in the tissue and which is responsible for the damage done to malignant tissue, this damage is localized at the point of the 400 kv x-ray absorption process. However, the long-range secondary electrons produced by the betatron's x-rays carry this destructive effect far beyond the point at which the x-ray is absorbed. The result is an ionization distribution shown in Figure 5. The ionization at a point beneath the surface of a tissue-like substance is called the depth dose. Percentages of the surface dose are plotted. At ordinary voltages the intensity of ionization falls off below the entrance surface because of x-ray absorption; but at 20 million volts the number of secondary electrons which reach a point beneath but near the surface is propor-

tional to the amount of matter which produces secondaries between this point and the surface. This is so because the long range of the secondaries carries many which are created at the surface to points well below the surface. The graph of the experimental results shows that a maximum ionization occurs at a depth of about 4 centimeters. Since many secondary electrons have low energy they do not penetrate to the extreme depth of 10 centimeters. This ionization distribution should assist in applying the x-ray dose where it is needed in the case of deep malignancies.[6]

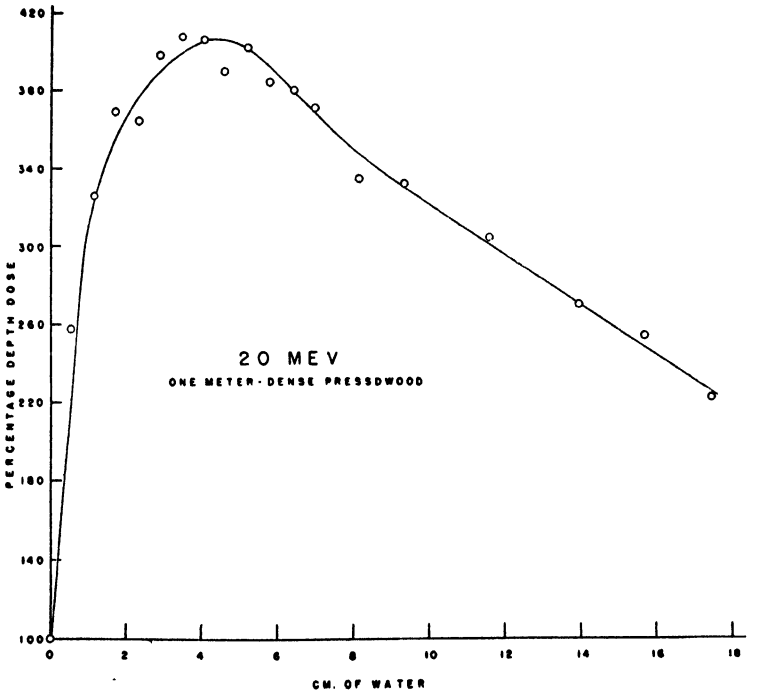


FIGURE 5. The distribution of ionization in a heavy Pressdwood phantom having a density equal to that of water.

Of more promise is the direct application of the original electron beam from the betatron to tissue under treatment. These electrons should all penetrate as far as the middle of the body and, according to the estimates of Dr. Philip Morrison of the University of Illinois Physics Department, should produce a maximum ionization 7 or 8 centimeters below the surface. Beyond 10 centimeters there can be no ionization, because this is the extreme range of 20-Mev electrons. This is a very advantageous distribution of ionization for therapeutic purposes.

While the x-rays and electrons from the betatron will undoubtedly find several practical uses, physical researches in new fields are now made possible. Tests have been made on the absorption coefficients of x-rays in various substances at very high energies.[7] Theoretical predictions have furnished these values in the past, except for a very few instances. Now the results obtained with the betatron confirm these calculations quite well. The important phenomenon observed is that x-rays become less penetrating at voltages above a certain point of minimum absorption coefficient. The minimum absorption for lead is at about 2 to 3 million volts, and for lighter substances this minimum occurs at somewhat higher energies. The rise of absorp-

tion coefficient with photon energy is a result of the increasing production of positron and electron pairs.

Another subject now presenting many problems is the nuclear photoeffect. A stable nucleus can be disrupted by the absorption of an x-ray photon. Usually a neutron is ejected and the remaining nucleus is either a stable isotope of the original nucleus or a radioactive isotope. Just as the photoelectric effect with ordinary light occurs only at quantum energies greater than a certain threshold value, so does the nuclear photoeffect occur only when the gamma-ray photon has an energy, $h\nu$, greater than the threshold for neutron emission. Because of the very precise control of the betatron's energy it has been possible to measure these thresholds for many elements with good accuracy.[8] One expects the average binding energy of a neutron, that is, the threshold energy, to be about 6 to 8 million electron-volts, but the thresholds found by detecting the radioactivity of the isotopes produced have so far all been higher than 8 million volts. New radioactivities have been found, as might be expected, since certain products may be formed in a betatron which cannot be formed in a cyclotron.

Even at 20 million volts some small cosmic-ray type showers should occur; but since there should be no difficulty in obtaining greater energies, the betatron should make easy some of the problems now being solved by experiments with the feeble cosmic rays. It was for this reason that the University's design for the 100-million volt betatron, which it was intended should be constructed on the University campus, was worked out. In this design the author has attempted to reduce the cost by allowing ¾ inch of pole face diameter per million volts. The 20 Mev betatron requires one inch per million volts. The design also reduces cost and simplifies the power problem by employing 60-cycle per second alternating current on the magnet.

References

1. Kerst, D. W., *Phys. Rev.*, **58**, 841 (1940).
2. Kerst, D. W., *Phys. Rev.*, **60**, 47 (1941).
3. Kerst, D. W., and Serber, R., *Phys. Rev.*, **60**, 53 (1941).
4. Wideröe, R., *Arch. f. Electrotechnik*, **21**, 400 (1928).
5. Walton, E. T. S., *Proc. Camb. Phil. Soc.*, **25**, 469-81 (1929).
6. Koch, H. W., Kerst, D. W., and Morrison, P., *Radiology*, February, 1943.
7. Adams, G. D., and Clark, R. K., Chicago Meeting of the American Physical Society, November, 1942.
8. Baldwin, G. C., and Koch, H. W., Chicago Meeting of the American Physical Society, November, 1942.

The Electrophoretic Study of Proteins and Related Substances

DUNCAN A. MACINNES, PH.D., AND LEWIS G. LONGSWORTH, PH.D.

The Laboratories of The Rockefeller Institute for Medical Research, New York, N. Y.

The subject of electrophoresis, also frequently called cataphoresis, is concerned with the motion of colloidal particles in an electric field. Since the size of colloidal particles varies from those visible at low magnification in a microscope to those but slightly larger than inorganic ions, such as Na^+ and Cl^-, it is evident that the experimental methods and the theory merge with those applicable to solutions of electrolytes. The connection, at least in its elementary aspects, has been outlined in a recent book by the senior author,[47] which will be used as a point of departure for the following treatment of recent developments of the subject. These developments have

been important and exciting enough to justify reserving for them all the space in this necessarily brief article.

Electrophoresis has been studied by two principal procedures: (a) the microscopic method and (b) the method of moving boundaries. The first of these has been adequately treated in recent books and review articles,[1, 2, 47] and will therefore not be described further here. Although it has decided utility in certain cases the microscopic method is giving way, at present at least, to the moving-boundary method, in spite of the fact that the former requires far less elaborate and expensive apparatus. Furthermore the microscopic method is limited to the study of the motion of particles visible in the field of a microscope, whereas the moving-boundary method is not subject to that limitation, and has other advantages that we hope to make evident in the following.

The early work on the moving-boundary method for studying electrophoresis was carried out by Burton,[8] who used the simple apparatus illustrated in Figure 1. By

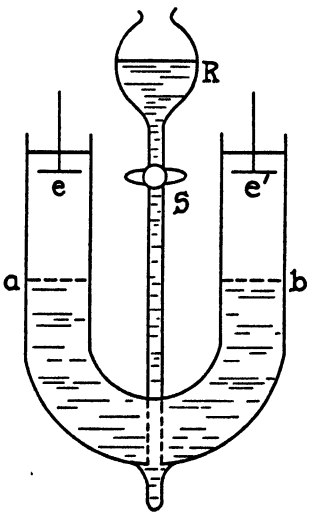


FIGURE 1. Early form of U-tube for the electrophoresis of colloids by the moving-boundary method.

first introducing the pure solvent or buffer solution into the U-tube, and then cautiously opening the stopcock *S* between the U-tube and the reservoir *R*, containing the somewhat denser colloidal solution, the electrodes *e-e'* can be surrounded by the solvent, and two fairly sharp boundaries, *a* and *b*, between the solvent and the colloidal solution are formed. On applying an electrical potential these boundaries will, in general, move. A difficulty with the early experiments with this method was that the boundaries were disturbed by convection effects arising from the heating effect of the electric current. Also, early experimenters were limited to the study of colored or turbid solutions, since boundaries involving colorless and transparent substances could not be readily seen. However, Tiselius [70] was able to obtain useful data on colorless protein materials with an apparatus similar in principle to that shown in Figure 1, by utilizing the ultraviolet absorption of the proteins for following the motion of the boundaries.

Tiselius has also made most of the more recent developments in the electrophoretic method,[72] with the results that it is now one of the most powerful instruments for research in the field of physical chemistry of proteins and related materials. Its influence is extending deeply into the domains of biology and medicine. Tiselius' main additions to the electrophoretic method are: (a) the use of the "shear" technique in the formation of the initial boundary; (b) the adaption of the Foucault-

Toepler schlieren effect to make the boundaries visible; and (c) the minimizing of convection effects by the use of a low-temperature thermostat. These developments will be discussed in the order given.

The Electrophoresis Cell and Accessories

The cell in which the boundaries are formed and observed during electrophoresis is shown in cross-section in Figure 2A, and consists of the sections I, II, III, and IV.

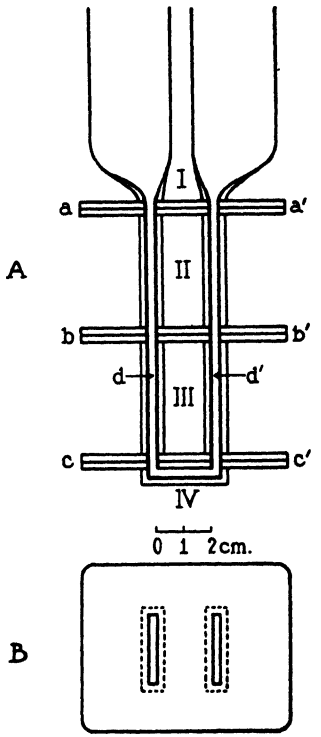


FIGURE 2. A, electrophoresis cells in cross-section; B, top view of one of the center sections.

These may be slid over one another along the planes *a-a'*, *b-b'*, and *c-c'*. Through the cell runs a U-shaped channel *d-d'* of rectangular cross-section. Figure 2B is a top view of one of the center sections. To form a boundary, the channel is filled with the buffer solution of protein to a level slightly above the plane *b-b'*, about 10 ml of solution being required with the cell of the size usually employed. Section III is then pushed to one side, the excess solution in section II is removed, and this section is rinsed with buffer. The remainder of the cell and the attached electrode vessels to be described are then filled with the buffer. On returning section III to the position shown in the figure, two sharp boundaries between the protein-bearing solution and the buffer solution are formed in the plane *b-b'*. These boundaries may then be displaced from behind the plates, where they cannot be seen, by a means to be described below. Recently cells have been constructed in which the two center sections, II and III, have been combined, yielding a single section double the length of those shown. Such a cell may be filled so that one boundary is formed at the plane *a-a'* and the other at the plane *c-c'*. The filling and manipulation of such a cell is somewhat more complicated than that just discussed, but is fully described in a paper by Longsworth, Cannan, and MacInnes.[87] This new design, which is also due to a suggestion of Tiselius, has the advantage that the channels are not inter-

rupted by the plates *b-b'*, and that electrophoretic patterns, to be described later, can be spread over twice the length of the earlier type of cell.

The support for the cell and the electrode vessels, as used at the Rockefeller Institute, together with the mechanism for moving the sections of the cell in relation to each other, are shown in Figure 3. In our apparatus a rack-and-pinion system

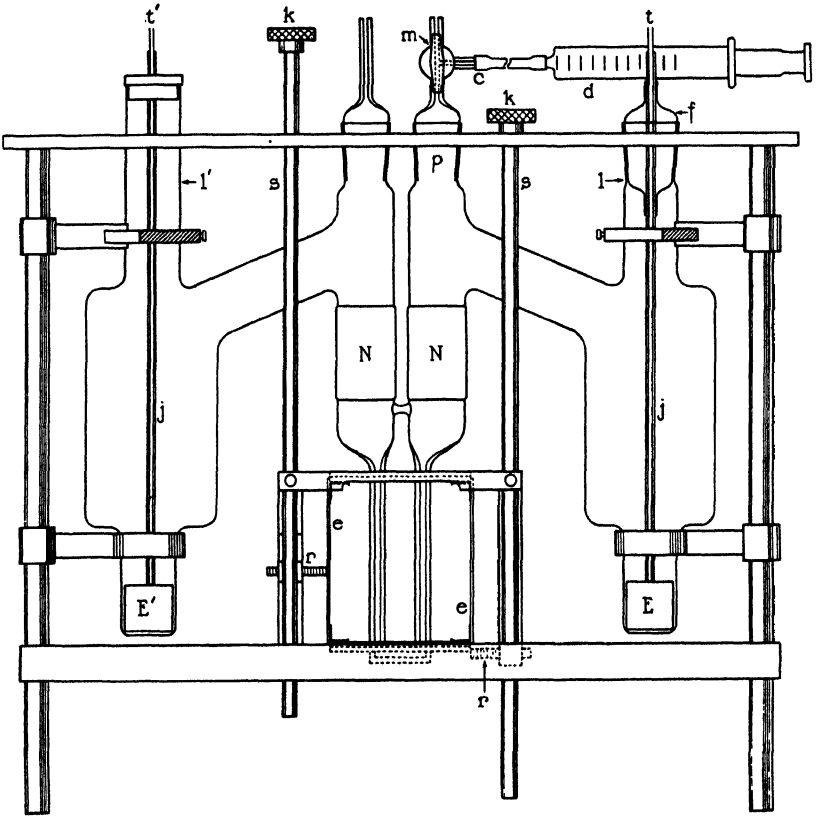


FIGURE 3. Eletrophoresis cell (new type), electrode vessels, and support.

replaces the pumps utilized by Tiselius for this purpose. The racks *r* are attached to the shafts *s* and each may be raised, or lowered, to a position opposite the section of the cell to be moved. Turning the knurled knob *k* then imparts a horizontal motion to the rack. The latter presses against a metal insert *e*, which communicates the pressure to the edges of the horizontal glass plates. The center section of the cell shown in Figure 3 is the new tall type mentioned above and, as can be seen in the figure, provision has been made for moving the bottom section as well as the center section. The short sections II and III of Figure 2 are interchangeable with the tall section shown in Figure 3 and, when substituted for the latter, either center section, as well as the bottom section, may be moved, as desired, with the rack-and-pinion system.

After filling the apparatus as described above, the silver-silver chloride electrodes (*E-E'* of Figure 3) are inserted. For effective operation of these electrodes they must be immersed in a strong chloride solution. This is accomplished by carefully introducing the solution through the silver tubes *t-t'*. These silver tubes, which are

insulated by the glass tubes (*j*), also serve as current leads to the electrodes *E-E'*. We have modified the procedure of Tiselius and our earlier procedure, in that one side of the apparatus is closed, care being taken to exclude air bubbles. This is accomplished with the ground-glass stopper *f* and the stopcock *m*. Closing one side is more convenient than having both sides open, as the latter procedure involves equalizing the liquid levels on the two sides of the system before forming the boundaries. It also permits the use of a convenient method for shifting the boundaries to be described below.

With the cell and electrode vessels filled as described and the two boundaries formed, a potential from a battery applied to the terminals *t-t'* will then, in general, cause the boundary in one side of the cell to rise and that in the other to fall; or, as will be shown later, a number of boundaries may form on both sides of the cell.

The Simple Schlieren Method for Observation of the Boundaries

As has been mentioned, another important contribution to the electrophoretic method made by Tiselius was the adaptation of the Foucault-Toepler schlieren method for observation of the boundaries. A diagram of the optical system is shown in Figure 4. The image of the horizontal slit *S*, illuminated by the lamp *L*, is

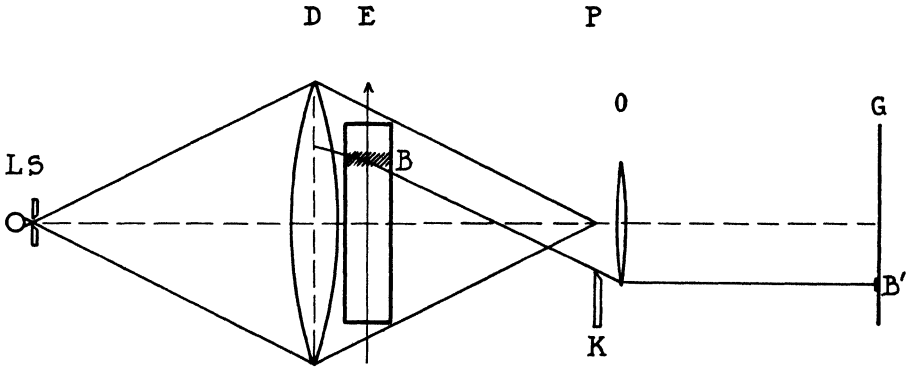


FIGURE 4. Arrangement for observation of the boundaries by the simple schlieren method.

brought to focus in the plane *P* by the lens *D*. The schlieren diaphragm *K*, which is a screen with a sharp, horizontal upper edge, is placed in the plane *P* and may be moved vertically, a micrometer adjustment being used. The cell *E*, in which the electrophoresis is carried out, is placed as near the lens *D* as the thermostat construction permits. The camera objective *O*, placed immediately behind the schlieren diaphragm, is focussed on the cell and forms an image on a ground-glass or photographic plate at *G*.

In the absence of refraction gradients in the electrophoresis cell, all the light is brought to focus in the image of the illuminated slit at *P* and enters the camera objective. If, however, a boundary, *B*, is present in the cell, the refraction decreases with increasing height through the boundary and a pencil of light through this region is deflected downward. If this deflected pencil is intercepted by raising the schlieren diaphragm *K*, it will fail to reach the screen. Thus the region at *G* conjugate to the boundary, *i.e.*, *B'*, appears as a dark band on a light background. A typical case of the use of the simple schlieren technique for locating and following electrophoretic boundaries is shown in Figure 5. The horizontal dark lines are the schlieren bands of the boundaries between an egg albumin solution in a sodium acetate buffer and the pure buffer. The upper and lower photographs are of the boundaries migrating in the anode and cathode sides of the channel, respectively.

Exposures were made at 30-min intervals, and after four exposures the current was reversed. It is of interest that the reversal brought the boundaries accurately back to their original positions. From the distance moved by the boundary in a given time, and the potential gradient in the solution, the *mobility* of the migrating material may be obtained. The potential gradient may be obtained from the conductance of the solution and the current in amperes. Some complicating phenomena must, how-

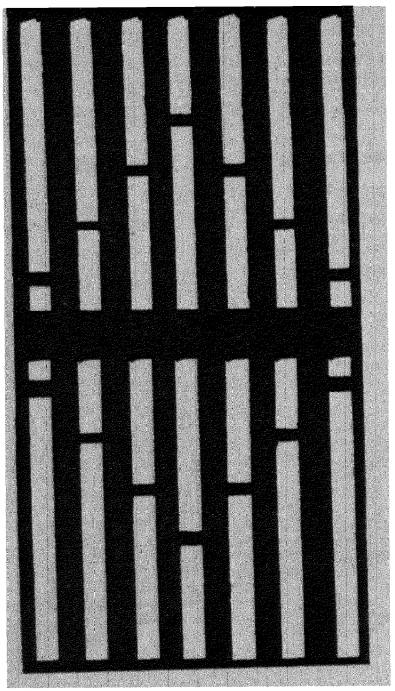


FIGURE 5. Schlieren bands of a single protein. Exposures were made at 30-minute intervals and the current was reversed after the fourth exposure.

ever, be considered, as will be described below. Modifications and improvements of this simple schlieren technique are considered later in this article.

The Elimination of Convection

As already mentioned, the early study of electrophoresis by the moving-boundary method was made difficult because the boundaries were disturbed by thermal convection. The greatest recent advance in the use of the method is largely due to a simple but brilliant idea which is also Tiselius'. If an electric current is passed through a tube containing a conducting solution, the distribution of temperature inside the tube, and in the wall of the tube, will be somewhat as is shown by curve *a* in Figure 6, *i.e.*, there will be a maximum of temperature in the center of the tube, on both sides of which the temperature will drop to that of the thermostat at the outer surface of the wall of the tube. For ordinary temperatures, and with aqueous solutions, this means that there will be a minimum of density of the solution in the center of the tube, as is also indicated for the special case of a 0.1*N* acetate buffer at 25° C in curve *b* of the same figure. As a result the heavier solution around the edges of the tube will tend to fall, and that in the middle to rise, and a convection current will result. However, although it is not possible to eliminate the temperature variation if the current is passed, one can choose a temperature at which the *density* variation is a minimum by working at a temperature near that of the maxi-

mum density of the buffer, which, for the solution under consideration, is 2.85° C. If the thermostat temperature t_0 is such that the mean temperature in the tube is 2.85°, the variation of the density across the tube is shown by the line *d* of the figure. Theoretically and also experimentally, this condition results in the minimizing of effects due to convection.

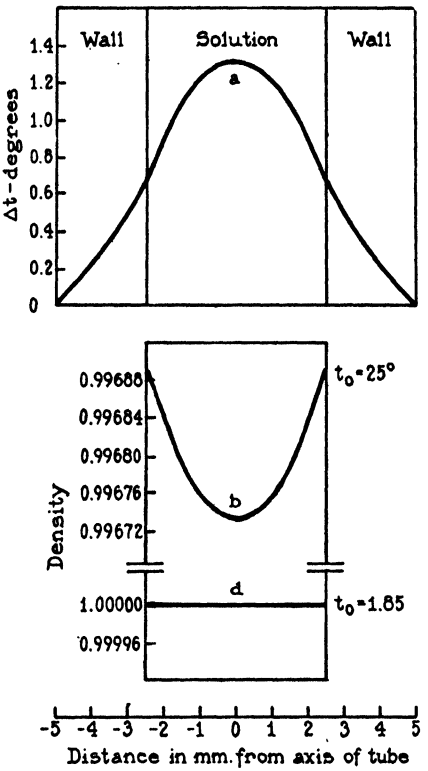


FIGURE 6. Distribution of temperature and density in a salt solution, in a cylindrical tube, during passage of an electric current.

Recently it has become evident, however, that the variation of density with height in the boundary itself has a decided influence in reducing convection effects. Thus Shedlovsky and Smadel [60] have shown that it is difficult, if not impossible, to prevent the convection of buffer solutions containing the elementary bodies of vaccinia, which have a particle size somewhat smaller than bacteria and do not appreciably affect the density of the buffer solution. On the other hand, protein solutions which are in molecular dispersion in appropriate buffers yield boundaries substantially free from convection effects (if the temperature is even approximately that of the maximum density of the solution), because of the fact that there is a definite decrease in density on passing from the protein-bearing solution to the buffer solution. Thus, although feeble convection currents may exist, even at the optimum conditions for suppressing them, such currents are apparently stopped by an appreciable gradient of density.

Applications of the Electrophoretic Method to Mixtures and to Separations

In addition to the measurement of mobilities, the electrophoretic method furnishes information as to the purity and homogeneity of protein solutions, and also may be adapted to separation of the components of a mixture. This latter applica-

tion is illustrated diagrammatically in Figure 7. Suppose a mixture of proteins A, B, and C, for which the (positive) mobilities are $u_A > u_B > u_C$, is placed in the cell as shown in Figure 7a. Movement of the bottom center section of the cell to the right will bring the protein and buffer solutions into contact in the plane α. On passage of a current, three boundaries will appear in the upper cathode and in the lower anode sections, as is shown in Figure 7b. It is evident from the figure that

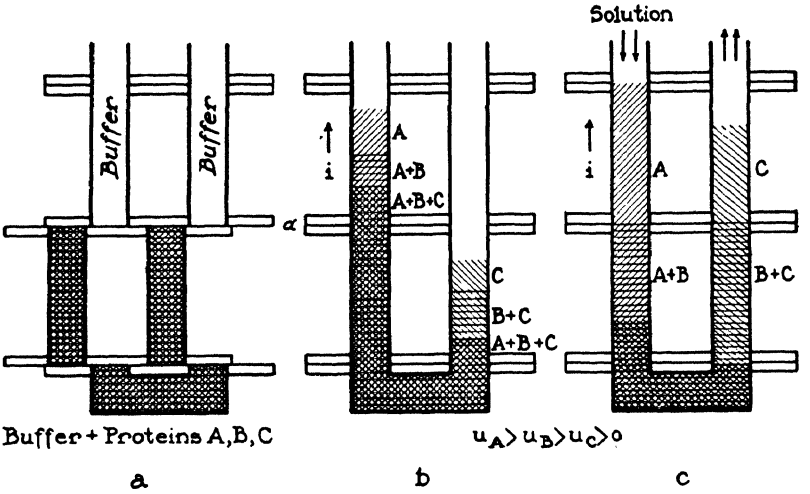


FIGURE 7. Idealized eletrophoresis of a protein mixture.

there is a separation of pure component A in the region between the two leading boundaries on the cathode side and of pure component C between the two slowest boundaries on the anode side. If the electrolysis is continued as indicated, before any large proportions of A and C have been separated, the boundaries will have migrated out of the cell in one case and into the bottom section in the other. However, if the second boundary is given an apparent velocity of zero, as indicated in Figure 7c, while the leading boundary moves through the length of one section, isolation of this section from the others would make possible the recovery of a solution of pure A from the upper cathode portion of the cell. Simultaneously, as indicated, the slowest component would have an apparent negative velocity, and pure C could be recovered from the upper anode section. As a matter of fact, a boundary can be given any apparent velocity desired by a displacement of the entire solution of the cell. In Tiselius' apparatus he displaces the boundaries by withdrawing, with clockwork, a loosely fitting plunger in one of the electrode vessels. We have modified this technique, as shown in Figure 3, by keeping the right-hand electrode vessel closed, and have displaced the solution in the cell by forcing buffer into this side from a syringe *d*, the piston of which is displaced, at the desired rate, by a threaded rod operated by a synchronous clock motor.

Recently it has been found possible to remove the solution from any portion of the cell by inserting a hollow needle and withdrawing the solution into a syringe. The material between any pair of boundaries can be obtained by watching the schlieren bands during the operation.

Improved Optical Systems for Studying Electrophoretic Boundaries

The simple schlieren method described above and illustrated in Figure 4 indicates whether or not boundaries are present in the electrophoresis cell. A boundary, how-

ever, does not consist of a single plane, but is a region in which the properties, such as the composition and the refractive index, vary from that of one solution to that of the other in contact with it. The recent developments of the schlieren method for studying and recording refractive index gradients have made possible the precise study of electrophoretic boundaries. From "electrophoretic patterns" thus obtained the concentrations of the various components, as well as their mobilities, may be determined. The available procedures are the scale method of Lamm,[27] the schlieren scanning method developed by Longsworth in the Rockefeller Institute laboratory,[34, 40, 42] and the cylindrical lens (or inclined slit) method of Philpot[52] and Svensson.[60]

The *scale method* has the advantage of involving relatively simple apparatus, but requires a large amount of tedious labor with a comparator and, having been largely superseded by the methods which follow, will not be described further.

In outline the *schlieren scanning method* is illustrated in Figure 8. As in the simple schlieren method, an image at P of the illuminated slit S-S is formed by the schlieren lens D. The camera lens O is focussed on the electrophoresis cell E and forms an image (full size in the Rockefeller Institute apparatus) on the screen at

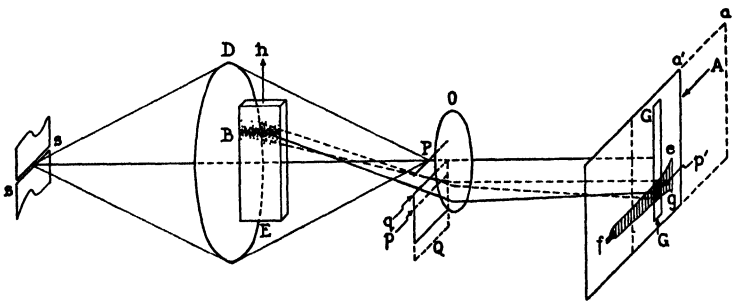


FIGURE 8. Diagram of the schlieren scanning method for the photographic recording of gradients of refractive index.

G-G. Now if the fluid in the cell is homogeneous this image will be uniformly illuminated. On the other hand, if there is a boundary, B, between, for instance, a protein-bearing solution and a buffer, there will be a region in which the refractive index varies with the height in the cell, and light which would normally pass to P is deflected downward, since the solution with the greater refractive index is on the bottom. If the schlieren diaphragm Q is raised to a point p where it intercepts the most deflected light, a dark band will appear on the screen, conjugate to the region of steepest gradient in the boundary B. Such a band appears in the image G-G at p'.

However, as previously noted, a boundary is a region in which the composition varies from that of one solution to that of the other. The refractive index, n, in such a region changes continuously with the height, h, of the liquid in the cell. The gradient, dn/dh, of refractive index, for each boundary, will thus, theoretically at least, vary from zero to a maximum, and back to zero. In Figure 8 the variation of the gradient, dn/dh, of the boundary B in the cell E, is represented by the density of the shading. The pencil of light passing through the layer having the maximum value of the gradient will be most bent from the normal path and will be the first to be intercepted when the schlieren diaphragm Q is raised. On lifting this diaphragm still further, *i.e.*, to the level q, less refracted pencils of light will also be intercepted and the schlieren band in the image G-G will widen. This obviously can be continued until the whole field has been covered. In the schlieren scanning method this process is made continuous and is recorded photographically. The image

of the cell at *G-G* is masked by a narrow vertical slit and a photographic plate *A* is moved in the direction of the arrow at a constant rate across this slit. Actuated by the same mechanism, the schlieren diaphragm *Q* is given a steady movement upward. The resulting (positive) photographic record for a typical single boundary is indicated by the area *e-f-g*. The displacement of the diaphragm *Q* from the position *P* is proportional to the gradient at levels in the cell *E* conjugate to the edges of the schlieren bands. Thus the contour of the area *e-f-g* indicates both the position and the magnitude of the refractive index gradients existing in the boundary. Since the photographic plate *A* was in position *a* at the time the schlieren diaphragm was at *p*, a section of the band *p'* appears at *f* when the plate has been moved to *a'*. It will be seen that the usual schlieren bands are narrow sections through the area *e-f-g*. By a system of gears the ratio of the rates of motion of the schlieren diaphragm *Q* to that of the plate *A* is given a constant value, such as one to three, and this ratio can be varied by changing the gears. For establishing the base line of the schlieren patterns the position of the diaphragm *Q* may be read accurately with a micrometer.

The *cylindrical lens method* is convenient in that it makes visible the whole electrophoretic pattern on a screen. It is shown diagrammatically in Figures 9a

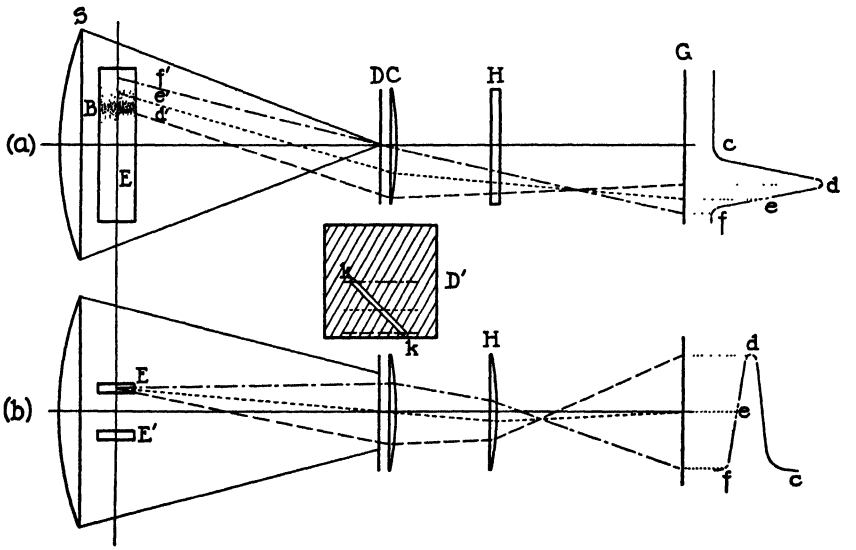


FIGURE 9. Diagram of the cylindrical lens method for observation of the electrophoretic patterns.

and 9b, the first being the arrangement as viewed from the side and the second as seen from above. Light from an illuminated horizontal slit to the left of the schlieren lens *S* is focussed upon a diaphragm *D*, at right angles to the page. In this diaphragm there is a slit cut at an angle, as is shown at *D'* in which it has been turned 90° from its position at *D*. The camera lens *C* normally projects an image of the electrophoretic cell *E* on the screen *G*. However, there is a cylindrical lens *H*, focussed upon the diagonal slit *D*, which is interposed between the lens *C* and the screen *G*. This cylindrical lens has no effect on the direction of the light beams in the (vertical) direction of its axis, as is shown in Figure 9a, but alters the paths of the light in the horizontal planes, as in the lower part of the diagram, Figure 9b. The diagonal slit *D* causes the light which is deviated downward by the boundary *B* in the cell *E* to be cut off sideways proportionally to that deviation. The cylindrical

lens focusses this sidewise shift on the screen G while the cell E is still in focus vertically. The result of this optical arrangement is that the line *c-d-e-f* is a plot (actually visible on the screen G at right angles to the page) of the refractive index gradients in the boundary B. In that boundary the values of such gradients are again represented by the density of the shading. A light beam from a given layer in the boundary, d' for instance, passes in the vertical direction, to its conjugate position d, and indicates on the plot *c-d-e-f* the position of the layer in the boundary, *i.e.*, the abscissas of the plot, as shown in Figure 9a. The downward shift of the beam of light (and therefore the sidewise shift by the angular slit and cylindrical lens) is proportional to the refractive index gradient in the boundary, and yields the values of the ordinates in the plot in the same figure. The principle will be made clear to the reader if he follows the beams f'-f, e'-e, and d'-d in both figures. These represent respectively light from layers in the boundary where there is no refractive index gradient, an intermediate value, and the maximum value.

Although this cylindrical lens method is convenient for a quick visual examination of the field and for control of electrophoretic separations, it has the disadvantage that adequately corrected cylindrical lenses are not commercially available. The electrophoresis equipment in our laboratory is now arranged [36] so that either the cylindrical lens or scanning procedure may be used interchangeably, the former for visual inspection of the patterns during an experiment and the latter for the permanent photographic records on which the analyses are based.

The Determination of Mobilities and Concentrations from Electrophoretic Patterns

A reproduction of a typical electrophoretic pattern obtained by the schlieren scanning method of normal human blood plasma is shown in Figure 10. The left and right sections of the figure are the patterns from the rising and descending boundaries respectively. The different peaks are those produced by albumin (A), the α_2, β, and γ globulins, and fibrinogen (ϕ) as marked. Incidentally, until the electrophoretic work of Tiselius it was not known with certainty that there exists more than one globulin in human blood.

From such electrophoretic patterns it is possible to obtain the mobility of the constituent forming the leading boundary if the following data are available: (a) the specific conductance, K_p, of the protein solution, (b) the current, I, in amperes, and (c) the time, t, during which the descending boundary has swept through the volume V. With these data the mobility, U, may be computed with the aid of the formula

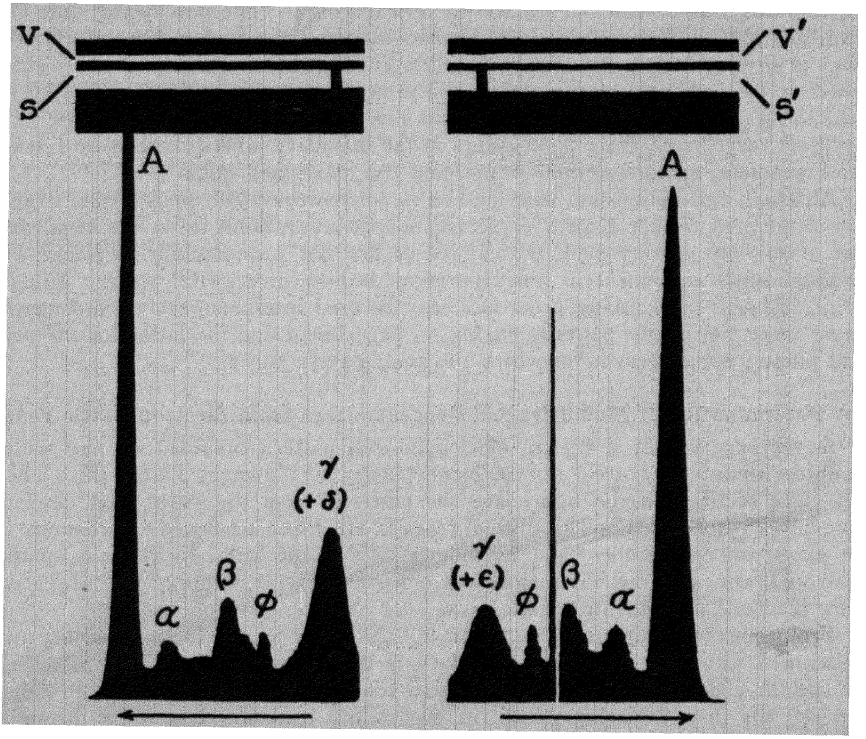
$$U = \frac{VK_p}{It}$$

"Mobilities" for the slower components may also be computed from the patterns and are useful for identifications. Quantitatively they are subject to small errors.

In order to obtain the volume it is necessary to know the starting position of the boundaries, and the following procedure is used. The initially sharp boundaries between the protein solution and buffer, which are out of sight because they are hidden behind the sliding plates of the cell, are brought into view by forcing a small volume of buffer into one side of the electrophoresis apparatus. The schlieren band due to the initial boundary is recorded on the photographic plate. The result is shown on the strips s, s' of Figure 10. The electrophoresis is then carried out for the measured time, t, the plate is shifted slightly in the camera, and the scanning process, as described above, is carried out. The distance, h, between the starting position and that of a given boundary may then be measured with a comparator. The position of a boundary, for instance, that of the albumin, A, in Figure 10b, may without serious error be assumed to be that of the maximum ordinate. Mobilities

computed from data obtained from rising boundaries are found to be, under the conditions of these experiments, over 10 per cent greater than those found from descending boundaries. The reasons for this discrepancy will be discussed below.

The concentration of a component may be obtained from the area formed, in the schlieren pattern, by a given boundary, if the coefficient K, connecting the refractive index increment with the concentration, is known. However, it is very important



(*a*) **Rising** (*b*) **Descending**

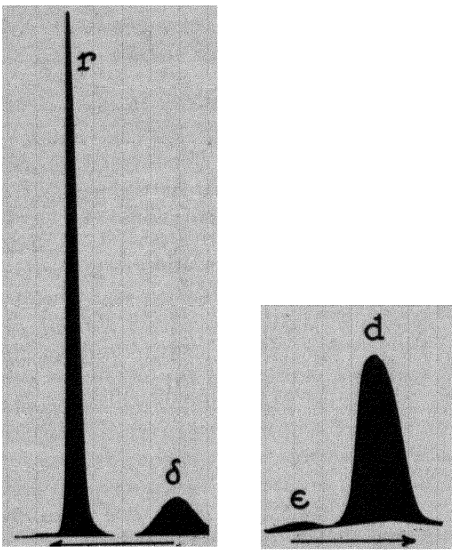
FIGURE 10. Electrophoretic patterns of a sample of normal blood plasma, including records of the base line of the pattern, and the starting position of the boundaries.

for this purpose to establish the base line of the pattern. This will be the position, on the pattern, of undeviated light, including any slight refractions due to irregularities of the cell walls, and is located as follows. Before the boundaries are in view, the schlieren diaphragm, Q, of Figure 8, is brought to a position just below P, *i.e.*, below the normal image of the slit S, and the scanning process, as described, is then carried out until the diaphragm has passed P. A typical result of this process is shown in the strips v and v' at the top of Figure 10. From the readings of the micrometer head registering the position of the schlieren diaphragm Q of Figure 8, and the ratio connecting the rates of motion of the diaphragm with the distance, z, along the plate, the reading of the micrometer head at which the undeviated light is intercepted may be interpolated. If there are irregularities of the cell wall this reading may be found as a function of the height, h, in the cell. When the protein pattern is taken by the schlieren scanning method, the positions along the vertical direction (as reproduced in Figure 10) are also known in terms of readings of the

micrometer head. The lower edge of the electrophoretic pattern, *i.e.* the base line, can thus be accurately plotted. In determining the area due to a given component with a planimeter it is necessary to make a more or less arbitrary separation, a vertical line being drawn from the lowest point between two adjacent peaks, because of the fact that the refractions due to the different boundaries frequently overlap.

Some "Boundary Anomalies"

The type of phenomenon to be discussed in this section is illustrated in Figure 11, which represents electrophoretic patterns of a 1.36 per cent solution of ovalbumin



(*a*) Rising (*b*) Descending

FIGURE 11. Electrophoretic patterns of a 1.36 per cent solution of ovalbumin in a 0.1*N* sodium acetate buffer at pH 3.92 and 0°.

at pH 3.92 in a 0.1*N* sodium acetate buffer. It will be seen in this figure that two peaks or maxima are evident in the patterns of both the rising and descending boundaries, but that the patterns are far from being mirror images of each other. This may possibly be an extreme case, but is all the more useful in illustrating deviations from ideal behavior and in testing the validity of the interpretation to be given below.

The deviations, or so-called "boundary anomalies," illustrated by these patterns are as follows. (1) The rising boundary *r* is much sharper than that, *d*, descending into the protein solution. (2) The boundary *r* has swept through a larger volume V_r, than that, V_d, of the boundary *d*. (3) The total shaded area of Figure 11a is equal to that of Figure 11b, but the partial area A_r under the rising boundary *r* is less than that, A_d, for the descending boundary *d*. Since these areas are proportional to the concentration changes at the boundaries, it is apparent that the change is less at the rising boundary than at the descending one. Also (4), the concentration distribution in a boundary is not, in general, symmetrical about the maximum, as is shown by the shapes of the peaks *d* and *r*.

These deviations may be explained, with the aid of Figure 12, as follows.[38, 40] In practice the protein solution is prepared by dialysis against a buffer solution. A difference of salt composition, due to the Donnan equilibrium, exists between the two solutions when dialysis is complete. Moreover, the protein solution contains conducting constituents, the protein ions, that are not present in the buffer solution. When passage of a current causes a boundary between two such solutions to move

from a to d of Figure 12, for example, there is found in the intervening volume V_d a buffer solution of composition [B′]. This composition has been "adjusted" by the passage of the electric current, in general, to a value different from that of B.* The boundary ε, which moves very slowly, thus forms between two solutions of the same buffer, but at different concentrations, and is quite evident in Figure 11. With the boundary r moving upward into the buffer solution, there is a similar but more complicated adjustment of the composition of the protein solution which replaces the buffer as the boundary rises. The resulting concentration boundary δ, between the

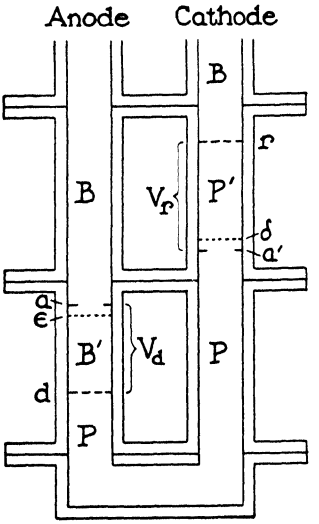


FIGURE 12. Diagram of the electrophoresis cell illustrating the relative positions of the boundaries.

solutions P and P', of Figure 12, also moves very slowly under the influence of the current. The δ boundary, also shown in Figure 11, is much more visible than the ε boundary, because the former involves a gradient of protein concentration whereas the latter does not. This also has the effect of reducing the area A_r under the rising boundary. These δ and ε boundaries were at first thought to be due to electrophoretically inert components present in the solution. There is, however, ample evidence at present of the adequacy of the explanation given above. In some cases the δ and ε boundaries may be suppressed by an appropriate selection of the concentration of the buffer solution with which the initial boundary is made, as has been shown by Longsworth and MacInnes.[40]

Since the specific conductance of the protein solution P has been found experimentally to be less, in general, than that of the adjusted buffer solution, the electric field is greater in the protein solution than in the buffer, and variations of this field thus exist at the descending boundary d. The protein ions in the dilute uppermost layers of the boundary d are therefore in weaker fields than are those in the concentrated layers and thus tend to lag behind, causing the boundary to become diffuse, as is shown in Figure 11b. In the case of the boundary r rising into the buffer solution, the specific conductance of the "adjusted" protein solution is less than that of the buffer, so that the dilute, slowly moving protein ions tend to be overtaken by the faster, concentrated ones, with the result that this boundary tends to remain sharp, as illustrated by Figure 11a. Due to these field gradients at the boundaries, and in some instances to pH gradients also, the distribution through a boundary may be unsymmetrical about the ordinate passing through the maximum of the re-

* This adjustment is understood for some simple cases,[16, 26, 33, 81] but the theory for the more complicated systems frequently encountered in practice has not been developed.

fractive index gradient curve. This lack of symmetry, shown clearly by the boundary *d* of Figure 11b, has been studied by Tiselius and Horsfall.[76] Assuming a proportionality between refractive index and protein concentration, it is apparent that in such cases the position of the maximum gradient is not identical with that of the ordinate which divides the area under the curve into two equal parts. The position of the latter ordinate is the better value to use in mobility computations, since it is approximately the position the boundary would have if it retained its original sharpness. The precise location of an unsymmetrical boundary requires an integration of the concentration-distance curve.[35] Computations for some typical cases show that the position of the ordinate dividing the gradient curve into equal parts differs from the true value by at most 1 per cent, whereas the position of the maximum ordinate may be in error by as much as 20 per cent.

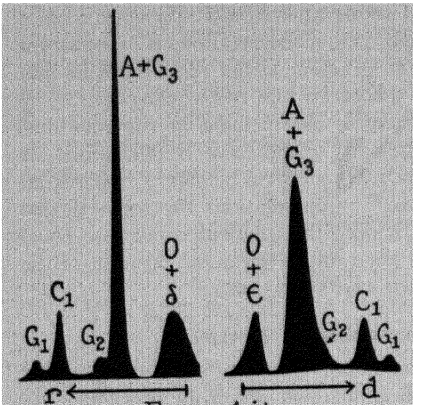
The Results of Investigations Using the Electrophoretic Method

In the short time since Tiselius announced his improvements of the electrophoretic method quite a number of investigations have been made in which the procedure has been applied to problems arising in connection with the chemistry of proteins and related substances, and in biology and medicine. Some typical examples are outlined below.

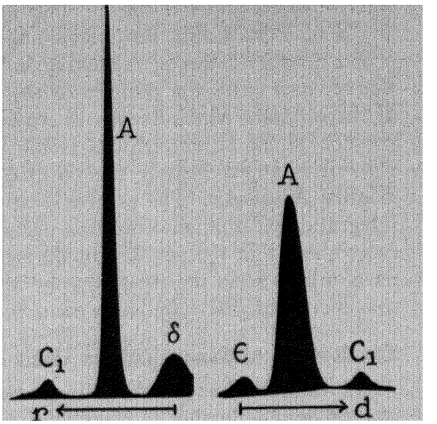
A word of warning to those considering the application of the method to specific problems may not be amiss. It is our experience that quick solutions of such problems are not to be expected. Usually it is found that a few electrophoretic patterns only suggest further work, and that the answer to a given question is reached only after many patterns have been collected and all the relevant chemical, physical, and biological information has been used. Usually the patterns suggest chemical or biological experiments which in turn require further electrophoretic experiments. However, the additional information yielded by the electrophoretic method frequently leads to solutions of research problems that cannot be obtained otherwise, or are arrived at with great difficulty.

(a) **The Purification of Ovalbumin.** An example of the utility of the electrophoretic method for obtaining information as to what is actually going on during a familiar procedure, *i.e.*, the purification of ovalbumin by the method suggested by La Rosa[28] will be illustrated with electrophoretic patterns, which are given in Figure 13. These are from a research on egg white by the authors and Dr. R. Keith Cannan.[37] In that figure (a) is for the untreated egg white at pH 3.92. Comparison with patterns obtained with the separated (or partially separated) components has made possible the identification of the sources of the various peaks as: O, ovomucoid (including the δ and ε effects); A and G_3, ovalbumin and one of the three globulins, C_1, conalbumin, and G_1 and G_2, globulins. At this pH value G_3 has the same mobility as ovalbumin, as will be shown below. This is only one of many examples we have encountered which indicate that the number of components in a mixture must not be judged from the number of peaks in the electrophoretic patterns (or the number of schlieren bands) obtained at a single pH value. If now the conventional separation is made into "albumin" and "globulin" fractions by adding to the egg white an equal volume of saturated ammonium sulfate and filtering, the resulting electrophoretic pattern of the soluble fraction is shown in (b) of Figure 13. The pattern shows peaks corresponding to all the constituents indicated in (a) with the exception of globulin G_3. The pattern for the corresponding insoluble or "globulin" fraction is given in Figure 13 (d). That material also shows peaks corresponding to all the constituents appearing in the pattern for egg white. However, there has been concentration of the albumin components in the filtrate and of the globulins in the precipitate.

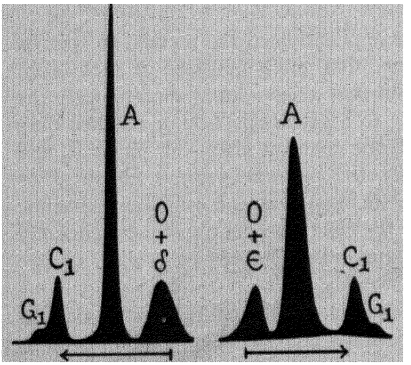
Figure 13 (f) shows the effect of crystallization of the "albumin" fraction, after bringing this solution to pH 4.6 by adding acetic acid. The globulin G_1 has dis-



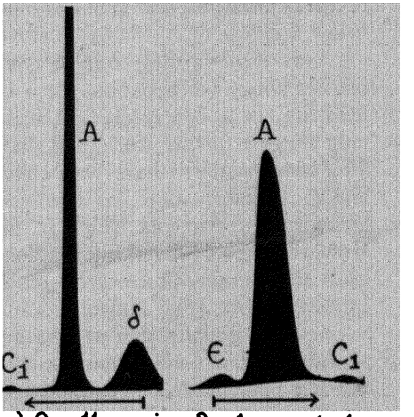
(a) Egg white



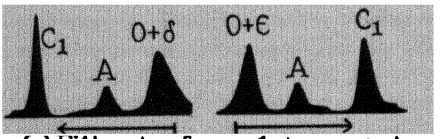
(f) Ovalbumin - 1st crystals



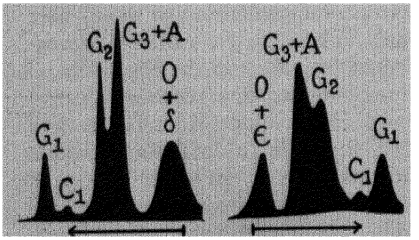
(b) Albumin fraction



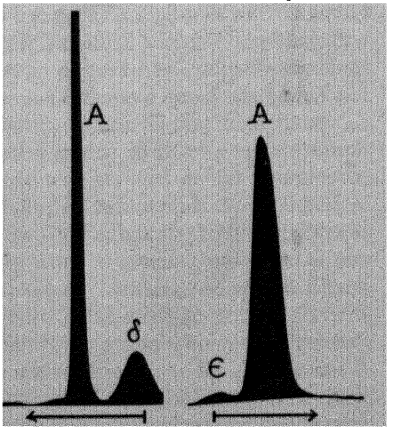
(g) Ovalbumin - 2nd crystals



(c) Filtrate from 1st crystals



(d) Globulin fraction



(h) Ovalbumin - 3rd crystals.

FIGURE 13. Electrophoretic patterns of the various fractions arising in the preparation of crystalline ovalbumin.

appeared, and conalbumin, C_1, is present in greatly decreased amount. That the ovomucoid, O, is reduced is shown by comparing the O + ε peaks in (a) and (b) with the ε peak in (f). The electrophoretic patterns, Figure 13 (g) and (h), show the effect of two further crystallizations. In the former, a scarcely perceptible amount of conalbumin is present, and in the latter it has practically disappeared. It will be observed that the peaks marked δ and ε persist unchanged in patterns (g) and (h). As already mentioned, these are due to concentration changes near the original position of the boundaries and do not arise from any single component. After three crystallizations, therefore, the patterns show that substantial purification of the ovalbumin has been attained. This is in agreement with the observations of Sørensen and Høyrup,[68] and of Hektoen and Cole.[15]

Of interest in this connection is the pattern of the filtrate obtained after crystallization of the "albumin" fraction, Figure 13 (b). This is shown in Figure 13 (c), and indicates relatively large proportions of the conalbumin, C_1, and ovomucoid, O, and a correspondingly small amount of ovalbumin, A, indicating once more that the crystallization serves as an effective means of separation of the chief constituent from its impurities.

(b) **The Complexity of the Proteins of Egg White.** Before the electrophoretic investigations it was not known that egg white contains at least three globulins. In addition, the method has shown that the other proteins, albumin, conalbumin, and ovomucoid, are complex. The type of complexity is different for the three substances.

Although *ovalbumin* forms small but well defined crystals, the electrophoretic patterns indicate that it has two components, A_1 and A_2, which have different mobilities at most pH values. This was first pointed out by Longsworth[84] and has been observed independently by other workers.[73] Some electrophoretic patterns of at least three-times recrystallized ovalbumin are shown in Figure 14. It will be seen that

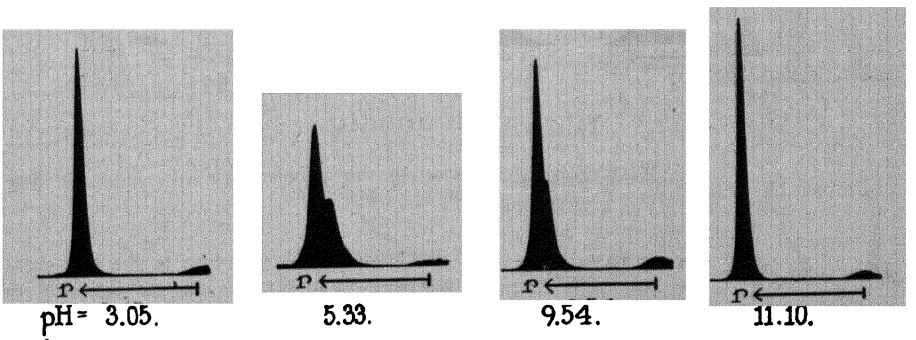


FIGURE 14. Electrophoretic patterns illustrating the complexity of crystalline ovalbumin.

at low pH values and at very high ones there is only one peak or maximum in the patterns, but at intermediate pH values there are two. There appears to be some variability in the relative amount of the component A_2, but it has been found in all the many samples of ovalbumin that we have examined, although the material has been prepared from eggs from diverse sources, and with some variations in the manner of preparation. The component A_2 does not appear to be, as was first conjectured, an intermediate step in the usual processes of denaturation. The native recrystallized ovalbumin remaining after a large part of the material had been denatured by means of heat, acid, alkali, and surface denaturation yielded patterns practically indistinguishable from the starting material.

Conalbumin was studied electrophoretically at a series of pH values, with the results, for the rising boundaries, shown in Figure 15. At pH 3.92 only one component is indicated, whereas at the higher pH values there are two maxima (other than that due to the δ effect), indicating two components which we will designate by C_1 and C_2. It will be seen that the proportion of the component C_2 decreases with increasing pH. Conalbumin in egg white itself, at least above pH 4, shows only one component with mobilities corresponding to the constituent C_1. If, however, the diluted egg white is dialyzed for three days at pH 3.6 and then again dialyzed at pH

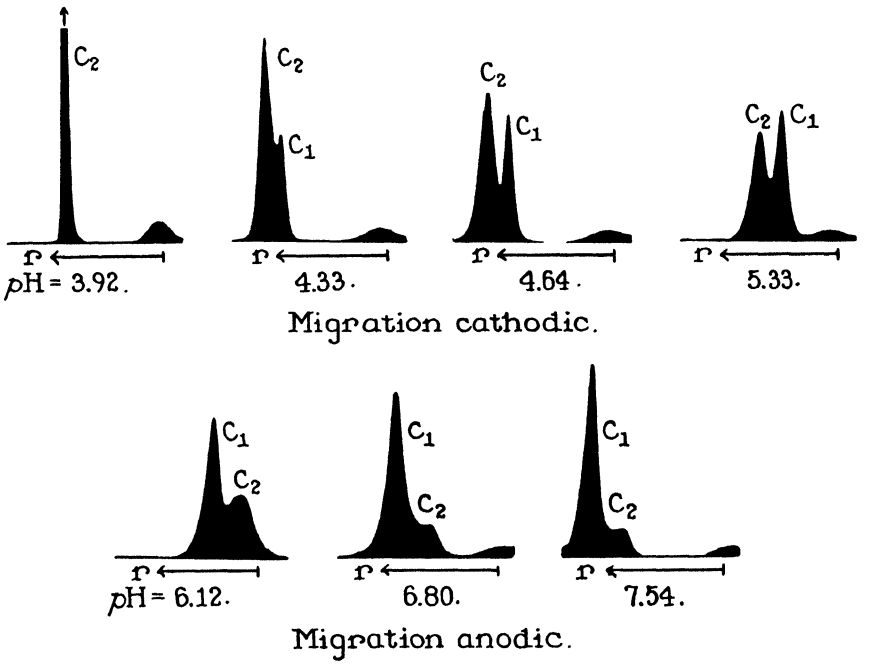


FIGURE 15. Electrophoretic patterns illustrating the complexity of conalbumin as prepared from egg white.

4.6 and an electrophoretic pattern obtained, maxima corresponding to the two components, C_1 and C_2, are observed. It thus appears that in egg white, as well as with the separated conalbumin, the transformation from C_1 to C_2 takes place below pH 4, the latter being the stable form below that pH value and the former being stable above it, at least in dilute salt solutions. Evidence that there are two components with different molecular weights in purified conalbumin was furnished by the ultracentrifuge. The experiments were carried out by Dr. A. Rothen of the Rockefeller Institute.

The indication that *ovomucoid* is complex is furnished by the observation that it shows "reversible boundary spreading" in an electric field. The phenomenon was described by Tiselius and Horsfall.[75] In such cases an initially sharp boundary will become diffuse as the electrolysis proceeds, but when the current is reversed will progressively recover some of its initial sharpness. The simplest explanation of this phenomenon is that the protein molecules forming the boundary have a range of mobilities instead of a single value. The boundaries of the ovomucoid of egg white exhibit this type of spreading to a market extent. This is illustrated by the superimposed tracings, shown in Figure 16, of the electrophoretic patterns obtained dur-

ing the electrolysis of a 1.18 per cent solution of ovomucoid in a 0.1*N* sodium acetate buffer at pH 5.32. The initial boundary positions in the two sides of the channel were at a and a'. The pattern contours after electrolysis for successive periods of 4500 seconds at 5.81 volts/cm are indicated by the full curves r_1, r_2, and r_3 for the rising boundary and d_1, d_2, and d_3 for the descending boundary, respectively. After the third period the current was reversed and the dashed curves represent the con-

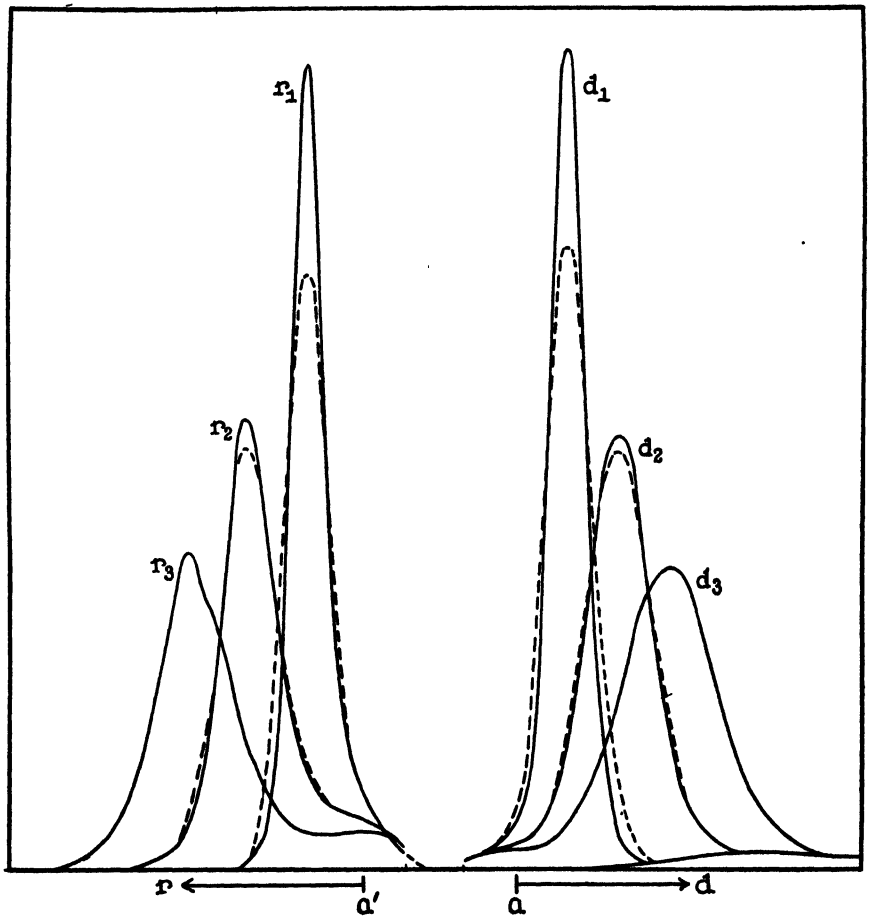


FIGURE 16. Electrophoretic patterns illustrating the reversible boundary spreading of ovomucoid.

tours after two more intervals of 4500 seconds. A comparison of the full and dashed curves for each time interval indicates that, although the boundaries spread as they migrate away from their initial positions, this spreading is almost completely reversed on changing the direction of the current. Since some spreading of a boundary, due to diffusion, occurs independently of the electrophoresis, the boundaries do not recover all their original sharpness. The substantial recovery that is effected indicates, however, that most of the boundary spreading is due to electrical inhomogeneity of the ovomucoid. Similar spreading of the ovomucoid boundaries has been observed at other pH values. However, in no instance has there been indication of resolution into two or more separate boundaries.

To summarize, the proteins ovalbumin, conalbumin, and ovomucoid illustrate three different types of electrophoretic complexity. Ovalbumin shows two components, their amounts being quite independent of pH and other variables. Conalbumin has two components, the relative amounts of which vary with the pH. Ovomucoid shows a range of mobilities, and possibly therefore a range of composition or of molecular shape.

(c) **The Electrophoretic Study of Blood.** Of the protein-bearing solutions the most interesting and important would appear to be human blood. Stenhagen [64] has obtained the mobilities of the main constituents of human blood plasma by the schlieren band method. However, by the schlieren scanning procedure it has been possible to obtain not only the mobilities but also the concentrations of the protein constituents in the sera and plasmas. The following is a brief outline of a portion of the work done in that field by the authors and by Drs. Shedlovsky and Scudder in the laboratories of the Rockefeller Institute.[42, 59] Figure 10 shows the electrophoretic patterns for the rising and descending boundaries of normal human blood plasma. As already explained, there are peaks for albumin A, the three globulins, α_2,* β, and γ, and for fibrinogen, ϕ. The electrophoretic patterns obtained from the blood of normal healthy individuals are strikingly alike. In the descending boundary pattern there is the interesting "β globulin disturbance," indicated by the sharp spike. The explanation for this phenomenon appears to be that when it is separated from the albumin and the α globulins a change takes place in the β globulin, producing convection currents and, in consequence, regions of total reflection. In some cases the β globulin, or accompanying lipoid material, precipitates, yielding a turbidity which travels with the β boundary.

In the course of this work the electrophoretic patterns of the plasmas and sera of patients suffering with a wide variety of diseases were studied. Some of these patterns differ so much from that of normal blood that any interpretation was at first very difficult. However, order has begun to emerge, and tentative conclusions as to the meanings of patterns may be made. In Figure 17 the patterns obtained from the sera of patients suffering from various pathological conditions are compared with that from normal serum. One condition common to all these pathological conditions was fever. It will be observed that in each case there is a marked increase of α globulin. This increase of α globulin has been invariably found in cases of infections accompanied by fever, and, more generally, in conditions involving inflammation or destruction of tissue.[5, 59]

Figure 18 contains the electrophoretic patterns of other samples of pathological sera and plasmas, including that of normal plasma, *a*, for comparison. Of these patterns, *b* is that of the plasma of a multiple myeloma patient. The most marked difference between this pattern and that of normal plasma is the great increase of the peak due to β globulin, or, more exactly, of material moving with the mobility ascribed to β globulin, since a peak may arise from two or more materials moving with essentially the same velocity. The more recent work of Kekwick [24] and of Gutman and his associates [12] indicates that the enhancement of the β globulin peak in Figure 18b is probably due to one of the Bence-Jones proteins frequently associated with this disease.

It is our observation and also that of Blix [6] that lipoid material may be carried with the β globulin, probably in loose chemical combination. This is indicated by the patterns *e* and *f* of the figure, from a case of obstructive jaundice. The first of these is directly from the plasma, whereas the second was obtained from the plasma after it had been extracted with ether to remove the fatty material. It will be ob-

* Recently another globulin, designated as α_1, has been observed [36] to separate from the albumin in patterns obtained at pH 8.6, using a buffer consisting of 0.1*N* sodium diethylbarbiturate and 0.02*N* diethylbarbituric acid. The use of this buffer also has the important advantage that there is a separation of the peak of the γ globulin from those due to the δ and ϵ boundaries (see reference 38).

served that the peak due to the globulin has been much reduced by this procedure.

(d) **The Isolation of a Pure Hormone.** An example from the work of Shedlovsky *et al.*[58] of the utility of the electrophoretic technique in connection with the identification and purification of a biologically active material is that of the isolation of metakentrin, *i.e.*, the luteinizing hormone of the anterior lobe of the pituitary gland. By application of the usual techniques a material was obtained that gave at pH 7.85 the electrophoretic pattern shown in Fig. 19 A, which indicates three

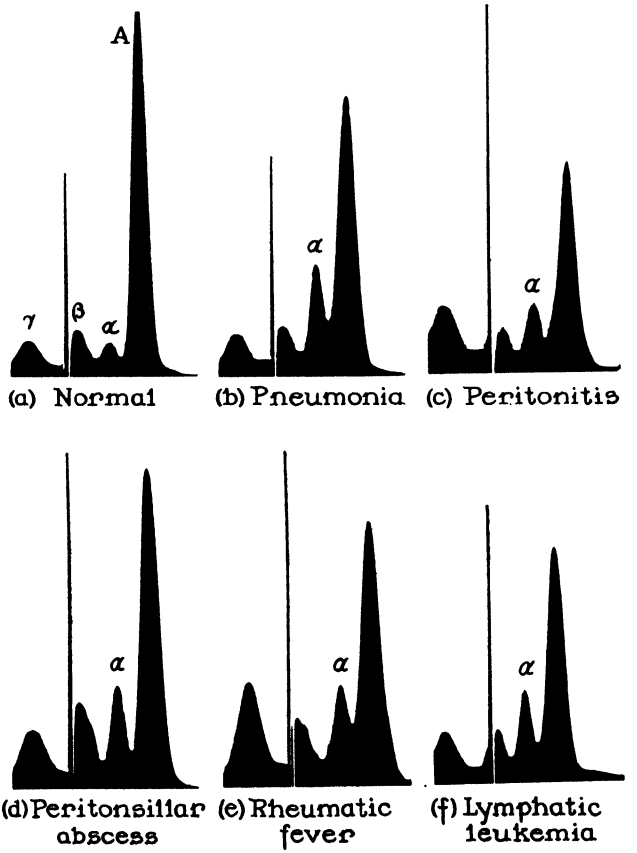


FIGURE 17. Electrophoretic patterns of some normal and pathological human blood sera.

components. After sufficient separation of the electrophoretic boundaries it was possible to isolate two fractions, one containing the major component and the other the two minor ones. This was done by inserting the needle of a syringe and withdrawing the material desired, using the schlieren bands to control the operation. The biological activity was found to be entirely in the major component. Electrophoresis experiments were performed over the pH range 4.5 to 8.0, and it was found from a plot of the mobilities against the pH values that the major component has an isoelectric point, *i.e.*, zero mobility, at pH 7.45. This information made it possible to secure the pure hormone by isoelectric precipitation, which yielded the electrophoretic pattern shown in Figure 19 B. The fact that the material shows only one peak in the pattern over its pH-stability range is a necessary, but not suf-

ficient, indication of its purity. Ultracentrifugal, solubility, and biological investigations, however, contributed additional evidence that a single pure hormone had been separated.

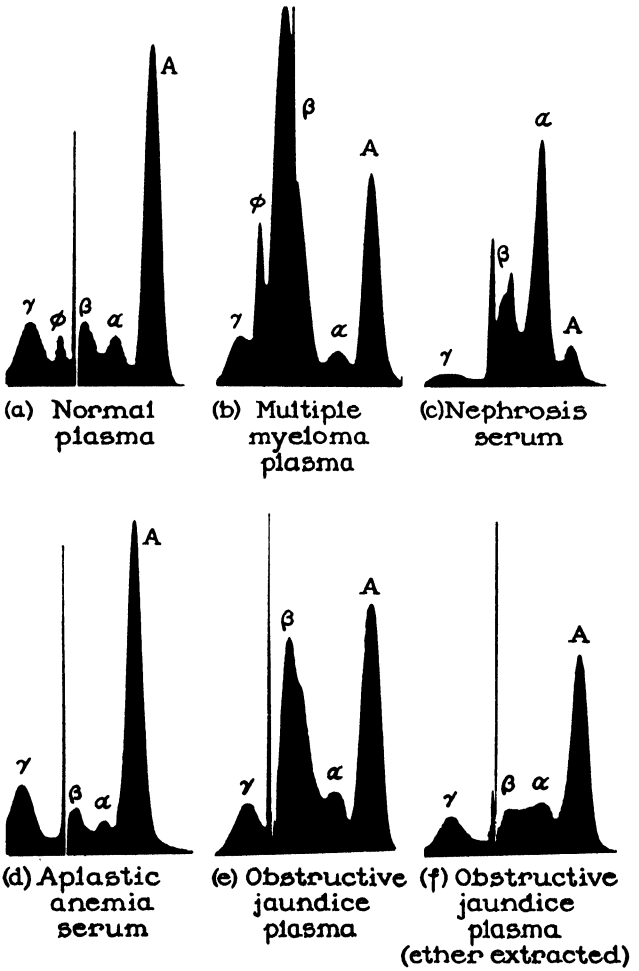


FIGURE 18. Electrophoretic patterns of some normal and pathological human blood sera and plasmas.

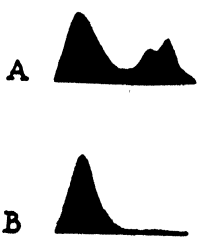


FIGURE 19. Electrophoretic patterns at pH 7.85 of a pituitary hormone, metakentrin. A, before final purification; B, after final purification.

Further Investigations Using Electrophoretic Technique

The following is a brief, but fairly complete, outline of investigations, using the Tiselius electrophoretic technique, not described above.

Electrophoretic analyses of normal and pathological human sera have been made by Kekwick,[22, 23] and by Luetscher.[43, 44] The proteins of other body fluids and tissues have been examined electrophoretically by Hesselvik,[19] by the authors,[89] and by Luetscher in the work mentioned above. In addition to the proteins normally present in such materials, high molecular weight nucleic acids and polysaccharides are frequently observed. These, too, are subject to electrophoretic characterization, as has been shown by Seibert, Pedersen, and Tiselius,[55] by Hesselvik,[18] by one of the authors in collaboration with Meyer and Chaffee,[48] and by Hall.[18] It is of considerable interest that the electrophoretic patterns of materials containing both protein and nucleic acid, or polysaccharide, frequently indicate the existence of complexes between the components. This is illustrated by the work of Stenhagen and Teorell [65] on mixtures of serum albumin and thymus nucleic acid, of Seibert and Watson [54, 56] on the components of the tubercle bacillus, of Chargaff, Ziff, and Moore [9] on mixtures of heparin and plasma, and our own work on ovalbumin-yeast nucleic acid mixtures.[41]

The application of electrophoretic analysis as a control of the salting-out methods for the separation and purification of proteins is illustrated by the work of Cohn and his associates.[10] They find a parallelism between the mobilities, at alkaline reaction, of the horse-serum proteins and the ease with which they may be salted out. The ranges of salt concentration through which precipitation of components occurs are, however, rather broad and overlap seriously. Svensson [67] has extended this observation to the sera of other animals and has further shown that the classical eu- and pseudo-globulins of serum contain all the electrophoretically separable globulins, *i.e.*, α, β, and γ globulin, in much the same proportions.

In the field of immunology the pioneer electrophoretic studies of Tiselius [71] and of Tiselius and Kabat [77] on hyperimmune sera showed that the antibody globulin generally migrated with a mobility similar to that of γ globulin, and hence appeared in the pattern as an enhancement of the latter. Their work and that of Pappenheimer, Lundgren, and Williams; [51] of Kekwick and Record; [25] of Fell, Stern, and Coghill,[11] and of van der Scheer, Wyckoff, and Clarke [78, 79] shows, however, that there are instances in which the antibody has a mobility intermediate between the mobilities of β and γ globulins. The last-mentioned work affords an excellent example of the use of electrophoresis in the analysis and production control of antisera. In antisera prepared by Thompson and Melnick [69] against a gonadotropic hormone, the antibodies were present in the γ globulin fraction. The work of Newell and his associates [50] indicates that the antibodies in certain allergies are also γ globulins. The "proteins" from ragweed responsible for the allergic reaction have been investigated electrophoretically by Abramson, Moore, and Gettner [3] (see also [4]).

Electrophoretic patterns of the sera of the more common laboratory animals have been obtained in connection with the immunological and salting-out studies mentioned above. Rat sera have been studied by Jameson and Tostado.[20, 21] Patterns of mouse sera have been obtained by Bourdillon and Lennette [7] in connection with their studies of the influenza virus.

We have already discussed an application of the electrophoretic method in the field of hormone chemistry, namely, in the purification of metakentrin. Other studies on the pituitary hormones are those of Shipley, Stern, and White [62] and of Li, Lyons, and Evans [30, 31, 32] on the lactogenic agent, and of van Dyke and his associates [80] on the oxytocic, pressor, and antidiuretic principles. Hall [14] has made an electrophoretic study of insulin.

With the aid of electrophoretic studies Tiselius, Henschen, and Svensson [74] have shown that the crystalline enzyme, pepsin, is an acid protein, *i.e.*, it migrates as an

anion at all accessible values of pH. Herriott, Desreux, and Northrop [17] have further shown that samples of pepsin can be prepared that are electrophoretically homogeneous and have a maximum enzyme activity. Ribonuclease and cytochrome C are other enzymes that have been characterized electrophoretically, the former by Rothen [58] and the latter by Theorell and Akesson.[68]

The difficulties that have been encountered in the application of the electrophoretic method to vaccinia (elementary bodies) have already been mentioned [60] and are further illustrated by the work of McFarlane.[45] Important results have, nevertheless, been obtained with this material by Shedlovsky and Smadel.[61] No comparable difficulties, on the other hand, have been experienced in the electrophoretic examination of the virus proteins, *i.c.*, bushy stunt virus by McFarlane and Kekwick,[46] of alfalfa mosaic virus by Lauffer and Ross,[29] and of papilloma virus by Sharp and his associates.[57] Miller and Stanley [49] have used electrophoresis in a comparative study of tobacco mosaic virus and certain of its derivatives.

. Conclusion

The authors have failed if they have not indicated that the field of electrophoresis is rapidly progressing both in the development of the method and in its application. They do not profess to have said the last word on the subject, and will be disappointed if this chapter is not somewhat out of date by the time it appears.

References

1. Abramson, H. A., "Electrokinetic Phenomena and Their Application to Biology and Medicine," New York, Chemical Catalog Co., Inc. (Reinhold Publishing Corp.), 1934.
2. Abramson, H. A., *Ann. N. Y. Acad. Sci.*, **39**, 121 (1939).
3. Abramson, H. A., Moore, D. H., and Gettner, H. H., *Proc. Soc. Exptl. Biol. and Med.*, **46**, 153 (1941).
4. Abramson, H. A., Moore, D. H., Gettner, H., Gargarin, J., and Jennings, L., *J. Am. Chem. Soc.*, **62**, 1627 (1940).
5. Blix, G., *Z. ges. exp. Med.*, **105**, 595 (1939).
6. Blix, G., *J. Biol. Chem.*, **137**, 495 (1940).
7. Bourdillon, J., and Lennette, E. H., *J. Exptl. Med.*, **72**, 11 (1940).
8. Burton, E. F., *Phil. Mag.* (6), **11**, 425 (1906).
9. Chargaff, E., Ziff, M., and Moore, D. H., *J. Biol. Chem.*, **139**, 383 (1941).
10. Cohn, E. J., McMeekin, T. L., Oncley, J. L., Newell, J. M., and Hughes, W. L., *J. Am. Chem. Soc.*, **62**, 3386 (1940).
11. Fell, N., Stern, K. G., and Coghill, R. D., *J. Immunol.*, **39**, 223 (1940).
12. Gutman, A. B., Moore, D. H., Gutman E. B., McClellan, V., and Kabat, E. A., *J. Clin. Investigation*, **20**, 765 (1941).
13. Hall, J. L., *J. Am. Chem. Soc.*, **63**, 794 (1941).
14. Hall, J. L., *J. Biol. Chem.*, **139**, 175 (1941).
15. Hektoen, L., and Cole, A. G., *J. Infectious Diseases*, **42**, 1 (1928).
16. Henry, D. C., and Brittain, J., *Trans. Faraday Soc.*, **29**, 798 (1933).
17. Herriott, R. M., Desreux, V., and Northrop, J. H., *J. Gen. Physiol.*, **23**, 439 (1939-40).
18. Hesselvik, L., *Skand. Arch. Physiol.*, **82**, 151 (1939).
19. Hesselvik, L., *Acta med. Scand.*, **101**, 461 (1939).
20. Jameson, E., and Alvarez-Tostado, C., *J. Phys. Chem.*, **43**, 1165 (1939).
21. Jameson, E., and Alvarez-Tostado, C., *Proc. Soc. Exptl. Biol. and Med.*, **40**, 476 (1939).
22. Kekwick, R. A., *Biochem. J.*, **33**, 1122 (1939).
23. Kekwick, R. A., *Trans. Faraday Soc.*, **36**, 47 (1940).
24. Kekwick, R. A., *Biochem. J.*, **34**, 1248 (1940).
25. Kekwick, R. A., and Record, B. R., *Brit. J. Exptl. Path.*, **22**, 29 (1940).
26. Kohlrausch, F., *Ann. Physik*, **62**, 209 (1897).
27. Lamm, O., *Nova Acta Reg. Soc. Sci. Upsaliensis*, Series IV, **10**, No. 8 (1937).
28. La Rosa, W., *Chemist-Analyst*, **16**, 3 (1927).
29. Lauffer, M. A., and Ross, A. F., *J. Am. Chem. Soc.*, **62**, 3296 (1940).
30. Li, C. H., Lyons, W. R., and Evans, H. M., *J. Gen. Physiol.*, **23**, 433 (1939-40).
31. Li, C. H., Lyons, W. R., and Evans, H. M., *J. Am. Chem. Soc.*, **62**, 2925 (1940).
32. Li, C. H., Lyons, W. R., and Evans, H. M., *J. Biol. Chem.*, **140**, 43 (1941).
33. Longsworth, L. G., *J. Am. Chem. Soc.*, **52**, 1897 (1930).
34. Longsworth, L. G., *J. Am. Chem. Soc.*, **61**, 529 (1939).
35. Longsworth, L. G., *Ann. N. Y. Acad. Sci.*, **41**, 267 (1941).
36. Longsworth, L. G., *Chem. Rev.*, **30**, 323 (1942).
37. Longsworth, L. G., Cannan, R. K., and MacInnes, D. A., *J. Am. Chem. Soc.*, **62**, 2580 (1940).
38. Longsworth, L. G., and MacInnes, D. A., *Chem. Rev.*, **24**, 271 (1939).
39. Longsworth, L. G., and MacInnes, D. A., *J. Exptl. Med.*, **71**, 77 (1940).
40. Longsworth, L. G., and MacInnes, D. A., *J. Am. Chem. Soc.*, **62**, 705 (1940).
41. Longsworth, L. G., and MacInnes, D. A., *J. Gen. Physiol.*, **25**, 507 (1941-42).
42. Longsworth, L. G., Shedlovsky, T., and MacInnes, D. A., *J. Exptl. Med.*, **70**, 399 (1939).
43. Luetscher, J. A., Jr., *J. Clin. Investigation*, **19**, 313 (1940).
44. Luetscher, J. A., Jr., *J. Clin. Investigation*, **20**, 99 (1941).
45. McFarlane, A. S., *Trans. Faraday Soc.*, **36**, 257 (1940).
46. McFarlane, A. S., and Kekwick, R. A., *Biochem. J.*, **32**, 1607 (1938).
47. MacInnes, D. A., "Principles of Electrochemistry," New York, Reinhold Publishing Corp., 1939.

48. Meyer, K., and Chaffee, E., *J. Biol. Chem.*, 133, 83 (1940).
49. Miller, G. L., and Stanley, W. M., *Science*, 93, 428 (1941).
50. Newell, J. M., Sterling, A., Oxman, M. F., Burden, S. S., and Krejci, L. E., *J. Allergy*, 10, 513 (1939).
51. Pappenheimer, A. M., Jr., Lundgren, H. P., and Williams, J. W., *J. Exptl. Med.*, 71, 247 (1940).
52. Philpot, J. S. L., *Nature*, 141, 283 (1938).
53. Rothen, A., *J. Gen. Physiol.*, 24, 203 (1940-41).
54. Seibert, F. B., *J. Biol. Chem.*, 133, 593 (1940).
55. Seibert, F. B., Pederson, K. O., and Tiselius, A., *J. Exptl. Med.*, 68, 413 (1938).
56. Seibert, F. B., and Watson, D. W., *J. Biol. Chem.*, 140, 55 (1941).
57. Sharp, D. G., Taylor, A. R., Beard, D., and Beard, J. W., *Proc. Soc. Exptl. Biol. and Med.*, 46, 419 (1941).
58. Shedlovsky, T., Rothen, A., Greep, R. O., van Dyke, H. B., and Chow, B. F., *Science*, 92, 178 (1940).
59. Shedlovsky T., and Scudder, J., *J. Exptl. Med.*, 75, 119 (1942).
60. Shedlovsky, T., and Smadel, J. E., *J. Exptl. Med.*, 72, 511 (1940).
61. Shedlovsky, T., and Smadel, J. E., *J. Exptl. Med.*, 75, 165 (1942).
62. Shipley, R. A., Stern, K. G., and White, A., *J. Exptl. Med.*, 69, 785 (1939).
63. Sørensen, S. P. L., and Høyrup, M., *Compt. rend. lab. Carlsberg*, 12, 12 (1917).
64. Stenhagen, E., *Biochem. J.*, 32, 714 (1938).
65. Stenhagen, E., and Teorell, T., *Trans. Faraday Soc.*, 35, 743 (1939).
66. Svensson, H., *Kolloid-Z.*, 87, 181 (1939).
67. Svensson, H., *J. Biol. Chem.*, 139, 805 (1941).
68. Theorell, H., and Akesson, A., *J. Am. Chem. Soc.*, 63, 1804 (1941).
69. Thompson, K. W., and Melnick, J. L., *Endocrinology*, 28, 723 (1941).
70. Tiselius, A., "The Moving Boundary Method of Studying the Electrophoresis of Proteins," Dissertation, Upsala, 1930.
71. Tiselius, A., *Biochem. J.*, 31, 1464 (1937).
72. Tiselius, A., *Trans. Faraday Soc.*, 33, 524 (1937).
73. Tiselius, A., and Eriksson-Quensel, I. B., *Biochem. J.*, 33, 1752 (1939).
74. Tiselius, A., Henschen, G. E., and Svensson, H., *Biochem. J.*, 32, 1814 (1938).
75. Tiselius, A., and Horsfall, F. L., Jr., *Ark. Kem. Mineral. o. Geol.*, 13A, No. 18 (1939).
76. Tiselius, A., and Horsfall, F. L., Jr., *J. Exptl. Med.*, 69, 83 (1939).
77. Tiselius, A., and Kabat, E. A., *J. Exptl. Med.*, 69, 119 (1939).
78. van der Scheer, J., Wyckoff, R. W. G., and Clarke, F. H., *J. Immunol.*, 39, 65 (1940).
79. van der Scheer, J., Wyckoff, R. W. G., and Clarke, F. H., *J. Immunol.*, 40, 173 (1941).
80. van Dyke, H. B., Chow, B. F., Greep, R. O., and Rothen, A., *Am. J. Physiol.*, 133, Proceedings P473 (1941).
81. Weber, H., "Die partiellen Differential-Gleichungen der mathematischen Physik," 5te Aufl., Vol. I, pp. 195-203, 503-527, Braunschweig, 1910.

High-speed Centrifugation

Ultracentrifugal Analysis and the Centrifugal Fractionation and Concentration of Molecular Particles

E. G. PICKELS

Laboratories of the International Health Division of The Rockefeller Foundation, New York, N. Y.

Introduction

Centrifugal force was used at least 1000 years ago for extracting tung oil, and today it is still serving industry in many important roles, a notable example being that of the common cream separator.[81] In certain fields of research, the familiar motor-driven laboratory centrifuge has become indispensable as a simple tool for fractionating the particulate matter suspended in liquid media. Within comparatively recent years there have appeared new implements of research in the form of high-speed centrifuges capable of separating from solution small particles of molecular dimensions, *e.g.*, proteins and filterable viruses.

With regard to the origin of this new field of investigation, the most important historical event was the development of a small, electrically-driven "optical centrifuge" by Svedberg and Nichols [132] at the University of Wisconsin in 1923. Svedberg and Rinde [134] used the machine to study size distribution among gold sols and proposed for it the name "ultracentrifuge," denoting an instrument by means of which sedimentation in a centrifugal field could be measured quantitatively. However, in re-

cent years it has become the practice of many investigators to associate the term "ultracentrifuge" with any type of high-speed centrifuge rather than with the original connotation of quantitative measurement.

All significant developments in the field of ultracentrifugation prior to 1940 have been reviewed in "The Ultracentrifuge," compiled by Thé Svedberg and Kai O. Pedersen, in cooperation with J. H. Bauer and E. G. Pickels, G. Boestad, E. O. Kraemer and J. B. Nichols, O. Lamm, A. S. McFarlane, and R. Signer.[188] A thorough bibliography of the literature pertaining to the construction and applications of high-speed centrifuges is included. At the end of the present paper will be found a less extensive but more up-to-date bibliography, along with an index arranged according to subject. Especial attention has been given to recent improvements in optical methods and to developments and experiments concerned with the newer electrical and air-driven centrifuges.

Theory of Convection-free Sedimentation and the Determination of Particle Size and Weight

The concentration distribution of small, monodisperse particles settling through a fluid in a uniform field of force has been investigated theoretically by Mason and Weaver [60] and by Weaver.[147] Expressions giving the concentration of diffusible particles at any time and position within a liquid subjected to a centrifugal field of force have been contributed by Faxen,[35] Lamm,[55] and Archibald.[8]

Svedberg [121] has derived mathematical equations which show the relationship existing between the molecular weight of homogeneous monodisperse particles and their sedimentation characteristics under certain special conditions in an ultracentrifuge. These equations apply strictly only to dilute solutions of particles which are large in comparison with the molecules constituting the medium, and which are electrically neutral with respect to the medium. Consider a suspension of particles of molecular weight, M, revolving in a centrifuge at an angular velocity of ω radians per second. Assume that the particles under observation are far enough from the bounding surfaces of the fluid to be uninfluenced by alterations of particle concentration in these regions. The net force acting on a mole of these particles at a distance x from the axis of rotation is simply the difference between their combined centrifugal weight and the buoyancy exerted by the displaced medium, *i.e.*,

$$F_c = M\omega^2 x - M\frac{\rho\omega^2 x}{\sigma} \text{ or } F_c = M\omega^2 x - MV\rho\omega^2 x \tag{1}$$

where ρ is the density of the medium and σ is the density of the particles. The latter is equivalent to $1/V$, where V is the partial specific volume of the particles. Disregarding the negligible forces due to the slight acceleration in the speed of the particles with time, the force F_c is exactly counterbalanced by a frictional force F_f which is proportional only to the sedimentation velocity of the particles and to a factor dependent upon the extent and shape of their surface. The frictional force on each mole per unit speed is called the frictional coefficient f_s. Thus,

$$F_c = M\left(1 - \frac{\rho}{\sigma}\right)\omega^2 x = F_f = f_s dx/dt \tag{2}$$

where t is the time. In this expression $\frac{dx}{dt}\left(\frac{1}{\omega^2 x}\right)$ is the sedimentation velocity per unit field of force, and is usually denoted by s. Rewriting the equation, we have

$$s = \frac{M\left(1 - \frac{\rho}{\sigma}\right)}{f_s} \text{ or } sf_s = M\left(1 - \frac{\rho}{\sigma}\right) \tag{3}$$

The diffusion constant D, which expresses the rate at which the particles will diffuse, is also inversely proportional to a molar frictional coefficient f_D:

$$D = \frac{RT}{f_D} \tag{4}$$

R is the gas constant and T the absolute temperature. A discussion of methods for measuring D are given in a recent paper by Neurath.[78] It should be noted that a change in the mass of particles does not affect the diffusion rate directly but only in so far as it is associated with a change in the frictional coefficient. This is true since the frictional forces are large in comparison with the inertial forces involved. Theoretical evidence, as well as experimental evidence based on sedimentation at different speeds, justifies the assumptions that the random orientation of molecular particles during either diffusion or sedimentation is not appreciably altered, and that for the same medium at the same temperature T, $f_s = f_D$ to a very close approximation. Eliminating the frictional coefficient from Equations 3 and 4, one obtains Svedberg's formula for the molecular weight as determined by what is known as the sedimentation velocity method:

$$M = \frac{RTs}{D\left(1 - \frac{\rho}{\sigma}\right)} \tag{5}$$

In general, sedimentation and diffusion experiments are performed at different temperatures, and for the sake of comparisons, it is customary to correct measured values of D and s to some standard temperature, usually $T = 293°$ Absolute. Svedberg has standardized conditions even further, and expresses each measurement of sedimentation in terms of a "sedimentation constant," s_{20}, which is the rate the material would have in a hypothetical medium having the density and viscosity of water at 20° C or 293° Absolute:

$$s_{20} = \frac{s\eta(\sigma - \rho_{w20})}{\eta_{w20}(\sigma - \rho)} \tag{6}$$

where η denotes viscosity. Sedimentation constants are usually given in S or Svedberg units, *i.e.*, in units of centimeters per second per unit field of force multiplied by 10^{13}. For example, in the case of hemoglobin:

$$\begin{aligned} s_{20} &= 4.4 \times 10^{-13} \text{ cm/sec/dyne/gm} \\ &= 4.4 \times 10^{-13} \text{ cm/sec/unit field of force} \\ &= 4.4 \times 10^{-13} \text{ sec} \qquad \text{(absolute units)} \\ &= 4.4S \qquad \text{(Svedberg units of sedimentation)} \end{aligned}$$

If the particles are known to be spherical and the value of σ to be unaffected by hydration or solvation, the molecular weight can be determined without the diffusion constant by substituting into Equation 2 the value of the molar frictional coefficient for spherical particles:

$$f_0 = 6\pi\eta N\left(\frac{3M}{4\pi N\sigma}\right)^{1/3} \tag{7}$$

in which N is the Avogadro constant. For this expression, we are indebted to G. G. Stokes,[120] who proved theoretically that the resisting force against a slowly moving small sphere is equal to a constant (6π) multiplied by the product of its radius, its velocity, and the viscosity of the medium. It is the linear relationship with velocity which renders the above expression valid even when the particles are subject to

Brownian movements, and the measured sedimentation rate is actually a resolved averaged rate of many short movements in many directions. A more direct method of computing the radius a of a spherical particle, after which the value of M can be easily obtained, is to apply Stokes' law in the following form:

$$a^2 = \frac{9\eta s}{2(\sigma - \rho)} \tag{8}$$

The sedimentation equations given above refer strictly only to particles isolated from the bounding surfaces of the fluid. However, Svedberg and Rinde [184] have shown that for the purposes of measurement use can be made of a small sector-shaped cell which has two flat walls oriented at right angles to the axis of rotation and two others which, if extended, would intersect along the axis. Particles originally close to any of these walls continue during sedimentation to pursue an average course parallel to the wall, so that their radial movement proceeds just as if the bounding surfaces were further removed. However, with the passage of time the concentration of particles does progressively increase near the bottom of the cell and decrease at the top; but if the centrifugal force is sufficiently high, measurements of rate can still be made directly, as will be shown. The particles originally at the top or inner surface of the fluid migrate together, forming a "rear line of march," or moving boundary, which demarcates the supernatant fluid and sedimenting solute. The sedimentation rate is determined by measuring the position of this boundary at intervals and using the integrated form of $\frac{dx}{dt}\left(\frac{1}{\omega^2 x}\right)$, *i.e.*, $\frac{\ln x_2 - \ln x_1}{\omega^2(t_2 - t_1)}$ to obtain a mean value for s. Generally the boundary becomes progressively less well defined with time because of a superimposed diffusion of particles across the boundary. However, the position which the boundary would have had in the absence of diffusion is simply the level at which the concentration is one-half that in the lower, unaffected sections. When the rate of sedimentation is rapid enough in comparison to the diffusion process, there exists for some time, between the diffuse boundary and the section of the cell in which the particles are accumulating against the wall, a region in which the concentration is nearly uniform throughout (see Fig. 1). The concentration in this region decreases at a steady rate during a constant-speed centrifugation, because of diverging radial movements of the particles and because of the fact that the centrifugal force, and consequently the speed of migration, increases with the distance from the axis of rotation. The relation between this concentration C_t at any time, t, and the original concentration C_0 has been shown by Svedberg and Rinde [184] to be given by:

$$C_t = C_0 \frac{x_0^2}{x_t} \tag{9}$$

where x_0 and x_t are the respective radial distances of the meniscus and the boundary at time t.

If sedimentation is allowed to progress for a sufficiently long time, usually at a rotational speed considerably lower than that used for the sedimentation velocity method, a state of equilibrium is reached in which the tendency of the particles to sediment is exactly counterbalanced by their tendency to diffuse back into the medium. The molecular weight can then be determined by comparing the concentrations C_1 and C_2 of particles at two radial distances x_1 and x_2. The formula for the

sedimentation equilibrium method as originally given by Svedberg,[121] and as later proved thermodynamically by Tiselius,[140] is:

$$M = \frac{2RT \ln (C_2/C_1)}{\left(1 - \frac{\rho}{\sigma}\right)\omega^2(x_2^2 - x_1^2)} \tag{10}$$

With proteins, for example, several days of continuous centrifugation are generally required before the concentration distribution becomes sufficiently constant for the equilibrium formula to be applied. An extension of the mathematical work

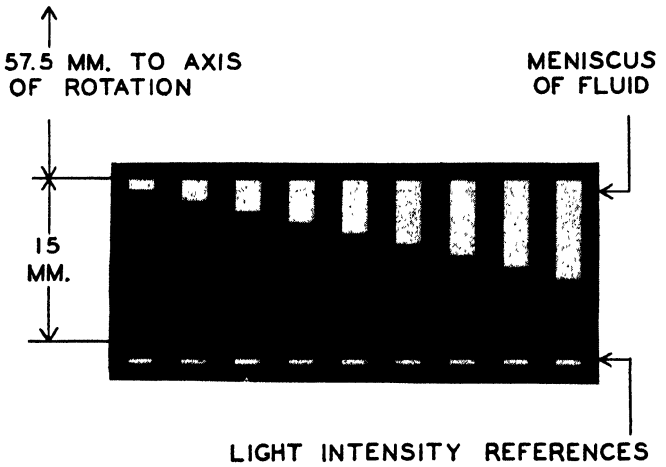


FIGURE 1. Sedimentation of monodisperse hemocyanin in the serum of *Limulus polyphemus*, as shown by the absorption method. Speed, 18,000 rpm; interval between photographs, 15 min; illumination, 3650 (Å) line of mercury; cell thickness, 3 mm; approximate concentration, 1.45 per cent.

of Archibald [8] in obtaining solutions for the general differential equations of sedimentation might permit direct determinations of M with considerably shorter runs. The question of sedimentation in fluid columns which are either not convection-free or not sector-shaped is discussed later in this article.

Density, Shape, Hydration, and Electrical Charge of Particles, and Their Influence on Sedimentation

In the equations given for the determination of molecular weight or particle size it has been assumed, as already stated, that the particles are electrically neutral with respect to the medium, *i.e.*, that they are not electrostatically attracted and influenced by smaller diffusible ions which sediment slower and diffuse faster. However, Tiselius [141] has shown that where electrical dissociation is present, the effect on the sedimentation rate of the principal particle can be rendered negligible by the addition of sufficient non-sedimenting electrolyte. About 0.2M sodium chloride is usually used to repress this Donnan effect.

Of the several quantities which enter into the computation of molecular weight from sedimentation data, the density of the particles is the one which has aroused the greatest uncertainty because it does not lend itself readily to direct measurement. The density of dried or crystallized material can be determined by measuring with a pycnometer the volume displaced by a known weight of substance. However, because of the possibility of hydration there is considerable doubt as to whether this represents the effective density of the particles sedimenting in a centrifugal field.

This problem has been treated mathematically by Lamm [55] and Kraemer,[58] and was recently discussed in simple physical terms by the author.[89] It has been found theoretically that only inappreciable error will generally be produced by unknown amounts of hydration or solvation when molecular weights are computed by either Equation 5 or Equation 10, even though σ is taken for unhydrated particles. Specifically, in dilute solutions the proportional error in a determination is $(Z - 1)$, where Z is simply the ratio between the true weight of a mole and the true weight increased by the difference between the weight of the bound material and the weight of an equal volume of the suspending medium. The reason for this relationship is that hydration changes the net centrifugal force acting on a particle only slightly, and increases the frictional coefficient equally in both sedimentation and diffusion. It should be noted that the molecular weight determined by either Equation 5 or Equation 10 is, then, the molecular weight of particles in such a state that their density is the same as that which characterizes the particulate material when the weight of a sample is determined.

When the diffusion constant D and the sedimentation rate s are known for the same temperature T, the values of M, f, and f_0 can be computed from the equations given above. The expression f_0 represents the minimum possible frictional coefficient, *i.e.*, that which particles having the given sedimentation rate would have if they were unsolvated and spherically shaped. The frictional constant is represented by f/f_0 which always has a value equal to or greater than unity. Most naturally occurring monodisperse proteins already studied appear to have constants considerably below 2. In the case of viruses particularly, however, values as high as 2.8 may be obtained.[86]

Frictional constants greater than unity can be caused by solvation or by lack of spherical symmetry. Usually both factors are present. Adair and Adair [1] have shown several proteins to be hydrated in amounts ranging from 14 to 20 per cent. With some of the larger viruses there is some evidence that the values may be somewhat higher.[112] When the frictional constant has been corrected for an assumed or measured degree of solvation by recomputation with properly adjusted values for σ and M, it is possible to gain some idea of the particle shape by the use of equations obtained from the works of Herzog, Illig, and Kudar [47] and of Perrin [84] (see also Ref. 133). These equations relate frictional constants with axial ratios for particles assumed to be either oblong or oblate ellipsoids. Observations on the double refraction produced by streaming may give information as to whether the particles are oblong or oblate, and to what extent. Axial ratios of many "globular" proteins are in the neighborhood of 4,[80] while values of the order of 40 have been found for some viruses.[86] Asymmetries may also be investigated by experiments based on measurements of viscosity or dielectric constant.[80]

The density of most proteins in the unhydrated state is very close to 1.33 gm/cc,[183] although for thymonucleohistone, as an example, values have been reported ranging from 1.41 [104] to 1.50.[100] Some large viruses have been estimated to have hydrated densities of about 1.16 and unhydrated densities near 1.25 to 1.30.[112] In certain cases, some idea of the hydrated density can be obtained by ultracentrifugation in media of various densities.[112]

From these considerations, it will be seen that Equation 8 can be used directly without supplementary information to determine from a measured sedimentation rate the minimum possible values for the volume and mass of the sedimenting particles. In the case of proteins, even the value of σ may be assumed with some assurance of a fair approximation.

Optical Ultracentrifuges of Svedberg and Associates

In order to obtain higher rotational speeds for sedimentation velocity measurements, Svedberg and Lysholm [181] in 1927 substituted an oil drive for the electric

drive of the centrifuge. In the present improved model,[29, 133] the steel rotor is driven by oil under pressure and rotates in a reduced atmosphere of hydrogen gas. The sector-shaped cell (see Fig. 2) generally accommodates a column of solution about 16 mm long and is spun in the rotor at speeds up to approximately 70,000 rpm. Smaller rotors with shorter columns have been run at much higher speeds but have been found less efficient for general purposes.[133] The thickness of the sector-shaped

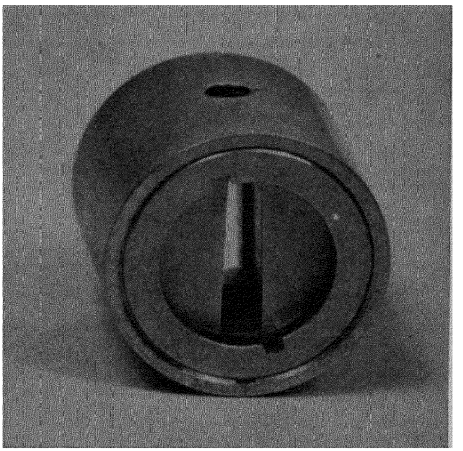


FIGURE 2. Photograph of transparent cell used in ultracentrifuges of the vacuum type. Diam. of cell, 25.4 mm; capacity, 0.6 cc; thickness of solution column in direction of light path, 1 cm. Quartz windows permit passage of ultraviolet light. Evaporation through the small filling hole is prevented by a rubber gasket and a screw plug.

solution column parallel to the axis of rotation is usually less than 12 mm, and the average width is only a few millimeters. The center of the cell is 65 mm from the axis of rotation, which is horizontal. The cell is provided with plane-parallel quartz windows which make it possible to direct light rays through the revolving solution in a direction parallel to the axis of rotation. Light passing through openings in a counterbalance or dummy cell at radial levels above or below the solution column is unaffected by the sedimentation and furnishes a convenient reference (see Fig. 1).

In 1929 Svedberg and Sjögren[135] improved the early low-speed electrical ultracentrifuge used for equilibrium measurements by substituting a direct motor drive for the gear drive.[124, 133] The motor is of a self-balancing type used for spinning viscose thread in the artificial silk industry (Siemens, Berlin). It is fed with a three-phase A. C. current of variable frequency from a special generator and has a short-circuited squirrel-cage rotor, moving in ball bearings. The 15-cm rotor spins in an atmosphere of hydrogen. The transparent cell and counterbalance are of the same general character as those used with the oil-driven ultracentrifuge. Speeds up to about 20,000 rpm can be employed.

Optical Ultracentrifuges of the Vacuum Type

Although the Svedberg ultracentrifuges have served their intended purpose admirably, they are comparatively complicated and too expensive to be of service to many investigators. Within the past decade there have been developed considerably simpler ultracentrifuges of the so-called vacuum type, which can spin rotors of desirable dimensions under experimentally favorable conditions. The available speed is limited only by the strength of the material from which the rotor is made. It has been the practice to use easily machineable Duralumin in preference to the special steels employed in the oil-driven ultracentrifuge. The resolving power, *i.e.*, the ability to separate materials with sedimentation rates close together, is consequently slightly less than that obtained with the Svedberg machine,[133] but is considered close enough

to the practical limit not to warrant more expensive rotors for general purposes.

The first arrangement for spinning large rotors at high speed in a vacuum was described by Pickels and Beams [95] in 1935. A drawing from the original paper is reproduced in Fig. 3. From a small cone-shaped, air-driven and air-supported rotor

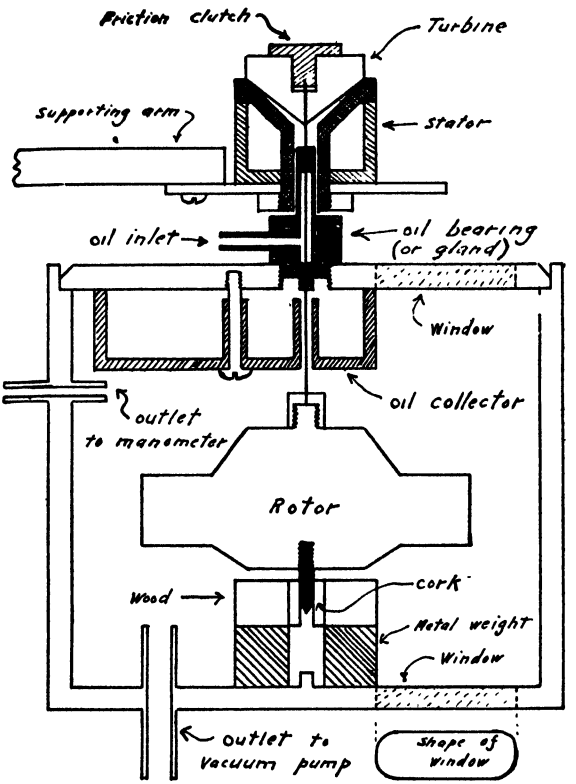


FIGURE 3. Drawing of first high-speed centrifuge of the vacuum type (Ref. 95). The turbine is supported and driven by compressed air. The motion is transmitted to the large rotor through a length of piano wire. Absence of air friction against the large rotor prevents heating and permits practical attainment of high speeds.

of the "spinning top" variety described later in this article, a length of straightened piano wire extended downward through an oil gland into a vacuum chamber and was there fastened to a larger rotor. This wire served as a drive shaft. With almost all the air resistance on the main rotor eliminated and with so little friction offered by the guiding bearings, high speeds could be obtained with little driving energy and without appreciable frictional heating. This was an important circumstance in the avoidance of thermal gradients which can cause convection currents within a liquid being centrifuged.

A little later Beams and Pickels [16] reported improvements in the apparatus and a successful sedimentation of hemoglobin. Adaptations were made with larger rotors by Biscoe, Pickels and Wyckoff, and they conducted a series of tests with circular rotors of the lighter alloys to ascertain what speeds could be safely tolerated.[26, 27] Bauer and Pickels [6, 7] introduced many refinements, including an independent air bearing, a system of decelerating by reversed air jets, and particularly an oval-shaped rotor (see Fig. 4) capable of withstanding considerably higher rotational speeds.

The sedimentation of egg albumin at 60,000 rpm was demonstrated. A more dependable air bearing was described in 1938 by the author.[88] References to improvements and alternative arrangements described by Beams, Hoxton, Chiles, and other investigators are listed in the subject index.

An ultracentrifuge of the modified Bauer-Pickels type [7, 88] has given consistently satisfactory service in our laboratory for several years. In Fig. 5 is shown a high-speed, vacuum-type centrifuge with the protective steel cylinder of the vacuum

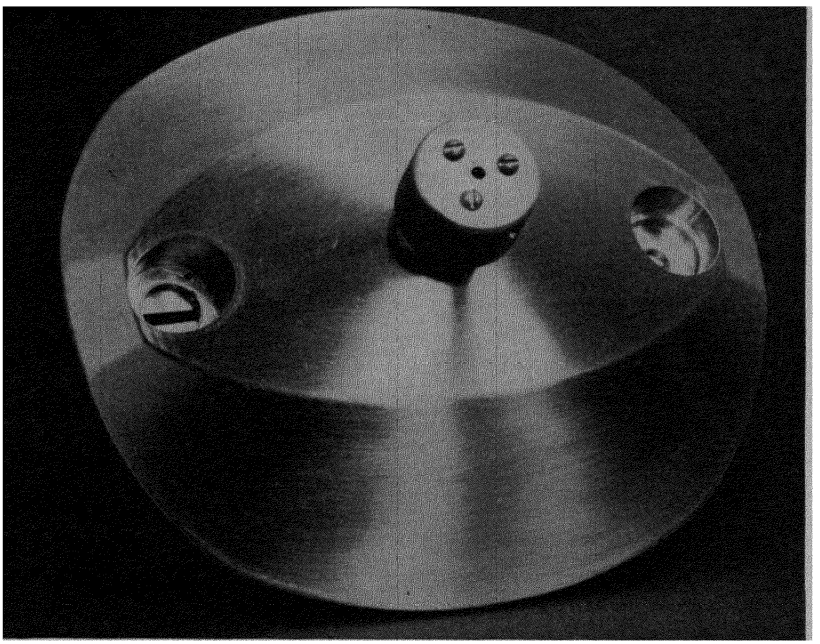


FIGURE 4. Oval-shaped Duralumin rotor of vacuum-type ultracentrifuge with cell and counterbalance shown in position. Maximum diameter of rotor is 18 cm. It withstands routine operation at 60,000 rpm, with an average centrifugal force in the cell 260,000 times that of gravity.

chamber removed. The driving mechanism is the "turret" type described by the author.[88] The rotor shown is designed for the concentration of biological material in inclined tubes. For ultracentrifugation work a Duralumin rotor such as that in Fig. 4 is used. Its maximum diameter is 18 cm. Rubber gaskets form vacuum-tight joints between the cylinder and the top and bottom steel plates of the chamber. The steel drive shaft is $\frac{1}{10}$ inch in diameter. Its flexibility permits the rotor to be self-balancing. Arrangements, not visible, are provided for collecting small amounts of oil passing through the main guide bearings. The cell [92] (see Fig. 2) accommodates a fluid column of 1.5 cm maximum length (radially to rotor) and 1 cm deep. Two sets of windows are provided in the top and bottom plates so that two beams of light can be directed vertically through the revolving cell for the purposes of photographing the sedimentation by two different optical methods to be described later. Front-surfaced mirrors inclined at 45° are arranged in line with the windows at a convenient distance above and below the centrifuge so that the camera and illuminating systems can be horizontal. With an air pressure of 70 lbs/in.^{2} the rotor can be accelerated to or decelerated from 60,000 rpm in about 13 minutes. At this speed the average centrif-

ugal force in the cell is 260,000 times gravity. It can be maintained at this speed with a pressure of 22 lbs/in^2 and a flow of 12 cubic feet of free air per minute. The temperature of the rotor rises less than 1° C per hour when a good vacuum is maintained. The speed is kept constant with little trouble when a system of air-pressure regulators is employed.[7] The speed can be measured by a variety of stroboscopic arrangements, although a comparatively inexpensive instrument such as the Strobotac (General Radio Co.) is well suited for the purpose, especially when multiples of the A. C. line frequency are used.

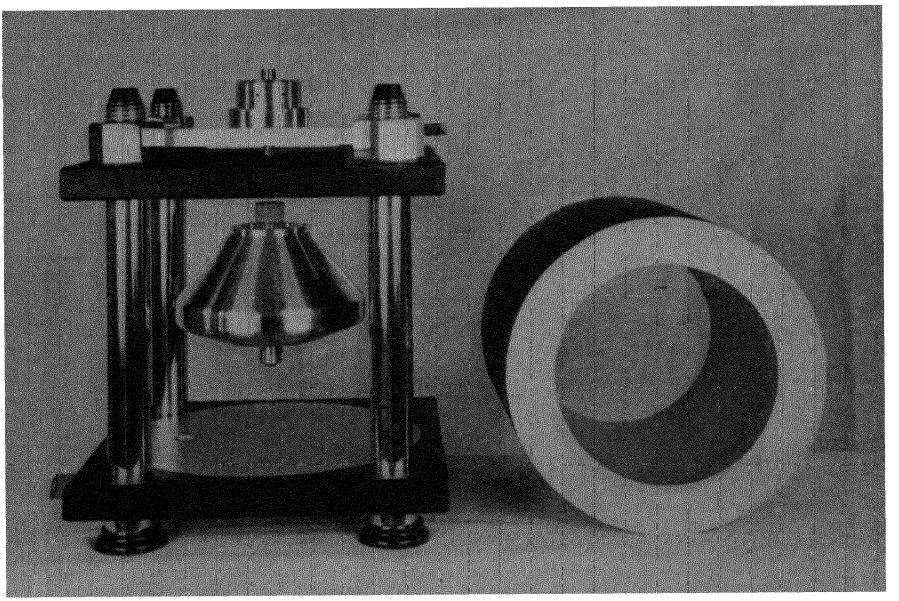


FIGURE 5. Photograph of a high-speed angle centrifuge of the vacuum type, with the cylindrical section of the steel vacuum chamber shown removed for the purpose of illustration. The turbine is driven and supported by compressed air. Rotors similar to the 20 cm one shown can be operated near 50,000 rpm. For the ultracentrifuge, top and bottom plates of the vacuum chamber are provided with windows to permit the passage of a light beam vertically through the revolving cell.

Beams and Black[12] and later Skarstrom and Beams[111] published descriptions of a high-speed electrical drive which can be substituted for the air drive. The weight of the rotating elements is supported magnetically. An elaborate vacuum-tube oscillator circuit is employed to supply a regulated electrical frequency to a small high-speed induction motor. Remarkably constant speeds can be maintained automatically with this machine. Apparently it is not as practical as the air drive for producing very rapid accelerations. Beams[11] has also reported a method for increasing the resolving power of an ultracentrifuge by displacing the solution column with more fluid progressively during centrifugation.

Optical Methods for Measuring Sedimentation

All the earlier sedimentation studies conducted by Svedberg and his collaborators were based on the light absorption method first described by Svedberg and Rinde.[184] Light of some suitable wave length which is absorbed by the sedimenting solute, but

not appreciably by the solvent, is directed through the cell. Filtered ultraviolet light from a mercury arc is usually employed for protein solutions. With visible light passing through a hemoglobin solution within the revolving cell, for example, the unaided eye sees a bright red annular band extending about the axis of rotation. The meniscus of the fluid appears as a fine line with the same center of curvature. The sedimenting boundary demarcating the clear supernatant and the solution is also concentric about the axis. To obtain a record of the sedimentation, it is necessary to photograph at intervals only a narrow cross-sectional strip of the total annular band of light. A camera of long focal length (100 cm) is used, in order to avoid errors of parallax and to give good focusing for all depths in the cell. A typical sedimentation of a large homogeneous protein is illustrated by the series of photographs in Fig. 1. An absorption photograph showing two sedimenting components is shown in Fig. 6.

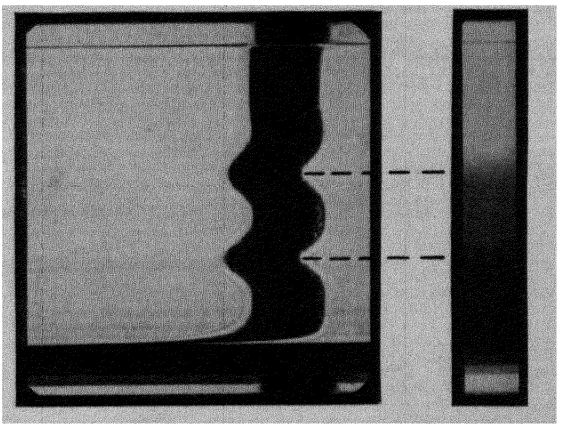


FIGURE 6. Sedimentation photographs illustrating correlation between absorption method and refractive index method. Material, polydisperse hemocyanin in diluted serum of *Limulus polyphemus;* speed 16,200 rpm; time 2.26 hrs after full speed; cell thickness, 1 cm; total protein concentration, 0.41 per cent.

Other optical methods which have to a great extent superseded the absorption method are based on the detection or measurement of deviations suffered by light rays passing through regions of refractive index gradient such as exist at sedimenting boundaries. The basis from which these methods have stemmed can be traced to the fundamental work of Weiner,[148] published in 1893. Lamm[56, 57] was the first to apply a refractive index method to the measurement of sedimentation. His so-called scale method can probably still be made subject to the fewest optical errors of any refractive index methods. It involves a great deal of tedious measurement and computation, and increasing attention is being given to other optical arrangements. Tiselius, Pedersen, and Eriksson-Quensel[142] have applied the principle of Toepler's schlieren method to the ultracentrifuge. A sedimentation boundary is located immediately by the appearance on the camera screen of a dark band at the region of steep concentration gradient. Concentration of a component cannot be measured in a practical manner, as with the scale method.

The cylindrical lens methods which have proved so useful originate in the work of Thovert.[189] Philpot[85] combined certain features of the Thovert method and the Toepler schlieren method to obtain the first automatically recorded refractive index curves with the ultracentrifuge. Andersson[2] later described an alternative arrange-

ment. Svensson [136, 137] modified Philpot's method for use with the Tiselius electrophoresis apparatus and made an extensive study of various theoretical aspects of the method. For example, he compared the pattern obtained with different inclined slits with those obtained by a simple inclined knife edge as used by Philpot. As with the absorption system, lenses having focal lengths of the order of 100 cm are used so that all depths within the cell are in focus.

Experience in our laboratory with a modification of Svensson's method has indicated that even with comparatively inexpensive optical parts a precision of measurement can be attained which is quite sufficient in view of the many minor uncertainties involved in the ultracentrifugal method. Special tests have shown that deviations of light produced at any level of the cell, at least deviations up to .02 radian, can be registered at the photographic plate with an error of less than 1 per cent, and that for all deviations in this range the image positions of every level in the cell can be produced with an accuracy of less than 1 per cent of the total cell height.

A schematic representation of the optical principles involved is given in the perspective drawing of Fig. 7. A camera lens is placed so as to form on a screen

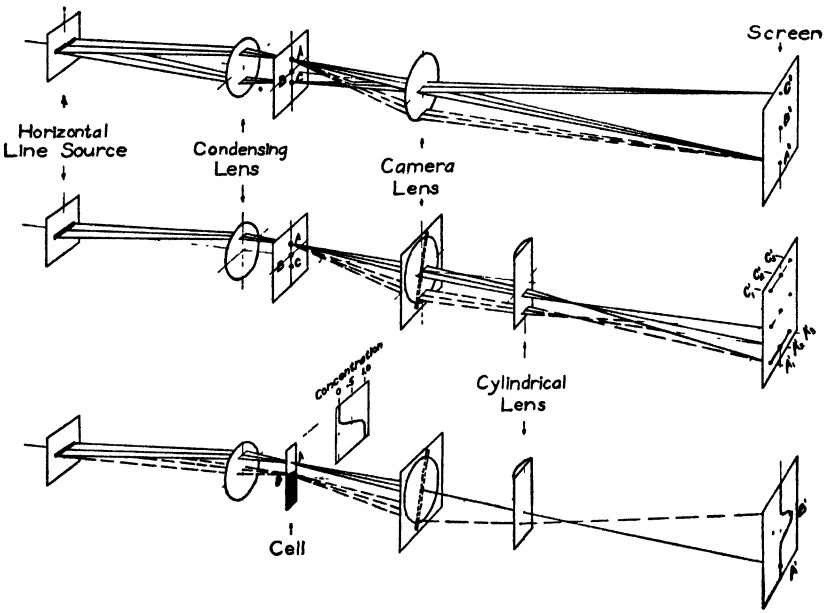


FIGURE 7. Perspective schematic drawing illustrating the working principle of a refractive-index method employing a cylindrical lens.

the images A′, B′, C′ of the three small holes A, B, and C in a thin diaphragm. A condensing lens is placed so as to focus the image of a horizontal line source on the front surface of the camera lens. Limit consideration to those light rays which pass through the small holes. It can be seen that the light eventually constituting each of the images A′, B′, C′ consists of rays which alternately diverge and converge in accordance with simple optical principles. Each set of rays may be thought of as making up a "light sheet" that varies in width but is always parallel to the horizontal line source. If the light sheet diverging from A toward the camera lens were deviated downward from the normal course at A, as indicated by the dotted lines, it would nevertheless converge at A′ by reason of the fact that the camera lens is

focused to produce such a result. An analogy holds for all other images. Suppose that a cylindrical lens with vertical axis is added to the system in such a position that a point light source at the front surface of the camera lens would be focused as a vertical line image at the screen. The effect of this lens on each of the original light sheets will be to converge it prematurely, so that it again diverges and forms a horizontal line image A_1' A_2' A_3' on the screen instead of a point image. The vertical positions of the images are not altered by the cylindrical lens. A downward deviation (dotted lines) of the sheet at A would not alter the position of the line image. Suppose that over the front surface of the camera lens is placed a mask provided with a narrow inclined slit. It can be seen that of the undeviated rays through A, only the central one is now able to pass through the camera lens and reach the screen, where it will appear as a point of light A_2', as contrasted to the previous horizontal line image A_1' A_2' A_3'. Since the undeviated light sheets through A, B, and C all converge in a common image of the light source at the mask, it follows that B and C will also be represented at the screen, when there is no deviation at the diaphragm, by light points B_2' and C_2', which are situated along a central vertical line passing through A_2'. However, if the light sheet from A, for example, is deviated downward sufficiently from its normal course, then only the ray at one edge of the light sheet with pass through the camera lens, and the light point corresponding to A will appear on the screen at A_3'. The horizontal distance by which any light point will be displaced from the central line C_2' A_2' is directly proportional to the downward displacement suffered by the corresponding light sheet at the diaphragm. The same holds true for light coming from B or from C, and thus the system is capable of translating a vertical deviation of light into a horizontal displacement of a light point on the screen without altering the point's vertical height which is characteristic of a certain level in the cell.

Consideration will show that the action of the system is not altered by substituting narrow horizontal slits for the holes A, B, and C. For any given deviation of a light sheet, more rays will pass through the camera lens, but since they do pass through at the same point the cylindrical lens converges them into a point focus again at the screen. Now suppose that a cell containing an immobilized sedimentation boundary is substituted for the diaphragm in such a way that the maximum refractive index gradient, *i.e.*, the center of the boundary, is in the position B. The corresponding point B′ will be the one suffering the maximum horizontal displacement from the center line on the screen. Light coming through the cell below and above the boundary, as at A, is not deviated and the corresponding light points fall on the center line. Displacements of light points corresponding to successive positions within the boundary are proportional to the respective refractive index gradients, with the result that the light points form a smooth continuous curve on the screen. The peak of the curve, if symmetrical, determines the position of the boundary; the area between the center line and the curve is proportional to the concentration of the material forming the boundary, assuming it is the only source of refractive gradients. With the ultracentrifuge there are extraneous gradients due to hydrostatic compression and distortion of the cell, and a curve incorporating these alone and called the base line must be obtained. The magnitude of this effect can be seen in the photographs of Fig. 9. If heavy salt ions are present in sufficient strength in the medium, they too produce gradients which must be taken into account by making separate ultracentrifugation runs with the medium alone.[188] In applying the method to the centrifuge it is important to limit the light passing through the revolving cell to a zone having a width of not more than a few millimeters in a tangential direction.

In our laboratory, the substitution of an opaque strip for the inclined open slit has produced certain advantages. In place of a line of light as described above, a dark band against a bright background indicates deviations of light at various

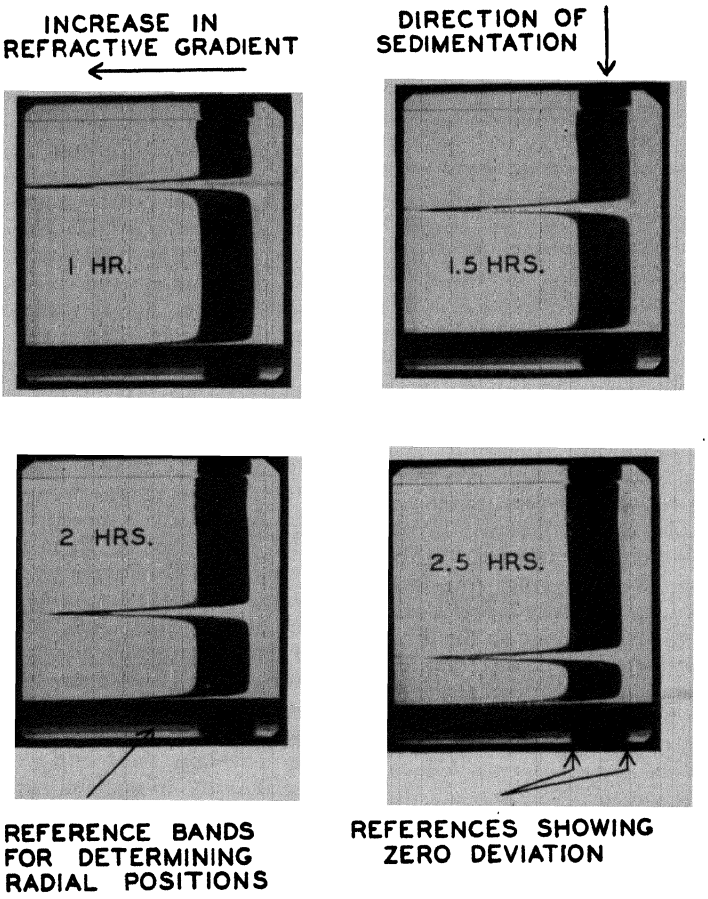


FIGURE 8. Pictures taken during the experiment illustrated in Fig. 1 and showing sedimentation in the same hemocyanin solution, as recorded by the refractive-index method.

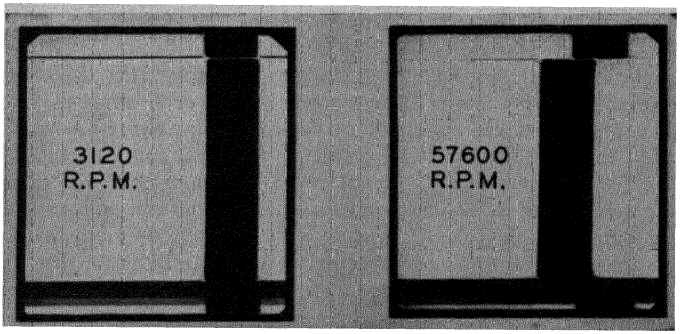


FIGURE 9. Photographs illustrating the gradients of refractive index introduced in a water-filled ultracentrifuge cell at high speed by hydrostatic compression and distortion of the cell.

depths in the cell. Measurements are taken from both edges of this band and averaged.

The photographs of Fig. 8 illustrate an application of the refractive index method employing a cylindrical lens. These were obtained during the same run as were those of Fig. 1. Fig. 6 is an example of the simultaneous sedimentation of two protein components and shows the correlation existing between photographs taken of the same material at the same time by the refractive index and absorption methods.

Experimental Considerations

For most cases the refractive index method is preferable to the absorption method, particularly in the study of mixtures and where it is desired to determine the absolute or relative concentrations of several components. There is a considerably greater proportional variation in the absorption coefficients of proteins in the ultraviolet region, for example, than there is in their specific refractive indices in the visible region.[89] As a matter of fact, good approximations of the concentrations of protein components of unknown optical properties can nearly always be made with the refractive index method. It is also an advantage to be able to observe the sedimentation on the screen with visible light during sedimentation. Furthermore, it is more expensive and less convenient to work with ultraviolet light, which requires quartz optical parts.

A good refractive index system is subject to fewer and less serious errors than is an absorption system. The pictures shown in Fig. 1 represent an unusually sharp boundary of a fairly concentrated, slowly diffusing material. Measurements can be made directly from the photographs. In actual practice, however, boundaries of smaller, more common proteins are more diffuse, such as those shown in Fig. 6. To analyze absorption photographs, it is then necessary to employ a microphotometer which will measure the photographic density of the plate at all levels in the cell picture. Results obtained must then be translated into measurements of concentration. The analysis is subject to all the usual photographic errors, as has been pointed out by Pedersen.[138] The apparent position of a diffuse boundary and the apparent concentration distribution throughout the cell can be greatly altered by photographic overexposure or underexposure and also by the use of a highly absorbing preparation in the cell. It is imperative to use monochromatic light of high purity and to illuminate the cell uniformly.

In some cases, however, the absorption method is to be preferred. For example, Pickels and Bauer[94] found that the virus of yellow fever in infected monkey serum showed a high differential absorption in the upper ultraviolet region, and they were able to study, by absorption, sedimentation boundaries which could not be detected with the refractive index system. Since the precision of the refractive method depends on the magnitude of the concentration gradients rather than on the concentration itself, the absorption method is sometimes more applicable with unusually ill-defined boundaries or with very inhomogeneous preparations. Infrared light can be employed in cases where visible light is too strongly absorbed by the solution.[145]

The sedimentation equilibrium method requires several days of continuous operation and is most useful for the investigation of purified preparations which are known to be homogeneous, since in a mixture, the contributions of several species to the distribution of total concentration are not easily differentiated.[188] The sedimentation velocity method, on the other hand, is ideal for the study of mixtures, since an individual boundary is obtained for each species whose rate is sufficiently different from that of the other components. The method is of great assistance in judging the purity and homogeneity or monodispersity of a sedimentable material. For most proteins, an experiment can be completed within a few hours. Runs should be made with several concentrations of material, and the results extrapolated to zero concen-

tration to determine the true sedimentation constant. It is known that in mixtures of proteins, for example, there is often an interaction between the different proteins, and the possibility of this effect must be taken into account. The problem has been discussed by Pedersen [83] and by Lundgren and Williams.[64]

In the earlier sedimentation velocity experiments of Svedberg determinations of the diffusion constant were attempted upon the material under study while it was sedimenting in the ultracentrifuge. An argument in favor of the procedure was that the environmental conditions were identical for both sedimentation and diffusion measurements. In later years it was found that the diffusion values obtained in this manner were not very reliable. Probable reasons for this have recently been discussed by the author.[89] Concentration is an important factor. In the photographs of Figs. 1 and 8, for example, the boundary shows with increasing time an increase in spread which is only a small proportion of the amount to be expected from measured values of the diffusion constant. The contrast with the spread of the more dilute boundaries in Fig. 6 is at once evident. The author [89] has tested a new method by which it is possible to make good diffusion measurements in the ultracentrifuge, even with mixtures when they give resolvable boundaries in the centrifugal field.

It is important to know accurately the temperature of the solution revolving in the ultracentrifuge so that its viscosity can be compared with that of the standard medium. With the vacuum-type ultracentrifuge, the temperature of the rotor is taken by thermocouple before and after each run. No appreciable error is introduced by taking the average of these readings as the temperature of the run, which is usually very near room temperature. For the investigation of materials which exhibit excessive instability under such conditions, Rothen [100] has surrounded his air-driven ultracentrifuge rotor with stationary refrigerated metal surfaces which permit him to maintain the material under study near 0° C. Hydrogen at a reduced pressure of about 0.1 mm of mercury furnishes the thermal conduction and permits the maintenance of a constant rotor temperature.

However, the maintenance of a constant *average* temperature within the rotor is of no consequence within itself in preventing disturbing convection currents within the solution. The principal potential sources of convection currents are temperature gradients which must exist within the cell to insure the passage of heat from the sections of the rotor exposed to the greatest proportion of frictional resistance to those exposed to smaller amounts, *i.e.*, from the faster moving periphery to the central zone. A rotor spinning in a vacuum represents the ideal case as far as convection-free sedimentation is concerned. It should be remembered that the convection action is enormously magnified by high centrifugal force and can interfere seriously with the sedimentation of small particles.[93] Beams [9] and the author [89] have shown that the compressional gradient produced by the centrifugal field is of little assistance in inhibiting convection. A method of testing for thermal convection currents has been suggested in a recent article.[98] Slight misalignments of the ultracentrifuge cell can also cause convective disturbances.[92]

While there is some uncertainty as to whether the viscosity of the solvent rather than that of the solution represents the correct value for use in the Svedberg formula, the problem is of little consequence if rates are extrapolated to zero concentration.

Typical Applications of the Optical Ultracentrifuge

Svedberg and his collaborators have investigated a great number of high molecular weight substances, especially proteins,[125] with the electrically driven and oil driven ultracentrifuges. Although the newer ultracentrifuges of the vacuum type have been employed for a comparable variety of purposes, their principal use to date appears to have centered around the study of filterable viruses. A complete bibliography for the whole field of ultracentrifugation will be found elsewhere [133] and the references to many typical experiments will be found listed in the subject index of

this article. To mention a few: McFarlane [69-78] has made an extensive study of the components of normal and pathological sera; Svedberg and Hedenius [130] have compiled a great deal of information regarding the respiratory proteins; studies have been made by others to determine the effects on various proteins of desiccation,[80] suspension in heavy water,[128] association with other proteins,[88] subjection to ultraviolet light and x-rays,[127, 102] and treatment with x-rays.[102] That molecular weights as low as 14,000 can be determined conveniently by the sedimentation velocity method with the vacuum-type ultracentrifuge is illustrated by the work of Rothen [98] with ribonuclease.

Among the viruses which have been studied in the optical ultracentrifuge are bushy stunt virus protein,[74, 58, 61, 59] the elementary bodies of vaccine virus,[96, 112] tobacco mosaic virus,[117, 58] chicken tumor agent,[119] the virus of equine encephalomyelitis,[138] alfalfa mosaic virus,[60] yellow fever virus,[94] the virus of mouse encephalomyelitis,[36] cucumber virus,[52] and rabbit papilloma virus.[79] Published accounts of these investigations describe many interesting experimental data regarding the size, shape, density, and state of dispersion of viruses, as well as their physical reaction to various hydrogen ion concentrations and other agents.

Certain naturally occurring carbohydrates [25, 129, 38] have been examined in the ultracentrifuge and some have been found by Svedberg and Gralén [129] to be nearly monodisperse. Pedersen [82] has applied the sedimentation equilibrium method to the study of such low molecular weight substances as small inorganic salts. Some detergents [114] have been found by ultracentrifugal analysis to form large micelles which sediment in the centrifuge. In the case of digitonin, the micelles were found to be quite monodisperse. Further references are given in the subject index of this article to studies of several other materials of special importance, such as antibodies, hormones, etc.

"Spinning Top" Centrifuges

In 1925, only two years after the first comparatively low-speed ultracentrifuge of Svedberg and Nichols was built, Henriot and Huguenard [45, 46] described a new method for obtaining very high rotational speeds without the use of mechanical bearings. Employing a whirling layer of air issuing under pressure from properly directed jets for both supporting and driving a small cone-shaped rotor only 1.17 cm in diameter, they were able to attain a speed of 660,000 rpm. A few years later the method was confirmed by Beams.[8] He and his collaborators [20, 17, 16] introduced many improvements and with a 0.9 cm rotor driven with hydrogen gas were able to reach 1,200,000 rpm.[17] They attempted, with larger rotors, to adapt the air-driven "spinning top" as ultracentrifuges of both the optical and opaque type. An illustration of a "spinning top" taken from one of the original papers [16] is shown in Fig. 10. In 1934, Beams, Pickels, and Weed [17] reported the successful sedimentation of haemoglobin and ascribed earlier failures in this direction to the disturbing action of convection currents caused by temperature gradients through the rotor. They had found that for any degree of sedimentation with molecular particles to be accomplished, the thickness of the solution layer had to be kept very small; *i.e.*, of the order of 0.1 cm or less. From these and other studies by Bauer and Pickels,[7] it was concluded that the "spinning top" centrifuge was not adequately suited to serve as an ultracentrifuge which would give results comparable in precision with those obtained in the Svedberg machine. Even with arrangements which would counteract the temperature gradients to some extent,[17] it was felt that experimental results would be too unreliable, especially with dilute preparations and thick layers of fluid. These are just the conditions that must be realized for the accurate determination of sedimentation constants with an optical ultracentrifuge. In a recent article,[93] the author has presented experimental evidence regarding the following items: the presence of convection currents in rotors exposed to air resistance but spinning at constant tem-

perature (average); the dependence of such disturbances on the concentration; and the fact that a sharp sedimentation boundary may migrate at an abnormal rate and is itself no proof of convection-free sedimentation.

Also, it was apparent that rotors simulating those of Svedberg in size and providing a theoretically comparable resolving power would be entirely impractical in the form of a "spinning top" because of the enormous driving energy required. These

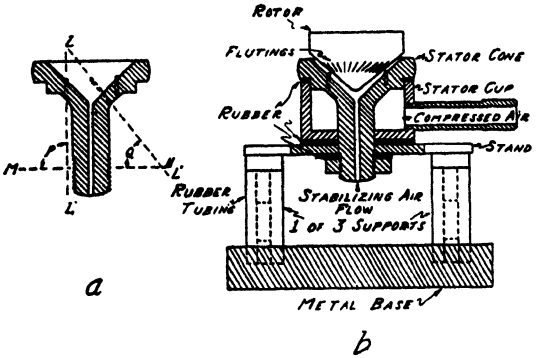


FIGURE 10. Schematic drawing of a typical "spinning top," which is supported and driven by jets of compressed air (see Ref. 16). Rotors are usually a few centimeters in diameter and available speeds are of the order of 100,000 rpm.

considerations led to the development of the vacuum-type ultracentrifuge which has been described earlier in the article.

McBain and O'Sullivan [67, 68] and a few other investigators [101] have continued to use the "spinning top" as an ultracentrifuge and have reported several successful sedimentations, using various improved forms of the apparatus. However, the reliability of some of their results has been questioned by Williams and Watson.[149]

In certain instances where precise quantitative measurement has not been the principal consideration, the "spinning top" has been employed to great advantage. For example, McIntosh [75] has successfully concentrated bacteriophage and fowl sarcoma virus, and Gratia and Goreczky [89] have demonstrated the sedimentation of immune hemolysin. Rotors with diameters as large as 20 cm have been used for preparative purposes by Bauer and the author.[5] Some very interesting observations have been made by Beams and King [21] and by Guyer and Claus [40] regarding the effects of high centrifugal force on biological cells. Harvey [41, 42] employed the "spinning top" for similar purposes even earlier than these investigators, and, in addition, he perfected an optical arrangement whereby an individual cell could be microscopically observed while it was being centrifuged. An alternative arrangement has been described by the author.[86]

High-Speed Angle Centrifuges for Purification and Concentration of Molecular Particles

In the ordinary preparative or low-speed laboratory centrifuge, freely swinging buckets carry tubes or cylindrical containers filled with fluid. For centrifugation at high speeds, it is desirable and usually necessary to provide fixed holes or receptacles for the tubes. This arrangement makes it possible to design rotors with the mechanical strength required for high speeds and to furnish the rotors with a symmetrical surface, with respect to the axis of rotation, which will minimize the resistance and consequent heating caused by the friction of the surrounding air. The tube holes are usually arranged at some angle between 20° to 45° to the axis of rotation for several reasons: a partially filled tube will not overflow either while the centrifuge is spinning or while it is at rest; the greatest number of tubes can be accommodated and a rotor

of given capacity can be made mechanically stronger when comparatively small angles are employed; [91] and such angles are known to be favorable to the rapid deposition of sedimenting material.[150, 151, 93]

The scheme of inclined tubes is used in certain motor-driven commercially available centrifuges, principally the comparatively slow-speed Swedish Angle centrifuge and the International centrifuge with high-speed attachment. The latter involves a step-up pulley-belt system which can be operated at about 18,000 rpm, but unfortunately the size and capacity of the rotor furnished with this equipment are comparatively small. Claude [33] has employed the apparatus for the sedimentation of chick tumor extracts and haemocyanin. The author has published descriptions [90] of a simply constructed angle centrifuge with direct electric motor drive. Interchangeable rotors fit directly on the shaft of a flexibly mounted universal motor which is commercially available. Fig. 11 shows a photograph of the apparatus. The rotor illustrated is almost 8 inches in diameter, accommodates 166 cc of fluid, and spins at 18,000 rpm with an applied potential of 180 volts or at 12,500 rpm with a potential of 115 volts. The original article [90] gives specifications and operational data for rotors having diameters ranging from 7 to 10¼ inches and capacities ranging from 112 to 610 cc of fluid. The temperature of a rotor reaches a steady value a few degrees above the temperature of the surrounding air. The apparatus can be operated inside a refrigerator when it is desirable to keep the rotor cold. A study of the sedimentation process in these centrifuges has been made by the author.[93] In common with all centrifuges which spin in the open air, they are subject to convective disturbances

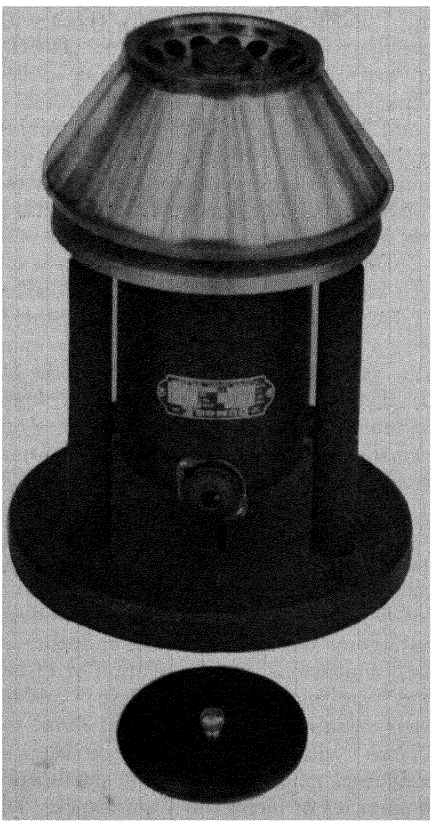


FIGURE 11. High-speed preparative angle centrifuge of the type designed by the author (see Ref. 90). Rotor fits directly on shaft of universal electric motor, which is flexibly mounted. For rotor illustrated: diameter, 20 cm; number of tubes, 18; capacity, 166 cc; speed, 18,000 rpm; maximum centrifugal force on fluid, 34,800 times gravity. Other rotors accommodate up to 610 cc of liquid.

but have been found convenient for rapid clarification and for the concentration and purification of very large molecules. Sedimentation boundaries [88, 98] of high molecular homogeneous materials can be detected in tubes which have been centrifuged in open-air machines when the concentration is sufficiently high. Work of the author [98] has shown that even with low concentrations, boundaries can be obtained by originally providing the solution column with a synthetic density gradient formed with sucrose or some other nonsedimenting solute. Such a gradient counteracts the effects of the temperature gradient which tends to produce convection.

The first high-speed angle centrifuge to be spun in a vacuum, and thus rendered convection free, was described in 1936 by Bauer and Pickels.[6] An improved form of this apparatus is shown in Fig. 5, with the cylinder of the vacuum chamber removed for the purposes of illustration. The arrangement of the inclined tube holes is similar to that shown in Fig. 11. Since glass tubes do not withstand high centrifugal forces without cracking, celluloid tubes of the variety introduced by Bauer and Pickels [5] are used. For work which must be performed under sterile conditions, tubes with screw caps are convenient.[91] They can be sterilized with ultraviolet light. The air drive illustrated in Fig. 5 is of the improved "turret" type which has been described by the author [88] and which is still used routinely. Except for the lid, the rotor is machined from a solid piece of Duralumin and is 8 inches in diameter. A rubber gasket forms a vacuum-tight joint between the rotor proper and the lid, which is held in place by six screws. Stress analyses for several different rotor designs by the author [91] have indicated that a correctly proportioned rotor accommodating several hundred cubic centimeters of fluid can be run at speeds as high as 50,000 rpm with attendant centrifugal forces of the order of 200,000 times gravity.

Alternative air drives for the vacuum type angle centrifuge have been described by other research workers, notably Beams, Linke, and Sommer.[15] Rawson, Scherp, and Lindquist [97] have employed successfully as a substitute for the air drive a commercially available, electrically operated "lathe grinder" drive. They report the attainment of speeds as high as 44,000 rpm. Masket [65] has described a simple arrangement for plugging each hole and its celluloid tube individually, so that the tubes can be more completely filled and inclined at smaller angles to the axis of rotation. Wyckoff and Lagsdin [150, 151] have investigated experimentally the efficiency of tubes inclined at various angles and have shown that small angles are best for concentrating molecular particles. The author [98] has investigated the sedimentation process in the vacuum type angle centrifuge with special reference to the following: the detection and measurement of partially sedimented boundaries; methods of improving the definition of such boundaries by the use of synthetic density gradients; the effects of convection caused by deceleration forces acting on the fluid; the influence of concentration; the effect of changing the tubes' dimensions; the correlation of results with those obtained in the electrical centrifuge [90] and in the ultracentrifuge; [7] and the reasons why small numbers of particles always remain in the supernatant fluid.

Since the concentration of yellow fever virus [6] with the first angle centrifuge of the vacuum type, similar machines have been used extensively for the purification and concentration of viruses and other small particles of biological significance. For example, Stanley [116] has concentrated the tomato bushy stunt virus, and Beard, Bryan, and Wyckoff [22] have isolated the rabbit papilloma virus. Purification through differential centrifugation involves no chemical treatment of the material under study and often offers a decided advantage in this respect.

High-Speed Centrifuges of Other Types

For research purposes, angle centrifuges are greatly favored for preparative work because of the convenient manner in which the material under study can be handled and the sediment collected. However, Schlesinger,[103] for example, has used hollow closed cylinders which are spun about their axes of radial symmetry. More recently,

McIntosh and Selbie [76] have described an adaptation of the Sharples air-driven tubular centrifuge. The cylinder of the centrifuge is about 5 cm in diameter and is usually operated at 40,000 rpm. The centrifuge is of the continuous-flow type and 15 to 20 cc are passed through the rotating cylinder per minute. Successful concentrations of several viruses, bacteria, and bacteriophage have been reported. The apparent absence of marked convection is probably due to the fact that the layer of fluid within the centrifuge is only about 0.55 mm thick. Beams, Linke, and Skarstrom [14] have experimented with tubular, continuous-feed centrifuges spinning inside a vacuum chamber. They reported a partial separation of hemoglobin molecules from solution. Other papers [19, 18] have described the use of similar apparatus for separating gases and gaseous isotopes. The scarcity of reported applications of the tubular vacuum centrifuge in biological research is evidence of the many problems which are peculiar to this type of machine. Its great potential value will probably be realized as further experience is accumulated.

For the estimation of sedimentation rates, several investigators have made use of the "inverted capillary" technique which apparently was originated by Bechhold and Schlesinger.[24] For example, Elford and Galloway [84] have determined the sedimentation rates of the viruses of vesicular stomatitis and foot and mouth disease. Scott and Elford [105] have similarly studied the virus of lymphocytic choriomeningitis. Their electrically driven centrifuge is an "Ecco Blitz" model of Collatz and runs in the open air at about 15,000 rpm. The material to be studied is placed in small capillaries and the sedimenting particles migrate into a larger reservoir filled with the same material. After centrifugation, the degree of sedimentation is estimated by comparing the average concentration of the liquid in the capillaries with that in the reservoir. The success of the apparatus undoubtedly depends in a large degree upon the inhibition of convection currents due to the small size of the capillaries.

A simple optical ultracentrifuge with a small rotor and cell made of transparent plastic has been described by Stern.[118] It rotates about a horizontal axis in mechanical bearings and is driven by air jets impinging at the periphery. The temperature distribution is perhaps somewhat more favorable than with the "spinning top," but the limitations of the two are, in general, the same. Also, there are objections to the use of plastic materials from the standpoints of both optics and mechanical strength.[91]

Sampling of Centrifuged Material and Correlation of Specific Properties with Sedimentation Rate

Some idea of the degree of sedimentation experienced by a particular substance under given conditions can be had by comparing the concentration of the sediment with that of the supernatant fluid, using some specific property as a basis of measurement. To relate more accurately the intensity distribution of some specific property with the sedimentation of an optically registered boundary, Tiselius, Pedersen, and Svedberg [148] have provided the optical ultracentrifuge cell with a porous partition near its center. After a boundary is observed to have passed through the filter, the ultracentrifuge is stopped and the concentrations of the material in the two halves of the cell are compared by chemical or biological tests. Obviously, several experiments with different centrifugation times may be required for a conclusive correlation. It has been found that with the vacuum-type ultracentrifuge, the rotor may be brought to rest and the cell removed without greatly disturbing the appearance of a partially sedimented boundary.[94] The cell can then be sampled at several successive levels by a simple displacement technique. Also, it can be placed in front of a spectrograph slit and the differential absorption across the boundary measured for different wave lengths. Correlation is more definitely established in a single run when concentrations at several levels can be compared by the different methods.

As shown by studies which the author has made,[98] the estimation of sedimentation rate is always hazardous unless a definite sedimenting boundary can be demonstrated.

Fortunately, boundaries can be obtained under certain circumstances with the convenient angle centrifuges.[98] Transparent tubes can be removed and photographed to demonstrate the boundary optically. By a technique which Hughes, Pickels, and Horsfall[49] have described, the fluid columns can be sampled with accuracy at any number of levels. With dilute materials which do not lend themselves to optical measurement, this technique had proven to be far superior to others tried in estimating homogeneity and sedimentation rate. Simple equations have been derived by the author[98] which can be used as approximations in determining the size of particles from the degree of sedimentation experienced in an angle centrifuge. The equations are based on the observation that a boundary in an angle centrifuge migrates at roughly the same rate, measured radially, as it does in an ultracentrifuge.

Subject Index to References

(References under each subject are listed in chronological order.)

A. Reviews, bibliographies, and general information: 122, 124, 16, 9, 125, 10, 126, 133, 80, 89
B. Oil-driven ultracentrifuge for measurement of sedimentation rate: 131, 29, 133
C. Electrically-driven ultracentrifuge for determination of molecular weights by equilibrium method: 132, 135, 124, 133
D. Vacuum-type ultracentrifuge: 16, 26, 27, 7, 11, 133, 92
E. "Spinning top" and related types of air-driven centrifuges: 45, 46, 8, 20, 17, 16, 68, 75, 5, 149, 39, 69, 76, 101, 118
F. Air and steam drives for vacuum-type centrifuges: 16, 95, 6, 7, 13, 48, 15, 31, 88
G. Electrical drives for high speed centrifuges: 132, 124, 12, 105, 97, 111, 133, 90
H. High-speed angle centrifuges for purification and concentration: 6, 33, 15, 65, 90, 91
I. Tubular, capillary-type, and continuous feed centrifuges: 24, 103, 14, 34, 81, 105, 76
J. Measurement and regulation of high rotational speeds: 7, 107, 87, 28, 99
K. Theory of sedimentation in the ultracentrifuge and the determination of particle size: 120, 66, 121, 123, 140, 147, 35, 55, 3, 126, 133, 89
L. Theory, experimental tests, and examples of sedimentation in the angle centrifuge: 6, 33, 150, 151, 23, 49, 106, 22, 51, 116, 93
M. Density, shape, hydration, electrical charge, and diffusion of particles, and their influence on sedimentation: 141, 47, 1, 84, 112, 53, 133, 80, 78, 89, 100
N. Experimental considerations, *e.g.*, convection currents, temperature, Donnan effect, cell alignment, concentration effect, stability of material, hydrostatic compression, etc. 4, 7, 9, 64, 83, 113, 133, 145, 92, 100, 93
O. Optical methods for the measurement of sedimentation: 144, 148, 139, 134, 56, 7, 57, 142, 85, 136, 32, 133, 137, 89
P. Estimation of sedimentation rates by other methods and the correlation of rate with specific properties: 143, 49, 94, 93
Q. Centrifugal separation of gases and isotopes: 19, 18
R. Ultracentrifugation of proteins: 130, 16, 30, 69, 70, 71, 72, 73, 62, 128, 37, 83, 102, 126, 127, 98, 133
S. Ultracentrifugation of viruses: 74, 96, 112, 22, 117, 119, 138, 58, 60, 61, 94, 36, 52, 79, 59
T. Ultracentrifugation of:
Amine (77); antibodies (44, 50, 146); carbohydrates (129, 25, 38); cellulose (54); chlorophyll compounds (115); collodion (113); detergents (114); hormones (109); linear high polymers (Kraemer and Nichols, 133); normal tissue extract (108); nucleic acid (104); polydisperse systems (Kraemer, 133); polystyrenes (110); salts (82); toxin (63); visual purple (43).
U. Centrifugation of biological tissue and cells and microscopic observation of small objects during centrifugation: 41, 42, 21, 86, 40

References

1. Adair, G. S., and Adair, M. E., *Proc. Roy. Soc. London, (Series B)*, **120**, 422 (1936).
2. Andersson, K. J. I., *Nature*, **143**, 720 (1939).
3. Archibald, W. J., *Phys. Rev.*, **54**, 371 (1938).
4. Bailey, E. D., Nichols, J. B., and Kraemer, E. O., *J. Phys. Chem.*, **40**, 1149 (1936).
5. Bauer, J. H., and Pickels, E. G., *J. Bact.*, **31**, 53 (1936).
6. Bauer, J. H., and Pickels, E. G., *J. Exp. Med.*, **64**, 503 (1936).
7. Bauer, J. H., and Pickels, E. G., *J. Exp. Med.*, **65**, 565 (1937).
8. Beams, J. W., *Rev. Sci. Instr.*, **1**, 667 (1930).
9. Beams, J. W., *J. Appl. Phys.*, **8**, 795 (1937).
10. Beams, J. W., *Rev. Modern Phys.*, **10**, 245 (1938).
11. Beams, J. W., *Science*, **89**, 543 (1939).
12. Beams, J. W., and Black, S. A., *Rev. Sci. Instr.*, **10**, 59 (1939).

13. Beams, J. W., and Linke, F. W., *Rev. Sci. Instr.*, **8**, 160 (1937).
14. Beams, J. W., Linke, F. W., and Skarstrom, C., *Science*, **86**, 293 (1937).
15. Beams, J. W., Linke, F. W., and Sommer, P., *Rev. Sci. Instr.*, **9**, 248 (1938).
16. Beams, J. W., and Pickels, E. G., *Rev. Sci. Instr.*, **6**, 299 (1935).
17. Beams, J. W., Pickels, E. G., and Weed, A. J., *J. Chem. Phys.*, **2**, 143 (1934).
18. Beams, J. W., and Skarstrom, C., *Phys. Rev.*, **56**, 266 (1939).
19. Beams, J. W., and Snoddy, L. B., *J. Chem. Phys.*, **5**, 993 (1937).
20. Beams, J. W., Weed, A. J., and Pickels, E. G., *Science*, **78**, 338 (1933).
21. Beams, H. W., and King, R. L., *Nature*, **135**, 232 (1935).
22. Beard, J. W., Bryan, W. R., and Wyckoff, R. W. G., *J. Infect. Dis.*, **65**, 43 (1939).
23. Beard, J. W., Finkelstein, H., Sealy, W. C., and Wyckoff, R. W. G., *Science*, **87**, 89 (1938).
24. Bechhold, H., and Schlesinger, M., *Biochem. Z.*, Berlin, **236**, 387 (1931).
25. Beckmann, C. O., and Landis, Q., *J. Am. Chem. Soc.*, **61**, 1495 (1939).
26. Biscoe, J., Pickels, E. G., and Wyckoff, R. W. G., *J. Exp. Med.*, **64**, 39 (1936).
27. Biscoe, J., Pickels, E. G., and Wyckoff, R. W. G., *Rev. Sci. Instr.*, **7**, 246 (1936).
28. Björnstahl, Y., *Rev. Sci. Instr.*, **10**, 258 (1939).
29. Boestad, G., Pedersen, K. O., and Svedberg, T., *Rev. Sci. Instr.*, **9**, 346 (1938).
30. Brosteaux, J., and Eriksson-Quensel, I.-B., *Arch. de Phys. Biol.*, XII, No. 4 (1935).
31. Chiles, J. A., Jr., and Severinghaus, A. E., *J. Exp. Med.*, **68**, 1 (1938).
32. Chiles, J. A., Jr., and Severinghaus, A. E., *Rev. Sci. Instr.*, **11**, 71 (1940).
33. Claude, A., *J. Exp. Med.*, **66**, 59 (1937).
34. Elford, W. J., and Galloway, I. A., *Brit. J. Exp. Path.*, **18**, 155 (1937).
35. Faxen, O. H., *Arkiv. Mat., Astro. Fysik*, **21B**, No. 2 (1929).
36. Gard, S., and Pedersen, K. O., *Science*, **94**, 493 (1941).
37. Gralén, N., and Svedberg, T., *Biochem. J.* (London), **32**, 1375 (1938).
38. Gralén, N., and Svedberg, T., *Biochem. J.* (London), **34**, 234 (1940).
39. Gratia, A., and Goreczky, L., *Zeitschr. Immunitätsforsch*, **93**, 18 (1938).
40. Guyer, M. F., and Claus, P. E., *Proc. Soc. Exp. Biol. and Med.*, **35**, 468 (1937).
41. Harvey, E. N., *J. Franklin Inst.*, **214**, 1 (1932).
42. Harvey, E. N., *Biol. Bull.*, **66**, 48 (1934).
43. Hecht, S., and Pickels, E. G., *Proc. Nat. Acad. Sci.*, **24**, 172 (1938).
44. Heidelberger, M., Pedersen, K. O., and Tiselius, A., *Nature*, **138**, 165 (1936).
45. Henriot, E., and Huguenard, E., *Compt. rend. Acad.*, **180**, 1389 (1925).
46. Henriot, E., and Huguenard, E., *J. phys. et radiol.*, **8**, 433 (1927).
47. Herzog, R. O., Illig, R., and Kudar, H., *Z. physik. Chem.* A, **167**, 329 (1933).
48. Hoxton, L. G., and Beams, J. W., *Phys. Rev.*, **51**, 690 (1937).
49. Hughes, T. P., Pickels, E. G., and Horsfall, F. L., *J. Exp. Med.*, **67**, 941 (1938).
50. Kabat, E. A., *J. Exp. Med.*, **69**, 103 (1939).
51. Kirschbaum, A., Stern, K. G., and Hooker, C. W., *Yale J. Biol. and Med.*, **13**, 1 (1940).
52. Knight, C. A., and Stanley, W. M., *J. Biol. Chem.*, **141**, 29 (1941).
53. Kraemer, E. O., *J. Franklin Inst.*, **229**, 393 (1940).
54. Kraemer, E. O., and Lansing, W. D., *J. Phys. Chem.*, **39**, 153 (1935).
55. Lamm, O., *Ark. Mat. Astrom. och Fysik.*, **21B**, No. 2 (1929).
56. Lamm, O., *Nature*, **132**, 820 (1933).
57. Lamm, O., *Nova Acta Regiae Soc. Sci. Upsaliensis* (4), **10**, No. 6 (1937).
58. Lauffer, M. A., *J. Phys. Chem.*, **44**, 1137 (1940).
59. Lauffer, M. A., *J. Biol. Chem.*, **143**, 99 (1942).
60. Lauffer, M. A., and Ross, A. F., *J. Am. Chem. Soc.*, **62**, 3296 (1940).
61. Lauffer, M. A., and Stanley, W. M., *J. Biol. Chem.*, **135**, 463 (1940).
62. Lungren, H. P., *Nature*, **138**, 122 (1936).
63. Lundgren, H. P., Pappenheimer, A. M., Jr., and Williams, J. W., *J. Am. Chem. Soc.*, **61**, 533 (1939).
64. Lundgren, H. P., and Williams, J. W., *J. Phys. Chem.*, **43**, 989 (1939).
65. Masket, A. V., *Rev. Sci. Instr.*, **12**, 277 (1941).
66. Mason, M., and Weaver, W., *Phys. Rev.*, **23**, 412 (1924).
67. McBain, J. W., *Chem. Rev.*, **24**, 289 (1939).
68. McBain, J. W., and O'Sullivan, C. M., *J. Am. Chem. Soc.*, **57**, 2631 (1935).
69. McFarlane, A. S., *Biochem. J.*, **29**, 407 (1935).
70. McFarlane, A. S., *Biochem. J.*, **29**, 660 (1935).
71. McFarlane, A. S., *Biochem. J.*, **29**, 1175 (1935).
72. McFarlane, A. S., *Biochem. J.*, **29**, 1202 (1935).
73. McFarlane, A. S., *Biochem. J.*, **29**, 1209 (1935).
74. McFarlane, A. S., and Kekwick, R. A., *Biochem. J.*, **32**, 1607 (1938).
75. McIntosh J., *J. Path. and Bact.*, **41**, 215 (1935).
76. McIntosh, J., and Selbie, F. R., *Brit. Jour. Exp. Path.*, **21**, 153 (1940).
77. Miles, A. A., and Pirie, N. W., *Biochem. J.*, **33**, 1716 (1939).
78. Neurath, Hans, *Chem. Rev.*, **30**, No. 3, 357 (1942).
79. Neurath, H., Cooper, G. R., Sharp, D. G., Taylor, A. R., Beard, D., and Beard, J. W., *J. Biol. Chem.*, **140**, 293 (1941).
80. Oncley, J. L., *Ann. N. Y. Acad. Sci.*, **41**, 121 (1941).
81. Pecker, J. S., *Trans. A. S. M. E.*, December, 1938.
82. Pedersen, K. O., *Z. physik. Chem.*, **A170**, 41 (1934).
83. Pedersen, K. O., *Proc. Roy. Soc.*, B, **127**, 20 (1939).
84. Perrin, F., *J. Phys. Rad.* VII, **7**, 1 (1936).
85. Philpot, J. St. L., *Nature*, **141**, 283 (1938).
86. Pickels, E. G., *Science*, **83**, 471 (1936).
87. Pickels, E. G., *Rev. Sci. Instr.*, **9**, 354 (1938).
88. Pickels, E. G., *Rev. Sci. Instr.*, **9**, 358 (1938).
89. Pickels, E. G., *Chem. Rev.*, **30**, 341 (1942).
90. Pickels, E. G., *Rev. Sci. Instr.*, **13**, 93 (1942).
91. Pickels, E. G., *Rev. Sci. Instr.*, **13**, 101 (1942).
92. Pickels, E. G., *Rev. Sci. Instr.*, **13**, 426 (1942).
93. Pickels, E. G., *J. Gen. Physiol.*, **26**, 341 (1943).
94. Pickels, E. G., and Bauer, J. H., *J. Exp. Med.*, **71**, 703 (1940).
95. Pickels, E. G., and Beams, J. W., *Science*, **81**, 342 (1935).
96. Pickels, E. G., and Smadel, J. E., *J. Exp. Med.*, **68**, 583 (1938).
97. Rawson, A. J., Scherp, H. W., and Lindquist, F. E., *J. Bact.*, **40**, 657 (1940).
98. Rothen, A., *J. Gen. Physiol.*, **24**, 203 (1940).
99. Rothen, A., *Rev. Sci. Instr.*, **12**, 355 (1941).
100. Rothen, A., *Ann. N. Y. Acad. Sci.*, **43**, 229 (1942).

101. Ruben, S., Kamen, M. D., and Perry, L. H., *J. Am. Chem. Soc.*, **62**, 3450 (1940).
102. Sanigar, E. B., Krejci, L. E., and Kraemer, E. O., *Biochem. J.*, **33**, 1 (1939).
103. Schlesinger, M., *Nature*, **138**, 549 (1936).
104. Schmidt, G., Pickels, E. G., and Levene, P. A., *J. Biol. Chem.*, **127**, 251 (1939).
105. Scott, T. F. M., and Elford, W. J., *Brit. J. Exp. Path.*, **20**, 182 (1939).
106. Severinghaus, A. E., Levin, L., and Chiles, J. A., Jr., *Endocrinology*, **23**, 285 (1938).
107. Shapiro, H., and Butt, C., *Rev. Sci. Instr.*, **8**, 35 (1937).
108. Sharp, D. G., Taylor, A. R., Finkelstein, H., and Beard, J. W., *Proc. Soc. Exp. Biol. and Med.*, **42**, 459 (1939).
109. Shedlovsky, T., Rothen, A., Greep, R. O., van Dyke, H. B., and Chow, B. F., *Science*, **92**, 178 (1940).
110. Signer, R., *Trans. Faraday Soc.*, **32**, 296 (1936).
111. Skarstrom, C., and Beams, J. W., *Rev. Sci. Instr.*, **11**, 398 (1940).
112. Smadel, J. E., Pickels, E. G., and Shedlovsky, T., *J. Exp. Med.*, **68**, 607 (1938).
113. Smadel, J. E., Pickels, E. G., Shedlovsky, T., and Rivers, T. M., *J. Exp. Med.*, **72**, 523 (1940).
114. Smith, E. L., and Pickels, E. G., *Proc. Nat. Acad. Sci.*, **26**, 272 (1940).
115. Smith, E. L., and Pickels, E. G., *J. Gen. Phys.*, **24**, 753 (1941).
116. Stanley, W. M., *J. Biol. Chem.*, **135**, 437 (1940).
117. Stanley, W. M., and Lauffer, M. A., *Science*, **89**, 345 (1939).
118. Stern, K. G., *Science*, **95**, 561 (1942).
119. Stern, K. G., and Duran-Reynals, F., *Science*, **89**, 609 (1939).
120. Stokes, G. G., Mathematical and physical papers, University Press, Cambridge, Vol. III, pgs. 59-60 (1901).
121. Svedberg, T., *Kolloid-Z., Zsigmondy-Festschrift, Erg. Bd. zu*, **36**, 53 (1925).
122. Svedberg, T., Colloid chemistry, theoretical and applied, Vol. I, Chemical Catalogue Co., (ed. J. Alexander) New York (1926).
123. Svedberg, T., *Z. physik. Chem.*, **121**, 65 (1926).
124. Svedberg, T., *Naturwissenschaften*, **22**, 225 (1934).
125. Svedberg, T., *Chem. Rev.*, **20**, 81 (1937).
126. Svedberg, T., *Proc. Roy. Soc.* B, **127**, 1 (1939).
127. Svedberg, T., and Brohult, S., *Nature*, **143**, 938 (1939).
128. Svedberg, T., and Eriksson-Quensel, I.-B., *Nature*, **137**, 400 (1936).
129. Svedberg, T., and Gralén, N., *Nature*, **142**, 261 (1938).
130. Svedberg, T., and Hedenius, A., *Biol. Bull.*, **66**, 191 (1934).
131. Svedberg, T., and Lysholm, A., *Nova Acta Regiae Soc. Sci. Upsaliensis*, vol. extra ord. (1927).
132. Svedberg, T., and Nichols, J. B., *J. Am. Chem. Soc.*, **45**, 2910 (1923).
133. Svedberg, T., and Pedersen, K. O., The ultracentrifuge, Clarendon Press, Oxford (1940).
134. Svedberg, T., and Rinde, H., *J. Am. Chem. Soc.*, **46**, 2677 (1924).
135. Svedberg, T., and Sjögren, B., *J. Amer. Chem. Soc.*, **51**, 3594 (1929).
136. Svensson, H., *Kolloid-Z.*, **87**, 181 (1939).
137. Svensson, H., *Kolloid-Z.*, **90**, 141 (1940).
138. Taylor, A. R., Sharp, D. G., Finkelstein, H., and Beard, J. W., *Proc. Soc. Exp. Biol. and Med.*, **42**, 462 (1939).
139. Thovert, J., *Ann. phys.* (9), **2**, 369 (1914).
140. Tiselius, A., *Z. physik. Chem.*, **124**, 499 (1926).
141. Tiselius, A., *Kolloid-Z.*, **59**, 306 (1932).
142. Tiselius, A., Pedersen, K. O., and Eriksson-Quensel, I.-B., *Nature*, **139**, 546 (1937).
143. Tiselius, A., Pedersen, K., and Svedberg, T., *Nature*, **140**, 848 (1937).
144. Toepler, A., *Ann. Physik Chem.*, **131**, ([*V*]11) 33, 180 (1867).
145. Treffers, H. P., and Moore, D. H., *Science*, **93**, 240 (1941).
146. Van der Scheer, J., Lagsdin, J. B., and Wyckoff, R. W. G., *J. Immunol.*, **41**, 209 (1941).
147. Weaver, W., *Phys. Rev.*, **27**, 499 (1926).
148. Weiner, O., *Ann. der Physik und Chemie*, N. F., **49**, 105 (1893).
149. Williams, J. W., and Watson, C. C., *J. Am. Chem. Soc.*, **58**, 2650 (1936).
150. Wyckoff, R. W. G., and Lagsdin, J. B., *Rev. Sci. Instr.*, **8**, 74 (1937).
151. Wyckoff, R. W. G., and Lagsdin, J. B., *Rev. Sci. Instr.*, **8**, 427 (1937).

Measurement of the Surface Areas of Finely Divided and Porous Materials by Low-temperature Adsorption Isotherms

P. H. EMMETT

Johns Hopkins University, Baltimore, Md.

The adsorption of gases on solids has been known and studied for many years. In a few instances [1] attempts have been made to utilize chemisorption or physical adsorption measurements in estimating the surface areas of finely divided or porous materials. Such attempts have met with two outstanding difficulties. In the first place, it is often difficult to find suitable adsorbates for chemisorption measurements on the particular solid being studied. Secondly, satisfactory criteria are usually lacking for judging the point on a chemical or physical adsorption isotherm corre-

sponding to the completion of a monolayer. Within the past six years a great deal of work has been done in an endeavor to find a way of utilizing adsorption isotherms for measuring accurately the absolute as well as the relative surface areas of finely divided porous or non-porous solids. A method has been evolved that appears to have much promise and that up to the present has been used in measuring the absolute surface area of metallic Fe,[2] Ni,[3] and Cu [3] catalysts; silica gel,[3] chromium oxide gels,[3] soils,[4] powdered bacteria,[5] carbon blacks,[6] paint pigments,[6] and a variety of other substances. In the present paper a brief description of the principle and the experimental nature of the method will be presented, together with a few results to illustrate its applicability.

The principle of the method is comparatively simple. If one can select from an experimental adsorption isotherm the volume and hence the number of molecules of some inert gas, such as nitrogen, corresponding to a monolayer, a simple multiplication by the average cross-sectional area of each molecule will yield the absolute surface area of the adsorbent. The accuracy of the method must obviously rest upon (a) the exactness with which the point corresponding to a monolayer of adsorbed molecules can be picked, and (b) the closeness to which one can approximate the average area covered by each adsorbed molecule.

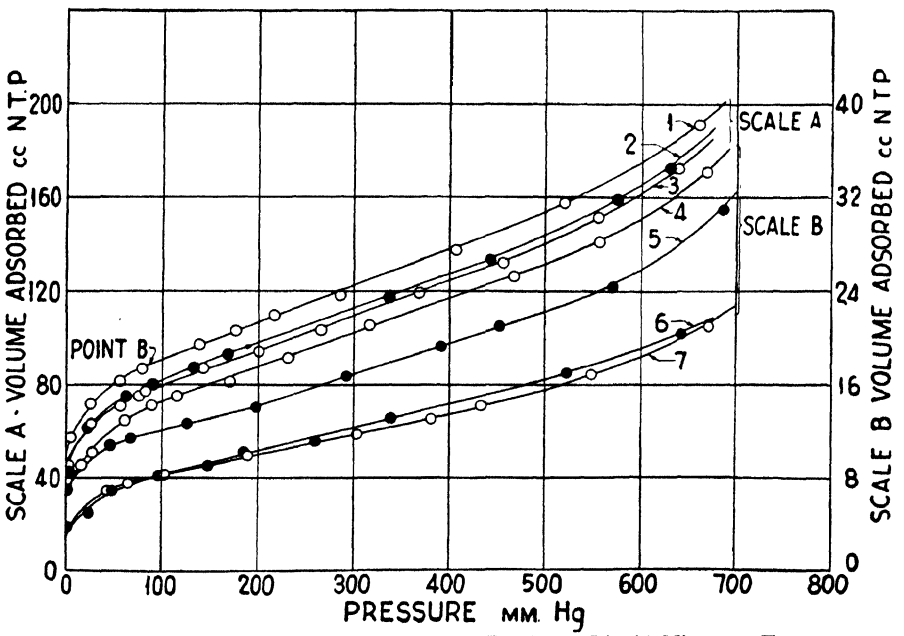


FIGURE 1. Adsorption of Nitrogen on Carbon Blacks at Liquid Nitrogen Temperatures.

1. 3.29 grams of Arrow black
2. 3.09 grams of Micronex
3. 2.95 grams of Wyex
4. 2.86 grams of P-33
5. 0.792 gram of acetylene black
6. 5.50 grams of thermatomic
7. 4.44 grams of Thermax

Early in the experimental work [2] the point corresponding to a monolayer of adsorbed gas (hereafter designated as "point B") was selected empirically, on the basis of accumulated evidence, as the lower pressure extremity of the long linear portion of the adsorption isotherm (see Figure 1, for example). The experimental evidence that led to the selection of point B on the iron catalysts was threefold. In the first place, the surface area for a given catalyst, as calculated from a series of isotherms

for argon, oxygen, nitrogen, carbon monoxide, carbon dioxide, and butane, showed better agreement if the volume of gas at point B was considered to be a monolayer than if any of the other likely points on the isotherms was chosen. The maximum deviation of the surface area values from the mean calculated from these adsorption isotherms was only about 10 per cent on the iron catalysts. Secondly, the heats of adsorption calculated by the Clapeyron equations from a series of adsorptions at different temperatures showed that the heat of adsorption at point B was intermediate between probable values for first-layer adsorption and for second-layer adsorption. Thus, for the adsorption of a volume of gas 25 per cent greater than at B, a molal heat of adsorption close to that of liquefaction of the gas was obtained, whereas for a volume 25 per cent smaller, the heat of adsorption was at least 50 per cent greater than the heat of liquefaction and was clearly characteristic of adsorption in the first layer. Finally, the chemisorption of a layer of carbon monoxide over the entire surface of the pure iron catalyst required a volume of carbon monoxide approximately equal to the volume of physically adsorbed gas at B on the low-temperature isotherms. These three separate types of evidence combined to convince the author that the volume of gas corresponding to point B on the isotherms was approximately that required for a monolayer.

An excellent confirmation of the selection of point B as the monolayer and a valuable new working tool in using this method of surface area measurements are contained in a recent paper [9] on the theory of multilayer adsorption such as is apparently being obtained in S-shaped isotherms of the type shown in Figure 1. It is shown that the isotherm can be plotted in such a manner as to yield a straight line whose slope and intercept yield directly the volume of gas required to form a monolayer. The complete equation is

$$\frac{p}{V(p_0 - p)} = \frac{1}{V_m C} + \frac{(C-1)p}{V_m C p_0} \qquad (1)$$

where V is the volume of gas adsorbed at pressure p, at a temperature at which the vapor pressure of the liquefied gas is p_0; V_m is the volume of gas in cc required to form a monolayer; and C is a constant related exponentially to the difference between the heat of liquefaction of the adsorbate and its heat of adsorption. It is evident that from the slope and intercept of a plot of $p/V(p_0-p)$ *against* p/p_0 both V_m and C can be evaluated.

Application of equation (1) to several hundred isotherms obtained during the last few years shows excellent agreement between the volume of gas adsorbed at point B and the volume V_m. A few examples of the closeness of the agreement are presented in Table 1.

Table 1. Volume of Adsorbed Nitrogen in a Monolayer on Various Adsorbents *

Absorbent	Volume V_m (ml per g)	Volume at Point B (cm^3 per g)
Unpromoted Fe catalyst 973 I	0.13	0.12
Unpromoted Fe catalyst 973 II	0.29	0.27
Fe-Al_2O_3 catalyst 954	2.86	2.78
Fe-Al_2O_3 catalyst 424	2.23	2.09
Fe-Al_2O_3-K_2O catalyst 931	0.81	0.76
Fe-Al_2O_3 catalyst 958	0.56	0.55
Fe-K_2O catalyst 930	0.14	0.12
Fused Cu catalyst	0.05	0.05
Commercial Cu catalyst	0.09	0.10
Cr_2O_3 gel	53.3	50.5
Cr_2O_3 glowe	6.09	6.14
Silica gel	116.2	127.0

* The values shown in columns 2 and 3 were obtained from nitrogen adsorption isotherms at about 77° K.

The linearity of the isotherms plotted according to equation (1) extends up to a relative pressure of about 0.35. This is sufficient for evaluating V_m. A typical set of such plots is shown in Figure 2 for the carbon black samples for which the adsorption isotherms were given in Figure 1.

One particular advantage of equation (1) for obtaining V_m arises from the fact that only a few experimental points are needed in the range 0.05 to 0.35 relative pressure, provided the solid is known to yield an S-shaped isotherm. In fact, a single adsorption point at a relative pressure of about 0.3 will, when connected with the origin on such a plot, give a line whose slope usually will differ by less than 5 per cent from that drawn with the help of a number of adsorption points. In practice it has become customary to determine only 3 or 4 experimental points in the relative pressure range below 0.35, since the slope of the straight line so obtained is indistinguishable from that resulting from a larger number of points.

The second factor upon which the absolute values obtained for surface areas will depend is the area occupied by each adsorbed molecule. In a previous publication it has been pointed out that two convenient and reasonable molecular area values are those corresponding to the packing of the molecules in the solidified [Area (S)] and the liquefied adsorbates [Area (L)]. The equation by which the area per molecule has been calculated is

$$\text{Area per molecule} = 4\ (.866)\left(\frac{M}{4\sqrt{2}\,AD}\right)^{\frac{2}{3}} \tag{2}$$

where M is the molecular weight of the adsorbate, A is Avogadro's number and D is the density of the liquefied or solidified gas. The equation assumes close packing of the adsorbed molecules. A list of the areas calculated for the various adsorbate molecules has been published.[2] It will suffice here to note that the area per molecule calculated from the density of the liquefied adsorbate is about 20 per cent larger than that from the solidified adsorbate. Hence absolute surface area measurements must be regarded as uncertain by at least this amount, though experience has shown that relative areas of materials are reproducible to a few per cent by the adsorption method. Taking 16.2 square Ångstroms as the average area occupied by an adsorbed nitrogen molecule as calculated from the density of liquefied nitrogen, one obtains a factor of 4.38 by which to multiply the volume of nitrogen (S.T.P.) in a monolayer on a gram of adsorbent to obtain the surface area in square meters per gram. This constant has been used in the surface area calculations in the present chapter.

Experimental

The detailed experimental procedure for obtaining the adsorption isotherms has been adequately described elsewhere [10, 1] and need not be repeated here. A standard adsorption apparatus and technique are used, consisting of the calibration of the dead space around the adsorbent with pure helium and then the measurement of the volume of nitrogen or other adsorbent required to fill the adsorption bulb containing the adsorbent at some given temperature and a series of pressures. At each pressure the helium calibration enables one to calculate the amount of adsorbate in the space around the adsorbent, and hence by difference to obtain the amount of adsorbate on the surface of the adsorbent. The experimental measurements are made at temperatures close to the normal boiling point of the adsorbate. The temperature of the bath thermostating the adsorbent during the run is carefully determined. The usual purification procedures are used on the various adsorbates employed.

Results

Up to the present time hundreds of experimental determinations of the surface areas of various porous and non-porous solids have been made by the low-tempera-

ture adsorption isotherm method with apparent success. For the purpose of illustrating the method, however, the results reported in the present chapter will be restricted to those materials for which some independent estimates of the surface area were available.

Carbon Blacks.[6] A standard set of six carbon blacks was measured using nitrogen as adsorbate and a liquid nitrogen bath at about 78° K as a cold bath. The isotherms are shown in Figure 1 and the linear plots of the data according to equation 1, in Figure 2. The values for V_m and the surface areas are given below.

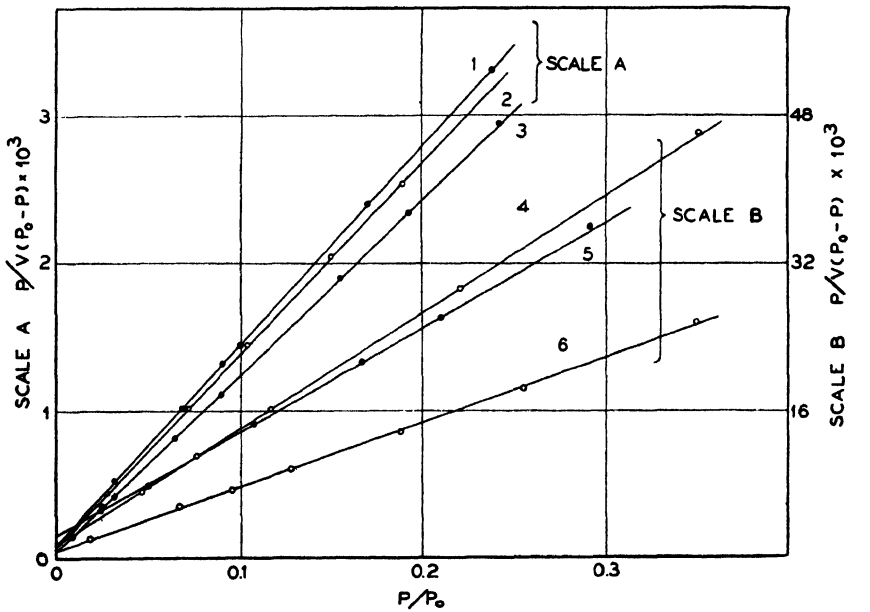


FIGURE 2. Adsorption Data for Nitrogen on Carbon Black According to Equation 1.

1. Wyex
2. Micronex
3. Arrow Black
4. Thermax
5. Thermatomic
6. P-33

Weights of samples same as in Figure 1.

Table 2. Surface Area Measurements on Carbon Black Samples

Material	Weight (grams)	V_m (cc)	Area (sq m/g)	Diameter of average particle By adsorption (microns)	Diameter of average particle By ultramicroscopic count* (microns)
Micronex	3.08	75.2	106.7	0.031	0.061
P-33	2.86	14.42	22.12	0.151	0.159
Arrow black	3.29	84.8	112.7	0.029	....
Wyex	2.95	74.2	110.2	0.030	....
Thermax	4.44	7.8	7.69	0.43	....
Thermatomic carbon†	5.497	8.54	6.81	0.49	1.12
Acetylene black	0.792	11.68	64.5	0.052	0.130

* Gehman and Morris.[7]

† Names here used were those on samples received by authors. However, thermatomic carbon and Thermax are probably the same, "thermatomic carbon" being used ordinarily to designate a class of carbons formed by thermal decomposition of hydrocarbon gases.

In column 5 of Table 2 are listed the diameters of the carbon black particles as calculated from the surface area measurements, on the assumption that the particles

are nonporous spheres and that their true density is 1.8. For comparison, in column 6 are shown the average particle diameters as determined by Gehman and Morris [7] on similar commercial samples in 1932. The agreement is probably satisfactory in view of the fact that the samples for the ultramicroscopic examination were prepared a number of years earlier than those for the nitrogen adsorption work.

Within the last year, measurements by means of an electron microscope on Micronex and acetylene black have been reported that are in almost exact agreement with the average diameter values obtained by the nitrogen adsorption method. For Micronex the electron microscope value is 0.028 micron;[8] that obtained by the nitrogen method varies from 0.028 to 0.031 micron depending on whether one uses 1.8 or 2.0 for the density of carbon. In the discussion attendant upon the ASTM Symposium on particle size measurements (March, 1941), it was disclosed that the electron microscope value for acetylene black is 0.050 micron; this compares with 0.052 micron that one obtains by the adsorption method using 1.8 for the density of carbon, and with 0.047, using 2.0 for the carbon density. This agreement is, in fact, much closer than we could consistently expect, since we have no way of knowing whether the correct value for the area covered by a nitrogen molecule adsorbed on the surface is more nearly 16.2 (Area L) or 13.8 square Ångstroms (Area S). Even an approximate agreement between the results of the two methods, however, is exceedingly gratifying.

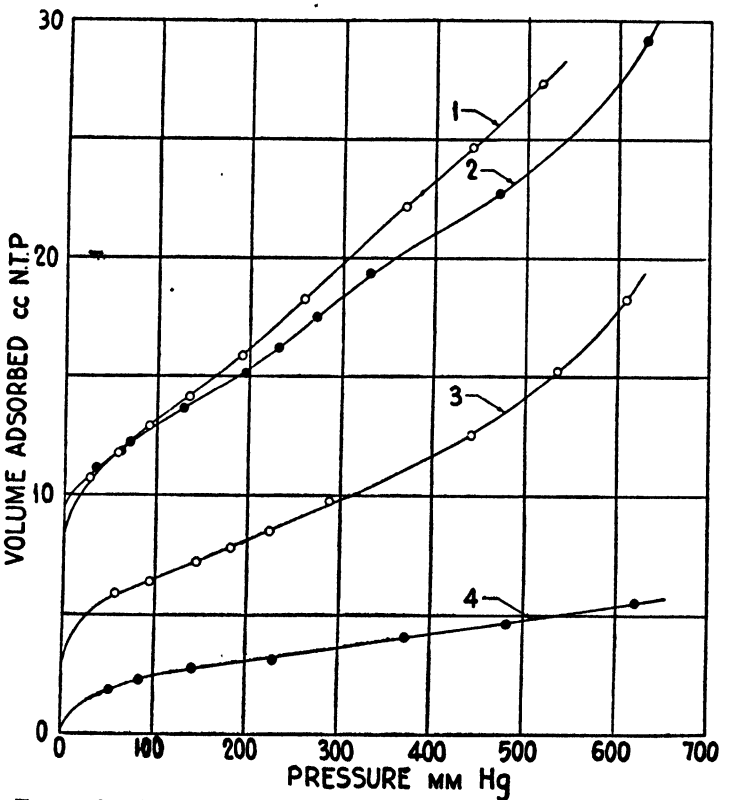


FIGURE 3. Adsorption of Nitrogen on Zinc Oxide Samples at Liquid Nitrogen Temperatures.

1. 6.44 grams of sample K-1602
2. 5.49 grams of sample F-1601
3. 7.40 grams of sample G-1603
4. 17.2 grams of sample KH-1604

Zinc Oxide.[6] Four samples of zinc oxide pigments on which measurements had been made by a number of different procedures were very kindly put at our disposal by the New Jersey Zinc Company. The nitrogen adsorption isotherms for the four samples are shown in Figure 3. The surface area values and calculated particle sizes are listed in Table 3.

Table 3. Surface Area Measurements on Zinc Oxide Pigments

Sample No.	Weight of Sample (grams)	V_m (cc)	Area (sq m/g)	Average Particle Diameter (microns)
F-1601	5.492	11.89	9.48	0.115
K-1602	6.444	12.95	8.80	0.124
G-1603	7.403	6.56	3.88	0.28
KH-1604	17.167	2.58	0.658	1.65

In Table 4 are summarized the results obtained on this standard set of zinc oxide

Table 4. Particle Diameters (microns) for Zinc Oxide Pigments Measured by Various Methods

	F-1601	K-1602	G-1603	KH-1604
d_s by direct Microscopic Observation	0.28	0.34	0.79	1.86
d_s by adsorption of Methyl Stearate	0.19	0.24	0.55	4.50
D by Microscopic Count	0.21	0.25	0.49	1.40
D by Ultramicroscopic Count	0.135	0.16	0.26	0.82
d_s by Permeability Measurements	0.12	0.15	0.25	1.25
d_s by Nitrogen Adsorption				
Using Area (L) values	0.115	0.124	0.28	1.68
Using Area (S) values	0.135	0.145	0.33	1.97

samples by about six different methods. The values for the microscopic method have recently been compared by Ewing[15] with his own results on the adsorption of methyl stearate. The ultramicroscopic values were furnished directly by the New Jersey Zinc Company and represent results obtained in their own laboratories. The measurements by the permeability method were made by Carman[16] of South Africa and were forwarded to the author by the New Jersey Zinc Company.

An examination of the results in Table 4 reveals an excellent agreement among the values for samples F-1601, K-1602, and G-1603 as determined by permeability, ultramicroscopic examination, and nitrogen adsorption methods. The slightly larger values for average diameters on these three oxides obtained by the direct microscopic examination probably indicate the presence of some particles that are invisible under the microscope. The results obtained by the adsorption of methyl stearate agree more closely with the microscopic than with the other methods of determination, as might be expected if any capillaries between particles in small agglomerates are present, which are not readily accessible to the big organic molecules diffusing through a liquid medium. The results obtained on sample KH-1604 are really not comparable with each other, since the various samples of KH-1604 had not been treated identically. This is a reheated oxide that is known to contain agglomerates of particles that had been sintered together in the reheating process. The rubbing down of such samples in preparing slides for microscopic examination, or the use of the dispersion technique in preparing the ultramicroscopic samples would on such a sample of oxide certainly be expected to reduce the average particle size compared to that of the undisturbed agglomerated samples such as were used in the gas adsorption method.

Sized Microscopic Glass Spheres.[10] The publication by Bloomquist and Clark[17]

of a method for preparing sized microscopic glass beads suggested a possible means of checking with considerable accuracy the nitrogen adsorption method for measuring surface areas. In Table 5 are listed the results of determinations on a sample of glass spheres furnished by Bloomquist. The sample was found to average about 7 microns in diameter as judged by microscopic examination. Furthermore, 90 per cent of the beads were supposed to be within two microns of the average. In agreement with this, the settling time in water according to our measurements was 84 minutes per 10 cm. Since the upper level of a mass of homogeneously dispersed settling particles would settle at a rate determined by the smallest particles present, the 5-micron value calculated from the settling time [18] observed is consistent with the claims made for the homogeneity of the bead samples.

The diameters of the beads calculated from the adsorption measurements are 4.5 microns using area L values for the area per nitrogen molecule, and 5.3 using area S values as shown in Table 5. It appeared from these results that the original treatment of the glass spheres with cleaning solution might have roughened the surfaces enough to cause the surface area measured by the nitrogen isotherms to exceed somewhat the surface expected from the average diameter of the particles. Accordingly, the beads were subjected to an additional two-hour treatment by cleaning solution and remeasured. The surface area had increased 40 per cent as a result of this treatment. If one assumes that the original cleaning of the glass spheres had etched them enough to increase the surface area 40 per cent, the measured values of the particle size by nitrogen isotherms are in almost exact agreement with those obtained by direct microscopic observation. It appears, therefore, that for glass spheres the particle size determinations by settling through liquids, by microscopic examination, and by nitrogen adsorption are all in substantial agreement.

Table 5. Surface Area Measurements on Sized Glass Beads

Area by nitrogen adsorption (area L values)	5.52 sq. m/g.
Average diameter calculated from adsorption of nitrogen (area L)	4.50 microns
Average diameter calculated from adsorption of nitrogen (area S)	5.30 "
Diameter by microscopic observation	7.20 "
Surface area after cleaning the beads with cleaning solution	7.48 sq m/g

Discussion and Summary

Space does not permit the detailed discussion [1, 19] of all the gas adsorption work on surfaces that has been reported in recent years. It may be stated by way of summarizing the present status of the method, however, that to the author's knowledge, no published data for surfaces for which independent area measurements are available can be correctly interpreted as contradicting the general conclusions of the present paper relative to the selection of V_m, the volume of gas corresponding to an adsorbed monolayer. On the other hand, a number of published results on surfaces for which independent area values were available seem definitely to confirm the ideas here expressed. As examples may be cited the results of Palmer and Clark [11] for the adsorption of acetone on a sample of powdered quartz glass whose surface area has been measured by the rate of solution in HF, and the results of Barrett, Birnie and Cohen [12] for the adsorption of water vapor on quartz glass particles whose area was obtained by a newly described method involving the projection areas of the individual particles.

The method here described for measuring surface areas is of course equally useful in some instances for calculating the particle size of finely divided substances. In conclusion, therefore, it may be well to set forth clearly the applications and

limitations of the low-temperature adsorption isotherm method for measuring surface areas and for estimating the size of microscopic or submicroscopic particles.

(1) Particle-size measurements of finely divided materials by the low-temperature gas adsorption method must be limited to non-porous solids; surface area measurements on the other hand, can be made for either porous or non-porous solids, since the "surface area" includes the inner as well as the outer or geometric surface of the particles.

(2) The particle-size calculations yield values for $d_s = \frac{6}{\rho S}$, where S is the total area in square meters per gram, ρ is the true density, and d_s is the average diameter of the particles in microns. Some assumption as to shape factors [18] is, of course, involved. Calculations in the present paper assume the particles to be spheres or cubes.

(3) With nitrogen as a measuring gas in the temperature range 77° to 90° K and at pressures up to 760 mm, the method is applicable to samples having aggregate surface areas of at least 2 square meters. Samples of larger particles can be measured provided one uses high-vacuum technique [20] and works with some gas the saturation pressure for which is only a few millimeters of Hg.

(4) The gas used for making the low-temperature adsorption measurements must be inert toward the solid being measured; if it is chemisorbed by, or reacts chemically with the solid, the resulting surface-area measurements will be considerably in error.

(5) Layers of chemisorbed oxygen, hydrogen or other gases on the surface of the solid particles will not, according to past experience, affect the validity of surface-area measurements by the low-temperature physical adsorption method. However, it is necessary to remove by evacuation at suitable temperatures condensed films of moisture or other liquids that might exclude the measuring gas from some of the otherwise accessible surface between the individual particles.

(6) The S-shaped isotherms such as are shown in Figures 1 and 3 are characteristic of all results obtained with nitrogen on non-porous, finely divided solids up to the present time. It appears very probable, therefore, that low-temperature adsorption measurements plotted with the help of Equation (1) will furnish a reliable method for estimating the size of particles that are smaller than about 20 microns. There appears to be no lower limit to the size that can be studied, though the smallest particles to which the method is known to have been applied are about 0.003 micron in diameter, as reported recently by Smith, Thornhill and Bray.[21]

(7) Most of the work so far done has used nitrogen or argon for a measuring gas. Caution must be used in predicting the behaviour of miscellaneous vapors that might be used in place of nitrogen. As pointed out by Brunauer, Deming, Deming and Teller,[22] some five different types of isotherms have so far been reported in the literature. It can be stated, however, that with nitrogen as a measuring gas only two substances have been found on which other than S-shaped isotherms of the kind shown in Figure 1 have been obtained up to relative pressures of about 0.5; these are, respectively, charcoal [8] and chabazite,[14] both of which yield flat, Langmuir-type isotherms, concave throughout to the pressure axis and supposedly characteristic of porous materials in which the capillaries are only a few molecular diameters in size.

(8) A detailed analysis of the present low-temperature adsorption method for measuring surface areas leaves little doubt as to its wide and general applicability, some gas such as nitrogen being used as an inert adsorbate. Nevertheless, there are some indications that big vapor molecules may yield surface-area values that differ somewhat from those given by nitrogen.[23] Accordingly, it is urged that the method should still be considered as one that is in the process of development and that may require some additional theoretical elaborations to account for results on all types of adsorbed molecules.

References

1. See "Advances in Colloid Science," pp. 1-36, Interscience Publishers (1942), for a full discussion of the subject.
2. Emmett and Brunauer, *J. Am. Chem. Soc.*, **59**, 1553 (1937).
3. Brunauer and Emmett, *J. Am. Chem. Soc.*, **59**, 2682 (1937).
4. Emmett and Love, *Soil Science*, **45**, 57 (1938).
5. Lineweaver, *J. Biol. Chem.*, **122**, 549 (1938).
6. Emmett and DeWitt, *Ind. Eng. Chem., Anal. Ed.*, **13**, 28 (1941).
7. Gehman and Morris, *Ind. Eng. Chem., Anal. Ed.*, **4**, 157 (1932).
8. News Edition, **18**, 492 (1940).
9. Brunauer, Emmett and Teller, *J. Am. Chem. Soc.*, **60**, 310 (1938).
10. Emmett, Am. Soc. Testing Materials, Symposium on new methods for particle size determination in the subsieve range, **41**, 95 (1941).
11. Palmer and Clark, *Proc. Roy. Soc. London*, **A149**, 360 (1935).
12. Barrett, Birnie and Cohen, *J. Am. Chem. Soc.*, **62**, 2839 (1940).
13. Green, *J. Franklin Inst.*, **204**, 713 (1927).
14. To be published
15. Ewing, *J. Am. Chem. Soc.*, **61**, 1317 (1939).
16. Private Communication. (See References 1 and 10).
17. Bloomquist and Clark, *Ind. Eng. Chem., Anal. Ed.*, **12**, 61 (1940).
18. Walker, Lewis, McAdams and Gilliland, "Principles of Chemical Engineering," p. 299, McGraw, Hill Book Co., 1937.
19. Twelfth Report of Committee on Catalysis, National Research Council, Chap. V, John Wiley and Sons, 1939.
20. Armbruster and Austin, *J. Am. Chem. Soc.*, **61**, 1117 (1939).
21. Smith, Thornhill and Bray, *Ind. Eng. Chem.*, **33**, 1303 (1941).
22. Brunauer, Deming, Deming and Teller, *J. Am. Chem. Soc.*, **62**, 1723 (1940).
23. Harris, B. L., Dissertation, Johns Hopkins University, 1941.

Adsorption and Crystal Habit Modification

WESLEY G. FRANCE

Ohio State University, Columbus, Ohio

The extensive and varied literature covering the subject of crystals is an excellent indication of the importance of the crystalline state of matter. Beginning with the observations of Steno,[47] in 1669, on the constancy of crystal angles, and continuing with the work of Haüy [28] leading to the "law of rational indices," on down to the present time, the volume of this literature has been growing at an increasingly rapid rate. This widespread interest is readily understandable when one considers the diverse fields in which a knowledge of the growth and properties of crystals is essential. Among these properties both size and shape have always been highly significant.

In some instances large, well-formed crystals are desired, as in crystal phonopick ups, crystal detectors, Nicol prisms, and infrared spectrographic prisms. Then again there is the opposite extreme, where exceedingly fine crystals are required, as in the production of the hydrous oxide gels and stable gold sols. Between these extremes many other examples are to be found. These include the production of paint pigments of increased hiding and covering powers; electrodeposition of metals made up of minute, strongly adherent crystals frequently giving rise to much-desired "bright plates"; the production of alloys and the control of their crystalline characteristics by heat treatment; the changes in crystal structure occurring during the setting of cements; and the need for fine crystals in various ceramic insulating and refractory bodies.

The organic chemist is frequently confronted with difficulties encountered in crystallizing certain organic compounds, and so also is the inorganic chemist in the preparation of pure compounds by recrystallization. The analytical chemist is often concerned with precipitations in which the crystals are too fine to be filtered or in which impurities have been adsorbed.[25b]

The formations of gall and kidney stones have been shown to be crystalline in

structure.[41, 51] The most common and direct complications which have occurred in the widespread use of some sulfonamide derivatives in therapeutics are anuria and/or hematuria, which result from the deposition of concretions of crystals of the drugs or their acetylated derivatives formed in the body.[80] The biochemist is interested in the crystallization of proteins [9] and of synthesized vitamins. Recently, virus proteins have been crystallized and described.[46]

Crystalline structure has been shown to have an important bearing on the properties of explosives. Cullen [11] states that detonating mixtures must have crystals of definite size and shape; also that lead azide can be rendered safe for handling by precipitation in the presence of adsorbable material so as to produce very small crystals.

With such a diversity of applications it is natural that investigations of the crystalline state should have proceeded along many lines. In this connection it is interesting to note that it was very early recognized that the various polyhedral crystalline forms were the result of some definitely ordered internal structure.[48, 14] Nevertheless, it was not until the development of x-ray methods of crystal analysis that exact information concerning the internal structure and its effect on the external symmetry of crystals became known. As a consequence of this development, an extensive x-ray study of the structure of crystals has been in progress during the past two decades; this has come to occupy a most prominent place in crystallographic literature. This knowledge of the internal arrangement of the ions, atoms or molecules has been of great value in accounting for both the external form and many of the physical and chemical properties of crystals. Thus, for example, the ionic or covalent character of inorganic compounds can readily be established from x-ray examinations.

While such information is of great value, it is nevertheless insufficient, especially if one is chiefly interested in the properties of the surface of a solid rather than with those of the crystal as a whole, as in the adsorption of impurities by growing crystals. Although surface properties can be fairly well predicted from the internal structure,[27] it is becoming increasingly apparent that the mechanism of crystal growth must also be known in order that a complete understanding of the crystalline state may be gained.

Among the earliest observations of the modification of crystal habit is that of Robert Boyle,[6] who noted that the normal shape of many crystals was changed by the "addition of other bodies." Somewhat later J. B. L. Romé de l'Isle [89] showed that NaCl grown in the presence of fresh urine formed octahedral crystals. In this connection it should be noted that Fourcroy and Vauquelin [16] and Beudant [4] produced octahedral crystals of NaCl from solutions containing urea.

Haüy [23] in 1801 observed that crystals of the mineral axinite colored violet by manganese have additional faces not present on those colored green by "chlorite." Senarmont [44] employed colored impurities to show that habit modification resulted from the adsorption of the impurity. Livizzari [28] noted that more carbon dioxide was evolved by the action of sulfuric acid on prism faces of calcite than on basic faces. Du Boisbaudran [5] was among the first to recognize differences in the solubilities of the faces of a given crystal.

Ord [36] seems to be the earliest worker to use the term "colloid" in connection with crystallization, although not in its modern sense. He studied the effect of gelatin, albumin and similar substances on calcium carbonate, calcium oxalate, uric acid and other salts, and concluded their effect was physical and not chemical. Becke [2] accounted for the habit of a crystal by assuming that the crystal surrounds itself with faces of least solubility. Gibbs [20] considered the equilibrium between a crystal and its saturated solution, and developed the mathematical relations, including the interfacial tensions of different faces. He showed that, for complete equilibrium, the summation, over all possible forms which may appear on the crystal, of the prod-

uct of the total area of all faces of a given form by the interfacial tension for that form in contact with the saturated solution should be a minimum. Therefore, the decrease in the total surface energy must result in a given crystal habit. A limitation of this principle was imposed by Gibbs in his statement that the tendency of similiar crystals to assume the habit required by this principle would be inversely proportional to the linear dimensions of the crystals.

A closely related and better known theory was proposed by Curie.[12] According to this theory, each crystal face has a "capillary constant" defined as "that amount of energy which must be used to enlarge the surface one unit of area against the surrounding medium." The stable form would therefore be one in which the sum of the products of the areas of the different faces and their "capillary constants" would have a minimum value. Faces having small "capillary constants" would have small external fields of force and should grow slowly. They would also have the greatest area and be most prominent. It was later pointed out by Marc and Ritzel [29] that this theory would be valid only for crystals having faces of like solubility. However, they derived an equation which provides for such differences in solubility.

Noyes and Whitney [33] developed a theory in which the rate of solution appears as a function of concentration difference and the diffusion constant. Nernst [31] further emphasized the importance of diffusion rates in his theory of reaction velocity in heterogeneous systems.

Gaubert [19] crystallized phthalic acid from a number of dye solutions and determined the amount of dye adsorbed by the crystals. Changes in crystal habit were thought to be due to the formation of a solid solution.

The investigations of Marc and co-workers [29] on crystallization velocity and solution rate established the fact that dyes retarded the crystallization velocity but did not affect the rate of solution. From this it was concluded that crystallization was not the reverse of solution. Changes in habit modification were attributed to selective adsorption at the different faces.

With the advent of specific information covering the lattice structure of crystals as determined from x-ray studies, a better understanding of adsorption forces became possible. In this connection Langmuir [27] evolved his "surface active" theory, in which the chemical activity of a solid surface is thought to be dependent upon the spacing and arrangement of the atoms or ions comprising the surface layer.

On the basis of the lattice structure of the sodium chloride type, Niggli [32] developed an interesting theory of crystal growth. According to his ideas, the residual valences of the sodium chloride lattice have components in the six directions of the cube, and planes passing through this lattice vary in the density of their ionic population. He considered the ionic layer, immediately below the crystal face, as possessing unsaturated or residual valences, and attributed the velocity of perpendicular face displacement to the thickness of that layer, this being zero for prominent and maximum for vicinal faces. With reference to crystals grown in the presence of adsorbable impurities, in accordance with this theory it appears likely that the adsorption is due only to those ions actually in the crystal face; the perpendicular face displacement results from the unused valences in the layer below the crystal face. Generally the growth forces are greater than the adsorptive forces, and vicinal faces disappear as the crystal develops.

Valeton [50] at about the same time suggested a theory of crystal growth similar to that of Niggli. Because of the electrostatic charges of the ions or atoms in the outer layer of a growing crystal, atoms or ions of unlike charge in the solution are attracted to the face and those of like charge are repelled. The cube faces of sodium chloride are made up of a checkerboard arrangement of positive and negative ions which tend partially to neutralize each other's fields. As a consequence, growth normal to these faces is slow. The octahedral planes are populated by ions of like

charge, and therefore normals to these planes should be directions of rapid growth. Furthermore, the net planes in forms of higher indices are spaced in such a way that the electrostatic forces are exerted through a number of layers, and the growth normal to such planes is also rapid.

An extension of this theory was presented later by Valeton.[50] He assigned a definite growth velocity to each face which was solely dependent on its ionic structure and independent of its size and environment. The rate of growth of any crystal face should therefore be the result of an equilibrium between its specific growth velocity and the rate at which ions can diffuse up to it.

Spangenberg [45] found, contrary to Valeton's theory, that the growth ratios of corresponding faces of potassium alum and sodium chloride were in many cases different. He thought that this was because of different lattice arrangements at the surface of the growing crystal.

Bentivoglio [8] measured the rate of growth perpendicular to a crystal face and found that similar faces of a simple form grew at the same rate independently of their size. With complex forms having a variety of faces, unlike faces grow with different rates and like faces at the same rate, provided they are not close to a large face of a faster-growing form, in which case the latter produces a deficiency of the ion in the layers of solution in contact with the crystal.

A considerable amount of work on the adsorption of foreign matter by crystalline precipitates has been carried out. Paneth [37] concluded that the adsorption of ions by a lattice consists in the replacement of a similar ion for another at the surface of the crystal, the principal requirement for this exchange being that the compound formed by the adsorbed ion and the lattice ion be insoluble. Dyes were used to estimate the surface of the precipitate on the assumption that the adsorptions were monomolecular.

Fajans regarded the residual valences of the ions in the surface of the lattice as responsible for the adsorption of foreign ions by the crystal lattice. He extended Paneth's adsorption controlling factors by the inclusion of hydration, deformability, and agreement of size of the adsorbed and lattice ions.

Kolthoff [25a] considers the adsorptive properties of ionic precipitates to be due to the residual valence forces of the ions at their surfaces. Furthermore the effectiveness of the ions is determined by their positions, those at the edges of the crystal being more effective than those within the plane and those at corners being more effective than either. He believes that crystal surfaces, instead of being plane, have submicroscopic pits and humps which became smoother on ageing.

The writer's interest in the adsorption of foreign substances by growing crystals developed from a series of studies of stability of suspensoid sols. With the establishment of the crystalline structure of the dispersed phase of a number of sols,[42, 22, 52] it seemed highly probable that the growth forces of these colloidal crystals have an important part in the adsorption of the ions responsible for the stability of the sol. Furthermore, since the lattice structure was found to be the same for both macroscopic and colloidal crystals, it was thought that worthwhile information relative to suspensoid sol stability could be obtained by investigating the factors influencing the adsorption of foreign substances by macroscopic crystals during their growth from saturated solutions.

The work in this laboratory [17] has been mainly concerned with the adsorption of foreign substances at the crystal-solution interface during growth and with the habit changes resulting from such adsorption. In this connection attempts have been made to learn both how the adsorbed ions, atoms or molecules are held in the crystal and the nature of the directive forces responsible for the adsorption.

These investigations have included: (a) the measurement of the changes in growth ratios due to adsorption; (b) the correlation of lattice structure with habit modification; (c) x-ray and electron-diffraction studies of crystals containing dyes

that have modified the crystal habit; (d) the determination of the quantity of dye adsorbed by growing crystals; and (e) the effect of the molecular structure of dyes on their adsorption by growing crystals.

At the outset it was clearly recognized that in order to establish a relation between crystal structure and the adsorption of foreign substances, a number of experimental conditions have to be met. These include: (1) the use of crystalline substances whose lattice structures are known; (2) the use of foreign substances of known molecular structure; (3) the selection of some crystallographic constant capable of accurate measurement whose value varies with the adsorption process; (4) in so far as possible the use of a single macroscopic crystal freely suspended in its saturated solution uninfluenced by the presence of other crystals.

These conditions were satisfied by the measurement of the growth ratio of a single crystal grown in the presence of dyes of known structure at constant temperature.

Studies of potassium and ammonium alums[17d, e] led to the conclusion that the adsorption of any given dye by a growing crystal is a specific process dependent upon both the structure of the crystal and the shape, size, presence and orientation of polar groups in the dye molecule. In general the adsorption was greater on those faces having the stronger fields of force. Thus, in the case of the alums, the adsorption appeared to be entirely on the cube planes populated by ions of like charge, leaving the octahedral planes, populated by mixtures of positive and negative ions, clear and colorless (Figure 1).

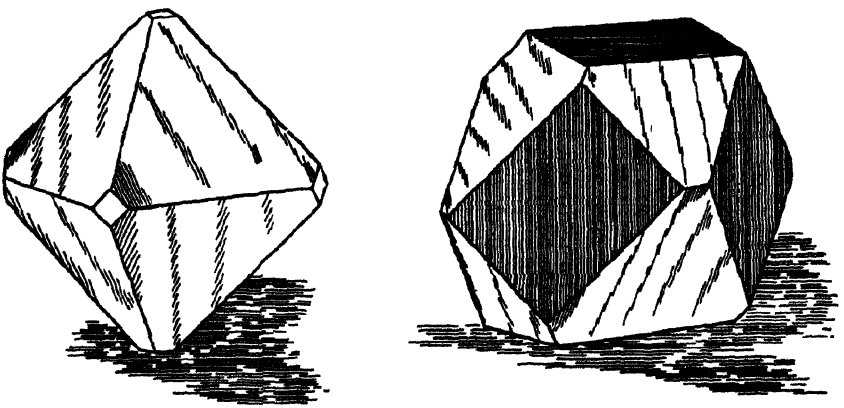


FIGURE 1. Alum crystals before and after growth in presence of an adsorbable dye.

As a result of this preferential adsorption, the growth ratio 100/111 of potassium alum was reduced from 1.61 to 0.00 depending upon the concentration of dye present in the solution (Figure 2).

These results indicate the possibility of predicting the habit modification of any given crystal from a knowledge of its lattice structure provided its growth occurs in the presence of an adsorbable foreign substance.

That such predictions are possible was verified by the observation of the habit modification of sodium, barium, lead, and lithium nitrates, sodium bromate, potassium chloride and bromide, potassium and sodium sulfates, cupric acetate, urea, and finally citric and tartaric acids, grown in the presence of adsorbable dyes.

Since the growth of a crystal takes place as the result of the adsorption of its ions into the lattice, it is evident that the various polyhedral forms of crystals result from differences in growth velocities in specific directions. From this point of view a consideration of the habit and lattice structure of the alums indicates a high

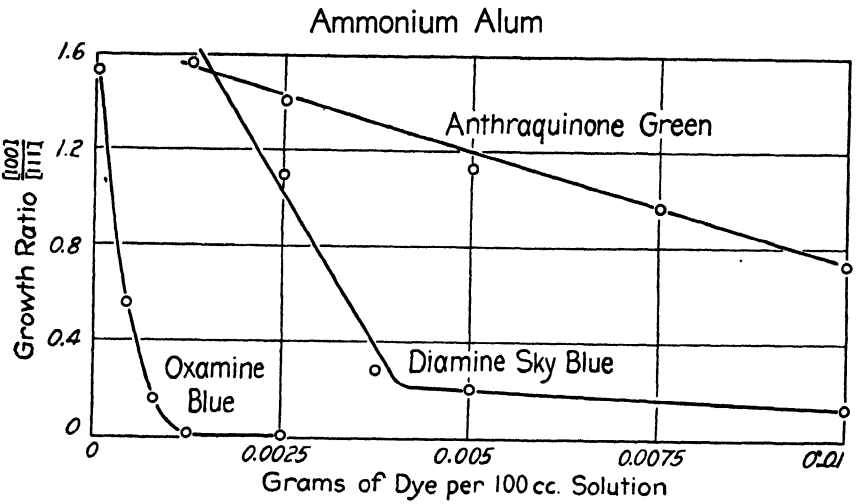


FIGURE 2. Growth ratio—concentration curves for ammonium alum.

specific growth rate for the cube faces such that the actual growth rate is strongly influenced by the rate of diffusion of the ions to the crystal face.[50] Therefore any condition such as stirring, or growth from supersaturated, rapidly cooling, or rapidly evaporating solutions, which increases the rate of diffusion, should increase the actual growth rate of the cube face and hence the growth ratio 100/111 of an alum

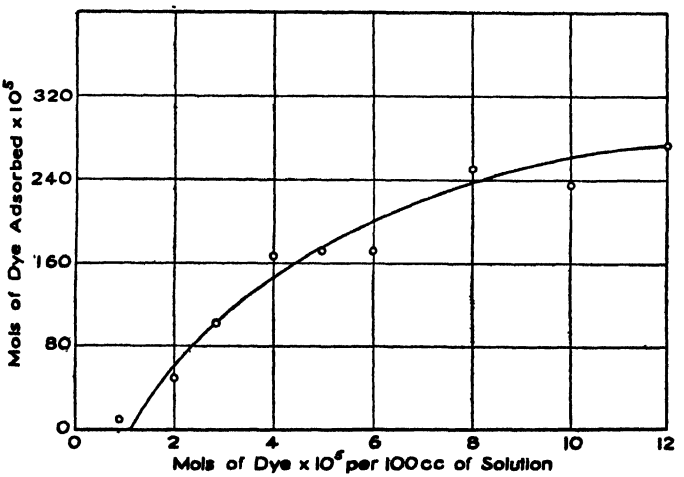


FIGURE 3. Adsorption of P-β OH-6 by potassium sulfate.

crystal. Precisely such effects were observed experimentally. Thus the growth ratio of pure potassium alum [171] was increased from 1.61 to 1.75 with the elimination of the cube faces by rapid stirring of the solution during growth. It was also found that when dyes were present, smaller quantities were adsorbed at the same concentration from stirred than from unstirred solutions.

In all cases, whether the solutions were stirred or not, neither Langmuir's adsorption equation nor Freundlich's adsorption isotherm adequately described the relation between the concentration of foreign matter in the solution and the amount

adsorbed by the growing crystal. The a-concentration curves (Figure 3) do not pass through the origin of the coördinate system but begin at a finite distance from the origin; that is, there is a minimum concentration of dye below which no adsorption occurs. However, at concentrations above this minimum a typical adsorption isotherm is obtained. This can be represented by a modified form of the equation of the adsorption isotherm as follows

$$a = kc^{1/n} - k'c^{-1/n'}$$

The second term of this equation appears to be in the nature of a "counter adsorption process" and is primarily a function of the dye concentration; it is infinite at zero concentration and even at relatively low concentrations rapidly approaches zero. It can be accounted for as follows. Normally, even in the presence of an adsorbable dye, the crystal takes up its own ions in preference to the dye, just so long as the crystal ions are readily available. If, however, the supply of lattice ions adjacent to the crystal face becomes depleted, because of rapid growth and slow ionic diffusion, the crystal will then satisfy its growth forces by taking up available and adsorbable dye ions, when their concentration becomes sufficiently great. The intercept of the curve on the concentration axis represents the minimum concentration of dye necessary to establish this condition.

The value of predicting habit modifications resulting from adsorption processes would be greatly increased if one could, at the same time, predict just what foreign substances would be adsorbed by a given crystal. However, attempts to make such predictions have thus far not been especially successful. This has presumably been caused, in part at least, by a lack of knowledge as to just how the foreign molecule or ion is adsorbed and held by the crystal.

The foreign ion or molecule may be held interstitially within the crystal interstices, or it may actually replace an ion in the crystal lattice, especially if there is an identity or close similarity in the size and shape of an active group in the foreign ion with that of the ion it replaces.

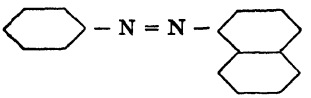
That the habit modification results from the adsorption of ions or molecules and not colloidal particles is indicated by the fact that solutions containing the dyes failed to show any relation between the number of colloidal particles present and the magnitude of the adsorption.[17f, 1]

Measurements of the quantity of dye adsorbed and the total area of the ionic planes involved indicated that the dye was not adsorbed in a continuous monomolecular layer over the ionic planes. Thus, in the case of diamine sky-blue, the total area of the cube planes of an ammonium alum crystal was estimated to be 4300 times greater than the area covered by the adsorbed dye molecules.[17f]

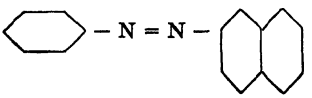
The marked dichroism frequently observed in the dyed crystals indicates a definite orientation of the dye. However, the absence of dichroism does not in itself suggest a complete absence of orientation, since adsorbed complex and elongated molecules may all be attached to the crystal in the same way, but may be coiled up so that rotation of the crystal makes no detectable difference in their refracting powers. Furthermore, there is the possibility of the adsorption of elongated molecules whose long axes are all parallel, but which nevertheless appear to be randomly distributed in polarized light due to rotation about one or more of the single bonds in the individual molecules.

Bunn [8] has suggested that pleochroism and birefringence are due to colorless impurities. Gaubert [19h] and Tamman and Laass [49] report pleochroism due to adsorbed impurities. Gaubert [19g] explained the pleochroism he observed in cubic crystals of lead nitrate containing adsorbed dye, on the assumption of the presence of microscopic striations of dye, resulting from the simultaneous crystallization of the dye and crystal at certain concentrations, followed for a time by the deposition of nearly pure host crystal.

In order to determine how the presence and distribution of strong polar groups in the dye molecule affect its adsorption by a growing crystal, investigations [17m, n, o] were undertaken in which a series of isomeric acid and basic monoazo dyes was used. These dyes were prepared by Griffith [21] and Eberhart [15] under the direction of Professor Wallace R. Brode for spectroscopic work, and were of an unusually high degree of purity. They were made by coupling aniline and the ortho-, meta-, and para-sulfonated anilines, by diazotization, with a series of mono- and di-sulfonic acid derivatives of α- and β-naphthols and α- and β-naphthylamines. For convenience these dyes are represented by abbreviations derived from the intermediates used in their preparation, as follows: (A) aniline; (O) ortho-sulfonated aniline; (M) meta-sulfonated aniline; (P) para-sulfonated aniline; (α-OH) α-naphthol; (β-OH) β-naphthol; (α-NH_2) α-naphthylamine; (β-NH_2) β-naphthylamine. Thus the scale model (Figure 4) is the structural representation of the sodium salt of the dye P-β OH. There were in all 38 acidic and 48 basic dyes. About half of these contained the nucleus



and the remainder the nucleus



Potassium sulfate grown in the presence of these dyes as impurity provided a number of interesting results.[17m] The adsorption generally occurred on the (110) and (111) planes, which according to the ideas previously expressed, should therefore be those having the stronger fields of force. In order to aid in the interpretation of the data, scale models (1Å = 1 cm) of the dyes and potassium sulfate were constructed (Figures 4, 5). For the latter the dimensions of the unit cell determined by Ogg and Hopwood were used.[84, 85] The potassium sulfate model clearly showed the (100) planes to consist of a mixture of K^+ and SO_4^{--} ions and the (111) and (110) to be alternately either K^+ or SO_4^{--}.

That the $-SO_3Na$ group of the dye molecule is active in determining whether the dye is adsorbed or not is evident from the fact that when this group is sterically hindered or absent from the benzene ring the dye is not adsorbed. Also when this group is in the 1-, 4-, 5- or 8-position of the naphthalene ring, structural hindrances are encountered. Thus the most strongly adsorbed dyes were those in which the $-SO_3Na$ groups were in the meta and para positions in the benzene rings and 6- and 7-positions of the naphthalene rings.

In view of these results it seems probable that the adsorbed dyes are attached to the potassium sulfate crystals by means of the oxygen triangle of the $-SO_3Na$ group. These dyes ionize in solution, furnishing a large anion having a charge equal to the number of substituent $-SO_3Na$ groups, together with a sodium ion for each of the latter. The dye anion thus has three oxygen atoms in an equilateral triangle attached to a sulfur atom as in the SO_4^{--} ion, the remainder of the dye molecule assuming the position occupied by the fourth oxygen atom. It is therefore possible for the $-SO_3$ group of the dye anion to be substituted for the SO_4^{--} ion of the potassium sulfate crystal when present in favorable concentration.

Buckley had previously suggested [7b, f] such a means of attachment of RO_4^{--} and similar ions to the sodium chlorate crystal. He found that most substances that ionize to give an RO_4^{--} type ion result in the development of (111) planes of sodium

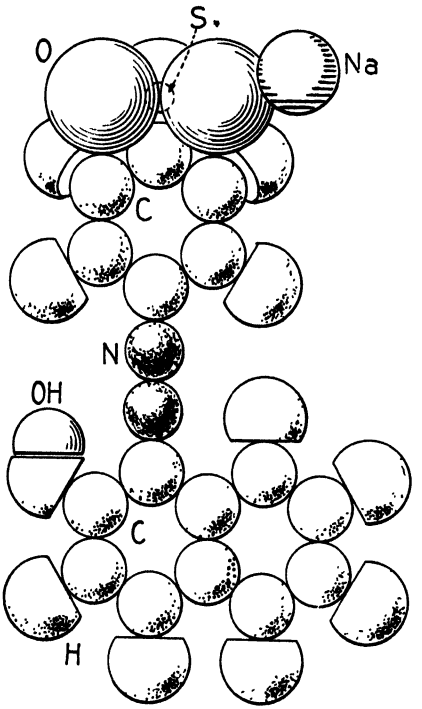


FIGURE 4. Model of dye P-β OH.

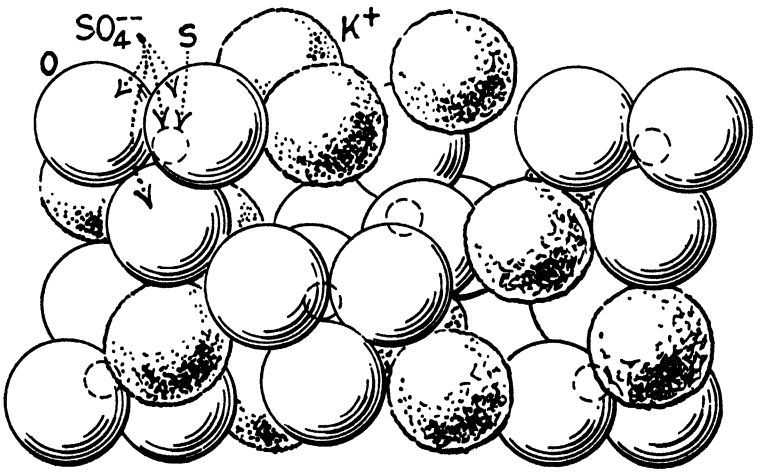
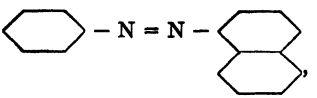


FIGURE 5. Model of potassium sulfate. Side elevation (100) plane.

chlorate at the expense of (100). He regarded the size and distribution of forces in the O_s portion of the impurity ion as the main factor in determining whether that ion would obstruct the growth of the crystal. He later included dyes in his investigations and concluded that the mode of attachment of the dyes and foreign inorganic ions is identical.

The results obtained with potassium sulfate [17m] suggested that the dipole moments

of the dye molecules might also be of considerable influence in accounting for their adsorption. Thus it was found that the dyes produced from the β-naphthol sulfonic acids were more effective than those from α-naphthol sulfonic acids. However, if NH_2-groups were substituted for the OH-groups, the reverse was true, and those dyes prepared from the α-naphthylamine sulfonic acids were then more effective than those from the β-naphthylamine sulfonic acids. These facts, in terms of the dyes having the nucleus



mean that the most effective position for the —OH group is ortho to the —N = N-group, and for the $-NH_2$ group it is the para or 4-position in the naphthalene ring. Williams [88] found the dipole moment of the —OH group to be —1.5 Debye units and that of the $-NH_2$ group to be +1.7. The dipole moments of the dye molecules would therefore be more nearly the same if the —OH group were ortho and the $-NH_2$ group para to the —N = N-group, assuming of course that the other groups remain in the same positions.

These dyes were also used as impurity with sodium nitrate, bromate and chlorate.[17m] In the case of the latter salt, no habit modification was produced; however, four of the dyes colored the crystal uniformly throughout. The results obtained with sodium nitrate and bromate were in agreement with those obtained with potassium sulfate, except that the influence of the dipole moment appeared to be less evident.

X-ray powder spectrographs were made of dyed and undyed crystals of potassium alum, sodium nitrate, lead nitrate, and urea-sodium chloride.[17e, i, l] In all cases the lattice constants were the same for the dye modified and the pure crystals. This identity of the lattice constants was evidently due either to the actual failure of the adsorption of the dye to affect the plane spacings, or to the inability of the x-ray method to detect such small changes as may have occurred. The former alternative was considered as indicating interstitial adsorption rather than ionic replacement.

However, because of the inconclusiveness of these x-ray studies, it was thought that electron diffraction investigations might be better, since electrons are more readily scattered than x-rays and therefore should be more satisfactory in a study of the surface structure of dyed and undyed crystals. An electron-diffraction camera was constructed [17p] and single crystal reflection pictures of dyed and undyed sodium bromate crystals were taken. The measurements of these plates showed that the lattice constant of the habit-modified crystal was slightly but definitely greater than that of the pure sodium bromate. If further results prove this to be a real lattice stretching, and not due to an inner potential effect, it may be attributed to the replacement of the BrO_3^- ion by the large dye anion.

The results obtained by the writer and co-workers in their studies of the adsorption of foreign substances by growing crystals have led to the following conclusions:

(1) The theories of Niggli, Valeton and Spangenberg are generally acceptable.

(2) The adsorption of foreign substances by a growing crystal is dependent upon (a) the lattice structure of the host crystal; (b) the residual valency force fields; (c) the ionic structure of the crystal face; and (d) the presence, size, shape, position and orientation of polar groups within the foreign molecule.

(3) There appears to be no simple rule deducible from these observations, whereby one can predict what foreign materials will be adsorbed by any given crystalline substance.

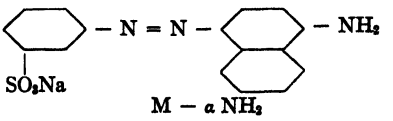
(4) From a knowledge of the lattice structure of a crystal it is possible to predict,

with some success, the habit modifications resulting from the adsorption of foreign substances during the growth of the crystal.

(5) Dyes in which $-SO_3Na$ groups are active in determining the adsorption are unlikely to be adsorbed if such groups are sterically hindered.

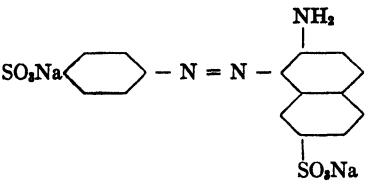
It should be pointed out in connection with the last of the foregoing conclusions that the converse is not necessarily true. There are many cases in which a dye is not adsorbed, even though its $-SO_3Na$ groups are not sterically hindered. This is illustrated in the case of sodium bromate grown in the presence of the following dyes:

(a)



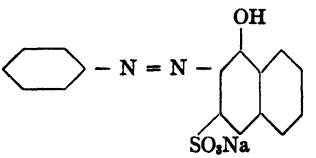
M – α NH_2

(b)



P – β NH_2 7

(c)



P – α OH_3

The first is readily adsorbed and modifies the habit of sodium bromate; the second is not adsorbed; and the third is but slightly adsorbed without habit modification.

A number of workers in this field have not concurred in these conclusions. That this is so is rather to be expected, especially in view of the almost complete lack of knowledge available concerning the mechanism of crystal growth.

Buckley[7] concluded that the adsorptions are more general and less specific than the writer considers them to be. He also questioned the importance attached to the residual valency force fields,[7k] believing that while these force fields may "draw in" the impurity, something else, presumably located in definite spots in the molecule to "fit a similar orientation in the surface," is required to hold the impurity. This criticism is met by the writer's use of Buckley's theory of the mode of attachment of the dye, in which the oxygen triangle of the sulfonate groups replace the SO_4^{--} ions of potassium sulfate. Even so, this mode of attachment is not general, as there are many cases in which there appear to be no common characteristics between crystal and adsorbed substance, especially in reference to structural similarity.[17l, 18]

His investigations led him to assume a growth process, as suggested by Kossel,[26] in which the crystal grows row by row, with the rows forming in a single direction only on a given crystal plane.

In order to account for the fact that frequently only a small amount of dye is included in a crystal, he thought a periodic resolution process occurred in which most of the impurity was removed.

He recognized three types of adsorption: (a) the simplest type consisting of those cases explained by the adsorption of a known ion or molecule on one set of planes; (b) a more complex type in which there is an acceleration of the rates of growth

of certain planes, leading to their disappearance; and (c) a type in which a number of planes associated about an axis or face normal are affected together. Most of the theories so far proposed apply only to the first and simplest type of habit modification.

Frondel [18] also questions the importance of the residual valence force fields. He cites as evidence the fact that out of 59 cases of dye adsorption by sodium and lithium fluorides, only four showed adsorption on the (111) planes, and that if the residual force fields were effective the adsorption should have been on the (100) planes.

One can also point to similar results by Reinders [88] with silver chloride and by Gaubert [19b] with silver bromide and chloride. This is not necessarily an argument against the residual valency force fields. Rather it means that the structure of the crystals and the dyes used were such that a structural fit was impossible on the (111) planes but possible on the (100), under which circumstances the adsorption occurred on the weaker rather than on the stronger planes. From the outset, the writer has maintained that the adsorption is determined not alone by the residual valency force fields, but also by some structural relation between the dye and the crystal lattice, so as to enable either a part or all of the dye to be fitted into the lattice. In the majority of cases studied in this laboratory the adsorption has been found to occur on planes populated by like ions; and in the much fewer cases where it has been on the checkerboard type of planes the adsorption has been decidedly weaker.

Another objection, offered by Frondel, to the writer's use of residual valency force fields is that he believes it requires the fundamental assumption that there are differences in the surface energy between different faces of a crystal that are meaningful with reference to crystal growth and adsorption. However J. Willard Gibbs [20] has shown that such an assumption is not valid as applied to macroscopic crystals. In support of his objection Frondel cites the work of Cottrell.[10] Since Cottrell was dealing primarily with equilibrium systems and not dynamic growth processes, the objection appears to be unwarranted. Furthermore, Cottrell's argument that the assumption of different diffusion rates to various crystal faces cannot be used to explain crystal habit is not pertinent, since the writer has never made use of such an assumption. The influence of diffusion is probably related to the specific growth rate of a given crystal face in the manner suggested by Valeton.[50]

Frondel's criticisms are based in part on the work by Kossel [26] and Stranski [48] on crystal growth. They analyzed a mechanism of growth of macroscopic crystals of the sodium chloride type from a consideration of the energy relations involved. Kossel has made calculations for the relative energies released by the deposition of an ion upon the (100) or the (110) planes of the sodium chloride type of crystal. It seems however that this treatment is rather limited in that it has proved satisfactory for only the external fields of electrically neutral crystal faces.

Certainly the Kossel-Stranski theory, in which the ions are deposited row by row on the cube faces of sodium chloride type crystals, provides a much more regular and uniform growth than can be offered by the deposition of ions on the (111) planes due to residual valency force fields. However, it has been pointed out by Davey [13] that there is much evidence indicating that crystal growth takes place along a sort of three-dimensional lattice work, which is later filled in with crystal material, rather than proceeding uniformly along a plane surface. Furthermore, according to Kleber,[24] the theory of Kossel and Stranski is not applicable when adsorption and other secondary effects predominate.

Bunn [8] has proposed a theory of adsorption and crystal growth employing the basic concepts developed by Gaubert [19a] that the inclusion of impurity in an oriented form, the parallel growths of one crystal on a larger crystal of another substance, and the habit modification of a crystal are all closely related phenomena. One point stressed is that information at hand concerning mixed crystal formation can be

advantageously applied, with necessary modification, to adsorption problems, since mixed crystals and adsorption bodies apparently differ only in degree. In this connection it is emphasized that while a complete identity of lattice structure is required for stable mixed crystal formation, a similarity of the lattice structure of the two substances along certain planes only is necessary for strong oriented adsorption and unstable mixed crystal formation. Bunn makes use of the instability of the mixed crystal to account for habit modification; he states that, from his point of view, the conditions for adsorption and habit modification and oriented overgrowth are alike, and that modification of habit should be reciprocal. He also found evidence in support of the opinion that a redissolving process takes place concurrently with growth.

While there are undoubtedly pairs of substances that support Bunn's views and exhibit reciprocal habit modification, nevertheless such an explanation, by its requisite of identity of lattice structures, must fail to account for a very large proportion of adsorptions and habit modifications. In other words Bunn's theory is highly restricted and therefore greatly limited in its application. Thus, Frondel [18] observed two types of adsorption, only one of which appeared to be in harmony with it.

Royer,[40] in a series of papers, expresses agreement with Bunn in reference to the necessity for similarity of lattice structure in the absorbed substance and growing crystal.

Balareff and co-workers [1] have developed an interesting theory of "inner adsorption" for crystal growth in which it is assumed that the growth of an actual crystal, as distinguished from an ideal crystal, occurs at a large number of unrelated nuclei distributed over the entire surface of the crystal, and that each nucleus ceases to grow when it comes in contact with neighboring nuclei. The resulting macro-crystal is therefore a mosaic of tiny crystals of colloidal dimensions. When growth has stopped, foreign substances can be adsorbed on the exposed crystal faces within the macro-crystal. While there is extensive evidence in favor of the general existence of imperfect crystals, it is difficult to see how Balareff's theory can be used to explain adequately the marked and preferential adsorption that takes place on certain crystal faces during growth in the presence of impurities.

A consideration of the various theories of crystal growth and adsorption presented here indicate that of Buckley, based on the Kossel-Stranski theory, and that of Bunn to be complementary. The writer has had somewhat the same viewpoint as Buckley, but differs in using the residual valency force fields in a way more nearly corresponding to Valeton's theory than to that of Kossel and Stranski.

Differences of opinion such as expressed by the various workers studying crystal growth and adsorption phenomena strikingly emphasize the need for exact knowledge concerning the way in which a crystal grows, in order that an understanding of adsorption and habit modification processes may be gained.

Admittedly, comparatively little is definitely known concerning adsorption by growing crystals, even though the volume of literature on this subject is rather extensive and varied. This situation is in itself a strong argument for further work in this interesting and important field.

Such work should certainly include additional studies of the Kossel-Stranski theory by extending the energy calculations to more complex crystal systems. The following investigations are either in progress or contemplated in the writer's laboratory: (a) a study of the adsorption of dyes with known dipole moments; (b) the use of the mono-azo series of dyes with ammonium sulfate which is isomorphous with potassium sulfate, and also with tartaric acid which crystallizes in the triclinic system; (c) an extension of the electron diffraction and x-ray studies of modified and unmodified crystals; and (d) use of the electron microscope in an attempt to learn more about the "fine structure" of the crystal surfaces.

Some preliminary studies have already been made with the electron microscope [17b]

and tiny crystals of apparently perfect form having dimensions smaller than 5×10^{-5} cm have been observed, together with some irregular forms difficult to account for. This preliminary study indicates the great need of further work, using crystals of known fine structure in order to interpret correctly the images resolved. In this connection it would no doubt be very much worthwhile to parallel the electron microscope investigations, using the same materials, with the ultraviolet light microscope. Furthermore, much needs to be done in the direction of the preparation and mounting of the samples for observation with the electron microscope. These and many other investigations will doubtless be carried on as the full power of the electron microscope in crystallographic studies comes to be realized.

References

1. Balareff, *Kolloid Chem. Beihefte*, **30**, 249 (1930); **32**, 205, 304, (1931); **33**, 279 (1931); **34**, 441 (1932); **37**, 180, 324 (1933); **50**, 1 (1939); *Z. anorg. Chem.*, **154**, 170 (1926); *Kolloid Z.*, **61**, 244 (1932); **66**, 51 (1934); **67**, 203 (1934); **72**, 25 (1935); *Z. Krist.*, **89**, 268 (1934); **100**, 167 (1938); *Z. anal. chem.*, **115**, 104 (1938); **118**, 411 (1940).
2. Recke, *Z. Krist.*, **21**, 185 (1888).
3. Bentivoglio, *Proc. Roy. Soc. London*, **A115**, 59 (1927).
4. Beudant, "Ann. des Mines," 1818.
5. du Boisbaudran, *Compt. rend.*, **80**, 888 (1875); **88**, 629 (1879).
6. Boyle, "Origin of Forms and Qualities," Oxford, 1666.
7. Buckley, (a) *Z. Krist.*, **75**, 15 (1930); (b) **76**, 147 (1930); (c) **78**, 412 (1931); (d) **80**, 238 (1931); (e) **82**, 31, 37, 47, 285 (1932); (f) **85**, 58 (1933); (g) **88**, 381 (1934); (h) **91**, 375 (1935); (i) **97**, 370 (1937); (j) *Mem. Proc. Manchester Lit. Phil. Soc.*, **83**, 31 (1938-39).
8. Bunn, *Proc. Roy. Soc. London*, **A141**, 567 (1933).
9. Cohn, Fankuchen, Oncley, Vickery and Warren, *Ann. New York Acad. Sci.*, **41**, 77 (1941).
10. Cottrell, *J. Phys. Chem.*, **10**, 52 (1906).
11. Cullen, *Trans. Inst. Chem. Eng.*, 17, 14 (1939).
12. Curie, *Bull. Soc. Min. Franc.*, **8**, 145 (1885); *Z. Krist.*, **34**, 512 (1901).
13. Davey, *Phys. Rev.*, **29**, 206 (1927).
14. Delafosse, *Mem. Acad. Belgique*, **8**, 641 (1843).
15. Eberhart, Dissertation, Ohio State University, 1935.
16. Fourcroy and Vauquelin, *Ann. de Chimie*, **32**, 130 (1810).
17. a. France and McBurney, *J. Am. Chem. Soc.*, **46**, 540 (1924).
    b. France, "Colloid Symposium Annual," III, 317 (1925).
    c. Eckert and France, *J. Am. Ceram. Soc.*, **10**, 579 (1927).
    d. Keenen and France, *ibid.*, 10, 821 (1927).
    e. Bennett and France, *ibid.*, **11**, 571 (1928).
    f. Lash and France, *J. Phys. Chem.*, **34**, 724 (1930).
    g. Foote, Blake and France, *J. Phys. Chem.*, **34**, 2236 (1930).
    h. France, "Colloid Symposium Annual," VII, 59 (1930).
    i. Weinland and France, *J. Phys. Chem.*, **36**, 2832 (1932).
    j. Paine and France, *ibid.*, **39**, 425 (1935).
    k. Davis and France, *ibid.*, **40**, 81 (1936).
    l. France and Davis, *ibid.*, **40**, 177 (1936).
    m. Rigterink and France, *ibid.*, **42**, 1079 (1938).
    n. France and Wolfe, *ibid.*, **45**, 395 (1941).
    o. France and Wolfe, *J. Am. Chem. Soc.*, **63**, 1505 (1941).
    p. Blomquist, and France, *J. Phys. Chem.*, **46**, 1044 (1942).
18. Frondel, *Am. Mineral.*, **25**, 91, 338 (1940).
19. Gaubert a. *Bull. soc. franc. mineral.*, **17**, 121 (1894).
    b. *ibid.*, **23**, 211 (1900).
    c. *ibid.*, **28**, 180, 286 (1905).
    d. Compt. rend., **142**, 219 (1906).
    e. *ibid.*, **147**, 632 (1907).
    f. *ibid.*, **157**, 1531 (1913).
    g. *ibid.*, **167**, 491 (1918).
    h. *ibid.*, **194**, 109 (1932).
20. Gibbs, *Trans. Conn. Acad.*, **3**, 489 (1878); "Collected Works," I, 221, 325.
21. Griffith, Dissertation, Ohio State University, 1934.
22. Haber, *Ber.*, **55B**, 1730 (1922).
23. Haüy, "Traité de mineralogie," Paris, 1801.
24. Kleber, *Kolloid Z.*, **94**, 39 (1941).
25. a. Kolthoff, *J. Phys. Chem.*, **40**, 1027 (1936).
    b. Kolthoff and MacNevin, *J. Phys. Chem.*, **44**, 1921 (1940).
    c. Kolthoff and Eggersten, *ibid.*, **46**, 616 (1942).
26. Kossel, "Handworterbuch der Naturwiss.," 2nd ed. **10**, 372 (1934); *Physik. Z.*, **29**, 553 (1928). See also paper by L. Lichtwitz in this volume. *Nach. Gess. Wiss. Göttingen Math. Physik Klasse*, **1927**, 135.
27. Langmuir, *J. Am. Chem. Soc.*, **38**, 221 (1916); *Phys. Rev.*, **6**, 79 (1915).
28. Livizzari, "Nouveaux phénomones des crops crystallisés" Lugano, Canton Printer, 1865.
29. Marc and co-workers, *Z. physik Chem.*, **61**, 385 (1908); **67**, 470, 640 (1909); **68**, 104 (1910); **73**, 685 (1910); **75**, 710 (1910); **76**, 58, 584 (1911); **79**, 71 (1912); **81**, 641 (1913); *Z. Elektrochem.*, **15**, 679 (1909); **16**, 201 (1910).
30. Myers, Wm. G., Private Communication.
31. Nernst, *Z. physik chem.*, **47**, 52 (1904).
32. Niggli, *Z. anorg. chem.*, **110**, 55 (1920).
33. Noyes and Whitney, *Z. physik. chem.*, **23**, 689 (1897).
34. Ogg, *Phil. Mag.*, **5**, 354 (1928); **9**, 665 (1930).
35. Ogg and Hopwood, *Phil. Mag.*, **32**, 518 (1916).

36. Ord, "On the Influence of Colloids upon Crystalline Form and Cohesion with Observations on the Structure and Mode of Formation of Urinary and Other Calculi," London, 1879.
37. Paneth, *Physik. Z.*, **15**, 924 (1914); *Z. physik. chem.*, **89**, 512 (1915); *Ber.*, **48**, 700 (1915); **57B**, 1215 (1924).
38. Reinders, *Z. physik. chem.*, **77**, 677 (1911).
39. Romé de L'Isle "Crystallographie" (Paris) **1**, 379 (1783).
40. Royer, *Compt. rend.*, **198**, 185, 585, 949 (1934).
41. Schade, *Kolloid Z.*, **4**, 175 (1909); "Colloid Chemistry," Alexander, Vol. II, 803 (1928).
42. Scherrer, *Nachr. Ges. Wiss. Göttingen*, 96 (1918).
43. Seeber, *Gilbert's Ann.*, **76**, 229 (1824).
44. Senarmont, *Ann. Chim. phys.*, **3**, 41, 319 (1854); *Compt. rend.*, **38**, 101 (1854).
45. Spangenberg, *Z. Krist.*, **59**, 375, 383 (1924); **61**, 189 (1925); **65**, 204 (1927).
46. Stanley, *J. Phys. Chem.*, **42**, 55 (1938).
47. Steno, "De solido intra solidum," Florentia, 1669.
48. Stranski, *Z. physik. chem.*, **136**, 259 (1928); **163**, 399 (1932).
49. Tamman and Laass, *Z. anorg. Chem.*, **172**, 65 (1928).
50. Valeton, *Physik. Z.*, **21**, 606 (1920); **59**, 135, 335 (1924).
51. Weiser and Gray, *J. Arch. Path.*, **17**, 7 (1934).
52. Weiser, "Colloid Chemistry," John Wiley and Sons, 1939.
53. Williams, *J. Am. Chem. Soc.*, **52**, 2352 (1928).
See also J. Alexander, *Koll. Zeit.*, 1909 and 1910, on effects of colloids on crystallization.

Some Applications of Selective Adsorption and Differential Diffusion in Chemical Analysis

BEVERLY L. CLARKE

Bell Telephone Laboratories, Inc., New York, N. Y.

Adsorption and Diffusion

If one begins with a homogeneous mixture of several chemical species, and establishes an interface, concentration gradients may spontaneously set themselves up with respect to one or more of the species, in the region of the interphase boundary: this is adsorption. Equilibrium is established when the interfacial energy has attained a minimum, which is predictable from the Principle of Least Energy; whence arises Gibbs' Law, that the intensity of the adsorption of a substance at an interface parallels its capacity to lower the interfacial energy.

The theory of adsorption presents the following qualitative picture: a molecule below the surface boundary of a solid, liquid or gas is completely surrounded by other molecules, so that the attractive forces emanating from it tend to be neutralized. But a molecule *at* the surface is not completely surrounded by its fellows, and therefore possesses unneutralized affinity, or energy. If now some substance dispersed in one of the contiguous phases has the capacity, if present at the interface, to neutralize some of this residual energy, it will spontaneously proceed to the interface. In other words, for the originally existing energy gradient across the interface there is substituted a concentration gradient.

The process by which material aggregates proceed to an adsorbent interface might be termed directed diffusion; as a molecule of the uniformly dispersed adsorbate-to-be just below the interface is adsorbed, another moves up to take its place, and this process continues until equilibrium, as expressed by the well-known adsorption isotherm, is reached. Thus diffusion, the natural function of which is to eliminate concentration gradients and produce uniform dispersal, is inveigled into becoming a party to the attainment of the precisely opposite result. Nevertheless, at equilibrium, the great principle that diffusion serves—the Second Law of Thermodynamics—still operates; and thus the state of equilibrium represents a dynamic balance between two opposing forces, that of adsorption and that of diffusion.

There is another important way in which the force of diffusion may be made to accomplish the attainment of heterogeneity, the most familiar exemplification of which is the process of dialysis. Here diffusion would pass both colloid and crystalloid through the membrane, as in all other directions, but the membrane retards the

passage of the colloid, so that the process becomes differential diffusion and results in a separation.

Similarly, adsorption is selective in that the forces by which different substances are drawn to and held by a given adsorbent are in general different. Thus the process of adsorption from a mixture of substances results in an adsorbate whose components are in a different ratio from that in the original mixture.

Capillary Effects

If we limit our consideration to adsorption at gas-liquid interfaces, we are not concerned with capillarity, for matter in these states, although discontinuous in the molecular sense, and possessing at any given instant in time a network of intermolecular voids, lacks entirely the rigidity and permanence of structure that characterize the solid state. (There are of course gradations, in respect to structure, between the ideal solid and the ideal liquid, as in the colloidal gel.) But the applications of adsorption and differential diffusion to chemical analysis that will be discussed in this chapter involve the use of solids having a porous or capillary structure, and capillarity plays a part in the processes.

Leonardo da Vinci first recorded observations of capillary phenomena, and as early as 1718 Jurin's Law was enunciated. This law can be expressed as follows:

$$rh = \text{constant},$$

where r is the radius of the capillary and h the height of rise. More recent work has evaluated this constant, leading to the formula:

$$rh = 2\gamma/dg,$$

where γ is the surface tension of the liquid against air, d the density of the liquid, and g the gravitational constant. It is clear that the effort of Nature to reduce interfacial energy underlies capillary phenomena, as it does adsorption.

Inherent Applicability of these Phenomena to Analysis

Selective adsorption, differential diffusion and capillarity are phenomena that are fundamental to many natural and technological processes. When rain soaks the soil and later on the sun and wind aid evaporation at the surface, the salts and other diffusates tend to become sorted out in the vertical profile of the soil according to their diffusibilities and adsorption affinities. Similarly, calcium sulfate in bricks may accumulate toward the surface on weathering, where it forms the familiar white excrescence; and sticks of "chalk" (really cast rods of plaster of paris) often show a hard skin which makes writing on the blackboard difficult until it is broken down. Alexander [2] pointed out that diffusion concentrations occur in the preparation of foods, such as in the crust of bread; and it is often a question whether or not beneficial accumulations of vitamins and salts may exist in other crusts and in the outer sheaths of edible fruits and vegetables. The role of these phenomena in biological processes is too well known to require elaboration here.

Our particular concern in this chapter is the application of these effects in chemical analysis. It has been pointed out that adsorption and diffusion, alone or in combination with capillarity, cause separations of mixtures into their components: this is analysis, in the sense that it is the opposite of synthesis. Moreover, such separations are an essential preliminary to many methods of qualitative and quantitative analysis, to all methods, in fact, that require the isolation of the substance under examination before it is identified or measured. If we have a solution to analyze for A and B, we may, by classical methods, fractionally precipitate as the hydroxides, obtaining for eventual weighing two precipitates, one containing all the A and the other all the B. By perfect analogy, we may alternatively shake the solution first with a selective adsorbent for A, then, after filtering, with one that adsorbs B.

Thomas Graham recognized the analytical importance of diffusion, as is shown

by the opening paragraph of his paper entitled, "Liquid Diffusion Applied to Analysis" [3]:

"The property of volatility, possessed in varying degrees by so many substances, affords invaluable means of separation, as seen in the ever-recurring processes of evaporation and distillation. So similar in character to volatility is the diffusive power possessed by all liquid substances, that we may reckon on a class of analogous resources to arise from it. The range also in the degree of diffusive mobility exhibited by different substances appears to be as wide as the scale of vapor tensions."

Goppelsroeder's "Capillary Analysis"

The use of filter paper as a medium for carrying out analytical reactions is a very old device indeed. Thus lead acetate paper for detecting hydrogen sulfide is described in textbooks published as early as 1850, and in 1861 Schönbein,[4] the discoverer of ozone, proposed the use of starch-potassium iodide paper for detecting that substance. But these employments of paper made only incidental use of the adsorptive and capillary properties of the network of cellulose fibers that is paper.

It was left to Goppelsroeder [5, 6] to devise a method of "capillary analysis" in which these properties were deliberately taken advantage of, and for which the claim [7] was made that it could detect the presence of 10^{-11} gram of methylene blue in 1 cubic centimeter!

Strain [8, p. 7] gives what is considered to be an adequate discussion, for the present purposes, of the Goppelsroeder technique:

"In the same year that Tswett first reported the use of adsorption columns [see section on Spot Tests], Goppelsroeder described a somewhat similar analytical method known as capillary analysis. According to this method, one end of strips of adsorptive paper are placed in solutions of the pigments or other materials to be resolved. As the liquid is drawn into the paper by capillary forces, the substances in solution gradually separate from one another forming a series of bands analogous to those observed in the Tswett columns. The rate of flow of the solvent into the paper strips can be increased by hanging them over the edge of the vessel containing the solution of the mixture. To some extent the bands may be separated from one another by placing the paper strips in a portion of the fresh solvent after some of the solution has been adsorbed.

"In principle the capillary analysis procedure is similar to the columnar technique. However the development of the chromatogram is rather difficult and the preparation of materials in quantity is virtually impossible. As a consequence, capillary analysis has not found extensive use. A recent modification of the Goppelsroeder technique in which the paper alone or paper dusted with adsorptive powders is placed between glass plates [9] is subject to the same limitations as the original method. Various applications of the capillary analysis methods have been ably reviewed by Rheinboldt." [10]

To this it is only fair to add that, although Goppelsroeder did not publish his full-fledged capillary analysis method until 1906, this publication [5] was essentially a summary of his previous work, extending back to 1861. It should also be noted that Goppelsroeder appears to have conceived his work in this field as an outgrowth of that of Schönbein; and that Freundlich [11, pp. 289-90] considered that the method had considerably greater potentialities than have materialized.

Chromatographic Analysis *

Historical. In 1906 the Russian botanist, M. Tswett, working at Warsaw, published the description of a method of resolving mixtures by selective adsorption on

* The task of the writer attempting to summarize the work on chromatography is greatly facilitated by the availability of two recent monographs on the subject, that by Strain [8] and that by Zechmeister and collaborators,[12] as well as an even more recent review by Strain.[13] The reader is referred to these for details and complete bibliographies.

porous columns of suitable materials. This technique was the culmination of a number of years of experimentation on the isolation of plant pigments, and may be generally described as follows:

If a solution is caused to percolate unidirectionally through a closely packed column of powdered adsorbent material, with or without electrophoresis, one or more of the components of the solution may be to some degree retained by the adsorbent; and on subsequent percolation of the pure solvent through the column, the adsorbates may be further separated and concentrated in discrete bands of the column. The substances thus separated, after chemical fixation if desired, may be examined *in situ*, or removed from the adsorbent by means of eluants.

It had, of course, long been known that certain substances, such as charcoal, if shaken with solutions of dyes and the like, remove some of the dissolved materials from solution. Practical applications had been made of this, as in the decolorization of molasses. But in all these cases one effected the separation of the solutes as a whole from the solvent and not from one another. Tswett seems to have been the first to conceive the possibility of making use of the selectivity of adsorption to bring about such desired separations. In the view of the writer, Tswett's fundamental discovery was that this selectivity could be taken advantage of if the flow of the solution through the adsorbent were unidirectional.

Theory. It must be stated that up to the present the chromatographic method has remained on a basis that is largely empirical. If one wishes to separate a new mixture, there are almost no guiding principles of theory to assist in the selection of the absorbent, the solvent or the eluant, and if the mixture has no discernible features of analogy with mixtures previously studied, one must resort to trial and error. There is no reason to doubt, however, that an adequate theory will emerge in time. On the other hand, as will be discussed later, there is gratifying correlation between the structure of certain series of organic compounds and the relative positions of their bands when adsorbed on the Tswett column.

If, however, a proper solvent and a proper adsorbent are assumed for the chromatographic separation of a given mixture, Wilson[14] offers a mathematical theory that accounts at least qualitatively for the experimental results. Wilson treats first the simple case of the adsorption of one solute. He assumes a long column containing M grams of adsorbent per centimeter length, and sets himself the task of formulating the distribution of adsorbed material in the column after a volume, v, of a solution containing solute at an initial concentration c_0 has passed through it. Two simplifying assumptions are made: (1) that equilibrium between adsorbed material and the solution in contact with it is always maintained, and (2) that the volume of the interstices between the particles of adsorbent per unit length of column is negligible.

If Q is the number of millimoles of adsorbed solute per centimeter length of column at a point where the solute concentration is c, an adsorption isotherm is assumed of the form $Q = Mf(c)$. If now the distance from the top of the column to any lower point in it is x, the following differential equation can be derived:

$$\left[\frac{\partial}{\partial v}(Mf(c)\right]_x + \left(\frac{\partial c}{\partial x}\right)_v = 0,$$

whose general solution is,

$$c = \Phi[v - Mxf'(c)],$$

where ϕ is an arbitrary function.

By inserting the appropriate boundary conditions, the following discontinuous solution is arrived at:

$$\text{For } 0 \leqq x < \frac{vc_0}{Mf(c_0)},\ Q = Mf(c_0);$$

$$\text{for } x > \frac{vc_0}{Mf(c_0)},\ Q = 0.$$

Wilson's assertion that "the discontinuous nature of the solution accounts in a satisfactory way for the sharpness of the bands observed in chromatographic experiments," appears reasonable. He also points out that the solution, $Q = Mf(c_0) =$ const., "is in accordance with the experimental observation that the intensity of color is approximately uniform throughout the band." Wilson also considers mathematically the case of n solutes.

In a discussion of Wilson's paper, Meyers [15, pp. 341-2] writes as follows: "When a sufficient volume of solution has been fed to the column to extend $vc_o/Mf(c_o)$, the length of the band, to l, the length of the column, 'breakthrough' occurs. Since breakthrough is expressed as grams adsorbate per gram adsorbent, and since

$$vc_0/lM = f(c') = \frac{\text{gms adsorbate}}{\text{gms adsorbent}} = \text{capacity},$$

a knowledge of adsorption isotherms should permit a calculation of breakthrough capacity of columns of adsorbents when breakthrough occurs."

Wilson's work is a very promising beginning on a comprehensive theory of chromatography, and it is to be hoped that its development will be carried forward.

Technique. The basic technique of chromatography is extremely simple and straightforward. It consists of the following steps:

(1) Formation of the chromatogram, *i.e.*, percolation of the solution through the column of adsorbent.
(2) Development of the chromatogram, *i.e.*, percolation of fresh solvent to widen and separate the bands.
(3) Elution of the substances separated by extracting the several sections of the column corresponding to the bands, with polar solvents.

Since the writer does not intend to give laboratory directions for chromatography, it will suffice to state that a number of variables and secondary factors have been recognized and studied, in addition to the chemical nature of the adsorbent, the solvent, and the eluant. Among these are: firmness of binding of adsorbates to adsorbent; particle size and porosity of adsorbent; avoidance of exposure of column to air in certain cases; poisoning of adsorbent; effect of one adsorbed compound on another; size and shape of adsorption columns; preparation of adsorption columns; concentration of solutions; rate of percolation; and techniques of development and elution.

Among the variations in the basic technique that have been published is the drawing of the solution from the bottom of the column upward, and the location of the boundaries between bands in the liquid emerging from the top of column by the use of the Toepler schlieren effect.[25] Other workers have eluted the successive bands without removing the column from the tube, by washing with fresh portions of solvent until each compound was carried in succession into a separately collected percolate. The modifications are useful when working with colorless compounds whose adsorption boundaries are difficult to locate. Ultraviolet fluorescence, and the brushing on of a thin line of a reagent that forms colored products with the adsorbates, have also been used to locate the boundaries.

Adsorbents. The outstanding fact about chromatographic adsorbents is that no

relation has been discovered between their chemical structure or composition and their adsorption capacities. Not only is there no approach to a universal adsorbent, but none has been found that is perfect even for a restricted application.

The ideal chromatographic adsorbent for a given mixture would have these characteristics, possibly among others:

(1) High adsorption capacity for all the solutes, yet not so high as to make development or elution difficult.
(2) Sufficiently different specific adsorption capacities for the several solutes to permit their separation by development of the chromatogram.
(3) Absence of capacity to cause chemical changes in adsorbate, solvent or eluant.
(4) Non-porous particles of uniform and suitable size.
(5) Whiteness or absence of color.

Since it is impossible to write a specification for an adsorbent as we would for a chemical reagent, in terms of desired chemical and other characteristics, the best we can do is to require that successive lots of a given adsorbent be prepared by the same process from the same raw material. Magnesia or alumina or soda ash prepared in different ways often show gross differences in performance in the Tswett column. This fact renders much of the older literature on chromatographic adsorbents difficult of interpretation. Tables 1 and 2 show the applications and relative adsorption capacities of some of the commonly used adsorbents.

Table 1. Adsorbents and Solvents Useful for Chromatographic Studies of Various Substances (Strain[8, p. 49])

Substance	Adsorbent	Solvent
Inorganic ions	Alumina	Water
" "	8-hydroxyquinoline	"
Enzymes	Bauxite	"
Sugars	"	"
"	Bone Ash	"
Flavins	Frankonit	"
Anthocyanins	Alumina	"
Alkaloids	"	" or benzene
Amino acids	Titania	Water
Chlorophylls	Saccharides	Pet. ether & benzene
"	Magnesium citrate	Pet. ether & ether
Chlorophyll derivatives	Talc or saccharides	Benzene, ether or alcohols
Nitro compounds	Talc	Benzene
Organic acids	"	"
Phenols	Lime	Alcohols
"	Alumina	Benzene
Ethers, ketones, esters	Alumina	Pet. ether & benzene
Xanthophylls	"	Benzene
"	Magnesia	1, 2-dichloroethane
"	Alkaline earth carbonates	Pet. ether, benzene or carbon disulfide
Aliphatic alcohols and sterols	Alumina	Pet. ether & benzene
Unsat'd hydrocarbons	Lime, alumina or magnesia	Pet. ether, carbon tetrachloride or benzene
Sat'd hydrocarbons	Alumina or magnesia	Pet. ether

It was stated that adsorbents should not interact chemically with the adsorbates. There are, however, modifications of the basic chromatographic technique in which chemical reaction seems to be necessary to effect the separation. Thus, separations of inorganic ions by adsorption on 8-hydroxyquinoline involve reversible chemical reactions with the adsorbent, fractionation of isotopes on permutit is accompanied by ion-exchange reactions, and the adsorption of certain dyes produces color changes indicative of complex formation.

Table 2. Adsorbents Arranged in Approximate Order of Adsorption Capacity (Strain [8, p. 50])

Sucrose, starch
Inulin
Magnesium citrate
Talc
Sodium carbonate
Potassium carbonate
Calcium phosphate
Magnesium carbonate
Magnesia (Merck)
Lime (fresh and partly slaked)
Activated silicic acid
Activated magnesium silicates
Activated alumina, charcoal and magnesia (Micron Brand)
Fuller's earths

Solvents. The choice of a solvent for a mixture to be resolved chromatographically is guided by the following considerations:

(1) The mixture should have sufficient solubility in the solvent.
(2) The adsorbent to be used must be capable of extracting the solutes from their solution rapidly and completely. Since the process of forming a chromatogram is a competition between the adsorption capacity of the adsorbent and the solvent power of the solvent, highly polar solvents are in general to be avoided.
(3) There must be available an eluant of higher solvent power.
(4) The solvent should not react chemically with the adsorbent, the solutes or the eluant and should be stable.

Table 3, taken from Strain,[8, p. 66] lists the important solvents according to the readiness with which they give up their solutes to adsorbents.

Table 3. Solvents Arranged in Approximate Ascending Order of Stability of Their Solutions to Adsorbents

Petroleum ether, b.p., 30-50°
" " " " 50-70°
" " " " 70-100°
Carbon tetrachloride
Cyclohexane
Carbon disulfide
Ether (anhydrous, alcohol-free)
Acetone (anhydrous, alcohol-free)
Benzene
Toluene
Esters of organic acids
1, 2-Dichloroethane, chloroform, dichloromethane
Alcohols
Water (variations with changes in pH and salt conc.)
Pyridine
Organic acids
Mixtures of acids or bases with water, alcohol or pyridine

Eluants. From what has been said about solvents, it will be clear that an eluant must be so chosen that in the competition between it and the adsorbent, in the elution process, the eluant must prevail, rapidly and completely. Thus, liquids arranged, as are those in Table 3, in descending order of usefulness as chromatographic solvents, are automatically arranged in *ascending* order of efficiency as eluants. Polar solvents are generally the best eluants.

It will be obvious from the discussion of adsorbents, solvents and eluants, that the choice of any one is dependent on the other two, and that the roles of solvent and eluant may interchange from one experiment to another.

Electrophoresis. In the usual application of electrophoresis, charged particles in a liquid are caused to migrate toward one or the other of a pair of electrodes, by the application of an EMF. Since particles chemically different move in general at different rates, this technique can be used to effect separations. A disadvantage is that convection currents exist in liquids and these tend to reduce the sharpness of the boundaries and the cleanness of separation. A combination of electrophoresis with chromatography has been used which in some cases produces better results than does either technique alone. A mixture of solutes that is adsorbable as a whole but whose components are difficult or tedious to separate by unaided chromatography can sometimes be more rapidly resolved, after adsorption on a Tswett column, by the insertion of electrodes at the top and bottom of the column and the application of an EMF. Convection currents are minimized and the boundaries remain sharp.

What chromatography can do. This technique has for its primary objective the resolution of mixtures. Such resolutions have been carried out for the following purposes:

(1) Qualitative analysis.
(2) Quantitative analysis.
(3) Determination of molecular structure.
(4) Testing substances for chemical homogeneity.
(5) Establishing identity or non-identity of substances.
(6) Purification of substances.
(7) Concentration of materials from dilute solutions.
(8) Control of technical products.
(9) Regeneration of substances from complex addition compounds.

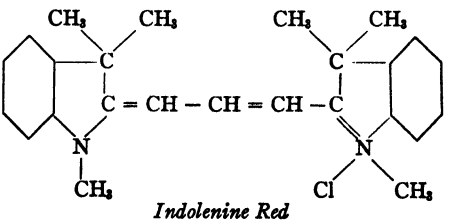
There will be found in Zechmeister's [12] and Strain's [8] books full details and many examples of these nine types of problems for which chromatography has been found useful. Perhaps the most interesting application, from the viewpoint of theoretical chemistry, is the determination of molecular structure of organic compounds, some examples of which may properly be reviewed here.

Generally speaking, adsorbability of organic compounds increases with the number of polar groups in the molecule. A classical example is the series of synthetic polyenes,[16]

$$C_6H_5 - (CH = CH)_n - C_6H_5$$

To summarize the experimental results, the larger n, the stronger the adsorption, and hence the higher in the column is the adsorption band.

The same order of adsorption is obtained if the phenyl groups in these polyenes are replaced by certain heterocyclic groups, to form, for example, the indolenine series of dyes:[17]



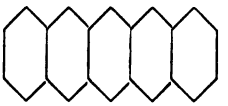
Indolenine Red

Extensive chromatographic investigations have been made of the carotenoid pigments, and much valuable information on structure has been obtained. These substances are yellow to violet fat-soluble pigments. The results illustrate the fact that, with complex organic molecules, not only the double bonds but other polar groups, such as hydroxyl and carboxyl, affect the order of adsorption, so that interpretation of the chromatograms in terms of structure may be very difficult. In this series of

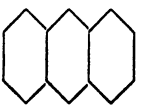
compounds it was found that the adsorption affinity increases, with compounds otherwise analogous, when [18]

(a) The number of conjugated double bonds increases.
(b) Of a given number of double bonds, all are conjugated.
(c) OH groups occur in unsaturated systems otherwise identical.
(d) OH groups occur in unsaturated systems otherwise identical.
(e) The number of OH groups increases.
(f) A carboxyl group is conjugated with a carbon double bond.

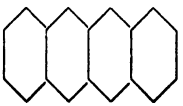
Table 4. Adsorption Series of Some Linear Polycyclic Compounds, in Descending Order of Strength of Adsorption.



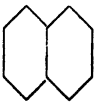
2, 3, 6, 7-Dibenzanthracene
(uppermost band)



Anthracene



Naphthacene

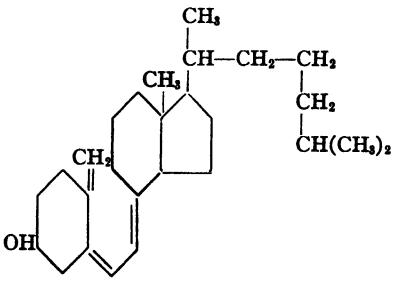


Naphthalene

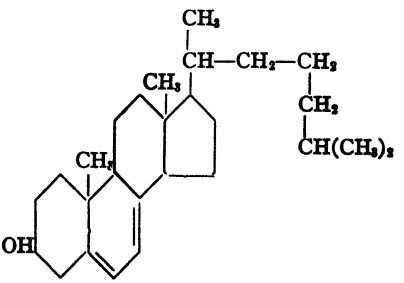


Benzene
(lowest band)

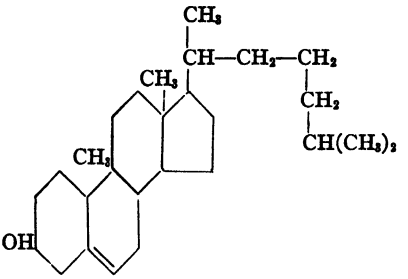
Table 5. Adsorption Series of Some Sterols.



Vitamin D_3

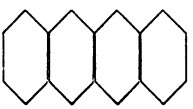


7-Dehydrocholesterol

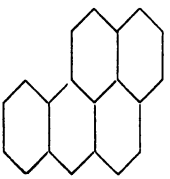


Cholesterol

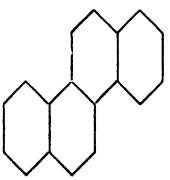
Table 6. Adsorption Series of Some Polycyclic Compounds.



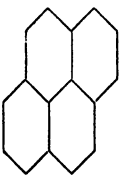
Naphthacene



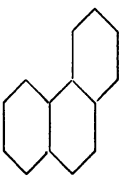
1, 2-Benzpyrene



Chrysene



Pyrene



Phenanthrene

As Strain[8, p. 15] puts it: "In homologous series of organic compounds, the adsorbability increases with an increase in the number of double bonds. . . . Among compounds of the same type, adsorbability also increases with an increase in unsaturation. . . . Among isomeric compounds such as the polycyclic hydrocarbons . . . there is no simple relation between the number of double bonds and adsorbability. The shape and polarity of the molecule, as indicated by color, appear to exert greater influences than the number of double bonds alone." These relationships are illustrated in Tables 4, 5 and 6.

Recent Literature. There will be found in the books by Zechmeister[12] and Strain,[8] and in the review paper by Strain,[13] a complete bibliography of the chromatographic technique through 1941. The following additional papers have become available during 1942:

Schuler and Yang,[26] using as adsorbents alumina, calcium carbonate, and mag-

nesia, separated certain isomeric nitro-compounds, nitration products, and homologous quinones.

Tiselius [27, 28] continued his work using the Toepler schlieren effect to locate boundaries, and applied the method to glucose and lactose solutions. He reports that the adsorption of saccharides from aqueous solutions by carbon is governed chiefly by the size of the molecule.

"Chromatographic Analysis in Reverse" was described by Lowman.[29] In this, a tube is mounted vertically and its bottom closed with a cork. Pure solvent is poured in, on top of which is poured the solution. A funnel dips several centimeters into the solution. The powdered adsorbent is added in successive small amounts. The compounds in the mixture are carried to the bottom in the order of their adsorbability. When the adsorption is complete more powder is added, and the column is pushed out of the tube, as in usual chromatographic analysis.

Bretschneider [30] separated androstenedione and progesterone from sterol oxidation products by adsorption on alumina from benzene-petrolic ether solution. He also separated cholestanone and cholestenone by adsorption from petrolic ether on silica.

Working in the inorganic field, Venturello and Agliardi,[31] using alumina as adsorbent, determined the following order of adsorption from aqueous solutions:

Ca, Pd, Pt (strongest adsorption)
Au
(Mo)
As
(Bi, Ga, Os, Th, ZrO)
Co, Cr, Cu, Fe, Hg, Ni, Pb, UO_2, V
Ag
Be, Cd, Mg, Mn, Tl, Zn (weakest adsorption)

Bielenberg and Goldhahn [32] determined pyrocatechol, resorcinol and phlorglucinol in the presence of each other in aqueous solution by treating with ferric chloride followed by deposition on alumina. Better results were obtained when they used the azo dyes of these phenolics in benzene-isoamyl alcohol solution.

Erlenmeyer and Schmidlin,[33] using the following adsorbent system:

cotton pad
starch plus 5-oxo-4-oximino-3-phenylisoxazoline
cotton pad
starch plus violuric acid
cotton pad,

obtained good separations, from the aqueous solutions of their acetates, of potassium (yellow zone) and sodium (red zone).

Erlenmeyer and Schönauer,[34] using violuric acid alone or with starch, barium sulfate or diatomaceous earth, as adsorbent, separated potassium, calcium, magnesium and sodium as acetates in water solution. Good separations were obtained with potassium-sodium, ammonium-sodium, potassium-magnesium, magnesium-sodium, ammonium-magnesium, barium-calcium, strontium-calcium, cupric-mercuric, plumbic-mercuric and cupric-plumbic. The results with potassium-ammonium and barium-strontium were poor.

Crowe [35] described a "Micromethod of Chromatographic Analysis." A Petrie dish was fixed in a tilted position, and a layer of adsorbent powder shaken down to cover part of the lower half of the bottom, in the form of a wedge very thin at the top and a few millimeters thick at the bottom. The solution was flowed drop-wise into the thin top of the adsorbent wedge, and broad zones of separated material were noted. This method is claimed to be very useful in the analysis of biologic substances where only small quantities are available, and was found to be more satisfactory than other micromethods in the analysis of certain biologic substances.

The Moving Boundary Method

This is a method that makes use of differential diffusion of ionized substances in an electric field, to effect separations. It was first developed by Lodge[19] as a means of measuring directly transference numbers and ionic migration velocities. "Lodge's apparatus consisted of an horizontal glass tube 40 cm long, bent vertical at each end so as to dip into beakers. A solution of sodium chloride containing a small amount of phenolphthalein and slightly alkaline with sodium hydroxide, was made into a jelly with gelatin, and the mixture allowed to set in the glass tube. In one beaker a platinum cathode was placed and a solution of cupric chloride poured in. The anode was of carbon and the anode liquid hydrochloric acid. The apparatus was allowed to stand for twenty-five hours to measure the effect of diffusion. The progress of the hydrochloric acid was rendered visible by the decolorization of the pink gelatin, that of the copper chloride by the bluish turbidity due to the interaction of the copper salt with the small amount of alkali present, forming copper hydroxide. The following day the speed of diffusion was noted, and a potential difference of 40 volts was impressed upon the electrodes. The decolorization now proceeded from both ends at unequal rates, the hydrogen ion reacting with the indicator directly, and the chlorine discharging the color by combining with the sodium ion of the alkali."[20]

James Kendall and students[20, 21, 22, 28] were the first to make extensive use of this technique for effecting separations. In general, different ions move with different velocities under a potential gradient. Thus, if a solution containing a mixture of two electrolytes having the same anion but different cations is placed in a long horizontal tube and electrolyzed, the two cations would be expected to move toward the cathode at different rates. If remixing by ordinary diffusion can be minimized, a separation of the two cations is possible. These workers incorporated their solutions into agar-agar gels to obtain rigidity and to reduce ordinary diffusion. It was observed that, while diffusion in these gels was indeed greatly reduced over that in water, the colloid had negligible effect on the speeds of electrolytic migration of the ions. They considered the agar inert with respect to their electrolytic process, but in the light of later knowledge about selective adsorption on gels it seems possible that some of their anomalous results might be explained by adsorption effects.

The Kendall method used a modified Lodge apparatus. The mixture to be separated was in a narrow section of agar gel formed in the tube near the opposite end of the apparatus from that to which the ions would travel under electrolysis. Behind this was gel containing an ion inherently slower than those to be separated, and in front of the sample, gel containing a faster ion. Upon electrolysis, if the concentrations had been properly adjusted, the ions in the sample section moved slowly toward the appropriate electrode, the section remaining constant in length and the boundaries sharp. At the end of the run the portion of the gel containing the sample was removed from the tube and sliced into sections, which were examined for evidence of separation. Or the gel containing the mixture could be run through the apparatus one or more additional times if this was considered necessary to obtain a separation. In some experiments the total length of run was as much as 100 feet.

Among the mixtures that were separated more or less completely by this method were iodide and thiocyanate, barium and calcium, barium and strontium, chloride and iodide, yttrium and erbium, gadolinium and samarium, praseodymium and neodymium, radium, barium and mesothorium, and certain isotopic mixtures.

As was pointed out, the role of adsorption in the Kendall form of the moving-boundary method was generally negligible, in contrast to electrophoretic chromatography, in which it is an important factor. Possibly these two techniques might be profitably combined for inorganic separations. It would appear that if, in the moving-boundary method, some substance were used in place of the agar that selectively adsorbed one of the ions, this would in effect widen the difference between the mi-

gration velocities and facilitate a separation. Lange and Nagel,[24] in a short note published in 1936, proposed that the chromatographic method be tried for separating mixtures of rare-earth salts. There is nothing in the literature to indicate that these authors or others have acted on this proposal. The present writer intends to attack this problem by a combination of the two techniques as indicated.

Electrography

In this technique the adsorbent properties of cellulose fibers in filter paper are utilized to obtain distributive patterns or "chemical maps" of the distribution of certain materials on the surfaces of conducting specimens. A square of filter paper is impregnated with an electrolyte (such as sodium nitrate or chloride) and a reagent for the metals to be detected (such as antimony sulfide for copper, silver, etc.), or with a compound like sodium carbonate that serves both purposes. This prepared paper, moistened, is sandwiched between the specimen surface and a plate of inert metal, and the assembly pressed firmly together. A small current is now passed between the two metals, the specimen surface being made anodic. Ions from the specimen are electrolyzed into the paper, where they are immediately fixed as reaction products which are adsorbed on the cellulose fibers. Lateral diffusion or "bleeding" is minimized, so that the distributive pattern of the reaction products on the print corresponds to that of the metals tested for on the specimen surface. If desired, the print can be further "developed" by immersion in reagent solutions that form new reaction products of greater color contrast or permanence.

This method has been used extensively in the writer's laboratory for such purposes as rapid identifications, study of metallic inclusions in surfaces, testing plated surfaces for porosity, etc.[36] Its great advantage, besides those of simplicity and rapidity, is that one is enabled to determine not only what is present on the surface but how it is distributed.

The electrographic method was published simultaneously and independently in 1929 by Glazunov and Fritz. The interested reader is referred to the excellent review paper by Jirkovsky,[37] and to the paper by Yagoda [38] which not only reports some of that author's original work on biological material but contains a complete and up-to-date bibliography.*

Spot Tests

The well-known Feigl spot-testing technique may be carried out either with or without the aid of adsorbent media like filter paper. When paper is used, the sample drop is placed thereon, where it is allowed to spread uniformly by capillarity until the surface forces are balanced. The wet area is then treated with a drop of reagent, and an identifiable colored product is produced. Feigl's book [39] discusses the theory and practice of these tests, and gives complete directions for carrying out several hundred of them.

Clarke and Hermance [40] contributed two new ideas to this technique. By the use of a specially designed capillary burette, the rate of flow of the sample solution into the impregnated paper was controllable. With this it was found possible to restrict to a minimum the area covered by the reaction product, by providing that the rate of flow did not exceed the rate of reaction.

This increased the sensitivity of the tests, because sensitivity is determined by the smallest quantity of reaction product per unit paper area that the observer can just detect. These workers pointed out, however, that "if the precipitating agent

* The present writer, in collaboration with H. W. Hermance, has in preparation a monograph tentatively entitled, "The Application of Impregnated Papers to Microanalysis," to be published by Interscience Publishers, Inc., New York. In this will appear a full presentation of the authors' original work in electrography as well as in other techniques employing adsorbent media in analytical reactions.

contained in the paper is soluble in the solution under test, effective, uniform fixation of the reaction product is often delayed or entirely prevented. As the solution flows into paper so impregnated, the first portions react with the soluble reagent, producing a small quantity of colored product. The clear liquid, its reactive ion or ions thus removed, spreads outward carrying with it much if not all of the impregnant, depleting the area around the point of entrance. Hence the next portions of solution will remain unprecipitated until they are brought into contact with the reagent, concentrated by evaporation, at the periphery of the wet area. A further disadvantage of many soluble impregnants is their instability. Many soluble alkali salts decompose rather easily in paper, while less soluble salts of other metals keep reasonably well." [40]

These considerations led to the use of reagents for paper impregnation that are slightly rather than readily soluble, *e.g.*, cadmium or antimony sulfide instead of sodium sulfide. The concentration of the reactive ion can be regulated automatically by proper selection of the impregnating salt, and the specificity of the test greatly improved by thus restricting the number of possible reactions.

Using their modified spot-testing technique, Clarke and Hermance were able to obtain hundredfold increases in sensitivity in some cases. They also effected differential precipitation of two or more solutes, in the form of concentric rings about the point of entrance of the drop into the reagent paper, the rings being analogous to the bands of a chromatogram.

A Method of Applying Reagent Papers to Large Volumes of Solutions

The spot test deals with drops of solutions. But there arises in microanalytical laboratories the problem of separating from much larger volumes of solutions metallic ions present in low concentrations. Clarke and Hermance [41] describe a method for achieving this by the use of their principle of low-solubility reagents. Their method is to percolate a sufficient volume of the solution at a slow rate through a flange assembly that holds a disc of impregnated paper across the liquid stream. Three fundamental requirements must be met:

(1) The impregnant must be sufficiently insoluble to withstand the action of the large volume of solution flowing through a small area of the paper.
(2) Sufficient difference must exist between the solubilities of reagent and reaction product to ensure reasonably quantitative precipitation of the ion sought.
(3) The reaction product must be capable of immediate fixation on the fibers of the paper.

A theoretical analysis of this method indicates that it should be capable of very complete separations. Thus it is calculated that, when separating traces of copper from a solution by percolation through cadmium sulfide-impregnated paper, even though the rate of flow is such as to allow the solution flowing through the disc to attain only 1 per cent saturation with cadmium sulfide, the copper remaining unprecipitated will be only 8.3×10^{-12} microgram per liter.

Future of These Methods of Analysis

Forces of diffusion and adsorption are active in most of the operations of chemical analysis. Sometimes they give rise to unwanted phenomena, such as occlusion. The techniques described in this paper have developed from the premise that, since these forces exist and cannot be avoided, they should be deliberately harnessed to useful purposes. One can say, in general, that the more we learn about the fundamental nature of diffusion and adsorption, the wider and more varied will be the analytical applications of these phenomena. But there is very much of these fundamentals yet to be learned. Paper, for instance, has proved to be a most useful and versatile adsorbent, but the mechanism by which it retains substances is by no means

clear. Freundlich [11, pp. 288-9] states that adsorption on cellulose is a pure ion-exchange reaction, and concludes that it is not the cellulose itself but rather the impurities like calcium salts that determine adsorption. This view is not accepted in all quarters, however. Comparatively little study has been made of other media than paper (except, of course, in chromatography). Undoubtedly such studies will receive emphasis in the future. The writer feels that an important future trend in these techniques will be the increasing employment of capillarity and electro-capillarity.

Finally, many other applications of these phenomena to analysis can and will be made, even with the theory in its present state. These developments are retarded by the fact that each is a problem in laboratory manipulation. Such problems cannot always be solved by systematic attack, but must await something akin to inspiration.

References

1. "Colloid Chemistry, Theoretical and Applied," Vol. I, p. 246, Ed. J. Alexander, Chemical Catalog Co. (Reinhold Publishing Corp.), 1926.
2. Alexander, J., *Science*, **77**, 489 (1933); see also *J. Am. Chem. Soc.*, **39**, 84–88 (1917), paper entitled "Selective Adsorption and Differential Diffusion."
3. Graham, T., "Liquid Diffusion Applied to Analysis," *Phil. Trans. Roy, Soc., London*, **151**, 183–224 (1861). Also Liebig's Annalen **121**, 1–77 (1862).
4. Schönbein, C. F., *Verh. d. Naturforscher Ges. i. Basel, III Teil*, 249 (1861).
5. Goppelsroeder, F., "Anregung zum Studium der Capillaranalyse," Basel, 1906.
6. Goppelsroeder, F., "Kapillaranalyse," Theo. Steinkopff, Dresden, 1910.
7. Goppelsroeder, F., *Kolloidzeit.*, **4**, 107 (1909).
8. Strain, H. H., "Chromatographic Adsorption Analysis," Interscience Pub. Co., N. Y., 1942.
9. Brown, W. G., *Nature*, **143**, 377 (1939).
10. Rheinboldt, H., in Houben's "Die Methoden der organischen Chemie," 3rd Ed., Vol. I, p. 291, Georg Thieme, Leipsig, 1925.
11. Freundlich, H., "Kapillarchemie," 3rd Ed., Akademische Verlagsgesellschaft, Leipsig, 1923.
12. "Principles and Practice of Chromatography," by L. Zechmeister and L. v. Cholnoky, trans. by Bachrach and Robinson, John Wiley & Sons, N. Y., 1941.
13. Strain, H. H., *Ind. Eng. Chem., Anal. Ed.*, **14**, 245 (1942).
14. Wilson, J. N., *J. Am. Chem. Soc.*, **62**, 1583 (1940).
15. "Advances in Colloid Science," Ed. E. O. Kraemer, Vol. I, Interscience Publishers, N. Y., 1942.
16. Winterstein, A., and Schön, K., *Z. physiol. Chem.*, **230**, 146 (1934).
17. Ruggli, P., and Jensen, P., *Helv. Chim. Acta*, **18**, 624 (1935).
18. Winterstein, A., in "Handbuch der Pflanzenanalyse," Vol. IV, pp. 1403–37, Vienna, Julius Springer, 1933.
19. Lodge, O., *Brit. Assoc. Reports*, 389 (1886).
20. Clarke, B. L., Dissertation, Columbia Univ., 1924.
21. White, J. F., Dissertation, Columbia Univ., 1924.
22. Kendall, J., and Crittenden, E. D., *Proc. Natl. Acad. Sci.*, **9**, 75 (1923).
23. Kendall, J., and Clarke, B. L., *ibid.*, **11**, 393 (1925).
24. Lange, E., and Nagel, K., *Z. Elektrochem.*, **42**, 210 (1936).
25. Tiselius, A., *Science*, **94**, 145 (1941).
26. Schuler, L., and H.-H. Yang, *J. Chinese Chem. Soc.*, **5**, 239 (1937) (Pub. 1940).
27. Tiselius, A., *Arkiv. Kemi, Mineral., Geol.*, No. 22 (1940).
28. Tiselius, A., *ibid.*, No. 32 (1941).
29. Lowman, A., *Science*, **96**, 211 (1942).
30. Bretschneider, H., *Monatsh.*, **74**, 53 (1941).
31. Venturello, G., and Agliardi, N., *Ann. chim. applicata*, **30**, 224, 230 (1940).
32. Bielenberg, W., and Goldhahn, H., *Brennstoff-Chem.*, **21**, 236 (1940).
33. Erlenmeyer, H. and Schmidlin, J., *Helv. Chim. Acta*, **24**, 1213 (1941).
34. Erlenmeyer, H., and Schönauer, W., *ibid.*, **24**, 878 (1941).
35. Crowe, M. O'L., *Ind. Eng. Chem., Anal. Ed.*, **13**, 845 (1941).
36. Hermance, H. W., *Bell Labs. Record*, **18**, 269 (1940).
37. Jirkovsky, R., *Mikrochemie*, **15**, 331 (1934).
38. Yagoda, H., *Ind. Eng. Chem., Anal. Ed.*, **15**, 135 (1943).
39. "Spot Tests," by F. Feigl, trans. by J. Matthews from 3rd German Ed., Nordemann Pub. Co., N. Y., 1939.
40. Clarke, B. L., and Hermance, H. W., *Ind. Eng. Chem., Anal. Ed.*, **9**, 292 (1937).
41. Clarke B. L., and Hermance, H. W., *ibid.*, **10**, 591 (1938).

Colloid Chemical Aspects of Photographic Development

S. E. SHEPPARD

Research Laboratory, Eastman Kodak Co., Rochester, N. Y.

There can be no such thing as a colloid chemical theory of photographic development. Rather do we meet in the varieties of this process certain colloids or colloidal systems, and the particular chemico-physical characteristics thereof. Such are modifications of the properties of solids produced by extremely fine subdivision, the associated effects of increased surfaces and interfaces, processes of swelling and membrane potential formation, the rôle of phase boundary potentials, both as rigid Helmholtz double layers, and as diffuse double layers (ψ- and ζ-potentials) and the effects of these upon reaction velocities. The essence of photographic development seems to be that a "latent image," an invisible impression of varying amounts of light upon a material substrate, should be made visible by a subsequent physical or chemical process. The French term "révélation" is just the word for this perception of the matter. It is qualitatively adequate, but quantitatively incomplete. The fact is, that quantitatively, development is essentially what we would describe in these days of radio technics as amplification. An extremely minute chemical change is amplified or magnified many times by a subsequent process. Also, to an increasing extent with the degree of amplification, the gain in power is made at the expense of exactness in rendering of definition, or minute representation. In colloid chemical terms, development consists in reducing the dispersity, or increasing the particle size, of the so-called "latent image," at any rate for all processes with silver salts. The application of the term "development" to processes with bichromated colloids is practically justified, but technically inexact; what happens is more nearly analogous to "fixation" in photographic processes with silver salts, that is the removal of material not affected by light.

It is not necessary that we should deal in advance with the formation of the latent image in, or mostly on, the silver halides. It must suffice to say that it is generally agreed that this consists chemically of ultra-ultra-minute specks of metallic silver on, and to some extent in, the grains of silver halide. In gelatino-silver halide emulsions, it appears fairly certain that these latent image nuclei are formed upon preëxisting "sensitivity specks" of silver sulfide. We shall return to this in connection with gelatino-silver bromide emulsions, but let us first glance at that prototype of photography termed after its originator, daguerreotype. Was daguerreotype a discovery or an invention? Certainly chance, "sa sacrée majesté, l'hasard" counted for much in its inception, as Frederick II of Prussia said of other things. That Daguerre experimented with silver plates fumed with iodine was due probably to his association with Nicèphore de Niépce and the latter's more consistent and persistent efforts to invent a process of photo-mechanical reproduction. But what a chance that Daguerre should have put away an insufficiently exposed plate in a darkened cupboard with some mercury! Thus happened the "discovery" of the latent image by its development.

What occurs in daguerreotype is as follows. The silver iodide film is decomposed by light to give silver and iodine, the latter being taken over by the under-

lying silver plate to form more silver iodide. With short exposures to light, the amount of silver is too minute to be visible. However, the vapor of mercury condenses upon and magnifies these silver embryos, producing a visible image. It was appropriate in this, so to say, alchemical forerunner of photography, that "ennobling" was effected by a further deposition of metallic gold upon the developed image. In contrast to this alchemical history or mystery, is the fact that it offers an interesting example of Langmuir's [1] fundamental theory of adsorption from the gaseous state. The mercury atoms which collide with unchanged silver iodide have "no abiding city"—the collisions are quasi-elastic and they return to their gassy chaos. But those that impinge upon the metallic silver particles of the latent image are held, and hold more of later comers, until a perceptible image is formed. How large must be an original silver nucleus: (i) to maintain itself; (ii) to hold incident mercury atoms? In the related case of sputtered metal (*e.g.*, silver) atoms upon a glass plate *in vacuo* Reinders and Hamburger [2] have argued for a nucleus of three or four silver atoms. The calculations of H. S. Taylor, H. Eyring, and A. Sherman [3] on "The Binding Energies in the Growth of Crystal Nuclei from Metallic Atoms" rather indicate that two atoms within a certain interval (a primitive twin) should be sufficient. We do not know whether the answer "by Gemini" is correct, and the problem as to the smallest size of a latent image nucleus has carried over into the modern theory of the gelatino-silver halides.[4] The "heavenly twins" may even be a single silver atom adsorbed to a silver ion—an immobilized atom as proposed by E. R. Bullock.[5] At any rate daguerreotype affords this interesting connection with a fundamental colloid chemical problem, that of adsorption from the gaseous state.

In "wet plate" collodion, the silver support of daguerrotype was replaced by collodion on glass, and mercury vapor development by "physical" development with silver nitrate solution and a reducer. The silver nitrate came from the sensitizing solution for the iodized collodion, and the reducer was either pyrogallic acid (acidified with acetic or citric acid) or ferrous sulfate, slightly acidified.

It is probable that silver ions are adsorbed to the silver of the latent image and thus reduced *in situ* by the reducer. A number of recent investigations [6] of the reduction of silver ions (in solution) indicate the adsorption of silver ions to an effective solid surface (*e.g.* gold, silver sulfide) as an initial step, followed by adsorption to silver metal as soon as available. In consequence, these reactions show marked catalysis by silver nuclei, and it is this that the development of the latent image illustrates. In "wet plate" collodion, and again in collodion dry plates, an essentially colloid * medium appears, viz. collodion. This is an incompletely nitrated cellulose very soluble in mixtures of about 50:50 ethyl alcohol and ethyl ether. These solutions on evaporation give clear transparent films of great strength and flexibility. They are insoluble in water, but can be made permeable to water and aqueous solutions.

This last property is very important for photographic use since it permits access or penetration of developers and fixing agents. Permeability would continue more or less so long as the film was still swollen with residual solvent, but on loss of this the collodion became "horny" and impermeable. Empirically, this was lessened by introduction into the solution or dope of various substances, generally hygroscopic, termed "preservatives." Examples are salts, such as magnesium nitrate, also honey, glucose and—as transitional to the gelatino-bromide dry plate—albumen. Some part of this technique overlaps with that employed in the preparation of collodion membranes of varying porosity, as for ultrafiltration.[7]

The reactions of silver ions in solution with various reducers have been the subject of recent important studies by T. H. James.[6] His investigations have brought out the important rôle of colloid silver in affording an adsorption site for silver ions,

* From the Greek κολλα, glue.

with further attraction of reducing ions followed by reduction *in situ*. In some cases, *e.g.*, hydroxylamine, *p*-phenylene-diamine, there may be more or less simultaneous adsorption of the reducer to the colloid silver.

The introduction of dry collodion emulsions of silver bromide, together with the application of alkaline organic developers by Russell [8] paved the way for the replacement of collodion by gelatin.[9]

We shall defer for later notice the specific colloid chemical behavior of gelatin in relation to development. For the moment it must suffice that its ability to swell in water and aqueous solutions while remaining an elastic gel, allows a variety of reactions to occur between dissolved reagents and the solid silver halide. As to this, in *negative* emulsions, we have to deal with a suspension of silver bromide crystallites, containing some two to five per cent silver iodide, and ranging in size from *ca* 0.3μ to 3μ. In *positive* emulsions we have more generally crystallites of silver chloride or chloro-bromide, of a greater dispersity, *ca* 0.05 to 0.3μ, or in some cases of silver bromide alone, of comparable dimensions. It is not, however, with the practice of photography that we have to deal, but with certain aspects of the theory of one phase of it.

After the formation of a latent image there is employed today almost exclusively an alkaline organic developing solution. For the chemical constitution of these developers reference must be made elsewhere.[10] Shortly after the introduction of gelatin plates, there was considerable use of Carey Lea's ferrous oxalate developer. This, together with certain other "inorganic" developers is now only of scientific interest. Their great importance for the theory of "chemical" development, as the direct reduction of silver halide grains is termed, in contrast to "physical" development employing solutions of silver salts, is that they have practically no greater solvent power than water itself. The alkaline organic developers in practice contain quantities of sodium sulfite, from 5 per cent upward, which can dissolve silver bromide to an extent increasing with concentration. However, except in the higher concentrations, this solvent power is of secondary importance compared with its intervention in the purely chemical aspect of development as a preservative. Its other potentiality as a silver halide solvent will be dealt with later. We may consider, therefore, that in normal chemical development the activity of silver ions in solution is very small, as instanced by the following figures * for 25° C.

Water	pAg = 6.08	$S = (8.3 \times 10^{-7})^2$
0.001 N KBr	" = 9.16	and $C_{Ag^+} \times C_{Br^-} =$ constant
.01 N KBr	" = 10.16	

$pAg = -\log C_{Ag^+}$, where C_{Ag^+} = concentration of silver ions. Hence pAg = 6.08 indicates a concentration of silver ions = 8.3×10^{-7}, which is the solubility of the AgBr in water at 25°C. S = solubility product. This equals $C_{Ag^+} \times C_{Br^-}$ = constant.

Photographic emulsions usually contain sufficient excess of bromide ions to be considered, when swollen with water, as at about pAg 8. Further, many photographic developers have sufficient bromide included to keep the concentration of free silver ions at about this figure. It will be seen that the free silver ion concentration must be extremely low. None the less, early hypotheses as to the mechanism of chemical development, under the spell of the "physical chemistry" of Ostwald-van't Hoff-Arrhenius were subservient to the adage "Corpora non agunt nisi soluta"—by which was understood, aqueous solution. This had its points, but only van't Hoff was already indicating the existence of solid solutions. Without too much thought as to the extraordinarily small concentrations of silver ions possible, only a reaction in aqueous solution was considered. It was supposed that in this phase reaction oc-

* From A. Moyse, A. Ballard and S. E. Sheppard: Unpublished work from the Eastman Kodak Research Laboratories.

curred between dissolved silver ions and the (organic) reducer according to such an equation as

$$\overset{+}{Ag} + \overline{R} = Ag_{met} + R$$

or, for such semi-inorganic reducers as ferrous oxalate

$$\overset{+}{Ag} + Fe(\overline{\overline{C}}_2O_4) + C_2\overline{\overline{O}}_4 = Ag + Fe\ (\overline{\overline{C}}_2O_4)_3$$

and such inorganic ones as hydroxylamine

$$2\ \overset{+}{Ag} + 2\ NH_2OH = 2\ Ag + N_2 + 2\ \overset{+}{H}_2O + 2\ \overset{+}{H}$$

As stoichiometric equations, these express the economics of the process, but actually give no indication of the mechanism. In this matter, it was supposed that a supersaturated solution of atomic silver is formed, which then deposits silver locally upon the latent image nuclei. This was considered to occur the more readily, the larger the nuclei. In connection with this there was stressed the conception of equilibrium or reversibility in development. This could, in fact, be demonstrated for the ferrous salts, and imputed *de jure* to organic reducers such as hydroquinone—although, in fact, it was evidently of vanishing moment for these substances. And its entire absence in the case of NH_2OH, (hydroxylamine), and some kindred reducers, was quite overlooked. None the less, an important rôle was assigned to the redox potentials of developers.[11] The influence of bromide ions on development was ascribed to their ability to reduce the concentration of free silver ions, according to the mass law $[\overset{+}{Ag}\ \overset{-}{Br}] = \text{constant}$, and thereby the free energy of the reaction $\overset{+}{Ag} + \overline{R} \rightleftharpoons Ag + R$. Hence it was deduced that the *relative* redox potentials of developers might be ascertained from the effect of bromide ion in depressing the density developed in unit time.[11] In the case of true equilibrium in development, which can be approached in the case of ferrous salts, these considerations have a certain weight. But with the usual organic reducers, development is not a reversible reaction.

It was W. D. Bancroft[12] who first expressed scepticism of the importance of an equilibrium in ordinary photographic development. He suggested that adsorption of the developer to the silver bromide grain is the essential thing. Provided that the "latent image" could be reduced more easily than the remaining silver halide, one obtained a *negative,* whereas if the unaffected silver halide were more easily reduced, as he supposed to occur with certain reducers, a *positive.* This corollary is a purely *ad hoc* interpretation, and disregards the evidence that the latent image consists itself of silver. None the less Bancroft first rescued the process from a watery grave. The idea of an adsorption was proposed in a more explicit fashion by Sheppard.[13] He suggested that the reducer, which is usually an anion, is adsorbed to the silver halide, forming a metastable complex or salt of silver. At the interface between the silver speck of latent image and the silver halide the complex decomposes giving an atom of silver to the nucleus. Certain experiments independent of development were produced in support of this conception, but it could not be regarded as conclusive. A change of ground was proposed by A. J. Rabinowitsch.[14] He suggested that the developer is adsorbed to the metallic silver nucleus itself, and with S. Peissakhowitsch[14] proffered experiments indicating adsorption of hydroquinone to colloid silver. The experiments of Rabinowitsch and Peissakhowitsch were repeated by Perry, Ballard and Sheppard.[15] They showed that the experimental error of the Russian investigators, consequent on not removing oxygen from the system, was much too great for their conclusions to be accepted. Actually, Perry, Ballard and Sheppard failed to calculate the degree of unimolecular adsorption possible under their experimental conditions. This will be done now to give an approximate idea of the possibilities.

The particle size of the silver sol was estimated from electron micrograms obtained by Mr. A. L. Schoen,* taken at 25,000x magnification. Micrometric readings were made on prints at a further optical enlargement of 3 to 4 times. The readings were made for about 100 particles. The diameters varied between 10 mμ and 40mμ, the arithmetical mean being 20mμ. Assuming the particles to be cubic, and the measured diameter that of the diagonal, from the size-frequency distribution, the total area of all class-sizes gave an average diameter of 202 Å, which is very nearly that obtained for the arithmetic mean, *viz.*, 200 Å. The surface area of 100 particles was calculated to be 8.2×10^6 Å^2. These values are subject to error from several sources. The number of readings was insufficient to give a good size-frequency estimation. Some agglomerates were observed in the electron micrographs, but it is not possible to say whether they existed in the sol, or were formed in the preparation for the electron microscope. Hence they had to be neglected.

From the radius of the silver atom ($r = 1.44$ Å) or a distance between centers in the surface of 2.87 Å, it is readily calculated that the number of silver atoms on the surface of the average particle is 1.20×10^4. According to the estimate of size, the specific area per grain is 4.9×10^5 cm^2. Again, the amount of silver per liter of the sol was 0.68 grams, corresponding to 4.1×10^{16} particles per liter, or to 5.2×10^{20} surface silver atoms.

The experimental data showed 1.03×10^{-4} equivalents of hydroquinone unaccounted for per liter, *i.e.*, 6.2×10^{19} molecules. Hence, the ratio of hydroquinone molecules not accounted for to the number of surface silver atoms is $\frac{6.2 \times 10^{19}}{5.2 \times 10^{20}}$, or approximately 1/10. However, the hydroquinone molecule is much larger than the silver atom, and would "cover" several if monomolecularly adsorbed.

Calculation of the van der Waals' area of the projection of a dihydroquinonate ion (i) lying flat, (ii) oriented edge-on, (see Fig. 1) gives for the former (more

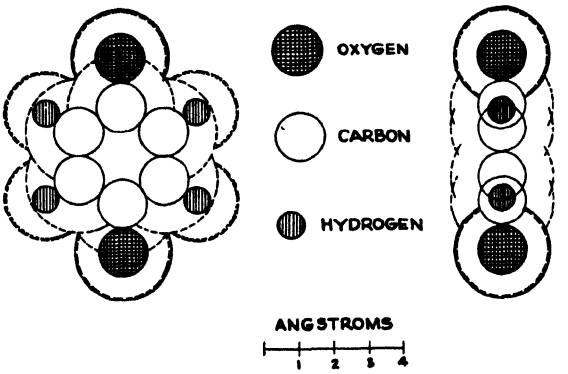


FIGURE 1. Van der Waal's projective areas of the dihydroquinonate ion (1) flat, (ii) oriented edge-on.

probable) orientation 43 Å^2, for the latter 35 Å^2. From this, and assuming *irreversible* adsorption, *i.e.*, total adsorption, one hydroquinonate ion could cover about 7 silver atoms when flat, or about 5.5 edgewise.

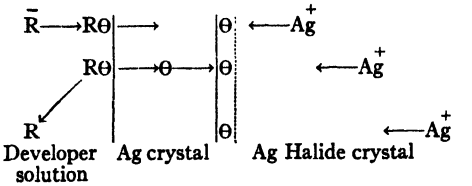
If *all* the hydroquinone not accounted for in the analyses were adsorbed, this would allow about 84 per cent coverage for flat molecules, or about 60 per cent, with edgewise molecules. However, the discrepancy was actually close to the experimental error.

Consequently, it has not been demonstrated that an unimolecular adsorption may

* Of the Eastman Kodak Research Laboratories.

not occur to some extent, whether on colloidal silver or colloidal silver sulfide, but it is equally probable that the amount possibly was consumed by residual oxidation.

Thus the question is not decided, and further experimental work is required. It is, moreover, decisively important in respect of a hypothesis of chemical development advanced by Gurney and Mott.[16] According to these authors, developers are substances, donors of electrons, the levels of which must be higher than the electronic levels in silver or silver sulfide. There is supposed to be some difference in the electronic levels in minute specks of silver, such as the latent image, the levels being lower, up to a point, as the speck is larger, and conversely. It is supposed further that the reducer attaches itself (is adsorbed) to the silver speck, charging it with its electrons. Thereby the speck tends to acquire a negative charge on the face remote from the developer. This face attracts silver ions of the lattice of silver halide in which the speck is imbedded; these are neutralized by the electrons of the hinterland, becoming as silver atoms part of the growing silver lattice. Schematically, the mechanism is indicated by the following diagram:



This indicates what is supposed to happen at the two interfaces. On the left hand side of the silver speck the reducing ions move up and are adsorbed to the silver, to which they contribute electrons. At the right hand side these accumulate on the rear face of the silver speck, where they attract interstitial silver ions from the solid silver halide in contact with the silver speck. These they neutralize, forming silver atoms which build on to the silver crystallite, and *da capo*. A less schematic and more topographic version of this hypothesis is shown in Figure 2, due to Dr. J. H. Webb of the Eastman Research Laboratories.[17] This is a very attractive picture of the mechanism of development, especially in respect of its continuity with Gurney and Mott's theory of the formation of the latent and visible images by a similar process. It has particular interest for colloid chemistry, in suggesting a mechanism for catalysis at interfaces.

In specific regard to development, the accretion of silver atoms at the rear interface Ag/AgBr has been supposed to develop a reaction pressure pushing the speck outward into the developing solution. This could provide an explanation of the observations, discussed in more detail later, that in the early phases of chemical development, protuberances and even filamentous processes may be seen—or rather photographed—to have been produced beyond the surface of the silver halide grain. It is true that the specific volume of silver metal is *less* than that of silver halide which it is replacing, and that the halide ions remaining from the reduction must go out into the solution, leaving void spaces to the rear of the speck. These considerations reduce the probability of a pressure occurring as suggested; there is also the fact that extra-granular protuberances do not appear to be an inevitable aspect of development, since it is possible for development to be effected[18] so that no protuberances are formed, the crystal maintaining its shape, giving a slightly shrunken pseudomorph. It has been suggested, however, by adherents to this hypothesis—one is tempted to term it the Januarian or two-faced hypothesis—that it provides an explanation of the infra-granular filamentous proliferation of the reduced silver in chemical development, as revealed by the electron microscope.[19] We shall come to speak of this in due course.

There is one factor which argues rather strongly against this hypothesis. Other things equal—same developer, same temperature, etc.—the rates of development should depend upon the mobility of silver ions in the respective crystals. The rates should be, therefore, in the order AgI>AgBr>AgCl. Now photographic experience has long opted for the reverse order of developability. Recent experiments of S. E. Sheppard and C. L. Graham [20] have shown that the relative rates of development under as nearly identical conditions as possible strongly indicate the order

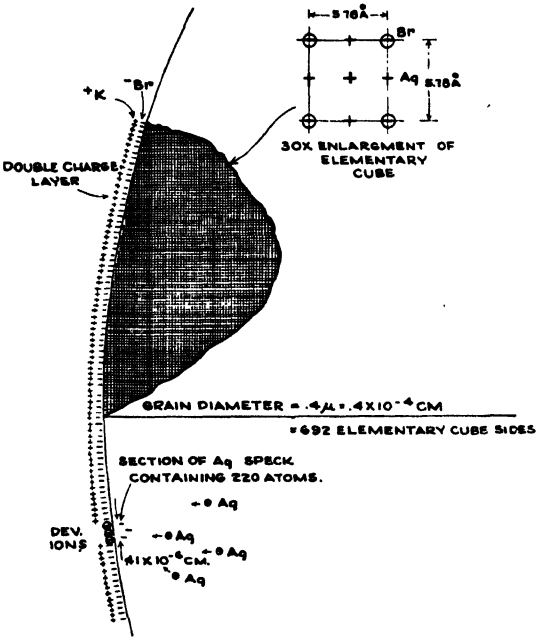


FIGURE 2. Schematic model of an exposed grain and latent image speck, illustrating Gurney-Mott hypothesis of development (according to J. H. Webb).

$AgCl \leqq AgBr \gg AgI$. The mobilities will be proportional to the electrolytic conductances since the evidence is that only the silver ion conducts in these crystals, and the movement is supposed to be directed by the electric field of the speck. Values for these conductances are given in the following table:—

Table 1

(Specific Conductivities in mhos)

Silver Halide	Conductance at 26° C. S. I.	Conductance at 25° C. S. S.	Observed Range of Transition from S. S. to S. I.
AgI	4×10^{-7} to 8×10^{-7}	4×10^{-7}	None
Ag(Br)(I) 1 to 5% ∓	1×10^{-4} to 1×10^{-7}	1×10^{-4} to 1×10^{-5}	None
AgBr	5×10^{-5} to 1×10^{-7}	0.5×10^{-7}	20° C. to 0° C.
AgCl	1×10^{-10}	3×10^{-7} to 1×10^{-7}	100° C. to 200° C.

It should be explained why to these substances there might be imputed more than one conductance at the same temperature. This depends upon whether the crystal is presented in a condition approaching the ideal lattice, or, broadly speaking, as an imperfect crystal. In the former case (S. I. = structure insensitive) the conductance

has a more constant and reproducible value, and work has to be done both to get a silver ion out of the lattice into an interstitial position and to mobilize the interstitial ion. In the S. S. (structure sensitive state) the work function may refer chiefly to mobilization of interstitial ions; although in this latter case it is possible that we are not solely concerned with the relatively simple contrast of order: disorder, where the disorder is of the Frenkel type. This means that in the disordered system, only strayed $A\dot{g}$ ions are involved.[21] It is possible that we are concerned in the S. S. state—qua state—with lattices, or orders, of more open type, *i.e.*, of somewhat higher energy.

It has been calculated by Webb[22] from purely photographic data * that the mobility of silver ions in the silver iodo-bromide crystals of ordinary negative photographic emulsions corresponds to that of S. I. silver bromide. That is to say, he concludes that the temperature coefficient of the mobility over the range from 50° C to −72° C accords with that for the S. I. type. This does not seem to agree too well with the direct electrolytic conductance measurements. However, should this be the case, and for the other halides, that is, should their crystals in gelatin emulsions be of the S. I. type, evidently we have a rather flagrant disagreement with the Januarian hypothesis in respect of the order of developability. On the other hand, should it be concluded that the grains are of the S. S. type for silver bromide, iodo-bromide, and iodide, this could hardly be the case for the chloride, according to all the measurements available. In fact, in any case, the actual order of developability disagrees definitively with that deduced from direct measurements of the conductivity.

It has been suggested[23] that in view of the rapid and localized nature of chemical development, the heat produced by the exothermic reaction $\bar{R} + A\dot{g} \rightarrow R + Ag$ may count for something in the actual process, that this may tend to becóme adiabatic rather than isothermal. There is much to say for this as an operative factor. However, the "heats of activation" of silver ion mobility derived from the temperature coefficients of electrolytic conductance are as follows:

	S. S.	S. I.
AgI	(10)k. cal./mol.	(10)k. cal./mol.
AgBr(I) (1 to 5%)	5 to 7	
AgBr	6 to 9	14
AgCl	7 to 9	20

If the grains are S. I. crystals, the same heat of reaction (in development) should produce a much greater acceleration for AgI than for AgBr than for AgCl, which is again contrary to the order of developability. If they are of S. S. type, there should be approximate equivalence for AgCl and AgBr, with perhaps some disadvantage to AgI, but hardly sufficient to explain the low developability of this. Moreover, the actual temperature coefficients of development velocity depend considerably upon the developer used, and bear little relation to the temperature coefficients of silver ion mobility in the silver halides. Hence, one may conclude at present that this latter can play but a small part in chemical development.

None the less, the evidence is strong that the classical model, reduction of silver ions in aqueous solution followed by deposition of silver on the "latent image" nucleus, is entirely inadequate. The process has this definite colloid chemical aspect, it seems to be essentially a reaction at the interface of solid and solution.

Retarding and Inhibiting Barriers

Thus far we have considered photographic development without notice of the most obvious colloid concerned, *viz.*, the gelatin in which the silver halide crystals are suspended. Let us retrace our steps to the beginning. In the earliest phase, gelatin has to absorb water and developing solution, and swell.

*** From the effect of temperature on the high intensity reciprocity failure.**

The swelling of gelatin in liquid water involves at least three processes, *viz.*: (i) Adsorption of water molecules to the gelatin macromolecules. This corresponds to the early stage of the sorption of water vapor by gelatin.[24] (ii) Release of orientation (strain) [25] of gelatin macromolecules and increase of entropy—a process involving a Joule effect similar to that observed on heating stretched rubber. This process—elastic swelling—occurs in the latter stages (high humidities) of vapor sorption. (iii) Adjustment of the ampholyte to the pH of a solution, with development of Donnan membrane equilibrium, giving excess swelling over that of isoelectric gelatin according as the pH departs from the isoelectric point.

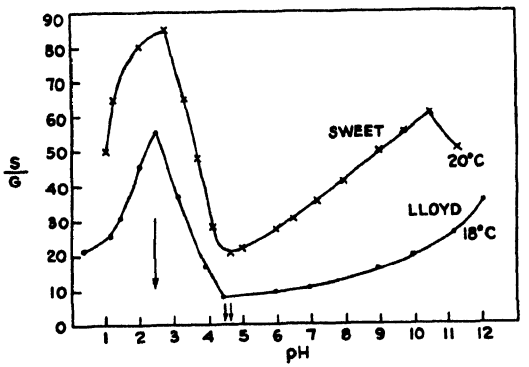


FIGURE 3. Swelling of gelatin in water as a function of pH.

The usual organic developing solutions are quite alkaline, ~ pH 11. As can be seen from Figure 3, gelatin in such solutions will tend to swell very considerably and indeed quite rapidly. This is, however, restrained by the presence of neutral salts, usually the sulfite and carbonate of soda of the developing solution, and on occasion, by the addition of sodium sulfate. The adjustment due to change of pH corresponds roughly to the process symbolized below:

$$\overset{+}{H_3}N - R - COOH \;|\; \overset{+}{H_3}N - R - C\bar{O}O \;|\; H_2N - R - C\bar{O}O$$

Hybrid ion at isoelectric point

pH decreasing ← → pH increasing

In the alkaline solution the gelatin framework develops a negative charge. According as the reducing molecule is negatively charged or neutral, this will or will not occasion a barrier to its penetration.[26] This is not the only barrier protecting the silver halide from the attack of the reducer, as we shall see. The alkalinity or high pH of the medium, however, has a twofold effect: it charges the gelatin negatively, but at the same time the relative freeing of $-NH_2$ groups of the gelatin enables these to coordinate with silver ions, as shown on titrating gelatin electrometrically with silver salts (cf. Fig. 4). This involves a lowering of silver ion activity, and therewith of reducibility. Considerable of the "protective" action of gelatin is no doubt due to this. It is limited, however, in ordinary photographic emulsions by another factor, namely, the excess of bromide ions usually present.* This will tend to restrain the amino groups from coordinating with the silver ions of the silver halide.

The negative charge on the gelatin will tend to hinder and restrain the entry of negatively charged reducing ions into the emulsion layer, as has been definitely

* The pAg of such emulsions is of the order of 8 to 9.

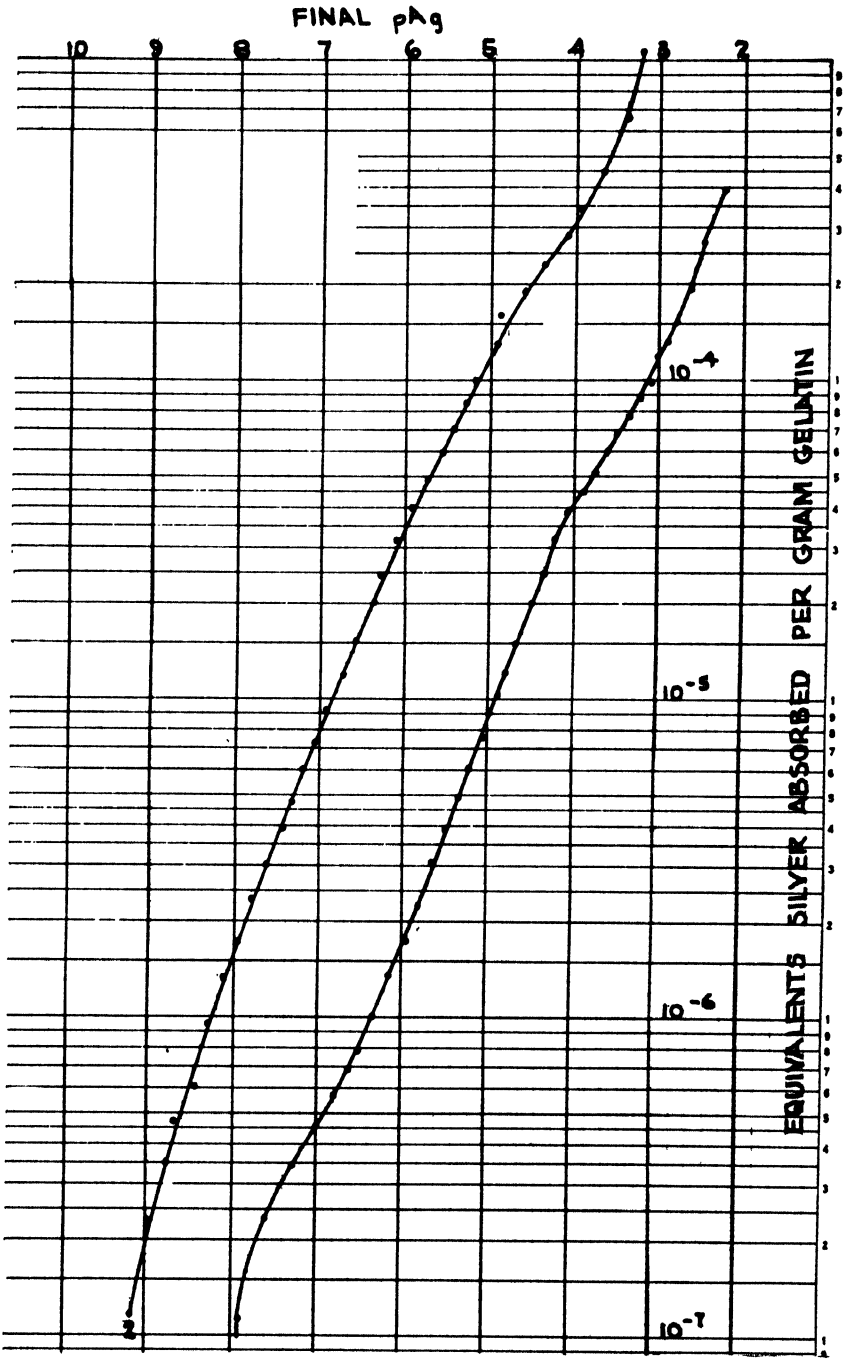


FIGURE 4. Combination of silver ions with gelatin as a function of concentration of silver ion and of pH; curve on right for pH 5; curve on left for pH 9.

shown to occur for negatively charged dye ions.[26] Although the molecular weight of such dyes exceeds that of the usual developers somewhat in the ratio of 4 to 1, the barrier must exist to some extent in the latter case also.

The effect of such a barrier may be reduced or "swamped out" by adding a considerable concentration of neutral electrolyte, *e.g.,* sodium sulfate. This has been demonstrated in the dyeing of gelatin [26] and is perhaps in some degree the reason for the acceleration of development by neutral salts with certain alkaline developers, which was observed by Lüppo-Cramer.[27] However, the "potential barrier" constituted by gelatin for negative ions is only an outer barrier. There exists an inner barrier of more fundamental importance, consisting of adsorbed halide ions, and in our particular case, bromide ions. Some two decades ago, Lüppo-Cramer [28] directed attention to two important facts. Precipitated and well washed silver bromide is immediately reduced in darkness by ordinary photographic developers. On the other hand, a hydrosol of silver bromide, stabilized by slight excess of bromide ions, is only slowly reduced in darkness by the same developer, but rapidly after exposure to light. It is, as he put it, a model of a nonfogging emulsion. This striking difference between "gel" and "sol" forms of silver bromide was made the basis of a hypothesis of latent image formation and developability * by Ungar and G. Schwarz.[29] They suggested that the essential action of light consisted in discharging electrons from bromide ions, adsorbed to silver bromide, and forming a "barrier" layer. This allowed a developer to attack unprotected silver ions. Without much doubt this exaggerated the rôle of the barrier ions, totally neglecting as it did, the demonstrable importance of silver, and on occasion, silver sulfide, specks.[30] The essential importance of the barrier ions for the kinetics of development, rather than for the formation of the latent image, has been shown by James,[31] in the first of the most searching and serious of recent investigations on development. In earlier studies on the kinetics of development [32] the existence of a more or less marked "induction period" was educed, as well as the fact that its magnitude depended upon both the nature of the developer and the concentration of bromide ions. This was interpreted, in terms of a thermodynamic equilibrium, as a period of *nucleation* consequent upon both the redox potential of the developer, and the active mass of the silver ion as controlled by that of the bromide ion (as we would say now, the pAg). This lead to much work on the empirical estimation of the redox potentials of developers *via* their bromide depression coefficients, since developers, like humans, differ in respect of the depressing effect of a dose of bromide. James made clear that the supposed connection between bromide susceptibility and redox potential was entirely fallacious. Proceeding from the viewpoint of reaction kinetics rather than from that of thermodynamic equilibrium, he associated bromide susceptibility with the sign and the multiplicity of the electric charge upon the reducing entity. In an investigation of Staude's observation [33] that *quinone* catalyzes development by *hydroquinone,* James showed that acceleration by oxidation products is not a general effect for organic developing agents, nor even for a majority of them. For inorganic developers it was evidently not the case. Further, he showed that quinone catalysis cannot even account for the induction period in hydroquinone development. In his words, "The induction period has its origin in an entirely distinct kinetic phenomenon, involving the penetration by the developer ions of an electrical barrier at the surface of the silver halide grains." [31] In support of this conclusion, James pointed out that "with some developers no induction period appears, with others a short induction period is present, and that with still others, the induction period is pronounced. A comparison of the curves obtained for all the developing agents tested reveals four distinct types. These types are not defined by the chemical nature of the developing agents, but rather by the *charge on the reducing ion or molecule.*"

* Curiously enough, an equivalent difference between "sol" and "gel" forms of colloidal silver was suggested for latent image formation (developability) by F. F. Renwick.[32a]

In Table II developing agents are grouped somewhat according to their chemical constitution (vertically) and (horizontally) according to the charge on the actual reducer.

Table 2

Classification of Developing Agents

Chemical Series	Charge on Active Particles			
	0	−1	−2	−3
Hydroquinone			Hydroquinone	Sodium hydroquinonemonosulfonate
p-Aminophenol		Elon	*p*-Hydroxyphenylglycine	
		p-Aminophenol	Elonmonosulfonic acid	
p-Phenylene-	*sym*-Dimethyl-*p*-phenylenediamine	*p*-Aminophenylglycine		
	p-Aminodiethylaniline			
	Diaminodurene			
	2-Amino-5-diethylaminotoluene			
Others		Hydroxylamine	Pyrogallol $(Fe)C_2O_4)_2^{--}$	

Figure 5 reproduces James' characteristic curve for each charge group. "When the reducing agent is uncharged (group 0), there is no induction period. If the agent possesses a single negative charge (group −1), a small induction period appears. The extent of the induction period increases progressively as we pass to two negative charges (group −2) and to three negative charges (group −3). This behavior was observed with every agent tested, regardless of the chemical series to which the agent belongs. Group 0 contained only members of the *p*-phenylene-diamine series which, as we have observed, should not undergo significant acid ionization at the pH employed. Group −1 contains some of the *p*-aminophenols, a substituted *p*-phenylene-diamine which has obtained a negative charge by virtue of dissociation of the acid carboxyl group it contains, and the inorganic substance hydroxylamine. Group −2 contains hydroquinone, which, in the pH range employed, is active as the divalent ion, two members of the *p*-aminophenol group which have acquired an added negative charge by virtue of ionization of a carboxyl or sulfonic acid group, and potassium ferroöxalate.* The active ionic species of pyrogallol has not been definitely established as yet, but the curve of the pyro developer fits fairly well into group −2. Group −3 has as its sole representative sodium hydroquinonemonosulfonate." For further quantitative evidence validating his theory, we must refer to the original paper.

James then shows that the behavior of bromide ion, and of change of concentra-

* "This developer had the composition M/40 ferrous sulfate, M/5 potassium oxalate, and 0.1 g. potassium bromide per liter. The solution was buffered at pH 8.56."

tion of bromide ion, with different developers is consistent with this "potential barrier" theory. The respective parts played by a presumed rigid element of a Helmholtz double layer—ψ-potential of Freundlich and by a diffuse or Gouy layer—so-called ζ-potential—have afforded a fruitful field for debate in colloid chemical problems. It is very possible that it is fundamentally a matter of total (irreversible) and reversible adsorption respectively. In the case of bromide ions, adsorption as distinguished from integration in the crystal lattice, must be regarded as essentially reversible, although there are probably certain degrees of lattice surface stability.

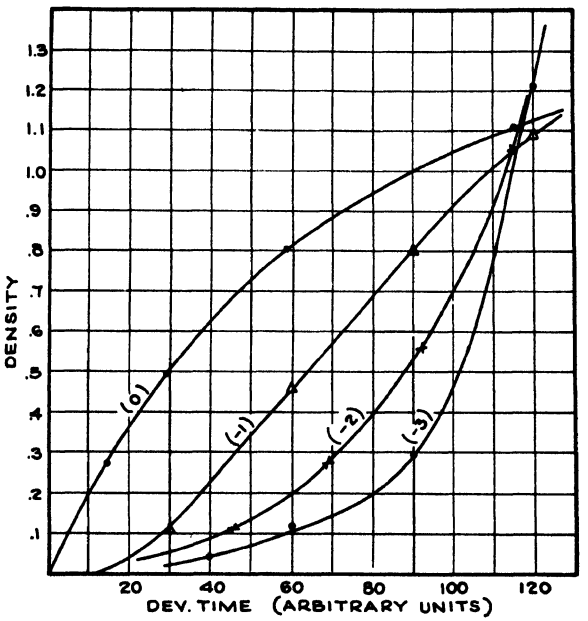


FIGURE 5. Effect of charge of reducing agent upon shape of development curve.
× Hydroquinone
△ *p*-Aminophenol
• Diaminodurene
o Sodium Hydroquinone Monosulfonate

The double layer for bromide ions and counter ions will, therefore, be of the diffuse type, and not the rigid type frequently illustrated (Cf. Fig 2). Owing to the mobile type of this adsorption, the rôle of the kinetic energy of the developer ions in penetrating the barrier, as postulated by James, becomes more plausible. This kinetic (ζ-potential) character of the bromide ion barrier is in agreement with the fact observed by Sheppard and Elliott[84] with a rather weakly alkaline (sodium carbonate) hydroquinone developer that agitation progressively reduced the induction period. It does this by washing out bromide ions into the surrounding liquid. The suggestion of an impeding or delaying rôle of negatively charged gelatin is quite consistent with and consequent upon James' theory of the rôle of bromide ions. Compared to these, its negative barrier function must be of minor importance. This is evident from the dominance of the bromide ion effect in photographic development. The gelatin barrier would be relatively *rigid,* and hence the smaller molecular weights of the developers would count in comparison with dyes. The "protective" colloid character of gelatin for silver halides towards all developers is, as previously stated, doubly a function of pAg, and indeed becomes decisively lessened for

pAg<6 and for pH<6. This is shown in the following figure (Fig. 6). This brings out two things. Spontaneous developability or "chemical" fog increases rapidly for pAg values<6, but also the amount is much greater the longer the time of keeping or of incubation of the emulsion layers. The latter effect indicates the acceleration produced by *silver nuclei* formed on incubation of the layers of low pAg. The correlative to these consequences of deliberately adjusting the pAg of emulsion layers is the frequently observed "fogging" effect of simple washing of photographic

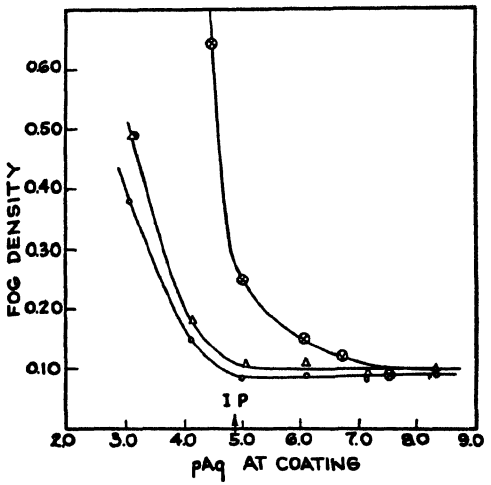


FIGURE 6. Protective behavior of gelatin for silver halide toward developers as a function of pAg.

plates and films. In recent years the reduction of the activity of silver ion effected by excess bromide has been replaced or supplemented by various socalled "antifoggants" such as nitrobenzimidazol, benzotriazole, which form compounds of low solubility or of high complexity with silver ions. It is possible that their action is similar in some respects to that of dyes which lower the reducibility of silver halides [35] and that in particular they tend to react with the silver halide at its interface with the silver speck of the latent image,[36] an action which would constitute a still deeper barrier than that by bromide ions.

Oriented Adsorption

The adsorption of halide ions, *e.g.*, of bromide ions to silver bromide, is a reversible process, involving a kinetic equilibrium similar to vapor pressure. There is possible with certain substances adsorption resulting in static equilibrium, because the adsorbed ions or molecules are so *oriented* on adsorption that that part of the ion or molecule which allows hydration, the hydrophile part, is attached to the solid surface, while the part presented to the water is hydrophobe and incapable of hydration. The picture is essentially that introduced by Hardy,[38] Harkins [38] and Langmuir.[39] Certain dyes, notably the cyanine dyes used as optical sensitizers in photography, and many desensitizing dyes, have a structure, illustrated for the cyanine dyes in Figure 7. It has been shown by James [40] that the reduction of precipitated silver chloride by hydroxylamine proceeds in two phases, a non-catalyzed one and a catalyzed one—the latter corresponding to the "nucleation" of the halide with metallic silver. The rate of reduction in the non-catalyzed phase is proportional to the surface area exposed to the hydroxylamine. This is shown by the progressive diminution of the effective area on increasing the degree of adsorption of a selected cyanine dye. When saturation with the dye is reached, the rate of reduction falls to a steady minimum. The existence of a minimum, rather than complete inhibition, derives no

doubt from the size of the dye molecule, and the fact that it "covers," but does not neutralize, several ions of the surface.[37] Hence it leaves interstices through which the reducer can penetrate and react.

Gelatin behaves in a similar manner, but with a considerably higher minimum, indicating a still more open adsorbed layer. The similarity, however, is notable. It was shown by Sheppard and Sweet[41] from experiments on the effect of gelatin on the interfacial tension between water and toluene, that gelatin must be attributed the double-edged * character of such molecules as the higher fatty acids, the soaps, many

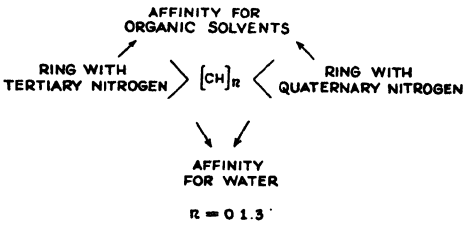


FIGURE 7. Schematic structure for cyanine dyes.

dyes. This has been confirmed recently in experiments on its adsorption to silver bromide. Probably because of the large size of the molecule or micelle, and the sparse distribution of the atom-groups attaching to the silver halide, the oriented adsorption layer is much more open, the "protection" even less complete than with dyes.

Both the basic cyanine dye and gelatin reduce the rate of reduction of *nucleated* silver chloride by hydroxylamine to minima, as in the case of the non-nucleated reaction. The behavior is consistent with the view that the reaction is between unprotected silver ions of the surface and the hydroxylamine,[18] and agrees with the conception previously put forward.

Another example of "oriented adsorption" in relation to development is due to A. Lottermoser and R. Stendel.[42] Proceeding from the Schwarz-Ungar hypothesis of developability by removal of negative charges at the surface, they tried the effect of surface-active negative and positive ions on the reducibility of silver halide sols, and on the development of gelatino-silver halides. As surface-active *anion* was used the sodium salt of duodecyl sulfuric acid ester—$C_{12}H_{25}OSO_3Na$ (abbreviated as C_{12}'), and as surface-active *cation,* the duodecyl-pyridinium bromide—$C_{17}H_{30}NBr_1$ (abbreviated as C_{17}'). These interesting experiments can be only briefly summarized. It was found with silver iodide hydrosols that in presence of the specified "surface-active" ions these alone determined the charge, irrespective of excess of silver or iodide ions. When positively charged silver bromide hydrosols were prepared with C_{17}', they were immediately reduced by reducing anions, in agreement with the "barrier" theory. Furthermore, the development of photographic emulsions (developing-out papers) was strongly accelerated by C_{17}', in the case of hydroquinone, definitely but less so for pyrocatechin, and for glycine but little. It is unfortunate that a (probably) uncharged reducer such as hydroxylamine was not tested. The work of James, with silver chloride, to which the cationic cyanine dye was adsorbed, showed that in this case, the rate of reduction was *lowered* by the adsorption of the dye. One would expect the same behavior for the duodecyl-pyridinium cation. The acceleration of development (of gelatino-silver chloro-bromide layers) in hydroquinone by C_{17}' is in accord with Lüppo-Cramer's discovery of the acceleration of development (in hydroquinone developers) by basic dyes, such as phenosafranin.

*G. S. Hartley has termed such bodies "amphipathic," but the term has a subjective implication. Perhaps "amphibolic" would be better.

Inhibitors

A good deal of the matter concerning inhibiting agents has been dealt with in the preceding sections. The conception of "inhibitors" nominally concerns substances which prevent or reduce "fog" in development, undesired "spontaneous" reduction of non-exposed silver halide grains. The use of bromide (halide) as a fog inhibitor did not lead to any extension of this category. The discovery by Lüppo-Cramer [43] of desensitizing dyes (consequent on his observation of the action of the oxidation products of certain developers, notably *amidol,* in protecting an emulsion against light during development) very possibly opened up this field. He found that many basic dyes, *i.e.,* cations, inhibited the formation of new "latent image" during development. By no means all dyes of this type are practically useful, and the theory of their action is not fully understood. However, it may be stated that fundamentally the dyes in question are such as have a definite redox potential sufficient to oxidize nascent silver—or its equivalent *—of a latent image. It is a curious fact that Lüppo-Cramer himself vacillated between two hypotheses as to their action. He proposed first an "oxidation" theory in accord with the foregoing statement, then a hypothesis of "nucleus isolation," according to which the dye was adsorbed chiefly at the interface between silver nuclei of the latent image, and silver halide.[44]

In 1927 Sheppard and Hudson [45] discovered the inhibiting action (upon development) of the thioanilides. These substances have no oxidation potential, and their action was interpreted in terms of Lüppo-Cramer's hypothesis of nucleus isolation—it was supposed that they were adsorbed at the interface between latent image silver and silver halide [46]—forming difficultly reducible complexes with silver ions. They were the precursors of a considerable multitude of "fog inhibitors," organic compound forming silver compounds of lower reducibility than silver halide.

It is generally characteristic of these substances that an optimum concentration is soon reached, beyond which not only is "fog" inhibited, but the developability suffers, with consequent lowering of photographic sensitivity and speed.[45] As an example, the following data obtained with nitrobenzimidazol added to a *p*-aminophenol developer [47] may be cited:

Table 3. Action of Nitrobenzimidazole in m/20 Developer

Development Time in Min.	Developer Fog at a Concentration			
	0	m/7500	m/3750	m/750
1	0.06	0.12	0.05	0.05
2	.12	.13	.06	.05
4	.16	.13	.07	.06
8	.23	.25	.09	.07
16	.28	.39	.12	.07
32	.37	.42	.15	.07
b	0	0	0	.4

Dispersity and Form of Developed Silver

After this compilation of factors preventing reduction in photographic emulsions, one might be surprised that it does occur. *E'pur si muove.* The mere abrogation or penetration of barriers does not appear in itself sufficient, the rôle of silver and silver sulfide specks seems undeniable. How do these act, what actually takes place?

Much is shown, yet thereby new questions posed, by visible record of the development of silver halide grains. Not to the naked eye, of course, but to this cyclops armed with microscope, camera, cine-camera, and electron microscope. The eye functions best with *green* light 5600 Å. The practical limit of resolution (with

* With reference to the hypothesis of latent image formation of M. L. Huggins noted later.

ordinary light) is conditioned by the numerical aperture of the lens and the wavelength of the light. With N. A. = 1.41 and light of *ca.* 4200 to 4400 Å, this will mean 0.16μ to 0.20μ. Slightly superior results can be attained with ultraviolet light (and quartz optics), but the gain is hardly worth the effort. Under these circumstances, the depth of focus is of the same order as for resolution. This means that for sufficient exposure the "appearance" is a mere optical section complicated by shadows of the out-of-focus regions.[48] For photomicrography A. P. H. Trivelli has sought to overcome this by cumulative exposures, each an underexposure *per se,* at different focal distances or image depths, but this is a tedious procedure seldom used. The useful magnifications under these conditions lie from 1000 to 2500 diameters. None the less, with these humble means, much was accomplished. Observations by Hodgson[49] showed that development started at discrete points on the grain surface. The qualitative observation was extended and made quantitative by The Svedberg.[50] He and his coworkers showed that these "development centers" were distributed among and over the grains at random. Mathematically: if we suppose q development nuclei are scattered haphazardly over a number of grains of total surface S, then the probability that a grain of area a will get r development nuclei is

$$p_r = \frac{(aq)}{r!} e^{-\frac{aq}{S}}$$

it being assumed that a grain receiving one nucleus is developable. Then the probability of a grain receiving no nucleus, or remaining undevelopable is

$$p = 1 - e^{-\frac{aq}{S}}$$

wherefore the percentage number of developable grains of a certain class-size a is given by the expression

$$p = 100\,(1 - e^{-ca})$$

where a is the class-size (projective area) and c is a constant. This expression in the logarithmic form

$$\underline{\log [100 - \log (100 - p)]} = \text{constant}$$

has been compared with experiment[50a] and verified.

From this evidence for discrete development centers arose a discussion as to whether they derived from a probability distribution of discrete light quanta[51] or from preëxisting "sensitivity centers."[52] The evidence in favor of the latter view was greatly strengthened by the discovery of the photographic sensitizing effect of silver sulfide specks.[53] The rôle of such nuclei in latent image formation and developability was shown to consist, not in increasing the actual photochemical yield of silver atoms, but in concentrating this silver about the speck,[54] a result incorporated in the "concentration speck" theory.

This theory proffered no explanation of the mechanism of concentration. In the "orientation" hypothesis, Sheppard[55] suggested that around a concentration speck was formed a "potential trough" to which electron-wavelets moved, and wherein photoelectrons combined with silver ions to form metallic silver. Intervention of a photoelectrolytic process, involving the migration of silver ions, was suggested by Sheppard,[56] and by Sheppard and Trivelli,[57] following the work of Tubandt and his associates on electrolytic conduction in silver halides. This view was developed more explicitly by Trivelli[58] in a photoelectrolytic hypothesis of

latent image formation. According to this, the essential sensitivity nucleus was a speck of silver sulfide *and* silver. Under the influence of light, this attracted silver ions which were discharged by photoelectrons, forming more metallic silver.

The suggestion of migration of silver ions was not at once followed up, and it was left to Gurney and Mott [16] and to Webb [59] and to Berg [60] to properly evaluate its importance, in terms of the Gurney-Mott theory of latent image formation. Briefly, since we can only discuss the latent image as a prelude (but an essential one) to development, they supposed that electrons released by light from the "filled levels" of the crystal—atomistically the process $Br^- + h\nu \rightarrow Br + \theta$—attained to upper "conductance levels" and then were "trapped" by silver sulfide or other sensitivity specks. The specks, becoming negatively charged by the electronic excess, attract silver ions which react to form silver atoms crystallizing on the speck. Again we cannot deal with the important experimental evidence [59] from reciprocity failure at high and low intensities of light, and at low temperatures, which is thought to support this hypothesis, but we have indicated already a corollary in regard to chemical development. We come now to further photomicrographs and electron micrographs of developing grains. It was observed by Scheffer [61] in 1907 that initially developing grains frequently showed protuberances, evidently of reduced silver. He supposed that adjacent silver halide grains (which he termed "nutrient grains") went into solution, and that silver reduced by the developer was deposited at specific points on the grain which then grew outward. It may be pointed out that calling in adjacent grains was unnecessary; it would suffice for adjacent silver halide of the same grain to perform the task of nourishing the growing points, and, in fact, in some way or other this conception is necessitated by the progress of the grain to complete reduction.

Further high power photomicrography of the developing grain did not support this localized outward process as necessary and sufficient. Studies with weak and retarded development showed that the developed grain may be a slightly shrunken pseudomorph of the original silver bromide grain.[18] A morphological survey of grain development with several developers, by Trivelli and Sheppard,[62] showed that outward processes did indeed occur with all these developers, starting apparently from localized development centers. However, while these may be connected with deformation in development, this is not certain, and many grains developed to more or less definite pseudomorphs of the original bromide grain. Some of these results are reproduced in Figure 8.

It is necessary to mention here that x-ray diffraction studies have shown the developed image, even in early stages, to consist in the main of regularly crystallized silver having the normal cubic lattice.[63] Consideration of this, and of the apparently compact state of the developed grain might have suggested that a principal matter to be explained in development is the exodus of bromide ions—at least no less important than the reduction of silver ions. But, in fact, this was little considered, in resemblance to the neglect in so many theories of the latent image to consider the fate of the halogen. But this at least is small in amount, whereas in development, numerically for every silver ion reduced a halide ion must be evacuated into the solution.

A turn of great importance to the problem was given by the application of the electron microscope. The degree in which this instrument has improved the elucidation of fine structures is dealt with elsewhere in this volume.* Its first application to the developed image was made by M. von Ardenne [18] and this has been much extended by Schoen and Hall.[19] We shall leave till later its revelations as to "fine grain" and "physical" development. Its momentous disclosure for chemical development of silver bromide was to show that in early and intermediate stages the developed grain consists of a tangle of rather flat filaments, recoiling on themselves, sometimes thickening and eventually more or less coalescing (cf. Fig. 9). That

* See paper on the electron microscope by A. E. Prebus in this volume. Editor.

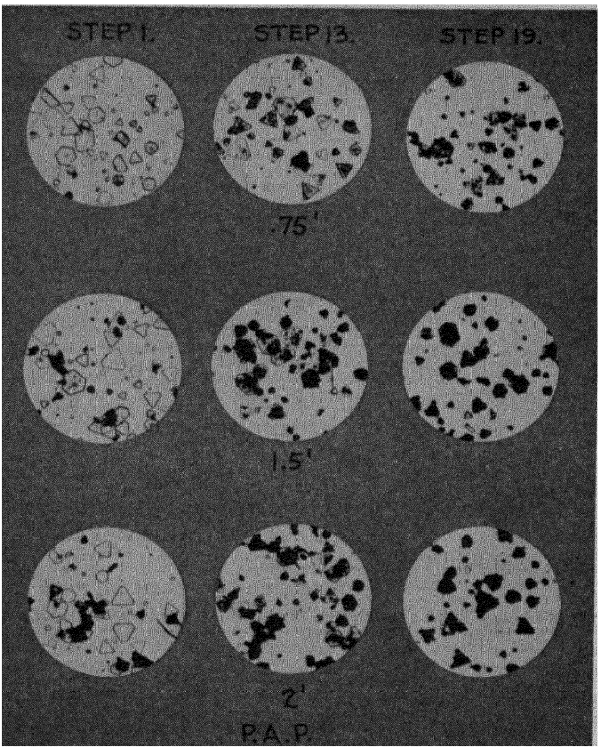


FIGURE 8. Morphological survey of development of silver bromide grains by *p*-aminophenol developer.

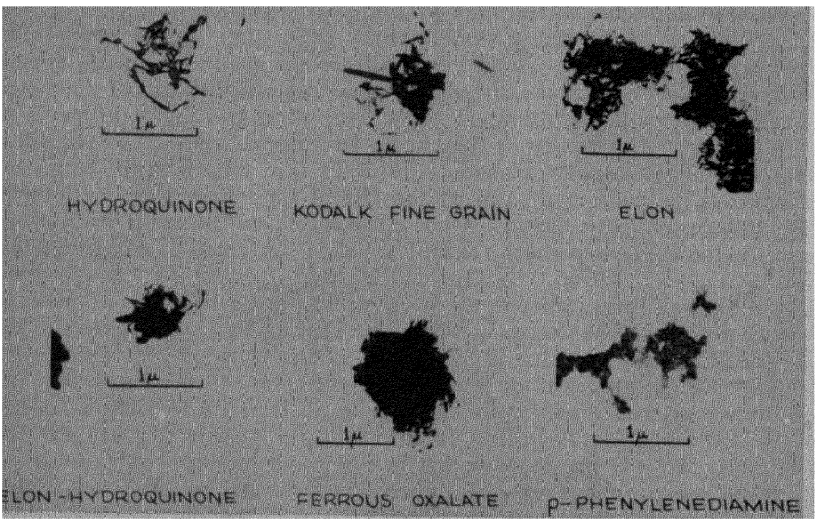


FIGURE 9. Electron micrographs of developing silver halide grains.

this quasi-linear progress of development is primary and fundamental is shown by the electron micrographs of a Lippmann emulsion before and after development (cf. Fig. 10). Each silver bromide grain—of the order of 0.2 to 0.4μ in diameter—is shown to give a single filament or hair. There is much to suggest from the electron micrographs of ordinary negative grains in various stages of development that this is the elementary or unit process in their case also, as shown by successive stages of development on one-grain layer preparations. However, certain electron micrographs obtained by Hall and Schoen with partially developed silver bromide crystals,

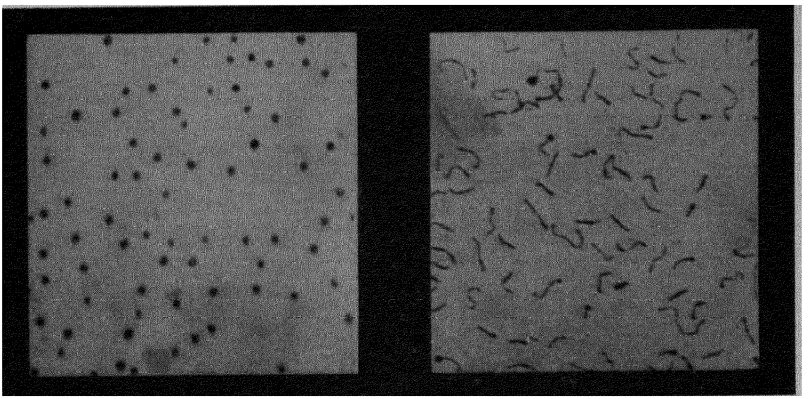


FIGURE 10. Electron micrographs of Lippmann emulsion, before and after development.



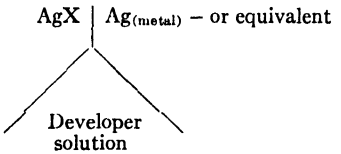
FIGURE 11. Electron micrograph of surface tangle of silver on developing silver halide grain.

fixed and washed on the object slide, suggest that quasi-filamentous growth may spread over the surface of the grain and throughout the crystal. This is illustrated in Figure 11. E. E. Jelley [48] has pointed out that "frequently, angular enlargements are seen in the filaments . . . , and many of these enlargements have an obviously triangular contour."

"This is a strong indication that the sides of the filament correspond to the octahedral (111) faces of the silver crystal, so that the filaments may be considered as silver crystals which have grown chiefly in one dimension." Before this consideration of the revelations of photomicrographic and even more of electromicrographic revelations of the development of *individual* grains of silver halide, we have dealt mainly with the *statistical* behavior of very large numbers of grains in development. We may regard the conclusions as to various "barrier" effects controlling the initiation of reduction in the grains as so far well founded. The optical and electron microscopes are not fitted to testify as to these conclusions. *Per contra,* these instruments give us the most definite evidence available as to the actual progress of reduction of a grain once the process is initiated. Reconciliation of the statistical with the individual data will give a definitive theory of development. At the point at which we have arrived, the question is by what mechanism is the filamentous growth of the developed grain to be explained?

Interfacial Conditions

Except for an extreme form of the Gurney-Mott hypothesis it is admitted that the process must commence and continue at the triple interface,



. . . provided, that is, that the original Ostwald, etc., and recent Reinders [11] theory of reaction of silver ions in solution is not accepted.

The principal objection to the former hypothesis is that it makes no satisfactory provision for the exodus of halide $--\overline{X}--$ ions into the solution. On the other hand, the older hypothesis by aqueous solution has too low a concentration of silver ions to offer to meet the demands of the reaction velocity.* Let us for the moment discard these two dualistic hypotheses, and consider only the advance of the *triple* interface—the metallic silver phase gaining at the expense of the silver halide phase by the energy of the oxidation of the developer by the silver ions. For the moment it is not of principal importance whether this occurs by way of the adsorption of reducer ions or molecules to the silver halide as suggested by Sheppard [13] or to the silver metal, as proposed by Rabinowitsch.[14] In both cases it is agreed that it is the triple interface which is the essential focus of the process. It is important that in either case continued contact of the developer solution with the solid phases is required. This is vital in regard to an aspect of the problem which has been much neglected in recent theorizing on reduction in the solid phase, *viz.,* the removal of the halide ions of the lattice. In the classical "solution" theory, this occurred so to say concurrently with the solution of the silver ions. The *order* of developability, $AgCl \gtrless AgBr > AgI$, observed, if not quantitatively at least qualitatively in both practice and experiment, follows of itself. However, for the newer hypotheses as to development in the solid phase, difficulties arise, since the mobility of halide ions in the solid lattices is admittedly very low, and requires very high "activation ener-

* Large grains—of 2 to 3μ—may be completely developed within a minute or two!

gies" compared with the silver ion. We may conclude provisionally that *the hydration and diffusion away of the halide ion is a dominant factor in the process of chemical development.* A condition for this is the continuous presence of the aqueous developer at the site of the elemental chemical reaction.

However, if we admit pure "adlineation" of developer ions at the triple interface, with resultant reduction of silver ions to silver atoms which "crystallize" on the speck, we do not obtain immediately a condition for filamentous processes. About a symmetrically circular speck of silver sulfide or silver in octahedral (111) surfaces of a silver halide crystal this should give a regularly increasing circular area of reduced silver. Rabinowitsch [14] has described such a phenomenon, observed with a single nucleus produced on special large silver bromide crystals—6 μ and greater—prepared according to a procedure of Meidinger.[63] The area is said to have increased according to the square of the time, in accordance with an autocatalytic reaction in two dimensions. However, this process is stated to have occurred within the "induction period"—which phrase is not defined—and it is further stated that with "fast" developers, a "stormy" process sets in with the formation of protuberances or "explosions" shot out from the grain.

The phase of superficial spread may occur and is in accord with observations of the formation of faint superficial films of reduced silver in the earliest stages of actual development,[62] when presumably several such zones from more than one nucleus might meet. However, it is evident that it does not help to account for filament formation. A. J. Rabinowitsch has suggested rather vaguely that the "stormy" period of development is occasioned by the necessity of formation of a new crystal lattice of metallic silver with different parameters—*e.g.*, $dAgAg = 4.08$ Å compared with $d\overset{+}{A}g\overset{+}{A}g = 5.78$ Å in silver bromide.

If the reaction takes place in the crystal, or at least in the silver halide crystal contiguous both with developer and metallic silver, this appeal to the principle of minimum work is important. It would have no bearing here if the silver ions were reduced from aqueous solution. Bernal [64] has pointed out certain pertinent applications of the principle of minimum work in connection with transformations in the solid phase. As he says "It is fairly obvious that any change inside a solid cannot involve the movement of any of its parts outside the range of the interatomic forces of its neighbors. Consequently, one of the main consequences of solid reactions must be a tendency to keep to the original configuration. Changes which leave this only slightly changed will occur very much more readily than those which require serious alteration." In general conformity with this, and somewhat contrary to the specific reference of A. J. Rabinowitsch, Dankov [65] has discussed both the formation of the latent image and its development. He postulates "the process

$$\text{crystal}\left[\overset{+}{A}g\,\bar{B}r\right] + h\nu = \text{crystal}\left[Ag\right] + \bar{B}r$$

should go on in that most simple direction which ensures a minimum change of the orientation of the silver atoms (ions)." This involves "minimum translocation of the atoms during the process." He goes on to show that *silver bromide* exhibits particular advantages in this respect. In Figure 12 it is evident that the lattice of silver bromide can transform itself into the silver lattice with hardly any change in the spatial disposition of the silver atoms. Also "the complex of cube planes (100) in the $[\overset{+}{A}g\,\bar{B}r]$ lattice will correspond to the octahedron (110) and cube (001) planes in the [Ag] lattice. At the same time, according to the principle of dimensional conformity,* the distances between the silver atoms in the cube planes (001), which retain their denomination even after the silver bromide has turned into the metal do not change at all after bromine [in the case of print-out and latent image forma-

* This is the form in which Dankov states the principle of minimum work.

tion, but of bromide ion in the case of development—interpolation by the writer] has been removed from the crystalline lattice." Simple calculation is in agreement. The diagonal of the cube plane for silver a = 4.077 Å equals $\sqrt{2(4.077)^2} = 5.75$ Å, which approaches closely to the constant of the silver bromide lattice, *viz.*, a = 5.76 Å.

There is less conformity in the case of silver chloride, where the diagonal of the cubic plane in the silver nucleus (5.76 Å) is 3.8 per cent greater than the edge of the cubic nucleus of silver chloride (5.54 Å). However, Dankov's supposition that

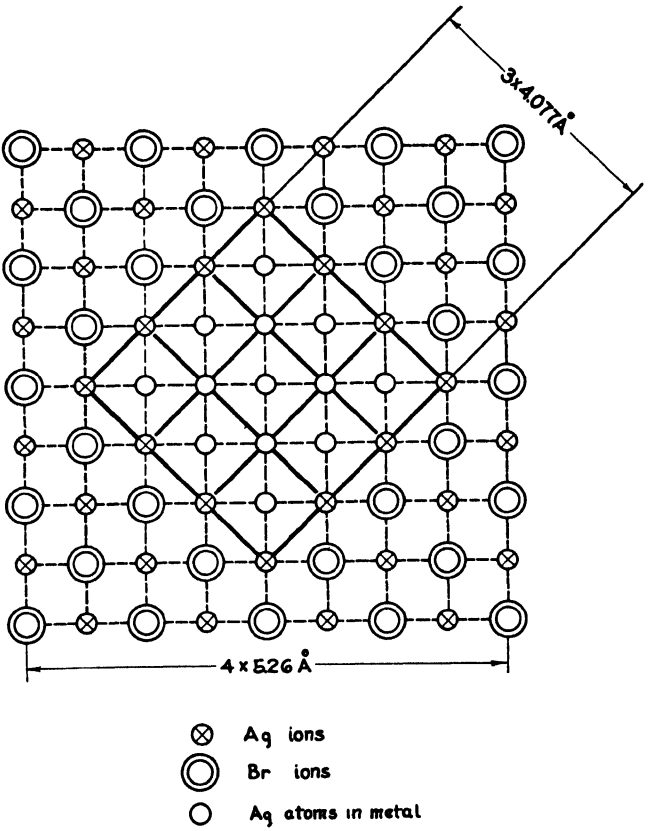


FIGURE 12.

this may account for a more restricted photographic utilization of silver chloride is unwarranted. The defects of silver chloride relative to the bromide are in the matters of spectral sensitivity, and of greater solubility and reducibility. So far as the postulate of minimum work is effective, the fact that silver chloride, *ceteris paribus*, develops at an equal or greater rate than silver bromide shows that some compensating factor or factors must intervene. This factor we consider to be the easier removal of the halide ion. All the recent hypotheses on chemical development, which predicate reaction in the solid state, have concentrated on the question of reduction of the silver ion, with little attention to the equally important matter of the removal of the halide ion. Where some attention has been paid to this, it has been by way of rather *ad hoc* suppositions, most of which involved larger motions

of the halide ions in the lattice than are consistent with the data on the conductance of the ions in the solid silver halides. These, as is well known, indicate that movement in an electric field is practically confined to the silver ion.

Thermal Energy Balance

To obtain a value for the "heat of development" the simplest procedure is to regard the process as composed of two steps, *viz.* (a) solution (b) reduction of hydrated silver ion to metallic silver. Whatever the mechanism, the products are metallic silver and halide anion. The unknown quantity is then the heat of reduction of the silver ion.

For three silver halides we have for (a)—heat of solution:

AgCl	15 k cals
AgBr	20 k cals
AgI	27 k cals

For (b) values have been obtained from the reaction

$$2\overset{+}{A}g + 2NH_2OH = [Ag_2] + N_2 + 2H_2O + 2\overset{+}{H}$$

Approximate calculation of the heats involved gave about 92 k cal/mole. This figure is likely to be somewhat high because certain specific heat effects were neglected. On the other hand, imperfectly corrected experimental determinations gave 72 k cal/mole. This is probably low, owing to disregarded heat losses, evaporation, etc. Pending values for other developers, it appears that for silver bromide, with a "heat of solution" of 20 k cals, there should be a net heat balance of some 50 k cals available per mole AgBr on reduction by hydroxylamine. A supposedly more energetic developer, as hydroquinone, should afford more heat on reaction. The calculated value, for one equivalent of silver ion was found to be 20 k cals and experimentally 28 k cals. Again subtracting the heat of solution of silver bromide, we have a favorable balance possible of *ca* 8 k cals.

These are larger amounts than postulated by Weissberger;[23] it remains to consider how they compare with "activation energies" likely to be important in the crystal.

The "activation energies" in question would be in the first place those for mobilization of the silver *and* halide ions—at any rate from the point of view of processes in the solid. The maximum possible values for these would be those deducible from the theory of the electrostatic potential energies in the lattice[66] uncorrected for polarization and other factors. These are for cubic (B1) type silver bromide:—

Silver ion 13.2×10^{-12} erg per atom or 200 k cal/mole

Bromide ion 42×10^{-11} " " " " 640 " "

These "limiting values" are all undoubtedly very much too high, because no account is taken of polarization and other factors. Experimentally, estimates of the activation energy, assumedly for the silver ion, are obtainable from the electrolytic conductance. If we assume[67] that the activation energy for the S. I. region:—

$$E = 1/2\ W + U$$

where W = activation energy for movement of $\overset{+}{A}g$ ion from a stable lattice position to an interstitial position,

U = activation energy for an interstitial silver ion,

then $W = 2\ (E - U)$,

and we have to have some measure of U to estimate W. It has been suggested that the activation energy for S. S. regions may be taken for this. We give in Table IV, some estimates for the work function or activation energy (supposed for the silver ion) for three halides. All values are in k cals/mole.

Table 4

Substance	U	E	W (a)	(b)	(c)
AgCl	8.0	20	25	22	24
AgBr	7.5	15	20	20	15
AgI	(5) *	10	—	19	(10) *

(a) E. Koch and C. Wagner, *Zeit. phys. Chem.*, B **38**, 295 (1937).
(b) H. Reinhold, *Z. Elektrochem.*, **39**, 580 (1933).
(c) Unpublished data of Sheppard, Vanselow and Happ.
* This value is based on the assumption U = 1/2 E.

Supposing that these values are more nearly correct than the "limiting values" obtained from the electrostatic potential, then the net energy balance of the reaction of developer with silver halide might well be enough in principle to support Weissberger's hypothesis. In fact, considerably more heat is then available than his tentative estimate provided. The "hot spot" hypothesis becomes more plausible, but there still remains the recalcitrant fact that the rates of development of the silver halides are *not* in the order apparently demanded by conditions for mobilization of the silver ions in the solid phase.

Another possible phase of the action which would be favored by rise of temperature is solution of silver halide in halide ion solution—locally provided by escaping halide. This will be considered in dealing with "fine grain" development.

These purely thermochemical considerations only indicate the maximum work possible in a process and not the true thermodynamic limiting conditions. However, the probability of a large net surplus of energy available for "activation" is roughly indicated by them, and, therefore, of localized or "hot spot" effects. Colloid chemistry has been defined by Bancroft as the chemistry of interfaces, of bubbles and filaments.[68] The colloid chemist knows of conditions where a reaction becomes evidently periodic in space, *e.g.*, in Liesegang ring formation. There are also examples of reactions becoming quasi-periodic or pulsating in time, such as the decomposition of hydrogen peroxide at a mercury surface observed by Bredig. In this case there is a periodic making and breaking of an oxide film, a transient "bubble." There are, moreover, other instances of filament formation in physical and chemical processes. Many substances crystallizing out in a colloid, such as a nitrocellulose film, do so, by the constraint of a rigid foreign medium, in plumes of spirally curved filaments or trichites.[69] * However, in such cases the whole medium is supersaturated with the material of the crystals. Much more similar to the present case is the "growth" of alumina filaments from pure aluminum specked with a mercuric salt.[70] † This is particularly interesting, because it is an oxidation-corrosion, whereas development is a reduction-corrosion. Also, it has been observed that mercury globules in an atmosphere of iodine vapor appear to eject filaments of mercuric iodide.[71] It is notable that in the great number of topochemical reactions, in a solid phase, as they were termed by Kohlschütter [72] there is but little mention of filamentation, the formation of "hair silver" by the reduction of silver sulfide by hydrogen at rather high temperatures (400°-500° C.) being the most noteworthy. It is possible that electron microscope studies of many of these, *e.g.*, the thermal decomposition of silver and mercurous oxalates, of silver and other azides, might reveal new examples.

The production of dendritic growths in electrolysis is well known, as also that of "inorganic vegetation" in the growth of precipitates in silicate solutions. Speaking broadly, these are all instances of a general principle. They tend to occur when the "ordering" of a new solid phase (in its state of minimum potential energy) is unable to keep pace with the rate of condensation (precipitation) [69] and particularly

* Various papers in earlier volumes of this series discuss the effects of colloids on crystallization. See also paper by W. G. France in this volume. Editor.
† See paper on "grown" alumina by H. Wislicenus in Vol I of this series. Editor.

in the presence of a mechanically resisting medium.* Examples of this in the recrystallization of silver bromide in gelatino-silver bromide layers have been given by Trivelli and Sheppard.[69]

It is likely that such a peculiar somatic modification as filament production in the chemical development of silver bromide—or rather, of silver iodo-bromide crystals—is the resultant of several conspiring factors. If we are not satisfied with the simple dualistic hypothesis of Gurney and Mott (*v. inf.*) mainly because it makes no provision for the exodus of halide ions from the solid crystal, but also because it requires that the reducer adsorb to metallic silver specks, and donate electrons, without considering that an oxidation product, *e.g.*, quinone, will equally attach itself and accept electrons from a silver speck,† we seem forced to look for some other operative process in the solid silver halide. Moreover, it would seem that such a process should have the possibility of a *pulsating* character in time, and, probably of a mechanically disintegrating character on the crystal. Such a process may be derived perhaps from a new theory of latent image formation and its chemical development due to M. L. Huggins.[73] The writer has had the privilege of a preview of Dr. Huggins' presentation of his theory. No more than the writer is in entire agreement with the conclusions which Dr. Huggins draws as to the latent image and its development, no more will the latter be as to certain consequences educed here. The foundation which Dr. Huggins presents we consider to be very truly laid and of great importance. Briefly, he shows that in solid silver bromide, which is found to crystallize predominantly in the B1 (NaCl, cubic) type, there exists the possibility of easy transition, under given conditions, to a B3—B4 (ZnS) type. This structure is shown to be only slightly higher in energy—*i.e.*, lower in stability—*per se* than the B1 type, with which it is *tautomeric*, to speak in the language of organic chemistry. The argument for this must be considered in Huggins' own papers; we shall consider only certain aspects and consequences. It should be said at first that a greater tendency for silver bromide to adopt this structure than for silver chloride is consistent with the conclusion that the former shows a greater degree of covalent bonding or of homopolar valency, than the latter.[74] The writer has indicated elsewhere the possible importance in the photographic process with silver halides of the transition from the ionic (electrostatic) to this other type of bonding. Huggins points out that the probability of a transition B1→B3—B4 would be increased by the presence of a small percentage of iodide, such as is usually introduced into gelatino-silver bromide emulsions.‡ A very important property of the B3—B4 structure brought out by Huggins is this. It contains "holes" of positive potential and "holes" of negative potential (cf. Fig. 13). He points out that "the former are possible sites for extra electrons, *e.g.*, photoelectrons—an electron in such a site stabilizes the structure by at least 1.10^{-12} erg or 0.6 e v." The presence of extra electrons favors the transition from a B1 to a B4—B3 structure. Huggins suggests that "the latent image consists of a small region of the crystal or crystal surface in which the atomic arrangement is—the B3, or cubic ZnS type—stabilized by the presence of extra electrons (photoelectrons)." In this writer's view, which on this point is not to be regarded as identical with that of Dr. Huggins—this region might be regarded as an aureole surrounding or adjacent to a sensitivity speck of silver sulfide, or its equivalent of other "noble" sulfide or metal. In the "orientation hypothesis" of the concentration speck the writer has suggested [55] that the speck is surrounded by a potential trough tending to attract electron-wavelets from any part of the crystal absorbing light. The Gurney-Mott hypothesis proposed that the photoelectrons migrated to the speck itself,

* One wonders what form H. G. Wells' time traveller and his machine might have taken if he had "precipitated" in some ancient volcanic upthrust long forgotten, or in the Tower of London. (Cf. H. G. Wells, "The Time Machine.")

† Electronically the condition of a silver electrode in a redox medium.

‡ From 1 to 5 per cent.

giving it an excess negative charge which attracted interstitial silver ions. Huggins' suggestion appears to embody important features of both these hypotheses, and to be superior to either.*

Let us return, prior to the latent image, to the formation of a sensitizing speck, *e.g.*, of silver sulfide, in a silver halide crystal. This is produced by the breakdown of

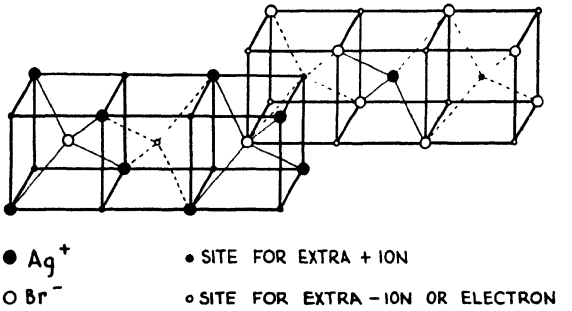


FIGURE 13. Extra charge capacity of Huggins' B3 structure for silver bromide.

an antecedent organo-silver halide complex[53] with elimination of halide ions and of organic radicles. It occupies, therefore, less volume than the earlier complex, and indeed than the silver halide which it replaces. In these relatively violent transformations it is reasonable to suppose that contiguous silver bromide would pass from the B1 to the B3—B4 structure—of somewhat higher energy, and specifically in the less stable structure, having a volume about 13 per cent greater than that of an equivalent

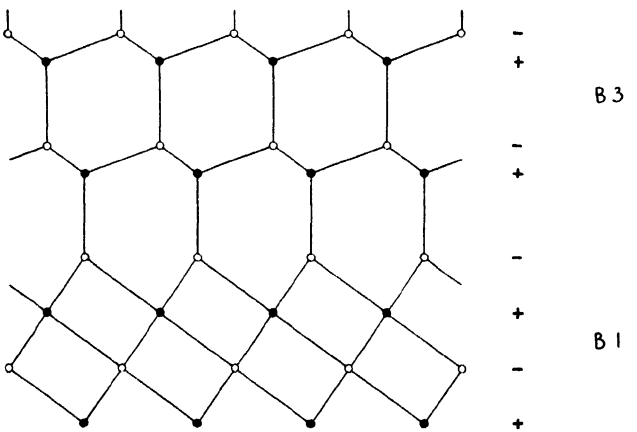


FIGURE 14. Continuity of B1 and B3 lattices of silver bromide.

amount of B1. The intercalation of such a region with neighboring B1 is fully described by Huggins; in particular, it may be noted that the bromide ion distribution would be continuous throughout the whole, "only the distribution of silver ions and extra electrons being different in the two parts." (cf. Fig. 14). This zone or aureole of B3 structure adherent to a speck is the destined trap for photoelectrons on ir-

* An important aspect of the "orientation hypothesis" was that it allowed for more than a chance accretion of photolytic reaction product around the speck. What it did not provide *per se*, which the G.—M. hypothesis did, was an attraction center for silver ions.

radiation of the grain. Thereby the dimensions of the structure would tend to contract to those of the adjacent B1 structure. We may suppose, however, that it is at least equally possible for the electrons to spread back into adjacent B1, producing more B3, a possible factor in the phenomenon of "growth of the latent image."[75] In any case, a more or less stationary state can be assumed to form about the speck for a given and relatively slight illumination; we are not interested at present in considering the progression under further irradiation to a "print out" image. Huggins has pointed out that the extension of the B3 structure may well be greater in the surface of the crystal than in the interior. In general, as he shows, "the boundary between the two structures (B1 and B3) . . . could shift very readily, one atom at a time. A silver ion on the B1 side . . . may shift only slightly . . . in order to take its place as a portion of the B3 part of the crystal. This shift would require only a very small activation energy—in the neighborhood of 10^{-12} erg or 14. 6 k. cals." It is important to realize that a suitable net plane of such silver ions retaining sufficient electrons is already equivalent to a (111) plane of a silver metal crystal. In explanation of filament formation, we are inclined to insist upon the tautomerism of the structures:

$$\text{B1} + \text{electrons} \rightleftharpoons \text{B3} \rightleftharpoons \text{B3}' + \text{electrons}$$

where B3′ is the B3 structure of greater volume, *i.e.*, less extra electrons. This B3′ structure may be that which adsorbs the developer, receiving electrons; partly, as extruding into the developer (losing bromide ions), and collapsing to a silver lattice, partly handing on electrons to form more B3→B3′ from B1. The rôle of a sensitivity speck of silver sulfide, soon reinforced by metallic silver, is to centralize this sequence of events and permit its continuance. It has been forgotten that a silver or silver sulfide speck can only receive electrons in so far as these are in defect, that it will, in general, at least equally attract an oxidant, *e.g.*, quinone, as—where electronically defective—a reductant. Provided reduction takes place in a B3′ zone, somewhat extruded—in virtue of its greater volume—from the surface of the crystal but adherent to the therewith growing silver, there seems no inherent difficulty in protrusion of filaments—initially at any rate. The bromide ions are thereby brought to the surface, to undergo hydration, and escape.

Within the crystal we may suppose that the variable expansion and contraction induced by the pulsation of B3 ⇌ B3′ formation proceeds faster than any large movements of Ag^+ ions, let alone of bromide ions. A splitting of the crystal grain, admitting developer, and renewed filament formation, will ensue, the process having some resemblance to the "branching chain reactions" discussed by Semenoff[76] and instanced by W. E. Garner[77] in reference to the purely thermal decomposition of solids. It must be remembered that the silver halide grain is enclosed in a mechanically resisting integument of gelatin, which may allow at the start some penetration by filaments, but not to an unlimited amount. In many cases, oxidation products of the reaction will tan this integument, increasing the resistance. It is to this factor possibly that we owe the relative conservation of figure of developed grains.[48] But, in reverse, we must either admit inward filament formation—which requires considerable bromide ion mobility in the solid crystal, or such a splitting as we have indicated. We cannot pretend that the mechanism of filament formation in the chemical development of gelatino-silver (iodo) bromide grains is yet fully elucidated. We do consider, however, that the theory of M. L. Huggins is likely to give the most satisfactory explanation. The reversible change of volume which it allows, according to surges of electron acceptance and donation, requires only small displacements of silver ions, of low activation energy. At the same time, this expansion and contraction affords the occasion for bromide ions to hydrate and escape. Flux and re-

flux of electrons could be assisted by the silver sulfide-silver speck acting as an electrode in a redox medium—this, however, being only of secondary moment.

Before leaving this aspect, it is necessary to consider the cases of silver chloride and silver iodide. The prospects for the B1→B3⇌B4 transformation with the former appear considerably less promising than for silver bromide. The electric conductance of AgCl in the S. I. region is considerably less than for AgBr, and the activation energy is higher. However, it seems to exist at temperatures below 200° C. chiefly in the S. S. state, having a much higher conductance—*i.e.*, mobility of silver ions, and lower activation energy, conditions assuredly favoring the possibility of a B1→B3⇌B4 tautomerism. Over and above this, however, we must consider that reduction from aqueous solution may play a considerable part with silver chloride. Its solubility in water is some hundred times as great as for silver bromide, and the tendency to covalent binding much less. The energy of hydration of the chloride ion is also greater, and these factors probably account for the more ready reducibility and developability of silver chloride.

The case of silver iodide is quite different. This crystallizes, at ordinary temperatures, in either of the two closely related "tetrahedral" structure, B3 (cubic) and B4 (hexagonal) in each of which each ion is surrounded tetrahedrally by four of the other kind. Huggins suggests that "the preference of the iodide for the tetrahedral structures may be explained as due to the greater tendency of iodine (than of bromine) to form covalent bonds with silver. Electron pair bonds between all adjacent atoms are possible in the B3 and B4 structures, but not in the B1 structure." With this preëxisting B3⇌B4 structure, it might be thought, according to the previous argument, that silver iodide should be even more readily chemically developable than silver bromide, whereas the case is quite the contrary. Pending a further investigation by Huggins of the energetic conditions for redistribution of electrons and ions in silver iodide, we would suggest the following aspect of the stability of silver iodide. It may be considered that the iodine atom-ions of the structure resonate between these configurations:—covalently bonded iodine atom, electrostatic iodide ion, and one-electron bonded iodine with the other electron in a "hole" of positive potential. This is in agreement with the behavior of silver iodide in regard to light (photolysis) and in regard to reduction by developers. There would be less incentive to the formation and reception of foreign electrons, and to any change of lattice dimensions. Whatever the fate of Huggins' theory, it is a tantalizing thought that there are questions which only the colloid chemist can put, and which only the crystallographer can answer.

Physical and Fine Grain Developments

The relatively early differentiation between "physical" development and "chemical" development had in it somewhat of the same logical, *i.e.*, verbal contrast, as that between the physical and the metaphysical, the latter being what was beyond or outside the physical. It may be remembered that physics was already an embryonic science when chemistry was still mostly a mystery, that is, alchemy. With the advent of physical chemistry, it was assumed that (cf. p. 475) the processes were essentially the same, the silver halide going into aqueous solution to be reduced. Then later, with chemical physics arising, the differentiation became reasserted, with increasing insistence that in "chemical" development the actual reduction of silver ions to silver and their aggregation to metallic silver took place within the solid phase, the silver halide crystal. We have discussed some of the possibilities and some of the difficulties of this conception. The "adsorption theory" of development seemed to offer a sort of middle way out of the difficulties by making the interface between the reducing solution and the solid silver halide grain the actual locus of the reaction—or rather the triple interface, silver halide: silver metal: developer. As we have seen, this conception does not explain too readily the revelations of the electron microscope,

and filament formation.* There has been a tacit assumption that "physical" development is so well understood that it would suffice for "chemical" development to be referred to that category to regard it as thereby "explained." But how, if "physical" development has its own problems? They certainly exist, photographically speaking, and perhaps, in some cases at least, chemically.

We shall return to the photographic aspect, meanwhile, let us consider again the simple matter of "physical" development by ferrous salts.

There is generally very rapid—so-called instantaneous—reaction, between "ions" in aqueous solution. But this is not always the case.[78] Thus the reaction between ferrous ions and silver ions in aqueous solution shows a definite induction. This was attributed by Roberts and Soper [79] to the time required for supersaturation with silver atoms. There are other possible causes of induction. In this case the reaction $(Fe^{++})_{aq} + (Ag^{+})_{aq} \rightarrow (Fe^{+++})_{aq} + [Ag]_{met}$ is saddled with a very considerable potential barrier. This is caused by the similarity of the sign of the charge, and the hydration of the ions. The free energy of hydration of the silver ion, 95.5 *k cals,* is very considerable, while that of the ferrous ion is probably of the order of 400 *k cals.* In both cases the water molecules of their hydrospheres will be strongly polarized and similarly oriented. There is a condition similar to that indicated by James as causing an induction period in chemical development. The problem for the ions is how are they going to get close enough together for the ferrous ion to hand over an electron to the silver ion. We disagree with Roberts and Soper's view that this phase of the reaction is *de jure* and *de facto* "instantaneous." We have always considered that the structure of ionic aqueous solutions is fundamentally micellar, due to the electrostriction of water molecules about the ions,[80] and this has been quantitatively reaffirmed in modern theories of the hydration of ions.[81] With this in mind, a tentative sketch of the process can be outlined which may indicate that it is not so unlike the process considered in "chemical" development.

The ferrous ion is not *per se* very stable, it interacts readily with hydroxyl ions to form hydrous ferrous oxide. This hydrolysis is repressed by increase of hydrogen ion, but the catalysis of silver ion reduction can hardly be due to this alone. Ferrous ions will tend to form ion-pairs, and we suggest that in the present case this involves the formation of a one-electron bonded dimer:—$(Fe^{++})_{aq} + (Fe^{++})_{aq} \rightarrow (Fe^{++} \cdot Fe^{+++})_{aq}$ which would be stabilized by charge resonance between the cores. A hydrogen ion, $\overset{+}{H}_3O$, consisting of a proton *inside* a water molecule, can penetrate readily enough the hydrosphere about the ferrous ion pair, and stabilize the complex against hydrolysis. Between the hydrosphere of such a ferrous ion complex and that of a silver ion or of a silver ion pair, there will exist chains of water molecules polymerized by hydrogen bridge formation. The initial stage of reaction or reduction we conceive to be effected as follows: A proton of the hydrion in the ferrous complex is neutralized by the lone electron of the latter, but receives a positive charge from the nearest bridge hydrogen connecting with a silver ion. By this type of "tunnel" at the other end of the chain a hydrogen atom in the silver ion pair hydrosphere takes a positive charge from a silver ion of an ion pair. There is formed another one-electron bond pair $Ag^{+} \cdot Ag^{+}$, which we may regard as a metallic silver atom adsorbed to a silver ion, or conversely. The "tunnel" process producing these may be regarded indifferently as a movement of an electron along the chain from the $(Fe^{++} \cdot Fe^{+++})_{aq}$. complex

$$\overset{+}{H}$$

to the silver ion pair, or as a movement of positive charges (positrons) from the silver ion pair to the ferrous complex; evidently it comes to the same thing, and this

* One is reminded of the conflict between the uniformitarian and the catastrophic schools of geologists in the 19th century. For the former, relatively slow and regular deposition of strata by sedimentation from aqueous suspension was all prevailing, for the latter volcanism, the upthrust and solidification of molten magma. There seems to have been room for both the isothermal and the adiabatic processes!

process we regard as constituting the primary phase of the "induction period."

The secondary phase may well consist in the coalescence of such silver ion-atom pairs to form binary and ternary nuclei of a metallic silver lattice [3] with an accompanying reduction of the hydrosphere barrier. There would even be possible a replacement or hydrogen bridges by silver ion-atom chains:—$Ag^+ \cdot Ag^+ \cdot Ag^+ \cdot Ag^+$ . . . with complete continuity between the reaction in the induction period and that in the "metallic silver" catalyzed reaction. In this stage, the hydrospheres of the ferrous ion complexes are dissipated, and ferrous, ferric and argentous ions finally establish equilibria on silver crystals. It is not necessary to suppose that the ions become naked gaseous ions at the surface of the crystal—though this must occur for silver ions incorporating in the lattice. It is sufficient that the hydrospheres become so thin that the probability of the transition discussed above becomes enormously increased, because the probability of such transitions is approximately inversely as the thickness of the barrier, and therewith the velocity of the reaction. If the structure of a liquid, of a solution, were so to say, static, instead of being only an averaged kinetic structure, generally isotropic because of the thermal energy, the chain mechanism might well give rise to filamentous forms of silver. This is disrupted because of the thermal energy, but when, as in electrolysis, a macroscopic anisotropy *is* applied to the solution, dendritic growths readily appear—so-called silver and lead "trees."

In the case of "physical" development with acid ferrous salt and silver nitrate solution, the "induction period" is abbreviated by the latent image. However, with this, and in all cases of "prefixation physical development" it is characteristic that a considerably lower "speed" or "threshold density" is manifested than that with chemical development. Partly this is because the "grain" of the reduced silver masses is finer, more compact, hence the covering power and optical density are lower. But partly also because, owing to the nature of a reduction-oxidation equilibrium, there is a high probability of the smallest nuclei being "oxidized" rather than silver ions being "reduced" upon them. This would be equally true if the "threshold" latent image nuclei consist of B3-trapped electrons as supposed by Huggins.* Correlatively, certain alkaline physical developers, to be noted shortly, give considerably higher threshold speeds than the acid ones.

Induction periods are manifest also in the reduction of silver ions by hydroxylamine and by hydroquinone. The kinetics of these reactions have been investigated by T. H. James [82] to whose studies we refer in respect of characteristic details. Broadly, however, it may be said that they exhibit induction periods, catalysis by colloidal gold, silver sulfide, colloidal silver, and therefore autocatalysis. Foreign catalytic nuclei are overlaid and replaced by silver. It is an obvious necessity for "physical" development that there be an induction period and catalysis by the latent image and by the added metallic silver. Otherwise, that topographic distribution which constitutes a photographic image would not be conserved. The reverse of this does not appear to hold. Not every silver reduction showing an induction period is effective for physical development. Again with hydroxylamine, we consider that the fundamental reason for the induction period is the low velocity of the uncatalyzed homogeneous reaction, and that this is an expression of the low transition probability for electron exchange between donor and acceptor particles. According to the pH, hydroxylamine can exist in the following forms:—

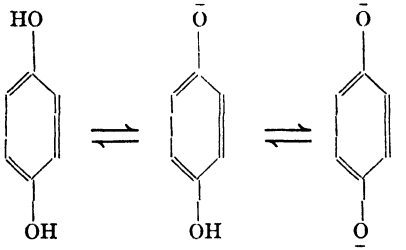
$$\underset{\longrightarrow \text{ increasing pH.}}{\overset{+}{N}H_3OH \leftrightharpoons NH_2OH(+\overset{+}{H}) \leftrightharpoons NH_2\bar{O}(+H)}$$

The equilibrium is such that even at very high pH values the $NH\bar{O}$ ion can be present only in very low concentration. Also it should be noted that, unlike the

* Photographically, the essence of a latent image is topographic. Disappearance at one point, though balanced at another, is effectively destruction or loss.

nitrate and nitrite ions, it cannot be stabilized by resonance, which would accord with relatively low activity as a developer.[83] The ammonia-like neutral molecule should have a considerable dipole moment and probably a considerable energy of hydration. It would, therefore, have a large hydration sphere, reducing the probability of encounter with $A\mathring{g}$ ions. Such encounters would occur, however, and, as in the case of ammonia, involve coördination between the silver ion and the hydroxylamine. Rearrangement in the coördination complex $A\mathring{g}(NH_2OH)$ could lead to reduction of the silver ion and the adsorption to this of a silver ion, forming $Ag^+ \cdot Ag^+$, which we have suggested as the protonucleus for formation of metallic silver. In the catalyzed reaction, both silver ions and hydroxylamine molecules would be readily adsorbed to metallic silver as suggested by James with great increase of the transition probability.

The reduction of silver ions in aqueous solution by hydroquinone over a pH range of 5.15 to 6.27 has been investigated also by T. H. James.[40] The reaction shows a very pronounced induction period. This is diminished, and the reaction catalyzed, not only by colloidal silver, gold, palladium and silver sulfide, but to a considerable extent by quinone. While the action of colloidal silver and the other nuclei might be interpreted, as by Roberts and Soper in the case of ferrous ions, as discharging supersaturation with metallic silver, this can hardly be assumed for quinone—and hence is an unwarranted assumption for the primary homogeneous reaction in any case. With increasing pH there will be a change in the hydroquinone molecule in the sense shown below:



⟶ increasing pH

At pH 5 to 6 the concentration of the monohydroquinonate ion will be very small, still less that of the dihydroquinonate ion. The latter will have no great probability of reacting with free silver ions, because of the inequality of charge number.[78] While James' results indicate that the heterogeneously catalyzed reaction is chiefly carried on by the monohydroquinonate ion, and only slightly at the lowest pH values by the neutral molecule, it is possibly this latter which is chiefly active in the homogeneous reaction. However, considerable catalysis of this is effected by quinone, and James regards this as due to the formation of a very active semiquinonate ion by interaction of the dihydroquinonate ion with quinone.

The kinetics of the silver catalyzed reaction (which shortly replaces catalysis by gold, etc.) are best accounted for, according to James, by adsorption of silver ions to the metallic nuclei followed by their reduction *in situ* chiefly by the *mono*hydroquinonate ion—the velocity of the catalyzed reaction being such that on correction for the rate of reaction between unionized hydroquinone and silver ions, a plot of the logarithm of the resultant rate against pH gives a straight line of slope 1.0.

There is a type of "physical" development which approaches "fine grain" development of negative emulsions, as well as the normal development of chloride and chlorobromide positive emulsions.* This is the reduction in alkaline solution of complex silver ions. The reduction of unprotected silver ions by alkaline organic de-

* As used for photographic papers and transparencies.

velopers is too rapid, the deposition of silver too immediate, to be photographically useful. Incidentally, this argues strongly against the view that induction periods are basically determined by silver supersaturation. It is the transition probability in the reaction between reducer and silver ion which is decisive. In the present case, where the activity of the reducer is very high, that of the silver ion must be greatly diminished to allow photographically useful induction. A photographically useful reaction is one in which the number of effective nuclei for the catalyzed reaction is substantially determined by the light exposure, no new nuclei being afforded by the reduction proper. Control of the silver ion activity in solution may be effected by formation of complex silver ions, as for example, silver sulfite

$$Ag^+ + \bar{S}O_3 \rightleftharpoons Ag\bar{S}O_3$$

As concentration increases, there are perhaps complexes of higher order formed, as happens with silver ions and thiosulfate

$$Ag\bar{S}_2O_3 + S_2\bar{O}_3 \rightleftharpoons Ag(\bar{S_2O_3})_2$$

Complex formation with the halides increases from the chloride to the bromide to the iodide. With the chloride the solubility data (of AgCl in KCl aq) permit interpretation in terms of

$$Ag^+ + 2\bar{C}l \rightleftharpoons A\bar{g}Cl_2 + \bar{C}l \rightleftharpoons AgCl_3$$

but with the bromide and the iodide this is no longer possible, and it appears that one must admit "solvation" by bromide or iodide essentially similar to the solvation of ions by water. But whether by coördination or solvation, the silver ion is increasingly protected from interaction with reducing ions or molecules. Not only is there induction, but the silver catalyzed reaction is also slowed down.

We have just considered reduction—and alkaline "physical" development—where the silver ions were introduced *ab extra*— as a prepared solution. But the solubility of the silver halides in solution of such coördinating and solvating ions is such that their presence in sufficient quantity in an alkaline (organic) developer must be considered to convert the process in part or wholly from "chemical" to "physical" development, the silver in (complex) solution being derived from the silver halide grains themselves. The principal silver halide solvent present in such developers is sodium sulfite*—though in early pyrogallol developers the alkali—ammonia—was such a solvent—and it is interesting to consider a few figures, for 25° C:—

Table V

Concentration Na_2SO_3/liter		Solubility of AgBr/liter	Solubility AgCl/liter
Grams	Moles	Grams	Grams
25	0.20	0.44	—
50	.40	.90	(2.1)
100	.80	1.90	4.4
125	1.00	(2.45)	—
200	1.60	—	9.5

It is evident that even within normal limits for organic developers (25 — 50 g./liter) the actual solubility is quite considerable. While this may not, for high alkalinity and activity, cause much direct participation, by reduction in aqueous solution, it may affect the process considerably by *etching* the surface of the crystal, and increasing surface irregularities.

* Unpublished data by A. Ballard and H. Yutzy of the Eastman Kodak Research Laboratories.

These effects become greater with increased concentration of sulfite, and decreased alkalinity and activity of the developer. On the other hand, the solubility in sulfite is decreased by increased concentration of soluble bromide. With the introduction and progress of the miniature camera, operating with 35 mm film the problem of "fine grain"—already glowing because of motion pictures—became incandescent. These midget negatives, not much larger than a postage stamp, must afford positive enlargements of say 2¼ by 3¼ inches upward. The *dispersity* of the developed silver was the controlling factor. Photographically, this dispersity falls into two levels, the basic level of individual, microscopic grain size, and a superimposed level of "graininess," which is an optical inhomogeneity visible at low magnifications, *e.g.*, 10 to 20 times. This "graininess" is a composite thing, partly due to the fact that there is a limit to the uniformity of "mixed uppedness" of a granular suspension, even of equisized grains, partly to optical superpositions in depth of any such clumped or nonuniform distributions, partly to actual infection of non-exposed by developing exposed grains.[84] The skill and art of the emulsion maker are employed to securing the best compromise in a given case of grain and graininess of the silver halide consistent with the specific requirements as to sensitivity and gradation. But it was found that something—though not everything—could be done by modification of the developer. The ultimate grain of the developed silver can be confined to the dimensions of the original silver halide particle, or even diminished, and the tendency to secondary clumping in development decreased.

Such results may be attained in various ways:

1. Increase in solvent power of the developer and reduction of its alkalinity and activity. The former may be obtained by increasing the sulfite concentration, *e.g.*, to 100 gms per liter, or by using *p-phenylene-diamine*, which is not only a reducer, but a fairly powerful solvent of silver bromide. Thiocyanate has been used also. Reduction of the activity of the reducer is effected chiefly by reducing the concentration of alkali and/or substituting a milder alkali such as borax. With such developers, increased time of development is necessary, in order that the same contrast (gamma) may be secured as with a normal developer. Comparing both developers *without* bromide * in the formulas, some loss of speed or sensitivity will be generally found with the "fine grain" developer. This is generally true, that the more the developer is modified in the direction of fineness of grain, the greater will be the loss of speed; to be compensated by increase of exposure.

2. The other procedure, which has not gained much application in practice, consists in a form of prefixation "physical" development, using an accessory silver solution and alkaline physical developer. A more direct idea of the dispersity of the developed image is obtained from photo- and electron micrographs of the grains in reduction of silver ions in solution.[48] Figures 15a, 15b and 15c show so-called "gray" silver produced by R. Liesegang and Lüppo-Cramer [85] by reduction of silver nitrate with elon (metol) and sulfuric acid. Except for a few large flat plates, it consists chiefly of small crystals, readily discernible at the highest magnifications. "Black" silver, however, produced with citric acid in place of sulfuric, could not be resolved with the optical microscope. Electron micrographs show it to consist of clusters of ultramicroscopic crystallites. (Fig. 15b)

It is probable that the difference is due primarily to the rate at which initial nuclei are produced, which rate appears to be much decreased in the presence of the stronger acid.

Related effects in the field of silver reduction, known to the colloid chemist, are the production of varicolored colloid silver by changing the amount of innoculating colloid metal, gold or silver, in a reducing system [86] and, known to the photographic chemist, the production of warm tones with gelatino-chloride papers by restrained

* Failure to observe this condition has led to erroneous conclusions as to the effect on speed at the same gamma: *e.g.*, G. Schwarz: *Phot. Korresp. Belge*, No. 1, p. 1 (1934).



FIGURE 15a. Photomicrograph of "gray" silver. b. Electron micrograph of "black" silver, showing clusters of minute silver crystals. c. Electron micrograph of clumps of silver crystals in precipitated "black" silver.

development—using ammonium carbonate and ammonium bromide—and greatly increased exposure.[87] The effect of the latter is to secure increased number of nuclei, of the former, to secure independent "physical" development of these.

A Note on Graininess and Granulation

It is perhaps unfortunate that the term "grain" had already been applied to the ultimate crystallites of the silver halide emulsion and the developed image before the coarser phenomenon of "granulation" or "graininess" received adequate recognition. Whereas the units in the former case are preëminently single crystallites from perhaps 0.001μ to 3—5μ in size, requiring either the electron microscope or the highest powers of the optical microscope for their discrimination, "graininess" is discernible at magnifications even of 10 to 30 times. It is a phenomenon above all of a statistical character—corresponding somewhat to the conception of second and third order aggregates in colloid chemistry.

Grain size distribution of *completely* chemically developed grains will tend to follow that of the grain size distribution of the original silver halide crystal,[88] although with some distortion, due to local deformations, enlargements, and coalescences. A generally valid distribution function for silver halide precipitates and simple emulsions is that for a skew probability distribution:—

$$y = Ae^{-\log (x-a)^2} dx$$

where y is the frequency of class-size x, dx or Δx is the width of the class-size unit used, a is the mean value, and A is a constant. It is not easy to test the relevance of such a function for even fully developed grains, because of (a) the irregular and indistinct shape of most grains with ordinary development (b) the tendency of many of the developed grains to break up and disperse on preparing a one-grain layer.

Again, any derived distribution for complete development ($\gamma\infty$) has only a limited relevance to the normal developed photographic image, at say $\gamma = 0.6$ to $\gamma = 1.2$, because so many will be incompletely developed. The fact is that *sensitivity* is not solely and simply proportional to grain size (or projective area) but that any class-size of grain has a two-fold variation of sensitivity due to (i) distribution of sensitivity centers, *e.g.,* silver sulfide nuclei (ii) position in depth of silver halide grains of emulsions. The latter can be plausibly represented by an empirical exponential

function [89] but while (i) is no doubt a probability function, there is no way at present of ascertaining it.

In most developers, the most sensitive grains will be the first to start to develop and therefore at a given stage of development (γ) will be giving the relatively greatest amount of silver. Since, on the whole (though with considerable variation, as stated) the largest grains are the most sensitive, the least exposed areas of the image tend to be composed of these, in diminishing proportion as the exposure increases. This is borne out by investigations on the "covering power" (C.P.) of chemically developed silver deposits.[90] The "covering power" is the reciprocal of the mass of silver per unit area which gives an optical density $D = 1$, where D is defined as $D = \log 1/T$, and T, the transparency, is the percentage light transmitted. However, since the developed image has more or less scattering power, distinction must be made between *diffuse* density measured by completely scattered light, $D_{\dagger\dagger}$, and *specular* density measured by and for parallel collimated rays $D_{\parallel}$. Only incomplete approach can be made practically to these extremes, but the usual densitometers measure approximately $D_{\dagger\dagger}$.

In the case of (chemical) development with hydroquinone James found over a considerable range of exposure when development was carried to completion a constant ratio for D/Ag. This agrees with the results of an examination of the image on ultimate development of a pure silver bromide emulsion by three different developers, *viz.*, *p*.aminophenol, pyrogallol, and metol (elon)-hydroquinone (cf. Fig. 16). It will be seen that there is no very considerable variation of the C. P. $= 1/$P.E. over the whole range of exposure. However, the significance of this must not be overestimated, since the fog values were high, notably with the pyrogallol and metol-hydroquinone, and this tends to equalize the C. P. values. With short development (Fig. 16) the C. P. shows a tendency to fall off to a shallow minimum, after which it rose slowly, approaching constancy in the overexposed region of the characteristic curve.

In a fuller investigation of covering power in relation to exposure and development [90] for two regular emulsions, (Eastman *Commercial* and *Process* films) developed with standard *p*-aminophenol developer, it was found that the C. P. at any exposure *decreases* with time of development at first rapidly, then more slowly. This is in agreement with increase in average size of the developing grain. When C. P. was plotted against the logarithm of the exposure, it was found to *increase* with exposure and approximately linearly with the logarithm (Fig. 17).

A relation between photographic density, transparency, and average projective area of (developed) grains has been obtained by P. G. Nutting.[91] According to this,* the transparency of multiple layer of n grains of average projective area a would be $T_m = (1 - a_1 n_1)\ (1 - a_2 n_2) - - - (1 - a_m n_m)$ the corresponding *absorption* $B_m = 1 - T_m$. If the layers are alike in number and size of grain,

$$T_m = (1 - a\ n)^m$$

when $D = -m \log (1 - a\ n)$. For sufficiently small values of a n this gives $D = m\ (n\ a)$ where a is the average size of grain, and P, the photometric constant, should be equal to k. d., where k is a constant and d the mean diameter of grains. Higson [92] from similar considerations, came to the same conclusion.

In a paper on "The So-called Photometric Constant" J. Eggert and A. Küster [93] determined "mean grain diameter" in an indirect manner. They had previously concluded that the Callier constant $Q \doteq D_{\parallel}/D_{\dagger\dagger}$

Where $D_{\parallel}$ = density by parallel light
and $D_{\dagger\dagger}$ = density by diffuse light

stood in a simple relation to the mean grain diameter d, which could be expressed

* Probability reasoning for the snow-flake problem.

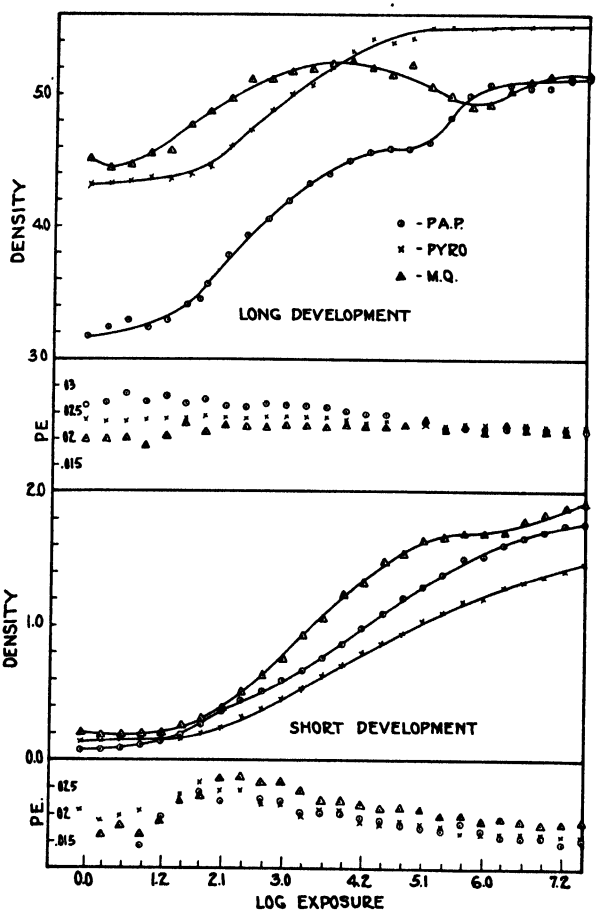


FIGURE 16. P.E.—photometric equivalent for different developers for long and short development.

as d (measured in μ) $= K \log Q$, where K was a constant for a given photometric apparatus. Using this relation to establish the "mean diameter" d of certain developed densities, they found, on measuring analytically the mass of silver, that $P = A + Bd$ where A and B are constants. This they established for different types of emulsion, and different times of development. The diameter d here considered is presumed to be the (arithmetic) mean for the ultimate silver grains. There is a rather peculiar corollary to this in respect of "graininess." In other investigations,[94] Eggert and Küster concluded that the "graininess" G_n (found subjectively by relative enlargement) could be equalled to a multiple of log Q; hence there should exist simple proportionality between "graininess" G_n and the ultimate mean grain diameter. This simple relationship has been considerably criticized chiefly in respect, however, of the asserted relation

$$G_n = 100 \log Q.$$

A physical measure of "granularity" might be expected by scanning a given area of a uniformly illuminated and developed layer with a microphotometric device of

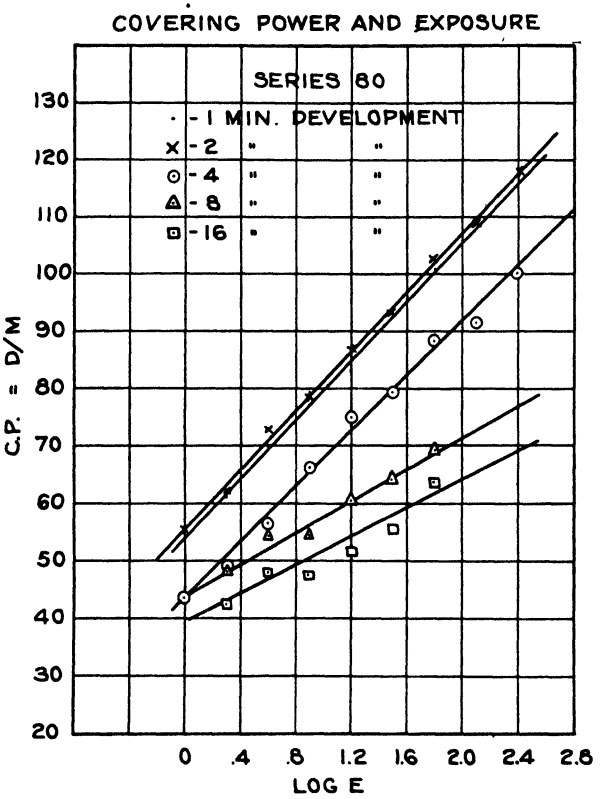


FIGURE 17. C.P.—covering power, as function of exposure and time of development.

sufficient resolving power.[95] Assumption of a Gaussian probability function for representing the transparency fluctuations could lead to a "granularity" constant expressing a standard deviation (Fig. 18).
If the distribution is Gaussian, then

$$\Delta T/T_m = x,$$

the relative transparency fluctuation, is expressible by an error function of the form

$$(a/G_a\sqrt{\pi})\int_0^x \exp. - (x/G_a)^2 dx$$

with the single parameter G_a measuring the "granularity." Actually, the distribution of fluctuations appears to be somewhat skew as in the case with the distribution of grain-size.



FIGURE 18. Micro-densitometric fluctuations of developed image.

The relations between the physically measured granularity and the subjectively perceived "graininess" are outside our subject. It may be pointed out, however, that "graininess," as a visual perception, is conditioned by various factors such as field area and field brightness, contrast and so forth. The psychological integration of the various and variously modified thresholds to a "gestalt" is unlikely to be representable by a physically measurable single parameter, nor even by an explicit function of physically defined variables. The position is similar to that discussed recently by G. W. Scott Blair,[96] for certain tactomuscular perceptions, such as "firmness" and "plasticity."

In conclusion there is some interest in contrasting the simplest form of "physical" development, with the simplest form of "chemical" development. The former we have noted as daguerreotype. It consists in the preferential adsorption of mercury atoms to the "latent image" of silver iodide on silver metal. It is hardly doubtful that we have in this case preferential adsorption to silver nuclei, the subjacent silver base having accepted the iodine photolytically released. The phenomena of mercury vapor sensitization[96a] and mercury vapor intensification of the latent image on gelatino-silver bromide are continuous with this, and indicate a direct bridge between this simplest type of "physical" development and the complications of chemical development. It must be assumed, in this latter case, that adsorbed *mercury atoms* form an effective catalyst for chemical development. That these mercury atoms are oriented and polarized by the original nucleus surface is probable[97] but equally must it be concluded that they deform and activate silver ions of the adjacent silver halide, and that it is just here that the external developer commences its attack. Very possibly this is concomitant with an extension of Huggins' B3 — 4 region in the silver chloride or bromide; this may be an important part of what is somewhat glozed over by the term "activation of the silver ion"—which would consist largely in its loosening from the constraint of the B1 lattice.

The smallest form of "chemical" development is exhibited in the *thermal* development of various endothermic compounds, such as silver oxalate, thallous oxalate, mercurous arsenite, and mercurous oxalate. These compounds do not absorb in the visible region, but after exposure in the near ultraviolet a latent image may be developed by heat alone, the rate of development increasing with the temperature.[98] There is no extraneous substance implicated, neither from a gaseous phase nor from a solution. The development or decomposition proceeds at the interface between the metal and the undecomposed crystal, the ion or radicle C_2O_4 disappearing as CO_2—gaseous carbon dioxide. The example of mercurous oxalate is perhaps outstanding, since there can be no question of a new lattice of mercury atoms—no more than in the comparable case of the chemical development of mercurous chloride.[99] There is little or no evidence for migration of mercury ions, nor of silver ions with silver oxalate. The probabilities are definitely in favor of movement of mercury or silver *atoms* in the extension or prolongation of the interface, both from the very poor or negligible electrolytic conductivities of these substances, and from the absence of any internal or impressed e.m.f.* Although migration of ions seems not to be necessary for development to occur in the solid crystalline phase, this is not to say that it may not occur to some extent in the silver halides, where the conditions may be more favorable for that mode of decomposition. So far as we know, there is nothing in thermodynamics which prescribes the kinetics or the path by which an inevitable increase of entropy shall be secured—nor the time in which it shall be attained. From this point of view, the chemistry and physics of colloids have to do

* This simplification of "chemical" development need not exclude the scheme of James and Kornfeld[100] for the reaction-complex in "chemical" development. In terms of this, D, the developer, is here the radicle or ion = C_2O_4, which is the equivalent of the various complex forming ions or molecules previously discussed, such as the di-hydroquinonate ion, hydroxylamine, phenylene diamine, ferroöxalate anion.

in the main with systems transitorily arrested in the attainment of the ultimate thermodynamic nirvana. Photographic development is an enlightening example.

Acknowledgment

Beside acknowledgment to many colleagues mentioned in this article, I wish also to express my thanks to Dr. R. H. Lambert of the Eastman Kodak Research Laboratories for assistance in the determinations of heats of development, to Dr. C. E. Kenneth Mees and the editors of various journals for permission to reproduce certain figures from papers published from the Eastman Laboratories.

References

1. I. Langmuir, *J. Am. Chem. Soc.*, **40**, 1368 (1918).
2. W. Reinders and L. Hamburger, *Z. wiss. Phot.*, **31**, 32 (1932); *ibid.*, **31**, 265 (1933); *Rec. trav. chim.*, **56**, 985 (1937).
3. H. S. Taylor, H. Eyring and A. Sherman, *J. Chem. Phys.*, **1**, 68 (1933).
4. S. E. Sheppard, *Phot. J.*, **55**, 313 (1931); also *Ber. VIII, Internat. Kongr. Photog.*, (Dresden, 1931), p. 13.
   J. H. Webb, *J. O. S. A.*, **29**, 309 (1939).
5. E. R. Bullock, "The Latent Image" (Eastman Kodak Company Monographs on the Theory of Photography, No. 6, D. Van Nostrand Company, New York, 1927).
6. T. H. James, *J. Am Chem. Soc.*, **61**, 2379 (1939); *J. Phys. Chem.*, **45**, 223 (1931).
7. E. Manegold and R. Hofman, *Koll Zeit.*, **50**, 22 (1930); R. Zsigmondy in Vol. I of this series.
8. Maj. Russell, *Cf.* "The Evolution of Photography" p. 117 by J. Werge (Piper & Carter, London, 1890).
9. R. L. Maddox, *ibid*, p. 91.
10. *Cf.* T. H. James and G. Kornfeld, *Chem. Rev.*, **30**, 1 (1942); also Chap. IX in "The Theory of the Photographic Process," C. E. K. Mees (The Macmillan Co., N. Y., 1942).
11. *Cf.* S. E. Sheppard, *J. Chem. Soc*, **87**, 1311 (1905); *ibid*, **89**, 530 (1906).
    A. H. Nietz, "The Theory of Development" (Eastman Kodak Company Monographs on the Theory of Photography, No. 2, D. Van Nostrand Company, New York, 1922).
    W. Reinders, *J. Phys. Chem.*, **38**, 783 (1934).
    R. M. Evans and W. T. Hanson, Jr., *Phot. J.*, **77**, 497 (1937).
    The later citations illustrate the vitality of this delusion concerning redox potentials of developers. It may be doubted that even the investigations of A. E. Cameron, (*J. Phys. Chem.*, **42**, 629 (1938)) have given it the *coup de grace*.
12. Meeting of the American Chemical Society, Rochester, N. Y., 1913, and *Trans. Farad. Soc.*, **19**, 243 (1923).
13. S. E. Sheppard, *Phot. J.*, **59**, 135 (1919); S. E. Sheppard and G. Meyer, *J. Am. Chem. Soc.*, **42**, 689 (1920); *cf.* also P. Wulff and K. Seidl, *Z. wiss Phot.*, **28**, 239 (1930).
14. A. J. Rabinowitsch and S. Peissakhowitsch, *Z. wiss. Phot.*, **33**, 94 (1934); A. J. Rabinowitsch, *Trans. Farad. Soc.*, **34**, 920 (1938).
15. S. E. Sheppard, A. Ballard, and E. S. Perry, *J. Am. Chem. Soc.*, **63**, 2357 (1941).
16. R. W. Gurney and N. F. Mott, *Proc. Roy. Soc., A* **164**, 151 (1938);
    N. F. Mott, *Trans. Farad. Soc.*, **34**, 822 (1938).
17. Described by C. E. K. Mees; *Trans. S. M. P. E.*, **37**, 18-19 (1941).
18. A. P. H. Trivelli, in *Amer. Ann. of Phot.*, **42**, 179 (1928).
    T. Yano, *J. Soc. Chem. Ind., Japan* (Supp. Bind.), **43**, 44B (1940).
19. M. von Ardenne, *Z. angew. Phot.*, **2**, 14 (1940).
    C. E. Hall and A. L. Schoen, *J. Optical Soc. Am.*, **31**, 281 (1941).
20. To be published.
21. *Cf.* W. Jost, *Trans. Farad. Soc.*, **34**, 860 (1938).
22. J. H. Webb, *J. Optical Soc. Am.*, **32**, 299 (1942).
23. A. Weissberger, Chap. VIII of "The Theory of the Photographic Process" (C. E. K. Mees, published by Macmillan Co., New York, 1942).
24. S. E. Sheppard, R. C. Houck and C. Dittmar, *J. Phys. Chem.*, **44**, 185 (1940).
25. S. E. Sheppard and J. G. McNally, *Colloid Symposium Monograph*, **7**, 17 (1929).
26. S. E. Sheppard, R. C. Houck and C. Dittmar, *J. Phys. Chem.*, **46**, 158 (1942).
27. Lüppo-Cramer, *Kolloid Zeit.*, **18**, 163 (1916).
28. Luppo-Cramer, "Die Grundlagen d. Photog. Negativverfahung," 3rd Edit., Chap. I (W. Knapp, Halle) 1927.
29. G. Schwarz and F. Ungar, *Z. wiss. Phot.*, **31**, 77 (1932).
30. J. E. de Langhe, *Z. wiss. Phot.*, **35**, 201 (1936).
31. T. H. James, *J. Phys. Chem.*, **43**, 701 (1939).
32. S. E. Sheppard, *J. Chem. Soc.*, **39**, 53 (1906); S. E. Sheppard and F. A. Elliott, *Trans. Faraday Soc.*, **19**, 355 (1923).

32a. F. F. Renwick, *J. Soc. Chem. Ind.*, **156T**, 39 (1920).

33. H. Staude, *Fot. Rundsch.*, **72**, 390 (1935); *Z. wiss. Phot.*, **38**, 65 (1939); *ibid.*, **38**, 137 (1939).
34. S. E. Sheppard and F. A. Elliott, *Trans. Farad. Soc.*, **19**, 355 (1923); *J. Frank. Inst.*, **195**, 211 (1923).
35. Lüppo-Cramer, *Trans. Farad. Soc.*, **19**, 389 (1923).
36. S. E. Sheppard, *Phot. J.*, **69**, 37, 206 (1929).
37. S. E. Sheppard, R. H. Lambert and R. D. Walker, *J. Chem. Phys.*, **7**, 265 (1939).
38. W. B. Hardy, *Proc. Roy. Soc.*, A **86**, 634 (1912).
    W. D. Harkins, *J. Am. Chem. Soc.*, **39**, 354, 541 (1917).
39. I. Langmuir, *J. Am. Chem. Soc.*, **39**, 1848 (1917).
40. T. H. James, *J. Am. Chem. Soc.*, **62**, 536 (1940); *ibid.*, **62**, 1649 (1940).
41. S. E. Sheppard and S. S. Sweet, *J. Am. Chem. Soc.*, **44**, 2797 (1922).
42. A. Lottermoser and R. Stendel, *Kolloid Zeit.*, **83**, 47 (1938).
43. Lüppo-Cramer, *Kolloid Zeit.*, **18**, 163 (1916); *ibid*, **19**, 17 (1916).
44. Lüppo-Cramer, *Trans. Farad. Soc.*, **19**, 389 (1923).
45. S. E. Sheppard and H. Hudson, *Phot. J.*, **67**, 359 (1927).

46. S. E. Sheppard, *Ber. VIII Internat. Kongr. Photog.*, p. 24 (1931, Dresden).
47. S. E. Sheppard and A. P. H. Trivelli, *Ber. VIII Internat. Kongr. Photog.*, p. 113 (1931, Dresden).
48. E. E. Jelley, *J. Phot. Soc. Amer.* **8**, 283 (1942).
49. M. B. Hodgson, *J. Frank. Inst.*, **184**, 705 (1917).
50. The Svedberg, *Z. wiss. Phot.*, **20**, 36 (1920).
The Svedberg and H. Anderson, *Phot. J.*, **61**, 325 (1921).
50a. S. E. Sheppard, A. P. H. Trivelli and E. P. Wightman, *Phot. J.*, **65**, 134 (1925).
51. L. Silberstein, *Phil. Mag.*, **44**, 257 (1922).
52. F. C. Toy, *Phil. Mag.*, **45**, 715 (1923).
W. Clark, *Phot. J.*, **63**, 230 (1923).
S. E. Sheppard, A. P. H. Trivelli and E. P. Wightman, *Trans. Farad. Soc.*, **19**, 296 (1923).
53. S. E. Sheppard, *Coll. Symp. Monog.*, **3**, 75 (1925); *Phot. J.*, **65**, 380 (1925).
54. S. E. Sheppard, A. P. H. Trivelli and R. P. Loveland, *J. Frank. Inst.*, **200**, 51 (1925).
55. S. E. Sheppard, Sixth Hurter and Driffield Memorial Lecture, *Phot. J.*, **68**, 397 (1928).
56. S. E. Sheppard, *Phot. J.*, **62**, 88 (1922).
57. S. E. Sheppard and A. P. H. Trivelli, *Phot. J.*, **61**, 400 (1921).
58. A. P. H. Trivelli, *J. Frank. Inst.*, **205**, 111 (1928).
59. J. H. Webb and C. H. Evans, *J. O. S. A.*, **28**, 249 (1938).
60. W. F. Berg and K. Mendelssohn, *Proc. Roy. Soc.*, A **168**, 168 (1938).
W. F. Berg, *Trans. Farad. Soc.*, **35**, 445 (1939).
61. W. Scheffer, *Brit. J. Phot.*, **64**, 116, 271 (1907).
62. A. P. H. Trivelli and S. E. Sheppard, *Proc. VII Congr. Photog.*, p. 184 (1928, London).
63. W. Meidinger, *Phys. Z.*, **36**, 312 (1935).
Phot. Ind., **34**, 1305 (1936).
64. J. D. Bernal, *Trans. Farad. Soc.*, **34**, 834 (1938).
65. P. D. Dankov, *C. R. (Doklady) de l'Acad. des Sci. de l'U.R.S.S.* **24**, 773 (1939).
66. M. Born and J. E. Mayer, *Z. f. Physik*, **75**, 1 (1932).
J. E. Mayer, *J. Chem. Phys.*, **1**, 327 (1933).
67. *Cf.* W. Jost, *Trans. Farad. Soc.*, **34**, 860 (1938).
68. W. D. Bancroft: *"Applied Colloid Chemistry"* (General Theory) p. 3 (McGraw-Hill Book Co., New York, 1932).
69. *Cf.* A. P. H. Trivelli and S. E. Sheppard, "The Silver Bromide Grain of Photographic Emulsions." (Eastman Kodak Company Monograph, No. 1 on the Theory of Photography D. Van Nostrand Company, New York (1921)).
70. H. Wislicenus, *Z. angew. Chem.*, **17**, 805 (1904).
Koll. Z., **2**, 2nd Suppl. XI, (1908).
71. R. Titus, Eastman Research Laboratory, private communication.
72. V. Kohlschütter, *Koll. Z.*, **42**, 254 (1927).
73. M. L. Huggins: presented in part at St. Louis meeting of Am. Chem. Soc., April, 1941. To be published.
74. L. Pauling, "The Nature of the Chemical Bond," p. 73 (Cornell University Press, 1939).
75. G. Jausseran, *Sci. et Ind. Phot.* (1) **9**, 85, 169 (1929).
E. R. Bullock, *ibid*, (2), **1**, 124 (1930).
76. Semenoff, "Chain Reactions,"
77. W. E. Garner, *Trans. Farad. Soc.*, **34** (2), 940, 979 (1938).
78. P. A. Schaffer, *J. Phys. Chem.*, **40**, 1021 (1936).
79. D. Roberts and F. G. Soper, *J. Chem. Soc.*, 2004 (1932).
80. S. E. Sheppard, "Photochemistry," (Longmans Green and Co., London, 1914).
81. J. Webb, *J. Am. Chem. Soc.*, **48**, 2589 (1926).
82. T. H. James, *J. Am. Chem. Soc.*, **61**, 2379 (1939).
J. Phys. Chem., **43**, 701 (1939).
83. E. C. Armstrong, T. H. James and A. Weissberger, *J. Am. Chem. Soc.*, **63**, 182 (1941).
84. A. P. H. Trivelli, F. L. Righter and S. E. Sheppard, *Phot. J.*, **46**, 407 (1922).
85. R. Liesegang and Lüppo-Cramer, *Koll. Z.*, **9**, 290 (1911).
86. *Cf.* H. N. Holmes, *Laboratory Manual of Colloid Chemistry*, p. 30 (J. Wiley and Sons, 1922).
87. *Cf.* C. B. Neblette, "Photography," p. 454 (D. Van Nostrand Company, 1927).
88. *Cf.* S. E. Sheppard and A. P. H. Trivelli, *Ch.* 17 in "Die Grundlagen d. photog. Negativverfahren," by Dr. Lüppo-Cramer (W. Knapp, Halle, 1927).
89. F. E. Ross, "Physics of the Developed Photographic Image," p. 55 (Monograph No. 5 on the Theory of Photography, Eastman Kodak Company, 1924).
90. S. E. Sheppard and A. Ballard, *J. Frank. Inst.*, **206**, 659 (1928).
91. P. G. Nutting, *Phil. Mag.*, **26**, 423 (1913).
92. G. I. Higson, *Phot. J.*, **60**, 161 (1920).
93. J. Eggert and A. Küster, *Kinotechnik*, **23**, 381 (1936).
94. J. Eggert and A. Küster, *ibid*, **16**, 127 (1934).
95. E. W. H. Selwyn, *Phot. J.*, **75**, 571 (1935).
ibid, **79**, 513 (1939).
For a full discussion *cf.* Chap. XXI, "The Physics of the Developed Image" of "The Theory of the Photographic Process." (C. E. Kenneth Mees, Macmillan Co., 1942).
96. G. W. Scott-Blair and F. M. V. Coppen, *Proc. Roy. Soc.* B **128**, 109 (1939).
96a. H. Dersch and H. Dürr, *J. Soc. Mot. Pict. Eng.*, **28**, 178 (1937).
97. H. E. Ives and A. R. Olpin, *Phys. Rev.*, **34**, 117 (1929).
98. J. T. MacDonald and C. N. Hinshelwood, *J. Chem. Soc.*, **127**, 2764 (1925).
S. E. Sheppard and W. Vanselow, *J. Am. Chem. Soc.*, **52**, 3468 (1930).
S. E. Sheppard and W. Vanselow, U. S. Patents 1,976,302; 2,129,242.
99. T. H. James, *J. Am. Chem. Soc.*, **63**, 1601 (1941).
100. T. H. James and G. Kornfeld, *Chem. Rev.*, **30**, 1 (1942).

Liesegang Rings

ANDREW VAN HOOK

Department of Chemistry, Lafayette College, Easton, Pennsylvania

In 1896 R. E. Liesegang [1a] discovered that if a gelatin gel of dilute potassium chromate is inoculated with a crystal or strong solution of silver nitrate, the resulting precipitate forms in a discontinuous and periodic pattern, which form has since been known generally as "Liesegang Rings." Although Liesegang was not the first to observe such rhythmic precipitations,[2] it was his extensive study and descriptions [1b-d] that directed attention to, and stimulated much speculation and work on, this rather spectacular and impressive phenomenon.

The general conclusion of the prodigious amount of work following Liesegang's original announcement is that the phenomenon is not specific, but seems to be general for any deposition of matter under constrained conditions. Thus, almost all precipitation reactions [3] may be conducted to form a rhythmic structure. Most workers agree that a rigid medium, such as that afforded by a gel, is not necessary,[4] but it assists by providing a framework to stabilize or fix the resulting structure. Not only precipitates formed by metathesis, but crystals formed from melts or solvents,[5, 6] solids settling from suspensions,[7] coagulants of colloidal suspensions [8, 9] etc.[10, 11, 12] are liable to periodic development. In quantitative aspect the time, t_n of appearance of any band, n, and its displacement, Yn, are quite accurately designated by the expressions:

$$\frac{Yn}{\sqrt{t_n}} = k$$

and

$$\frac{Yn}{Y_{n-1}} = k'$$

These ratios and various other relations are known by various names,[17d] but are perhaps best designated as the Morse and Pierce [20] and Jablynski [29] constants respectively. Slight deviations from constancy are observed in the early and late parts of any particular experiment, but these variations may be accounted for.[17b] The abundant literature up to 1932 is presented in review by Hedges [10] and by several others.[11, 12]

With the plentiful data available on the Liesegang phenomenon, most of which is, unfortunately, only qualitative in nature, it was quite natural that no single proposed theory and mechanism was adopted universally. Hedges [10] mentions four principal theories extant at the time of his review. Since serious objections to all four of these theories are recognized, he prefers not to emphasize any one explanation, but suggests a general theory which describes and incorporates the essential features of all. He suggests that for the deposition of a rhythmic structure there must first be realized a critical condition (which he does not specify) followed by a mobilization of material forming the structure. It is apparent that this situation is common

to the following four theories and suffers by its generality and lack of exact mechanism.

(1) **Supersaturation theory of Wilhelm Ostwald.**[18] Ostwald suggested that as the diffusion wave of inoculant advanced through the solution a concentration of precipitate in excess of saturation was formed as a consequence of the tendency to supersaturate. When this concentration reached the "supersolubility," a concentration of spontaneous crystallization, the precipitation spread backward in its course to the solubility limit. Simultaneous depletion of reactant by counter-diffusion in advance of the precipitation front clears this area, thus preserving a well-defined precipitation front, and creating a clear area which must be traversed by the inoculant before the action is repeated.

(2) **Wolfgang Ostwald's theory**[14] is essentially an extension of his father's, in which the diffusion of the soluble products of reaction play a magnified role. These effects are of great significance in some cases[15] (*i.e.*, $NH_4OH + MgCl_2 \rightarrow Mg(OH)_2 + NH_4Cl$), but they may be considered as secondary. Thus this theory reduces essentially to the disturbed diffusion mechanism of the supersaturation theory.[42, 25]

(3) **The Adsorption Theory of Bradford**[10] attributes the clear spaces between the rings to the lack of reactant removed by adsorption. The objections to this theory are that adsorption of this kind has never been demonstrated;[10] many experiments are known in which adsorption effects are unlikely;[10] rhythmic precipitates may be formed when the rate of adsorption is much less than the rate of formation;[17a] and lastly, there is no necessity to invoke such a marshalling effect as adsorption to account for the clear spaces in a periodic structure.[17c]

(4) **A Modified Adsorption Theory** has been suggested by Dhar and Chatterji,[1, 18] in which they relate the periodic nature of the structure to the periodic peptizing influence of the gel on the collodial reaction product. This explanation is limited in its scope, since it is generally recognized that the phenomenon is quite independent of the medium, and also since the structures as laid down are often definitely crystalline,[19, 5, 17b] and of appreciable particle size.

Attention is thus focused on the supersaturation theory of Wilhelm Ostwald, the first proposed to account for Liesegang rings. The three principal objections which have been raised to this theory are: (a) ring formation proceeds in the presence of seeds or nuclei of the ring-forming material; (b) supersaturation, prior to ring formation, has not been demonstrated; (c) supersolubility and a definite supersolubility product, H, as postulated by Wilhelm Ostwald and confirmed by quantitative measurements of Liesegang Rings,[21] is a dubious concept.

Concerning the first objection, it was Hatschek[21] who first observed that the lead iodide bands could be generated in the presence of seeds of the same material. He reasoned that supersaturation was impossible under this condition, and hence Wilhelm Ostwald's theory was ruled out. However, it has been observed[17b, 22] that if sufficient seeds are present, ring formation is prevented. The more concentrated the reactants, the greater is the seed density required to prevent ring formation. Or, in a given circumstance with uniform seed distribution, extent of ring formation, X, (as measured by number or distance of ring formation) will be quite definite as a consequence of dilution of reactants as diffusion proceeds. Over a limited concentration range X depends upon n (seed density) in a hyperbolic manner, $nX^2 = \text{constant}$. This relation has been deduced[17c] from Wilhelm Ostwald's formal theory[20] upon ascribing a definite rate to the growth of precipitating crystals.

Bolam,[23] Yanagihara,[37] and the author[17b] have demonstrated in various direct and indirect ways that extensive supersaturation is realized in systems which form Liesegang Rings. For silver chromate in gelatin, ion products 1000-2000 times the ordinary solubility product are realized. Lead iodide and other systems in various media behave similarly.

In his original paper, and in applications[20] thereof, Wilhelm Ostwald suggest

that precipitation commenced when a very definite concentration in excess of saturation was reached. This "supersolubility" [24] was implied to be entirely analogous to the ordinary solubility, and signified the concentration above which a metastable supersaturated solution became labile or spontaneously crystallizing. A supersolubility ion product, H,—a counterpart of the ordinary solubility product—has been assigned to define this concentration.[20, 25, 17c, 26, 42]

Many objections [14, 26] have been raised to this concept of a definite, static supersolubility limit, and a more general interpretation has been substituted. This condition consists of the rapid growth of existing and/or generated nuclei.[28] This kinetic interpretation has been applied to explain the Liesegang phenomenon.[17f] In place of the supersolubility product H as definitive of crystallization, there is substituted the condition that the rate of crystallization, at any point, increases as the concentration builds up by diffusion. This rate is augmented by the presence of nuclei of the precipitant. The positions and times of appearance of the bands are correlated with the maxima, with respect to space and time respectively, of the net concentration resulting from the opposing tendencies of diffusion and crystallization. To carry through this computation, information on the rate of crystallization of the precipitate is required. Such data have been determined for silver chromate in water and gelatin,[17g, h] and the calculations performed in an approximate manner.[17b] The results confirm the usual quantitative description of the Liesegang phenomenon, and also describe the behavior in the presence of seeds. This interpretation also indicates that any effect of a gelled medium will be primarily through its effect on the rate of crystallization, since diffusion rates remain essentially unaltered. This has been confirmed qualitatively [17a] by observing the time and spacial characteristics of definite precipitations in various media. Quantitative confirmation would require the kinetic data on the rates of crystallization in the various media.[17i]

In 1934 Christiansen and Wulf indentified * the average motion of the diffusing particles with the velocity in the de Broglie equation, $\lambda = \frac{h}{mv}$, where λ is the wave length or distance between rings, v the diffusion velocity, m the mass of the diffusing particle, and h Planck's constant. Further application of Schrodinger's equation results in approximate formulas which seem to account for the observed Liesegang phenomenon, although Christiansen and Wulf in their original publication did not find the agreement too good. The discrepancy is explicable [17d] if an extrapolated number of rings, instead of the uncertain number observed, is compared with the number calculated. The agreement is also improved if one eliminates hydration of the salts used,[31] or corrects for it in the computation.[32]

This new interpretation of the Liesegang phenomenon has been extensively and successfully pursued by Shemyakin and his co-workers,[33] and others.[34, 36] The wave-mechanical analogy is complete, and the analysis profits by eliminating a detailed model and exact mechanism. Shemyakin finds that the product (λv) retains its value within close limits for all periodic reactions in gels. The interpretation is a complete indorsement of the supersaturation theory already described, with the advantage of suggesting methods of computing the various characteristic constants of a series of bands. Various secondary systems of rings [36] and distorted figures [35] have been explained qualitatively and quantitatively as interference and refraction patterns. The formal and detailed supersaturation theory is perhaps superior in explaining results when the rate of crystallization is strongly influenced by nuclei already present.

The rather common occurrence of the deposition of matter in a periodic fashion

* That is, if the Morse and Pierce and Jablynski constants are combined in the Bauer form,[17d] $\frac{\Delta y}{\sqrt{\Delta t}} = k''$, and then considered as $\frac{\Delta y}{\Delta t} \cdot \Delta y = (k'')^2$; the first factor is identified with v, and the second with λ.

has suggested some very interesting and at times fanciful correlations with the Liesegang phenomenon. An immediate and apparent example of a naturally occurring periodic structure is the agate [1c] and other geologic formations.[38] The banded structure of gallstones and other concretions, the markings on wasps' wings, the rhythmic swarming of bacilli, beet root markings, (annual) tree rings, the structure of boli or pasture droppings ("fairy rings"), etc.,[10, 17a] have all been cited as examples of

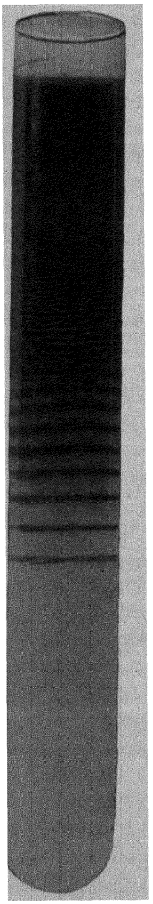




FIGURE 1. Regular bands of Silver Chromate in gelatin gel. From A. W. Thomas—*Colloid Chemistry*, N. Y., 1934.

FIGURE 2. Spiral form of Liesegang rings. Silver Chromate in gelatin gel. From A. W. Thomas—*Colloid Chemistry*, N. Y., 1934.

Liesegang rings. One cannot deny that all these structures are the result of disturbances in the generating systems; but if one takes as the criterion for the true Liesegang phenomenon the constancy of the Morse and Pierce and Jablynski ratios, then all but the agate and other geologic formations are definitely excluded from this classification. And in all these examples there is no particular need to involve the Liesegang phenomenon in explanation, since more reasonable explanations are evident.[17a]

The rather peculiar forms sometimes assumed by periodic precipitates, such as

spirals,[39] incomplete bands,[40] repeating bands,[17b] etc., may be explained as the result of particular conditions of interrupted diffusion or other disturbances. It is not unlikely that temperature or concentration gradients may cause some of these unusual effects.

A very good application of the Liesegang phenomenon has been suggested by

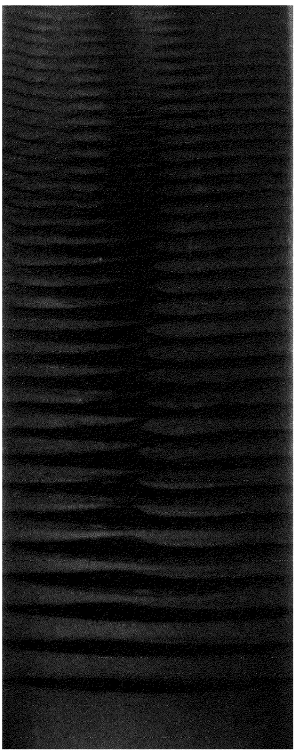


FIGURE 3. Partial bands of Cobalt Hydroxide in gelatin. T. Moeller—J. Chem. Ed. Nov. (1940) p. 519.

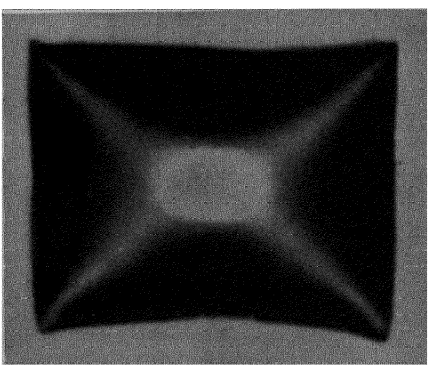


FIGURE 4. Zone pattern cut from center of ferrocyanide impregnated cake of gelatin after immersion in copper sulfate solution. E. R. Riegel and L. Widgoff—J. Phys. Chem. 29 872 (1925).

Bucher,[41] who has devised a new method of blood analysis depending on the sensitiveness of Liesegang Ring formation to very slight variations in blood composition and quality. Liesegang [10] has discussed the production of artificial pearls by diffusing alkali phosphate into a calcium salt. The writer has produced a few pieces of novelty ware by generating periodic systems in media which could be hardened and fixed by tanning.

References

1 a. Liesegang, R. E., *Phot. Archiv.*, **21**, 221 (1896).
b. —, *Chemische Reaktionen in Gallerton*, Düsseldorf, 1898.
c. —, *Geologische Diffusionen*, Dresden, 1914.
d. —, *Z. Wiss. Phot.*, **37**, 218 (1938).
e. —, *Koll. Z.*, **12**, 181 (1913).
2. Runge, G., *Der Bildungstrieb der Stuffe*, 1855 (cited by Hedges [10]).
3. Lloyd and Moravek, *J. Phys. Chem.*, **35**, 1912 (1931).
4. Many workers claim that the effect of the gel is not merely secondary in providing a rigid medium; *i.e.*,
Desai, W. B., *Current Sci.*, 5, 584 (1937).
Chatterji, A. C., "Liesegang Rings and the Influence of Media on their Formation," Lucknow, India, 1937.
Marriage, *Koll. Z.*, **11**, 1 (1912).
5. McMasters, *J. Am. Chem Soc.*, **57**, 2504 (1935).
6. Ermolenka, N. and Burankova, K., *J. Phys. Chem. (U.S.S.R.)*, **11**, 149 (1938).
7 a. Mendhall and Mason, *Proc. Nat. Acad. Sci. U. S.*, **9**, 199 (1923).
b. Ungerer, *Koll. Beih.*, **14**, 227 (1922).
8. Holker, J., *Path. and Bact.*, **25**, 522 (1922).
9. Gore, V., *Koll. Z.*, **82**, 203 (1938).
10. E. S. Hedges, "Liesegang Rings and Other Periodic Structures," Chapman and Hall, London, 1932. "Problem of Physico—Chemical Periodicity," Longmans, Green, New York, 1926.
11. Kuster, E., Über Zonenbildung in Kolloidalen Medien, Jena, 1931.
12. Veil, S., *Rev. Gen. Sci.*, **43**, 535 (1933).
13. Ostwald, Wilhelm, "Lehrbuch," 778 (1898).
14. Ostwald, Wolfgang, *Koll. Z.*, **36**, 208, 380 (1925).
15. May, *Koll. Z.*, **72**, 217 (1935).
16. Bradford, *Koll. Z.*, **30**, 364 (1922).
17 a. Van Hook, A., Unpublished experiments.
b. —, *J. Phys. Chem.*, **42**, 1191 (1938).
c. —, *Ibid.*, **42**, 1201 (1938).
d. —, *Koll. Z.*, **87**, 125 (1939).
e. —, *Northwest Science*, Dec., 1940.
f. —, *J. Phys. Chem.*, **45**, 879 (1941).
g. —, *Ibid.*, **44**, 751 (1940).
h. —, *Ibid.*, **45**, 1194 (1941).
i. —, *Northwest Science*, **13**, 45 (1939).
18. Dhar and Chatterji, *Koll. Z.*, **31**, 15 (1920); **40**, 97 (1926).
19. Dreaper, *Koll. Z.*, **14**, 163 (1914).
20. Morse and Pierce, *Z. Phys. Chem.*, **45**, 589 (1903).
21. Hatschek, *Koll. Z.*, **10**, 124 (1912).
22. Morse, *J. Phys. Chem.*, **34**, 1554 (1930).
23. Bolam, T. R. and co-worker, *Trans. Far. Soc.*, **22**, 151, 162 (1926); **24**, 463 (1928); **26**, 133 (1930); **29**, 864 (1933); **37**, 549 (1941).
24. a. Miers, H. A., *J. Inst. Metals*, **37**, 331 (1927).
b. —, Rice Inst. Pamphlet, 1919.
25. Hughes, *Koll. Z.*, **71**, 100 (1935); **72**, 216 (1936).
26. Desch, "Metallography," New York, 1939.
27. Smith, T. B., "Analytical Processes," London, 1940.
28. Avrami, M., *J. Phys. Chem.*, **7**, 1103 (1939); **8**, 212 (1940); **9**, 177 (1941).
29. Jablynski, M., *Bull. Soc. Chem.* (*4*), **33**, 1592 (1923).
30. Christiansen and Wulf, *Z. Phys. Chem.*, **26**, 187 (1934).
31. Kharamonenko, S., *Coll. J. (U.S.S.R.)*, **4**, 235 (1938).
32. Mikhalev and Shemyakin, *J. Appl. Chem. (U.S.S.R.)*, **12**, 568 (1939).
33 a. Shemyakin, *et al.*, *Koll. Z.*, **66**, 197 (1934).
b. —, *Compt. Rend. (U.S.S.R.)*, **25**, 201 (1939).
c. —, *J. Gen. Chem. (U.S.S.R.)*, **5**, 943 (1943).
34. Neumann and Costenau, *Koll. Z.*, **84**, 130 (1938).
35. Cherches, K. A., *Coll. J. (U.S.S.R.)*, **4**, 407 (1938).
36. Raman, C. V., and Ramaiak, K. S., *Proc. Acad. Sci. India*, **94**, 455, 467 (1939).
37. Yanagihara, A., *Bull. Phys. Chem. Res., Tokyo*, **19**, 1251, 1432 (1940).
38. Hartman and Dickey, *J. Phys. Chem.*, **36**, 1129 (1932).
39. *Koll. Z.*, **87**, 57 (1939).
40. Moeller, T., *J. C. Erl.*, 519 (1940).
41. Bucher, "Die Diffusion Analyze Blutplasmagel," Basle, Switzerland, 1937.
42. Fricke, R., *Z. Phys. Chem.*, **107**, 41 (1923); **124**, 359 (1936).

Fixing the Dates of Past Events by Tree Rings and Varves *

CHESTER A. REEDS

During the first four decades of the 20th century tree-ring and varved clay chronologies were developed and are now being used in dating various events. These time-tables are based on various kinds of natural evidence—different colloidal deposits developed under conditions quite remote and distinct one from another. They have the year, however, as their unit of measurement and in this respect are of importance to science. They overlap and supplement one another, in part, and have been joined to our own historical (Julian) calendar. Seasonal changes affect plants and trees and even sediments laid down under water. The tree-ring and varved clay chronologies are based on seasonal changes and represent much field and laboratory work by various men working in widely separated areas.

The growing period of trees occurs in spring and summer and their resting or maturing stage in autumn and winter. Soft white cells grow at a rapid rate in the spring. The growth is dependent upon the relative amounts of snowfall and rainfall of the preceding winter, as well as upon the nature of the soil—whether porous or compact—and its depth. In the autumn, because of lowered temperature or diminished water supply, there is a gradual cessation of the activities of the tree. This change is recorded by the deposition of denser and darker material in the cell walls. During the winter, growth practically stops. The seasonal changes are thus recorded in the annual rings of trees.

Years deficient in rainfall or lowered temperature are more noticeable and more widespread than favorable years, for a deficient year is characterized by an individual ring that is thin compared to those beside it. A group of thin rings represents a drought period. Thick rings are more likely to come in groups, and are not so extensive geographically as thin rings. Variations in climate can thus be detected in the growth rings of trees. Successive years are not alike, for variable factors, such as the annual amount of snowfall, rainfall, degree of temperature and length of growing season, affect the amount of tree growth.

The tree-ring chronology is primarily the work of Dr. A. E. Douglass, Professor of Physics and Astronomy in the University of Arizona. For many years, as an astronomer, Douglass made a study of sun-spots. There is a periodicity in this phenomenon which is reflected in tree growth. As an aid to sun-spot investigation Douglass studied trees, for, as he says, solar changes affect our weather and weather in turn affects the trees. In regions where moisture is abundant in the soil there is often noticed a beautiful rhythm of annual rings which corresponds to the sun-spot cycle of 11.4 years. Other cycles of 6 years, 22 years, 35 years and 100 years have been noted. In fact, successive centuries may have different combinations of climatic cycles.

In the pine trees of the semi-desert areas of Arizona, Douglass found evidence of

* The deposition of tree-rings and of varves (laminated deposits of glacial clays) involve colloid-chemical principles. Scientific interpretation of these deposits has revealed important chronological data. Editor

the eleven year sun-spot cycle recurring periodically for 500 years, except for the years 1650 to 1725. During this interval the tree rings afforded no evidence of sun-spots. This was separately confirmed by the English astronomer E. W. Maunder, who discovered that there were no sun-spots between 1645 and 1715.

The next tree-ring problem undertaken by Douglass was a study of the pines of northern Arizona for evidence of climatic effect of drought. This work resulted in the identification, dating and measurement of more than 100,000 rings in nearly 400 different trees. In pine trees near Prescott, Arizona, Douglass, in 1911, established cross-identification of tree rings, the yearly identity of the rings and the climatic character of their major variations. Of the climatic character he says: "The extent of area over which this extreme similarity is found may be only a half mile, as in the mountain regions near Prescott, or 50 miles as found between groups in the Sequoia region, or even more than 200 miles as shown between Flagstaff, Ariz. and Durango, Colo. Occasionally rings of extreme character are found to be alike in the Sequoias of California and in the pines of Arizona, 450 miles away. Very rare rings have been traced across 750 miles of country."

In 1915, Douglass, in response to a letter from the anthropologist Dr. Clark Wissler of the American Museum of Natural History, New York, studied some beams from cliff dweller ruins in New Mexico. These beam sections were compared with surrounding trees and from trees growing in ruins, and it was evident that pine and spruce beams were very good for this study, but that juniper was often very disappointing.

In 1918 additional beam specimens were studied: six from Aztec, N. M., and three from Pueblo Bonito, in Chaco canyon. Three or four of those from Aztec afforded cross-identifications. Later, in 1919, Douglass visited Aztec and found that the sections which he had examined had come from a pile of loose beams whose exact location in the ruin was not known. Since this was desired, Douglass devised a set of core-boring tools from steel tubing, and with this apparatus cores were obtained from some 26 beams, located in 16 different rooms, without injury to the ancient building. In time, tree-ring records of some 37 different beams from Aztec were obtained, cross-identification established and the relative dates of the cutting ascertained. He says: "Nine years covered the time of cutting the various timbers examined from Aztec." These were then compared with beam sections from Pueblo Bonito, with the result that cross-identifications between these ruins were established. Of this study Douglass says (1921): "There is no doubt that the beams in Aztec and Pueblo Bonito were living trees together during more than a hundred years and that the cutting of the timbers for Aztec followed that for Pueblo Bonito by from forty to forty-five years." While the relative dates of these two ruins were thus determined, their actual dates in our chronology were not to be announced until much further work had been done.

From 1921 to 1929 the National Geographic Society conducted eight and Douglass three expeditions in the study of the cliff-dweller ruins of the Southwest. After codifying the results of the study of tree rings in many ancient beams, some of which had been burned and were in charcoal form, Douglass, in a most interesting article in the December number of the *National Geographic Magazine* (1929), announced that he had succeeded in developing a continuous tree-ring chart for the Southwest, starting with 1900 A.D. and running back to 700 A.D., that is, for 1200 years. He also dates in terms of our own calendar the great prehistoric communal dwelling of Pueblo Bonito, containing some 300 rooms, and also some 40 other ruins in that desert-plateau-canyon country surrounding the four corners of Utah, Colorado, Arizona and New Mexico. The age of these ruins has long been a problem that the archaeologist would like to have solved, for he has found much of interest therein.

In endeavoring to find the date of construction of the communal dwellings in the various ruins, Douglass found that some of them were under construction at differ-

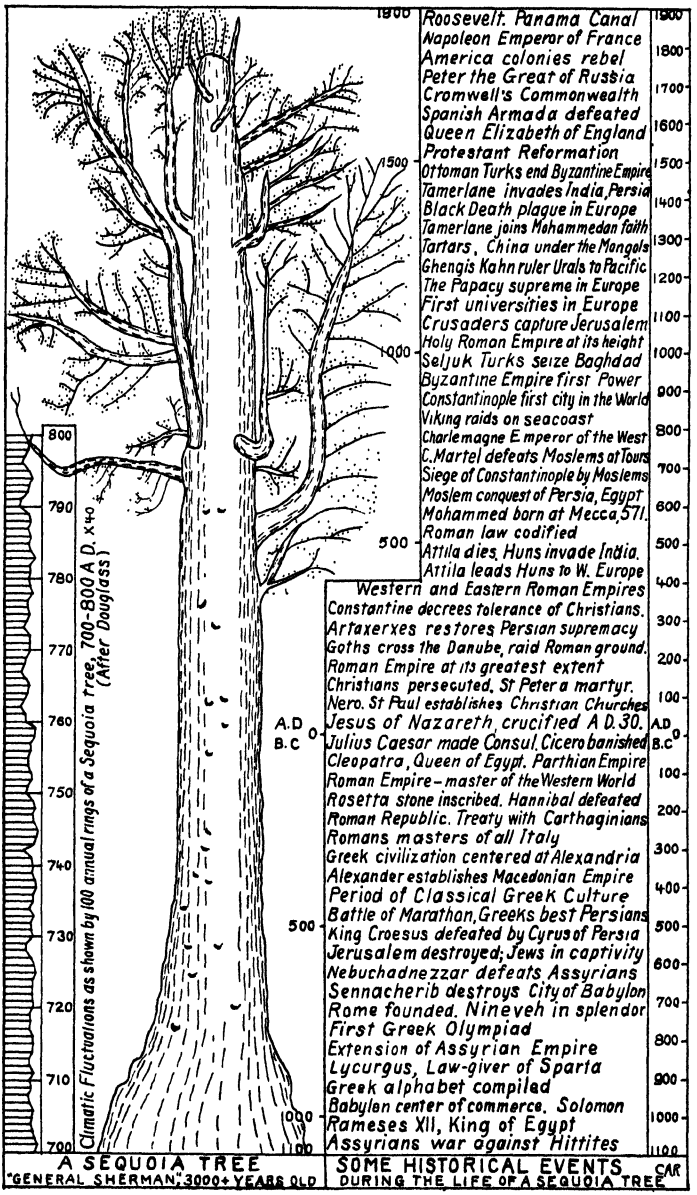


FIGURE 1.

ent times, as in Pueblo Bonito, in Chaco canyon, New Mexico. This is also the oldest of the ruins according to the tree-ring story; for one of the beams, cut in 919 A.D., shows that it started to grow as a tree in 700 A.D. Douglass gives the age of the following named ruins: Pueblo Bonito, N. Mex., 919-1130, under construction at different times: 919, 1017, 1033-1092, 1102 and 1130; Sliding Ruin, Ariz., 936-957; Pueblo del Arroyo, 1052-1103; White House Pueblo, Ariz., 1060-1096, 1219 and 1275; Mesa Verde, Colo., 1073-1262, including: Cliff Palace, 1073, Oak Tree House, 1112, Spring House, 1115, Balcony House, 1190-1206, Square Tower House,

1204, Spruce Tree House, 1216 and 1262; Wupatki, Ariz., 1087-1197; Solomon, N. Mex., 1089; Aztec, N. Mex., 1110-1121; Cliff dwellings in Grand Gulch, Utah, 1133-1135; Klag-E-Toh, Ariz., 1162; Sholow, Ariz., 1174-1383; Fewkes Ruin J, Ariz., 1192; Citadel, Ariz., 1192-1260; Pinedale, Ariz., 1200-1320; Turkey Hill Pueblo, Ariz., 1203-1278; Mummy Cave, Ariz., 1253-1284; Betatkin, Ariz., 1260-1277; Keet Seel, Ariz., 1274-1284; Kintyel (Wide Ruin), Ariz., 1275; Kawaiku, Ariz., 1284-1495; Oraibi, Ariz., 1370-1800; Chaves Pass, Ariz., 1381; Shongopovi and Shipaulovi, Ariz., 1390-1800; Kokopnyama, Ariz., 1410; Walpi, Ariz., 1427-1605.

It should be noted here that, while Douglass was fitting his tree-ring calendar together and for a long time before that, the historical period in the Southwest began with the date 1540, the year when the first Spanish expedition came into that region. In that year Indians in the Southwest were living in large communal houses just as the Hopi, Zuni, and other tribes about Santa Fé dwell in their pueblos today.

While assembling the parts of his tree-ring chronology, Douglass found that the years 840, 1067, 1276-1299, 1379, 1632 and 1880-1904 were years of excessive drought in the plateau country of the Southwest. These drought-year registrations were of great assistance to him in determining the cutting date of the numerous timbers used in the construction of the various cliff houses and communal buildings. The people who erected and lived in these abodes were primarily agriculturists with weaving and pottery making as important industries. The women were masters of the home and the men of the fields. Turquoise was not only highly prized for ornamental purposes but also as a means of exchange, and for appeasing the spirits.

The longest record of tree growth is that found in the "Big Trees" of California, the *Sequoia washingtoniana*. These trees average 275 feet in height with a diameter near the ground of 20 feet; and specimens from 300 to 320 feet tall, with trunks 25-35 feet thick, have been found. There is a moderate-sized section of one of these trees in the American Museum of Natural History, New York. The tree was cut in 1894. A count of its annual rings, most of which are about a millimeter in thickness, shows that it started to grow in 550 A.D., and was thus 1344 years old when cut. The growth rings in the stumps of some large Sequoias have been counted by A. E. Douglass and Ellsworth Huntington. Huntington counted the annual rings in 79 trees that were more than 2000 years old and in 4 that exceeded 3000 years. In regions which have climates similar to California, such as the Mediterranean, it is possible to compare the "Big Tree" record with the rise and decline of the great governments of the countries bordering the Mediterranean from about 1300 B.C. to the present. This has been done by Ellsworth Huntington for the American Museum exhibit. In his book "The Climatic Factor," Huntington makes frequent reference to the rings of growth of the great Sequoia trees, which he states give an approximate record of climatic changes that is fairly reliable for two thousand years.

A longer annual record than that afforded by the living Sequoia trees has been obtained in Sweden, and other parts of the world, from the glacial clays deposited in fresh-water lakes which bordered the retreating ice front of the last continental glacier. On close inspection these glacial clay deposits show distinct seasonal layers or bands: a summer layer, which is the thicker, of more sandy material, and of lighter color, usually gray; and a winter layer, which is the thinner, of very fine clay and of dark bluish or reddish color, depending upon the color of the rock from which the clay particles were derived. In passing upward from a dark winter layer to the succeeding coarse gray summer layer, the change is abrupt; from the summer layer to the winter layer, however, the change is gradual. The coarse summer layers have very fine lines of bedding, while the winter layers are homogeneous and uniform in appearance. The seasonal layers alternate in position without exception

throughout a clay bed. A pair of such layers representing an annual deposit is called a varve.

The clays are of glacial age, since each layer had its origin in an annual retreat stage of a melting glacier. They were developed as follows: As the ice melted and retreated slowly northward during the warm summer months of each year, the swollen rivers which flowed out from under the ice mass picked up the fine sand and clay particles and transported them to fresh-water lakes which occupied the lower portions of enclosed basins in front of the glacier. As the stream currents, on entering the still waters of the lake, gradually lost their power to transport their load of sediment, the fine sand and coarse clay particles settled down over the lake bottom to form the sandy summer layer. During the cold winter months of each year the ice-front became stationary, the englacial and subglacial stream courses either ran dry or froze, and little, if any, sediments were transported by the rivers flowing into the lake. The surface of the lake also became encrusted with ice and snow, and the fine clay particles, which had been held in suspension in the milky waters following the summer incursions, slowly settled to the bottom to form the bluish, reddish or dark winter layers composed of pure clay. Before the end of the winter season, the lake waters cleared, and a sharp line of demarcation was established between the top of the winter layer and the base of the succeeding summer layer. This well-defined line is of value in grouping the seasonal layers into varves.

Variations in the thickness of succeeding varves indicate that there were fluctuations in the amount of melt water and volume of sediment delivered to the lake basin. These features were governed no doubt by the daily, seasonal and annual changes in the weather. The varve for a particular year, however, is of approximately the same relative thickness at any designated level in the lake basin. Where there are high and low levels covered by the lake waters, the varves are thinner on the higher elevations than on the lower. Another circumstance of considerable note is that the varves overlap one another very much like shingles on a roof. This too was brought about by the amount of summer melting and the annual retreat of the ice front. The location of the limit of each varve, that is, where it touches the bed rock, enables one to determine, where practicable, the position of the ice front for a particular year, as well as the rate of retreat.

In Sweden the rate of glacial retreat was irregular—in the south in Scania and Beleking about 75 meters a year. Before reaching the two great Fennoscandian moraines near Stockholm, which represent distinctly adverse climatic conditions for a century each, it increased to 100 meters. North of the great moraines the retreat fluctuated from 100 to 300 meters or more a year, and only occasionally was it interrupted by a stoppage or small advance.

The retreat of the last glaciation in Sweden may be subdivided and summarized as follows: The dates and time given are according to Gerard De Geer, Stockholm, 1940.

(1) *Daniglacial*: Part of Denmark, part of Scania, and north central Germany south of the Baltic moraine. Time undetermined.

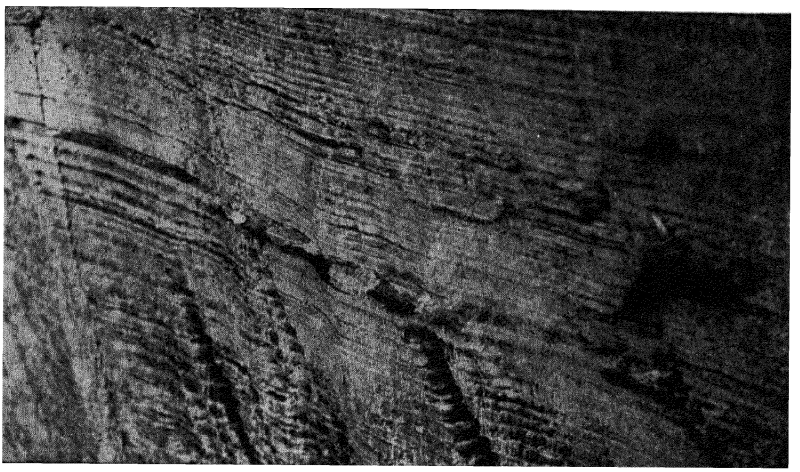
(2) *Gotiglacial*: Retreat from the terminal moraines in middle Scania to the southern border of the great Fennoscandian moraines south of Stockholm, 13100 B.C. to 6721 B.C., or 6379 years.

(3) *Finiglacial*: The retreat from the southernmost of the Fennoscandian moraines to the parting of the land ice into two parts in the Ragunda district, 6721 B.C. to 5642 B.C., or 1073 years.

(4) *Postglacial*: of Swedish geologists, based on the work of R. Lidén in the valley of the river Angermanälven, 5642 B.C. to 1900 A.D., or 7542 years. The above figures give a total of 15,000 years for the retreat of the land ice from central Scania to the present small ice caps in north central Sweden.

necticut river valleys, the shores of the Great Lakes and southern Canada were studied and the party returned to Sweden at the end of the year. Dr. Antevs remained behind to conduct further investigations in the United States and Canada. Many excellent exposures of varve clays were examined, and connections were established between long series of varves in Sweden and North America.

Although the last glaciation covered a larger area in North America than in Europe, the varve clay record, so far as studied, is represented by De Geer as enduring for about 7200 years, scarcely half that of the Swedish Time Scale. The American record is distributed as follows: 5000 years through the upper four-fifths



Courtesy Univ. of Toronto Press

FIGURE 2. Varves from clay belt in northwestern Quebec.

of the Gotiglacial stage, all of the Finiglacial stage (1073 years) and about 1200 years in the lower Postglacial stage. The terminal moraine on Staten Island and Long Island is interpreted by De Geer to be of Gotiglacial rather than Daniglacial age. The ice retreat from New York City to the northern line of New York State was accomplished, according to De Geer, in about 5000 years, that is, during the Gotiglacial stage.*

In 1924, De Geer organized the Geochronological Institute to house in Stockholm the already large collections and to further the varve clay investigations. Varve-collecting expeditions were also organized and sent out to distant countries as follows:

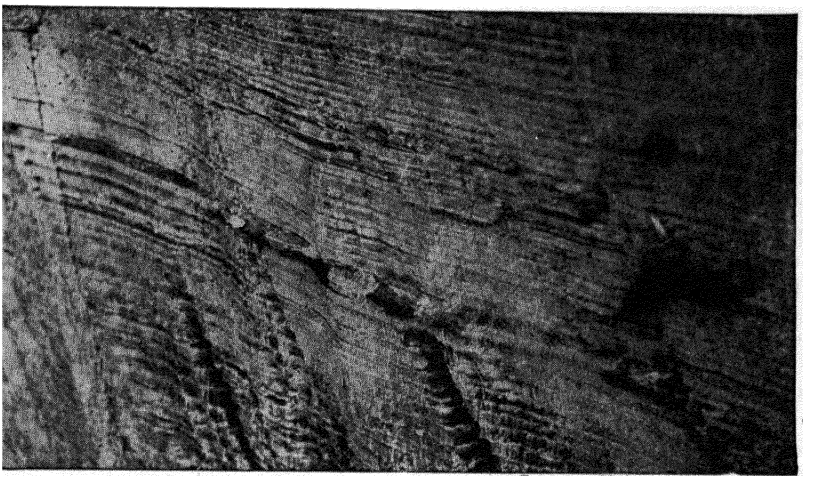
Dr. E. Norin in 1924 and 1925 conducted expeditions to the northwestern Himalaya mountains and collected varve series which correspond to the upper part of the Gotiglacial stage (−1318 to 1332). Dr. Carl Caldenius from 1925 to 1929 collected a large series of varves in Argentina and Chile which correspond to late Gotiglacial and early Finiglacial stages of the Swedish Time Scale. From 1932 to 1934 Caldenius found varves in New Zealand which tie in with middle and upper Gotiglacial varves in Sweden.

In 1927-1928 Dr. Erik Nilsson in British East Africa investigated and mapped the glaciations of Mt. Kilimandjaro, and secured from Rift valley lakes varves which could be identified with the early Gotiglacial, middle and late Finiglacial stages in Sweden. The earliest varves are associated with the archaeological finds of Dr.

* Considerable information about varves is to be found in "The Last Million Years, a History of the Pleistocene in North America," by Prof. A. P. Coleman (Univ. of Toronto Press, 1941).

necticut river valleys, the shores of the Great Lakes and southern Canada were studied and the party returned to Sweden at the end of the year. Dr. Antevs remained behind to conduct further investigations in the United States and Canada. Many excellent exposures of varve clays were examined, and connections were established between long series of varves in Sweden and North America.

Although the last glaciation covered a larger area in North America than in Europe, the varve clay record, so far as studied, is represented by De Geer as enduring for about 7200 years, scarcely half that of the Swedish Time Scale. The American record is distributed as follows: 5000 years through the upper four-fifths



Courtesy Univ. of Toronto Press

FIGURE 2. Varves from clay belt in northwestern Quebec.

of the Gotiglacial stage, all of the Finiglacial stage (1073 years) and about 1200 years in the lower Postglacial stage. The terminal moraine on Staten Island and Long Island is interpreted by De Geer to be of Gotiglacial rather than Daniglacial age. The ice retreat from New York City to the northern line of New York State was accomplished, according to De Geer, in about 5000 years, that is, during the Gotiglacial stage.*

In 1924, De Geer organized the Geochronological Institute to house in Stockholm the already large collections and to further the varve clay investigations. Varve-collecting expeditions were also organized and sent out to distant countries as follows:

Dr. E. Norin in 1924 and 1925 conducted expeditions to the northwestern Himalaya mountains and collected varve series which correspond to the upper part of the Gotiglacial stage (−1318 to 1332). Dr. Carl Caldenius from 1925 to 1929 collected a large series of varves in Argentina and Chile which correspond to late Gotiglacial and early Finiglacial stages of the Swedish Time Scale. From 1932 to 1934 Caldenius found varves in New Zealand which tie in with middle and upper Gotiglacial varves in Sweden.

In 1927-1928 Dr. Erik Nilsson in British East Africa investigated and mapped the glaciations of Mt. Kilimandjaro, and secured from Rift valley lakes varves which could be identified with the early Gotiglacial, middle and late Finiglacial stages in Sweden. The earliest varves are associated with the archaeological finds of Dr.

* Considerable information about varves is to be found in "The Last Million Years, a History of the Pleistocene in North America," by Prof. A. P. Coleman (Univ. of Toronto Press, 1941).

S. A. Leakey, described as being of Aurignacian type. The age of these specimens is thus about 15,000 years. In 1928 varve series from the southwestern part of Iceland, collected in 1919, were correlated with the Swedish Time Scale and found to correspond to varves near the Goti-Finiglacial limit of the year —1073. In 1933 H. Lundberg in Newfoundland examined ice-lake varves which are of middle Gotiglacial age.

In 1928 and 1929 De Geer and his wife, Ebba Hult De Geer, made a series of varve measurements along the northern side of the Alps Mountains from Austria to France. Connections with the Swedish Time Scale were made for the early, middle and late Gotiglacial stage and the early part of the Finiglacial stage.

In 1940 De Geer published his "Geochronologia Suecia Principles," a more elaborate presentation of geochronology than has appeared in the 90-odd shorter papers by De Geer. The investigations of De Geer and eighty other geologists have been summarized here, the Swedish Time Scale has also been presented more fully, and workers in distant countries may now obtain a better understanding of this intricate subject. In the plates and tables all the different chronologies from various parts of the globe have been united into one common chronology of about fifteen millenia. It may be added that this volume deals primarily with Finiglacial material; the Gotiglacial series will be more fully represented in a volume devoted to that subepoch.

In the early part of this paper the tree-ring chronology, largely the work of Professor A. E. Douglass, was outlined. In the latter part the varve clay chronology, as developed by De Geer and his numerous collaborators, has been presented. In 1933 Ebba Hult De Geer made a comparison of a *Sequoia* graph, comprising 3020 years (1120 B.C. to 1900 A.D.), which was presented to the Geochronological Institute by Professor Douglass, with the corresponding part of the Swedish Time Scale. Several connections between the two chronologies have been reliably certified by this comparison, and the tree-ring *Sequoia* time scale, which falls within the historical period, has been included by De Geer in the Swedish Time Scale. The comparison of these two chronologies appears in the report of the 16th International Geological Congress, Washington, 1933. Although still in their youthful stages these two chronologies based on natural phenomena have already proved to be of value in dating past events. As time rolls on they will be found still more useful.

Part II

Biology and Medicine

Proteins *

W. T. Astbury

Textile Physics Laboratory, University of Leeds, Leeds, England

There is probably no subject more difficult to do justice to in one short article than that of the proteins, and we shall not attempt the task. Our aim must be simply to convey a quick idea of recent happenings, particularly from the viewpoint of x-ray analysis.

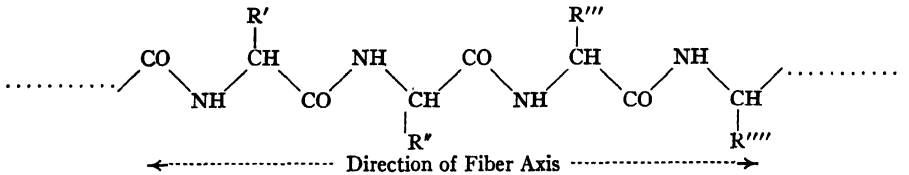
It is now generally agreed that proteins are polypeptide chain systems, alone or in combination with accessory groups or molecules, such as nucleic acids, polysaccharides, porphyrins, etc. The chains are linked by a variety of interactions between their side-chains and by less specific attractions between their backbones; above all, each protein owes its special properties not merely to the kinds and numbers of the amino-acid residues from which it is built, but also to the way the chains are folded or otherwise set up in space. Proteins are *organized* polypeptide chain systems: when the configuration of the chains is destroyed they become "denatured" and lose their specificity, even though accompanying changes in chemical constitution are usually of the slightest.

A first broad classification of the proteins would divide them into those that are visibly fibrous but not visibly crystalline, and those that are visibly crystalline but not visibly fibrous. To the first group belong such things as silk, hairs, tendons, muscle fibers, etc.; among the second we find egg albumin, insulin, hemoglobin, pepsin, etc. We should expect all proteins to be fibrous in some sense, just because there is so much evidence that they are composed of molecular chains, and because natural fibers are shown by x-rays to be of the nature of molecular yarns; but though the visibly fibrous proteins conform to expectations, as their description implies, the visibly crystalline ones seem not to do so, for their molecules in the native state are found by ultracentrifugal, x-ray, and other physico-chemical methods to be great rounded bodies—hence the name "corpuscular" or "globular" proteins. They betray themselves on denaturation, however, because then they all give similar x-ray photographs of the kind we have learned from fiber studies to associate with polypeptide chains, and the inference is that these chains have been liberated from their specific configurations and so manifest themselves simply as such. There is much other physico-chemical evidence, such as usually arise in viscosity, for this setting-free of peptide chains on denaturation; but the argument is clinched by the fact that it is possible to spin actual fibers from solutions of denatured corpuscular proteins, and these fibers when stretched give x-ray photographs analogous to those given by certain types of natural protein fibers.[1] What should be emphasized at once, therefore, is that it is now doubtful indeed whether there is any really fundamental distinction between the fibrous and the nonfibrous proteins: similar structural principles are used in the architecture of both, and as far as we can see at the moment, the difference is rather one of degree than of kind.

The fibrous proteins are the only ones so far that have yielded up to x-rays any appreciable *intramolecular* information: the inner structure of the corpuscular proteins is still mostly a matter of inference from that of the fibrous proteins. Many

* For a thorough-going treatment of this subject see "Proteins, Amino Acids and Peptides," by E. J. Cohn and J. T. Edsall, Reinhold Publishing Corp., 1943.

examples of the latter have been examined by x-rays and accessory methods, and it is remarkable into how few configurational groups they fall. The classical polypeptide chain, fully extended in the *trans*-form:

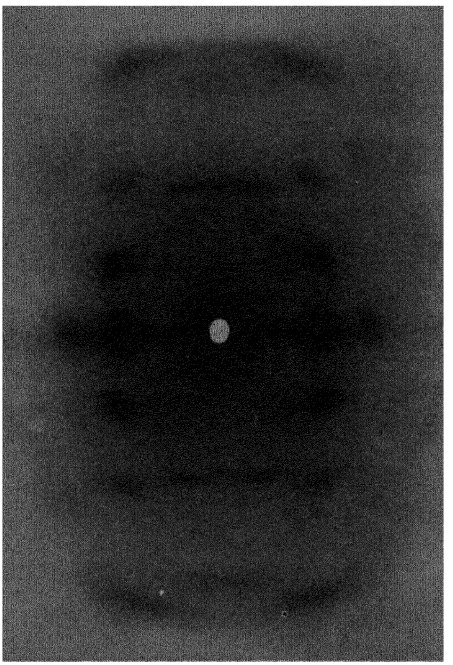


turns out to be uncommon. It is the normal configuration in silk fibroin, for instance, but most other protein fibers call for a wider interpretation. Even so, they still fall into probably only two main groups, the keratin-myosin group and the collagen group.

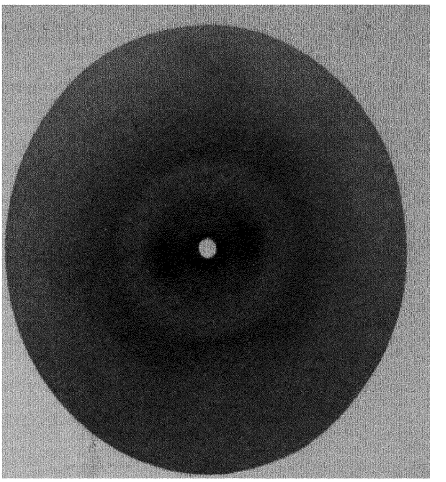
The problem of the keratin-myosin group is by way of being a prototype—or perhaps the prototype—problem in protein structure. The group includes not only the chief muscle protein, myosin,* but also the fibrous proteins of the epidermis of mammals, amphibians and certain fishes, and the fibrous structures, such as hair, horn, nails, etc., that arise from the mammalian epidermis.[2] All these are endowed with reversible, long-range elasticity, and all give two kinds of x-ray diffraction pattern according as they are unstretched or stretched. The normal form we have called α, and the stretched form β. As examples of the corresponding diffraction patterns, those of myosin, shown in Figures 1b and 1c, will serve as conveniently as any others, because the two types of pattern are constant throughout the group. This is not to say that the finer details are always the same—they are not—but it does imply that the main outlines of molecular structure are conserved, just as are certain characteristic elastic properties. The β-pattern is analogous to that of fibroin (Figure 1a) and corresponds to almost fully extended chains cross-linked through their side-chains to form polypeptide "grids" of the kind illustrated (purely diagrammatically) by Figure 2. A β-pattern may be obtained, too, from stretched artificial fibers made from denatured proteins,[1] and a beautiful elaboration of the β-pattern (Figure 1d) is typical also of the feather keratin subgroup of the keratin-myosin group, which includes, besides feathers, such structures as tortoise-shell, reptilian scales and the like. The *normal* state in the feather keratin subgroup is one in which the polypeptide chains are constricted to a length that is some 7 per cent short of their possible maximum, though it is possible to stretch them continuously and reversibly over this remaining 7 per cent.

The precise nature of the α-form and what happens during the reversibly elastic transformation to the β-form are the source of the great interest and importance of the keratin-myosin group, for the first promises the key to the intramolecular folds of the corpuscular proteins, and the second is probably no less than a particular manifestation of the molecular mechanism involved in the contraction of muscle.[3] *Whatever the answers, they have to be independent of the marked variations in chemical constitution that occur throughout the group.* The α-form must consist of regularly folded polypeptide chains whose effective length, from other evidence, is approximately half that of the extended β-chains, and the folds in the main-chains must lie in planes perpendicular to the average direction of the side-chains. The latest and most generally satisfactory model[4] is pictured diagrammatically in Figure 3: a detailed atomic counterpart to Figure 3 has also been constructed,[4] but the diagram will suffice to illustrate the present discussion. It will be seen that the side-chains are grouped in triangular sets of three that lie alternately on one side and the other of the main-chain. The reasoning leading to this solution takes

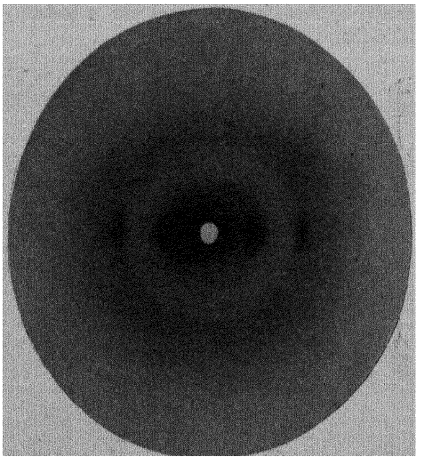
* *Footnote added in proof*: It now appears that fibrinogen and fibrin also belong to the keratin-myosin group (Bailey, Astbury and Rudall—in press).



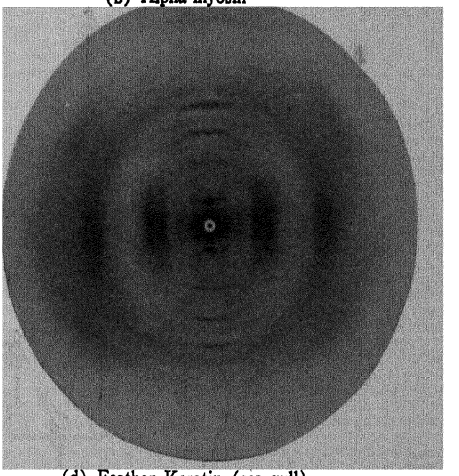
(a) Silk fibroin (Bombyx mori)



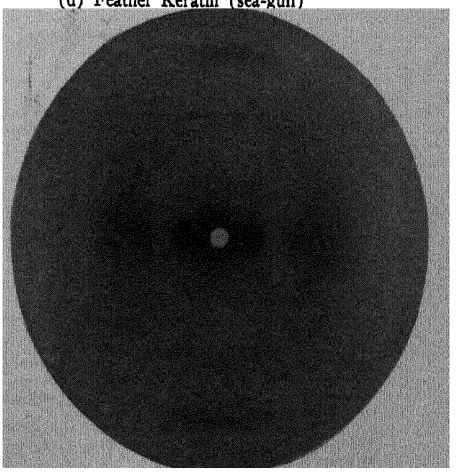
(b) Alpha-myosin



(c) Beta-myosin



(d) Feather Keratin (sea-gull)



(e) Collagen (rat-tail tendon)

FIGURE 1. Five typical x-ray diffraction patterns of the fibrous proteins.

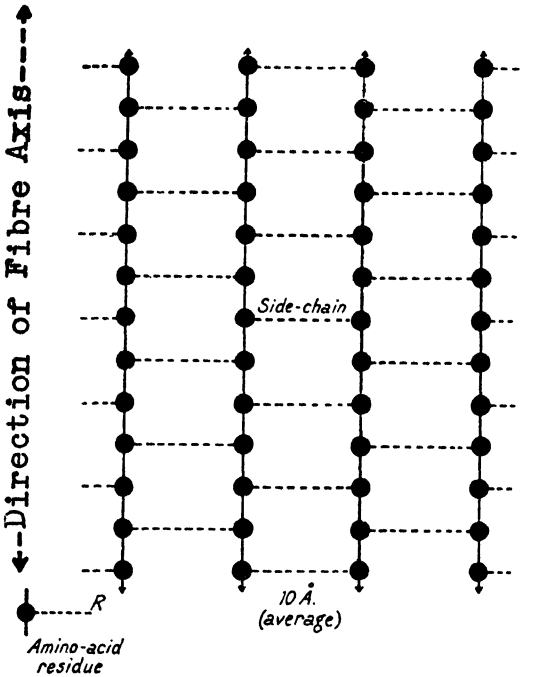


FIGURE 2. Diagrammatic representation of a polypeptide "grid" in the extended, or β-, form. The α-form may be derived schematically by throwing the paper into a series of regular folds with crests parallel to the side-chains.

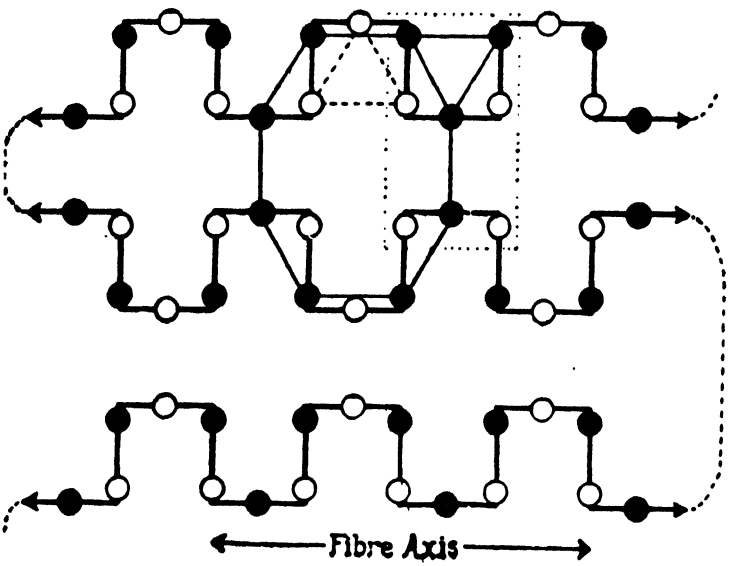


FIGURE 3. Packing of side-chains in the α-fold of the keratin-myosin group. — represents the direction of the main-chains; ● represents a side-chain pointing *up* from the plane of the diagram; ○ represents a side-chain pointing *down* from the plane of the diagram.

no immediate cognizance of the nature of the side-chains, which in a sense is all to the good; but it might be expected at least that the members of each triad should be similar to the extent of permitting or actively promoting aggregation: that is to say, that all three should be polar or all three non-polar. If this is so, then it follows also that *polar and non-polar side chains follow one another alternately along the main-chains.*[5]

Here in so many words is what promises to be the root idea of the whole keratin-myosin group. It is still far from being proved, of course, but clearly we must look for something of the sort if we are to understand how Nature has succeeded in erecting such a variety of structures on one foundation. There must be something comparatively simple underlying it all, something simple and comprehensive, and we suggest, on the basis of a long series of x-ray and related studies,[6] that the truth does indeed lie in the direction indicated. Add to the single postulate just formulated the principle of close packing long recognized by crystallographers, and at least the broad characteristics of the group seem to follow inevitably. Similar remarks apply to the collagen group: the collagen fibers rest on another foundation, but again it must support the whole immense group.

To the molecular biologist the problem of muscle is one of almost unrivalled fascination. We see it now as an exercise in the properties of proteins, or more exactly, in the properties of one particular protein, myosin, which turns out to belong to the same molecular family as the hair protein, keratin. Keratin too has the power of contraction * once certain cross-linkages that help to differentiate it from myosin have been broken; and in fact all the principal x-ray and elastic properties of myosin find their counterparts in those of the epidermis and the supercontracting form of keratin. Unfortunately, x-rays show no new *regular* pattern for the state that is shorter than the α-form, but such changes as are observed are always similar, whether in keratin, isolated myosin *or living muscle,* and never do they correspond quantitatively to a simple disorientation of the α-form, but rather to a further folding of the polypeptide chains. It is hard to resist the conclusion that muscular contraction is only another example—an exquisitely controlled example, but still only an example—of the kind of thing that occurs during the β-α transformation, the thermal contraction of the collagen fibers, or the elastic retraction of elastin and stretched denatured proteins. It is a shortening by intramolecular rearrangement, and *it takes its place logically between the extended, or β-, proteins on one hand and the corpuscular proteins on the other.*[7]

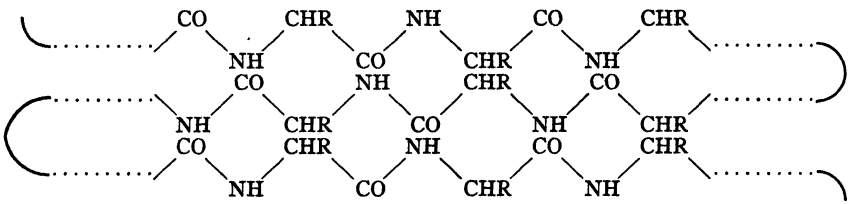
The scheme of folding illustrated in Figure 3 provides as yet the merest sketch of a full interpretation of the activities of the keratin-myosin group, and nothing is easier than to question it on points of detail. Details, however, from the very nature of the case, are largely the responsibility of individual members of the group, and in any event will not be forthcoming till we have many more data at our disposal. For the present, if we invoke only the two principles of roughly equal proportions of polar and non-polar side-chains (the evidence for which will be given below) and the close packing of these in such a way that polar is drawn to polar and non-polar to non-polar, then already the outlook is wider and more promising than anything before. The contraction of muscle appears as a consequence of increased facilities for aggregation induced by comparatively simple reversible changes in the relative distribution of polar and non-polar side-chains, and the fact that myosin is not only the working elastic mechanism but also a principal enzyme (adenosine triphosphatase [8]) in the elastic cycle becomes at once intelligible. It must mean that among the polar side-chains are included specific reactive groupings that leave the main elastic properties substantially unchanged while increasing the ver-

* To avoid confusion with contraction from the β- to the α-form on release from stretching, we have given the name "supercontraction" to the power that may be developed in keratin of contracting below its unstretched length.

satility or sensitivity of the system in relation to a particular chemical environment. The possibility of such a compromise is implicit in the scheme here proposed.

The state of affairs inside the molecules of the corpuscular proteins is still only an extrapolation from what has been discovered about the fibrous proteins. We know that denaturation ultimately liberates peptide chains—sometimes spinnable, fiber-forming chains of great length, sometimes much shorter chains (see below)[9]; but in only one instance so far has anything interpretable been observed by x-rays of the actual process taking place. Single crystals of the seed globulin, excelsin (one of the proteins from which, when fully denatured, artificial fibers can be spun), were found to give *in the same x-ray photographs* not only diffraction spots corresponding to a rhombohedral lattice of corpuscular molecules of weight about 300,000, but also three superposed fiber patterns with the fiber axes parallel to the hexagonal a-axes and the side-chains almost parallel to the basal plane.[9, 10] The crystals had partially degenerated into what Osborne called "excelsan," though outwardly and optically they appeared quite sound. Obviously in this case, to produce so little disturbance, the transition from corpuscular to fibrous must involve an almost negligible change: internally the corpuscular molecules must be for all practical purposes fibrous, and the probability is that they are built from chains running backward and forward parallel to the crystal axes along which the fibers make their appearance.

All the indications are that denaturation is often an easy, unobtrusive process of this kind, and it seems likely that the intramolecular folds in many corpuscular proteins, especially those like the seed globulins without any highly specialized function, are no more complicated than the following:[9, 11]



in which the side-chains (the R-groups) stand out alternately above and below the plane of the folds. The principle is the same as that of the proposed α-folds shown in Figure 3, except that the α-folds are approximately square instead of being long loops. It will be noticed that each loop is in the β-configuration; the α-loop, in fact, is merely the shortest possible β-loop and there is no formal difference between the two configurations: both α- and β-keratin have the same density, 1.3 gm/cc, and so have all proteins, roughly speaking.[12]

A further elaboration would consist of loops formed not from extended chains, as just described, but from α-chains, again as illustrated in Figure 3. In any case we should expect the folding to take place, in general, in planes transverse to the side-chains, and the next step in the argument leads automatically to the idea of laminae with side-chains projecting from both faces. A laminar structure is indeed already surmised for a number of corpuscular molecules, the most familiar being that of egg albumin, which is believed, as an inference from x-ray and monolayer data,[13] to comprise four superposed laminae. Possibly even in highly specific corpuscular proteins the bulk of the molecule is built on some such straightforward lines, with the more specialized part, or prosthetic group, as an added complication. It is to be feared that it will be a considerable time, particularly amid the demands of war, before x-rays are in a position to adjudicate on these points with authority, for they present a formidable problem in diffraction analysis. The most promising single crystal studies in progress at the moment are those on insulin[14] and hemo-

globin,[15] which are being prosecuted chiefly through the medium of Patterson projections. It may or may not be significant as a direct link with the keratin-myosin group that the Patterson-Harker section of wet zinc insulin parallel to (0001) at $z = 0$ is characterized by a hexagon of high and well-defined peaks at 5 Å, that is to say, at practically the distance at which the strong intramolecular pattern of the α-fibers repeats; but at least it will be granted that it is exceedingly stimulating and worthy of the closest investigation. We might expect *a priori* some system of close packing among the side-chains if it is at all possible, and the suggestion here is that the same system is used in insulin as in the keratin-myosin group. It is no more than a suggestion, however, and much work remains to be done before a decision is reached.

Questions of the folds or of almost any internal feature of the proteins are of course inseparable from that of chain length, but it is only recently that light of a clearer sort has been thrown on the latter. In the fibrous proteins, as a rule, it can hardly be doubted that the chains are of great length, and the same is true of the artificial fibrous proteins spun from denatured seed globulins; but for the majority of corpuscular proteins we have only such evidence as comes from ultracentrifuge[16] and dissociation experiments, indicating that the greater molecules seem to be built up of lesser units, and therefore that a fair number of shorter chains instead of a few longer chains might be more usual than had hitherto been supposed. Some years ago, as part of an x-ray study of the fiber-forming properties of denatured proteins,[9] the experiment was tried of stretching strips cut from films of "poached" egg-white; it was found very surprisingly that, whereas the chain-bundles in denatured seed globulins tended to set along the direction of stretching, those in denatured egg-white tended to set *across* the direction of stretching. There could be only one interpretation of this, that the chains in denatured egg-white are much shorter than in denatured seed globulins; and similarly for the native molecules, presumably. The observation fits in nicely with the inference, mentioned above, that the native egg albumin molecule (weight about 43,000) consists of four superposed laminae and therefore possibly of four more or less discrete chains, but it takes on a new interest and importance now that Chibnall's[17] chemical analytical data point to the same number. Chibnall has shown, furthermore, that in edestin (a seed globulin) there can only be one chain per weight of 50,000. These results mean that the edestin chains are of the order of five times as long as the egg albumin chains, and the surprise engendered by the poached egg experiment, though innocent, was perhaps unmerited!

Chibnall's method has been to make exceptionally painstaking analytical determinations of the numbers of acidic and basic residues, and then to compare them with the data given by titration and Van Slyke's nitrous acid procedure: the discrepancies, if such exist, lead to an estimate of the numbers of free carboxyl and amino groups at the ends of chains. In this way he has shown, in addition to the results just quoted, that there are probably eighteen chains in the insulin molecule and nine in the lactoglobulin molecule. Below we shall consider the fundamental bearing these findings have on the central problem of protein stoichiometry—the problem of the laws governing the proportions and distribution of the residues—but in the meantime it is gratifying to note, as a further stimulating approach to the crystallographic solution of insulin, that Crowfoot's Patterson projection on (0001) also suggests a compounding of the molecule from eighteen sub-units. There is nothing yet to identify Chibnall's eighteen peptide chains with Crowfoot's eighteen sub-units, but again it will be granted that here is something very promising to work upon.[14, 18] Riley[19] has carried out a lengthy x-ray study of the crystal structure of lactoglobulin, and he too has found evidence of sub-units—not so clear nor in such direct correspondence to Chibnall's results as in the case of insulin, but still definitely hopeful.

Everything in protein studies turns ultimately on the numbers and distribution

of the amino-acid residues. Their weights range from that of glycine (57) to that of tryptophan (186), giving an arithmetic mean of about 120. Roughly then, 100 gm of protein should contain something of the order of $100/120 = 0.83$ gram-residue of amino-acids; and if the yield of any particular acid is large, it may be possible at once to hazard a sound guess as to what fraction of the total residues arises from that acid. For example, the leucines yield from hemoglobin has been reported [20, 21] as 27.8 gm and 29.5 gm, corresponding to 0.21 — 0.22 gram-residue; and from insulin the leucines yield [22] of 30 gm corresponds to about 0.23 gram-residue; so that in both these cases the suggestion is that one-quarter of the residues are leucine or isoleucine residues. Similarly, from gliadin, 43 gm of glutamic acid [23] corresponds to rather more than 0.29 gram-residue (all in the form of glutamine residue, to judge by the amide nitrogen), with the suggestion that one-third of the residues are glutamine residues: and again, from silk fibroin, the reported yields [24, 25] of 40.5 gm and 43.8 gm of glycine, corresponding to 0.54 and 0.58 gram-residue respectively, indicate that as many as half the total residues in fibroin are glycine residues (since a high glycine content must give a low average residue weight). With single *low* yields we can of course feel no such confidence; nevertheless it has been observed for some time that certain yields, expressed as gram-residues, at least bear a simple numerical relation to one another, and that again, as with the examples just quoted, powers of 2 and 3, or products of such powers, seem to be common. Block [26] pointed out, for instance, that the hemoglobins from the horse, the sheep and the dog all contain iron, histidine, arginine and lysine in the ratio 1:8:3:9, and that in the keratins the ratio of histidine, arginine and lysine is about 1:15:5 (since modified [27] to 1:12:4); and Atkin [28] pointed out that Dakin's results indicate that glycine and hydroxyproline residues are present in gelatin in the ratio of 3:1; and so on. Other examples could be cited, but these will serve to illustrate a growing impression during the past ten years or so.

In 1933 the writer, with H. J. Woods, made the first attempt to link up the ratios of gram-residues with x-ray and physico-chemical data on wool keratin.[29] This was followed by the application of similar ideas to the problem of gelatin; [30] and on the basis of chemical analyses then available it was concluded that "the residues of the two chief acids, glycine and oxyproline (hydroxyproline), account, respectively, for about one-third and one-ninth of the total number of residues; that is to say, every third residue could be a glycine residue and every ninth an oxyproline residue." [31] The way this conclusion was reached was as follows. The total nitrogen content of gelatin was reported as 18 per cent, and the reported yields of amino-acids corresponded to a nitrogen content of 15.64 per cent: the difference, 2.36 per cent, came from acids still unaccounted for, and these were very probably monoamino-monocarboxylic acids. The undetermined acids amounted therefore to $2.36/14 = 0.16_8$ gram-residue, which brought the 0.87_8 already reported up to 1.04 gram-residues. Now 1.04 is, within the limits of experimental error, three times the number of gram-residues of glycine (0.34) and nine times the number of gram-residues of hydroxyproline (0.11).

Since that time Bergmann and Niemann [32] have proposed two fundamental principles that are generalizations of these early suggestions. They postulate that *all* proteins are built according to the following rules: (1) the numbers of the different kinds of residues and also the total numbers of residues in the molecule are always of the form 2^n3^m, where n and m are positive integers or zero; and (2) the residues of any one kind always occur at a regular periodic interval along the polypeptide chain or chains.

The most striking evidence for the validity of the 2^n3^m rule under certain conditions has recently been obtained by Chibnall,[17] working with edestin. Chibnall estimates the average residue weight (R) from a consideration of the nitrogen distribution, after the manner just described for gelatin: [31] his value for edestin is 115.7, and

the number of gram-residues (G) in 100 gm of protein is $100/115.7 = 0.864$. Table 1 is a modified expression of Chibnall's results.[18]

Table 1. Edestin.

Amino-acid	$R = 115.7$ Yield (%)	$G = 0.864$ Gram-residues Obs.	Calc.	Frequency *
Histidine	2.41	0.0156	0.0160	54 ($2^1 3^3$)
Arginine	16.71	0.0961	0.0961	9 (3^2)
Lysine	2.37	0.0162	0.0160	54 ($2^1 3^3$)
Tyrosine [33]	4.34	0.0240	0.0240	36 ($2^2 3^2$)
Tryptophan [33]	1.50	0.0073	0.0080	108 ($2^2 3^3$)
Methionine [33]	2.35	0.0158	0.0160	54 ($2^1 3^3$)
Cystine/2	1.44	0.0120	0.0120	72 ($2^3 3^2$)
Glutamic acid	20.7	0.1408	0.0720	12 ($2^2 3^1$)
Glutamine			0.0720	12 ($2^2 3^1$)
Amide NH_3	2.15	0.1265		
Asparagine	12.0	0.0902	0.0540	16 (2^4)
Aspartic acid			0.0360	24 ($2^3 3^1$)
				L.C.M. = 432 ($2^4 3^3$)

* The "frequency" is simply the reciprocal of the fraction of the total residues, without prejudice to the question whether the residues of any one kind are equally spaced along the chain or chains.

The agreement shown in Table 1 is so excellent that there can be no doubt now that for edestin at least, for the amino-acids quoted, the $2^n 3^m$ rule holds strictly. The position with regard to the other amino-acids must remain open pending further exact analyses, but in the meantime it is clear that there are no fewer than 432 ($2^4 3^3$) residues in the molecule, and therefore the molecular weight must be $432 \times 115.7 = 50{,}000$ approximately, or a multiple thereof. The multiplying factor is presumably 6, since the ultracentrifuge [16] gives a molecular weight of 310,000.

Table 2.

I, Percentage yield. II, Observed gram-residues in 100 gm of protein. III, Frequencies referred to G. IV, Minimum molecular weight calculated from I and II. V, Assumed number of residues. VI, Calculated molecular weight of protein.

Amino-acid	I	II	III	IV	V	VI
	Egg Albumin $R = 111.5$		$G = 0.897$			
Arginine	5.63	0.03234	27.7	3,092	14	43,280
Histidine	1.45	0.00935	96.0	10,070	4	42,800
Lysine	5.06	0.03463	25.9	2,886	15	43,310
Tyrosine	4.10	0.02264	39.6	4,417	10	44,170
Tryptophan	1.32	0.00647	139	15,460	3	46,400
Cystine [34]	1.79	0.00745	120	13,420	3	40,270
Methionine [34]	5.10	0.0348	26.2	2,925	15	43,900
Glutamic acid	16.1	0.1109	8.1	902	48	43,300
Aspartic acid	8.1	0.0612	14.6	1,638	27	44,100
Amide NH_3	1.23	0.0722	12.4	1,395	31	43,250
	β-Lactoglobulin $R = 112.4$		$G = 0.890$			
Arginine	2.89	0.0166	53.6	6,115	7	42,105
Histidine	1.54	0.0099	89.8	10,080	4	40,290
Lysine	9.75	0.0667	13.3	1,498	28	41,970
Tyrosine [35]	3.78	0.0209	41.7	4,790	9	43,110
Tryptophan [35]	1.94	0.0095	93.7	10,520	4	42,090
Methionine [35]	3.22	0.0216	41.3	4,633	9	41,700
Cystine [35]	2.29	0.0095	93.5	10,500	4	41,980
Cysteine (as ½ cystine) [35]	1.10	0.0092	97.3	10,920	4	43,670
Glutamic acid	21.51	0.1463	6.1	684	62	42,400
Aspartic acid	9.88	0.0743	11.6	1,346	31	41,730
Amide NH_3	1.29	0.0759	11.7	1,318	32	42,160

Chibnall's equally careful results for egg albumin and lactoglobulin, on the other hand, give no such unequivocal support for the 2^n3^m rule. They are given in Table 2. The discrepancies are still no proof, however, that the structures of these two proteins do not rest fundamentally on a 2^n3^m basis, because on the one hand recrystallized egg albumin has been shown by electrophoresis to contain two components [86] and each component contains probably four chains,[17] and on the other, lactoglobulin, though homogeneous, probably contains nine chains per molecule.[17] Each component or chain, as the case may be, could conform separately to the 2^n3^m rule, but groups of components or chains need not so conform: to what extent agreement is found in group analyses must depend on the number and kinds of dissimilar components.

The literature of protein analysis reveals many exact agreements, approximate agreements, and seemingly flat contradictions like those just quoted.[18] Data such as Chibnall's on edestin, egg albumin and lactoglobulin are just the sort of thing needed to make sense of the accumulating concordancies and discrepancies; for undoubtedly there are enough of the former to demonstrate the existence of a 2^n3^m rule of some kind or other, yet quite enough of the latter also to show that it cannot be of general validity in the form in which it was first expressed. In the light of what we know now, the most reasonable point of view would appear to be that the rule holds at least for structural sub-units, and hence sometimes partially, sometimes completely for structures as a whole, according to the number and kinds of dissimilar sub-units. The fact that it holds so extensively—perhaps absolutely—for edestin is thus a stoichiometric indication that the edestin molecule should consist of only a few long chains, and this is confirmed not only by the pronounced fiber-forming properties and x-ray photographs of the denatured seed globulins [9] but also by Chibnall's estimate of the chain-length.[17] Similarly, from Table 2, we might expect egg albumin to consist of a number of shorter chains, and this again is confirmed both by x-rays and by Chibnall's findings. Experiments are now in progress to try to extend the x-ray method to other corpuscular proteins.

Table 3. Comparative Chemical Constitutions of Wool Keratin and Rabbit Myosin for the More Reliable Amino-acids.

Amino-acid	Observed approximate numbers of residues in 576: Wool Keratin	Rabbit Myosin
Cystine/2 + cysteine	67	8
Methionine	3	15
Serine	66	23
Threonine	36	21
Tyrosine	17	13
Tryptophan	6	3
Aspartic acid + amide	34	45
Glutamic acid + amide	65	101
Arginine	40	27
Lysine	12	47
Histidine	3	7
Amides	55	57

Proposed "ideal" distribution of amides

	Prop.		Obs.	Prop.		Obs.
Aspartic acid	12	→ 36	34	24	→ 48	45
Asparagine	24	→ 56	55	24	→ 56	57
Glutamine	32	→ 64	65	32	→ 96	101
Glutamic acid	32			64		

The stoichiometry of the fibrous proteins, though no doubt amenable ultimately to some form of the 2^n3^m rule, must rest also on a wider basis, as already pointed out.

Table 3 [18] shows the range of variation in chemical constitution between wool keratin and rabbit myosin for the more reliably determined amino-acids. There are good grounds for believing that what is aimed at with regard to the acidic and amide residues is that keratin and myosin shall have the same numbers of asparagine and glutamine residues (24 and 32, respectively), while myosin shall have twice as many of both aspartic and glutamic acid residues (24 and 64), as keratin has (12 and 32);

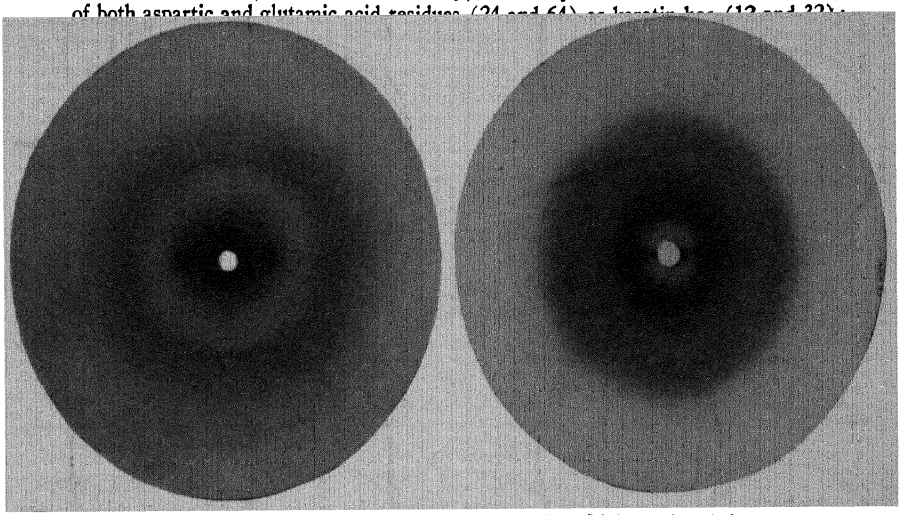


FIGURE 4a. Stretched foot retractor muscle of *Mytilus edulis* showing partial transformation into β-myosin.

FIGURE 4b. Living relaxed foot retractor muscle of *Mytilus edulis*.

but otherwise there seems to be no obvious correlation between the two sets of data, beyond the fact that the polar residues amount in each case to practically half the total (neglecting cysteine residues, of which in myosin there are about 2 in 576, the polar residues add up to 273 in keratin and 284 in myosin). The situation recalls mixed crystallization and alloy formation, in which it is the crystallographic *sites* that are significant, rather than the specific atoms or groups that occupy them; and we get the impression that what counts in the keratin-myosin series is that certain sites within the structure shall be occupied by certain *types* of residues. Figure 5 is an example of what is meant: [5, 18] it postulates three main types of polar cross-linkage, acid-base, amide-hydroxyl and hydroxyl-hydroxyl, and it adapts, within the framework of Figure 3, the well-established concept of "ideal" or limiting structures, one at either end of the series. Of the polar cross-linkages in the limiting keratin, one-third are salt linkages, one-third amide-hydroxyl bridges, and one-third hydroxyl-hydroxyl bridges; while in the limiting myosin (from which rabbit myosin itself is not far removed), two-thirds are salt linkages and one-third amide-hydroxyl bridges. The experimental evidence suggesting such a scheme lies in the following analytical figures, [5, 18] the "ideal" limiting figures being shown in brackets for comparison:—

	Type of residue (Nos. in 576)			
	Basic	Acidic	Amide	Hydroxyl
Wool keratin.....	55 (48)	44 (48)	55 (48)	119 (144)
Rabbit myosin.....	81 (96)	89 (96)	57 (48)	57 (48)

Another way of realizing the interchangeability of residues that runs through the keratin-myosin series could be by means of a variable grouping-together of chains or grids of constant composition, with each chain or grid perhaps conforming strictly to the 2^n3^m rule. On the whole, bearing in mind the multiple-chain structure of egg albumin, insulin and lactoglobulin, the evidence is possibly more in favour of this

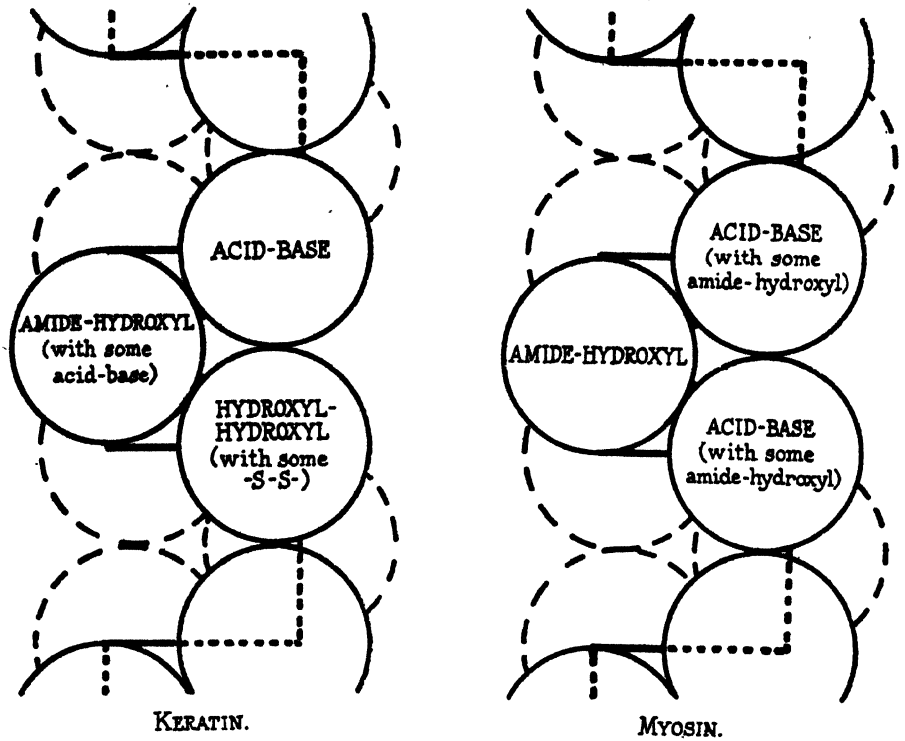


FIGURE 5. Proposed plan for the packing and cross-linkage of polar side-chains in the keratin-myosin series.

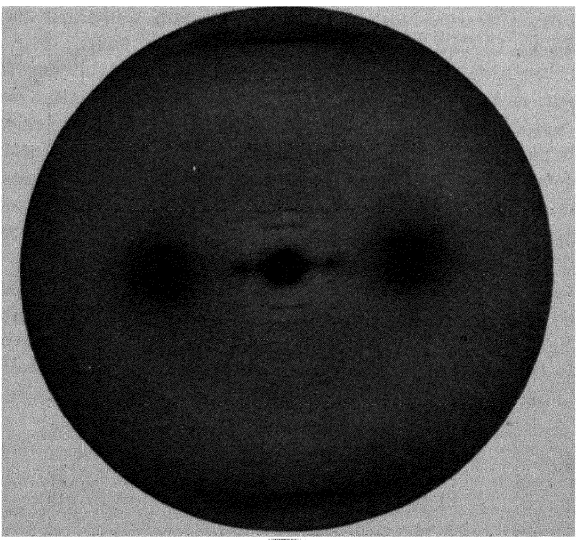


FIGURE 6. X-ray fiber photograph of the tip end of a porcupine quill, dried in vacuum over phosphorus pentoxide. Cu K_α rays. Specimen-to-film distance = 9.7$_6$ cm. Collimator 13 × 0.025 cm (MacArthur).

solution than the one just outlined, though it is quite reasonable, for all we know to the contrary, to suppose that both methods are used according to circumstances. An important crystallographic fact with regard to the fibrous proteins that may well find its true explanation in terms of this second solution is the occurrence in their x-ray photographs of large side-spacings, which show that the structural units must be composed of a number of different, or differently oriented, chains lying side by side. This strongly reminds us of the plagioclase feldspars, built from continuously varying proportions of albite ($NaAlSi_3O_8$) and anorthite ($CaAl_2Si_2O_8$), and the recent x-ray interpretation of their structure as a system of fine lamellae.[37]

At the moment, the final elucidation of the structure of the keratin-myosin group looks most promising by way of porcupine quill, which gives much the best x-ray diffraction patterns of the series. Figure 6, due to MacArthur,[38] is a reproduction of the central region of the porcupine quill photograph: the prominent group of reflections near the top and bottom corresponds to the arcs—the only arcs—seen in Figure 1b. Figure 7, also due to MacArthur, is a composite diagram that includes most of the α-keratin pattern as we know it now.

The molecular pattern along the fiber axis of porcupine quill repeats at probably 658 Å, and the strong meridian reflection of spacing 5.14 Å that is so characteristic of the keratin-myosin group is the 128th (2^7th) order of that period. If 5.14 Å represents the length of an intramolecular fold comprising three residues, as we believe, then a single chain of the pattern-length revealed by x-rays would contain 384 ($2^7 3$) residues, or a multiple thereof. Chemical analysis indicates 576 ($2^6 3^2$) residues,[5, 18] but there is not necessarily any conflict, because both the x-ray and the chemical analytical numbers quoted must be considered for the present only as the smallest compatible with the experimental data, and because the chain may conceivably be folded into long loops—it is easy, in fact, to suggest a type of loop that would conform with both numbers. The important thing to notice is that 384 is of the form $2^n 3^m$, and, what is more impressive still, that *almost all the dominant reflections along the fiber axis are orders of 658 Å that are exactly or nearly of the form* $2^n 3^m$.

Chemical analytical data for porcupine quill are still impossibly scanty, so there is no opportunity yet of trying to link up this pregnant discovery with the residue frequencies. In wool keratin,[5, 18] however, these frequencies are mostly quite close to numbers of the form $2^n 3^m$, and without doubt they will be found to be so in porcupine quill too. It follows from the porcupine quill diffraction pattern that such chemical relationships can no longer be thought of as amusing coincidences arising out of a combination of experimental errors and wishful thinking. The structural stage is now set for the third act.

Outside the keratin-myosin group almost all other natural protein fibers—all such things as the white connective tissue fibers, tendons, cartilage, the scales and fins of fishes, the ichthyocol of swim-bladders, jelly-fish and many other structures[2]—fall into the collagen group, for they all give diffraction patterns of the type shown in Figure 1e. The pattern is not in every case identical, it must again be emphasized, but all the photographs share certain features that point to a common molecular plan conserved in spite of variations in chemical constitution, just as in the keratin-myosin group. Unfortunately, very few good chemical analyses are available for pure collagen fibers, most analyses having been carried out on gelatin, the protein derived from collagenous structures by breakdown of a hydrolytic nature. Oriented gelatin, however, still gives the typical main outlines of the fiber pattern, and it is permissible to make use of gelatin analyses in at least a preliminary skeleton interpretation.

The collagen problem has been discussed at length elsewhere,[2, 18] and the barest summary of things must suffice here. The fibers are normally not elastic in the sense of the keratin-myosin fibers, and stretching them does not alter the x-ray

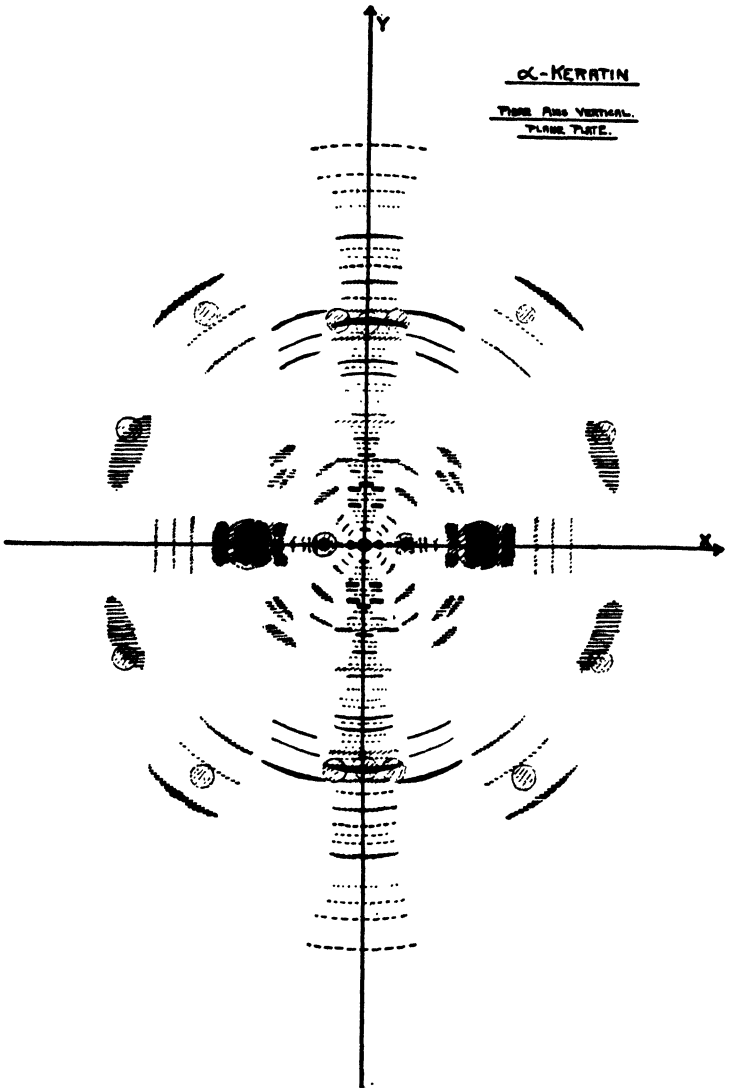


FIGURE 7. Composite diagram of the α-keratin diffraction pattern given by porcupine quill (Cu Kα rays). The spacing of the most prominent arc on the meridian is 5.14_7 Å (MacArthur).

photograph except for orientation effects; but they contract spontaneously at temperatures above about 60-70° C and then show long-range elasticity from their shortened length. Everything points to the polypeptide chains being normally in some configuration that is stereochemically fully extended—a configuration that is folded up irregularly when the temperature is raised high enough in the presence of water, or even without raising the temperature in the presence of certain swelling agents—yet the diffraction pattern seems to demand chains that are considerably shorter than fully extended; in fact, it suggests an average residue length about equal to the spacing (2.86 Å) of the strong meridian arc that is so characteristic of the group.

To resolve this conflict of evidence it is inevitable to look for peculiarities in the chemical analytical data, and there indeed the most reasonable answer lies. The solution is once more one of stoichiometry, a question of the proportions of the amino-acid residues, or more precisely, of certain kinds of residues.

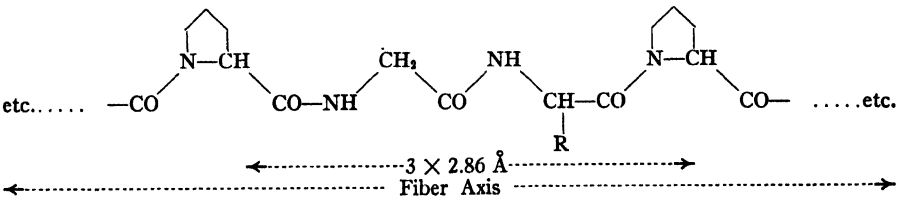
Table 4.[18] Gelatin.

$R = 94$ $G = 1.06$

Amino-acid	Yield (%)	Gram-residues Obs.	Calc.	Proposed Frequency
Glycine	25.5 [39]	0.3399		
	25.5 [40]	0.3399		
	* 26.5 [41]	0.3533	0.3533	3 (3^1)
Alanine	8.7 [39]	0.0978	0.1178	9 (3^2)
Leucines	7.1 [39]	0.0542		
	8.5 [42]	0.0649	0.0589	18 ($2^1 3^2$)
Proline	19.7 [40]	0.1713		
	* 17.5 [41]	0.1522	0.1767	6 ($2^1 3^1$)
Hydroxyproline	14.1 [39]	0.1060		
	14.4 [40]	0.1083		
	14.7 [43]	0.1105	0.1178	9 (3^2)
Arginine	9.1 [44]	0.0523		
	8.7 [45]	0.0500	0.0589	18 ($2^1 3^2$)
Lysine	5.9 [39]	0.0404	0.0442	24 ($2^3 3^1$)

* For collagen also.

Table 4 collects various determinations made on the more abundant amino-acids in gelatin (with two results for collagen proper): they account for some six-sevenths of the total. One-third and one-ninth, as mentioned earlier in this article, must be allotted to glycine and hydroxyproline, respectively; but now we know too that about one-sixth must go to proline. Thus nearly one-third (actually 5/18) of all the residues are *imino* residues, which in itself is sufficient to diminish the average residue length without invoking intramolecular folds. When a model is constructed, the decrease is found to be just about what is required, and the most probable configuration is of the form:



This is no more than the general scheme, it should be remembered, on which many variations are possible and are doubtless realized in nature: one way, for example, must be by virtue of the fact that once in every 18 residues or so the imino ring is lacking, and the gaps are available for a variety of other residues. We have to explain the occurrence of a great family of structures of differing chemical constitution but built to a common plan; and just as in the keratin-myosin group a single master key may be found in an alternation of polar and non-polar residues, so in the collagen group the key seems to lie in the proportions and preponderance of glycine and imino residues.

The full repeat of molecular pattern along the fiber axis in collagen is of great length. It had been inferred for some time [2] that it was a multiple of approximately 209½ Å, but quite recently Bear,[46] working with dried beef tendon, has recorded ten of the innermost orders of a fundamental period that he estimates to be about 640 Å. The overall average from Bear's data and those of other workers is about 636 Å, which on the interpretation favored here is equivalent to the length of something like 216 ($2^8 3^3$) residues.

This article could hardly close better than with a reminder of what was said in the opening paragraph; but in spite of the confessed inadequacy of the present discussion it is to be hoped that one point at least has been made clear; that far-reaching developments are maturing. The story of the proteins is as long and involved as any with which the human mind has grappled, but never was there a more worthy theme, and surely now a crisis is near.

References

1. Astbury, Dickinson and Bailey, *Biochem. J.*, **29**, 2351 (1935); Lundgren, *J. Am. Chem. Soc.*, **63**, 2854 (1941).
2. Astbury, *J. Int. Soc. Leather Trades' Chemists*, **24**, 69 (1940) (First Procter Memorial Lecture).
3. Astbury and Dickinson, *Nature*, **135**, 95, 765 (1935); *Proc. Roy. Soc.*, **B129**, 307 (1940).
4. Astbury, *Chem. & Ind.*, **60**, 41, 491 (1941); Astbury and Bell, *Nature*, **147**, 696 (1941).
5. Astbury, *J. Chem. Soc.*, 337 (May, 1942).
6. See, for example, Astbury, "Fundamentals of Fibre Structure" (Oxford, 1933); Astbury and Street, *Phil. Trans. Roy. Soc.*, **A230**, 75 (1931); Astbury and Woods, *ibid.*, **A232**, 333 (1933); Astbury and Sisson, *Proc. Roy. Soc.*, **A150**, 533 (1935); Woods, *ibid.*, **A166**, 76 (1938).
7. Astbury, *C. R. Lab. Carlsberg, Ser. Chim.*, **22**, 45 (1938) (Sörensen Jubilee Vol.).
8. Engelhardt and Ljubimowa, *Nature*, **144**, 668 (1939); Ljubimowa and Engelhardt, *Biochemia*, **4**, 716 (1939); Engelhardt, Ljubimowa and Meitina, *C. R. Acad. Sci. U. R. S. S.*, **30**, 644 (1941); Needham *et al.*, *Nature*, **147**, 766 (1941); *ibid.*, **150**, 46 (1942); Needham (D. M.), *Biochem. J.*, **36**, 113 (1942); Bailey, *ibid.*, **36**, 121 (1942).
9. Astbury, Bailey and Dickinson, *Biochem. J.*, **29**, 2351 (1935).
10. Astbury and Bell, *Tabulae biologicae*, **17** (1), 90 (1939); and unpublished data.
11. Astbury, *Trans. Faraday Soc.*, **36**, 871 (1940).
12. Astbury, *ibid.*, **29**, 193 (1933).
13. Astbury, *Nature*, **137**, 803 (1936); *Chem. & Ind.*, **60**, 491 (1941).
14. Crowfoot, *Proc. Roy. Soc.*, **A164**, 580 (1938); *Chem. Rev.*, **28**, 215 (1941); Crowfoot and Riley, *Nature*, **144**, 1011 (1939); Riley, D. Phil. Thesis (University of Oxford, 1942); also Crowfoot, private communications.
15. Perutz, Ph.D. Thesis (University of Cambridge, 1940); *Nature*, **149**, 491 (1942); *ibid.*, **150**, 324 (1942).
16. Svedberg and Pederson, "The Ultra-centrifuge" (Oxford, 1940).
17. Chibnall, Bakerian Lecture, *Proc. Roy. Soc.*, **B131**, 136 (1942).
18. Astbury, "X-rays and the Stoichiometry of the Proteins," Vol. III, "Advances in Enzymology" (in press).
19. Riley, Ph.D. Thesis (University of Oxford, 1942).
20. Abderhalden, *Z. physiol. Chem.*, **37**, 484 (1902-3).
21. Ussing, *Nature*, **144**, 977 (1939).
22. Jensen and Wintersteiner, *J. Biol. Chem.*, **98**, 281 (1932).
23. Jones and Moeller, *ibid.*, **79**, 429 (1928).
24. Abderhalden, *Z. physiol. Chem.*, **120**, 207 (1922).
25. Bergmann and Niemann, *J. Biol. Chem.*, **122**, 577 (1937-8).
26. Block, *ibid.*, **105**, 663 (1934); *Yale J. Biol. Med.*, **9**, 445 (1937). See also Vickery and White, *Proc. Soc. Exp. Biol. Med.*, **31**, 6 (1933).
27. See, for example, Block and Horwitt, *J. Biol. Chem.*, **121**, 99 (1937).
28. Atkin, *J. Int. Soc. Leather Trades' Chemists*, **17**, 575 (1933).
29. Astbury and Woods, *Phil. Trans. Roy. Soc.*, **A232**, 333 (1933).
30. Astbury and Atkin, *Nature*, **132**, 348 (1933).
31. Astbury, *Cold Spring Harbor Symposia*, **2**, 15 (1934).
32. Bergmann and Niemann, *J. Biol. Chem.*, **115**, 77 (1936); **118**, 301 (1937); *Science*, **86**, 187 (1937).
33. Lugg, *Biochem. J.*, **32**, 114, 775 (1938).
34. Kassell and Brand, *J. Biol. Chem.*, **125**, 145 (1938).
35. Brand and Kassell, *Federation Proc.*, Pt. II, **1**, No. 1 (1942).
36. Longsworth and MacInnes, *J. Am. Chem. Soc.*, **62**, 2580 (1940).
37. Chao and Taylor, *Proc. Roy. Soc.*, **A176**, 76 (1940).
38. MacArthur, *Wool Rev.*, Jan. 1940, p. 9; Rep. of the Int. Wool Secretariat; and in press.
39. Dakin, *J. Biol. Chem.*, **44**, 499 (1920).
40. Bergmann, *ibid.*, **110**, 471 (1935).
41. Bergmann and Stein, *ibid.*, **128**, 217 (1939).
42. Ussing, *Nature*, **144**, 977 (1939).
43. Fürth and Minnibeck, *Biochem. Z.*, **250**, 18 (1932).
44. Kossel and Gross, *Z. physiol. Chem.*, **135**, 167 (1924).
45. Vickery, *J. Biol. Chem.*, **132**, 325 (1940).
46. Bear, *J. Am. Chem. Soc.*, **64**, 727 (1942).

Catalysis As a Biological Factor

JEROME ALEXANDER

Introduction

The complex and multiform phenomena of life and living units (bionts) confront us with two apparently incongruous and irreconcilable facts. Most notable is the tendency of bionts to breed true. Regardless of wide differences in food and surrounding conditions, the fundamental processes of life appear, as a rule, to be unerringly shepherded along definite paths and sequences; so that despite the vagaries of wind and wave the ship of life maintains a fixed course. Closer study, however, reveals the opposing fact that the course of life, whether of the individual or of the mass, never does run smooth. We have ample evidence of disease, abnormalities, and idiosyncrasies, of mutation and evolution, no matter how we may differ in our attempts to explain these deviations from ideal normality.

Catalysis, involving the direction of definite chemical changes by surfaces of definite structure under definite conditions, seems to be the only basic principle at present known which can mainly account for both the normalities of life and the many deviations from normality. This does *not,* of course, mean that no other factors or principles are operative, especially in relatively large and complex organisms and at structural levels lower and higher than the molecular. But in final analysis it will be seen that these other factors influence the functioning of the biocatalysts—genic, enzymic, structural or symbionic—which dominate the chemical changes whereby life exists, persists, and proceeds.

If biocatalysts and the conditions of catalysis remained unchanged, we would not expect any change in the biochemical output, which underlies all larger structure, morphology, and function. Where we find a change in the biochemical output under apparently unchanged conditions, we are strongly justified in suspecting that some catalyst or catalyst system has undergone an intermolecular or interparticulate readjustment, that something has been added to or subtracted from active catalyst areas, or that new catalyst areas have been formed. These catalyst changes recall the well-known commercial practices used in preparing very active and specific catalysts, which aim not only to keep harmful substances from the catalyst surface, but also to insert there very small amounts of substances termed "promotors," because they direct or favor the formation of desired end products.

It is believed that a very wide range of biochemical and biological phenomena become easily understandable on the basis of specific catalysts and specific catalyst formation and/or changes; and also that no other simple, basic, and ubiquitous phenomenon is equally consonant with the experimental evidence over so wide a range.

Catalysis in General

Time is of the essence of all biological happenings. It is, therefore, purely academic whether, (as many texts put it), catalysts simply speed up reactions which would occur over the lapse of indefinite time, or whether (as seems more probable) catalysts initiate, in many cases, at least, reactions which would not otherwise occur.

A relatively tiny quantity of catalyst may determine the building up or breaking

down of enormous numbers of molecules. The "key" of a "zipper" fastener, commonly used on bags, tobacco pouches, sweaters, dresses, and other articles of clothing, has some analogies to a catalyst.[6a] Drawn in one direction, it massages the opposing "hooks" of the closure so that they engage each other and hold firmly; on reversing the motion of the "key," the opposing "hooks" are disengaged and the closure opens. As far back as 1898, Croft Hill [37] showed that the action of catalysts may be reversible. Using a biocatalyst, the yeast enzyme maltase, he built up from glucose what was first thought to be maltose, a disaccharide, but what was later shown to be isomaltose.

The analogy between "key" and catalyst may be carried a bit further. Recalling the dictum of Emil Fischer [56] that enzyme and substrate must be adjusted to each other like lock and key, we see that there is a mechanical specificity between any zipper key and the hooks that it can influence. At temperatures approaching the softening point of the metal of which the key and the hooks are formed, we would expect the opening and closing processes to fail. Prolonged use might wear down the key so that it would no longer function; or a deformation of the key or the hooks, an adhering drop of paint or solder, or a bit of wire might bind the key fast. Similarly catalysts are highly specific in their action, and they may be utterly ruined, or temporarily or permanently obstructed, masked, or "poisoned."

How Catalysts Function

The view that catalysts function by virtue of their externally directed electronic fields of force becomes clearer when we consider recent revelations of molecular electronic contour by the x-ray spectrometer. Figure 1 shows what may be termed a surveyor's map of a molecule of nickel phthalocyanine, and Figure 2 shows an electronic surface contour map of the same molecule, foreshortened because it is set at an angle in its crystal. Compare this with Figure 3, showing a similar map for phthalocyanine itself, and note the changes in electronic contour about the center of the molecule, brought about by the introduction of a single nickel atom. The importance of such a single atom is indicated by the fact that A. H. Cook [36] found that iron phthalocyanine, alone of the metallic phthalocyanine derivatives tried, possesses marked catalase properties (catalysing the decomposition of hydrogen peroxide). This is especially interesting because of the similarity of the structure of phthalocyanine to that of porphyrin, a constituent of hemoglobin, of chlorophyll, and of cytochrome and other enzymes.

If instead of a single molecule we consider a small group * of molecules, the complications of structure are much greater. Figure 4 shows a number of naphthalene molecules assembled as they are in a single surface cleavage layer of a naphthalene crystal, and Figure 5 shows the type of electronic surface presented if a naphthalene crystal is cut across.

Berzelius,[21] who coined the word *catalysis*, keenly anticipated the modern view that catalysis is consequent upon distortion of reactants by the electronic fields of the catalyst, for he said:

"It is then proved that several simple and compound bodies, soluble and insoluble, have the property of exercising on other bodies an action very different from chemical affinity. By means of this action they produce, in these bodies, decompositions of these elements and different recombinations of these same elements to which they themselves remain indifferent.

"This new force, which was hitherto unknown, is common to organic and inorganic nature. I do not believe that it is a force quite independent of the electrochemical affinities of matter; I believe, on the contrary, that it is only a new manifestation of them; but, since we cannot see their connection, and mutual dependence, it will be

* The group may be small, though the number of molecules may run into the hundreds or thousands.

more convenient to designate the force by a separate name. I will therefore call this force the catalytic force, and I will call catalysis the decomposition of bodies by this force, in the same way that one calls by the name analysis the decomposition of bodies by chemical affinity."

As the electronic patterns of catalysts are more or less thermolabile, temperature

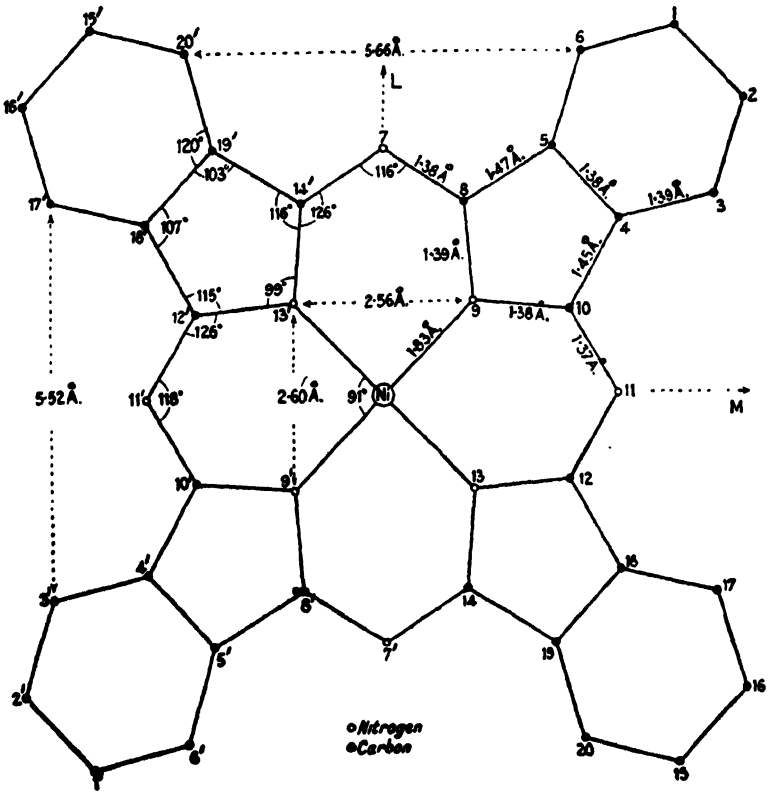


FIGURE 1. Dimensions of the Nickel Phthalocyanine Molecule (Robertson).

is an important factor in catalysis. What we measure as the "temperature" of any system represents an integration of the kinetic activity of its particles over an interval of time long enough to affect a measuring instrument (thermometer, thermopile). During this interval, individual atomic or molecular units of the system will undergo wide variations in their internal electronic configurations and therefore in their outwardly directed fields of force; and individual particulate kinetic units will vary widely in their rotation and free path motion from one instant to another. This raises the question as to whether any *individual* particle has a *measurable* temperature, or, indeed, any lasting temperature at all. Various kinds of radiation (light, ultraviolet, infrared, x-rays) may produce an effective activation of atoms or of molecules without any determinable increase in temperature, although some of the atoms or molecules would be momentarily brought into a state of "excitation" constituting a prelude to chemical change.

After completion of a catalytic change a statistical chemical analysis gives us the final results, but an understanding of the mechanism whereby the results were reached will often aid in controlling or modifying them.[2]

In his monograph on mechanisms of biological oxidations, David H. Green states (p. 9):[60] "The next problem is to explain how the substrate undergoes oxidation when combined with the enzyme. On that point experiment is silent though theory is eloquent. It is stated that the enzyme 'activates' the substrate or that a 'strain' is

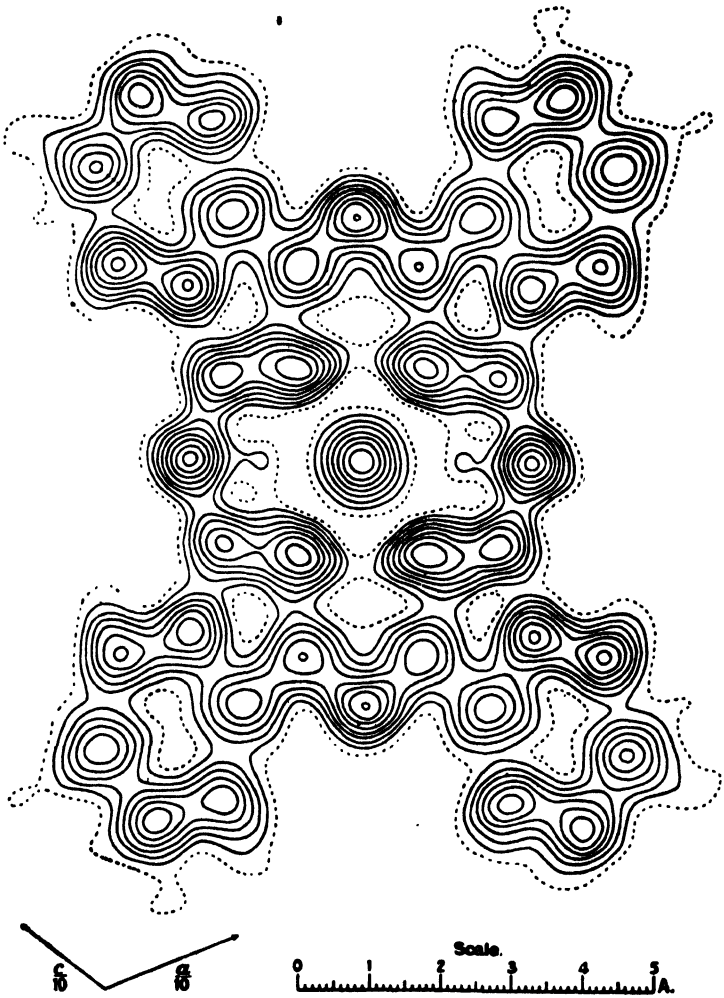


FIGURE 2. Projection of Nickel Phthalocyanine Along the *b* Axis, which makes an Angle of 44.2° with the Molecular Plane. (J. M. Robertson.) Each contour line represents a density increment of one electron per Å.^{2}, except the nickel atom where the increment is five electrons per Å.^{2} for each line. The one-electron line is dotted.

imposed on the substrate molecule. These are terms of ignorance and merely express the experimental fact that the substrate undergoes an oxidation in presence of the enzyme which it does not undergo spontaneously. The nature of activation or straining of the substrate molecule is the spearhead of present research in biological oxidations."

However, when we envision the activation of a substrate as due to a specific strain exerted upon its specific electronic areas by the specific electronic areas of the

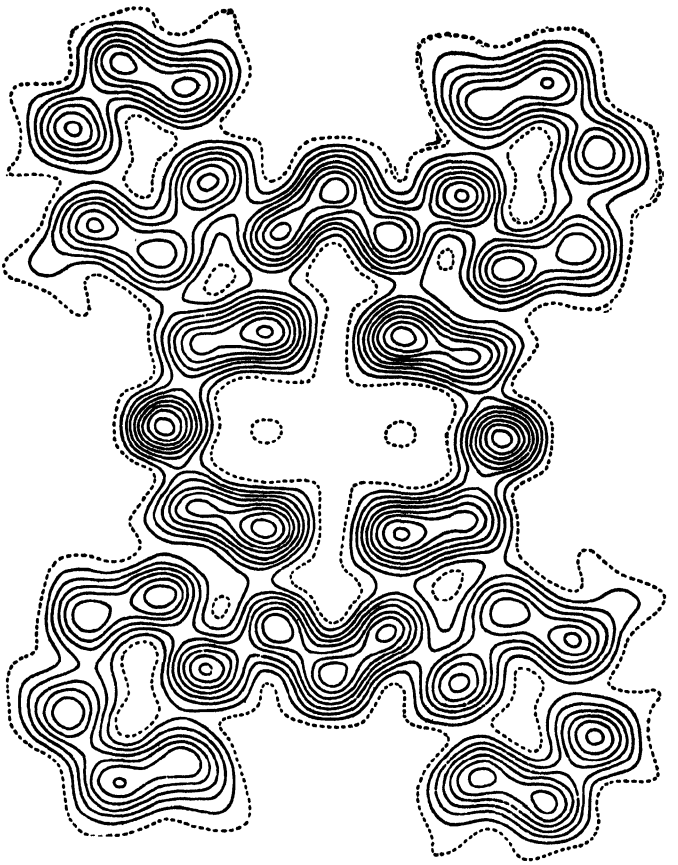


FIGURE 3. Projection of Phthalocyanine along the *b* Axis (J. M. Robertson).

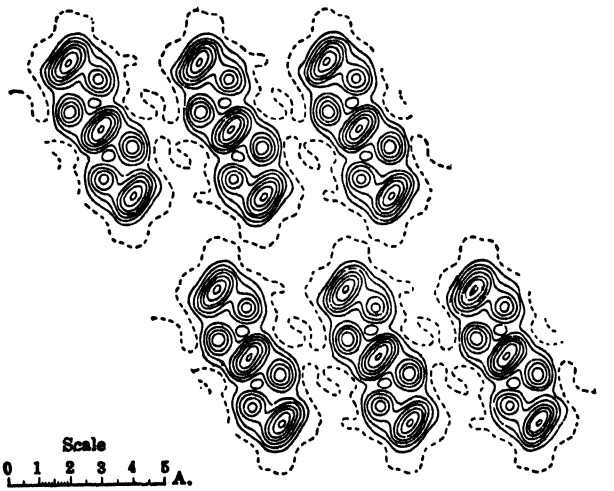


FIGURE 4. Projection of Naphthalene Molecules Along the *b* Axis, showing Mutual Relation of Molecules. Each Line Represents One Electron per $A.^2$ (J. M. Robertson.)

catalyst, we begin to approximate the "screw and nut" analogy of Pasteur and the "lock and key" analogy of Emil Fischer, and do somewhat more than merely express the experimental fact. We see open before us a simple physico-chemical principle, closely dependent on chemical structure and therefore indefinitely variable, which presents, nevertheless, a readily comprehensible mechanism for understanding both the mode of functioning and the extreme specificity of enzymes and other catalysts. Naturally, we are then confronted with the enigmas of electronic contours, which physicists can measure and delineate in some cases. Referring to the violent evolution of O_2 when H_2O_2 is added to a strong catalase solution, Green says (p. 17): "The available evidence indicates that the excessively short life period of the complex [between the substrate and the enzyme] explains why addition of H_2O_2 does

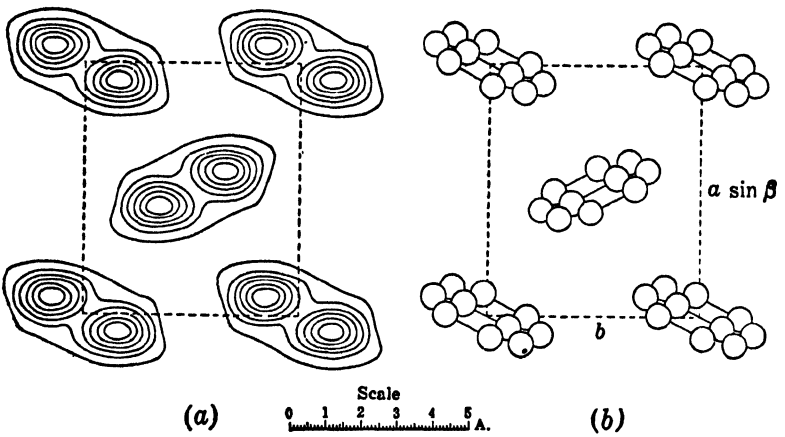


FIGURE 5. Projection of Naphthalene Along the *c* Axis. Each Contour Line Represents two Electrons per Å.^{2} (J. M. Robertson.)

not affect the [spectral] absorption bands of catalase." But these bands are greatly changed by substances that are fixed by the catalase, *e.g.*, KCN, NaF, H_2S. Whether the catalase-H_2O_2 complex is to be considered a "chemical compound" or not will depend upon our definition of this expression. But the taxonomic tag will not change the physical fact.

The Industrial Importance of Catalysis *

A constantly increasing percentage of the annual output of sulphuric acid, a basic raw material of chemical industry, is being produced by passing a mixture of air and sulphur dioxide over vanadium or platinum catalysts.† The chamber process, which for so many years dominated the field, may soon survive mainly in textbooks.

At the Eighth International Congress of Applied Chemistry (New York, 1912), German chemists announced the perfection of the Haber process, whereby hydrogen and nitrogen are catalytically combined to form ammonia gas, which can be catalytically oxidized to nitric acid. Germany was thereby rendered independent of importations of sodium nitrate (Chile saltpeter) for explosives, agriculture, dyes, etc.; and she began the long-planned war in 1914, as soon as the new Haber plants were in full operation. To illustrate how large an amount of chemical change may be

* See "Catalysis, Inorganic and Organic," by Sophia Berkman, Jacques Morrell & Gustav Egloff. Reinhold Pub. Co., New York, 1940. In 1941 about 60 million gallons of methanol were made catalytically in U. S. A.

† Apart from lower cost, vanadium catalysts have the advantage over platinum of not being "poisoned" by arsenic, a common impurity in the sources of sulphur dioxide.

effected by a small amount of catalyst, it is said that 1½ to 2 ounces of platinum distributed on silica gel can produce daily one ton of 100 per cent sulphuric acid, and that under optimum conditions one ounce of 10 per cent rhodium-platinum will catalyze the formation of over one million pounds of nitric acid.

In the organic field, enormous tonnages of petroleum, natural gases, and by-product gases from cracking petroleum are treated catalytically to produce a great variety of organic chemicals, many of which, even if known, had never before been produced in commercial quantities. Because of an abundant supply of coal tar in Europe, technical organic chemistry there revolved largely about aromatic substances derivable from benzene, phenol, naphthalene, anthracene, etc. The abundance of natural gas and petroleum in America beckoned our newly developing organic chemical industry into the aliphatic field, where extensive use is made of catalysis. Formaldehyde is made by the catalytic oxidation of methane as well as of methanol; and methanol may be synthesized by passing carbon monoxide and hydrogen over chromium and manganese oxide catalysts.* Large tonnages of unsaturated fats and oils are hardened by hydrogenation with the aid of nickel catalysts.

The chemical nature of the catalyst may determine the course of chemical change. To take an extremely simple case, the facts with formic acid are as follows:

(1) Pd, Pt, Cr, Ni, Cd, and the oxides of zinc and tin catalyze the following type of breakdown:

$$H{\cdot}COOH = H_2 + CO_2$$

With Pd and Pt sponges total decomposition occurs at 245° and at 215° respectively.

(2) TiO_2 and the blue oxide of tungsten give a different result:

$$H{\cdot}COOH = CO + H_2O$$

(3) SiO_2, ZrO, and uranous oxide catalyze the decomposition thus:

$$2H{\cdot}COOH = H{\cdot}CHO \text{ (formaldehyde)} + CO_2 + H_2O$$

Thoria can catalyze any of these three types of decomposition, depending on temperature.

Catalyst Carriers, Poisons, Promotors, and Modifiers

The technological and patent literature has long reflected the fact that the introduction of new atoms or molecules into the active surface of a catalyst may stimulate, inhibit, or alter the nature and/or rate of the catalyzed reaction.[76] If a desired result is accelerated, the added substance (as mentioned above) is termed a promotor; if it is inhibited, the added substance is called an inhibitor or poison. In cases where a catalyst can direct the formation of several substances and an inhibitor is found which depresses undesired reactions, we speak of "beneficial poisoning." In a general sense any substance which changes the behavior of a catalyst may be called a modifier, a term free from implication as to the result of the change. The importance of *catalyst modification* in biology will be referred to later on.

One factor in the efficiency of a catalyst is the amount of free active surface it exposes to the reactants. It is common commercial practice to distribute the catalyst substance on carriers, or supports, such as asbestos, charcoal, alumina, diatomaceous earth (kieselguhr), etc., which were supposed to act merely mechanically. It has been found, however, that the carrier often exercises a marked effect, either by contributing small amounts of impurities to the catalyst surface (which may thus be either promoted or poisoned), or else by forming with the catalyst molecules some new composite surface of different activity.

To illustrate the importance of the carrier, it may be mentioned that Adkins, Richards and Davis [8] found marked carrier effects in the catalytic dehydrogenation of hydromatic compounds containing a completely or partially saturated benzene,

* Other catalysts are also used commercially.

naphthalene or phenanthracene nucleus. In the case of decahydronaphthalene, the yields of naphthalene varied not only with the catalyst but also with the carrier.

Catalyst	% yield of naphthalene from deca hydronaphthalene
Pt on charcoal	87
Ni on charcoal	34
Ni on chromium oxide	78
Ni on kieselguhr	62
Ni on Alumina I	0
Ni on Alumina II	11
Ni on Alumina III	0
Ni on Alumina IV	36

Note: The charcoal was purified Norit. Alumina I was prepared through sodium aluminate; alumina II, III, and IV, by hydrolyzing alumina isopropoxide by three different procedures.

The authors remark that nickel on alumina and on kieselguhr are better catalysts for the dehydrogenation of substituted cyclohexanols than is platinum, because they show less tendency to induce condensations. Nickel on chromium was the most active catalyst in the case of resistant saturated hydrocarbons. In all, 28 representative compounds, including hydrocarbon alcohols, ketones, and ethers, were heated at 300° to 350° in the liquid phase under nitrogen, in the presence of the catalyst and also of benzene, which serves as a hydrogen acceptor or oxidizing agent.

The potent effect of catalyst poisons is brought out by the remarks of H. Bernthsen at the Eighth International Congress of Applied Chemistry (1912) relative to the iron catalyst used in the first step of the Haber process:

"Extremely minute quantities of these bodies (impurities), which are almost always present even in the purest commercial products or in so-called pure gases, suffice to render the catalysts absolutely inactive or at least to diminish their activity very seriously. Thus iron, for example, prepared from ordinary iron oxide with a content of one per thousand of sodium sulphate is, as a rule, inactive. Iron containing 0.1 per cent sulphur is generally quite useless, and even with 0.01 per cent is of very little use, although in appearance and when examined with the ordinary physical and chemical methods no difference at all can be detected as compared with pure iron.

"The recognition of these facts gave rise to two problems: (A) The preparation of contact masses free from poisons or the removal of poisons from them; and (B) freeing the gases to be acted upon catalytically from all contact poisons. A trace of sulphur, one part per million, in the gas mixture, can under certain conditions be injurious, so that electrolytically prepared hydrogen must generally be further specially purified."

Since a small amount of a catalyst may, if given sufficient time, direct an enormous amount of chemical change, we can understand how minute may be the quantity of promotor or modifier needed to produce a great change in the final output.

Experimenting with the catalytic production of methanol from carbon monoxide and hydrogen, Sir G. T. Morgan[96] observed that while catalysts made by calcining equimolecular mixtures of manganese and chromium nitrates gave methanol containing only traces of higher alcohols, catalysts prepared by precipitating a mixed solution of oxides of manganese and chromium in caustic potash gave a product containing appreciable percentages of higher alcohols. This led him to try the effect of regulated additions of alkali metals to the catalyst. To quote one outstanding case, when the catalyst contained 15 per cent of rubidium hydroxide, the carbon in the gases passing over the catalyst was distributed as follows: (with equimolecular proportions of the pure Mn and Cr oxides the yield of methanol was 80.5 per cent)

Methanol	41.5%
Ethanol	1.6
Higher alcohols	36.7
Aldehydes, acetals, ketones	15.5
Methane	2.0
Carbon dioxide	2.0

The higher alcohols in this case consisted chiefly of isobutanol, but contained besides normal propanol, 2-methylbutanol, 2-methylpentanol, and 2, 4-dimethylpentanol.

Morgan found that the addition of cobalt to the catalyst gave mainly straight-chain alcohols, rather than the branching-chain alcohols formed by alkalized catalysts. This is shown by the following table which gives in *parts per thousand,* the alcoholic content of the *liquid* formed by passing the same mixture of carbon monoxide and hydrogen over different catalysts, under like conditions of temperature and pressure.

Alcohols Formed	Catalyst Used					
	Rb-Cr-Mn		Co-Mn-Zn-K		Co-Cu-Mn	
Methyl	420	—	198	—	220	
Ethyl	12	—	86	—	200	
n-Propyl	43	—	17	—	50	
iso-Butyl		69	—	11	—	3
n-Butyl	—		4	—	16	
β-Methylbutyl	—	8	—	1.5	—	2
n-Amyl	—		1		6	
β-Methylamyl	—	6.5		—		—
n-Hexyl	—		—		2	—
n-Heptyl	—		—		1	—
Unidentified	89		1.5			

In concluding his address before the Société Chimique de Belgique [97] Morgan said:

"All these experiments are of special significance, for they permit us to see the enormous variety of organic compounds which can be synthesized by secondary reactions following primary condensation of carbon monoxide and hydrogen, a gaseous mixture well known under the name of 'water gas,' made on a large scale from coal and steam.

"Up to the present time these condensations have been studied at temperatures higher than the decomposition of many organic compounds. With higher pressures and more efficient catalysts, we may some day expect to lower the reaction temperature very materially. When we shall have reached this point, the synthesis of organic compounds of great complexity should be possible and among these products we will find organic substances which up to now can be formed only by the chemical activity of living beings."

Some Kinetic Aspects of Catalysis; Zone of Maximum Colloidality

The catalysts used in industry are sometimes aggregated colloids (*e.g.,* platinum sponge, nickel used in hydrogenation); but generally they are extended on carriers. This not only exposes a large active surface, but it also permits the finely divided catalyst to remain in place when gases or liquids are passed over it, and to be filtered off if it has been mixed into a batch.

Some biocatalysts are fixed, like genes in the chromosomes; but most of the enzymes, which, with the genes, mainly direct the course of chemical change in bionts, are free particulate units of colloidal dimensions. The establishment of colloidal dimensions may come about in several ways. Some molecules are born col-

loids; some achieve colloidality by molecular growth or aggregation; and some have colloidality thrust upon them by adsorption or chemical fixation on a colloidal carrier or an extended surface.

Whatever the underlying cause, increase in particle size results in decrease in mean kinetic free path motion, as may be readily seen in an ultramicroscope. Particles at the lower ranges of microscopic resolution (about 0.25 micron) show only the feeble Brownian movement. On the other hand the kinetic motion of small molecules is so rapid that if we could see a hydrogen molecule we could not see it, any more than we can see a rifle bullet in flight. Between these two extremes lie all degrees of particulate kinetic motion, which increases at an accelerated rate as molecular dimensions are approached.

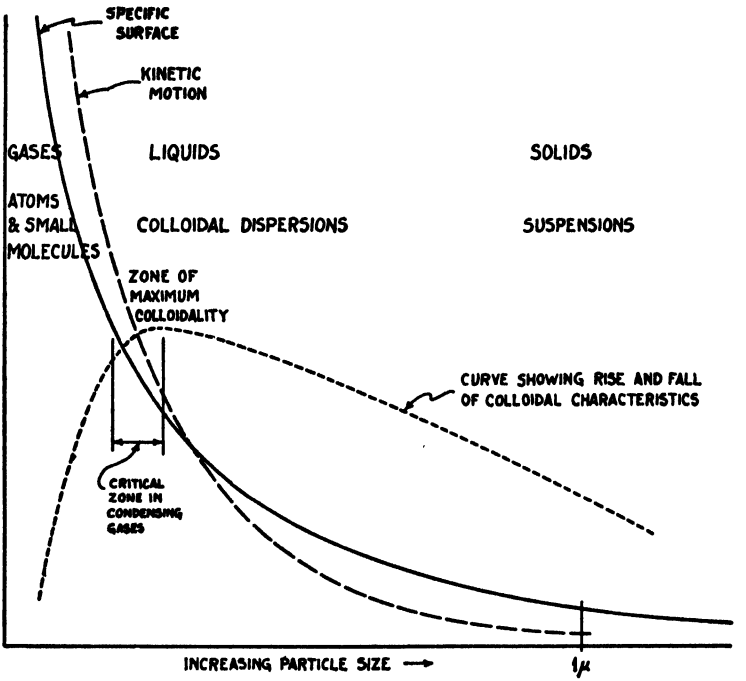


FIGURE 6. Degree of Colloidality Related to Particle Size.

But decrease in particle size is accompanied by great increase in specific surface —the amount of surface exposed per unit weight. That degree of dispersion where specific surface may exert its most effective influence before rapidly increasing kinetic activity dominates has been termed the zone of maximum or optimum colloidality.[6] Figure 6 gives a *diagrammatic* sketch of these relationships, based on the *assumption* that the particles are spheres, which, of course, is seldom the case. Particles have various, and even varying contours (threads, rods, plates, aggregates), so that the diagram simply illustrates the principle.

When a particle approaches an active catalyst area, two factors are of outstanding importance:

(1) the momentary electronic contours of the surfaces in apposition, for this determines the *possibility* of particulate union or influence;

(2) the particulate kinetic velocities of translation and rotation of moveable areas, which determine whether the possibility may become a *reality*.

The importance of this kinetic factor came to the fore in the course of a lecture experiment * made at the Toronto meeting of the British Association for the Advancement of Science (1924) by Sir Ernest Rutherford, and was designed to illustrate how a positively charged atomic nucleus will repel an alpha particle emitted by radium. An electromagnet, with its positive pole up, was fastened to a table so that another electromagnet, swinging from the ceiling with its positive pole down, could just pass over it. The mutual repulsion of like magnetic poles caused the swinging magnet to take a parabolic path in any *off-center* approach. To illustrate the relatively rare recoil which occurs when an alpha particle makes a *direct* approach to an atomic nucleus, Sir Ernest took careful aim—but the swinging magnet passed completely over and beyond the fixed one, probably giving a slight unseen "jump" as it did so. Quickly retrieving the swinging magnet, he swung it again from a lesser distance and got the expected recoil.

From this it is evident that if a certain relative critical velocity is exceeded, particles or surfaces having the power to cohere, will *not* do so. In catalysis, a certain period of *time* is necessary for the electronic fields of the reactants to become fitted to each other, so that apart from the effects of particulate rotation, internal kinetics, and mode of presentation, the particulate kinetic velocity of translation is an important factor.

Increase in thermal agitation increases the *total number* of encounters *per unit of time* between reactant catalyst and substrate areas, and this works in favor of increasing the number of *fruitful* encounters. On continuing to increase the temperature, however, a point is reached where so many particulate units have so high a relative velocity, that the number of *unfruitful* encounters increases to such an extent that the catalyst efficiency per unit of time *tends* to fall off. Many other intercurrent factors may influence the thermal optimum in catalysis besides the kinetic factors; *e.g.*, the catalyst, reactants, or end products may undergo decomposition; adsorbed substances may "lame" the catalyst; specific substances or ions may affect the degree of dispersion or electronic integrity of catalyst or reactants. Thus Svedberg found that many proteins are dispersed into smaller sub-units by change in pH, some of them being reconstituted from the fragments when the pH is shifted to the stability zone; and many substances will "dissolve" or disperse glue at room temperature, *e.g.*, calcium chloride, sodium nitrate, sodium naphthalene sulphonate. The union of carrier and prosthetic group in enzymes is often delicately balanced (see below).

Pepsin is "activated" by hydrochloric acid. What this means kinetically may in part be followed ultramicroscopically.[12] A dilute solution of egg white heated nearly to boiling gave an opalescent dispersion full of bright, rapidly moving ultramicrons. On allowing a droplet of pepsin solution (Fairchild's, containing 15 per cent alcohol) to diffuse under the cover glass of the slide, the albumin ultramicrons immediately coagulated into large, motionless masses. When a droplet of 0.1*N* hydrochloric acid was introduced under the cover glass the coagulated masses burst into small groups and isolated ultramicrons, which resumed their active kinetic dance. But almost immediately the albumin particles began to grow fainter and to disappear, the field meanwhile becoming brighter as smaller ultramicrons or amicrons were formed. The addition of pepsin to the opalescent albumin solution caused it to clear gradually at room temperature.

The kinetic motion of diastase particles may likewise be followed ultramicroscopically, as they gather about and actively rub or "gnaw" holes or cavities in starch granules.[12] The chemical changes occurring are, of course, far below the level of visability.

Recently, Haurowitz and Schwerin [67] found that the catalytic auto-oxidation of linoleic acid by hemin in *heterogeneous* emulsions ceases when the system is made

* This was referred to in Vol. II.

homogeneous by addition of acetic acid, dioxane, pyridine, alkali or bile, but recommences when homogeneity is abolished by addition of water or stronger acid. These authors believe that this reversal demonstrates that the catalysis takes place only in the interfacial film between the water and oil phases, and attribute the observed diminution of the velocity of oxidation by excess of hemin to displacement of linoleic acid molecules from the interfacial film. They conclude that orientation of substrate and hemin molecules in the interfacial film is of decisive importance for the establishment of the catalysis. This is, of course, true, and it involves the electronic contour factor (1) mentioned above; but the kinetic factor (2) is also to be reckoned with.

Biocatalysts; Prosthetic Groups; Symplexes *

The course of chemical change in organisms is directed by catalyst units of colloidal dimensions—enzymes and genes, and perhaps also by units in some symbionts, and units fixed at or on various surfaces. Whereas with primal bionts gene-like units were probably the sole directors of life chemistry, in higher forms we find thread-like aggregations of genes (chromosomes) and a great variety of secondary but highly specific catalysts (enzymes), whose production is in some way interwoven with the activities of the truly living, self-reproducing catalysts. The experimental evidence indicates that large numbers of biocatalyst units are destroyed or inactivated in the course of their functioning, so that there must be a continual replenishment of the supply.

How do biocatalysts arise? The broadest, and therefore the most indefinite answer to this question would be: by the assemblage of smaller units (atoms, molecules, colloidal particles) so as to constitute an exposed surface which has a suitable electronic structure and contour. Genes in the chromosome string thus form duplicates of themselves (autocatalysis),[18] and must therefore be considered as living. On the other hand, an active catalyst area may be formed by being built into or upon a larger fixed surface, *e.g.*, a cell wall; or else there may be formed by adsorption and/or chemical combination, colloidally dispersed units—the multitudinous enzymes—which can flow and kinetically "swim" about in cytoplasm, sap, or body fluids, and may even diffuse slowly. Texts often give the erroneous impression that colloids will not diffuse; but Thomas Graham [59] clearly pointed out that they may diffuse, though they "are slow in the extreme."

The catalytic effect of the walls of the containers in which chemical reactions are carried out is not, as a rule, sufficiently appreciated, and traces of substances existing in, forming on, or introduced into these surfaces may powerfully influence results. Thus a food product processed in aluminum vessels was deleteriously affected by tiny fragments of iron left in the aluminum surfaces when these were scoured with steel wool. G. Bredig [20] found that by fixing amino groups to fibers (cellulose, wool, silk) organic catalysts were produced which split off carbon dioxide from bromcamphocarbonic acid. Although inorganic catalysts may appear to be simple, *e.g.*, platinum black, those having first-hand experience with commercial catalysis know how potent is the influence of structure, carriers, and impurities on the functioning of such catalysts.

Biocatalysts are composed mainly of complex organic molecules, assembled into a delicate but specific structure. When a certain molecular group enters a biocatalyst to form and characterize the active catalyst area, it is termed a *prosthetic* group. The word "prosthetic" is derived from a Greek root meaning to *add to*, or *to insert*. Thus prosthetic dentistry deals with the insertion of missing teeth. The prosthetic group, so to say, puts "teeth" into the biocatalyst. Often the prosthetic group demands for its functioning the presence of a single atom, *e.g.*, magnesium with chlorophyll, copper in the polyphenol oxidase of mushrooms and in tyrosinase,[16] iron in

* See paper by A. E. Axelrod and C. Elvehjem in this volume.

catalase, cytochrome, etc. Where "homeopathic" doses are effective, their action may often be understood along these lines.

Richard Willstätter [146] proposed the term *symplex* for compounds where high-molecular substances are bound by residual valencies, *e.g.*, a prosthetic group and a colloidal carrier. Symplexes are distinguished from mere mechanical mixtures by one or more of the following characteristics: (1) alteration or enhancement of specific reactivity of one component; (2) change in solubility or dispersion of one component; (3) change in optical properties; (4) change in stability; (5) change in toxicity; (6) change in reactions, *e.g.*, color reactions.

True chemical combination is not necessary to make separate units acquire new properties when combined. Neither an arrow-head, an arrow shaft, nor a goose feather is an arrow; but the three, when properly fitted together, will make an arrow and will function as one.[10, 135] (See also J. Alexander, in Allen's "Commercial Organic Analysis," 5th ed., Vol. 10, 1933.)

Structures of the "symplex" type are very loosely bound together. E. L. Smith [121] found that solutions of phyllochlorins (chlorophyll-protein compounds extracted from spinach leaves) may be split into free chlorophyll and protein by the detergents sodium desoxycholate, bile salts (mainly sodium glycocholate) and digitonin. In the presence of sodium dodecyl sulphate (SDS), the prosthetic group remains attached to the protein, but the compound is split into smaller units, the protein properties and absorption spectrum being modified. Tobacco mosaic virus is also split by SDS into smaller fragments, nucleic acid being simultaneously separated from the protein.[126] In the presence of SDS chlorophyll loses magnesium and becomes phaephytin; and this substance, or the chlorophyll (depending on pH), remains attached to the protein, since the prosthetic group is not separated by ultrafiltration, dialysis, or fractional precipitation. Smith believes that much previous work on chlorophyll dealt only with the prosthetic groups of extremely complex specific catalysts. Emerson and Arnold [50] concluded from photochemical studies that 2,500 chlorophyll molecules form one functional unit in photosynthesis.

The importance of the carrier in the biological field is indicated by the following two examples. Keilin and Mann [88] report that peroxidase, which shows great resemblance to methemoglobin, can be considered as a compound of protohematin with a native protein. Peroxidase forms two highly unstable compounds with H_2O_2, whose decomposition is much accelerated by the acceptor present in plant extract or by the addition of other acceptors, *e.g.*, ascorbic acid, hydroquinone, or pyrogallol. "The same hematin nucleus combined with three different native proteins forms three distinct compounds: methemoglobin, catalase, and peroxidase, which have many properties in common but show, however, striking differences in the nature and magnitude of their catalytic activities." More recently Warburg and Christian [143] report that the old "yellow enzyme" may really be an artifact resulting from the loss of adenylic acid during its preparation. They describe five different yellow enzymes, some with similar proteins but different prosthetic groups, others with the same prosthetic group but different proteins. The protein/prosthetic group combination is reversible, and it is estimated that one molecule of alloxine dinucleotide can transfer 1440 molecules of O_2 per minute. Vegetables and fruits are "blanched" by steam before being dehydrated, to inactivate enzymes which cause undesirable changes.

This gives us some insight as to the potency and flexibility of the enzymic catalyst systems of living cells, and effectively answers the antiquated gibe of orthodox organic chemists that whenever a biological reaction is to be explained, a new enzyme is invented.

The formation of chlorophyll seems to demand the presence of iron, just as the formation of erythrocytes in man demands the presence of copper. The mode of formation of prosthetic groups is generally obscure,[86] but it has recently been shown that thiamin (vitamin B_1) is combined with pyrophosphoric acid,[134] and D. E.

Green states:[60] "There is ample evidence of the existence in animal tissues and microörganisms of enzymes which catalyze the phosphorylation of thiamin as well as the dephosphorylation of diphosphothiamin." Some animals (man) must eat vitamin C; others (rats) can synthesize it, or in any event do without it in their food.

Nature, with infinite time and opportunity for experiment, accomplishes results in manifold and devious ways. Thus the ascidian *Phallusia mamillata*[70] has an acid blood (3 per cent H_2SO_4) which contains as chromogen a non-dialysable organo-vanadium compound. On plasmolysis this gives a brown solution, which yields, on drying, a dark-blue powder showing over 10 per cent vanadium (two analyses gave 10.36 and 15.4 per cent, but the latter figure is doubtful). Mussels have a manganese-containing catalyst, pinnaglobin. Most crustaceans, including *Limulus* (the horse-shoe or king crab, of antediluvian ancestry) are literally blue-blooded because of copper-containing hemocyanin. A group of African birds known as the Touracous or Plantain-eaters have in their pinion feathers red or crimson patches or portions from which weak alkali or soap solutions will extract a pigment, which, after precipitation by acid, dries to a rich crimson solid[84] called turacin. This has been recently proven to be a copper porphyrin compound.[20] Over sixty years ago Professor James F. W. Johnston (University of Durham, England) made the following comment:[80]

"The existence of an animal pigment so rich in copper as turacin (about 8 per cent), offers many interesting problems for study. Traces of this metal seem generally diffused in most vegetables and many animals; but here are more than traces —weighable and visible quantities. It is true that these plantain-eaters have been seen to pick up in their native countries grains of malachite, the green mineral carbonate of copper; but we must rather look to the vegetable food they consume as the true source of this metal. And when the copper is ingested, how does it find its way, in the complex pigment of which it is an essential part, precisely to those feathers, and to those barbs of feathers, and to those parts of such barbs, which are red, and not to the black portions? For if one of these feathers is burnt in a Bunsen gas-burner, not till the red part of the feather is reached will the green flash of the copper tinge the flame. However, in the *crest* of the violet plantain-eater (*Musaphaga violacea*) and perhaps traces in blood of all these birds, turacin, and therefore copper, does occur. Still the whole mystery of this strange pigment is far from being understood."

It has recently been reported that certain molds will not grow in the absence of traces of gallium; and Aguilhon and Sazerac[5] found that the addition of 0.0001 per cent of uranium acetate to a fermentation mixture of *acctobacter suboxidans* and sorbitol increased the yield of sorbose by 76 per cent.

In an article on the biochemistry of microörganisms C. B. Van Niel[187] refers to many other cases where trace substances are vital: *e.g.*, molybdenum for the nitrogen-fixing organism *Azobacter;* copper for common molds as well as for higher plants; boron, apparently needed to form borocitrin, related to the flavin pigments and found in microörganisms by Kuhn. Van Niel also tabulates the recent work of various investigators showing the diverse ways in which different organisms split acetate and normal butyrate.

E. C. Auchter,[18] Chief of the Bureau of Plant Industry, U. S. Dept. of Agriculture, in discussing the interrelation of soils and plant, animal and human nutrition, lists some of the "physiological troubles" of plants which can be cured by supplying the necessary small amounts of certain missing but essential elements. It must, of course, be remembered that larger amounts than the tiny optima may be very harmful, as is the case, *e.g.*, with boron. *Magnesium* cures sand drown of tobacco in the soils of the coastal plain; *manganese* cures chlorosis of tomatoes on some calcareous Florida soils, and permits such soils to give improved yields of potatoes, snapbeans,

cabbage, lettuce, peppers, carrots, beets, citrus and corn; *zinc* cures pecan rosette in the South, citrus leaf mottle in California, South Africa, and Florida, little leaf of apples and other deciduous fruits in the West and Northwest, and white bud of corn on the Norfolk and Hernando fine sands of Florida; *boron* cures internal browning of cauliflower and dry rot of sugar beets in Michigan, crown rot of sugar beets in Ireland, die-back of citrus in Africa, internal cork and drought spot of apples in British Columbia, West Virginia and elsewhere, and cracked stem of celery in Florida; *copper,* added to muck soils in Holland and parts of western New York and to organic soils of Florida, improves the growth of several crops; *sulfur* corrects yellows of tea in Nyasaland (Africa), and improves the growth of field crops in Oregon and other areas.

Since animals depend upon plants, "the soil is the mother of all living things." * Deficiency of cobalt [55] in certain New Zealand soils causes "bush sickness" in sheep, which seems identical with the "pine" of sheep in Scotland. Iodine deficiency leads to goiter in man and beast and caused an annual loss of thousands of pigs in Montana until iodine feeding was practiced. Selenium, taken from the soil by plants, causes disease and malformations in animals eating the plants.[97a] Marco Polo, on his journey to Tartary, observed hoof deformations of this type, which may have had the same kind of origin. According to A. L. Moxon traces of arsenic compounds, fed in water or in salt, tend to protect animals against the toxic action of selenium, while bromo-benzene aids in its excretion.

The eighth scientific meeting of the Nutrition Society held in London, England, Oct. 17th, 1942, was devoted to "Trace Elements in Relation to Health." Among the rarer deficiency diseases mentioned are: enzootic ataxia (swayback of lambs), which may be prevented by feeding traces of copper during pregnancy; molybdenosis (called teart of Somerset), which affects ruminants, due to molybdenum taken up by plants where the soil content of molybdenum was about 100 ppm, and helped by traces of copper sulfate; fluorosis, which may cause osteosclerosis, though teeth with mottled enamel due to fluorine are relatively immune to caries.

It must not be supposed that trace substances always operate in one way, for the complications in any case may be great. Thus sulphur may be oxidized by soil bacteria and affect the local pH—from which a chain of other consequences may follow. However, much of the evidence points to the conclusion that the inclusion of trace substances in catalyst surfaces is a frequent factor that must always be considered; for visible results develop from the products of catalysis.

Biocatalyst Systems or Chains

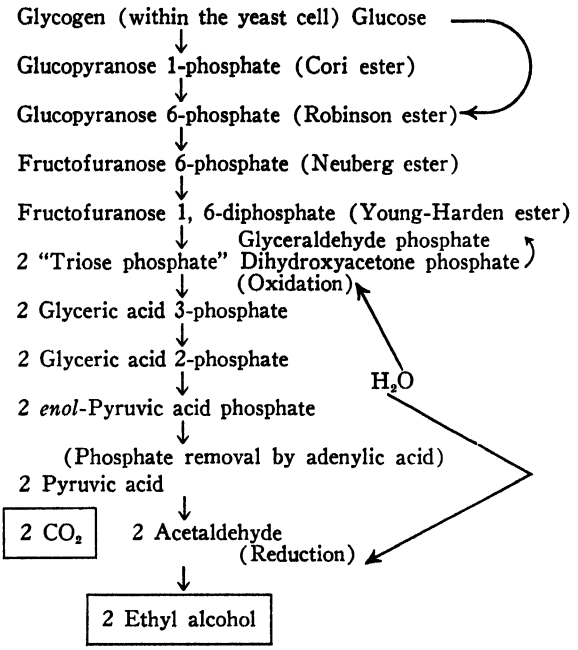
It is we who are simple, not nature. The physico-chemical happenings at various structural levels, which underlie the phenomena we observe, are numerous, intricate, often obscure, sometimes unsuspected. The complications involved in the structure, lability, and functioning of individual enzymes are magnified in cells, tissues, and organisms, where groups of catalysts and coöperative chemicals form chains or systems which have not yet been constructed for experiments *in vitro.* Isolated enzymes, supposed to be pure, have sometimes proven to be mixtures, and it is always a question as to how far experiments with pure enzymes may be applied to an intact biological unit. Liver slices energetically oxidize certain of the lower fatty acids, but when the liver is minced this power immediately disappears.[60, p. 165]

As an illustrative instance, we may take the fermentation of glucose by yeast,[99] widely practiced and long studied. Present experimental data justify the belief that underlying the extremely naive equation

$$C_6H_{12}O_6 = 2CO_2 + 2C_2H_5OH$$

* This statement is too sweeping, even though many forms of sea-life depend upon substances washed into the seas from the land.

the following catalyzed system is operative (end products shown in rectangles):



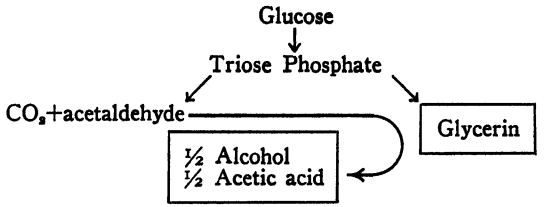
The successive steps tabulated are catalyzed by specific enzymes,* and many of the changes have been shown to be reversible. Besides the final products (ethyl alcohol and carbon dioxide), a small amount of glycerin is also produced. A somewhat more detailed table is given by O. Meyerhof in *Biological Symposia,* 5, 141-155 (1941).

Since the formation of Cori ester from glycogen is inhibited by glucose, enzymic formation of glycogen within the yeast cell, along the lines shown by C. S. Hanes,[68] may serve to restrict glucose concentration as glucose diffuses into the yeast cell.

If sodium sulfite is added to a yeast fermentation, the formation of alcohol practically ceases and the small percentage of glycerin which occurs in normal fermentation increases so that it becomes a main product, with acetaldehyde and carbon dioxide. The acetaldehyde is fixed by the sulfite, the CO_2 is liberated, and the glycerin remains; during the war which began in 1914, Germany produced glycerin by fermentation processes. Bell[20] states that the traces of glycerin in normal fermentation are due to the "mutase" system in the yeast cells. It seems that, normally, acetaldehyde acts as a hydrogen acceptor in the anaerobic oxidation of triose phosphate; but when the acetaldehyde is removed by the sulfite present, another molecule of triose phosphate can act instead. Since the former reaction is the more rapid, little glycerin is formed when acetaldehyde is present.

By carrying out the fermentation at an alkaline pH, Neuberg found that the breakdown of the glucose yielded glycerin, alcohol, and acetic acid, according to the following scheme,[60] which is not yet understood in detail:

* According to A. B. Gutman, F. B. Warrick and E. B. Gutman [*Science,* 95, 461 (1942)] three enzymes are present in calcifying cartilage, *viz.* phosphorylase, phosphoratase and phosphoglucomutase, which act on glucose-1-phosphate. The calcification process is carried further by a glycolytic enzyme system [G. M. Hills, *Biochem. J.,* 28, 864 (1934)].



Not only does the apparently simple yeast cell contain a considerable number of specific enzymic catalysts and collaborating molecules, but it can be "trained" to ferment galactose,[48] by adding small and continually increasing percentages of this sugar to the fermentation mixture. Evidently yeast can "manufacture" new enzymes to meet new situations; probably the new substances add their own specificity to a system of molecules to make a new prosthetic group.[8]

Karström [81] divided the enzymes of microörganisms into two groups; (1) *constitutive enzymes* whose formation is common to all media; (2) *adaptive enzymes,* which are formed only when the cell is "stimulated" by the presence of a specific chemical substance in the milieu, a procedure which seems to be analogous to the formation of antibodies in animals.[118, 8] Experimenting with nine strains of yeast (*Saccharomyces*), Rhoades [115] found that each one could ferment glucose, mannose, sucrose, and raffinose "regardless of the carbon source in which the cells had grown; that is, these sugars were fermented by constitutive enzymes." On the other hand, maltose, trehalose, alpha-methylglucoside and galactose * were fermented only if the organism was "trained" by growth in the presence of the specific substrate or a closely related substance, that is, after the formation of adaptive enzymes.

Crystallization Phenomena

The crystallization of substances formerly believed to be non-crystallizable has astonished many who fail to remember that orderly arrangement in a space lattice is a potential minimum which molecules *tend* to assume, although conditions often enforce the persistence of some intermediate state. Furthermore, misleading texts engendered the erroneous notion (which still persists) that crystalloids constitute a class of *substances* quite apart from the "non-crystallizable colloids," even though Thomas Graham, who coined the word "colloid," had stressed the mutability of the colloidal *state* of matter, and had pointed out that the same substance, *e.g.,* silica, can exist both in crystalline and in colloidal form.

In their natural state many "natural colloids" are probably prevented from establishing their tendency to form a visible crystal lattice by their own inherent and variable complexity, by the presence of protective or deflocculating substances, or by the ionic concentrations of their milieu *e.g.,* pH. By suitable purification and choice of conditions, even rubber hydrocarbon has been crystallized,[122] and P. P. von Weimarn held the view [144] that any substance can be made to assume both the colloidal and the crystalline states. The x-ray spectrometer shows that many colloids are cryptocrystalline, and the ultramicroscope and the electron microscope bring out the assemblage of units at higher structural levels.

According to Professor G. W. Stewart,[128] the x-ray and other evidence indicate that the molecules in the interior of a liquid are found, roughly, in one of two classes: (1) those having comparative freedom; (2) those in semi-orderly array, strongly imitating fragments of crystals. He terms these fluctuating molecular "mobs" *cybotactic groups,* which, in the case of *liquid* para-azoxyanisol, may consist of perhaps 100 to 1000 molecules. In the *liquid crystal* of this substance these companies are ordered into regiments composed of perhaps millions of molecules.

* One strain, *S. cerevisiae Hansen,* possessed the inherent ability to ferment galactose.

Crystallization involves orderly molecular assemblage, and often occurs in the following stages: (1) with drop in temperature or evaporation of solvent, molecular aggregates of colloidal size form under the laws of chance, just as the Faraday-Tyndall haze forms in a condensing gas;[49] (2) these aggregates, which may be crystalline, have a diminished kinetic activity and tend to form spherulites under the influence of surface forces; (3) owing to residual polarity, the spherulites tend to form chains or quasi-crystalline groups which strive toward and often attain the potential minimum of a crystal lattice.[7]

All these stages of crystallization may be found, here and there, in window-pane ice, but they may be produced at any time with a solution of sulfur in carbon disulfide to which has been added some Canada balsam to slow down the velocity of the steps and to fix the results. A drop of the solution is placed on a microscope slide, spread out with the finger and blown upon to speed evaporation of the solvent. On microscopic examination of various fields one generally finds areas showing spherulites, margarites (denditric, pearl-like chains of spherulites), spherulites in crystal-like array, and well-defined crystals. Upon ultramicroscopic examination, the crystals often show at their exterior ultramicrons or spherulites which failed to find lodgment in the crystal lattice.[7, 19, 52]

Molecular Assemblage in Proteins *

Proteins are subject to aggregative changes apart from actual crystallization. Thus I.-B. Eriksson-Quensel and The Svedberg[58] found that the molecules in a solution of the hemocyanin from the snail *Helix pomatia,* show a molecular weight of 6,740,000 when slightly acid (pH 6.8); but when made slightly alkaline (pH 8.0) the protein splits into three components whose molecular weights are 6,740,000, 3,370,000, and 842,000 respectively. On restoring the pH to 6.8, all the fragments of dissociation completely unite to form the original component. Dissociation of proteins may be brought about by high dilution or by the addition on an amino acid or another protein.[109] So minute an amount as 0.001 per cent of thyroxin causes an appreciable dissociation of thyroglobulin (H. P. Lundgren, from T. Svedberg[130]).

The Svedberg states[130] that the action of a dissociating compound on a protein is more or less specific. Thus in the presence of ammonium chloride, arginine dissociates serum albumin but not *Helix* hemocyanin, while lysine plus ammonium chloride splits the latter protein but not the former. Guanidine chloride affects *Helix* hemocyanin very strongly, but serum albumin only slightly. Clupein splits both proteins, but arginine, in the absence of ammonium chloride, affects neither.

These and other experiments of Svedberg and his school show how delicate is the physico-chemical mechanism controlling the particle size of proteins, and indicate how slight changes in biocatalysts, by determining the formation of even traces of specifically potent substances, may exert profound effects on protein structure and function, *e.g.,* as carriers or as prosthetic groups. Particle size and structure will, in turn, determine kinetic activity, diffusion, and the extent and nature of the outwardly directed electronic mosaic. With the biologically important colloid proteins, as indeed with most other substances, we must, at a certain stage of organization, envisage the possibility of particulate unions and dissociations which are not explicable in terms of simple stoichiometry, even though they are due to residues of those same sub-atomic potencies which account for chemical attraction. There comes a zone in the successive levels of increasing complexity of material structure, where the external electronic mosaics of the particles may be somewhat variable and affected by adsorbed "impurities," and yet the attraction—or repulsion—of the particles may be biologically important, although not as powerful or as precise as with true chemical combination between smaller, less highly organized particulate units.

Svedberg and his school have given particular attention to the physiologically

* See paper by W. T. Astbury in this volume.

important *respiratory proteins,* which are also of great interest from the physicochemical as well as the taxonomic point of view. In epitomizing this work, Svedberg [180] distinguishes between: (1) respiratory proteins active in cells—*viz.,* Keilin's cytochrome-C, Warburg's yellow enzyme,* and myoglobin, often called muscle hemoglobin, for it takes up and gives off oxygen along a dissociation curve and probably serves as an oxygen reservoir for the organism; the first two are true respiratory cell proteins, taking part in the enzymic oxidation-reduction reactions; (2) respiratory blood proteins active in external respiration, which fall into the following classes:

(a) red pigments (erythrocruorins, hemoglobins)
(b) green pigments (chlorocruorins)
(c) blue pigments (hemocyanins)
(d) reddish-brown pigments (hemerythrins)

Classes (a) and (b) have similar prosthetic groups containing iron, and are called hematic chromoproteins; they can take up one molecule of oxygen for each atom of iron. With class (c) the prosthetic group contains copper, and these proteins can take up one molecule of oxygen to two atoms of copper. With class (d) the prosthetic group also contains iron, but the oxygen-binding capacity is one oxygen molecule to three atoms of iron.

The cellular respiratory proteins, cytochrome-C and myoglobin, both contain iron in their prosthetic group and have almost the same molecular weight, about 17,000, which equals that of *Lampetra* erythrocruorin and is one-fourth the weight of the normal hemoglobin of vertebrates.

With one exception, the respiratory pigments enclosed in blood corpuscles have small molecular weights, whereas those dispersed in the plasma have large ones. All vertebrates, except species belonging to the lowest class, the *Cyclostomata,* have in their blood corpuscles a protein pigment of the same molecular weight, 68,000 (hemoglobin) with four iron atoms per molecule. With *Cyclostomata* the corpuscle pigment has a molecular weight only one-fourth that of hemoglobin, and according to A. Hedenius, certain invertebrates have a corpuscle erythrocruorin one-half the molecular weight of hemoglobin. Mammalian hemoglobin dissociates reversibly into half-molecules upon addition of certain amino compounds such as urea, acetamide and formamide. The hemoglobin synthesized by Anson and Mirsky from globin and heme, though it had the same molecular mass as the native protein, showed a slightly lower isoelectric point (N. Gralén). The chemical processes employed, therefore, had registered some lasting effect. Erythrocruorins of high molecular weight are found in the blood plasma of the crustacean *Daphnia,* (M = 400,000), of the snail *Planorbis* (M = 1,600,000), and of certain worms, *Arenicola, Lumbricus* (M = 3,200,000).

"The hemocyanins form an interesting class with a number of inner connections. The molecular weights of the hemocyanin molecules found in the blood of certain species are always simple multiples of the lowest well-defined component. Thus, for the *Malacostraca* the relationship is 1:2 and for the *Gastropoda* 2:8:16:24. Moreover, the weights of all the well-defined hemocyanin molecules seem to be simple multiples of the lowest among them. In most cases the hemocyanin components of certain species are interconnected by reversible, pH-influenced dissociation-association reactions. At certain pH values a profound change in the number and percentage of the components takes place. The shift in pH necessary to bring about reaction is not more than a few tenths of a unit. Consequently, the forces holding dissociable parts of the molecule together must be very feeble.

"Not only the molecular weights of the hemocyanins but also the mass of most protein molecules—even those belonging to chemically different substances—show a

* This has recently been separated into five enzymes, the original yellow enzyme having proven to be an artifact.[148]

Molecular Constants of Proteins

According to measurements carried out in the Institute of Physical Chemistry, University of Upsala, Sweden. As of Nov., 1942.
Kindly supplied by The Svedberg and Kai O. Pedersen, Upsala University, Sweden.

V_{20} = partial specific volume of the protein. (The values in brackets are assumed values based on experimental determinations on closely related proteins.)

s_{20} = sedimentation constant = $\frac{dx}{dt}\cdot\frac{1}{\omega^2 x}$, in units of 10^{-13} reduced to water at 20° C.

D_{20} = diffusion constant in units of 10^{-7} reduced to water at 20° C.

M_s = molecular weight computed from sedimentation velocity and diffusion measurements according to the formula:

$$M_s = \frac{RTs_{20}}{D_{20}(1 - V_{20}\rho^\circ)}$$

M_e = molecular weight computed from sedimentation equilibrium measurements according to the formula:

$$M_e = \frac{2RT \ln (c_2/c_1)}{(1 - V\rho)^2\omega^2 (x_2^2 - x_1^2)}$$

M_{calc} = molecular weight calculated from the rule of simple multiples.

f/f_o = ratio of experimentally determined molar frictional constant $f = \frac{M(1 - V\rho^\circ)}{s_{20}}$ to molar frictional constant calculated for an unsolvated spherical particle of the same mass $f_o = 6\pi\eta^\circ N\left(\frac{3MV}{4\pi N}\right)^{1/3}$

$\frac{du}{dpH_o}$ = slope of mobility curve in the vicinity of the isoelectric point.

R = gas constant = 8.313×10^7 erg grad^{-1}.

N = Avogadro constant = 6.06×10^{23}.

T = absolute temperature.

ρ° = density of water at 20° C = 0.9982.

ρ = density of solution.

η° = viscosity of water at 20° C = 0.010.

c = concentration.

x = distance from axis of rotation.

$\frac{dx}{dt}$ = sedimentation velocity.

ω = angular velocity of rotating solution.

Name	V_{20}	s_{20}	D_{20}	M_s	M_e	M_{calc}	f/f_o	Iso-electric Point [1]	$\frac{du}{dpH}\times 10^5$	References
Cytochrome C	0.702	1.83	12.0	12 400	12 000		1.17			22
Erythrocruorin (Lampetra)	0.751	1.87	10.7	17 100	19 100		1.16	5.6	3.2	38, 25, 34, 20
Myoglobin	0.741	2.04	11.3	16 900	17 500		1.11	7.0	7	25, 34, 46
Lactalbumin	(0.751)	1.9	10.6	17 400		17 600	1.16	5.12	6.7	12, 25
Bacillus Phlei protein	0.748	1.8	10.2	17 000			1.22			28
Gliadin	0.722 [2]	2.1	6.72	27 500 [2]	27 000		1.60			14, 15
Hordein	(0.729)	2.0	6.5	27 500			1.64			27
Barley globulin α	0.72	2.49	8.4	26 000			1.31	5.00	2.7	26
Erythrocruorin (Arca)	(0.751)	3.46			33 500		(0.99)	5-6		41, 38, 20

Name	V_{20}	s_{20}	D_{20}	M_s	M_e	M_{calc}	f/f_0	Iso-electric Point [1]	$\frac{d\alpha}{dpH}\times 10^5$	References
Erythrocruorin (Chironomus)	(0.751)	2.0			31 500		1.63	5.40	3.6	41, 38, 20
Ovalbumin	0.749	3.55 [3]	7.8	44 000	40 500		1.16	4.55	10.4	15, 20, 48
Lactoglobulin	0.751	3.12	7.3	41 500	38 000		1.26	5.20	11.9	19, 25
Bence-Jones α	0.749	3.7 [4]			35 000		(0.96)	5.20	5.8	43, 48
Bence-Jones β	(0.749)	2.8	7.3	37 000			1.31	5.46	3.5	21, 25, 20
Human tuberculosis bacillus	0.70	3.3	8.2	32 000		$2\times 17\ 600 = 35\ 200$	1.25	4.5 *		28
Zein	(0.71)	1.9	4.0	40 000			2.4			52
Insulin	0.749	3.58	7.53	46 000			1.18			16
Pepsin	(0.750)	3.3	9.0	35 500	39 000		1.08			23
Peroxidase	0.69	3.85	6.85	44 100			1.36			47
Concanavalin B	0.73	3.5	7.4	42 000			1.25			32
Crotoxin	0.704	3.14	8.6	30 000	30 500		1.22			7
Hemoglobin (horse)	0.749	4.41	6.3	68 000	68 000 [5]		1.24	6.92	7.2	21, 50, 40, 20
Hemoglobin (man)	(0.749)	4.48	6.9	63 000			1.16	7.07	6.3	20, 15
Serumalbumin (horse)	0.748	4.46 [6]	6.1 [6]	70 000 [6]	68 000	$4\times 17\ 600 = 70\ 400$	1.27	4.80	9.1	11, 42, 20
Yellow enzyme	0.731	5.76	6.3	82 000	78 000		1.17	5.22	6.4	13
Concanavalin A	0.73	6.0	5.6	96 000			1.26			32
Canavalin	0.73	6.4	5.1	113 000		$6\times 17\ 600 = 105\ 600$	1.31			32
Serumglobulin horse [7]	0.745	7.1	4.05	167 000	150 000		1.44	$\alpha = 5.06$ * $\beta = 5.12$ * $\gamma = 6.0$ *		17, 25, 34, 4, 49
Serumglobulin (man) [8]	(0.745)	7.1	3.84	176 000			1.49	$\gamma = 6.4$ *		9, 29
Antipneumococcus serum globulin (rabbit)	(0.745)	6.5 [9]	3.9 [9]	158 000 [9]		$8\times 17\ 600 = 140\ 800$	1.52			9, 51
Antipneumococcus serum globulin (man)	(0.745)	7.4	3.60	195 000			1.53			9
Myogen A	0.735	7.86	4.78	150 000	136 000		1.26			6
Phycocyan (Ceramium, dissociation component)	(0.746)	6.2	4.58	130 000	146 000		1.38			3, 25, 34
Barley Globulin γ	0.72	8.30	4.4	166 000			1.35	5.70	3.2	26
Serumglobulin (Lampetra)	(0.745)	12	3.2	360 000			1.41			35
Catalase	0.73	11.3	4.1	250 000			1.25			30
Phycoerythrin (Ceramium)	0.746	12.0	4.00	290 000	290 000		1.21	4.25	14.2	3, 50, 39, 48

Molecular Constants of Proteins (cont'd)

Name	V_{20}	s_{20}	D_{20}	M_s	M_e	M_{calc}	f/f_0	Iso-electric Point[1]	$\frac{du}{dpH}\times 10^5$	References
Phycocyan (Ceramium, main component)	(0.746)	11.4	4.05	270 000	275 000	16×17 600= 282 000	1.22	4.85	10.2	3, 50, 33, 20
Edestin	0.744	12.8	3.93	310 000			1.21			45, 25, 34
Excelsin	0.743	13.3[4]	4.26	295 000			1.13			44, 25, 39
Amandin	0.746	12.5[4]	3.62	330 000	330 000		1.28			44, 24
Erythrocruorin (Daphnia)		16.3								38
Hemocyanin (Pandalus)	(0.740)	17.4			400 000		1.07			5
Hemocyanin (Palinurus)	(0.740)	16.4	3.4	450 000	450 000		1.23			5, 25, 39
Hemocyanin (Eledone, dissociation component)	(0.740)	10.6	2.16	460 000		24×17 600= 422 000	1.93			5, 25
Urease	0.73	18.6	3.46	480 000			1.19			31
Thyroglobulin (pig)	0.72	19.2	2.65	630 000	650 000		1.43	4.58	11	8
Hemocyanin (Nephrops)	(0.740)	24.5	2.79	820 000			1.23	4.64	13	5, 25, 34, 20
Hemocyanin (Homarus)	(0.740)	22.6	2.78	760 000	800 000		1.27	4.95	18	5, 25, 20
Antipneumococcus serum globulin (horse)	0.715	19.3	1.80	910 000		48×17 600= 845 000	1.86	4.4 *		9, 51
Antipneumococcus serum globulin (cow)	(0.715)	18.1	1.69	910 000			1.98	4.8 *		9, 10, 51
Antipneumococcus serum globulin (pig)	(0.715)	18.0	1.64	930 000			2.02	5.1 *		9, 10, 51
Hemocyanin (Helix pomatia, dissociation component)	(0.738)	19.7	1.77	1 030 000		64×17 600= 1 130 000	1.79			1
Hemocyanin (Paludina vivipara, dissociation component)	(0.738)	21.8	1.79	1 130 000			1.72			2
Erythrocruorin (Planorbis)	0.745	33.7	1.96	1 630 000	1 540 000	96×17 600= 1 690 000	1.30	4.77	10.6	38, 41, 15, 4, 18
Hemocyanin (Calocaris)	(0.740)	34			1 330 000		1.22			41, 5
Hemocyanin (Octopus)	0.740	49.3	1.65	2 800 000			1.38			36, 5, 24
Hemocyanin (Eledone)	(0.740)	49.1	1.64	2 800 000			1.39	4.6		5, 25, 39, 20

Erythrocruorin (Arenicola)...	(0.740)	57.4			3 000 000		1.25	4.56	16	37, 39, 18
Erythrocruorin (Lumbricus)..	0.740	60.9	1.81	3 150 000	2 950 000	192×17 600= 3 380 000	1.21	5.28	12.6	37, 25, 34, 39, 20
Chlorocruorin (Spirographis).		55.2								39
Hemocyanin (Rossia).....	(0.740)	56.2	1.58	3 300 000			1.36			25, 39
Hemocyanin (Helix pomatia, dissociation component)....	(0.738)	65.7	1.41	4 300 000		256×17 600= 4 510 000	1.40			1
Hemocyanin (Helix pomatia, main component).....	0.738	103.0	1.07	8 900 000	8 500 000	512×17 600= 9 010 000	1.45	5.05	8.1(7.3)	1, 18 (48)
Hemocyanin (Paludina vivipara, main component).....	(0.738)	102.5	1.09	8 700 000			1.43	4.71	10.8	2, 18

[1] Except when otherwise stated, the values refer to experiments carried out at 20° C in acetate or phosphate buffers with ionic strength = 0.02. The determinations marked with * were carried out under the same conditions, but at 0° C.

[2] The specific volume of gliadin varies strongly with the protein concentration according to Krejci and Svedberg (1935). For the calculation of M, Lamm and Polson (1936) used $V = 0.722$.

[3] Mean value obtained by a number of different observers in Upsala.

[4] Recalculated to water basis.

[5] Mean of two determinations on CO-hemoglobin and two on methemoglobin.

[6] Mean value for the A and B fractions prepared by Kekwick (1938).

[7] Total serumglobulin obtained by half-saturation with ammonium sulfate.

[8] Electrophoretically prepared γ-globulin from normal human serum.

[9] Mean of four determinations on four different specimens.

1. Brohult, S., *Nova Acta Regiae Soc. Sci. Upsaliensis, IV*, **12**, Nr. 4 (1940).
2. Brohult, S., and Ekwall, P., (unpublished).
3. Eriksson-Quensel, I.-B., *Biochem. J.*, **32**, 585 (1938).
4. Eriksson-Quensel, I.-B., Cf. Svedberg and Pedersen, "The Ultracentrifuge."
5. Eriksson-Quensel, I.-B., and Svedberg, T., *Biol. Bull.*, **71**, 498 (1936).
6. Gralén, N., *Biochem. J.*, **33**, 1342 (1939).
7. Gralén, N., and Svedberg, T., *Biochem. J.*, **32**, 1375 (1938).
8. Heidelberger, M., and Pedersen, K. O., *J. Gen. Physiol.*, **19**, 95 (1935).
9. Kabat, E. A., *J. Exp. Med.*, **69**, 103 (1939).
10. Kabat, E. A., and Pedersen, K. O., *Science*, **87**, 372 (1938).
11. Kekwick, R. A., *Biochem. J.*, **32**, 552 (1938).
12. Kekwick, R. A., Cf. Svedberg and Pedersen, "The Ultracentrifuge," 1940.
13. Kekwick, R. A., and Pedersen, K. O., *Biochem. J.*, **30**, 2201 (1936).
14. Krejci, L., and Svedberg, T., *J. Am. Chem. Soc.*, **57**, 946 (1935).
15. Lamm, O., and Polson, A., *Biochem. J.*, **30**, 528 (1936).
16. Miller, G. L., and Andersson, K. J. I., *J. Biol. Chem.*, **144**, 459 (1942).
17. Mutzenbecker, P. von, *Biochem. Z.*, **266**, 250 (1933).
18. Pedersen, K. O., *Kolloid-Z.*, **63**, 268 (1933).
19. Pedersen, K. O., *Biochem. J.*, **30**, 961 (1936).
20. Pedersen, K. O., Cf. Svedberg and Pedersen, "The Ultracentrifuge," 1940.
21. Pedersen, K. O., and Andersson, K. J. I., Cf. Svedberg and Pedersen, "The Ultracentrifuge," Oxford, 1940.
22. Pedersen, K. O., and Andersson, K. J. I., (unpublished).
23. Philpot, J. St. L., *Biochem. J.*, **29**, 2458 (1935); Philpot, J. St. L., and Eriksson-Quensel, I.-B., *Nature*, **132**, 932 (1933) and Eriksson-Quensel, I.-B., (unpublished); cf. Svedberg and Pedersen, "The Ultracentrifuge."
24. Polson, A., *Nature*, **137**, 740 (1936).
25. Polson, A., *Kolloid-Z.*, **87**, 149 (1939).
26. Quensel, O., "Inagural Dissertation," Upsala, 1942.
27. Quensel, O., and Svedberg, T., *Compt. Rend. Lab. Carlsberg*, **22**, 441 (1938).
28. Seibert, F. B., Pedersen, K. O., and Tiselius, A., *J. Exp. Med.*, **68**, 413 (1938).
29. Stenhagen, E., *Biochem. J.*, **32**, 714 (1938).
30. Sumner, J. B., and Gralén, N., *J. Biol. Chem.*, **125**, 33 (1938).
31. Sumner, J. B., Gralén, N., and Eriksson-Quensel, I.-B., *J. Biol. Chem.*, **125**, 37 (1938).
32. Sumner, J. B., Gralén, N., and Eriksson-Quensel, I.-B., *Ibid.*, 45 (1938).
33. Svedberg, T., *Chem. Rev.*, **20**, 81 (1937).
34. Svedberg, T., *Nature*, **139**, 1051 (1937).
35. Svedberg, T., and Andersson, K., *Nature*, **142**, 147 (1938) and unpublished.
36. Svedberg, T., and Eriksson, I.-B., *J. Am. Chem. Soc.*, **54**, 4730 (1932).
37. Svedberg, T., and Eriksson, I.-B., *J. Am. Chem. Soc.*, **55**, 2834 (1933).
38. Svedberg, T., and Eriksson-Quensel, I.-B., *J. Am. Chem. Soc.*, **56**, 1700 (1934).
39. Svedberg, T., and Eriksson-Quensel, I.-B., *Tabulae Biologicae*, **XI**, 351 (1935/36).
40. Svedberg, T., and Fåhraeus, R., *J. Am. Chem. Soc.*, **48**, 430 (1926).
41. Svedberg, T., and Hedenius, A., *Biol. Bull.*, **66**, 191 (1934).
42. Svedberg, T., and Sjögren, B., *J. Am. Chem. Soc.*, **50**, 3318 (1928).
43. Svedberg, T., and Sjögren, B., *J. Am. Chem. Soc.*, **51**, 3594 (1929).
44. Svedberg, T., and Sjögren, B., *J. Am. Chem. Soc.*, **52**, 279 (1930).
45. Svedberg, T., and Stamm, A. J., *J. Am. Chem. Soc.*, **51**, 2170 (1929).
46. Theorell, H., *Biochem. Z.*, **252**, 1 (1932).
47. Theorell, H., *Arkiv Kemi, Mineral. och Geol.*, **15B**, No. 24 (1942).
48. Tiselius, A., *Nova Acta Regiae Soc. Sci. Upsaliensis, IV*, **7**, No. 4 (1930).
49. Tiselius, A., *Biochem. J.*, **31**, 1464 (1937).
50. Tiselius, A., and Gross, D., *Kolloid-Z.*, **66**, 11 (1934).
51. Tiselius, A., and Kabat, E. A., *J. Exp. Med.*, **69**, 119 (1939).
52. Watson, C. C., Arrhenius, S., and Williams, J. W., *Nature*, **137**, 322 (1936) and unpublished.

similar relationship. This remarkable regularity points to a common plan for building up the protein molecules. Certain amino acids may be exchanged for others, and this may cause slight deviations from the rule of simple multiples, but on the whole, only a very limited number of masses seems to be possible. Probably the protein molecule is built up by successive aggregation of definite units, but only a few aggregates are stable. The higher the molecular weight the fewer are the possibilities of stable aggregation. The steps between the existing molecules, therefore, become larger and larger as the weight increases. These statements are borne out by Table 1, in which are collected recent data for the various constants of protein molecules as determined in Upsala."

Modification of Biocatalysts

The writer has taken the view that biocatalysts are subject to *modification,* which involves the fixation, at an active catalyst area of a gene or other catalyst unit, of some particle (electron, ion, atom, molecule, or colloidal particle) which changes the nature and/or the rate of the catalytic change occurring there.[8] A *limiting case* under this general principle would be the *formation of a new catalyst area* by the fixation, *e.g.*, at a protein surface, of a prosthetic group. In the case of genes a change in catalytic output may arise from an intraparticulate change * in the gene itself (known as a *point mutation*), or because of some chromosomal upset (*e.g.*, inversion, translocation, crossing over) which places the gene in different surroundings (*position effect*). If heritable and non-lethal, both of these types of change generally lead to new mutant forms of plants or animals. The modification of enzymes within a cell, or the formation or introduction of new enzymes there, could lead to catalytic variations of chemical output which might simulate the effects producible by a point or a chromosomal mutation; but the result would be transient unless sufficient of the modifiers and prosthetic groups to maintain it were supplied or produced. Another possibility is that a specific area may serve as a template or mold against which may be formed specific catalytic molecular structures or plaques. Still another possibility is the local establishment of ionic or trace-substance conditions favorable to the formation of new enzymes, *e.g.*, by adsorptive fixation of a prosthetic group by a carrier. Viruses and bacteriophages reproduce themselves in cells, thus (where they exert catalytic action) behaving *as if* they were free-living genes. What is said of genes may be applied with equal force to mitochondria; for whether these are symbionts or cytoplasmic inclusions, they apparently reproduce and are efficient in directing chemical changes within the cell (either directly or through enzyme formation), and behave somewhat as if they are free-living genes or gene groups.

The stability of genes and the efficiency of the protections surrounding them are evidenced by the regular and orderly sequences normal to life, which show that gene mutations are relatively quite rare, and abnormal gene or catalyst modifications unusual. Since non-lethal gene mutations may be transmitted by heredity, they are basic factors in evolution; for beneficial mutations tend to survive and dominate, whereas harmful ones tend to die out. If mutation makes a gene (or other catalyst) more susceptible to an abnormal modification, that is one way in which the effect of a mutation may become evident.

Bacterial Dissociation and Transformation[98b]

The colonies of most bacteria that have been studied may appear in a rough, corrugated form (R), or in a smooth, glistening, drop-like form (S). The change of a culture from R to S, or *vice versa,* is known as *dissociation.*[62] The methods used to produce dissociation involve, mainly, change in the usual growth medium: (1) by the addition of sera, normal or immune; (2) by addition of definite chemical agents

* With or without particulate additions.

(*e.g.*, $LiCl_2$, $FeCl_3$, sodium taurocholate); (3) by change in pH; (4) by animal passage, a method much used, involving considerable, though unknown, changes in the culture medium. For example, Alexander-Jackson[14] was able to change two strains of human tubercle bacilli of R form over into S forms by the addition of 0.0004 per cent of ferric chloride to Bordet-Gengou medium. As a rule, the S form is relatively more virulent than the R form of the same strain. At the laboratory of the New York City Board of Health the virulence of test strains of pneumococci is maintained by daily passage through mice, to give an example of the well-known effect of animal passage, which also exerts changes in the virulence of viruses; *e.g.*, the virulence of smallpox to man is reduced by passage through cattle.

On the other hand Herald R. Cox[86a] found that the virulence of a *Dermacentor variabilis* strain of Rocky Mountain spotted fever was enhanced, as against guinea pigs, upon growth in the yolk sacs of developing chick embryos. The virulence decreased after about 50 passages, and after about 100 passages the strain, while causing slight or no reaction in the guinea pigs, made them "solidly immune to massive doses of highly virulent strains." See, also, R. E. Green[60a] on the nature of virus adaptations.

Griffith,[61] working with white mice, discovered the interesting fact that S forms of pneumococci can be transformed from one specific type * into other specific types through the intermediate stage of the R form. Since the specificity of the types involves the formation of specific substances, (especially, as Heidelberger[69] has shown, specific carbohydrates), it seems obvious that the catalysts of the pneumococci in Griffith's experiments must have undergone a change; either old catalysts were transformed, or new ones were formed.

Dawson and Sia,[41] working *in vitro*, found that R forms of pneumococci derived from S forms of one specific type (Type II), can be transformed into the S forms of other specific types (Types I or III), by growth of small inocula of the R form of Type II in media containing vaccines prepared from heterologous S cultures. It seems possible that in this case the R form contains an enzymic *carrier* which has been deprived of its prosthetic group, but which takes up *a different prosthetic group* from the new vaccine-containing medium. That these R forms are more labile is brought out by Sia and Dawson,[119] who state: "R cultures possessing only slight degree of R stability are most suitable for transformation purposes by *in vitro* procedures."

In cultures of *Hemophilus influenzae, Escherichia coli,* and *Streptobacillus moniliformis*, the germination of "large bodies" has been observed, and the descendants of these "L type" colonies, instead of resembling the parent organism, are similar in both morphology and development to the pleuropneumonia group of organisms. Such "L type" colonies come from strains which are rare in most species, but they were also observed in cultures of a *Flavobacterium,* of *Bacteroides funduliformis,* and of the gonococcus, and have been isolated in pure cultures from the two former. L. Dienes[44a] believes that they represent a variant type, in which case the variability of bacteria extends much further than is commonly supposed. E. Kleinberger (references in Dienes' paper) thought that they represented symbionts, a view that Dienes opposes. The question is not finally decided.

Speaking of the enzymatic synthesis of polysaccharides, M. Stacey † states: "There is good evidence for the theory, here advanced for the first time, that the synthesizing enzyme remains *in combination with the polysaccharide* it has synthesized, and there is progressively built up a complex macromolecule in which com-

* At present, 50 specific types of pneumococci (40 numbered types and 10 sub-types) have been identified. Besides these, other bacteria may produce pneumonias; *e.g.*, Friedlaender's bacillus, streptococci, and the influenza bacillus of Pfeiffer. (J. G. M. Bullowa).

† "Mucopolysaccharides and Related Substances," *Chemistry and Industry*, **62**, 110-112 (1943).

paratively short polyglucose chains are 'cemented' together by a nucleoprotein. In the case of one dextran from *Betabacterium vermiforme* (Ward-Meyer), the aggregating process can go on until the macromolecule is so large that it settles out of the solution in the form of granules. The separation of granular synthetic starch takes place in precisely the same manner and the product contains significant amounts of a nitrogenous constituent which could only come from the phosphorylase preparation."

Evidence that bacterial catalysts direct the formation of specific polysaccharides appears in the statement of E. J. Hehre [68a] and J. Y. Sugg that a serologically reactive polysaccharide of dextran nature was produced from sucrose by sterile filtered extracts derived from sucrose broth cultures of *Leuconostoc mesenteroides,* a Gram-positive coccus widely distributed on plants. "Rigorous controls were included to prove that this reaction occurred in the absence of microörganisms." The polysaccharide was recognized chemically and also by its ability to react with the antiserums of types 2 and 20 pneumococci, as well as with the antiserum of the homologous bacteria.

These and many other facts support the view that the specificity of all species, both plant and animal, stem from, or at least involve, differences in the basic biocatalysts. The visible forms with which taxonomy deals, and even the specific substances which biochemists isolate and identify, are mainly an aftermath. The basis of species specificity and even of evolution rests largely on biocatalysts and their changes.

Immunology

Immunological specificity, like all other kinds of biological specificity, depends on the outwardly directed specific electronic fields of the units involved in agglutination, precipitation, lysis, anaphylaxis, etc. With egg albumin the minimum sensitizing dose is about 0.000,05 milligram; and in general, the tiny quantities of antigen demonstrable by immunological methods cannot be detected specifically by other known methods. How shall we account for these potent and highly specific effects of incredibly minute quantities, and also for the fact that, when an immunized animal is bled, the temporary drop in the titer of antibodies in the remaining blood is soon retrieved and often surpassed? The reactivity of a sensitized animal or even of its tissues is not impaired by replacement of its blood by the blood of a non-sensitized animal.

The experimental facts point to *the formation, within the cells or tissues, of new specific catalyst surfaces which are able to direct the formation of antibodies.*[8, 12] Marrack [94] states that the properties of antibodies indicate that they appear to consist of modified serum globulin; and that on immunization there may be an increase in serum protein, mainly in the globulin fraction. While nature may utilize any conceivable or unthought of method, it seems most likely that the antigen particles serve as prosthetic groups in the formation of biocatalysts which can determine the formation of units having *essentially the reverse electrostatic charge pattern of the antigen* [8, 12] in their exposed surfaces, when they are detached or removed, perhaps by the dominant attraction of a carrier, or by action of a "detergent" substance in the milieu, or by reaching a critical thickness.[58]

The specificity and lasting effects of minute quantities of antigens become comprehensible on the basis of this view, for in theory one single molecule would suffice to form or convert a biocatalyst particle into a producer of specific antibody. In fact, Kuhn, Moewus, and Jerchel [87] report that one single crocin molecule is capable of inducing a sexual change * in a whole alga. There is no reason why many different antigens may not thus simultaneously or successively produce their own

* Oysters change sex as they grow older. Sex change in plants is also known,[111a] *e.g.,* in the Florida silver palm.

specific catalysts and antibodies, even in the presence of numerous enzymes and genes; and this checks with the experimental facts. The specific "mold," once formed, could function continuously, despite bleeding, unless irreversibly injured. Variations in the duration of immunity would follow variations in the persistence in functioning of the new "antigen catalyst," while inability to establish immunity could be due to destruction of antibody, or to inability to form a biocatalyst for its production, *e.g.*, because of lack of a suitable carrier substance. All these phenomena appear in "vaccination," a general term indicating introduction of antigen with the hope that immunity will result.

In discussing the nature of the intermolecular forces operative in biological processes, L. Pauling and M. Delbrück state:[106] "It is our opinion that the processes of synthesis and folding of highly complex molecules in the living cell involve, in addition to covalent-bond formation, only the molecular interactions of van der Waals attraction and repulsion, electrostatic interactions, hydrogen-bond formation, etc., which are now rather well understood. These interactions are such as to give stability to a system of two molecules with complementary structures in juxtaposition, rather than to two molecules with necessarily identical structures; we accordingly feel that complementariness should be given primary consideration in the discussion of the specific attraction between molecules and the enzymic synthesis of molecules." Pauling has worked out in further detail a theory of the structure and process of formation of antibodies, and has discussed the mechanism of their union with antigens.[107]

Later, Pauling and Campbell described the formation of antibodies in vitro.[107a] "The procedure consists in subjecting normal globulin or other protein to the action of denaturing reagents or conditions in the presence of the antigen. The protein molecule unfolds, and then refolds in such a way as to assume a configuration complementary to that of the antigen, thus acquiring the properties of a specific homologous antibody."

It is not necessary to think of antibodies as simple molecular units, or to imagine that the *whole* of the antibody area is specific to the antigen. Thus, as Landsteiner has shown, atoxyl became a "determinant group" in each of the antibodies synthesized by combining it with various proteins by a diazo reaction. The protein serves as a carrier, and the atoxyl (though not of itself an antigen) becomes a determinant group or *hapten,* which, in some respects, exercises a determinative function analogous to that of the prosthetic group in an enzyme.[135]

A hapten impresses upon its antibody a specificity which leads to the combination of the antibody with its specific antigen; but in some cases closely allied determinant groups are immunologically equivalent. The antibody, and perhaps also the antigen after it enters the animal, may be considered as a *plaque* having surface specificities, wherein the subunits may be held together by forces less potent than those of primary valency.

The functioning of the antigen "mold" or "template"[95a] may be crudely illustrated, at a much higher structural level, by pressing a piece of tinfoil against the surface of a coin.* The surface of the foil *next* to the coin acquires an impression which is the specific *opposite* or reverse intaglio of the coin pattern, while the *upper* surface of the foil acquires a surface impression which *duplicates* the coin pattern. If something analogous occurs when a molecular plaque is formed against an antigen mold, the subsequent influence of the detached plaque would depend upon *which of the plaque surfaces remains exposed to the milieu,* if the plaque serves as adsorbent or is itself adsorbed. There could thus be formed a new surface like the mold, or a surface with reverse contour, or modifications of either of these surfaces if the

* The writer has used this as a lecture illustration for over ten years. See J. Alexander, "Colloid Chemistry," 4th ed., p. 385, D. Van Nostrand Co., 1937.

plaque is distorted on adsorption or is subject to enzymic or other biochemical attack.

The separation of duplicated chromosomes during mitosis shows that forces exist which separate the duplicated gene-strings from each other. As N. K. Koltzoff [85] and C. B. Bridges [80] independently showed, the huge salivary gland chromosomes in the small grub that develops into *Drosophila,* which are enlarged or swollen to about 200 times normal size, appear to consist of a number, possibly 16 or more, gene strings which, instead of separating, remain coherent and, when stained, show specific bands and structures at the loci of specific genes. On comparing the locations of these bands with the gene maps developed from the data of geneticists, their matching indicates that we have before us what has been termed a "genetic spectrum." [12]

What forces normally separate each template gene from the new gene formed against it, so that the new chromosome may separate lengthwise from its originator? And what forces would determine the separation of our hypothetical plaque from the surface against which it was formed? While no simple or positive answer can be given, it must be recalled that small changes in ionic concentration (pH with protein units, C_{O_2} in the developing zygote) could be effective, as may also be the presence of small amounts of specific substances which act as "detergents," as Svedberg found with proteins and as Smith found with natural chlorophyll. Another possibility is suggested by the experiments of Goranson and Zisman,[58] who found that when about 500 successive X-multilayers of calcium (or barium) stearate were deposited upon an ebonite "probe," the polymolecular layer spontaneously detaches itself.* Possibly the cohesive surface forces diminish as the layer becomes thicker, and are no longer effective when the deposit reaches a critical thickness. It is not unreasonable to envisage the possibility that specific natural proteins, carbohydrates etc., may be thus formed at the surface of specific catalysts as templates, and float away to become effective units elsewhere. A contrary effect seems to be produced by colchicine, which causes chromosome doubling or polyploidy, apparently by interfering with spindle formation during mitosis.†

On the other hand C. C. Lindegren and C. B. Bridges [89a] advanced the hypothesis that the surface of each chromomere (not necessarily the gene surface), may stimulate the protoplasm to form specific antibodies, which on being specifically adsorbed at the chromomere interface, renders it capable of adhering specifically to its partner chromomere in synapsis. Any two allelic chromomeres happening to touch "would be cemented together by the antibody junctions specific to themselves. The chromomeres which are on each side of the already agglutinated ones would then be more likely to touch and fuse, so that synapsis would proceed, zipper-like, from the first points of homologous contact throughout the entire lengths of the chromosomes." (See below, Topley, Wilson and Duncan).

Differentiation and the Orderly Course of Life

The self-duplication of a particle of molecular or near-molecular dimensions, *e.g.,* a virus, a gene, or a hypothetical moleculobiont, can be understood as an assemblage of subunits ordered by the "progenitor" particle into its own precise structure. However, when we consider the development of a plant or an animal from a zygote,

* The authors say: "In plating X-multilayers it has been observed that after about 500 layers have been deposited, and the electrostatic repulsive field of the multilayer has thus reached a certain value, the upper portion of the submerged probe has a silvery appearance which gradually moves down the probe with increasing number of dips, and film does not adhere to the multilayer over this portion. This electrostatic repulsive field thus sets a limit to the thickness of X-multilayers on insulators." Dr. Goranson informs me (private communication) that "if plating is made on a metal, more layers can be put on because of the oppositely induced charge on the metal."

† See paper by W. D. Harkins in this volume.

a single fertilized cell resulting from the union of a male with a female gamete, we are confronted with the astounding fact that there emerge, in regular order, a variety of cells, tissues, organs, and bodily structures, generally possessed of highly specialized functions, and thus *differentiated* from the original zygote cell.* Under normal conditions, the developing zygote follows, with surprising precision, the devious transformations leading to the completed organism, the possibility of whose formation was foreordained subject, of course, to the presence of a suitable milieu, by the genes and cytoplasmic inclusions entering into the zygote from the parents.

It seems evident that the definitely directed course of differentiation cannot be accounted for by assuming that there occurs a suitable series of gene point mutations, each appearing at the proper time; for point mutation is sporadic and more or less indefinite. The view has been advanced,[8] and is here maintained, that *the formation and/or the modification of definite biocatalysts,* at definite times or stages, by definite chemical substances mainly included in the zygote cytoplasm, or subsequently formed catalytically, will reasonably account for these orderly happenings. While many other factors are operative, and intercurrent substances (or lack of them) in the milieu are important, no other basic phenomenon seems to offer so satisfactory an explanation of differentiation as does catalyst formation and/or modification.

In attempting to outline certain physico-chemical principles which are operative in the clone † of cells developing from the original zygote, it must be emphasized that not only is each system in itself highly complex, but also that very wide differences appear in the details and order of development with the numerous varieties, species, and genera of plants and animals. With many forms it appears that in the zygote itself an initial segregation of modified or of modifying substances occurs. The slight differences often found in monozygotic twins [93a] do not seem to be due to the precisely duplicated genes, but rather to inequalities in the apportionment of the zygote cytoplasm upon mitosis, involving unequal distribution of catalyst modifiers between the cells, which later develop into separate individuals.

As the clone of cells arising from the zygote increases in size, certain physico-chemical situations must naturally develop in the rounded mass of cells termed the blastula. Whereas with a single cell diffusion of materials to and from the cell is relatively rapid and equalized, with a mass of cells a *diffusion gradient* tends to appear, cells and even portions of cells at the exterior having a more favored position relative to cells on the interior. Ions, molecules, and even colloidal particles differ greatly in diffusibility, and as a consequence of differential diffusion [9] there appear zones of variation in ionic and molecular concentration which can exert profound influence on such factors as the formation and functioning of biocatalysts, the adsorbability of prosthetic groups and modifiers, the structure of proteins, and the aggregation of colloids. The intense particulate activity revealed in cells by the ultramicroscope may in part be due to diffusion at catalyst surfaces, the convection being intensified in interior cells if a thermal gradient exists there.

Kenneth B. Raper ["Third Growth Symposium," 41-76 (1941)] describes the remarkable behavior of simple slime molds, which live and multiply by fission when

* As Prof. R. G. Harrison points out this ever-recurring wonder is thus expressed in the 139th Psalm:

"14. I will praise Thee; for I am fearfully and wonderfully made: marvelous are Thy works; and that my soul knoweth right well.

15. My substance was not hid from Thee, when I was made in secret, and curiously wrought in the lowest parts of the earth.

16. Thy eyes did see my substance, yet being unperfect; and in Thy book all my members were written, which in continuance were fashioned, when as yet there was none of them."

† The term *clone* (sometimes spelled *clon*), indicates a cluster of cells descended from a common progenitor. It is derived from the Greek word *klon*, meaning *a twig*, which is apparently the root word for the Gaelic *clann*.

plenty of food is available. Following this vegetative stage, masses of cells, which may number as high as 150,000, gather together to form a pseudoplasmodium which migrates as a unit but stops, in *Dictyostelium discoideum,* just preceding formation of the sorocarp (spore-bearing stalk). The constituent myxamoebae collect into a rounded apiculate mass whose basal part is first differentiated near the center of the mass. "Somewhat above the level of the substratum a hyaline tube is developed and within this the cells become vacuolated and compacted together. It is not known whether this membrane, or sorophore sheath, is formed by the myxamoebae contained within it or by those surrounding it, for the myxamoebae on either side appear equally undifferentiated. In either case, however, it is the product of multiple cells and represents a function of the whole mass, for its diameter is regularly proportional to the size of the entire body. As the myxamoebae within it begin to differentiate, the sheath is extended upward and downward. . . . The upward elongation of the sorophore continues by extension of the slime sheath and by the differentiation within it of myxamoebae into stalk cells. In the area of sorophore formation progressive stages in the progressive differentiation and vacuolation of myxamoebae into parenchyma-like stalk cells occur from the apex downward. As the sorophore elongates the mass of myxamoebae ascends or is carried up by the developing sorophore . . . the myxamoebae at its periphery and in contact with the air begin to differentiate into spores. Spore maturation is progressive from the periphery inward. The two processes of cellular differentiation now proceed simultaneously and the formation of the sorocarp is completed by the transformation of all of the myxamoebae either into supportive cells, forming the sorophore, or into spores, forming the sorus."

Raper points out that after aggregation, there is no increase in the size or number of myxamoebae, but all further development consists entirely of the integration and subsequent differentiation of myxamoebae already present, which may be of wholly different spore origin and of initially equal potentialities. If myxamoebae are dissociated in the presence of bacteria, they return to the vegetative self-reproducing stage, though if no bacterial food is present, they re-aggregate and develop fruiting structures, whose pattern is constant, though the size is in proportion to the mass. Species thoroughly intermixed in the vegetative stage aggregate to separate and distinct centers; and though two species of Dyctiostelium may initially enter the same fruiting organization, they later draw apart and form separate sorocarps, which are of a characteristic form for each species. "What is inherited is not a specific type of structure, but the ability of like but discrete and independent units to cooperate in the formation of such a structure."

The notion of catalyst modification, following upon differential diffusion which must develop when the pseudoplasmodium forms, could readily account for most of the observed facts. The segregation of cells of like species is observed with many organisms, and is dealt with in the paper of H. S. Jennings in this volume.

According to Topley, Wilson and Duncan [185a] when a heterogeneous mixture of bacteria is agglutinated by a heterogeneous mixture of specific antisera, each cluster of bacteria is homogeneous. Apparently each kind of bacterium becomes coated by a layer of its own specific antibody, and the bacteria are so specifically conditioned that each kind forms a lattice or clump of identical unions through the adsorbed antibodies. A. Lazarow [88a] isolated from finely dispersed liver by fractional ultracentrifugation, a submicroscopic particle containing glycogen. The particle has a sedimentation constant of *ca* $4{,}000 \times 10^{-13}$, and appears to be an aggregate of smaller glycogen units held together by about 1 per cent. of protein, a coacervating agent which "seems to parallel the action of insulin, because insulin is known to lower blood sugar and facilitate glycogen storage in the liver." These two cases seem to be instances of *cohesive colloids,* mentioned in the first paper in this volume.

The question why the cells of the blastula tend to cohere was answered for some

cases by C. Herbst,[71] who found that, while all of the salts in sea-water are important for the development of echinid eggs, if calcium was absent the cells of the developing embryo would fall apart, though they could live and develop further. An analogous role is played by calcium in soils, where calcium "humate" binds soil particles into the crumby condition consonant with what agriculturists term good tilth. Alkali soils, besides being pasty and "sticky," tend to wash away. It may be that local, perhaps temporary, deficiency of calcium may be a factor in polyembryony, which seems to have occurred in the case of the Dionne quintuplets, and is common in some insects, especially wasps,[181] where hundreds of individuals may thus arise from a single egg. Polyembryony in insects was first discovered by Paul Marchal (1898), and Prof. J. T. Patterson informs me that probably Marchal was the first to suggest that polyembryony in insects might be explained on the basis of Herbst's experiments with Ca-free sea water to separate the cells of sea-urchins. With the Texas armadillo which typically produces four identical quadruplets from a single egg, and the South American Mulita (*Dasypus hybridus*) in which one egg gives rise to from seven to twelve individuals of the same sex, some other explanation must be found, because the primordia of the embryos arise rather late in development. (See "Polyembryony in Animals," by J. T. Patterson, in The Quarterly Review of Biology, *2*, 399-426 (1927)).

Because of their favored location relative to food supply and waste elimination, exterior cells tend to grow and to duplicate more rapidly than cells within the blastula. If the cells stick together, this demands an increase in the external cell surface of the blastula mass, and this can happen only in two ways: (1) by formation of a *diverticulum,* whereby the "skin" cells form a root-like projection; (2) by *invagination,* whereby the surface of the blastula bends inward to form a pocket or sheath. In animals the "skin" cells cohere so strongly to those within, that invagination occurs, leading to the formation of the gastrula. Thus, a hollow rubber ball immersed in gasoline superficially absorbs this rubber solvent, and the swelling surface tends to force an invagination at the weakest area.

With the formation of the gastrula a new set of physico-chemical conditions arises. The gastrula cavity may be regarded as a "bay" or inlet from the "ocean" of the milieu; and this bay will have, or will immediately develop, ionic and molecular concentrations different from those of the exterior "ocean" of the milieu. Cells facing the "bay," though originally "skin" cells, may be expected to react to their new situation by the adsorption and/or the elution of prosthetic groups or of specific catalyst modifiers; and modified catalysts mean *a change in chemical output of the cells,* which we finally recognize as *differentiation.* Furthermore, the interior cells of the gastrula and those in an intermediate position are differently situated relative to the exterior cells. There thus tend to develop three typical environmental zones, corresponding to the *ectoderm, mesoderm,* and *endoderm.* Just what may happen in and within these zones as a result of catalyst changes and/or catalyst formation enforced by ionic or concentration changes, or because of alteration in conditions governing catalysis, will vary in each case and will become increasingly more complicated as the growing structures evolve.

The viewpoint of biologists may be seen from the following quotations. H. S. Jennings states:[78] "At a certain time the egg of such a creature as a frog has become a mass of small cells. Under normal conditions, when we examine this mass, we can predict what part of the adult each part will produce. Cells here will produce the brain, there the sides of the eyes, here the ear, there the spinal chord, here parts of the skin. If we leave the egg to itself, these are indeed the parts that will be produced.

"But this is not because each cell can produce only that part and nothing else. . . . What happens is that from a certain spot on the egg—a recognizable spot—an organizing influence starts out, so that this spot is known as the organizer, or the

organization center. This organizing influence, whatever its nature, creeps from cell to cell, causing each cell to alter internally—through the interaction of its genes and cytoplasm—in such a way as to produce the structures of the embryo. Each cell that is reached later transforms in such a way as to fit the cells that have gone before—in such a way as to make the next proper part in the pattern of the body."

Ross G. Harrison states:[84] " . . . the internal environment of a cell (is) the sum of all those factors with which it is in relation that lie within the organism but outside of the cell in question. Strictly speaking, a single cell has only an external environment. However, the germ cells that unite to form the fertilized egg do have in a certain sense an internal environment inside the body of the parent organism. When they are cast loose they cease for a time to have such an environment, and it is only after cleavage begins that the resulting cells acquire a new internal environment in their sister cells. For a short period this is the only internal factor that counts, there being as yet no general *milieu intérieur*. However, during this period, which in the vertebrate embryo includes segmentation, gastrulation, and neurulation, the most important differentiations occur. Later, after the establishment of a circulating medium and the primitive nerve paths, an internal environment of the ordinary sort is again present. . . .

"Since cellular differentiation takes place in the cytoplasm, we are concerned here more directly with this constituent of the cell, bearing in mind, however, that the cytoplasm is accompanied and ultimately controlled by the genic complex in the chromosomes. Since the latter is presumably the same * for all cells of the organism, differences between cells must arise through interaction between the constant genom and the locally variable cytoplasm, in which they ultimately become visible. . . .

"In the so-called mosaic eggs, *i.e.*, those with 'precocious segregation,' to use Lankester's[88] term, local differentiation arises before cleavage and a minimum of modification in response to internal environmental factors is involved in subsequent development. In other eggs the protoplasm long remains labile with respect to its prospective differentiation, and in these many organs of the embryo develop only through the interaction between the cells producing them and other elements in the internal environment, *i.e.*, other cells and the internal medium. In such eggs local predetermination is reduced to a minimum. Their members in continuance † *are fashioned,* when as yet there is none of them.

"The most striking case of dependent differentiation is that of the central nervous system, which is formed in the outer layer of the embryo by reaction with the material underlying it, which is turned in during gastrulation. This material, because of its marked general effect upon the development of the embryo, has been termed the organizer by Spemann,[124] its discoverer. The material which is invaginated and which acts upon the overlying tissues forms the pharyngeal roof, head mesenchyme, axial musculature and notochord. . . .

"What it is that gives the organizer cells these properties, which are so effective as an internal environment to the overlying cells, and what the nature of the reaction in the surrounding tissues is, has not been ascertained. Spemann[125] himself experimented with organizer material the structure of which had been destroyed. Later Holtfreter[75] discovered that a great variety of animal tissues, both adult and embryonic, and from almost any phylum could, when placed in a blastocoel so that it ultimately came to lie under the ectoderm, stimulate the formation of a thickened area (neural plate) in the overlying tissue, which might fold over and close in as the neural tube does. Even pure chemical substances held in an agar matrix may

* We must here envisage the possibility that genes may be modified by specific substances, and that this modification may be persistently recurrent or even heritable. In this latter case the modification would be tantamount to a chemically induced mutation. J. A.

† Psalm 139.

have a similar effect. There is a difference, however, between the response to living normal organizing material and that to dead material, the former stimulating the formation of a complete neural plate and tube with normal regional differences, while the latter produces simply a thickened plate which may close to a tube without, however, showing such regional differentiation.

"Much effort has been spent to find out the nature of the chemical substance that produces these effects, and several different conclusions have been tentatively reached. The fact that certain embryonic tissues which have no organizing effect when living do so act when dead has led to the view that the substance responsible for the effect is free only in the active regions, but exists in a bound condition in other parts of the embryo. Needham, Waddington and co-workers [141, 142, 25] have isolated sterol derivatives from the organizer and have found that synthetic sterols when applied in an agar matrix do induce the formation of neural-platelike structures. Another fact that may be of significance is the occurrence of an intense carbohydrate metabolism in the organizer.[28, 68, 147]

"Differences between the action of the natural organizer in its normal position and that of dead material and chemical substances may be due to the circumstance that the former is distributed in a gradient having different local concentrations. Another factor that may have some influence is the possible variation of the intensity of action with respect to time. A normal induction might be achieved only through the action of a living system, because this might require the action to start with a low intensity, increase to a maximum, and then taper off; or it might require several cycles of variation in intensity, which dead material could not afford. Regional differences in the reacting material must also be taken into consideration, as shown by the different respective reactions of presumptive head and trunk regions to similar implanted organizer material." [125]

"What happens in the ectodermal cells that become transformed into neural plate is of much importance in any theory of organizer action, but it is almost entirely unknown. The neural plate contains relatively more water than the embryo as a whole [57] and it seems to be the inner portions of the neural plate cells that are the most swollen. We might assume, then, that the organizer produces a chemical change, which leads first to an orientation of particles that makes the cells columnar, followed by a greater hydration of the protein lattice localized in the basal halves of the cells. This would lead to the folding of the plate to form a groove and its ultimate closing into a tube. The changes underlying the process would thus be of a chemical and paracrystalline nature.

"It must also be borne in mind that the agencies effecting the differential changes in the central nervous system are heterogeneous in composition. Underlying the fore brain is the pharyngeal roof and some mesenchyme, and under the rest of the brain and spinal chord the notochord occupies the midline with the myotomes on each side. All of these factors are known to have a specific effect on the overlying neural plate.[89]

"Since the word 'organizer' connotes a master regulator which creates the organization, and since there are in the course of development many actions of the same general character that could hardly be accorded such a role, it is perhaps more appropriate to use the word 'induction' to denote processes of this kind. Outside the area of the central nervous system there are many other ectodermal organs—balancer, gills, hypophysis, etc.—that are dependent upon other structures for their differentiation. Whether the ectodermal structures just referred to arise *in situ* through some localizing factor in the ectoderm itself, or whether they are produced by reaction with the underlying tissue is still uncertain in some cases. . . .

"In growth, as well as in differentiation, cells and cell groups are dependent upon one another. The proportions of an organism are acquired through the maintenance of the normal relative growth of its parts. Delicate adjustments are here necessary, but they are little understood, though known to be governed by both genetic and en-

vironmental factors. Rates of growth are specific, and each organ or part has its characteristic rate relative to that of the whole. Relative growth rates are susceptible to external and internal environmental influences, though in a much less degree than the general growth of the whole organism. . . .

"As stated above, there is almost no evidence that regions which are far apart affect one another during early stages of development, but as soon as the circulation and nervous connections develop, factors influencing differentiation become effective through these media."

Differentiation presents us with an endless variety of new substances and new structures, whose nature is mainly predetermined by the minute proportions of genic material in the zygote, and the tiny weight * but huge numbers of specific molecules in the zygote cytoplasm. It is maintained that these complicated, variable, but orderly procedures may be best understood as due to the directive activities of specific catalysts, genic and non-genic, the order, quantity and nature of whose chemical output is specifically varied, in time and place, by changes in them induced by specific catalyst modifiers, or established in them by prosthetic groups and carriers.

The basic modifier particles seem to exist in the inherited zygote cytoplasm, but other modifiers may come from the milieu or from food, or may be formed by the activity of the initially existing catalysts. Modification may involve total or partial inhibition of a catalyst, or else a promoter action, whereby certain potentialities of the catalyst are favored relative to other potentialities. Prosthetic groups and specific carrier substances formed by new or by modified catalysts, or set free by elution or desorption, may unite to form still other specific catalysts, whose chemical output underlies the specific structural pattern of the developing biont.

Associated with catalyst modification, and perhaps efficient in their own right as conditioning catalyst activities, are changes in the milieu immediately adjacent to a catalyst particle or surface. Such changes, especially in the absence of a circulatory system, would follow upon selective adsorption and differential diffusion of substances formed catalytically.

The catalytic view of life and of life processes accounts for the otherwise mysterious fact that from tiny spores or cells there arise, in regular order, such a host of different plant and animal structures. The initial "seed" contains merely the *determiners* of chemical changes that lead to the relatively huge mass of the individual. The Old Testament stressed *seed* rather than *blood,* genetics, rather than consanguinity, which is a consequence of catalysis.

Chemical Activators †

A great mass of evidence has accumulated to indicate that many important biological changes are directed, or powerfully influenced, by *exceedingly minute proportions* of highly specific substances, even though the *actual number of molecules* (or other particulate units) involved may be enormous. Some activators come from outside the organism, *e.g.,* vitamins from food, trace substances from the soil; others are formed within the organism, *e.g.,* auxins in plants, hormones and neurohumors in animals.

In 1921 Otto Loewi (Nobel Laureate) discovered [92a] that nerve impulses provoked by stimulation of the nerves leading to the heart, act in such a way that they release at the nerve ends chemical substances (acetylcholine or adrenaline), and

* H. J. Muller calculates ("Hors de la Nuit," Gallimard, Paris, 1938) that within the volume of half an aspirin tablet could be packed the sperm-heads required to produce two billion human beings.

† In discussing the activation of papain, J. S. Fruton and M. Bergmann [*J. Biol. Chem.,* 133, 153-156, (1940)] state: "The fact that the various activators may produce different specificities should play a significant role in the biological action of the intracellular enzymes."

that these substances are responsible for the changes in heart activity produced by the nerve impulses. Subsequently it was proven that this "chemical transmission" of nerve impulses to effector organs is generally valid within the peripheral nervous system. Torda and Wolff [185b] report that there is an actual defect in the synthesis of acetylcholine in *myasthenia gravis,* a disease involving a slowly progressive and fatal fatiguability and weakness of the muscles. Their investigation was suggested by the observation of Walker [145a] that prostigmine, an inhibitor of cholinesterase, aids patients suffering from this disease, apparently by husbanding the lowered supply of acetylcholine.

Sir Henry H. Dale and collaborators found that each of the hundred (more or less) nervous impulses sent per second down a motor nerve fibril liberates at the nerve-muscle contact a cloud of molecules of acetylcholine, estimated by Langmuir to contain about 30,000 molecules. But these contraction-causing molecules are promptly de-esterified by an enzyme (cholinesterase), so that the gross muscular contraction is made up of a series of rapid "quantum" twitches, which may become evident as tremor if fatigue, age, or disease sufficiently lowers the twitch tempo.

Prof. George H. Parker [103] states that present evidence indicates that the terminal of a discharging neurone acts as a miniature gland, and produces a chemical substance able to excite the tips of the next neurone and thus initiate another impulse. Such chemical activators are called *neurohumors,* a term introduced by Henri Fredericq; and Parker would here include not only acetylcholine (Dale) and sympathin (Cannon), but also adrenalin and pituitary gland products known to activate animal effectors. Parker states:

"Neurohumors may act over distances of microscopic if not ultramicroscopic proportions or over large ranges in the animal body. Such substances, irrespective of their extent of spread, are in principle nervous activators and their grouping under one head, namely that of neurohumors, is fully justifiable. Some students of this subject would class all neurohumors as hormones; others would include under this heading only such as act over long distances, but this does not seem to me to be a matter of serious import. Neurohumors really act as hormones over shorter or longer ranges. The precise problem we have to face is not the classification of neurohumors, but the extent to which these substances actually exist."

Parker then discusses the nervous control of the color of fishes, going back to Pouchet (1876), who showed that if the integumentary nerves of a turbot are cut, the melanin granules in its melanophores undergo a *dispersion,* spreading out the color and darkening the skin. Parker further observes [104] that there are two classes of neurohumors: *hydrohumors,* soluble in water and therefore found in blood, lymph, etc., and *lipohumors,* soluble in lipoids, fats, oils, fat solvents, etc. He finds [105] that the color changes of catfishes are controlled by three chief neurohumors: intermedin from the pituitary gland; acteylcholine; and a concentrating neurohumor (probably adrenalin) from nerve fibers controlling, respectively, concentration and dispersion. Acetylcholine induces dispersion of melanophore pigment; adrenalin causes the reverse. Both are lipoid-soluble, and acetylcholine, in its fatty retreat, may thereby be protected from destruction by cholinesterase; in fact, it accumulates to such an extent that its effects may persist after nerve action has ceased, and it may be extracted in measurable amounts from dark fish skins. "Thus the fatty or lipoid substances in the animal body may serve as storage reservoirs for agents that may be of first consequence in the animal economy."

Francis B. Sumner [129] has discussed the quantitative changes in pigmentation resulting from visual stimuli in fishes and amphibia. We thus get an idea of how and why flat-fishes, chameleons, etc. change color in response to their surroundings. Measurable electric phenomena are associated with nerve functioning, and may be picked up from the heart (electrocardiograph), the brain (Berger rhythm), and in individual nerves (Adrian).[4] D. Nachmansohn and collaborators [56a,98a] found

that at neuro junctions and synapses the distribution and high concentration of cholinesterase is such as to split within milliseconds amounts of acetyl choline which, if released at those foci, would be in sufficient concentration for a stimulating action. They also advance the view * that "acetyl choline metabolism is intrinsically connected with the electric changes occurring everywhere at the neuronal surface. Hence, it is only quantitatively more important at the synapse where the neuronal surface increases considerably due to the extensive end-arborisation. . . . Bioelectrical phenomena occur at surfaces. The localization of the enzyme at the surface and the high rate of acetyl choline metabolism are particularly pertinent in view of experiments in which a close parallelism could be established between the electromotive force of the action potential and the concentration of cholinesterase." This new concept reconciles the "electrical" and "chemical" theories of transmission of nerve impulses, for the main question at issue was whether or not transmission at the synapses was of a different kind than that in the axon.

It must be most strongly emphasized that no *one* mode of action can be invoked to explain the influence of all "trace" substances, especially with increasing complications in biont structure and organization. As indicated above, and as Svedberg found with the proteins (see p. 562), small amounts of specific substances may not only greatly affect the *degree of dispersion* (with corresponding results on kinetic activity and diffusibility, as pointed out on p. 554), but they may also greatly influence the *permeability of membranes or septa.* Thus moistening the gut with bile facilitates the passage of emulsified fats.[95, 189] Many years ago R. Zsigmondy and Fr. N. Schulz [163] showed how a trace of egg albumen may permit colloidal gold to pass through Chamberland, Maassen, and Pukall filters, evidently because the filter pores preferentially adsorb a layer of albumen, thus preventing adsorption of gold particles which could clog the pores. Besides, the albumen exerts a deflocculating and protective action on the colloidal gold.

Another and most important function of trace substances is that they may serve as *catalysts,* or as *carriers* or *prosthetic groups for catalysts.* The direct catalytic results may not be recognizable in the often far-removed phenomena actually observed. For example, from 5 to 60 parts of copper in one hundred million of sea-water are essential for the transformation of a free-swimming oyster larva into a stationary "spat"; in brackish water a colloidal precipitate of copper oxychloride is formed which initiates the change about 4½ minutes after ingestion.[112] A trace of copper seems essential to hemoglobin formation,[65] and Florida orange trees whose leaves had lost most of the green color (a condition known as "frenching") recovered when copper sulfate was added to the soil, their leaves yielding 4.6 times as much chlorophyll as the controls.[100] When one gram of $MnSO_4 \cdot 4H_2O$ was added to 15,000 grams of Sassafras sandy loam soil (pH 7.1), tomatoes cultured in pots of the treated soil gave a pulp showing an increase in ascorbic acid (vitamin C) from 142 to 243 milligrams per liter.[72]

Relative to differences in sex of plants,† Prof. W. F. Loehwing states: [92] "More than a decade of research by the writer and his students upon dioecious species suggests that, among the most important physiological differences of the sexes, are the oxidation-reduction enzymes, marked quantitative if not qualitative differences in soluble carbohydrates, and qualitative contrasts in proteins. Mild hydrolysis of proteins in sexually differentiated hemp plants discloses arginine and lysine but no proline or histidine in staminate individuals, while pistillate plants are the exact converse. Tests for these amino acids in foliar tissues of vegetative plants without flower primordia enable one to forecast the eventual sex of the mature plant (Kiesel and Pachewitsch, 1938; Loehwing, 1939). Inception of staminate organs seems to be characterized by a preponderance of soluble sugars and presence of oxidase. Pis-

* *Science,* 1943, paper by Fulton and Nachmansohn.[56a]

† Re sex in unicellular organisms see [124a].

tillate loci have a preponderance of nitrogen and marked reductase activity (Loehwing, 1937; Stanfield, 1937). Though the foregoing appear to be significant physicochemical differences between the sexes, we have not yet successfully distinguished cause and effect, nor have we the desired degree of voluntary control over carpo- and andro-genesis."

Since the chemical units represent mainly consequences of catalytic activity, it would appear that the differences described probably stem back to the original cellular biocatalysts.

In quite another field, we find remarkable evidence of chemical control in bees. In full season a queen bee may lay in excess of 2,000 eggs daily, more than her own weight. Unfertilized eggs give rise only to drones (males), but the fertilized eggs may give rise to workers (sexually immature females) or to queens: nutritional difference decides the outcome. Townsend and Lucas state:[186]

"All female larvae are fed on royal jelly for the first 2-3 days after hatching and during this period their anatomical development is similar. Only the queen continues to receive this special diet. Any larva from a fertile egg, if given royal jelly throughout its larval period, develops sexually so that it becomes a perfect or true female bee, or what is called a queen; otherwise, the larva develops into a sexually immature worker. The queen is structurally much the same as the workers but with these important differences: the pollen-gathering apparatus remains undeveloped, the mouth parts and sting are modified, while the spermotheca and ovaries are highly developed."

These authors separated royal jelly into four fractions and found some evidence that the physiologically active material responsible for the sexual development of the queen bee is in the ether-soluble fraction. More recently Pearson and Burgin[108] report that bee royal jelly is the richest known source of pantothenic acid, its 35.8 per cent dry substance containing from 378 to 618 (average, 511) mg of pantothenic acid per gram.

In discussing ants, G. H. Carpenter stated:[38]

"One of the most interesting features of ant-societies is the dimorphism or polymorphism that may often be seen among the workers, the same species being represented by two or more forms. Thus the British 'wood-ant' (*Formica rufa*) has a smaller and a larger race of workers ('minor' and 'major' forms), while in *Ponera* we find a blind race of workers and another race provided with eyes, and in *Atta, Ecitron* and other genera, four or five forms of workers are produced, the largest of which, with huge heads and elongate trenchant mandibles, are known as the 'soldier' caste. The development of such diversely formed insects as the offspring of the unmodified females which show none of their peculiarities, raises many points of difficulty for students of heredity. It is thought that the differences are, in part at least, due to differences in the nature of the food supplied to larvae, which are apparently all alike. But the ovaries of worker ants are in some cases sufficiently developed for the production of eggs, which may give rise parthenogenetically to male, queen or worker offspring."

Discussing the role of chemoreception in the ability of rattlesnakes to recognize ophidian enemies, C. M. Bogert[26] states that although rattlesnakes when confronted by man or by domesticated animals, normally assume a coiled defense attitude, with head raised ready to strike, immediately after exposure to ophiphagous king-snakes they react to contact and to visual stimuli only by assuming a characteristic "king-snake defense posture": head and tail flat on the ground, with the body arched to a high loop with which to strike a defensive blow. Rattlesnakes assumed this characteristic posture when placed in a receptacle that had previously held king-snakes, and also when a clean stick that had been rubbed on the back of a king-snake was held near. The activating substance is apparently taken up by the tongue and conveyed to Jacobson's organ, which contains olfactory cells; and the response was

evoked by visual or by contact stimuli as long as three hours after receipt of the original olfactory stimulus. Apparently some trace of a volatile substance powerfully conditions the behavior of the rattlesnake.

H. S. Raper and collaborators have shown [114] that the enzyme tyrosinase catalyzes the conversion of tyrosine (an amino acid constituent of some proteins) into 3, 4-dihydroxyphenylalanine, called *dopa* for short. Dopa is then oxidized by the same enzyme to a red indole derivative which spontaneously changes to *melanin,* a pigment which gives hair and skin a dark brown or black color. Factors affecting the activation of the precursor of tyrosinase have been studied by J. H. Bodine and collaborators.[28] The pro-enzyme separates in the aqueous layer obtained by ultracentrifuging mashed grasshopper eggs, and is activated when added to the oily or lipid layer. Since chloroform, urethane, urea, heat and detergents (Duponol, Areosol) all produce the same final effect (active tyrosinase), Bodine states:

"The enzyme, tyrosinase, is protein in nature and as such shows all the characteristic properties of this chemical group of compounds. It, therefore, would seem logical to expect that activation may possibly be related to or dependent upon some physicochemical properties of this protein molecule and that the various activators employed bring about just such changes. Without adding unnecessary details in the way of experimental data, it may be stated that the activation of the melanin-producing enzyme in the present studies is thought to be brought about by the selective adsorption and orientation of the pro-enzyme molecules by the activating agents." Activation ceases to increase when the activator surface is occupied, either by pro-enzyme, or by another adsorbed protein.

The black spots on a Dalmatian dog appear where the local cells enable the melanin-producing process to occur, the remaining hairs being white. L. Earle Arnow [17] found that in the presence of oxygen, tyrosine may be converted by ultraviolet light into dopa, which is then changed into melanin by an oxidase. Sun-tan appears to be so caused in some persons. A curious case of inhibition of skin pigment formation appeared in a large tannery, where a considerable number of workers (about 50 per cent, many of them negroes) developed white patches on their skins.[117] Investigation showed that the rubber in some new gloves used to protect the workers' hands from an acid solution, had contained the monobenzyl ether of hydroquinone as an antioxidant, and that this substance also inhibits melanin formation; so that when the melanin originally present in the skin is absorbed or otherwise removed, the skin turns white (leukoderma). When the cause was removed the skin slowly returned to its normal shade, indicating that the pigment-forming mechanism was merely inhibited by the antioxidant, not destroyed. Normal skin color thus seems to represent a balance between pigment formation and removal.

E. L. Tatum and G. W. Beadle state: [183] "The development of eye color in *Drosophila* is known to be controlled by specific diffusible substances designated as v^+ and cn^+ hormones." By growing certain bacteria on an agar medium containing dead yeast, sugar, and *l*-tryptophane, "this bacterially produced v^+ hormone has now been obtained in a pure crystalline state," with "an activity of approximately 20,000,000 *v* units per gram when a solution is injected into vermillion brown test larvae."

An indication of how chemical control by trace substances may dominate fertility and thereby contribute an important factor to the course of evolution, is found in the Golden Rose strain of Petunia, which is completely self-sterile under natural conditions. Microscopic examination shows that its pollen tubes grow slowly, and that even before the most rapidly growing tubes reach half way down to the ovary, an abscission layer forms and blocks the way. Yasuda [151] found that the placenta in the ovary of *Petunia violacea* secrets a "special substance" which diffuses into the style and completely inhibits pollen germination and tube formation. W. H. Eyster [58a] found evidence that the ovarian secretion of the Golden Rose Petunia "which renders the plant self-sterile, can be transferred to other plants and renders them cross-sterile

with pollen from self-sterile plants." Golden Rose Petunia can be self-fertilized by two methods: (1) "If flower buds which are beginning to develop anthocyanin in the petals are opened and pollenated with pollen from fully opened flowers from the same plant, seed capsules containing viable seeds are produced"; similar results were found by Yasuda,[152] who calls this homo-pollination; (2) by spraying the flowering plants with a solution of ten parts of α-naphthalene acetamide in one million parts of water.

"Flowers which are sprayed with this solution immediately before or shortly after they have been self-pollinated produce seed capsules filled with viable seeds in exactly the same way that normal self-fertile plants of other strains produce seeds. Obviously α-naphthalene acetamide neutralizes the effect of the ovarian secretion which diffuses into the style and inhibits or greatly retards the growth of the pollen tubes."

Dwarfism in certain plants (*e.g.*, maize, pea) has been shown to be due either to deficiency of auxin (growth hormone) or to its destruction by oxidative enzymes. J. van Overbeek[101] further found that "laziness" in maize (a prostrate habit of growth) is consequent upon higher concentration of auxin in the upper part of the plant, whereas normally the reverse is true. He concludes: "It is thus evident that the lazy gene interferes with the auxin distribution in the stems which normally takes place under the influence of gravity."

While it has for some time been known that over about 1.5 parts per million of fluorine in drinking water generally produces a discoloration of the teeth known as "mottled enamel" (endemic dental fluorosis), recently Dr. H. Trendley Dean and collaborators[41a] found that fluoride levels of less than 1.0 ppm were accompanied by a corresponding increase in dental caries. While the function of fluorine has not yet been established, it seems possible that the liberation of fluorine ions locally, by acid-producing bacteria would result in the inhibition or death of the bacteria if the fluorine concentration became sufficient before material attack on the tooth had resulted. The fluorine probably also gives the tooth a denser and less readily attacked structure. "Bone gives an x-ray diffraction pattern similar to that of the mineral apatite, the unit structure of which contains $Ca_{10}(PO_4)_6F_2$. Various substitutions, such as $(OH)^-$ for F^- and Mg^{+2} for Ca^{+2}, are known to occur in the apatite lattice without producing significant changes in the diffraction pattern."[69a]

Wide use is made of ethylene gas (about 1 part per 1000 of air) to "color up" fruits and vegetables which must be shipped more or less "green" in order to stand handling and avoid spoilage. Bananas, melons, pineapples, persimmons, tomatoes, oranges, apples, pears, etc., are thus treated, the natural coloring processes being greatly accelerated. Celery may be thus blanched, and the hulls loosened from "stick-tight" walnuts. The diverse changes seem to be dependent upon initiating or facilitating specific enzymic changes.

Spraying apple trees with soluble thiocyanates, though it causes "spray burn" and a chlorotic condition of the leaves, tends to increase the red color of the fruit (blush) and to turn the green ground color toward yellow, both features desirable to consumers.[48] The red color is due to idaein, a glucoside which, on hydrolysis, yields cyanidin and galactose. There is some evidence that spraying apples with naphthaleneacetic acid and related compounds tends, with some varieties, to reduce the percentage of windfalls, apparently by firming the stems.[74a]

D. W. Wooley has shown[149] that the mouse requires a new "vitamin" for normal growth and maintenance of hair. The facts indicate "that the mouse anti-alopecia factor is inositol or its derivatives. They suggest that inositol exists in liver in alkali-labile combination with a large molecule which renders the former non-dialyzable."

These scattering instances, which might be multiplied many times, illustrate how devious, various, and potent may be the effects of small amounts of substances, which,

until comparatively recent times, were generally considered "negligible" in reports of chemical analyses. Their effects are often expressed through catalysts.

Deviations from Normality

Genic and Chromosomal Abnormalities, Natural and Induced

Darwin's great book, "The Origin of Species" (1859), is based on the notion that species were and are slowly and continuously changing, and that "natural selection," working upon the relation between these specific variations and the simultaneously changing environment, led to the "survival of the fittest" and the slow serial changes in plants and animals which we term "evolution."

In 1894 William Bateson in his book, "Materials for the Study of Variation," * collected data which indicate that some variations are *discontinuous,* or to use an expression which followed the publication in 1901 of H. de Vries's "Mutation Theorie," that mutation is *a physiological event.* De Vries distinguished *mutations* from the common *variations,* found even in identical twins,[98a] by the fact that mutations are transmissible to descendants and are therefore more permanent.

The term "mutation" should be, but is not now limited to changes in single genetic units (genes), which, according to our view, are the basic self-duplicating catalysts of the living unit (biont). The changed gene, commonly called a *mutant gene,* would lead (directly, or indirectly through the formation of catalysts, carriers or prosthetic groups) to alterations in the rate and/or the nature of chemical changes, that is, changes in the *chemism* of the biont, which may, in turn, lead to detectable consequences. In the case of *Drosophila melanogaster* it is estimated that about 2 or 3 per cent of the flies show detectable mutations, and of these about one-sixth are lethal. In view of the relatively huge molecular size and complicated structure of even a single gene, it is more surprising that precise self-duplication should occur so regularly, rather than that there should be occasional lapses, leading to changed effective catalyst surfaces. Such catalyst surface changes may follow (1) some form of "intramolecular change"; (2) the addition of or the subtraction of some physical unit (ion, atom, molecule, colloidal particle); (3) a change in position of a gene relative to some carrier substance in the gene-string, corresponding to the "position effect" well recognized by geneticists.

The causes of spontaneous or naturally occurring point mutations are obscure. Obviously, the gene must be sufficiently unstable to respond to such probable natural environmental impacts as radiation, activated atoms or molecules, or even the kinetic laws of chance. There is some evidence that mutations occur more frequently in the presence of increased natural radioactivity. Cosmic and solar radiations are possible but still unweighted factors, as are also the effects of such other natural physical agencies as heat, cold, desiccation and darkness. The effects of naturally occurring elements and compounds must also be considered. Most of us have observed, as did Faraday, the development of a violet color in white glass exposed to the sun, presumably caused by aggregation of manganese dioxide. Colorless gold ruby glass develops its red color when radiation (*e.g.,* by radium) causes its amicronic gold particles to aggregate to visible ultramicrons. Commercially, the color in gold ruby glass is developed by a heating process, and the development and growth of the ultramicrons can be followed in an ultramicroscope. (Zsigmondy) [158]

The artificial production of mutations was initiated by the researches of H. J. Muller with x-rays. In two recent reviews,[98] one on induced mutation in *Drosophila,* Muller stated that at present radiation is the only known external agent capable of producing mutations in abundance and with predictable frequency. He

* "Bateson's call to a new study excited but little interest, and in a few years his book was to be found among the list of cheap remainders." Prof. R. C. Punnett, in "Background to Modern Science," p. 205 (The Macmillan Co., N. Y., and Cambridge Univ. Press, 1938).

pointed out that the effectiveness of radiation can be increased by the use of superior apparatus, but that even now it far exceeds usual calculations. For if a dose of 10,000 *r*-units acting for an hour * can give about 100 times as many lethals per X-chromosome in *Drosophila* spermatozoa as ordinarily occur in untreated material, we must remember that the "natural" mutations, used as a datum, represent the accumulation of a *Drosophila* generation covering two weeks or more—say 350 hours. The x-ray treatment has, then, increased the lethal mutations at least 35,000 times. To the lethals we must add the large number of detrimental and undetectable mutations, apart from the "desirable" ones sometimes found.

There is no room here to discuss the extensive evidence as to the nature of the physico-chemical changes involved in mutations induced by radiation. The following facts are now acceptable to geneticists: (1) "As to the mechanism of production of those radiation mutations which affect individual genes or narrowly circumscribed chromosome regions, the most important fact that has come to light is the dependence of any given mutation upon a change that was initiated in a single individual atom, by its ionization or other excitation." (Muller, *loc. cit.*, p. 151). (2) Structural changes occur in the chromosomes, because of breaks in the chromatin and a two-by-two fusion of the ends derived from different breaks. This cohesion of the loose ends suggests the presence of some substance which functions as a *cohesive colloid,*[11] and recalls the behavior of some bacteria (especially those with capsules), and of the two rubber latex layers on "self-sealing" envelopes, which cohere when lightly pressed together. The breaks may be gross, or else too minute to be microscopically demonstrable, and the phenomenon of "position effect" is evident. Thus, referring to the work of Panshin [102] and of Dubinin and Sidoroff [46] Muller states: "In these works it was shown that, when a normal allel derived from a chromosome of normal structure is substituted by crossing over for a gene which, since the occurrence of a rearrangement in its vicinity, has functioned like a mutant gene, the substituted normal gene now takes on the properties that the apparently 'mutant' gene had, while, conversely, the removed 'mutant' gene reverts to normal functioning after its transference back into the chromosome of normal structure."

Reviewing seven years work with x-rays in the white-notch region of the X-chromosome of *Drosophila melanogaster,* M. Demerec and collaborators [44] conclude that there is an intimate relation between the various parts of the chromosome, the activity of a gene depending greatly upon its immediate environment. "The activity of a gene may be changed or suppressed by chromosomal rearrangements through which a foreign section of a chromosome is brought into contact with the gene."

To the writer, these findings indicate that the gene functions somewhat like a prosthetic group in an enzyme, for the activity of a prosthetic group depends upon the nature of its "carrier," and *vice versa.*[143]

Muller believes that chromosomal breaks occur higher up in the "ascending hierarchy" of structure than peptide bonds. Rather than the simple plus-and-minus relationships suggested by Muller, the specificities involved in re-joining chromosomal fragments appear to involve pattern charges or mosaics of electronic areas, especially on close approach of units.

Mutations have been produced by ultraviolet radiation [51, 90] as well as by x-rays, and M. Demerec and collaborators [44] report that neutrons are effective in inducing chromosomal rearrangements similar in frequency and general nature to those induced by x-rays. Mutation of bacteriophage has also been reported.[82]

In reviewing a number of reports of gene mutations produced by such chemicals as iodine, copper sulfate, ammonia, potassium permanganate, mercuric chloride and lead salts, T. Dobzhansky [45] expresses the hope that the experiments will be repeated

* I am informed that gene or point mutations depend mainly upon the total *quantity* of radiation; chromosomal breaks on both *quantity* and *time* of radiation J. A.

and extended. Steinberg and Thom [127] have reported the production of mutations in fungi (*Aspergilli*) by the use of sodium nitrite in acid solution (*i.e.*, nitrous acid) and of colchicine in the presence of excess of calcium carbonate. "The presence of high concentrations of *d*-lysine, the amino acid presumably affected by nitrite, brings about a reversal in mutation. Chemical conditions conducive to re-introduction of amino groups also bring about reverse mutation. Nevertheless, the precise nature of the underlying process is by no means certain as yet in the absence of cytological and analytical data. Nor is the formation of similar type mutants by colchicine particularly helpful to the interpretation of amino-group destruction. The mode of action of colchicine on fungi itself needs elucidation, since polyploidy need not necessarily or entirely enter into the observed responses of fungi."

The production of chromosome mutants by chemical means has been extensively demonstrated, principally by the use of colchicine, which by inhibiting spindle formation leads to cells with doubled chromosome number.[15, 22, 31, 47, 66, 91, 98] Successful results have also been obtained with acenaphthene and β-indole acetic acid (heteroauxin).

The transformation of Type II pneumococcus into Types I or III, mentioned previously may be a case of a chemically induced mutation, (even if we do not know the chemical nature of the effective mechanism), or else there is formation of a new and self-duplicating cytoplasmic catalyst, whose formation may in some cases be established by conditions favoring the union of existing carriers and prosthetic groups. Another case appears in the crown gall ("plant cancer") produced by inoculation of a host plant with the bacterium *Phytomonas tumefaciens,* which was shown by E. F. Smith and collaborators [120] to develop true secondary tumors in sunflower plants several internodes away from the primary tumor. In confirming this work recently, P. R. White and A. C. Braun [145] found that the secondary tumors were generally sterile; and using recently developed techniques, they grew secondary tumor tissue *in vitro*. They found that it bred true; and the medium was bacteria-free, though capable of supporting a profuse growth of *Phytomonas tumorfaciens*. The authors conclude that the affected tissues have undergone a drastic change, indicated (1) by the abnormal capacity to produce galls, and (2) by their abnormal growth habits *in vitro*. "That this change was originally brought about by some stimulus from the crown gall organism seems clear. That its maintenance is not dependent on the continued presence of the bacterium is equally clear."

Such results are understandable on the basis of chemically induced gene point or chromosomal mutation, or of the heritable modification of a gene or other cellular catalyst, by substances originally introduced by presence of the bacterium. In the case of modification, the modifying substance must be continually reproduced thereafter in the clone of cells, presumably by newly formed specific catalyst areas; or else the modifying particles originally introduced must form an autocatalytic (that is, self-duplicating) specific catalyst area or template. With modifying particles of complicated chemical structure, like many of the cancer-producing compounds, the latter effect seems more likely, unless there is a chemically induced mutation.

It seems possible that some viruses, entering the body as infective agents, may persist there and be transmitted to progeny, possibly forming a permanent addition to the self-duplicating units of the cytoplasm of body cells or of the body fluids. Mice experimentally inoculated with a non-fatal dose of choriomeningitis virus develop an immunity in 5 to 7 days and rapidly free themselves of demonstrable virus; but pregnant mice when inoculated seem to pass the virus on to their offspring, which develop after birth almost as normal mice, but continue to carry the virus and pass it on to their progeny.* In his Kober Lecture,† Dr. Charles Armstrong reported that a strain of choriomeningitis virus had been passed "in an unbroken series from an

* E. Taub, *J. Exp. Med.*, **68**, 229 (1938).
† *Military Surgeon*, **91**, 129-146 (1942).

infected mother through 11 generations of her offspring with no apparent tendency of the virus to weaken or disappear. It has also been observed by Haas,* that these congenitally infected mice are far more efficient transmitters of the disease to normal cage mates than are mice inoculated after birth." Here is another possible mechanism of inheritance, apart from the genes.

Abnormal Catalysts and Modifiers

We consider here only a few illustrative examples of the effects on the organism of small percentages of abnormal or strange catalysts and catalyst modifiers, disregarding the gross solvent or dispersing action which may be exerted by high local concentrations of such substances as phenol, acids, alkalies, etc. If the introduction into, or the formation within the body of foreign molecules results in harm to the organism, the material is generally classed as a poison; but here the quantitative factor is very important. Many poisons (*e.g.*, arsenic, strychnine) may be beneficial in small dosages; and essential trace elements (*e.g.*, iodine, copper) may be toxic if present in amounts which may still be considered small. The intense specificity of action of various drugs and poisons is the basis of their uses; and many of them, irrespective of the clinical effects or symptoms which they elicit, operate essentially by their influence upon biocatalyst systems, *e.g.*, cyanides, the "sulfa" drugs.†

Strange molecules are in some cases directly *eliminated* in the excretions (urine, feces, breath, sweat); in other cases they are *detoxicated* by combination with substances normally present, or by subjection to chemical change, presumably through the agency of specific catalysts, whereby they are oxidized, reduced, or otherwise converted into some harmless compound which may then be eliminated or destroyed. Many of the chemical detoxication processes appear to take place in the liver, that hotbed of potent biocatalysts. D. D. Van Slyke,[138] in discussing the physiology of the amino acids, points out that the liver may transform these compounds into urea, or into glucose, or into a storage form of protein. "The liver also appears to be the place where plasma fibrin and albumin are formed."

The mode of detoxication varies with the changing catalytic and chemical structure of the animal. Thus phenylacetic acid and its homolog benzoic acid, are combined, in fowls, with ornithine. All mammals combine benzoic acid with glycine and eliminate it as hippuric acid; and the lower animals, including monkeys, combine phenylacetic acid with glycine, eliminating it as phenaceturic acid. But Sherwin[118] found that man couples this latter poison with glutamine and eliminates it as phenylacetylglutamine. Experimenting on the chimpanzee, F. W. Power[110] found that this "man-ape" eliminates phenylacetic acid in the same manner as does man. Here too, as in the biochemistry of muscle, there is a definite relationship between chemical mechanisms and the position of the organism in the evolutionary or taxonomic scale.

Disease

The influence of heredity factors on individual health and behavior was recognized in the four "temperaments" of the ancients: (1) sanguine, (2) phlegmatic, (3) bilious or choleric, (4) melancholic. Later on medical men spoke of "diathesis," which generally meant a predisposition to some disease, *e.g.*, gout. Even Aesculapius had described the structural type of individuals who were predisposed to tuberculosis.

In his recent book on medical genetics, L. H. Snyder[128] points out that the familiar 3 : 1 ratio of Mendel no longer covers the major portion of the field of heredity, for geneticists have demonstrated varied types of heredity transmission—dominant, recessive and blending—as well as varied relationships and behaviors of

* V. H. Haas, *Pub. Health Rep.*, 56, 285 (1941).

† See, *e.g.*, "The Interaction of Drugs and Cell Catalysts," by Prof. Frederick Bernheim (Burgess Pub. Co., Minneapolis, 1942).

genes, *e.g.*, autosomal, sex-linked, sex-influenced, lethal, epistatic and combined genes as seen in multiple allels and multiple factors. As an example of epistasis,* it has been observed that in rabbits and mice the factor for gray, if present, prevents the development of the factors for black and chocolate. Many traits and abnormalities are transmitted by heredity, *e.g.*, polydactylism (supernumerary digits on hands or feet), syndactylism (webbed fingers or toes). Strains of Dalmatian dogs have been selected and bred which take definite positions—some close to the horse, some beneath the wagon, and some behind it.[82]

Among the many diseases transmitted by heredity are allergies, hemophilia, endemic goiter, and a variety of nervous conditions including insanity and Huntington's chorea.[140] The violent convulsive movements in this last-named disease generally appear in adults and follow degenerative changes of ganglion cells of the frontal cortex and of the cordate nucleus and putamen. Davenport[40] states that traits known to have a chemical basis have the cleanest-cut genetics, *e.g.*, blood groups, cretinism (thyroid deficiency), amurotic family idiocy (a recessive associated with the storage of phosphatides).

A striking instance of inherited metabolic error is the rather rare alkaptonuria (recessive) first described by von Boedecker,[24] and recently observed in an American Negro family.[1] In these cases the urine contains homogentisic acid, shown by Wolkow and Baumann[148] to be 2, 5-dihydroxyphenyl acetic acid, which is labile and oxidizes, turning the urine black. Apart from occasional blackening of cartilage (ochronosis) no other trouble is evident. Homogentisic acid apparently arises from the blocking, at this stage, of the normal oxidation of tyrosine and phenylalanine (constituents of many proteins), for normal persons fully oxidize this acid and perfusion through normal liver converts it into acetoacetic acid. Dakin[89] showed that alkaptonuria patients can fully catabolize *p*-methylphenylalanine and *p*-methoxyalanine, which cannot form quinoid derivatives; he believes that in these cases inability to catabolize homogentisic acid is associated with increased formation of the quinoid intermediate from which it is derived.

Davenport[40] states: "The relation between the somatic expression of a trait and its chemical basis may be remote. Thus hardness of hearing seems to depend on a defect in calcium metabolism such as causes abnormal bone formation at the oval window of the inner ear and simultaneously in other parts of the temporal bone.† In the latter case there is reason for concluding that the result depends on a dominant factor in an autosome which modifies the reaction of the mesenchyme and a sex-linked gene which perhaps affects calcium metabolism. Indeed it seems probable that in time chemical errors in the body may throw light upon the chemical processes of development.

"The fact that so many mutations have a known chemical basis and that development is, indeed a biochemical process, raises the question whether all mutations have not, and leads us to seek the chemical basis of any defect. As the chemical bases of mutations are discovered the mutations may well be classified on a chemical basis rather than a morphological."

An instance of the importance of environmental conditions in aiding to establish the effects of stranger molecules appears in the photosensitization of organisms. In 1897, O. Raab observed that small amounts of acridine, while without effect in the dark, exerted a lethal action on paramecia in the presence of light. Later, similar effects were seen with eosine, chlorophyll, etc. In a recent book on this subject, H. F. Blum[22a] describes a number of diseases which develop in grazing animals (sheep, horses, cattle) when they eat certain plants, *e.g.*, fagopyrism, due to buckwheat (*Polygonum fagopyrum*); hypericism, due to St. Johnswort (genus

* Epistasis, in Mendelian inheritance, means "the expression of one character to the exclusion of another not of the same allelomorphic pair" ("Standard Dictionary").

† There are other forms of deafness than otosclerosis. J. A.

Hypericum); geeldikkop ("yellow thick head") caused by the South African plant *Tribulus*, sometimes called "devil's weed." In geeldikkop the active material seems to cause liver injury and occlusion of bile, with jaundice, leading to retention of phylloerythrin (a chlorophyll derivative). This substance, accumulating in the skin, makes the animal light-sensitive and results in edema of the head and other places exposed to light.

Blum believes that the only way to account for all the evidence is to assume that when a quantum of light strikes a dye unit in a substrate-dye combination, the dye portion is activated and thereupon transfers its activation to the substrate, which then reacts with oxygen and injures cell structure. Having found that proteins which form the solid structure of the cell are fibrous, Prof. A. Szent-Györgyi [132] (Nobel Laureate) advanced the idea that activated electrons may travel some distance within their own specific band or energy level, "and can fall to a lower level and give off energy only at a place where they can do work (*e.g.*, a synthesis), analogous to the ZnS phosphors of Riehl," [116] where "the electron, raised to a higher level by a collision with an α particle, can travel relatively long distances and will fall back to a lower level, giving up its energy, where it meets with a Cu atom present as an impurity."

While agreeing with the importance of Szent-Györgyi's view that there exists a mechanism permitting the transfer of the energy of absorbed light or chemical reaction, without degradation or dispersion, for chemical reactions in relatively distant portions of the system, E. W. and F. H. Pike [109a] state: ". . . the whole chemical and physical behavior of proteins and other biochemical substances would suggest a van der Waals' binding, which does not have extended electron states, rather than the electron band binding typical of salt or the diamond. It is well known that neutral or excited atoms, molecules, or free radicals may be adsorbed on solid or liquid surfaces as a mobile two-dimensional gas. Such excited mobile entities constitute a second possible mechanism for the effects which Prof. Szent-Györgyi discusses. Others are known, and it would be a daring biologist who would suggest that there are no more undiscovered mechanisms."

Cancer *

Cancer has been defined diagnostically as a growth of cells which is progressive and invades healthy tissue; but a much clearer understanding of the basic cause of this protean group of diseases is to be had if we descend to the catalytic level. For catalysts (genes, enzymes, symbionts) dominate the chemical and physico-chemical changes which underlie morphology and physiological function.

The development of cancers in man and in animals has long been associated with exposure to certain influences and, more recently, to certain definite chemical substances. Thus the natives of Kashmir who carry a charcoal brazier (*kangri*) about their waists and often suffer burns, are prone to develop *kangri* cancer. The British surgeon Percivall Pott observed that chimney-sweeps frequently develop scrotal cancers, and later it was noted that cotton spinners exposed to oils and workers in aniline factories † have a high incidence of cancer. Seamen exposed to tar and to insolation, as well as sufferers from "x-ray burns" are likely to develop cancers, although now x-rays of certain wave length and intensity are helpful in attacking the disease.

In 1915 Yamagawa and Ichikawa [150] were able to produce cancers in the ears of rabbits by continued application of certain tars. In the valuable book of Louis F. Fieser [54] there is reviewed the fascinating story of the team-work of chemists,

* See paper by Leo Loeb in this volume.

† On dehydrogenation and condensation two molecules of β-naphthylamine yield 3, 4, 5, 6-dibenzcarbazole, which causes proliferation of the gall ducts in mice. If given orally or subcutaneously α-amidoazotoluene produces liver and bladder cancers.

physicists, physiologists and others, describing the isolation from carcinogenic tar of a previously unknown but strongly active hydrocarbon, 3, 4-benzpyrene, present to the extent of about 0.003 per cent. After identification, this substance was synthesized; and the synthetic substance was found to be identical in its action with that extracted from tar. After high-boiling tar fractions proved to be non-carcinogenic, Hieger [78] observed that active tars and oils gave a characteristic fluorescence spectrum, with bands at 4,000, 4180, and 4400 Å. It was then found that 1, 2-benzanthracene gave similar bands, and J. W. Cook prepared a large number of allied compounds, of which 1, 2, 5, 6-dibenzanthracene proved to be actively carcinogenic—the first pure compound of known structure to be thus identified. Finally, in 1933, Cook, Hewett and Hieger [85] extracted from about two tons of tar the very small percentage of 3, 4-benzpyrene, which produces epitheliomas in mice in about 5.7 months on application of dilute benzene solutions, whereas 1, 2, 5, 6-dibenzanthracene takes about 8 months to produce this effect. Later it was found that methylcholanthrene, closely related to certain sex hormones and first made by degrading desoxycholic acid (a bile constituent), is still more potent, acting in about 5 months. These substances may cause cancers to develop in various tissues, *e.g.*, epitheliomas when applied to the skin, sarcomas when injected subcutaneously. Fieser states (*lib. cit.*, p. 86):

"While proof is entirely lacking, it appears possible that many forms of cancer may originate in the metabolic production of methylcholanthrene or related substances from the bile acids, or perhaps from sterols or sex hormones, of the body. . . . Entirely unknown is the mechanism whereby certain hydrocarbons start normal cells on a career of malignancy. If a chemical reaction is involved, the nature of the change is entirely obscure. . . . Possibly the molecular dimensions and the surface activity of the substances are as important as their chemical characteristics." (p. 92).

Dr. Charles Huggins (Dept. of Surgery, Univ. of Chicago) has just reported * that "it is possible by reducing the amount or the activity of circulating androgens to control, more or less but often extensively, far advanced prostatic cancer in large numbers of patients. In this special case, androgen control seriously disturbs the enzyme mosaic of the cancer cells at least with respect to the important energy producing protein-catalysts, the phosphatases. As a contribution to the problem of cancer treatment, it is well to emphasize that any interference with an important enzyme system of a cell, normal or malignant, will cause in that cell a decrease of size and function."

The common denominator in all cancer cells, however they may first be formed or appear, is that their changed properties and behavior are heritable, even in tissue cultures. These changes stem back to a changed chemism, which, in turn, is dominated by heritable changes in biocatalysts (genes, enzymes)† whose catalytic specificities may not even exist in the smaller chemical entities into which they may be resolved or broken down when subjected to chemical examination and analysis.

As far back as 1902 Boveri [27] suggested that cancer might be due to genetic factors, and since 1929 J. Alexander [8, 12] has repeatedly advanced the view that cancer is consequent upon heritable gene mutations and/or gene or other catalyst modifications, of such nature that the specifically effective abnormal catalyst surfaces are heritably reproduced. While specific catalyst modifiers, carriers, or prosthetic groups, may themselves be reproduced (that is, produced catalytically), there is no evidence and little probability that a cancer cell primarily brought into existence by exposure to 3, 4-benzpyrene, methylcholanthrene, or 1, 2, 5, 6-dibenzanthracene, will continue to reproduce these specific and complicated compounds. What must be

* *Science*, 97, 541-44 (1943).

† The well-known chicken sarcoma of Rous follows invasion of the cells by a specific virus; but it is not known whether this virus or some of its by-products act as modifiers, prosthetic groups, carriers, or catalysts.

heritably reproduced are *the specific catalyst areas* which enforce the formation of substances that determine cancer. It is the *diagnostic or clinical consequences* and not any basic physico-chemical difference of process which distinguish the heritable changes in cancer cells from those other heritable changes in cells which underlie differentiation and also evolution, which latter we shall next briefly consider.

Physico-chemical Mechanisms in Evolution

In the popular mind at least, the unanimity of scientific opinion in affirmation of the evolutionary origin of new species of plants and animals is beclouded by divergencies of opinion as to the underlying causal factors and their relative importance. The notion that all the species first created, according to the Biblical Chronology of Archbishop John Usher,* in the year 4004 B.C., have persisted unchanged since then, was founded upon interpretations of Holy Writ by certain theologians, and is still maintained by some individuals and sects, as is also the notion that the earth is flat. The records of the rocks, read by geologists, paleontologists, and other scientists, give ample evidence of the immense age of the earth and demonstrate the reality of an *"échelle des êtres"* extending back for millions of years; and observations of biologists on the variations developing in plants and animals, both under domestication and in a state of nature, show that these variations are so wide and so numerous as to offer ample scope, *if and when they are heritable,* for the operation of natural selection (Darwin) or the survival of the fittest (Spencer) in the formation and establishment of new species.

"The term *species* corresponds to what is popularly called one 'kind of organism'; generally speaking, a species may be said to consist of individuals all of which interbreed freely under natural conditions. No very exact definition, however, can be given to this or any of the other categories of classification; they are all essentially arbitrary groups." (Charles R. Plunkett [111]).

Because of differences in *habit* (psychological and reproductive isolation) or of *habitat* (geographical and ecological isolation), some species (certain rodents) that *can* interbreed rarely do so under natural conditions, and may become cross-sterile later because of small divergent heritable changes.† Artificial insemination has produced the tiglon (a cross of lion and tiger). Breeders produce the hinny (stallion × she-ass), and the well-known mule (jackass × mare), which is generally sterile, though a few cases of fertility in mules have been reported. But the crossing of species may fail under natural conditions because the offspring is non-viable or is sterile. Non-viability may appear at any time, from the stage of fertilization (if this occurs) to an advanced stage of development; and all intermediate stages exist between complete interspecific fertility and complete interspecific sterility, which is sometimes the result of non-development of the gonads, but which may result in much simpler fashion. Thus the sterility of the Golden Rose Petunia (p. 582) is consequent upon the action of some substance which prevents the pollen tubes from reaching the ovary; and the fig trees imported into California failed to produce fruit until there was also imported a certain insect which accomplishes fertilization of the flower. Many plants are dependent upon specific kinds of insects for fertilization.

What factors underlie sterility, which ordinarily acts as a barrier between species? The proximate causes may not, of themselves, give a satisfactory answer.

* According to Usher's "Annales" (1560), the Creation began on Sunday, Oct. 23rd, with the creation of light and the Angels, and ended at sundown of Friday, Oct. 28th, 4004 B.C., after the creation of the animals and of man. Saturday was thus the first Sabbath, meaning in Hebrew *rest from labor.* It is interesting to note that the relative order in which conditions and organisms appear in the Biblical story correspond in the main to what is scientifically accepted.

† Dobzhansky, ref. 45, p. 257, gives a more detailed discussion of physiological isolation mechanisms.

Thus a cross between a diploid and a tetraploid race yields the usually sterile triploids, and sterility may also follow other chromosomal aberrations, *e.g.*, some translocations. If we follow the effects of chromosmal aberrations to a lower level of material organization, it seems reasonable to believe that they may lead to changes in the catalytic output, *e.g.*, by action of "position effect" or by upset in "genic balance," with the result that new substances may appear, or else that old substances may be formed in larger proportions and thereby reach new places and exert new effects in the developing biont. The formation of even traces of a new substance that could affect protein aggregation or serve as the carrier or as the prosthetic group for enzymes having a new functional ambit, might initiate far-reaching changes in chemical output and in physical and physio-chemical structures and function. Point mutations can produce similar results, and so may changes in, or the formation of other biocatalysts, if there exists a mechanism for their heredity continuance.

This raises the question as to whether there may not be some non-genic form of inheritance, apart from the well established genic inheritance.

In his "Philosophie Zoologique," first published in 1809, J. B. Lamarck enunciated two "laws" which may be summarized as follows: (1) Frequent and continued use of any organ gradually develops it, enlarges it, and gives it a power proportionate to the duration of its use. Disuse has the contrary effect. (2) Whatever nature has allowed individuals to acquire or to lose by influence of conditions to which their race has long been exposed, nature preserves in the generation of new individuals, providing that the acquired changes are common to both sexes, or to those who have produced the new individuals.

The sweeping generality of these "laws" was not supported by the fantastic explanation that Lamarck gave for the development of horns in cattle and of antlers in deer, which he attributed to an "interior sentiment in their fits of anger, which directs the fluids more strongly to that part of their heads." This explains certain objections to Lamarckism, expressed in the amusing lines of James Russell Lowell ("The Biglow Papers," First Series, No. 4, line 31, *et seq.*):

> "Some flossifers think thet a fakkilty's granted
> The minnit it's proved to be thoroughly wanted. . . .
> Ez, fer instance, thet rubber-trees fust begun bearin'
> Wen p'litikle conshunces came into wearin'—
> Thet the fears of a monkey, whose holt chanced to fail,
> Drawed the vertibry out to a prehensile tail."

As knowledge of the physical basis of heredity developed, it became evident that the chromosomes are the main vehicle of heredity; and they are by many believed to be the only vehicle. On the other hand, paleontologists continue to point to the definite though fragmentary fossil record, which to them indicates that results simulating the inheritance of acquired characteristics commonly occurred over the immense periods of the geological past. But they suggest no causal mechanism, and as Thistleton Dyer said in his address to the British Association for the Advancement of Science (Section D) at Bath in 1888: "Science will always prefer a material *modus operandi* to anything so vague as a tendency." On the other hand, Prof. E. G. Conklin stated:[34a] "The classic argument of the Weismannians was that *we can conceive of no mechanism* by means of which somatic changes can be carried back into the germ *cells, and therefore there is no such mechanism.* Now the fallacy of this argument is obvious; even if we could conceive of no suitable mechanism for this purpose, this does not preclude the existence of such a mechanism." With the failure of repeated attempts to bring satisfactory experimental proofs of the in-

heritance of acquired characteristics, this view became and still remains *tabu* in biological texts and teachings.*

Nevertheless, the view had its sponsors, among whom was J. T. Cunningham. He suggested [88] that the mechanism whereby "modifications produced in the soma by external stimuli could affect the determinants in the gametes in such a way that the modifications would be inherited," involved the action of hormones, that is, "special chemical compounds (which) take the place of the imaginary gemmules of Darwin's theory of pangenesis of the 'constitutional units' of Spencer." Stressing the fact that the theory of the heredity of somatic modifications is not in opposition to the mutation theory, Cunningham states: "there are two kinds of variation in evolution, one somatic and due to external stimuli, acting either directly on passive tissue or indirectly through function, and the other gametogenic and due to changes in the chromosomes of the gametes which are spontaneous and not in any way due to modifications of the soma. Adaptations are due to somatogenic modifications, non-adaptive diagnostic characters to gametogenic mutations. It is a mistake to attempt to explain all the results of evolution by a single principle." (p. 241).

In a recent paper entitled "Environmental Transformation of Bacteria," Prof. W. H. Manwaring [98a] stated: "These studies suggest a Lamarckian rather than a Darwinian world, or at least a world in which both Lamarckian and Darwinian evolutionary mechanics are operative." Henry Fairfield Osborn [100a] stated in the third of a series of addresses on the origin of species, that the theories of Buffon,† Lamarck and Darwin, usually regarded as contradictory, are really complementary, for they all turn on the question of inheritance or transmission of individual adaptation, which may furnish the key to evolution. "On the affirmative side paleontology proves that in the long run of geologic or secular time both Buffon and Lamarck, as well as Darwin, were right in their main conceptions: organs starved by unfriendly environment finally disappear; organs which do not pay their way or are starved by disuse slowly drop out of the germ-plasm; vitally essential organs are either absolutely stable or progressive. Why not therefore concede the truth of the great conceptions of Buffon and Lamarck, even if immediate inheritance by the germ is disproved in the great majority of cases? Why not concede the still greater conception of Darwin, misled as he was as to time by the marvelously rapid evolution of the germ-plasm witnessed in artificial selection?"

Catalyst modification, which may in turn lead to the formation of new carriers or new prosthetic groups, offers a simple but potent mechanism whereby non-genic as well as genic changes may be heritably transmitted. This does not mean that any and all non-genic changes are necessarily heritable; but the unequivocal demonstration of a single case of heritable non-genic catalyst change would compel us to regard the principle as demonstrated, or, to use a legal analogy, as permissive though not mandatory in function. In fact, nature represents a balance among various factors

* In the preface of his book "Germ Plasm" (1893), Prof. A. Weismann wrote: "I finally became convinced that an epigenic development is an impossibility. Moreover, I found an actual proof of the reality of evolution." Epigenesis is the theory that the germ is created by union of the fecundating principles of the male and female, and that it does not pre-exist in the ovum or spermatozoon; and in this quotation the word *evolution* had its original biological meaning, *i.e.*, an unfolding of the germ that pre-exists in the parent, and contains all the parts of the fully formed organism in minute form, somewhat like the homunculus of Paracelsus.

According to Weismann, *idants,* the main divisions of the germ plasm, are made up of *ids;* ids, of *determinants;* determinants, of *"biophores."* A certain portion of the immortal germ plasm is told off to carry on the unaltered heredity, while the rest is sorted out to direct differentiation and development of the somatoplasm. Weismann's view that the germ plasm could not be changed by the somatoplasm was supported by the failure of various experimental attempts to demonstrate the inheritance of acquired characteristics.

† Direct action of environment.

and therefore a compromise of tendencies. Once a *modus operandi* is established, a tendency ceases to be vague even in the many cases where it is not obviously realized. Many demonstrable facts of astronomy do not create a demand for their experimental duplication, because their magnitude transcends our ability to establish conditions that could result, for example, in a solar system or a galaxy. Although the exceedingly tiny atoms, molecules and ultimate biological units are apparently within our grasp, we often lack both the knowledge and the ability to control the infinitude of their structures, relations and behaviors. Furthermore, most experiments in the biological field are judged at the diagnostic level; that is, they involve the reading of end results at higher structural levels rather than observations of the behaviors of the basic units or mechanisms. The behaviors are rationally deduced or mathematically calculated under conventional assumptions.

Definite chemical substances demonstrably produce heritable changes in bionts. With colchicine the changes are attributed to heritable increase in chromosome number. With some carcinogenic substances the heritable changes seem to be at least in part nuclear. In the transformation of bacteria and in tissue cultures of plant and animal cancers, gene changes are possible though not demonstrated.

Knight and Stanley [84] found a marked difference between the amino acid content of ordinary tobacco mosaic virus and the Holmes rib grass strain. Unless it can be shown that still smaller units produce "virus," or that viruses act as catalysts to form other substances than themselves, these figures reflect a chemical difference in the virus substance, which Knight and Stanley consider a mutation. The relation of

	Tyrosine	Tryptophane	Phenylalanine
Ordinary Tobacco Mosaic Virus	3.8	4.5	6.0
Holmes Rib Grass Strain	6.4	3.5	4.3

the activity of tobacco mosaic virus to the presence and cogency of certain chemical groups or structures is shown by the conclusions of Anson and Stanley: [16a] Tobacco mosaic virus whose SH groups have been oxidized beyond the S-S stage by iodine, but whose tyrosine groups have not been converted into di-iodotyrosine groups, still retains its normal biological activity against tobacco plants; but inactivation follows conversion of the tyrosine into di-iodotyrosine. Biological activity fluctuates with chemical changes in biocatalysts.

Considering the interaction of cytoplasm and chromosomes in genetics and in development, Prof. H. S. Jennings [79] points out that these two cellular divisions "are not separate in fact and in substance, though they are often kept isolated in concept—to the great detriment of sound understanding. There is a continual cycle of interchange between them, a transformation of material from one into the other . . . the cytoplasm is not merely passive; it actively determines what shall happen in the chromosomes. . . . Every cell retains all the chromosomal materials, but which of these materials comes into action in each cell depends upon what cytoplasm is there present, as well as upon the conditions under which the reactions occur. . . ." Jennings then lists three classes of heritable modifications in protozoa: (1) degenerate changes induced by unfavorable conditions acting for many generations; (2) acclimatization to high and low temperatures or to injurious concentrations of chemicals (As, Sb, quinine, methylene blue and various organic substances); and (3) alterations in form and structure, not involved in (1) and (2), which may be retained for hundreds of generations.[80a] Referring to the observations of De Garis [42] on *Paramecium caudata* through a long series of asexual generations, Jennings states: "At every succeeding fission the original cytoplasm is diluted to one-half, so that after ten generations it is diluted to less than $\frac{1}{1000}$ part, the remainder being new

cytoplasm produced by growth. Yet after ten generations the nature of the original cytoplasm still has a marked effect on size. The original cytoplasm seemingly must therefore have to some extent the power of reproducing itself in its distinctive nature, at the time that growth occurs. In this respect it partakes of the character of a gene or genetic material, in that it affects the character of the individuals and reproduces itself in some degree true to type. But in time it is made over by the nucleus."

When a race of Paramecium averaging in length 198 μ was mated with another true-breeding race averaging 73 μ, the influence of the cytoplasm was evident in one experiment for 36 generations, for only then did the descendants reach about the same size, although they all had the same chromosomal content. Referring to experiments of Jollos [80a] indicating that such environmental modifications as acclimatization (see above) are inherited for as much as 800 generations, Jennings says: "If they are merely modifications of the cytoplasm one might anticipate that long before so many generations had passed the cytoplasm would have been made over by the nucleus, and its modifications would have disappeared. Yet we do not know that the time required for the nucleus to dominate the cytoplasm would be subject to the same limits in all cases. The question whether inherited environmental modifications are exclusively cytoplasmic, or whether they affect the nucleus, the chromosomes, must be left open for the present."

If persistent heritable nuclear and cytoplasmic changes were the usual, rather than the unusual fact, nature would present a confusion of bionts. And among the relatively few heritable changes of all kinds that actually occur, most are either detrimental or else represent no outstanding advantage to the inheritors. On the other hand, nature is continually conducting a vast multitude of experiments, over endless eons, and under an enormous range of ever-changing conditions. Even where a newly developed individual is of a type possessing the potentiality of persisting and/or dominating under conditions existing at the time of its emergence, it would have but slight chance of surviving and establishing a new strain in the face of established forms and predators. We must remember, however, that each new form of biont established adds to the amount and variety of chemical substances available to other bionts, either through the products of its metabolism or through its own body, living or dead. In general, each higher form of life depends upon simpler forms to supply it with essential molecules, which are molded into its own species-specificity by the analytic and synthetic action of its own catalysts. Thus man depends on simpler forms of life for his supply of the principal food molecules (fats, carbohydrates, proteins) and also for many essential trace substances (vitamins) some of which supply carriers or prosthetic groups for vital catalysts. Therefore food has an important evolutionary aspect [8] and there is truth in the French aphorism: "Rien n'est la proie de la mort; tout est la proie de la vie."

Geological and meteorological factors have also played an important role in evolution, for they determine where and under what conditions bionts must live. Thus it is believed that great outbursts of volcanic activity during the Upper Cretaceous period kept the atmosphere so filled with colloidal dust that glaciation resulted through failure of normal solar heat to reach the earth's surface. Volcanic dust and lava also establish new chemical and physical soil conditions. A subtle, potent, but obscure evolutionary factor lies in the exposure of bionts to unaccustomed chemical elements or compounds by virtue of soil changes, and because of their enforced change of habitat and food. Highly specialized dinosaurs are supposed to have perished upon drying up of the swamps which furnished them with food and shelter; but some bionts (*e.g.*, brachiopods) survived over geological epochs apparently because they lived in relatively stable ocean zones, while others (*e.g.*, ants, cockroaches) seem to have survived because of their adaptability. Great earth changes,

over long periods of time, appear to have left some heritable impress upon surviving bionts.

How can heritable factors, apart from genic or chromosomal changes, enter into and affect the germ cells which carry heredity?

It is reasonable to believe that the cytoplasm of germ cells, like that of other cells, tends to receive an average supply of molecular flotsam and jetsam, that is, of all particulate units which are carried throughout the biont by circulation and diffusion, and that among these molecular or near-molecular units are those that determine, or aid in determining, differentiation. Apart from other evidence, the fact that some differentiated cells can breed true in tissue culture warrants belief that specific tissues or organs contribute to the molecular melée in the organism units capable of determining the formation of their own specific types of cell.* Excessive use of a part or of an organ commonly leads to its hypertrophy, and may thus tend to increase the quantity of hypothetical specific modifier substances arising from such part or organ and reaching the gametes. This would represent a hypothetical physico-chemical mechanism which could in some cases lead to the inheritance of acquired characteristics based on the use or disuse of parts.

Without suggesting a mechanism for their action, J. T. Cunningham [88] advanced a Lamarckian view involving hormones as active factors in heredity. He pointed out that special glands are not essential for the production of these internal secretions, for his colleagues at the University of London, W. M. Bayliss and E. H. Starling, had shown (1902) that the wall of the intestine produces a hormone, secretin, which, carried by the blood, causes the pancreas to secrete. Cunningham stated:[88] "There is nothing improbable in supposing that a tissue stimulated to excessive growth by external irritation would give off special substances to the blood. We know that living tissues give off waste products, and that these are not merely pure CO_2 and H_2O, but complicated compounds. The theory proposed by me in 1908 was that we have within the gonads numerous gametocytes whose chromosomes contain factors corresponding to the different parts of the soma, and that these factors or determinants might be stimulated by waste products circulating in the blood and derived from the parts of the soma corresponding to them. There is no reason to suppose that an exotosis formed on the frontal bone as a result of repeated mechanical stimulation due to the butting of stags would give off a special hormone which was never formed in the body before, but it would probably in its increased growth give off an increased quantity of intermediate waste products of the same kind as the tissues from which it arose gave off before. These products would act as a hormone on the gametocytes, stimulating the factors which in the next generation would control the development of the frontal bone and adjacent tissues. . . . If the factors in the gametes were thus stimulated they would, when they developed in a new individual, produce a slightly increased development of the part which was hypertrophied in the parent soma. No matter how slight the degree of heredity effect, if the stimulation was repeated in every generation . . . the hereditary effect would constantly increase until it was far greater than the direct effect of the stimulation."

Starling formed the word "hormone" from a Greek root meaning "to stir up" or "excite," and Cunningham uses the term in its original meaning; but in speaking of "endocrines," produced by "glands of internal secretion," the original meaning is limited. Certain substances, ordinarily regarded as waste products of cellular metabolism, are known to affect the activities of organs, and have been called *parahormones*. Thus CO_2 activates the respiratory center of the medulla and is administered with oxygen to avoid respiratory collapse; and the normal heart beat of some lower vertebrates (*e.g.*, sharks) seems to depend upon a certain concentration of urea. The potency of some hormones may be gauged from the finding of J. J. Abel, about twenty years ago, that a purified posterior pituitary hormone is effective in a

* Cells grown in tissue culture often de-differentiate in the course of time.

concentration of 1 part in 18,750,000,000.* Though much is known as to the *effects* produced by hormones, little is known as to the basic *mechanism* whereby these effects are produced. It is suggested that the principle of catalyst modification (including the formation of new catalysts) offers a simple, reasonable, and probable explanation in many cases.

Experimental demonstration of the evolutionary aspect of this principle might appear if tissue cultures of undifferentiated cells could be made to become heritably differentiated by specific substances or extracts.† But past experimental failures do not constitute disproof. The factors involved in most biological processes, though individually simple after discovery, are generally so numerous and obscure that it is very difficult to duplicate all of them in one experiment. Thus many viruses which will not affect adult fowl multiply on the chorio-allantoic membrane of the developing hen's egg. Among those that will *not* grow there are the viruses of poliomyelitis and of foot-and-mouth disease. The culture of orchids from seed became possible after it was found that in nature a symbiotic fungus surrounds the seed with special moisture and other conditions. Lichens commonly found growing on rocks represent a symbiotic system, where a green unicellular alga supplies carbohydrate to a fungus that takes up salts and conserves moisture. Termites would starve to death on their diet of wood, were it not for a parasitic endameba which solubilizes the carbohydrates, as Cleveland found by subjecting termites to a slightly increased oxygen pressure which killed the ameba without injuring the termite itself. Some bionts (*e.g.*, viruses) are termed "obligate" parasites only because we have not yet discovered the conditions essential to their growth and cannot, therefore, yet raise them *in vitro*. Recognition of obscure conditions and of the importance of trace substances or "impurities" is aiding progress in the culture of special forms of life, as well as in an understanding of biological mechanisms.

A Hair perhaps divides the False and True;
Yes; and a single Alif were the clue—
Could you but find it—to the Treasure-house,
And peradventure to THE MASTER too.

References

1. Abbott, Lynn, D. F., Jr., *Science*, **94**, 365 (1941).
2. Adkins, H., "The Role of the Catalyst," *Ind. Eng. Chem.*, **32**, 1189 (1940).
3. Adkins, H., Richards, L. M., and Davis, J. W., *J. Am. Chem. Soc.*, **63**, 1320-5 (1941).
4. Adrian, Sir Edgar, "The Mechanism of Nervous Action," 1932.
5. Aguilhon and Sazerac, *Compt. Rend.*, **155**, 1187 (1912).
6. Alexander, J., *J. Am. Chem. Soc.*, **43**, 434 (1920); also Vol. I, this series.

6a. ——, *J. Chem. Education*, **9**, 2026 (1932).

7. ——, First Colloid Symposium Monograph, 1923; "Colloid Chemistry," 4th ed., Van Nostrand Co., 1937.
8. ——, *J. Heredity*, **27**, 139-150 (1936); *Protoplasma*, **14**, 296 (1931); *Biodynamica*, No. 54 (Dec., 1939); *Scientia*, **53**, 165 (1933); *Arch. di Sci. Biol.*, **14**, 409 (1934); *Science*, **96**, 252 and 446 (1942).
9. ——, "Selective Adsorption and Differential Diffusion," *J. Am. Chem. Soc.*, **39**, 84-8 (1917).
10. ——, *Science*, **80**, 79 (1934); **81**, 44 (1935).
11. ——, *J. Soc. Chem. Ind.*, **55**, 206-9 (1936); "Colloid Chemistry," 4th ed.
12. ——, "Colloid Chemistry," Van Nostrand Co., 3rd ed., 1924; 4th ed., 1937; also references mentioned under (8).
13. Alexander, J. and Bridges, C. B., Vol. II of this series, pp. 9-58.
14. Alexander-Jackson, E. G., *Rev. Tuberculosis*, **33**, 767-97 (1936).
15. Allen, E., Smith, G. M., and Gardner, W. V., *Am. J. Anatomy*, **61**, 321-9 (1937).
16. Allen, T. H. and Bodine, J. H., *Science*, **94**, 443-4 (1941).

16a. Anson, M. L. and Stanley, W. M., *J. Gen. Physiol.*, **24**, 679-90 (1941).

* In the Annual Report of The Rockefeller Foundation for 1941, p. 39, the following is stated about biotin: "In 1901 at the University of Louvain it was 'bios'; in 1920 in England it was 'protective factor X'; in 1931 in Hungary it was 'vitamin H'; but in 1936 in Utrecht, Holland, the organic chemist Fritz Kögl isolated a few crystals of the vitamin, and found its potency so great that one part in 400 billion would influence the growth of yeast. Professor Kögl named it 'biotin,' and this is the accepted name today." Its structure has just been established by V. du Vigneaud. *Science*, **96**, 455 (1942), and is given elsewhere in this volume.

† Something analogous happens in chemically induced cancer.

17. Arnow, L. E., *Science*, **86**, 76 (1937).
18. Auchter, E. C., *Science*, **89**, 421-7 (1938).
19. Barnes, H. T., Vol. I, this series, pp. 435-43.
20. Bell, D. J., "Introduction to Carbohydrate Chemistry," London, 1940.
21. Berzelius, J., *Jahresberichte*, **15**, 237 (1836).
22. Blakeslee, A. F., *Compt. Rend.*, **205**, 476-9 (1937).
22a. Blum, H. F., "Photodynamic Action and Diseases Caused by Light," Reinhold Pub. Corp., 1941.
23. Bodine, J. H., *J. Gen. Physiol.*, 24, 99-103, 423-32 (1941); *Am. Naturalist*, **75**, 97-106 (1941).
24. Boedecker, von, *Annalen*, **117**, 98 (1861).
25. Boell, E. J., Needham, J., and Rogers, U., *Proc. Roy. Soc. London (B)*, **127**, 374-87 (1939).
26. Bogert, C. M., *Ann. N. Y. Acad. Sci.*, **41**, 331-42 (1941).
27. Boveri, "Zur Frage der Entstehung maligner Tumoren," Jena, 1914.
28. Brachet, J., *Arch. Biol.*, **50**, 233, 267 (1939).
29. Bredig, G., *Biochem. Z.*, **250**, 414 (1932).
30. Bridges, C. B., *J. Heredity*, Feb., 1935.
31. Brues, A. M. and Cohen, A., *Biochem. J.*, **30**, 1363-8 (1936).
32. Burnet, F. M. and Lush, D., *Australian J. Exptl. Biol. Med.*, **14**, 27-37 (1936).
33. Carpenter, G. H., *Encyclop. Brit.* (11th ed.), **2**, 86 (1916).
34. Church, *Trans. Roy. Soc.*, 627-36 (1869).
34a. Conklin, E. G., "Heredity and Environment," Princeton Univ. Press, 1919; also in H. H. Newman, "Evolution, Genetics & Eugenics," p. 464, Univ. of Chicago Press, 1927.
35. Cook, J. W., Hewett, C. L. and Hieger, I., *J. Chem. Soc.*, **1933**, 395.
36. Cook, A. H., *J. Chem. Soc.*, **1938**, 1761-80, 1845-7.
36a. Cox, Herald R., *Science*, **94**, 399-403 (1941).
37. Croft-Hill, A., *J. Chem. Soc.*, **73**, 634 (1898).
38. Cunningham, J. T., "Hormones and Heredity: A Discussion of the Evolution of Adaptations and the Evolution of Species," Constable & Co., London, 1921.
39. Dakin, H. D., *J. Biol. Chem.*, **9**, 151 (1914).
40. Davenport, C. B., *Science*, **95**, 99-100 (1942).
41. Dawson, M. H. and Sia, R. H. P., *J. Exptl. Med.*, **54**, 681-99 (1931).
41a. Dean, H. T., Arnold, F. A. and Elvove, E., *Pub. Health Repts.*, **57**, 1155-79 (1942); H. T. Dean, *J. Am. Water Wks. Assn.* **35**, 1161-86 (1943).
42. De Garis, G. F., *J. Exptl. Zoöl.*, **71**, 209 (1935).
43. Deinert, F., *Ann. Inst. Pasteur*, **14**, 139-89 (1900).
44. Demerec, M., Carnegie Inst. of Washington, Year Book No. 38, 185-91 (1938-9); No. 39, 211-7 (1939-40).
44a. Dienes, L., *J. Bacter.*, **44**, 37-72 (1942).
45. Dobzhansky, T., "Genetics and the Origin of Species," Columbia Univ. Press, 2nd ed., p. 46, 1941.
46. Dubinin, N. P. and Sideroff, B. N., *Biol. Zh. (Moscow)*, **4**, 555-68 (1935).
47. Dustin, A. P., *Bull. Acad. roy. med. Belg.*, **14**, 487-8 (1934).
48. Dustman, R. B. and Duncan, I. J., *Plant Physiol.*, **15**, 343-8 (1940); *Science*, **90**, 233 (1939).
49. Einstein, A., Vol. I, this series, pp. 323-339.
50. Emerson, R. and Arnold, W., *J. Gen. Physiol.*, **16**, 191 (1930).
51. Emmons, C. W. and Hollaender, A., *Am. J. Bot.*, **26**, 467-75 (1939).
52. Endres, H., Vol. I, this series, p. 808-13.
53. Erikson-Quesnel, I.-B. and Svedberg, T., *Biol. Bull.*, **71**, 498 (1936).
53a. Eyster, W. H., *Science*, **94**, 144-5 (1941).
54. Fieser, L. F., "The Chemistry of Natural Products Related to Phenanthrene" and Supplement, Reinhold Pub. Corp., N. Y., 1936.
55. Filmer, J. F. and Underwood, E. J., *Australian Vet. J.*, **10**, 83 (1934); **11**, 89 (1935).
56. Fischer, E. and Thierfelder, A., *Berichte*, **27**, 2031-7 (1894).
56a. Fulton, J. F. and Nachmansohn, D., *Science*, **97**, 569-71 (1943).
57. Glaser, O. C., *Anat. Rec.*, **8**, 525-51 (1914).
58. Goranson, R. W. and Zisman, W. A., *J. Chem. Phys.*, **7**, 492-505 (1939).
59. Graham, T., *Proc. Roy. Soc. London*, June 16th, 1864.
60. Green, David H., "Mechanism of Biological Oxidations," Cambridge Univ. Press, 1940.
60a. Green, Robert G., *Science*, **95**, 602-3 (1942).
61. Griffith, F. G., *J. Hyg.*, **27**, 113 (1928).
62. Hadley, Philip B., *J. Infect. Dis.*, **40** (1), 1-312 (1927); **60** (2), 129-256 (1937); *Proc. Inst. Med. (Chicago)*, **9** (4), 1-22 (1932); *J. Infect. Dis.*, **65**, 267-272 (1939); also Chapter on Bacterial Variability in "Pathogenic Organisms," Park and Williams, N. Y., 1939. The original expression was "microbic dissociation," now commonly called "dissociative variation."
63. Hanes, C. S., *Nature*, **145**, 348 (1940).
64. Harrison, R. G., "The Cell and Protoplasm," A. A. A. S. Pub. No. 14, 1940. The paragraphs quoted are not in the same order as in the original, which should be consulted for further information, as should the other papers in this interesting and authoritative symposium.
65. Hart, E. B., Steenbock, H., Waddell, J. and Elvehjem, C., *J. Biol. Chem.*, **77**, 797-812 (1928).
66. Havas, L., *Nature*, **139**, 371-2 (1937).
67. Haurowitz, F. and Schwerin, P., *Enzymologica*, **9**, 193-7 (1941).
68. Heatley, N. G. and Lindahl, P. E., *Proc. Roy. Soc. London (B)*, **122**, 395-402 (1937).
68a. Hehre, E. J., *Science*, **93**, 237 (1941); Hehre, E. J. and Sugg, J. Y., *J. Exptl. Med.*, **75**, 339-53 (1942).
69. Heidelberger, M. and Avery, O. T., *J. Exptl. Med.*, **38**, 81 (1923).
69a. Hendricks, S. B. and Hill, W. L., *Science*, **96**, 255 (1942).
70. Henze, M., *Z. Physiol. Chem.*, **79**, 215 (1913).
71. Herbst, C., *Arch. f. Entwickelungsm. der Org.*, **9**, No. 3 (1900).
72. Hester, J. B., *Science*, **93**, 400-1 (1941).
73. Hieger, I., *Biochem. J.*, **24**, 505 (1930).
74. Hill, A. Croft, see Ref. 37.
74a. Hitchcock, A. E. and Zimmerman, P. W., *Proc. Am. Soc. Hortic. Sci.*, **38**, 104-110 (1940).
75. Holtfreter, J., *Roux' Arch.*, **117**, 421-510 (1929); **124**, 404-66 (1931); **128**, 584-633; 766-70 (1933); **132**, 225-306 (1934); **138**, 522-656, 637-738 (1938).
76. Ipatieff, V. N., *Science*, **91**, 605-10 (1940). Willard Gibbs Medal Address entitled: "Mixed Catalysts."
77. Jackson, E. A., see Ref. 14.
78. Jennings, H. S., "Genetics," N. Y., 1935.
79. ——, "The Cell and Protoplasm," Pub. No. 14, A. A. A. S., 1940, pp. 44-55.
80. Johnston, J. F. W., "The Chemistry of Common Life," p. 489, London, 1879.
80a. Jollos, V., *Arch. Prot.*, **43**, 1 (1921); **83**, 197 (1934).
81. Karström, H., Helsingfors Univ. Doctor's Thesis, 1930, "Ueber Enzymbildung in Bacterien."

82. Keeler and Trimble, *J. Heredity*, **51**, Feb. 1940.
83. Keilin, D. and Mann, T., *Proc. Roy. Soc. London (B)*, **122**, 119-39 (1937).
84. Knight, C. A. and Stanley, W. M., *J. Biol. Chem.*, **141**, 39-49 (1941).
85. Koltzoff, N. K., *Science*, **80**, Oct. 5th, 1934.
86. Kuhn, R., *Réun. Int. de Physique*, VII, 1, Paris, 1938. ("Relations chimiques entre Vitamines et Ferments.")
87. Kuhn, R., Moewus, F. and Jerchel, D., *Berichte*, **71**, 1541 (1938).
88. Lankester, E. R., "Notes on Embryology and Classification," *Quart. J. Micro. Sci.*, **17**, 399-454 (1877).
88a. Lazarow, A., *Science*, **95**, 49 (1942).
89. Lehmann, F. E., *Roux' Arch.*, **108**, 243-282 (1926); **113**, 123-71 (1928); **138**, 106-59 (1938).
89a. Lindegren, C. C. and Bridges, C. B., *Science*, **87**, 510-511 (1938).
90. Lindegren, C. C. and Lindegren, G., *J. Heredity*, **32**, 435-40 (1941).
91. Lits, F. J., *Compt. Rend. Soc. Biol.*, **115**, 1421-3 (1934).
92. Loehwing, W. F., *Science*, **90**, 553 (1939).
92a. Loewi, Otto, "Uebertragbarkeit der Herznervenwirkung," *Pflueger's Arkiv.*, **189**, 239 (1921).
93. Ludford, R. J., *Arch. expt. Zellf.*, **18**, 411-41 (1936).
93a. Mac Arthur, J. W., "Genetics of Quintuplets," *J. Heredity*, **29**, 323-29 (1938).
93b. Manwaring, W. H., *Science*, **79**, 466-470 (1934).
94. Marrack, J. R., "The Chemistry of Antigens and Antibodies," 2nd ed., Brit. Med. Res. Council, London, 1939.
95. Mendel, L. B., *Science*, **78**, 319 (1933).
95a. Miall, S., *Chemistry and Industry*, **51**, 806 (1932).
96. Morgan, Sir G. T., Hardy, D. V. N. and Procter, R. A., *J. Soc. Chem. Ind.*, **52**, 1T (1932).
97. Morgan, Sir G. T., *Bull. Soc. chim. Belg.*, **45**, Apr. 19th, 1936.
97a. Moxon, A. L., S. Dakota Agr. Expt. Sta. Bull. No. 311, 1-91 (1937); S. F. Trelease, *The Scientific Monthly*, **54**, 12-28 (1942).
98. Muller, H. J., *Cold Spring Harbor Symp.*, **9**, 151-65; 290-308 (1941).
98a. Nachmansohn, D. and Meyerhof, O., *J. Neurophysiol.*, **4**, 348 (1941); Nachmansohn, Coates, and Cox, *J. Gen. Physiol.*, **25**, 75 (1941).
99. Neuberg, C., "Handbuch der Biochemie" and Supplements, 1925-1933; *Am. Brewer*, **75**, 22-30 (1942).
100. Orth, O. S., Wickwire, G. C. and Burge, W. E., *Science*, **79**, 33-4 (1934).
100a. Osborn, Henry Fairfield, *Scientific Monthly*, **22**, 185-192 (1926).
101. Overbeek, J. van, *J. Heredity*, **29**, 339-41 (1938).
102. Panshin, I. B., *C. R. (Dokl.) Acad. Sci. U. S. S. R.*, n. s. 1, (10), 83-6 (1936).
103. Parker, G. H., "Neurohumors: Novel Agents in the Action of the Nervous System," *Science*, **81**, 279-83 (1935).
104. ——, *Proc. Am. Phil. Soc.*, **75**, 1-10 (1935); *Cold Spring Harbor Symp.*, **4**, 358-70 (1936).
105. ——, *Proc. Nat. Acad. Sci.*, (1940); *Science*, **91**, 456 (1940).
106. Pauling, L. and Delbrück, M., *Science*, **92**, 77-91 (1940).
107. ——, *J. Am. Chem. Soc.*, **62**, 2643-57 (1940).
107a. ——, and Campbell, Dan H., *Science*, **95**, 440 (1942).
108. Pearson, P. B. and Burgin, C. J., *Proc. Soc. Exptl. Biol. Med.*, **48**, 415-7 (1941).
109. Pedersen, K. O., *Nature*, **138**, 363 (1936); *Skand. Arch. Physiol.*, **77**, 67 (1937); *Compt. rend. trans. lab. Carlsberg, Ser. Chim.*, **22**, 427 (1938).
109a. Pike, E. W. and Pike, F. H., *Science*, **94**, 325 (1941).
110. Power, F. W., *Proc. Soc. Exptl. Biol. Med.*, **33**, 598 (1936).
111. Plunkett, C. R., "Outlines of Modern Biology," N. Y., 1930.
111a. Presley, J. T., *J. Heredity*, Dec., **25**, 485 (1934).
112. Prytherch, H. F., *Science*, **73**, 429 (1931); *Ecolog. Monographs*, **4**, 47-107 (1934).
113. Rahn, O., "On the Nature of the Adaptive Enzymes," *Growth*, **2**, 363-7 (1938).
114. Raper, H. S., *Physiol. Rev.*, **8**, 253 (1928).
115. Rhoades, H. E., *J. Bact.*, **42**, 99-115 (1941).
116. Riehl, N., *Naturwissen.*, **28**, 601 (1940).
117. Schwartz, L., Oliver, E. A. and Warren, L. H., *U. S. Public Health Reports*, **55**, 1111-30 (1940) (Reprint No. 2173).
118. Sherwin, with Thierfelder, *Berichte*, **47**, 2630 (1914).
119. Sia, R. H. P. and Dawson, M. H., *J. Expt. Med.*, **54**, 101-10 (1931).
120. Smith, E. F., Brown, N. A. and McCulloch, U. S., *U. S. Dept. Agr. Bull. 255* (1912).
121. Smith, E. L., *Science*, **88**, 170 (1938); **91**, 199 (1940); *Proc. Am. Physiol. Soc.*, **129**, 466 (1940); *J. Gen. Physiol.*, **24**, 565-82, 583-96 (1941).
122. Smith, W. H. and Saylor, C. P., *J. Res. Nat. Bur. Stand., Res. Paper No. 719* (1934).
123. Snyder, L. H., "Medical Genetics," Duke Univ. Press, Durham, N. C., 1941.
124. Spemann, H., *Verh. Deutsch. Zool. Ges.*, 129-32 (1931).
124a. Sonneborn, T. M., "Sexuality in Unicellular Organisms," Chapter 14 in "Protozoa in Biological Research," by Gary N. Calkins and Summers, F. M., pp. 666-709 (1941).
125. Spemann, H., *Roux' Arch.*, **123**, 389-517 (1931).
126. Screenivasaya, M. and Pirie, M. W., *Biochem. J.*, **32**, 1707 (1938).
127. Steinberg, R. A. and Thom, C., *J. Heredity*, **31**, 61-3 (1940); *Proc. Nat. Acad. Sci.*, **26**, 363-6 (1940); **25**, 329-35 (1939).
128. Stewart, G. W., *Trans. Farad. Soc.*, **29**, 990 (1933); *J. Phys. Chem.*, **4**, 231-6 (1936); see also G. Friedel in Vol. I, this series.
129. Sumner, F. B., *Biol. Rev.*, **51**, 351-75 (1940).
130. Svedberg, T., *Ind. Eng. Chem., Anal. Ed.*, **10**, 113-29 (1938).
131. Sylvestri, F., Harvard Tercentenary Address, 1936.
132. Szent-Györgyi, A., "Towards a New Biology," Korányi Memorial Andress, Budapest; *Science*, **93**, 609-11 (1941).
133. Tatum, E. L. and Beadle, G. W., *J. Gen. Physiol.*, **22**, 239 (1938).
134. Tauber, H., *J. Am. Chem. Soc.*, **60**, 2263 (1938).
135. Theorell, H., "Structure and Function of Some Enzymes," A. A. A. S. Pub. No. 14, 136-46.
135a. Topley, W. W. C., Wilson, J. and Duncan, J. T., *Brit. J. Exp. Path.*, **16**, 116 (1935).
135b. Torda, Clara, and Wolff, H. G., *Science*, **98**, 224 (1943).
136. Townsend, G. F. and Lucas, C. C., *Biochem. J.*, **34**, 155-62 (1940).
137. Van Niel, C. B., "Biochemistry of Microorganisms," A. A. A. S. Pub. No. 14, pp. 106-119 (1940).
138. Van Slyke, D. D., *Science*, **95**, 259-63 (1942).
139. Verzär, F., *Nutrition Abst. Rev.*, **2**, 441 (1933); also Ref. 12.
140. Vessie, P. R., *J. Nerv. Mental Dis.*, **76**, 553-73 (1932).
141. Waddington, C. H., *Proc. Roy. Soc. London (B)*, **125**, 365-72 (1938).
142. Waddington, C. H., Needham, J., Nowinski, W. W. and Lemberg, R. L., *Proc. Roy. Soc. London (B)*, **117**, 289-310 (1935).

143. Warburg, O. and Christian, W., *Biochem. Z.*, **298**, 150-68, 368-77; **139**, **301**, 221-2 (1938).
144. Weimarn, P. P. von, Vol. I of this series, pp. 27-101.
145. White, P. R. and Braun, A. C., *Science*, **94**, 239-41 (1941); *Cancer Research*, **2**, 597-617 (1942). See also Martin, J. P., *Science*, **96**, 39 (1942).
145a. Walker, *Proc. Roy. Soc. Med.*, **28**, 759 (1935).
146. Willstätter, R. and Rhodewald, G., *Z. Physiol. Chem.*, **225**, 103-24 (1934).
147. Woerdeman, N. W., *Proc. Acad. Sci. Amsterdam*, **36**, 189-93 (1933).
148. Wolkow, M. and Baumann, E., *Z. physiol. Chem.*, **15**, 228 (1891).
149. Wooley, D. W., *J. Biol. Chem.*, **136**, 113 (1940); *Science*, **92**, 384-5 (1940).
150. Yamagawa and Ichikawa, *Mitteil. med. Fakultät kaiser. Univ. Tokyo*, **15**, 295 (1915).
151. Yasuda, S., *Proc. Crop Sci. Soc. Japan*, **2** (2), 122-6 (1930).
152. ——, *Bot. Mag. (Tokyo)*, **46**, (548), 510-7 (1932).
153. Zsigmondy, R., "Colloids and the Ultramicroscope," Eng. trans. by J. Alexander, N. Y. 1909.
The reader is referred to three papers by the late Prof. Leonard T. Troland (Harvard), from which extensive quotations were given by J. Alexander and C. B. Bridges in the first paper of Vol. II of this series, pp. 18-21. These papers are:
"The Chemical Origin and Regulation of Life," Monist, Jan., 1914.
"The Enzyme Theory of Life," Cleveland Medical Journal, **15**, 377-89 (1916).
"Biological Enigmas and the Theory of Enzyme Action," Am. Nat., 1917, **15**, 321-50.
The paper in Vol. II also contains two paragraphs quoted from an unpublished Master's Thesis by J. Alexander, entitled: "The Importance and Trend of Recent Work on the Chemistry of Life and the Products of Life," submitted early in 1899.
Recently, Prof. Alwin Mittasch (Heidelberg), who did much of the chemical and trace-substance work on the Haber process catalysts (referred to in the quotation from Bernthsen in this paper), published three booklets (Julius Springer, Berlin):
"Ueber katalytische Verursachen in biologischen Geschehen" (1935).
"Ueber Katalyse und Katalysatoren in Chemie und Biologie" (1936).
"Katalyse und Determinisimus, Ein Beitrag zur Philosophie der Chemie" (1938).
The views expressed by Dr. Mittasch seem consonant with those here advanced. In a private communication (March, 1938), following an exchange of publications, he stated:
"Die 'Modifikation' von Katalysatoren wird ebenso wie die 'Autokatalyse' in der Wissenschaft immermehr Bedeutung erlangen. Wenn man aus Erfahrung weiss, was *kleine Aenderungen an Katalysatoren* für ein organisches Substrat zu bedeuten haben, so ist klar, dass 'Mutationen' an Katalysatoren für Erkrankungen, Artwandel, u.s.w. von ganz wesentlichem Einfluss sein können. Die Aufgabe für den Wissenschaftler ist dabei immer zweifach:
(1) Wie *entsteht* die Aenderung von Katalysatoren?
(2) Welche *Wirkungen* hat das für das Substrat?
Die Physiologen werden darauf sehr zu achten haben."

Photosynthesis

PAUL ROTHEMUND

The C. F. Kettering Foundation for the Study of Chlorophyll and Photosynthesis, Antioch College, Yellow Springs, Ohio

The term *photosynthesis* refers, in general, to the synthesis of chemical compounds effected with the aid of radiant energy, especially in the visible region of the spectrum; more specifically the name is applied to the formation of complex organic substances in the tissues of plants exposed to light. Usually the statement is made that the green plant is the converter of the supply of solar energy into the chemical energy of the structural material of the plants on earth. This is true for a relatively large number of compounds produced; but some lower plants, which do not contain chlorophyll, also show photosynthetic activity. Other lower plants effect strictly chemical syntheses, without the necessity of being exposed to light.

In the following discussion an attempt has been made to summarize and coordinate recent publications on photosynthesis, omitting, however, numerous tentative explanations as well as papers of purely speculative character; it is hoped that within the limitations of space the essence of our present knowledge is adequately presented. Chemical formulas for the chlorophylls *a* and *b*, and for other pigments possibly involved in the process, and a discussion of the reactions of these colored compounds are also not included; instead, the reader is referred to recent reviews on the chemistry of chlorophyll.[1-8] The term *chlorophyll* will be used for the green coloring matter, extractable from green parts of plants as a mixture of the two pigments chlorophyll *a* and *b* (usually in the proportion $a:b = 3:1$).

The many interpretations proposed to elucidate the mechanism of the process may be classified into four groups:

(1) Chemosynthesis (chlorophyll and light not necessary)
(2) Photosynthesis in the absence of chlorophyll (other pigments in the cells may be necessary)
(3) Photosynthesis in the presence of chlorophyll, the latter acting as a sensitizer, as a light filter, or as an absorber and transmitter of radiant energy.
(4) Photosynthesis with chlorophyll participating in one or more chemical reactions. In this case chlorophyll may react either as such, or in a complex compound consisting of chlorophyll as the prosthetic group on a large molecule as carrier; the carrier is supposedly not directly engaged in the reaction.

Reaction mechanisms for each of these four types of fundamental concepts are given in the more recent books on photosynthesis [4-9] and in the "Annual Review of Biochemistry." It should be remembered that many conclusions drawn from biological experiments hold for only one family of plants, or even for only one species; at present investigators more frequently select pure cultures of green algae or of bacteria for their work, while leaves or other chlorophyll-bearing parts of higher plants are used to a much lesser extent. Thus far no generalization on the photosynthetic behavior of these different plants can be made, although many interesting and important individual facts have been recognized. Thus it has been found that a number of algae may be trained by varying the conditions of growth; the qualitative and quantitative results of photosynthetic performance depend on such previous training.

The question as to the degree of efficiency attained by the plant in the use of energy from the sun for the reduction of carbon dioxide seemed to have been definitely answered by Warburg and Negelein.[10-12] These authors used monochromatic radiation on *Chlorella* (sp. ?; see footnote to Table 2) and found the following quantum efficiency:

Table 1

Wave Length ($m\mu$)	Spectrum Region	Corresp. Energy (Cal/Mole)	Quanta required per mole CO_2	Quantum Yield (Mole CO_2 per quantum)
660	Red	43,000	4.4	0.23
578	Yellow	49,200	4.3	0.23
436	Blue	65,100	5.1	0.20

The value for the quantum yield (or quantum efficiency) for the red region

$$\frac{\text{decomposed } CO_2}{\text{quanta per mole}} = \frac{1 \text{ mole } CO_2}{4} = 0.25$$

seemed to explain the basic reaction

$$CO_2 + 4H = CH_2O + H_2O$$

so well that a large number of authors accepted Warburg's value of 0.25 mole carbon dioxide per quantum as "classical" * and based their own calculations on it. Reinvestigations of Warburg's work in different laboratories established the fact that the values for the quantum yield deviate considerably from the value of 4; a tabulation of some of the results of different authors is given on p. 602.

The table shows that at present no reliable numerical statement for the quantum yield in photosynthesis can be made. Even an evaluation of the individual data for

* This epithet has been attached by some investigators in the field of chlorophyll and photosynthesis to the concept of 4 quanta, and to a number of other statements in the older literature. Since it has meaning only as of a specific period of time, and is of necessity arbitrary, its use has always been confusing rather than defining, and should, therefore, be discontinued.

comparison is not permissible, since the limits of error for the measurements by the different authors are not in every case known, and since technique and method of computation vary also to some extent. In spite of this wide discrepancy of data, some investigators believe it justifiable temporarily to assume 12 quanta as the most probable value.

Table 2. Determination of the Quantum Yield in Photosynthesis

Author	Year	Reference	Organism	Quantum Yield
Warburg, Negelein	1922-3	10-12	*Chlorella vulgaris* *	4-5
French	1936-7	13	*Streptococcus varians*	4
"		14	"	4-16
"	1937-8	15	*Spirillum rubrum*	17
Manning, Stauffer, Duggar, Daniels	1935-8	16	*Chlorella pyrenoidosa* (low light intensity)	20
Manning, Juday, Wolf	1935-8	17	*Chlorella pyrenoidosa* (high light intensity)	20
Eymers, Wassink	1938	18	*Chromatium*	12-18
Rieke	1939	19	*Chlorella pyrenoidosa*	4.2
Petering, Duggar, Daniels	1939	20	"	10-25
Magee, DeWitt, Smith, Daniels	1939	21	"	12.5
Eichhoff	1939	22	"	4
Eichhoff, Noddack	1939	23	"	4
Emerson, Lewis	1940	24	"	10-11
Rieke	unpublished	25	*Chlorella Scenedesmus*	11-14

* Although Warburg states specifically that his values were obtained with *Chlorella vulgaris*, Emerson and Green [26] conclude, on the basis of their taxonomical and physiological studies of *Chlorella vulgaris* and *Chlorella pyrenoidosa*, that Warburg's material must have been *Chlorella pyrenoidosa*.

Photosynthesis in the plant involves at least two steps:

(a) the photochemical reaction, for which the chemical effect produced by the unit of absorbed radiant energy is characteristic,

(b) a chemical reaction, often also called "dark reaction," "thermal reaction," or "Blackman reaction," [27] which precedes or follows the photochemical reaction, and is entirely independent of irradiation.

The duration of the Blackman reaction is obviously a limiting factor for the photochemical part of photosynthesis; on the green alga *Chlorella pyrenoidosa* Emerson and Arnold [28, 29] determined the time required for this reaction as 0.02 second at 25°, when light flashes of 10^{-5} second duration were alternating with dark periods. The same authors calculated that for each molecule of carbon dioxide reduced per flash of light approximately 2,500 molecules of chlorophyll are present. In following up this work Arnold and Kohn [30] found that in six species of plants, representing four phyla, the minimum number of chlorophyll molecules available for the reduction of one molecule of carbon dioxide is about 2,000. This result suggested the existence of a "chlorophyll unit," and gave rise to the hypothesis of a photosynthetic unit. This hypothesis aims to show that through a cooperative effort of a large number of chlorophyll molecules enough light is absorbed at any given time by the unit to reduce one molecule of carbon dioxide..

The assumptions concerning the nature of this unit vary considerably. Gaffron and Wohl [81] calculated (on the basis of Warburg's photosynthetic yield of 4 quanta) the number of chlorophyll molecules actually involved in the process of carbon dioxide reduction and found a value of 1,000. They postulated, however, that the plant may be able to adjust itself to the conditions of illumination by varying the number of chlorophyll molecules in the unit. Shortly afterwards Wohl [82] changed this picture by introducing the concept of an energy carrier. The individual chlorophyll molecules transfer radiant energy to other kinds of molecules ("energy carriers") and the latter carry it to reduction centers. The number of such centers is supposed to be several thousand times smaller than that of chlorophyll molecules.

Van Niel [83] considers photosynthesis by green plants "a complex of photochemical and dark reactions in which the former consists of a photodecomposition of water," (as hydrogen donor) "with the aid of chlorophyll and enzymes of unknown nature." A combination of Wohl's and Van Niel's ideas can be found in a recent paper by Ornstein and co-workers.[84]

The assumptions of a photosynthetic unit rendered it necessary to determine whether the known facts on photosynthesis could be satisfactorily explained without resorting to the concept of such a unit. Here the work of Franck and co-workers must be mentioned. Franck and Gaffron [85] replaced their former theories on photosynthesis by a new one, which assumes a number of coupled reactions, involving the change of chlorophyll to monodehydrochlorophyll, and the formation of stable and unstable reaction intermediates by means of hypothetical catalysts "A," "B," and "C." Some of the intermediate products, supposedly of peroxide nature, furnish the oxygen evolved during the reaction, and at the same time restore the hydrogen necessary for the regeneration of chlorophyll. This theory incorporates portions of the concepts by Fischer, Kautsky, Richter, Stoll and Willstätter, and takes into consideration such well-established experiments as the change of fluorescence and the behavior of "poisons" in the process of photosynthesis. But the authors are well aware of the shortcomings of their attempted explanations, especially since the preparation and isolation of the assumed reaction intermediates is completely lacking.

It is important to keep in mind that thus far all experimental attempts to perform photosynthesis *in vitro* have been failures. This fact has often been used as an indication that the special structure of the chloroplast may be the decisive factor in the assimilation of carbon in plants. The comparison of chlorophyll with the hemoglobin in mammalian blood suggests itself: chlorophyll may be present in the plant as a protein-chlorophyll complex compound similar to hemoglobin, which contains 96 per cent of the histone globin, and 4 per cent of the prosthetic pigment group. The application of this concept to the photosynthetic unit renders photosynthesis a problem in colloidal chemistry. The photosynthetic unit would then be a macromolecule made up of a large number of protein molecules to which chlorophyll molecules have been added. Several experiments seem to indicate the existence of such a protein-chlorophyll complex.

As early as 1912 Herlitzka [86] prepared water suspensions of spinach cell material without change of the spectral properties of the green pigment; similar suspensions were reported by Wurmser,[87] who obtained them from ground cultures of purple bacteria. These suspensions probably contained a portion of the protein and of the pigment in form of particles of colloidal size. The presence of a protein-chlorophyll complex in the green plant was suggested by Lubimenko,[88] and later by Stoll.[89]

For compounds formed through action of residual affinities from high molecular substances or complex compounds Willstätter [40] suggested the term *symplex*. This is supposed to include systems consisting of a high molecular carrier and a prosthetic group, as well as systems composed of several components of high molecular weight; it is applicable also to organic substances to which an inorganic compound is linked. Stoll assumes that the chlorophyll in the chloroplasts is linked to a colloidal carrier

to form a highly active symplex. In analogy to hemoglobin the name "chloroplastin" was suggested for this symplex. Chloroplastin has not been isolated from the plant so far in spite of numerous attempts in this direction. Stoll believes that chlorophyll owes its activity in photosynthesis to the linkage to a certain colloid, and that the chloroplastin symplex thus formed becomes the "assimilatory enzyme" of the plant. The long branched hydrocarbon chain of phytol accounts for the solubility in lipoids, and may anchor the chlorophyll to fats or related substances.

Since the chloroplastin hypothesis touches on the problem of the condition of chlorophyll in the leaf, a few studies on this subject may be mentioned. Baas-Becking [41] says "The theory is proposed that chlorophyll is present in small particles in the periphery of the plastid, adsorbed, on its lipophilic side, on lecithin, and on its hydrophilic side, on protein." In searching for a technique to prepare photochemically active water solutions of the cell contents, French [42] found that the cells of purple bacteria will break up in a few minutes, when they are exposed to sound waves of a frequency of 15,000 to 21,000 cycles/sec. Using *Streptococcus varians* and *Spirillum rubrum* in the vibrator, the decrease of the photosynthetic activity ran parallel with the process of pigment liberation from the cells, and the protein-containing pigment was not appreciably changed by the extraction method with sound waves. French suggested a distinguishing nomenclature: the name *chloroplastin* should be reserved for compounds with ordinary chlorophyll from higher plants and yellow green algae; *photosynthin* is suggested as a general term for compounds of photosynthetic activity, such as the combination bacterio-chlorophyll and protein.

The protein-chlorophyll combination undergoes very interesting changes when it is exposed to the action of proteolytic enzymes. Such experiments were performed by Inman [43] with trypsin in buffered solutions in the presence and absence of carbon dioxide, and the spectroscopic changes of the pigment component were measured. It was concluded that chlorophyll may be linked to the substrate for the enzyme by a magnesium-protein linkage. The absorption spectra of individual chloroplasts in living cells have been determined for *Protococcus, Spirogyra,* and *Euglena* by Albers and Knorr.[44] The results indicated either that there are other pigments present, which give absorption similar to that of chlorophyll *a,* or that chlorophyll *a* enters into the photochemical reaction in photosynthesis in several steps, forming reaction products with absorption bands at slightly different wave lengths. The first assumption may have significance in the light of the investigations of Strain and Manning; [45] they report that diatoms and brown algae normally contain chlorophyll *a,* and another chlorophyll-like pigment, which is different from chlorophyll *b.* For this pigment the name *chlorophyll c* or the old names *chlorofucine* (Sorby, 1873), or *chlorophyll* γ (Tswett, 1906) are suggested; the absorption spectrum and data for the isolation of the pigment are given. Since in certain regions of the spectrum chlorofucine absorbs considerably more light than chlorophyll *a,* it is considered possible that chlorofucine may play an important role in the process of photosynthesis.

The evolution of oxygen has been used in many investigations as the measure of photosynthetic activity. Recently this method of determination was applied to isolated chloroplasts by Hill and Scarisbrick.[46] At pH 8 the irradiation of isolated chloroplasts of *Stellaria media* in potassium ferric oxalate caused reduction of the trivalent iron under evolution of oxygen in the absence of carbon dioxide. This interesting result can, of course, not be interpreted as a successful performance of photosynthesis *in vitro.*

Undamaged chloroplast material from spinach leaves was isolated and analyzed by Menke; [47] the results are given in Table 3, together with the data for chloroplast substance.

From the difference in composition Menke calculated a content of cytoplasm material in the chloroplasts of approximately 15 per cent, or for the dry leaf 17.9 per cent

chloroplast substance and 16.7 per cent cytoplasm substance. The protein fraction of chloroplasts consists of about 80 per cent protein and 14.6 per cent nitrogen. After removal of the lipoids from the chloroplasts, a protein can be extracted, which contains phosphorus in a pentose nucleic acid. These experiments show the localization of phosphorus in the chloroplasts, not in the cytoplasm. A manometric study of the

Table 3

	Chloroplasts entire (%)	Chloroplast substance (%)
Protein (N × 6.25)	47.7	56.4
Ether-soluble ("Lipoids")	37.4	31.9
Ash	7.8	4.7
Undefined	7.1	7.0

gas metabolism of isolated chloroplasts demonstrated their inability to carry on photosynthesis; they were, however, able to perform a photoöxidation, since they used up more oxygen during irradiation than they did in the dark.[48]

Smith[49, 50] prepared aqueous solutions or suspensions of the green pigments by means of digitonin, and—following Mestre's suggestion[51] of a name for this type of chromoprotein—called the resulting material *phyllochlorin*.* The spectra of these solutions resemble spectroscopically those of chlorophyll in the leaves, and differ from the spectra of extracted chlorophyll.

Table 4. Absorption Spectra in Ethyl Ether

"Phyllochlorin"	675	470	437 mμ
Extracted Chlorophyll	660	465	420 mμ

A preliminary investigation of the molecular weight of "phyllochlorin" by means of the ultracentrifuge gave a value higher than 70,000 for this chromoprotein. Further purification showed a constant relation of about 16 parts of chlorophyll per 100 parts of protein, indicating a little over 3 molecules of chlorophyll per Svedberg protein unit of molecular weight 17,000; this fraction over 3 was interpreted as being possibly an extra molecule of chlorophyll *b*. "Phyllochlorin" is rendered water-soluble by the action of various detergents, such as bile salts, sodium desoxycholate, or sodium dodecylsulfate. The latter compound clarifies the solution completely, at the same time converting the chlorophyll quantitatively into pheophytin. The rate of this conversion depends on the hydrogen-ion concentration; it takes place even at pH 8 – 9, and is, at constant pH, proportional to the sodium dodecylsulfate concentration. Pheophytin formation does not occur in 4 per cent digitonin, or in 10 per cent bile salt solution at pH 4.5. At this pH a suspension of pure chlorophyll would be converted into a suspension of the magnesium-free pheophytin. In contrast to Inman (see[48]), Smith considers it improbable that magnesium plays a part in linking chlorophyll to the protein, but he believes that it may be concerned with the binding of larger units. "Phyllochlorin" does not carry on photosynthesis *in vitro*.

Special attention was paid to the size of the parenchyma cells used in a set of analyses of chloroplast material from tomato and tobacco leaves, performed by Granick.[52] At ⅛ maximum size the cells contain only about ⅒ of the total nitrogen and protein, which they contain at ⅜ maximum size; the difference between a cell of the latter size and one of maximum size is, however, very small: the cell of maximum size has only 0.25 per cent more nitrogen than the cell of ⅜ maximum size. Granick concludes from his analyses that for a molecular weight of 100,000 of the chloroplast protein there are approximately 30 chlorophyll molecules per protein molecule in the chloroplast. Per Svedberg protein unit of molecular weight 17,000,

* This term must not be confused with the chemical term phyllochlorin for the chlorin (= 7.8 – dihydro-derivative) of phylloporphyrin (see[1-3]).

this would mean approximately 5 chlorophyll molecules as compared with the foregoing value of 3 to 4.

The availability of radioactive elements in suitable compounds renders it possible to use such "tagged" elements in tracing reaction products. Thus the rate of photosynthesis by *Chlorella pyrenoidosa* with radioactive carbon dioxide ($*CO_2$) was determined by Ruben and co-workers.[58] Using the Warburg manometric method and simultaneously the radioactive technique, they find that carbon dioxide reduction is markedly affected by such poisons as hydrogen cyanide or phenyl urethane. At hydrogen cyanide concentrations of 10^{-2} to 10^{-3} molar, the respiration remains unchanged, but photosynthesis is strongly inhibited. *Chlorella* cells reduce $*CO_2$ reversibly in the dark (thermal reaction), and the radioactive intermediate products contain most of the radio carbon in form of carboxyl. The reaction may be written:

$$*CO_2 + R - H \rightleftharpoons R - *COOH$$

The authors suggest that it is the presence of the radioactivity in the carboxyl group which renders the formation of this group energetically and structurally possible. Later Ruben [54] investigated this reaction more thoroughly; from a comparison of the equilibrium constant for carbon dioxide fixation in *Chlorella* with the known equilibrium constants for simple carboxylations he concludes that the primary reaction in photosynthesis does not consist solely of a carboxylation as given above, but must involve a more complicated sequence of reactions. He bases his explanation on the studies of phosphorylation reactions,[55] and assumes a reaction scheme in which an energy-rich phosphate donor phosphorylates the carbon dioxide acceptor R — H (possibly an aldehyde), and the latter then reacts with carbon dioxide with formation of a carboxylic acid; if R — H is an aldehyde—and the author claims to have some indirect evidence for this assumption—an α-ketonic acid would be the reaction intermediate. In summary the scheme is represented by the following equations:

$$\begin{matrix}\text{Energy-rich}\\ \text{phosphate}\\ \text{donor}\end{matrix} + R - H \overset{\text{Enzyme}}{\rightleftharpoons} \text{"free" donor} + \begin{matrix}\text{phosphorylated}\\ R - H\end{matrix} \tag{1}$$

$$\begin{matrix}\text{phosphorylated}\\ R - H\end{matrix} + CO_2 \overset{\text{Enzyme}}{\rightleftharpoons} R - COOH + \begin{matrix}\text{inorganic}\\ \text{phosphate}\end{matrix} \tag{2a}$$

or, for the assumption of an aldehyde

$$\begin{matrix}\text{phosphorylated}\\ R - C - H\\ \quad\quad\|\\ \quad\quad O\end{matrix} + CO_2 \overset{\text{Enzyme}}{\rightleftharpoons} \begin{matrix}\quad\quad\quad\quad\text{inorganic}\\ R - C - COOH + \text{phosphate}\\ \quad\quad\|\\ \quad\quad O\end{matrix} \tag{2b}$$

The photochemical process is formulated by Ruben only in a general form:

$$A + B + H_2O + h\nu \overset{\text{chlorophyll}}{\rightleftharpoons} \begin{matrix}AH + BOH\\ (\text{or}\\ AH_2 + BO)\end{matrix} \tag{3}$$

where AH, or AH_2, and BOH, or BO merely symbolize oxidizing and reducing agents; it is claimed that "10 to 12 quanta are more than sufficient to fill the requirements of the above 'theory.'"

With the assumption of enzymes as reaction partners in photosynthesis it has always been important to determine the range of hydrogen-ion concentration within which the process is performed in nature. A large number of investigations deal with this subject; it must suffice here to mention that determinations in terms of pH have been made of the external hydrogen-ion concentration, *i.e.*, of the reaction medium in which the cells were grown, and also for the surface and for internal

conditions of the cells. The latter measurements are, of course, more complicated, and offer a greater possibility for errors.

At saturation concentrations of carbon dioxide the rate of photosynthesis by *Chlorella* does not change, when the external pH value varies between 4.6 and 8.9.[56] The changes of pH of cells and cell suspensions, and of suspensions of cell contents, and also of the oxidation-reduction potentials during photosynthesis have been studied by widely differing techniques, and under many variations of conditions, *e.g.*, under constant illumination, under alternating light flashes and dark periods, both of varied duration, in atmospheres of air, oxygen, or carbon dioxide, and of mixtures of these gases. For the measurements of the rates of photosynthesis and of respiration the dropping mercury electrode,[57] the glass electrode,[58] a platinum electrode,[59] and a bright platinum electrode polarized cathodically [60] have been used. A very interesting change was noticed by Blinks (see [58]) on *Stephanoptera* suspensions: intermittent light flashes of 0.02 second duration produced an acid gush, before an alkaline drift occurred. Upon darkening, a brief alkaline reaction was followed by the slower acid drift of respiration. Blinks discusses several possibilities as explanations for this acid effect:

(a) Release of carbon dioxide by some photolabile holding mechanism;
(b) Increased respiration in light in excess of photosynthesis;
(c) Production of an acid stronger than carbon dioxide as the first product of photosynthesis;
(d) Photodecomposition of compounds;
(e) Increased consumption of a base during the first moment of illumination.

For the first assumption, the existence of a photolabile holding mechanism, Shafer [61] furnishes a number of additional experimental data. In his opinion chlorophyll is not a factor concerned in the light-activated adsorption of carbon dioxide by leaves; it is the special holding mechanism which retains part of the carbon dioxide, either by adsorption, or in loose chemical linkage.

An important factor in discussions on photosynthesis is the assimilation of carbon dioxide by chlorophyll-free plants. There the absorption of radiant energy may be performed by other pigments; thus bacteriochlorophyll in the purple bacteria may be a parallel to chlorophyll in green plants. The two pigments are closely related chemically. Fischer [62] demonstrated the presence of magnesium in bacteriochlorophyll, and established its structural formula. For the assimilatory performance of bacteria reference must be made to the article by Van Niel already mentioned.[38] Van Niel concludes from the evidence now at hand that there is no reason to suppose a fundamental difference of photosynthesis in purple bacteria and in green plants. He assumes only the necessity of the presence of "appropriate hydrogen donors" (symbolized by H_2A in the following scheme) in the purple bacteria, and formulates his working hypothesis, with carbon dioxide as the substance to be reduced, for

Green Plants	$4(H_2O + h\nu$	$\longrightarrow H + OH)$
	$4H + CO_2$	$\longrightarrow (CH_2O) + H_2O$
	$2(2OH$	$\longrightarrow$ Peroxide $\longrightarrow H_2O + O_2)$
Purple Bacteria	$4(H_2O + h\nu$	$\longrightarrow H + OH)$
	$4H + CO_2$	$\longrightarrow (CH_2O) + H_2O$
	$2(2OH + H_2A$	$\longrightarrow 2H_2O + A)$

In the light of the work of Ruben and co-workers, the reduction process may, however, involve another substance, and in that case formaldehyde as the reaction intermediate is replaced by an intermediate with more than one carbon atom in the molecule.

The question of chlorophyll development in green plants under the influence of radiation has remained unanswered thus far in spite of many studies in this field.[63] In most plants chlorophyll development requires a photochemical change of a colorless precursor of chlorophyll; for this precursor a number of names are used in the

literature: colorless chlorophyll, chlorophor, leucophyll, etiolin, protophyllin, and protochlorophyll.[64] Protochlorophyll has been identified by Fischer [65] as the magnesium complex salt of the vinyl-pheoporphyrin a_5 phytyl ester; but the problem whether protochlorophyll is actually the precursor, or whether it is a decomposition product of chlorophyll, remains still unsolved (see Ref. 1, vol. II/$_2$, pp. 321-4).

It has been pointed out by Kostytschew [66] that photosynthesis in the green plant during the day is irregular, and that sometimes an intensive carbon dioxide elimination replaces the normal uptake of this gas. He suggests measurements of the total daily production of organic materials by the plant under natural conditions in order to remove this possible source of error.

The determination of the amount of radiation actually taken up by green leaves is at present also a subject of controversy. Lasareff [67] states that leaves absorb twenty times as much light as chlorophyll solutions, and that the energy falling on a leaf is absorbed with a loss of only 1-2 per cent; hence leaves may be regarded as absolutely black bodies. Other authors give data for much higher losses, and point out that structural and physiological conditions (*e.g.*, hairiness and age), or pathological conditions (like presence of mildew, or excessive radiation) may change the rate of light absorption. In his discussion of this subject Mestre [68] analyzes the photometric problem involved, and concludes that the present unsatisfactory state of accumulating "first approximation data" by a large variety of techniques on an increasing number of species must be changed to a thoroughgoing agreement on instrumental methods.

A much discussed problem is the fluorescence of photosynthesizing leaves. According to Kautsky,[69] the total amount of energy necessary in the process is absorbed by chlorophyll and transferred to the assimilatory system. This transfer is accompanied by a quenching of the fluorescence; and since oxygen is known to extinguish chlorophyll fluorescence, it is assumed that oxygen is acting as a collecting and transferring agent of the energy in photosynthesis. There is, on the other hand, much evidence that the presence of oxygen is not an absolute requirement in the process. The fluorescence of the leaves is very low, indeed, as compared with that of chlorophyll in solution. For such solutions under different atmospheres (air, oxygen, carbon dioxide, and nitrogen) detailed studies are available.[70, 71] But the exact correlation of absorption of light by the green leaf and re-emission in form of fluorescence is still lacking.

Only a fraction of the total incident radiation from the sun is used by the plant in photosynthesis; the balance is expended in other ways such as re-radiation, transpiration, thermal transmission, and respiration. The basic equation for the reduction of carbon dioxide (p. 606), in the form usually given, namely with a hexose (glucose) and oxygen as end products, is:

$$6CO_2 + 6H_2O \longrightarrow C_6H_{12}O_6 + 6O_2$$

The heat of combustion of glucose (mol. weight 180.16) is 673.0 kg-calories$_{15}$ per gram molecular weight (2,816.5 kilo Joules). This value, the amount of energy fixed in the process and reclaimable by combustion,[72] is obviously the minimum requirement for performing the reaction. The determination of the amount of energy necessary above this minimum to photosynthesize glucose has not yet been performed. Actually the reaction is highly improbable in the simple form given above, since carbon dioxide most certainly does not react as gas, but dissolves in water with formation of H^+ and HCO_3^- ions in equilibrium with undissociated H_2CO_3. Baly [8] pointed out that for the reaction

$$CO_2 + H_2O + 112{,}000 \text{ cal} \longrightarrow \tfrac{1}{6}C_6H_{12}O_6 + O_2$$

performed in one step, the wave length of light necessary would be 253.5 mμ, while the light of the sun reaching the earth does not contain wave lengths below approxi-

mately 290.0 mμ. This consideration illustrates one of the greatest difficulties in the present status of our experience concerning photosynthesis, namely the lack of information as to the number of chemical and photochemical reactions, and the quantity of solar energy used to bring about these reactions.

The assumption of a photosynthetically active symplex of a colloidal nature seems to explain the energetic difficulties very well, but the isolation of chemically defined intermediates is still an unfinished task. The validity of any of the numerous hypotheses and theories on photosynthesis can be tested only by the ultimate experimental performance of obtaining such intermediate compounds, and by accomplishing the process itself *in vitro*.

References

1. Fischer, H., and Orth, H., "Die Chemie des Pyrrols," 3 Volumes (I, II/$_1$ and II/$_2$) Akademische Verlagsgesellschaft m. b. H., Leipzig.
2. Steele, C. C., "Chlorophyll"; in "Organic Chemistry," H. Gilman, Editor-in-Chief, John Wiley and Sons, Inc., New York, 1943.
3. Rothemund, P., "Chlorophyll," in "Handbook of Medical Physics," Otto Glasser, Editor.
4. Bose, Sir J. C., "The Physiology of Photosynthesis," Longmans, Green and Co., London, 1924.
5. Stiles, W., "Photosynthesis," Longmans, Green and Co., London, 1925.
6. Spoehr, H., "Photosynthesis," Chem. Catalog Co., (Reinhold Publishing Corp.) New York, 1926.
7. Wurmser, R., "Oxidations et Réductions," Vol. XV of "Les Problèmes Biologiques"; Chapter II, 30-63. Les Presses Universitaires de France, Paris, 1930.
8. Baly, E. C. C., "Photosynthesis," D. Van Nostrand Company, Inc., New York, 1940.
9. Rabinowitch, E., "Photosynthesis."
10. Warburg, O., and Negelein, E., *Naturwiss.*, **10**, 647-53 (1922).
11. Warburg, O., and Negelein, E., *Z. physik. Chem.*, **102**, 235-66 (1922).
12. Warburg, O., and Negelein, E., *Z. physik. Chem.*, **106**, 191-218 (1923).
13. French, C. S., *Science*, **84**, 575 (1936).
14. French, C. S., *J. Gen. Physiol.*, **20**, 711-35 (1936-7).
15. French, C. S., *J. Gen. Physiol.*, **21**, 71-87 (1937-8).
16. Manning, W. M., Stauffer, J. F., Duggar, B. M., and Daniels, F., *J. Am. Chem. Soc.*, **60**, 266-74 (1938).
17. Manning, W. M., Juday, C., and Wolf, M., *J. Am. Chem. Soc.*, **60**, 274-8 (1938).
18. Eymers, J. G., and Wassink, E. C., *Enzymologia*, **2**, 258-304 (1938).
19. Rieke, F. F., *J. Chem. Phys.*, **7**, 238-44 (1939).
20. Petering, H. G., Duggar, B. M., and Daniels, F., *J. Am. Chem. Soc.*, **61**, 3525-9 (1939).
21. Magee, J. L., DeWitt, T. W., Smith, E. C., and Daniels, F., *J. Am. Chem. Soc.*, **61**, 3529-33 (1939).
22. Eichhoff, H. J., *Biochem. Z.*, **303**, 112-31 (1939).
23. Eichhoff, H. J., and Noddack, W., *Atti X° Congr. Internat. Chim.*, **4**, 427-39 (1939).
24. Emerson, R., and Lewis, C. M., AAAS Meeting, Seattle, Wash., June 20, 1940.
25. Rieke, F. F., Unpublished experiment, quoted by Franck and Gaffron, Reference 35.
26. Emerson, R., and Green, L., *Plant Physiol.*, **12**, 538-45 (1937).
27. Blackman, F. F., *Ann. Botany*, **19**, 28 (1905).
28. Emerson, R., and Arnold, W., *J. Gen. Physiol.*, **15**, 391-420 (1932).
29. Emerson, R., and Arnold, W., *J. Gen. Physiol.*, **16**, 191-205 (1932).
30. Arnold, W., and Kohn, H. I., *J. Gen. Physiol.*, **18**, 109-12 (1934-5).
31. Gaffron, H., and Wohl, K., *Naturwiss.*, **24**, 81-90, 103-7 (1936).
32. Wohl, K., *New Physiologist*, **39**, 33 (1940).
33. Van Niel, C. B., "Advances in Enzymology," Nord, F. F. and Werkman, C. H., Editors, Volume I, 263-328, Interscience Publishers, Inc., New York, 1941.
34. Ornstein, L. S., Wassink, E. C., Reman, G. H., and Vermeulen, D., *Enzymologia*, **5**, 110 (1938).
35. Franck, J., and Gaffron, H., see Reference 33, pp. 199-262.
36. Herlitzka, A., *Biochem. Z.*, **38**, 321.(1912).
37. Wurmser, R., Levy, R., and Tessier, G., *Ann. Physiol. et Physico—Chimie Biologique*, **1**, 298 (1925).
38. Lubimenko, V. N., *Rev. Gén. Bot.*, **39**, 547 (1927).
39. Stoll, A., *Naturwiss.*, **24**, 53-9 (1936).
40. Willstätter, R., and Rohdewald, M., *Z. physiol. Chem.*, **225**, 103-24 (1934).
41. Baas-Becking, L. G. M., *Proc. VIth Intern. Botan. Congress, Amsterdam* **II**, 265 (1935).
42. French, C. S., *Science*, **88**, 60-2 (1938).
43. Inman, O. L., and Crowell, M. *Plant Physiol.*, **14**, 388-90 (1939).
44. Albers, V. M., and Knorr, H. V., *Plant Physiol.*, **12**, 833-43 (1937).
45. Strain, H. H., and Manning, W. M., *J. Biol. Chem.*, **144**, 625-36 (1942).
46. Hill, R., and Scarisbrick, R., *Proc. Roy. Soc. London B*, **129**, 238-55 (1940).
47. Menke, W., *Z. physiol. Chem.*, **257**, 43-8 (1939).
48. Experiments by Dr. A. Pirson, quoted in Reference 47.
49. Smith, E. L., *Science*, **88**, 170-1 (1938).
50. Smith, E. L., *Science*, **91**, 199-200 (1940).
51. Mestre, H., "The Investigation of the Pigments in the Living Photosynthetic Cell," Contr. to Marine Biology, Stanford University Press, 1930.
52. Granick, S., *Am. J. Botany*, **25**, 561-7 (1938).
53. Ruben, S., Kamen, M. D., Hassid, W. Z., and De Vault, D. C., *Science*, **90**, 570-1 (1939).
54. Ruben, S., *J. Am. Chem. Soc.*, **65**, 279-82 (1943).
55. Review on the subject of phosphorylations: Lipmann, F., "Metabolic Generation and Utilization of Phosphate Bond Energy," in "Advances in Enzymology," Nord, F. F., and Werkman, C. H., Editors, Volume I, pp. 99-162, Interscience Publishers, Inc., New York, 1941.
56. Emerson, E., and Green, L., *Plant Physiol.*, **13**, 157-68 (1938).
57. Vitek, V., *Collection Czechoslovak. Chem. Communic.*, **7**, 537 (1935).
58. Blinks, L. R., and Skow, R. K., *Proc. Nat. Acad. Sci. U. S.*, **24**, 413-9 (1938).
59. Tang, P. S., and Ling, C. Y., *J. Cell. Comp. Physiol.*, **9**, 149 (1936).

60. Blinks, L. R., and Skow, R. K., *Proc. Nat. Acad. Sci. U. S.*, **24**, 420-7 (1938).
61. Shafer, J., Jr., *Plant Physiol.*, **13**, 141-56 (1938).
62. Fischer, H., Lambrecht, R., and Mittenzwei, H., *Z. physiol. Chem.*, **253**, 1-39 (1938).
63. Inman, O. L., Rothemund, P., and Kettering, C. F. "Chlorophyll and Chlorophyll Development in Relation to Radiation," Chapter 34 in "Biological Effects of Radiation," Duggar, B. M., Editor McGraw-Hill Book Company, New York, 1936.
64. Rothemund, P., *Cold Spring Harbor Symposia on Quantitative Biology*, **3**, 71-9 (1935).
65. Fischer, H., Mittenzwei, H., and Oestreicher, A., *Z. physiol. Chem.*, **257**, IV-VII (1939).
66. Kostytschew, S., *Planta*, **13**, 778-82 (1931).
67. Lasareff, P., *Biochem. Z.*, **182**, 131-3 (1927).
68. Mestre, H., *Cold Spring Harbor Symposia on Quantitative Bilology*, **3**, 191-209 (1935).
69. Kautsky, H., *Berichte*, **65**, 1762-70 (1932).
70. Albers, V. M., and Knorr, H. V., *Cold Spring Harbor Symposia on Quantitative Biology*, **3**, 87-97 (1935).
71. Knorr, H. V., and Albers, V. M., *Cold Spring Harbor Symposia on Quantitative Biology*, **3**, 98-107 (1935).
72. Kharasch, M. S., *Bur. Standards J. Res.*, **2**, 359-430 (1929).

Plant Cell Membranes

WANDA K. FARR *

The nature of the processes by which plants produce complex, organic substances from simple, inorganic materials was a clearly defined biological and chemical problem at the beginning of the nineteenth century. Shortly before 1840 two important advances toward its solution were made. The classification of organic substances into carbohydrates, proteins, and fats by Prout [204] occurred almost simultaneously with the recognition by Schleiden [226] and Schwann [232] that "All organized structures are composed of essentially similar units, namely *cells*." The approach to a clearer understanding of the synthetic power of living plants was thus changed from a state of more or less general speculation to one of direct observation and experimentation. The methods of study used by these nineteenth-century investigators were largely microscopic and chemical, and their inheritance, in terms of apparatus and procedures, from scientists of the preceding eras, was of great value.

Development of Chemical Techniques

The foundation of organic chemistry had been built by Lavoisier in the eighteenth century (1743-1794), and among his immediate successors in this field of research were Berzelius (1779-1848), Gay-Lussac (1778-1850), Liebig (1803-1873), and Wöhler (1800-1882). During the early part of the nineteenth century their methods of identification and analysis were used to promote the understanding of the chemical and physical properties of many different organic substances.

Kirchhof [148] in 1816 described the reactions which alter starch and transform it partially into gummy material and sugar. Bracconot [80] in 1825 removed the sugars from finely divided vegetable material and heated the residue with dilute HCl followed by dilute NaOH. A thick, mucilaginous solution was filtered from the residue. When acidified, it was transformed into a voluminous jelly which he called *pectic acid*. In a later publication Bracconot [81] reported the extraction of at least a portion of the pectic compounds from plant tissues by treatment with water alone. This opaque, mucilaginous solution was precipitated by alcohol in the form of a firm jelly. He suggested that this water-soluble, gelatinous compound should be named *pectin*. Fremy [69] was the first to recognize the presence of an insoluble substance of pectic nature in plant tissues; to it he gave the name *pectose*. He believed this to be an insoluble salt of calcium or potassium which, when acted upon by acids, gives rise to free pectin and bases. He was not able to remove the pectose from plant tissues without bringing about its decomposition.

* A contribution from the former Cellulose Department of the Chemical Foundation, Boyce Thompson Institute for Plant Research, Inc., Yonkers, New York.

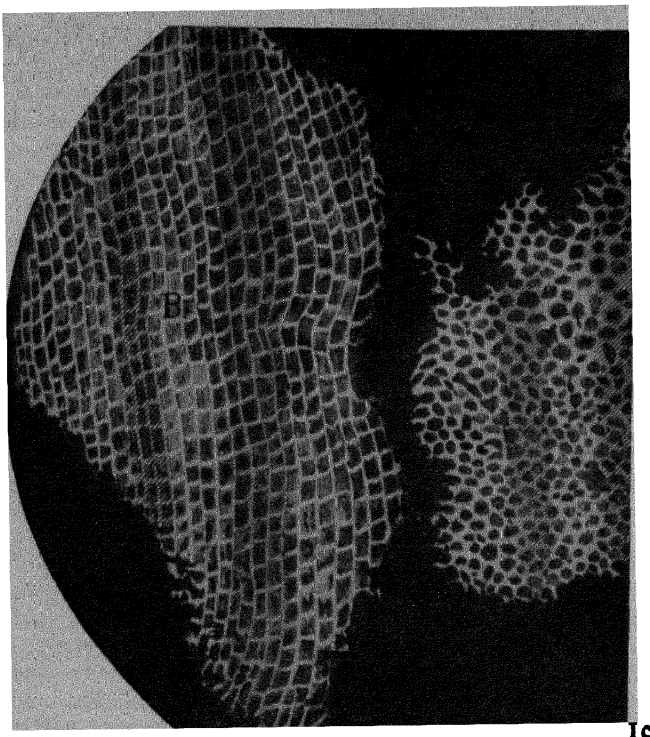
Prout [204] obtained a firmer grasp upon the more general aspects of organic analyses through his recognition, in 1827, that the principal alimentary substances employed by man can be classified as *saccaharine, oily,* and *albuminous.* In his discussion of carbohydrates (p. 375) he states, "Starch differs from sugar by containing minute portions of other materials which, we may presume, prevent its constituent particles from arranging themselves in crystalline form and thus cause it to assume totally different properties."

Prout was at first at a loss to form any notion of the *modus operandi* of these minute admixtures of foreign bodies. A paper published by Herschel [112] in 1824, dealing with certain motions produced in fluid conductors when transmitting an electric current, threw new light upon the subject and led Prout to comment, "That such minute proportions of extraneous matter should be found capable of communicating sensible mechanical motions and properties of a definite character to the body with which they are mixed, is perhaps one of the most extraordinary facts that has yet appeared in chemistry. When we see energies so intense exerted by the ordinary forms of matter, we may reasonably ask what evidence we have for the imponderability of any of the powerful agents to which so large a part of the activity of material bodies seems to belong." He expressed the belief that small quantities of any substance may be capable of functioning in such a capacity, and suggested that water is first in importance, in this respect, in organized substances. Prout's treatise contains methods of separation and combustion analyses of a large number of organic substances such as *lignin, starch, mannan, gum arabic, oxalic acid, citric acid,* and *tartaric acid;* it serves as a fitting example of achievements in the field of organic chemistry up to and including the first quarter of the nineteenth century.

The classification of chemical substances into crystalloids and colloids by Graham [94] was of fundamental importance in the interpretation of the composition and behavior of organic materials. Early applications of his findings are to be found in the work of Rainey,[205] who pointed out that crystalline substances, normally angular in form, are made to assume globular form by deposition in viscous or gummy substances. He considered the possibility that the viscosity of the gum annuls the polarity of the crystal, and allows the molecules of the crystal to obey simple laws of common and mutual attraction. He showed that, when calcium carbonate crystals are formed in a solution of gum arabic, a portion of the gum is enclosed or intermixed. He termed this phenomenon *molecular coalescence* and produced its counterpart, *molecular disintegration,* by immersing the crystals so formed in solutions of higher viscosity than those in which they were produced. Harting [108] found, likewise, that crystals of calcium carbonate, formed in a solution of albumen, were in the form of "calcosphaerites" and that the albumen contained in the calcosphaerites had undergone a chemical change which caused it to resemble *chitin.* Ord [191] extended the studies of Rainey to other crystalline substances, including uric acid and such colloids as albumen, gelatin, amyloid, and pectin; he found that when a crystalloid is deposited from solution in the presence of a colloid the crystals are broken down into spheres, or spheroids, or molecules, by the action of the colloids. The colloidal properties of both cell membrane materials and the protoplasts from which they are formed made these considerations of great importance in both the chemical and microscopic analyses of plant tissues. It is somewhat surprising to find that they received little immediate attention in this field of research and that this situation continues, in a measure, up to the present time.

Development of Microscopic Techniques

During the later Middle Ages magnifying glasses were made consisting of a single glass lens in the form of a segment of a sphere, mounted in a wooden tube, with provision for focusing. Small lenses of short focal distances, fashioned in the hands of skilled artisans, and mounted in silver or brass, magnified properly mounted



I could exceeding plainly perceive it to be all perforated and porous, much like a Honey-comb, but that the pores of it were not regular; yet it was not unlike a Honey-comb in theſe particulars.

Firſt, in that it had a very little ſolid ſubſtance, in compariſon of the empty cavity that was contain'd between, as does more manifeſtly appear by the Figure A and B of the XI. *Scheme*, for the *Interſtitia*, or walls (as I may ſo call them) or partitions of thoſe pores were neer as thin in proportion to their pores, as thoſe thin films of Wax in a Honey-comb (which encloſe and conſtitute the *ſexangular cells*) are to theirs.

Next, in that theſe pores, or cells, were not very deep, but conſiſted of a great many little Boxes, ſeparated out of one continued long pore, by certain *Diaphragms*, as is viſible by the Figure B, which repreſents a ſight of thoſe pores ſplit the long-ways.

I no ſooner diſcern'd theſe (which were indeed the firſt *microſcopical* pores I ever ſaw, and perhaps, that were ever ſeen, for I had not met with any Writer or Perſon, that had made any mention of them before this) but me thought I had with the diſcovery of them, preſently hinted to me the true and intelligible reaſon of all the *Phenomena* of Cork; As,

Firſt, if I enquir'd why it was ſo exceeding light a body? my *Microſcope* could preſently inform me that here was the ſame reaſon evident that there is found for the lightneſs of froth, an empty Honey-comb, Wool, a Spunge, a Pumice-ſtone, or the like; namely, a very ſmall quantity of a ſolid body, extended into exceeding large dimenſions.

FIGURE 1. Drawing of cellular structure of cork and accompanying description from Hooke's "Micrographia," 1665.

objects from 50 to 300 times their original diameters. Such instruments were called *microscopes*.

The *compound microscope*, invented by Hans and Zacharias Janssen [28] in 1590, consisted of two lenses, one placed near the eye of the observer, the other near the object, a principle of construction fundamental in microscopes used at the present time. Images formed with the compound microscopes left much to be desired, skillful observers seeing as much, if not more, with the simple microscopes. This was

Obſerv. 4.. Of Tabby.

A ſhort deſcription of it. A conjecture about the reaſon why Silk is ſo ſuſceptible of vivid colours: and why Flax and Hair is not. A conjecture, that it may perhaps be poſſible to ſpin a kind of artificial Silk out of ſome glutinous ſubſtance that may equalize natural Silk.

trant, juyces or liquors. And thereby thoſe tinctures, though they tinge perhaps but a ſmall part of the ſubſtance, yet being ſo highly impregnated with the colour, as to be almoſt black with it, may leave an impreſſion ſtrong enough to exhibite the deſir'd colour. A pretty kinde of artificial Stuff I have ſeen, looking almoſt like tranſparent Parchment, Horn, or Iſing-glaſs, and perhaps ſome ſuch thing it may be made of, which being tranſparent, and of a glutinous nature, and eaſily mollified by keeping in water, as I found upon trial, had imbib'd, and did remain ting'd with a great variety of very vivid colours, and to the naked eye, it look'd very like the ſubſtance of the Silk. And I have often thought, that probably there might be a way found out, to make an artificial glutinous compoſition, much reſembling, if not full as good, nay better, then that Excrement, or whatever other ſubſtance it be out of which, the Silk-worm wire-draws his clew. If ſuch a compoſition were found, it were certainly an eaſie matter to find very quick ways of drawing it out into ſmall wires for uſe. I need not mention the uſe of ſuch an Invention, nor the benefit that is likely to accrue to the finder, they being ſufficiently obvious. This hint therefore, may, I hope, give ſome Ingenious inquiſitive Perſon an occaſion of making ſome trials, which if ſucceſsfull, I have my aim, and I ſuppoſe he will have no occaſion to be diſpleas'd.

FIGURE 2. Photographic reproduction of Hooke's statement.

due to the presence of chromatic and spherical aberrations in the objectives of the compound microscopes. A simple lens of extremely short focal distance is not defective in these respects. It was not until the first quarter of the nineteenth century that the necessary corrections to the objectives were made, and the supremacy of the compound microscope over the simple microscope thus established.

Before the close of the seventeenth century, the first important observations of the microscopic structure of plants, animals, and inorganic substances, at magni-

fications as high as 500 diameters, had been made by the Dutch microscopist, Leeuwenhoek,[154] the English mathematician and microscopist, Hooke,[133] (Fig. 1), the English physician and botanist, Grew,[95] and the Italian physician and physiologist, Malpighii.[161] In the works of these early writers occasional comments as to the relation of microscopic structure to function and gross physical behavior also occur. One of these instances, of important prophetic significance, is to be found in Hooke's *Micrographia*.[133, p. 7]

Having recognized, in the course of his observations, the plastic properties of plant tissues, Hooke thus covered, in one sweeping statement, the fundamental principles of artificial-silk manufacture, an industry not developed until more than two centuries later.[89]

Rapid improvements in the design and construction of the compound microscope transformed it from a two-lens into a many-lens instrument in which spherical aberration was largely overcome by means of suitable forms and combinations of lenses. The important researches of Dolland (1758), Euler (1771), Fuss (1774), and Fraunhofer (1815) laid the foundation for the correction of chromatic aberration. Transmitted light, obtained by placing the illuminating reflector under the object, was introduced in the year 1735 by Culpepper and Scarlet.[107] Instruments, in which microscopic observations were no longer hindered by either marked spherical or marked chromatic aberrations, were available, therefore, early in the nineteenth century. The impetus to the microscopic study of plant tissues during this period was thus linked inextricably with improvements in the microscopes themselves.

Under these favorable conditions many of the chemical techniques were adapted to microscopic observation and the science of *microchemistry* had its beginning. Carbohydrates, fats, proteins, and other organic materials were identified not only separately in the test tube, but also microchemically in plant tissues, thus furnishing added information concerning their relative functions in building up the highly complex organ, endowed with synthetic powers of unknown nature—the plant cell. The microscopic and chemical data, which were rapidly accumulated by botanists and chemists during the first half of the nineteenth century, constitute the foundations upon which we have continued to build to the present time. Many of their methods of identification are in use today with little modification.

Plant-cell Membranes

In the greater number of living plant cells there are visible two distinctly different parts: the delicate, more or less viscous protoplast, and the comparatively rigid cell membrane which surrounds it. The general organization and functioning of the protoplast presents problems of the greatest difficulty. Its appearance is constantly changing, and one hundred years ago but little was known of the cellular changes which accompany metabolic and reproductive processes. Cell membranes are somewhat more stable, and it is not surprising to find the earliest interest of botanists directed toward a more complete understanding of their microscopic structure and chemical composition.

The first observations of plant cell membrane structure antedate the recognition of the importance of the cell as an organic unit.

In 1819, Lyngbye,[160] a Danish clergyman, wrote the following description of a unicellular marine plant: *"Substantia membranacea, non gelatinosa, sed pro tenuitate, sub lente minutissime punctata."* A photomicrograph of this plant cell membrane [Figs. 3b, d] has the essential characteristics of the drawing which accompanied Lyngbye's description,[Figs. 3a, c] and was made about a century and a quarter earlier.

Schleiden, who helped to evolve the "cell theory" in 1838, made careful studies of cell membranes as well as of protoplasts during cell membrane formation. In this connection he states:[226, p. 287] "In the cell there is also another distinct, perfectly

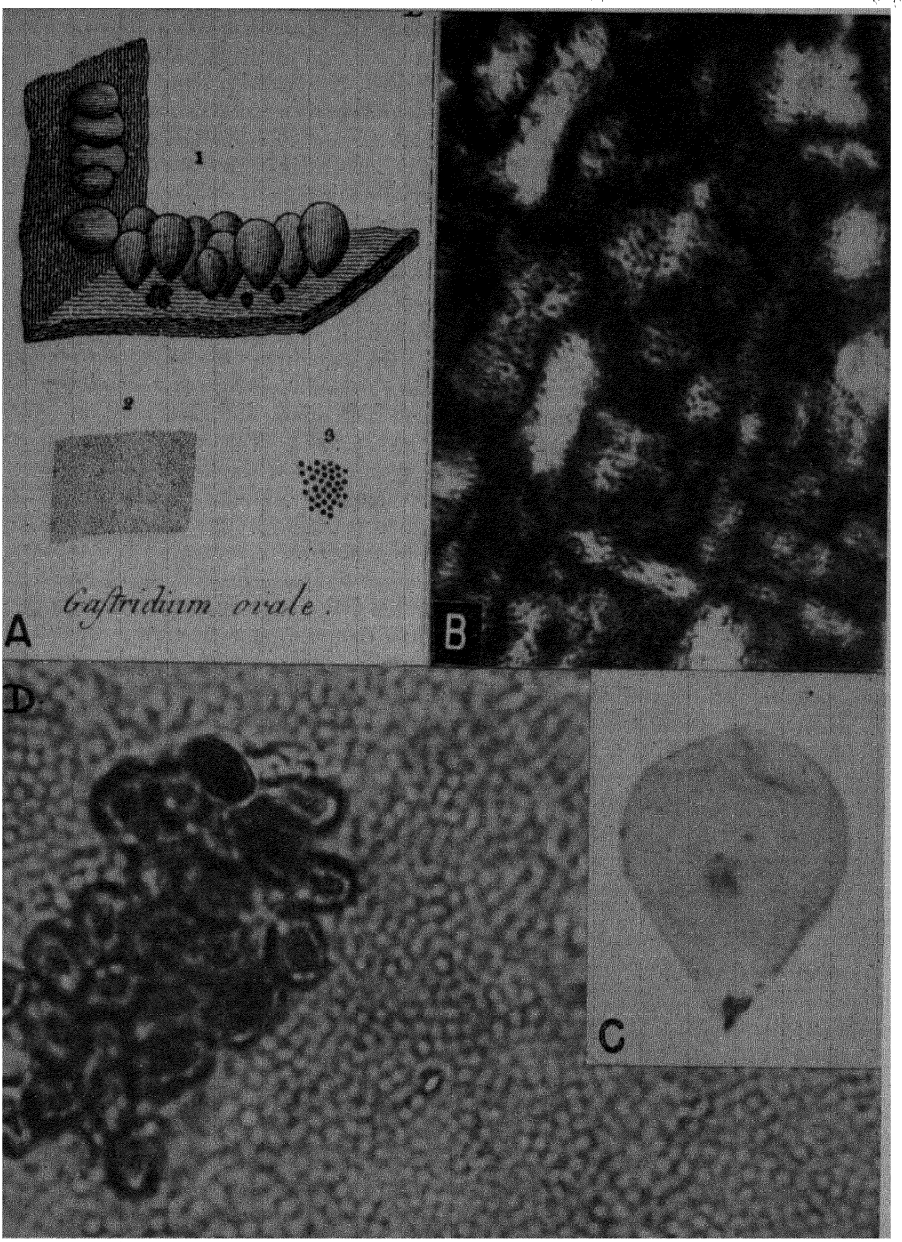


FIGURE 3. A. Copy of a plate from Lyngbye (1819) showing drawing of entire plants of *Gastridium ovale* (renamed *Halicystis ovalis* in 1901) and a detailed sketch of the "minutissime punctata" in the cell membrane. B. Photomicrograph of the same cell membrane in polarized light showing patch-like orientation of doubly refractive material in the wall (X 690). C. Photograph of single plant of *Halicystis* (X 2). D. Photomicrograph of membrane structure of *Halicystis* showing cellulose particles originally described by Lyngbye and chloroplasts in which the formation of cellulose particles is taking place (X 1460).

transparent substance which presents a homogeneous appearance when subjected to pressure; when dried it imbibes water and swells. For brevity's sake I shall call it vegetable gelatin, and am inclined to class under this head, as mere slight modifications, pectins and many of those substances which are usually enumerated under the term 'vegetable mucous.' It is this gelatin which is ultimately converted by new chemical changes into the actual cellular membrane, or structures of which it consists in a thickened state, and into the material of the vegetable fiber." [226, p. 244-245] "This jelly finally is converted at its outer surface into vegetable fiber, following the direction of a spiral line. . . . The unusual and altogether absolute fact at which we arrive is that these spiral fibrils are never found free, but are developed in the interior of the cells; and that the walls of these cells in the young state are simple, and generally very delicate. . . . I have reason to believe that complete consumption of the fibrils takes place in the formation of the membranes of the spiral vessels, and is the cause of their subsequently conveying only air. More frequently, however, one or more fibrils are unused; but then a great portion of the jelly has still remained unconsumed, which, when the cell is moistened with water, comes forth in the form of an intestine, and, in swelling, expands itself over the fibers."

From observations of earlier stages of protoplasmic behavior he adds: [226, p. 286] "In the further process of organization (in which process, the gum-like jelly is always the last immediately preceding fluid) a quantity of minute granules appears in it, most of which, on account of their minuteness, look like mere black points."

Likewise, von Mohl [175, p. 86] found in 1836: "If a tissue composed of young cells be left some time in alcohol, or treated with nitric or muriatic acid, a very thin, fine, granular membrane becomes detached from the inner wall of cells in the form of a closed vesicle. This membrane colors yellow with iodine."

Valentin observed at about the same time: [262, p. 287] "We see that at first a simple cell membrane exists in which there are neither spiral lines nor cross-lines. In the interior of the cell one finds a granulous substance whose particles exhibit no orderly arrangement. Later they build cross-lines, then spiral lines, in which, early in the development, one may still distinguish the discrete particles, and which finally form an unbroken continuity. In this way filaments are formed which, in some situations, have an appearance like strings of pearls, in others, on the contrary, are less sharply indented. A transformation of the cell contents thus takes place, the granules gradually disappearing; the wall of the cell at the same time becomes thicker at the expense of the protoplast."

In the studies of Payen,[194] particular attention was given to the chemical composition of the membranes and the protoplasts of many types of cells. He described differences between tissues which are more or less "pure" cellulose and those which contain larger proportions of "incrusting materials." In 1839 he published the rough formulas for these substances obtained from wood:

Incrusting substance: $C_{35}H_{24}O_{10}$

Cellulose: $C_{24}H_{20}O_{10}$ or $C_{24}H_{18}O_9$ plus H_2O

He found that cotton fibers contain less incrusting substance and are more easily purified than linen. He pointed out also evidence for the isomeric relationship of *starch, cellulose,* and *dextrine,* stating that these products differ among themselves only in their remarkable states of aggregation, which greatly modify their physical properties. Payen's "Mémoires" [195] published in 1842 contained the important information that, after extensive analyses, he had demonstrated that the membrane substances from all parts of the plant kingdom, when purified from foreign deposits, exhibit the same composition in terms of relative proportions of carbon, hydrogen, and oxygen. Purification treatments consisted in successive extractions with hydrochloric acid, ammonia, water, alcohol, and ether. Material thus prepared gave an analysis reasonably constant, isomeric with starch, and was identified by the author as *cellulose.* Payen also reports an experiment, "curious and easy to repeat," upon

the microscope slide which permits the recognition of the parts of the tissues which are made of cellulose, without purification. His directions are as follows:[195, p. 216] "(a) Place a small amount of tissue upon a slide in a drop of water. Add two or three drops of iodine. Absorb excess with a piece of filter. A light yellow coloration is produced. (b) Add a drop of concentrated sulphuric acid and again remove excess with filter paper. The cells with which it first comes into contact begin to separate, then acquire a blue violet shade, brought about by the iodine in contact with the cellulose, which in the course of the reaction is reduced to a state of aggregation of amylaceous particles."

This microchemical method of cellulose identification has been used continually, even to the present time. Through its careful application, both alone and combined with other micro- and macrochemical treatments, Payen was able to recognize the presence of pectic substances, wax, proteins, minerals, and many other materials, in intimate association with the cellulose in plant-cell membranes of many types. In 1856 he reported[196] the confirmation of these earlier findings through his own continued studies as well as those of Bronginart, de Mirbel, von Mohl, Trecul, and others. Three years later, honoring the presence of the distinguished botanist of Munich, Carl von Nägeli, he presented before the Paris Academy[197] his more recent observations upon the analogies and differences between starch and cellulose. Among other reagents, the effect of cuprammonium hydroxide, recently introduced as a solvent for plant fibers by Schweizer,[238] was described. He pointed out that the obstacles to solution in "cuprammonium" are often the cell cuticle and the incrusting substances, and that when these are removed, dissolution of the cellulose can be effected. He had found that solutions of cellulose thus prepared, to which a slight excess of hydrochloric acid or acetic acid had been added, precipitated the cellulose in white, granular flocks. When washed with water these represented the pure elementary particles of cellulose whose division, after precipitation, he considered to be more mechanical than chemical, for not only were they insoluble in water, but iodine would not color them directly. He found, however, after treatment with an aqueous solution of slightly alcoholized iodine, which gave them a pale-yellow color, that it sufficed to touch them with sulfuric acid to bring about the blue coloration with iodine characteristic of cellulose. A more complete reaction with sulfuric acid gradually converted the cellulose granules into dextrine and sugar. Payen found also that cellulose is precipitated from cuprammonium solution not only by mineral and vegetable acids, but also by an excess of ammonia and by a large excess of water. The elementary analysis of the precipitated cellulose, made with the assistance of M. Billequin, gave "the composition of pure cellulose, $C_{12}H_{10}O_{10}$."

At a later period in this same year, Payen[198] summarized for the Paris Academy his many years of fruitful research in the field of cell-membrane analysis with the following remarks: "Cellulose of the pith of underground stems and aerial stems of herbaceous and ligneous plants does not differ from that which makes up the textile fibers and other cells of plant tissues. Differences which exist between the properties of cellulose in plant organisms depend largely upon the variable degrees of its gradually accrued cohesion, and the influence of the organic substances or minerals with which it is injected or combined. Without doubt there remains much to do with respect to the precise determination of these substances in the different ligneous and herbaceous tissues. Perhaps there will be found many states truly isomeric with pure cellulose; but that demonstration, strongly interesting, without doubt, is not to date accomplished."

During this same period the question of the isomeric nature of cell-membrane constituents had been studied intensively by Fremy.[71] He had shown, over ten years earlier,[70] that vegetable tissues contain an insoluble substance, to which he gave the name *pectose*, which accompanies the cellulose almost constantly and which, under very weak influences, is converted into soluble *pectine*. The knowledge of this con-

version was applied to the better understanding of the remarkable modifications which he saw in cell membranes during vegetative growth. In certain cases their thickness was greatly augmented; in others, greatly diminished. The latter phenomenon was illustrated by the wall of a cell of a green fruit, which is very thick and constituted of many concentric layers. At the time of maturation the membranous material in pears, *e.g.,* was found to be reduced from 17.7 parts in 100 to 3.4 parts in 100, and this alteration was indicated by the changes which the fruit undergoes in hardness and transparency.

By treating cell membranes successively with cuprammonium hydroxide, an acid, and later potassium hydroxide, Fremy [71] found that cellulose, nitrogenous material, minerals, and pectic material can be identified in most plant cell membranes. He also found a substance, which he believed to be derived neither from cellulose nor from pectic material, which was soluble in water, had an acidity comparable with malic acid, was non-volatile, formed soluble compounds with all the bases, reduced salts of gold or silver with great ease, and was closely associated with the pectic material in the membrane. He named this substance *acide cellulique.*

In a later communication, Fremy [72] summarized his findings with the statement that the differences which one finds in the various types of cell membranes in the different parts of the vegetable kingdom are probably due to foreign materials deposited upon the surface of the cellulose, or infiltrated into its thickenings. While he could not entirely exclude the possibility that these substances in the membrane have, for a base, a body which is isomeric with cellulose, he expressed the belief that between them, at least, there are differences which are comparable to differences which exist between *albumen, fibrin,* and *casein.*

The physical structure of vegetable cell membranes likewise received attention from botanists and microscopists. The studies of Schleiden,[226] Valentin,[262] and Meyen [167] had shown that the membranes are not built of a homogeneous substance, and that many cell membranes are distinctly fibrillar. Von Mohl [176] reported in 1845 that while it was clear that these cell membranes possessed a definite inner structure, it did not seem to be conclusive that they were made up of definite fibrils separated from one another. He suggested that many appearances in the fiber might be due to irregular arrangement of associated molecules in the membrane, which resulted in the easily discernible divisibility of the membrane.

After the publication of the work of J. G. Agardh,[1] von Mohl repeated this phase of his work, disproving his former conception and corroborating Agardh's findings that the fibrils of the cell membrane are distinct, separable entities and that it is therefore not possible to call them "faults." In *Conferva aerea* the lamellae of the membrane separated in a clean-cut fashion, showing no connection between the sheaths. He concluded, therefore, that the spiral fibrils do not pass from one lamella to another. He found also that the fibrils within a given lamella could be separated with needles. Schacht [221] likewise found that these cell membranes could be disintegrated into fibrils by either mechanical or chemical means, and showed that the fibrils, when arranged in opposite directions in the membrane, have opposite properties in polarized light. Wigand [269] in 1856 corroborated von Mohl's conclusions that the crossing fibrils, so evident in many types of cells, belong to different layers of the cell membrane.

With full knowledge of these studies in which abundant evidence of the chemical and physical heterogeneity of plant cell membranes had been presented over a period of more than twenty years, Carl von Nägeli,[182, 183, 184, 185] having studied cell membranes and starch grains in polarized as well as ordinary light, drew the following conclusions: (a) The presence of crystalline material in the membranes is indicated by their double refraction in polarized light; (b) these crystallites are in molecular sheaths which are parallel with the surface of the cell as well as in molecular rows which extend from the inside out; (c) in one direction they are sheaths or lamellae,

in the two other directions, the visible membrane striations; (d) both lamellation and striation are due to the alternation of "water-rich" and "water-poor" areas; (e) the dry colloid is made up of diminutive, microscopically invisible, crystalline micellae which are nearly contiguous and almost fill the space; (f) by moistening with water or aqueous fluids, the surfaces of these micellae take up water and the previously hard substance becomes soft; (g) upon evaporation the condition is reversed and the micellae again lie alongside.

Nägeli's use of the micellar hypothesis in attempts to explain various phenomena was extensive. Protoplasm, as well as the cell wall, was assumed to be made up of these fundamental units or micellae. Upon the basis of their arrangement in the protoplast there were two kinds of protoplasm, *trophoplasm* and *idioplasm*. The idioplasm was designated as the physical basis of heredity, micellae or "ids" constituting an elaborate network throughout the organism. He assumed that the hereditary characters were brought about by various types of groupings within the network. This "highly speculative" [236] and "fragile" [54] *idioplasm theory* is now of historic interest only, its extreme assumptions having given way to the large accumulation of observational data associated with the cellular aspects of heredity. The continued use of the micellar hypothesis in connection with the structure of plant cell membranes has been discussed by Fischer [66] who states, "For more than thirty years I have observed each new revival of Nägeli's 'swelling theory.' The more consideration I give to the matter, the more I am impressed with the unsolvable contradictions between the well-known hypothesis and the easily observable facts."

In commenting upon disregard of the chemical heterogeneity of the plant cell membrane, such as was shown by Nägeli in the formulation of the micellar hypothesis, Branfoot [33, p. 86] states, ". . . the conclusions of Bracconot, Payen, Fremy, and many others were overlooked, mainly because of the widespread conviction which prevailed at that time that the cell wall was homogeneous, being constituted entirely of cellulose. The recognition of the part played by pectic compounds in plant structures was also hindered by the lack of coöperation between the botanists and chemists of the day, and by the fact that the properties and reactions of these compounds were frequently attributed to cellulose. A state of confusion was therefore inevitable, and in spite of efforts on the part of Guthowsky (1883) and Tschirch (1889), the heterogeneous character of cell membranes was not accepted until Mangin had published his valuable histological memoirs (1888-1894)."

Despite the "unsolvable contradictions" referred to by Fischer, the micellar hypothesis has served as a basis for continued speculation in relation to the structure of plant-cell membranes. An attempt to explain its widespread acceptance in the interpretation of cell-membrane structure and behavior, and its concurrent rejection in the interpretation of the phenomenon of heredity, leads one to the possible conclusion that it is due, not to the "lack of coöperation between botanists and chemists," [33, p. 86] but to comparatively full coöperation of botanists with chemists. In the field of chemical research the workers are accustomed to think and interpret in terms of molecular structure, and thus have been afforded an opportunity to select and emphasize those findings in botanical science which could be most conveniently correlated with their own data. As a consequence, detailed microscopic and microchemical analyses have been obscured in the development of theories of the structure of cell membranes which would account for their gross chemical behavior. In the development of the science of heredity, microscopic structural considerations have taken precedence over interest in the chemical factors involved, although the latter have not been entirely neglected. As a result, the intricate structural changes of nuclei and protoplasm, now recognized as fundamentally important in both vegetative and sexual phases of reproduction, would have little in common with the conception of "idioplasm" advanced by Nägeli.

The micellar hypothesis, as applied to cell membranes, involving the arrangement of submicroscopic cellulose micellae and water in such a way as to produce "water-rich" and "water-poor" areas, brought forth comments of various sorts from Nägeli's contemporaries. Kabsch [140] called attention to the presence of various "incrusting substances" closely associated with the cellulose in the cell membrane. He suggested the possibility that these associated substances are transformation products of cellulose. Hungerbühler,[135] commenting upon the chemical complexity of the membrane suggested, since the growth of cell walls occurs only in the immediate presence of proteins, that cellulose is probably a decomposition product of protein. Dippel [55] and Krabbe [147] could not corroborate Nägeli's findings concerning the "water-rich" and "water-poor" areas. Pfitzer [200] and Müller [181] observed that crystals of calcium oxalate from the protoplasm sometimes adhere to the inner surface of cell walls and are slowly imbedded by the deposition of new layers of cellulose. Klebs [145] and Noll [189] observed that dead protoplasmic particles became imbedded in the cell membrane in a similar manner, and Correns [46] found that, in some instances, even entire chloroplasts are to be found between the wall lamellae. Figures 4b and c illustrates two types of cell-wall inclusions.

Schmitz [227] asserted that the fibrils in the cell membrane are formed from granules in the cytoplasm. Van Wisselingh [270] and Krasser [148] accumulated microchemical data showing that cell membranes may contain proteins, pectins, hemicelluloses, and other materials, and that the lamellae and striae are due to chemical differences in the membrane constituents. Bowman [29] found that certain chemical reagents act unequally upon different parts of the cell membranes of cotton fibers, some portions of which were more readily soluble than others. He concluded [p. 67] that "No cotton fiber is ever perfectly pure cellulose . . . it is probable that on the average there is not more than 87 per cent of pure cellulose." He states,[p. 72] "If a quantity of cotton fiber be subjected to boiling acid for some time, and then dried, it can be rubbed into a fine powder which, under the microscope, exhibits all the features of mechanical and scarcely any chemical disintegration. It seems indeed as if the action were exerted not upon the pure cellulose itself, but upon the various foreign matters which are associated with it in the fiber wall." Again, [p. 64] "Some chemists have supposed that water uptake depends entirely upon the tubular or cellular structure of the fiber so that it is only an instance of capillary attraction; but the fact that if we precipitate the cellulose from its solution in ammonia-cupric oxide solvent, the amorphous jelly which has no capillary structure also possesses this same property, militates against this view."

Although these various workers had approached the problem of cell-membrane structure and composition from many different angles, one gathers that their chief difficulty in adopting the micellar hypothesis rested in its conception of chemical homogeneity of the membrane. The greater number experienced no difficulty in demonstrating the presence of crystalline material in the membrane by means of polarized light, and shared with Nägeli the inability to distinguish smaller microscopic structures than the *fibrils* or *striae*. Their microchemical studies, however, showed a chemical composition of the cell membrane more complex than the cellulose-and-water postulated by Nägeli. Although Russow,[215] Berthold,[25] and Terletzki [256] added further information concerning this chemical heterogeneity, the more comprehensive studies in this connection were those of Tschirch [261] and Mangin.[162, 163] Together these compilations, published shortly before 1890, constituted a fitting close to the half-century of microchemical work which had opened with the researches of Payen [198] and Fremy.[70] Mangin found *lignin*, *suberin*, and other substances in intimate association with the cellulose in many cell membranes which he examined microchemically. A neutral, insoluble substance of pectic nature was invariably present, for which he suggested the name "pectose", on the assumption that it was composed of pectin in combination with cellulose. He found difficulty, however, in

determining the chemical properties of pectose, although microchemically it was distinct from cellulose in its color reactions, behavior toward acids and alkalies, polarized light, and Schweizer's reagent. Mangin concluded that pectose exists in cell membranes along with cellulose from the earliest stages of formation of the

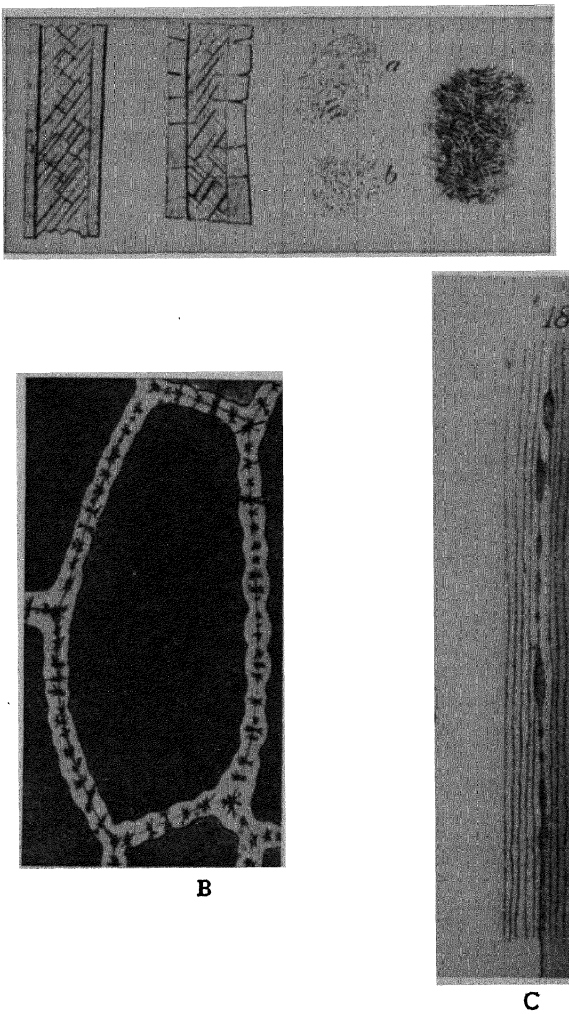


FIGURE 4. A. Disintegration of a cotton fiber into fibrils and *dermatosomes* (after Wiesner). B. Crystals imbedded in the cell membrane of *Allium cepa* (after Molisch). C. Chloroplasts between the lamellae of the cell wall of *Valonia* (after Correns).

membrane. He also suggested that various phenomena associated with cell-wall modifications, such as *gelatinization, cutinization,* and *liquefaction* are due to pectose transformation.

Nägeli's inability to discern microscopic structure in the cell membrane below that of the lamellae and striae led to his conception of the submicroscopic nature of the *micellae*. This was in disagreement with the observations of Lyngbye,[186] Valentin,[363] Schleiden,[226] and others who had preceded Nägeli in this field of research.

Strasburger,[254] after an intensive study of the swelling behavior of mature cell membranes and of the living protoplast during cell-membrane formation, reported that in cells, in which the wall thickening had not yet begun, *microsomes* were distributed at random in the protoplasm. Later on, when there was scarcely a trace of thickening in the wall, the microsomes showed an orderly arrangement in the outer region of the protoplast. He found that this arrangement was sometimes in the form of a spiral which corresponded to the spiral fibrils of the mature wall. In the regions of active wall-thickening the microsomes were quickly used up as they were deposited in successive layers to form the cell membrane: [p. 80] "The final membrane structure produced is similar to a net, the microsomes corresponding to the mesh and not to the lattice."

Strasburger's attempts to determine the chemical nature of the microsomes led to the conclusion that, since they stain yellow with iodine, they are presumably *protein.* A careful comparison of the results of Strasburger with those of the earlier workers shows clearly that the microsomes whose presence and behavior he observed are identical with the *"minutissime punctata"* of Lyngbye,[160] the *"granula minuta"* of Link,[157] the "spheroids" of Agardh,[1] the "ellipsoidal or spherical granules" of Kützing,[152] the "granules" of Valentin,[202] and the "membrane granules" of von Mohl.[176] These observations were concerned chiefly with the membrane structure, and the results differed with those of Nägeli in that they showed, in the course of the formation of cell-membrane lamellae, the deposition of microscopically visible granules from the protoplasm, instead of the intercalation or deposition of submicroscopic *micellae.*

In 1886, Wiesner [268] corroborated Strasburger's findings and suggested that the name *microsomes* be changed to *dermatosomes* because of the important role played by these granular structures in building up the cell membrane. He concluded that the cell membrane is an aggregate of granules or dermatosomes, between which is binding material, and that the dermatosomes are protein while in the cytoplasm, but turn into cellulose after they have been oriented and deposited in the cell wall. The binding substance he believed to be only partially converted into cellulose at the same time. In the mature membranes he found the dermatosomes arranged to form fibrillae, or merely sheaths, through the function of the binding material. By treatment with hydrochloric acid, he succeeded in reversing the process of formation and bringing about the separation of the membrane into lamellae, and the lamellae, in turn, into dermatosomes or into fibrillae composed of single rows of dermatosomes (Fig. 4a).

In 1889, Palladin [193] observed that the growth of cell walls and the formation of starch grains take place only in the immediate presence of proteins or leucoplasts. He concluded that starch and cellulose appear to be transformation products of proteins. This transformation, according to Palladin, takes place in the protoplasm, not after deposition in the cell membrane, as had been supposed by Wiesner.[268]

Schulze and co-workers,[229, 230] upon the basis of extensive chemical investigations of cell membrane constituents, classified them roughly into two groups: the *hemicelluloses,* which can be extracted by 1 per cent HCl or H_2SO_4 and, upon hydrolysis, yield *galactose;* and *true cellulose* which is not soluble in 1 per cent acid and, upon hydrolysis, yields *glucose.* Their detailed analyses revealed the presence, in plant-cell membranes, of hydrates of carbon of many different types. They reserved the name *cellulose* for the particular membrane components, which dissolved in cuprammonium hydroxide, were not attacked by weak acids, and were colored blue with sulfuric acid and iodine and with chlor-zinc-iodide. They emphasized the amorphous nature of cellulose and sometimes designated it as an uncrystallizable body.

Behrens' [22] "The Microscope in Botany," along with treatises by Tschirch [261] and Reiss,[207] expresses clearly the state of confusion concerning the structure and chemical composition of plant-cell membranes which existed at the beginning of the

last decade of the nineteenth century. As Gilson explained in 1893:[86, p. 397] "Certain workers believed that the membrane consists of cellulose, a single substance which was itself a chemical individual; others believed that the membrane consisted of many physical modifications of cellulose; still others designated under the name cellulose substances which were very different chemically."

Few workers who used the microscope in the course of their investigations failed to observe the double refraction of most cell membranes in polarized light. They were thus convinced of the presence of crystalline material and, as had Nägeli, associated this double refraction with the cellulose component. The fact remained, however, that even membranes composed largely of cellulose, as shown by both chemical and microchemical reactions, displayed typical *colloidal* behavior when treated with reagents of many types. Swelling reactions thus produced were often completely reversible, but the phenomena were so striking that such workers as Schulze and his associates were led to emphasize the amorphous nature of cellulose itself.

Gilson[86] was concerned with the growing belief in the amorphous nature of cellulose and attempted to obtain pure crystalline cellulose from its solution in cuprammonium hydroxide. He hoped, in the course of his investigations, to settle the following points: (a) Which of the carbohydrates is it that crystallizes from a solution of the material of vegetable tissue in Schweizer's reagent, when strong ammonia is added? (b) Is cellulose a distinct chemical individual, or are there several compounds in the cell wall that are insoluble in dilute acids and alkalies, and give the blue color with the iodine reagents? (c) Is the cellulose free or in combination in the cell membrane with other constituents? (d) Is the cellulose distributed through all three layers of the cell membrane or localized in certain parts?

Moderately thick sections of the swollen root of *Beta vulgaris* were soaked in 1 per cent KOH (or eau de javèlle) to dissolve out the cell contents. The sections were then washed with distilled water and placed in Schweizer's reagent for 4 to 5 hours. The dissolved cellulose was precipitated in crystalline form by putting the sections into ammonia, after removal from Schweizer's reagent. The size of the crystals were said to vary with the strength of the ammonia, 5 per cent giving only sphero-crystals, and 20 per cent giving beautiful arborescent or radially arranged, spicular crystals. The crystals could be seen best after clearing the sections of copper compounds by washing them with water, treating them with dilute HCl, and then coloring them with chlor-iodide of zinc (or Congo red before treatment with HCl). The crystals were formed in the cell membrane and within the cell cavity. If Schweizer's reagent acted long enough, no evidence of the cellulose remained in the cell membrane, the only blue-coloring substance being the crystals within the cell cavity.

From these studies of the tissue of *Beta vulgaris* and fifty other plants from various parts of the plant kingdom, in which all except the fungi gave satisfactory crystals of cellulose, Gilson[86] concluded that: (a) the crystals precipitated from solution in cuprammonium hydroxide by the addition of strong ammonia are pure cellulose; (b) all that portion of the cell wall that is colored blue by chlor-zinc-iodide, and that portion only, goes to make up the crystals; (c) all cell walls of tissues contain cellulose, but while the internal layer consists almost entirely of it, the intermediate layer has only a small portion, and the intercellular layer only a trace; (d) the cellulose is found crystallized within the cell, showing that its solution in Schweizer's reagent is not very diffusible and gives additional proof that it is derived from the inner layers of the cell wall.

Johnson[189] repeated and corroborated Gilson's work. The crystalline cellulose was readily obtained in cells of *Beta, Dahlia, Lactuca, Typha, Ceratozamia, Equisetum,* and *Chara.* He found that, after keeping sections of the beet in Schweizer's fluid for 20 days, the crystals were as plentiful in the cells as at first, and concluded

outer and inner membrane in the nitrated fiber wall. Between these membranes he found the "intercuticular" substance which appeared to constitute 85 per cent to 90 per cent of the fiber. In an illustration showing a longitudinal section of the fiber this substance is shown to consist of nearly uniform granules about one micron in

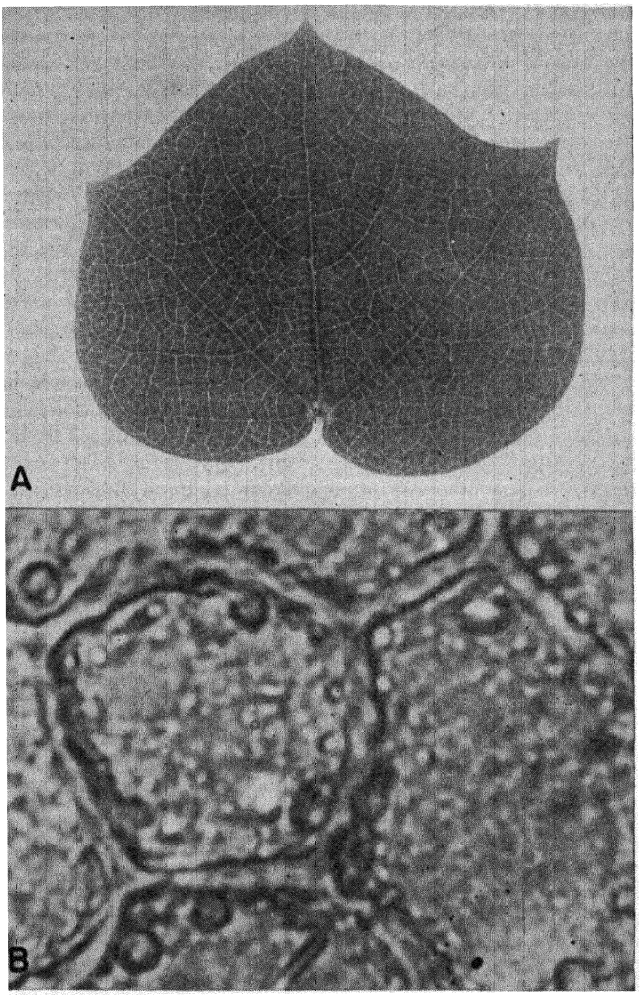


FIGURE 5. A. Young maple leaf showing the venation system which serves the dual purpose of strengthening and conducting water to the leaf cells where photosynthesis is in progress. B. Cells from a maple leaf taken from a maple seed, before germination, showing green chloroplasts in place in the outer regions of the protoplasm.

size:[p 202] "The intercuticular granules lie closely packed, kept in position by the cuticles, and it is only when the tension of these is sufficiently relaxed that the granules can escape."

De Mosenthal found that nitration produced little visible change in the cotton fiber. The outer cuticle wrinkled, but the intercuticular matter remained unchanged. The granules appeared to be somewhat larger, but he pointed out that this was dif-

outer and inner membrane in the nitrated fiber wall. Between these membranes he found the "intercuticular" substance which appeared to constitute 85 per cent to 90 per cent of the fiber. In an illustration showing a longitudinal section of the fiber this substance is shown to consist of nearly uniform granules about one micron in

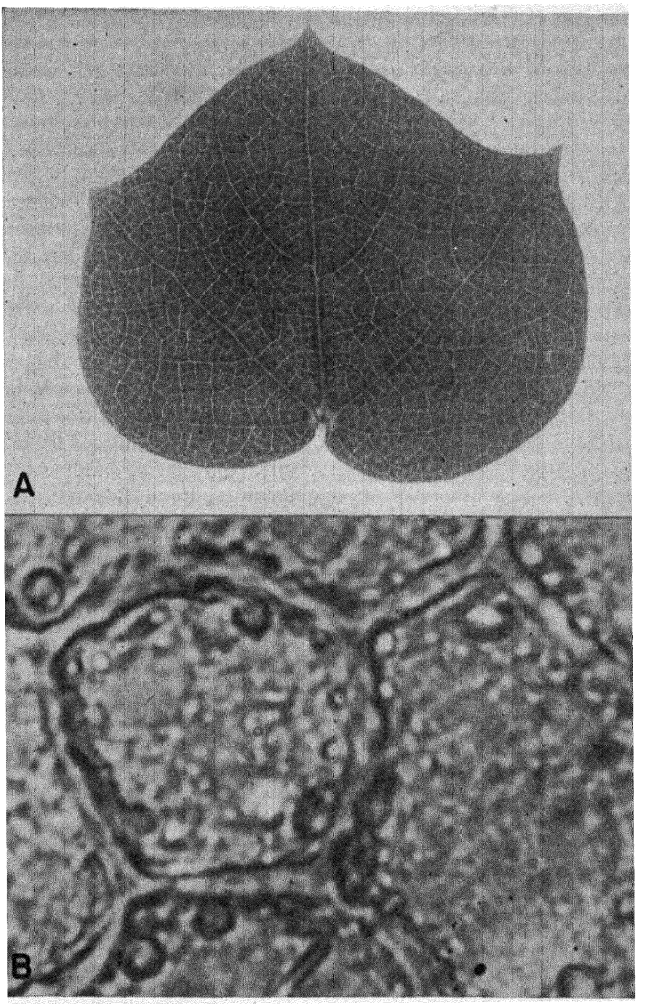


FIGURE 5. A. Young maple leaf showing the venation system which serves the dual purpose of strengthening and conducting water to the leaf cells where photosynthesis is in progress. B. Cells from a maple leaf taken from a maple seed, before germination, showing green chloroplasts in place in the outer regions of the protoplasm.

size:[p 202] "The intercuticular granules lie closely packed, kept in position by the cuticles, and it is only when the tension of these is sufficiently relaxed that the granules can escape."

De Mosenthal found that nitration produced little visible change in the cotton fiber. The outer cuticle wrinkled, but the intercuticular matter remained unchanged. The granules appeared to be somewhat larger, but he pointed out that this was dif-

stage in the formation of proteins and other more complex organic materials. The name *photosynthesis* was applied to the process by Hansen as related by Barnes [18] in 1898.

Studies of the structure, composition, and methods of formation of plant-cell membranes played an important role in bringing about a clearer understanding of the more general aspects of photosynthesis. The chemical processes involved require energy which the living cell obtains from light. Evidences of adaptation to the utilization of this source of energy are to be found in almost every green plant. If, for example, the cells of a tree, above its roots, were arranged in the form of a sphere, it would be able to produce very little food, since the light rays could reach only the cells near the surface. The tree increases its effective light-receiving area many thousand-fold by bearing high above the ground, upon its strong, tapering trunk, branches which, in turn, bear thousands of fan-like leaves. The cells of the leaves are thus held in positions in which they are able to receive sufficient sunlight to carry on their usual production of food materials from substances provided by the air and soil. The strength and rigidity necessary to maintain such a structure (sometimes hundreds of feet above the surface of the earth) could not be produced either with sugar, which is usually in solution in the cell, or with soft, loose granules of starch into which some of the sugar is transformed and stored as reserve food. Other portions of the sugar are converted, therefore, into more durable cellulose, which is used, for the most part, in the construction of sturdy walls around the cells.

The requirements placed upon these cell walls by the living organisms, of which they are a part, are exacting. When young, they must have sufficient plasticity to permit cell enlargement in one or more directions. As the cell matures they usually become more rigid. In all stages of development they must be somewhat flexible, must permit the transportation of dissolved substances to and from the protoplast and, in the food elaborating regions, at least, must be translucent. The cell walls of the tree leaf meet these requirements, as shown in Fig. 5. Fiber cells, such as those which grow upon the cotton seed, and bast or wood fibers, which are produced abundantly in stems of many of the higher plants, do not carry on the primary carbohydrate synthesis. Sugar, which is translocated from the green tissues of the plants in which it is formed, is converted largely into cellulose by the protoplasts of these cells and is used up in the construction of the thick, fiber cell walls. In the mature cotton fiber the proportion of cellulose to other cell-membrane constituents is great, amounting, in different varieties, to from 80 per cent to 90 per cent of the total mass. In mature wood and bast fibers the proportion of cellulose is much less, sometimes being reduced to from 40 per cent to 50 per cent of the wall substance. This comparatively high proportion of cellulose, as compared with the cellulose content of cell membranes in general, has led, however, to a synonymous use of such terms as *cotton fibers, linen fibers, spruce wood,* and *cellulose.* Properties of these fiber masses are commonly referred to as properties of *"cellulose."*

Some of the characteristics of cell membranes which have come to have singular industrial importance first functioned in the development of the living cell. From this viewpoint the transformation of a mass of dried fibers into a continuous plastic matrix is a form of a reversible reaction reminiscent of the time when each fiber membrane itself was capable of easy deformation. Many of the reactions which bring about this change in physical state of the mature cell membranes are currently described as "cellulose derivative formations." At the opening of the present century these reactions were the subject of considerable thought and discussion. Representative of the current thought of the time was the paper presented before the Society of Chemical Industry by de Mosenthal.[180] His studies had covered the microscopic structure of raw cotton and the miscroscopic properties of cotton fibers which had been transformed into a plastic mass either by the usual methods of nitration or by treatment with cuprammonium hydroxide. He found evidence of an

stage in the formation of proteins and other more complex organic materials. The name *photosynthesis* was applied to the process by Hansen as related by Barnes [18] in 1898.

Studies of the structure, composition, and methods of formation of plant-cell membranes played an important role in bringing about a clearer understanding of the more general aspects of photosynthesis. The chemical processes involved require energy which the living cell obtains from light. Evidences of adaptation to the utilization of this source of energy are to be found in almost every green plant. If, for example, the cells of a tree, above its roots, were arranged in the form of a sphere, it would be able to produce very little food, since the light rays could reach only the cells near the surface. The tree increases its effective light-receiving area many thousand-fold by bearing high above the ground, upon its strong, tapering trunk, branches which, in turn, bear thousands of fan-like leaves. The cells of the leaves are thus held in positions in which they are able to receive sufficient sunlight to carry on their usual production of food materials from substances provided by the air and soil. The strength and rigidity necessary to maintain such a structure (sometimes hundreds of feet above the surface of the earth) could not be produced either with sugar, which is usually in solution in the cell, or with soft, loose granules of starch into which some of the sugar is transformed and stored as reserve food. Other portions of the sugar are converted, therefore, into more durable cellulose, which is used, for the most part, in the construction of sturdy walls around the cells.

The requirements placed upon these cell walls by the living organisms, of which they are a part, are exacting. When young, they must have sufficient plasticity to permit cell enlargement in one or more directions. As the cell matures they usually become more rigid. In all stages of development they must be somewhat flexible, must permit the transportation of dissolved substances to and from the protoplast and, in the food elaborating regions, at least, must be translucent. The cell walls of the tree leaf meet these requirements, as shown in Fig. 5. Fiber cells, such as those which grow upon the cotton seed, and bast or wood fibers, which are produced abundantly in stems of many of the higher plants, do not carry on the primary carbohydrate synthesis. Sugar, which is translocated from the green tissues of the plants in which it is formed, is converted largely into cellulose by the protoplasts of these cells and is used up in the construction of the thick, fiber cell walls. In the mature cotton fiber the proportion of cellulose to other cell-membrane constituents is great, amounting, in different varieties, to from 80 per cent to 90 per cent of the total mass. In mature wood and bast fibers the proportion of cellulose is much less, sometimes being reduced to from 40 per cent to 50 per cent of the wall substance. This comparatively high proportion of cellulose, as compared with the cellulose content of cell membranes in general, has led, however, to a synonymous use of such terms as *cotton fibers, linen fibers, spruce wood,* and *cellulose.* Properties of these fiber masses are commonly referred to as properties of *"cellulose."*

Some of the characteristics of cell membranes which have come to have singular industrial importance first functioned in the development of the living cell. From this viewpoint the transformation of a mass of dried fibers into a continuous plastic matrix is a form of a reversible reaction reminiscent of the time when each fiber membrane itself was capable of easy deformation. Many of the reactions which bring about this change in physical state of the mature cell membranes are currently described as "cellulose derivative formations." At the opening of the present century these reactions were the subject of considerable thought and discussion. Representative of the current thought of the time was the paper presented before the Society of Chemical Industry by de Mosenthal.[180] His studies had covered the microscopic structure of raw cotton and the miscroscopic properties of cotton fibers which had been transformed into a plastic mass either by the usual methods of nitration or by treatment with cuprammonium hydroxide. He found evidence of an

ficult to verify through measurement. If a "solution" of the nitrated cotton were allowed to evaporate upon a glass slide, the resulting film appeared granular under the miscroscope, the individual granules being about one micron in size. He found a difference between the cotton regained from solutions of $ZnCl_2$ or $ZnCl_2 + HCl$ as compared with that which had been treated with cuprammonium hydroxide. In the former the granules were destroyed, in the latter they were preserved: [p. 298] "Microscopic examination shows the granules freed and washed out as the cuticles yield to the swelling of the fiber." De Mosenthal inquires: [p. 295] "Are we to regard these solutions as colloids, as suspensoids of the finely divided particles? The microscopic examination of films and precipitates certainly points in that direction. That no particles can be seen in the liquid does not militate, I think, against that view. According to Lottermoser, particles must be smaller than 0.5 micron to be microscopically invisible in solution and our particles are much larger. But surely the detection of minute particles in a liquid does not only depend upon their size, but also on the relative refractive index of the suspended matter and the liquid." During the discussion of de Mosenthal's paper the Chairman remarked that they had been wrong from the beginning in treating cellulose as a chemical entity; that it is a complicated substance; and that cellulose from various sources behaves very differently. De Mosenthal closed the discussion with the statement that the members had discussed *cellulose,* while he had discussed *cotton.* There is little doubt that he recognized their failure to regard the fiber as an organic unit, designed to meet biological requirements according to a microscopic plan which had been observed by Valentin,[262] Lyngbye,[160] Link,[157] Agardh,[1] Strasburger,[255] Wiesner,[268] and others, as well as himself. Since his interest was focussed upon the colloid behavior of treated fibers and his methods of approach primarily chemical, he probably was not aware of the concurrent interest of plant physiologists in the microscopic structure and behavior of living cell membranes.[199]

Contemporary publications showed the effect of the impetus to research in the field of cell-membrane structure and composition which had been given by the nineteenth-century investigators. In 1901, Allen [2] published upon the origin and nature of the middle lamella. One of the few general reviews of the botanical aspects of cell-membrane structure was contributed by Gaucher [83] in 1904. Remec,[209] Herzog,[114, 113] and Behrens [21] extended the studies of fiber structure by means of polarized light. Tröndle [260] made a detailed microchemical study of the cell membrane of *Spirogyra.* Gaidukov [82] reported advances in the use of the ultramicroscope and the microspectral photometer. Lester [155] extracted a "brown hygroscopic substance" from American cotton with water. Herzog [113] counted twenty-eight concentric layers in a fiber wall which had been swollen with cuprammonium hydroxide. Bowman [29] and Flatters [67] discussed various developmental aspects of cotton fibers, and Hanusek [104] summarized the microscopic properties of fibers which were of importance to the rapidly growing textile industries. He comments: [104, p. 61] "Like all epidermal cells, cotton hairs are covered with an exceedingly thin, structureless cuticle which is insoluble in cuprammonia. The presence of cuticle serves to distinguish cotton from all bast fibers. On the inner surface of the cuticle are fine granules and striations which are particularly conspicuous when the fibers are examined after immersion in dilute ammonia water." In 1903 a treatise by Ewart [56] "On the Physics and Physiology of Protoplasmic Streaming in Plants" appeared in the Oxford Press and is worthy of mention here because of the more recent considerations of the influence of protoplasmic streaming upon the deposition of materials from the protoplasm.[188, 285]

During the second decade of the twentieth century the need for a better understanding of fibrous materials, particularly with respect to their reactions to dyes and swelling agents, led to numerous studies of their chemical composition. Schunck [231] had found in 1868 that less that 0.5 per cent of the substances other than cellulose

were contained in unbleached cotton yarn. From this small amount of material he had identified cotton wax, margaric acid, alcohol-soluble material, alcohol-insoluble material, pectic acid, and albuminous matter. In 1915 Levine[156] obtained results of chemical analyses at different stages of fiber development showing much larger quantities of non-cellulosic material. Fats and waxes alone amounted to 2.2 per cent, and protein to 0.18 per cent. In 1918 Knecht and Hall stated:[146, p. 220] "Since the amount of impurities in raw cotton was known to be at least ten times greater than that accounted for by Schunck, and it was considered that prolonged boiling with alkali might bring about more or less drastic changes in them, it was decided to proceed as far as possible by way of systematic extraction with purely mechanical solvents." By these methods of extraction the amount of "impurities" found in American cotton totaled 4.327 per cent and in Egyptian cotton 4.124 per cent. The alcoholic extract (0.533 to 0.744 per cent) was an amorphous, brown, very hygroscopic, faintly acid substance containing as much as 43.4 per cent of ash and exerting a powerful reducing action on Fehling's solution. The aqueous extract (1.509 to 1.656 per cent) was likewise amorphous and brown, slightly less hygroscopic, and less effective in reducing Fehling's solution. This extract was also faintly acid and contained as much as 54.5 per cent of ash. The lime extract (2.1 per cent) was found to consist of protein, pectic acid, and fatty acid, as well as two resinous substances. A contemporary report upon the physical and chemical properties was made by Harrison.[105] Studies carried out by Minaev[173] upon developing cotton fibers showed that in the membranes 5-day fibers contained no cellulose deposits; 10-day fibers assumed a violet coloration with chlor-zinc-iodide; 20-day fibers contained cellulose in abundant amounts.

A paper announcing the existence of daily growth rings in the cell walls of cotton fibers was published by Balls[17] in 1919. Egyptian cotton of different stages of development preserved for five years in 30 per cent alcoholic solutions of glacial acetic acid were not sufficiently swollen to make clear the lamellate structure of the wall. Therefore, the Cross and Bevan[48] method involving treatment with sodium hydroxide and carbon disulfide was used; this enabled him, through correlations with climatological data, to draw the following conclusions: (a) the primary wall of seed hairs in cotton contain very small amounts of cellulose; (b) secondary thickening of the wall proceeds intermittently under normal Egyptian field conditions, being arrested every afternoon; (c) the cellulose of the hair consequently consists of a number of concentric shells, layers, or growth rings, each one representing one day's growth, with the exception of the primary wall; (d) counts of rings in the lint could never be proved to exceed 25; the smallest figure obtained was 20; (e) secondary thickening begins in the wall on the 21st to 25th day and continues until 25 daily depositions have taken place; (f) the thickness of each of the 25 "growth rings" can only be about 0.4 micron; (g) the hairs are covered outside the cellulose wall by a cuticle, bearing wax, which is structurally and historically identical with the cuticle of the testa; (h) the secondary wall, but not the primary wall (cuticle) is traversed obliquely to the axis by simple pits which are rarely visible except in the living hair, and to these pits is due the twisting of the hair and its characteristic convolutions after death.

Reissek[208] in 1852 had observed the wall "lamellae" in cotton fibers and shows in his illustrations at least nine, in longitudinal view. About the same time, Crum[50] cross-sectioned the fibers, and made drawings showing four and five concentric layers. Butterworth[86] and Minajeff[174] had also preceded Balls in the study of cross-sections of fibers; and Herzog,[116] five years earlier, had reported as many as 28 lamellae in the cotton-fiber wall after it had been swollen with cuprammonium hydroxide. Balls made the first attempt to correlate the lamellate depositions in the wall with a daily periodicity. His original premises were that no cellulose is deposited in the wall until after 20 days of fiber development and that during the re-

maining 20 to 25 days of fiber growth, one layer is deposited every day. More recent studies, to be described later, have shown the presence of cellulose in the protoplasts of 1-day fibers and its lamellate deposition in the walls of 3- to 5-day fibers. These observations have called for drastic alterations in Balls' original premises. With the total period of fiber development covering approximately 50 days, there is needed almost double the number of "daily growth rings" postulated by Balls.

A few of the many papers which are representative of the cell-membrane studies of this period are those published between 1909 and 1920 by A. Herzog [113, 114, 115, 116, 117] and Harrison.[105, 106] Their studies included microscopic observations in both ordinary and polarized light and reactions in cuprammonium hydroxide and other swelling agents. The effects of mechanical treatments were described by Cross and Bevan,[47] who found that cotton rope used for power transmission was converted, at the core, into rounded, hyaline masses of small dimensions; thus the "fibrous colloid had been transformed through a phase of hydrate fusion into the structureless form in question." Fort,[68] who followed the breakdown of cotton fabrics under "beetling" treatment, reported that a "powder" of cellulose was obtained which was soluble in alkali and possessed microscopic features different from "shredded" fabric. The importance of the colloidal properties of various types of cell membranes, emphasized earlier in the work of Chardonnet,[59] was evidenced anew in the work of Ost,[192] Gibson, Spencer, and McCall,[85] Wheeler,[264] and Schorger.[228] Among the books of the period destined to play an important part in stimulating and shaping investigations of the future were "The Textile Fibers" by Matthews,[166] "Researches on Cellulose" by Cross and Bevan,[48] "Researches on Cellulose. IV," by Cross and Dorée,[49] "Mikrochemie der Pflanzen" by Molisch,[179] and "Biochemie der Pflanzen" by Czapek.[51] Botanical researches dealing with cell-membrane structure and composition carried on during the first quarter of the present century are summarized by Sharp [236, p. 218] as follows: "As a general rule, the primary layer seems to be largely pectose, the secondary and tertiary layers being made up of cellulose together with pectose and other materials. It is only rarely, however, that pectose and cellulose exist in anything approaching the pure form. The pectose of the middle lamella soon changes to insoluble pectates, chiefly that of calcium, while the other layers become greatly altered in composition by chemical transformation and the addition of new materials. A certain amount of protein matter seems to be present in all the layers. In the apical meristems of *Vicia faba* Tupper-Carey and Priestley (1923) find the middle lamella to be probably a mixture of pectin and protein and that the thickened walls, although they all contain cellulose, may not react as cellulose because of the presence of proteins and other substances. . . . The presence of proteins is noteworthy in connection with the question of the relation of the protoplast to the wall."

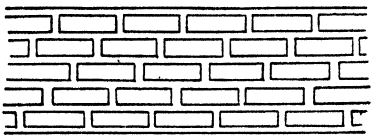
The investigations upon which these conclusions are based are thoroughly reviewed in the writings of Czapek,[51] Molisch,[179] Gleisberg,[87] Grafe,[89, 90] van Wisselingh,[271] and Hansteen-Cranner.[100, 101, 102, 103] The last author, in a series of investigations of the toxic effects of K, Na, and Mg ions on root growth, found that the toxic effect was due to changes produced in the cell walls and not in the cell protoplasm. In the course of subsequent investigations, the cell walls were found to contain lipoid substances, and walls of all living plant cells to contain comparatively large amounts of phosphatides. Grafe [91, 92] and Grafe and Magistris [93] corroborated these findings. Hansteen-Cranner,[102] unable to find any break in the continuity of phosphatides in the protoplasm and the wall of a cell, developed the conception that the wall as well as the protoplast is a living structure. Kisser [144] studied the relation of *cutin* to *cellulose* in cell membranes; van Wisselingh [271] demonstrated an intimate relation between *suberin* and *cellulose;* and Anderson [8, 9] showed that in collenchyma cell walls there is an alternation of pectic and cellulose

lamellae; he demonstrated an intimate relation between *cellulose, pectic material,* and *cutin* in the epidermal cell walls of *Clivia nobilis.* He concluded,[9, p. 312] "From the standpoint of microscopic structure, the wall is a complex of several chemically different lamellae. The lamellae may consist of one chemical substance, and alternate with lamellae of other cell wall constituents, or the various constituents may be present in the same lamellae."

The chemical heterogeneity of cell membranes, thus evidenced at all stages of their development, constituted the basis for the rejection of the concept of chemical homogeneity as advanced by Nägeli in the micellar hypothesis. The inadequacy of the conception of a structure composed of crystallites separated by films of water was commented upon at length by Fischer [66] in 1927, who pointed out that the phenomena of dyeing, swelling, and semi-permeability are not explainable upon the basis of the membrane structure proposed by Nägeli.

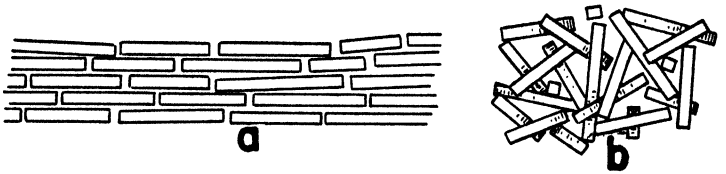
From the standpoint of the presence of crystalline cellulose in the cell membrane, Nägeli's hypothesis received, in the meantime, excellent corroboration. The extensive researches of Ambronn,[5] Ambronn and Frey,[6] and Frey,[70, 77, 78] by means of polarized light, led to the following conclusions: (1) by optical methods with the help of double refraction one can find that there are present in the cell membrane distinct cellulose micellae which must have an elongated form; they are strongly anisotropic and behave optically as a rhomboid crystal which shows the main refractive indices n_γ, n_β, n_α; n_β is only weakly differentiated from n_α and, upon the basis of many measurements, the values $n_\gamma = 1.594$ and $n_\beta \backsim n_\alpha = 1.533$ were obtained; (2) the micellae themselves are doubly refractive and this property permits detailed studies of their orientation in polarized light (Fig. 6a to d) [Ambronn's previous work [4] with "Stabchendoppelbrechung" and "Eigendoppelbrechung" provided information which was used in checking the nature of the double refraction produced by the cellulose micellae]; (3) the cellulose micellae possess normal dispersion of the double refraction and permit the passage of ultraviolet light. Cutin does not permit the passage of ultraviolet light and this method of study is valuable in differentiating these two membrane components; (4) the relative positions of cellulose, cutin, pectin, and lignin in the cell membrane were determined by means of observations involving both polarized and ultraviolet light; (5) the theory of Nägeli that the cell membrane is built of long, submicroscopic, optically anisotropic, crystalline micellae is valid.

The crystalline nature of the cellulose constituent of cell membranes received additional confirmation from the studies made by means of the newly developed technique, x-ray diffraction analysis. The methods described by Laue,[153] Bragg,[82] and Debye and Scherrer [58] were quickly adapted to the study of cell membranes and, as early as 1913, Nishikawa and Ono [188] had obtained symmetrical diffraction patterns from wood fibers and bamboo. Scherrer [222] corroborated these findings in 1920 and more complete confirmation was provided in a series of studies by R. O. Herzog and co-workers [124, 118, 119, 123, 122, 121] between 1920 and 1926. Herzog studied both untreated and treated fibrous materials in swollen and unswollen states, with and without tension. The definite crystalline nature of cellulose was thus established and, in addition, the significant point that cellulose from various sources in the plant kingdom produces the same x-ray diffraction pattern. From comparisons of many diffraction patterns Herzog pointed out that the results of x-ray diffraction analysis confirmed the detailed studies of fibers in polarized light; that the cell walls are made up of fibrillae which, in turn, are made up of a number of parallel crystallite chains; and that the crystallite is the unit revealed by x-rays. He believed that the fibrillar structure is the result of tension developed in the living fiber during growth and that the $C_6H_{10}O_5$ groups are linked together according to Werner's theory of secondary valence. Microscopic studies of colloidal materials reported by Freundlich and co-workers [75, 76] during the same period furnished information which has



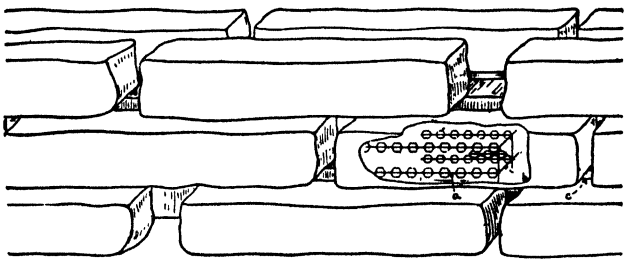
A

A. Diagram of micellar orientation according to Nägeli (after Frey).



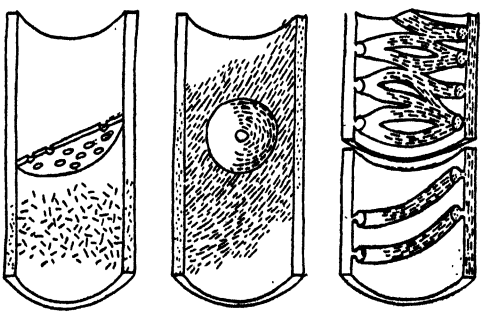
B

B. Parallel and random arrangement of micelles (after Mark).



C

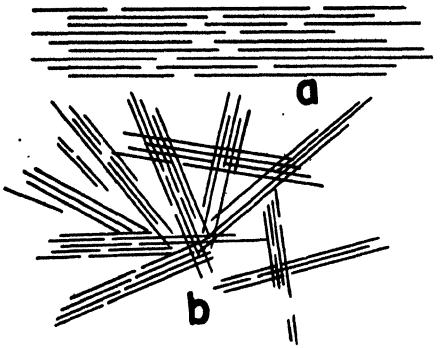
C. Model of cellulose structure as deduced from x-ray data showing brick-like micells, the interior of one of which is exposed and enlarged to show the chains of glucose residue units; primary valence forces, secondary association forces, and tertiary micellar forces are indicated by a, b, and c, respectively (after Clark, diagram by Seifriz).



D

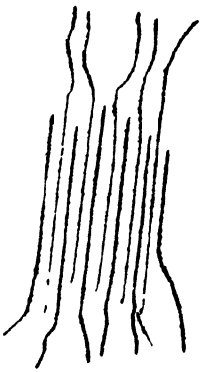
D. Orientation of crystalline micelles in different cell walls (after Frey).

FIGURES 6A, B, C, D



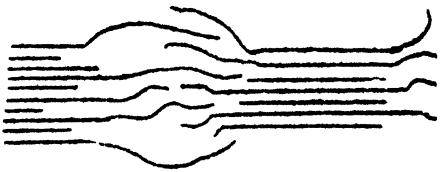
E

E. Parallel arrangement of chain molecules (a), and random arrangement of molecular aggregates (b) (after Carothers).



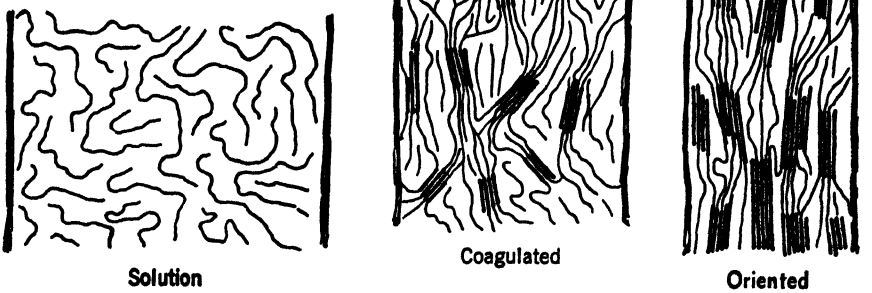
F

F. Fringed micells according to the new theory (after Kratky).



G

G. Joining of two micelles by intertwining of fringes through Van der Waal's forces (after Kratky).



H

H. Behavior of cellulose chains in a rayon spinning solution, coagulated and oriented (after Mark).

FIGURES 6E, F, G, H

proved to be helpful in attempts to correlate data obtained by means of x-rays and the microscope in dark field, and in ordinary and polarized light.

These various physical measurements of cell membranes which had served to establish the crystalline nature of their cellulose component had also contributed further evidence toward their chemical heterogeneity. Even the x-ray diffraction patterns bore evidence of amorphous material in addition to the cellulose crystallites. Herzog [120] in 1925 described its relation to the cellulose crystallites as an "imbedding substance." Anderson [9, p. 306] pointed out in this connection in 1928: "These facts have made necessary a modification of the original Nägeli theory. It is highly probable that the spaces between the micellae are not empty, as Nägeli believed, but are filled with a highly hydrophilic colloid. The general presence in cell walls of the hydrophilic colloidal phosphatides discovered by Hansteen-Cranner is of undoubted importance in this respect. Other workers have suggested other colloidal substances. In young membranes this colloidal intermicellar substance may be of pectic nature. In young *Closterium* walls one can see the double refraction of the cellulose gradually appear and increase in the isotropic membrane (Frey). In cellulose walls an amorphous cellulose may form the colloidal matrix (Herzog). In cutinized and suberized membranes, it may be the cutin and suberin (Frey), and in lignified walls the lignin. That lignin is *between* the cellulose micellae and not *in* them is proven by the fact that lignified walls give the x-ray pictures of pure cellulose (Frey)."

During this same year a paper by Hess and co-workers [128] reported the presence of a nitrogenous foreign substance in ramie and cotton fibers which was found to have a high ash content and to swell strongly in pyridin. They stated that the material is frequently held fast during the purification of the fibrous mass and questioned whether or not it is arranged inside the fiber wall. This substance was found to be the bearer of the ash in the fiber, and an analysis of the ash showed that it was almost entirely calcium oxide.

These accumulated data were sufficiently impressive to warrant the modification of the micellar hypothesis of Nägeli, according to the suggestions of Frey [77] and Anderson.[9] With the presence of a colloidal matrix between the cellulose crystallites, many of the observed physical and chemical properties of the membrane could be explained, and the objections to the hypothesis as a "working" conception largely overcome. During the development of these modifications the idea of the original hypothesis that the crystallites of cellulose are submicroscopic remained unchanged. The extensive studies of Ambronn and Frey revealed no microscopically visible crystalline units below the fibril, and the x-ray diffraction analyses could furnish no conclusive evidence concerning the existence of a structural unit between the cellulose unit cell and the fibril. Spierer [241] reported in 1926 that fragmentation of the wall striations into "supermicellae" of microscopic dimensions had been observed by means of the Spierer lens. The optical properties of this lens were said to be of such a nature that both microscopic and ultramicroscopic structure of the wall were revealed. Seifriz [235, p. 152] writes: "The supermicells seen with the Spierer lens are 0.5 to 0.7μ in thickness and 1.2μ in length. Such particles are above the lower limit of microscopic visibility by direct light, and are, therefore, not ultramicroscopic in the true sense, yet they are not visible with direct light; in this latter sense only are they ultramicroscopic. Lack of optical differentiation between particle and matrix makes dark-field observation necessary to show the particles." These studies with the Spierer lens were extended by Thiessen [258] to various types of cellulosic material including coal. Clifford and Cameron [41] questioned the optical properties of the lens and presented evidence to show that the supermicells in cell membranes described by Spierer, Seifriz, and Thiessen were artifacts. Although by this time, the existence of wall striations and of separable wall fibrils was evidenced from many types of study, the conception of the submicroscopic units, as the building blocks of the fibril, prevailed (Figs. 6a, b, c, d).

The existence of submicroscopic micellae in the cell membrane was, however, not without question. Fischer [66] commented upon the uncertainties as follows: "One sees upon an object a special optical effect, as for instance double refraction, and one is persuaded, upon the authority of Nägeli, that the material is made up of small parallel-arranged bodies of crystalline nature. Therein is the premise: *The objects of this internal structure lead to this special effect of double refraction.* Now nothing lies nearer than the converse: *This phenomenon of double refraction shows that such an inner structure must exist.* This, however, may be a fallacy; we cannot prove the matter experimentally; we cannot produce such a body which is made up of diminutive parallel arranged crystallites and then study with polarized light or x-rays, a body, of which we know, with full certainty, that it has this microstructure, because we ourselves have produced it."

Sponsler [242, 243] avoided many of these structural difficulties through the development, in 1925 and 1926, of a conception of cell-membrane structure which would account for its known properties upon a molecular basis. In the course of his investigations he obtained the first approximate dimensions of the cellulose unit cell and expressed the belief that the diffraction data obtained from x-ray analysis of cellulose may be accounted for upon the basis of chains of cellulose unit cells of indefinite length, thus eliminating the necessity for consideration of micellar aggregates. In a later study of the growth of cross walls in certain algae,[245] he found that the deposition of cellulose takes place in the form of molecular layers at the rate of 20 molecular layers per second. Sponsler and Dore [247] developed more fully the conception of the cellulose chain structure, and Sponsler [244] revised their conclusion regarding the nature of the union between glucose residues in favor of a chain structure which would accommodate the cellobiose linkage. Sponsler and Dore [248] pointed out later, however, that the problem of accounting for the odd order reflections of the long identity period still remained unsolved. Staudinger,[250] on the basis of viscosity measurements, likewise postulated the existence of very long chains of cellulose which he believed to be of sufficient size to warrant the name *macromolecule*. Hengstenberg and Mark,[110] recognizing the existence of cellulose chains, questioned their indefinite length, however. They expressed the opinion in 1928 that the cellulose chains are comparatively short, and are arranged in the form of a bundle or micellar unit. Their x-ray measurements indicated crystalline dimensions of $50 \times 50 \times 500$ Å. These measurements were corroborated in 1930 by the x-ray analyses of Clark [40] and in the chemical analyses of Haworth and Machemer,[109] but in 1933 Sponsler [244] suggested alternative interpretations of their data involving a long-chain rather than a micellar structure.

Carothers,[38] in a discussion of polymeric compounds, suggested a new type of organization of chain molecules in the walls of plant fibers in which the long axis of the chains is approximately parallel to the long axis of the fiber, with pronounced overlapping of the ends of the chains (Fig. 6e). In a review by Kratky [151] the evolution of the most recent concept of cell-membrane structure, which has points in common with both micellar hypothesis and the chain hypothesis, is presented. The idea of *fringed micells* (Fig. 6f) is based upon the work of many authors and has been defined by Thiessen [257] as follows: "The cellulose micelle is an ultramicroscopic mixed crystal of cellulose chains, differing in length. Ends of chains project beyond the micelle ends formed by shorter chains (fringed micelle). Because of the shorter filaments interposed in the micelle core the fringes have such large lateral distances that van der Waal's forces no longer hold them together. For this reason they tend to separate."

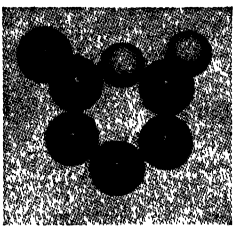
As shown in Fig. 6f, the chain molecules, according to this conception, overlap in such a way that in certain areas they are entirely regular. In other areas they are irregular and less dense. Alternation of these regular and irregular chain areas, in the fashion shown in Figs. 6g and h, results in the distribution of crystalline and

amorphous areas throughout the membrane—an interpretation in increasing demand as data concerning the heterogeneous physical and colloidal properties of cell membranes continue to accumulate. Freund and Mark [73] state, "All experimental evidence of the last years points in the direction that all high polymers in the solid state are complicated two (or even multi-) phase systems." The behavior of such systems during solution, coagulation, and orientation is shown diagrammatically in Fig. 6h.

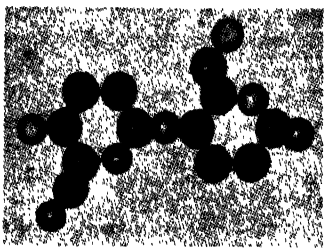
The voluminous literature which has accumulated during the past twenty-five years upon the physical and chemical properties of plant-cell membranes cannot be adequately represented within the limits of this presentation. It must suffice to cite briefly only certain aspects of the experimental approach to an understanding of the complex problem, and certain conclusions which have been drawn from the data obtained. Various viewpoints of the structural elements within *micellae* were published between 1929 and 1939 by Staudinger and Singer,[252, 249] Gerngross and coworkers,[83, 84] Frey-Wyssling,[80] Kratky,[149, 150] Mark and Meyer,[164] Guth and Rogovin,[97] and Sauter.[220] In a paper presented before the members of the Society of Dyers and Colourists in 1932, Neale,[186] however, maintained that the conception of *micellae,* or separate units between glucose residues and the cotton fiber, is not necessary; that a discontinuous micellar structure does not explain phenomena such as axial shrinkage and lateral swelling; and that such a viewpoint necessitates an arbitrary distinction between inter- and intra-micellar swelling. In the discussion following Neale's presentation, Katz pointed out that the fact that cohesion is not destroyed when liquid penetrates into the interior of a body capable of swelling is perhaps one of the most surprising facts that we know about the phenomenon of swelling. He raised the question why, since a precipitate of barium sulfate, for instance, does not swell, every precipitate of cellulose, starch, protein, etc. shows swelling and cohesion? Fischer [66] had made a similar comment in 1927, using colloidal platinum as an example of a non-swelling precipitate. To these questionings Khouvine [142] added the comment, in 1934, that an attempt to define the word *cellulose* merely serves to remind us that it is a substance which we are not able to crystallize, which has no fusion point, which will not dissolve in ordinary chemical solvents, whose structural formula has been somewhat evasive, and whose molecular weight is definitely uncertain.

Many of these difficulties were avoided by the development of new trends in physical-chemical investigations of natural "high polymeric" organic substances, clearly described in 1930 by Meyer and Mark.[172, Chap. 5] They pointed out that while the usual goal in classical organic chemistry is the development of a structural formula which would lead to knowledge of a method of synthesis, one is not satisfied with this in the study of high polymers; that it throws too little light upon the mechanical properties of these important substances; and that the present trend in physical chemistry is to model the structural formulas of organic chemistry in a way which is accurately descriptive of the size and form of the molecule and is also representative of the physical characteristics of the crystalline and fluid combinations under consideration. Beginning with the structural conception of the molecules of such a substance as cellulose or starch (Fig. 7), hypotheses were to be formulated as to their relative positions and the forces operating between them, and models were to be constructed which, if possible, would correlate with the characteristics of the substances. This method of approach has led to the current conception of plant-cell membrane structure upon a molecular basis. Before discussing concurrent developments in the field of cellulose analysis, it may be well to repeat these general results, as applied to cell membranes, summarized by Frey-Wyssling [81] in 1938: "A revision of the Nägeli hypothesis of discontinuous cellulose micellae was found to be necessary. Evidence was found of a system of connected cellulose threads or layers. Cellulose, in the form of long, chain-like molecules formed a skeletal framework.

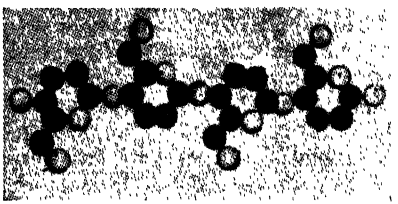
The molecular weight and length of the chains are not definitely known. They give a crystalline x-ray pattern, but are never seen in crystalline form. They are insoluble in all ordinary chemical solvents, but in reagents which will throw them into a viscous matrix, their viscosities show that a fairly close relation exists between the strength of the whole fiber and the length of the cellulose chain molecules; maximum strength is believed to be reached in chains of about 2000 glucose residues." These concepts were quickly adapted to the interpretation of the behavior of industrial materials.[99, 37]



Skeleton model of glucose



Cellobiose skeleton



Two cellobiose units assembled as in cellulose

FIGURE 7. Structural models (after Haworth).

In the meantime, data of various types were accumulating in other laboratories which were to lead to different conceptions of membrane structure. Boehm [26] reported in 1933 that woods or woody materials could be disintegrated and remade by using as a binder the *non-cellulosic* material, which bonds the fibers in the original wood, resulting in a product with the original strength of wood and with fibers so oriented that expansion and swelling were greatly reduced. Butler [85] found the crude fiber content of the cell walls of certain marine algae sometimes used as cattle food, fertilizer, and a source of nitrogen, to be very low. In only one instance was it over 5 per cent (*Fucus* 5.21, *Laminaria* 3.61, *Porphyra* 2.75, *Chondrus* 2.15, *Gigartina* 2.05, *Rhodymenia* 1.50). Protein, ash, and other carbohydrates constituted the greater part of these sturdy membranes of interesting physical and colloidal properties. Russel-Wells [214] likewise established the presence of cellulose in certain red

and brown algae (*Corallina officinalis, Bostrychia scorpioides, Chondrus crispus, Rhodymenia palmata, Laminaria saccharina, L. digitata, Fucus serratus, F. vesiculosus, Ascophyllum nodosum,* and *Pelvetia canaliculata*). Guilliermond [96] reported the observation of granules of various sorts, both at rest and in motion in the cells of *Saprolegnia,* by means of both ordinary and polarized light; in the epidermal cells of tulip and iris he found both granules and chains of granules. Bonner and Heyn [27] prepared suspensions of cell-wall particles of oat coleoptiles by grinding the membranes in distilled water or 5 per cent NaCl. Heavier cell-wall fragments were removed by repeated centrifuging in distilled water. The particles were found to be rich in pectin and protein. It was found that the protein was tightly bound because, after the removal of the pectic material with ammonium oxalate, hot dilute HCl and hot dilute NaOH, the protein reaction of the particles and their cataphoretic behavior were not affected. These observations were of special interest in the light of the studies of "protective colloids" by Loeb,[158] who found that paraffin particles coated with gelatin behaved as though they were gelatin; by Reinders and Bendien,[200] who found that a gold sol sufficiently protected with protein can exhibit positive or negative signs of charge depending upon the pH value of the solution, and suggested that besides the formation of a protective colloid, the possibility exists that the molecules of protein are oriented at the interface; and by Freundlich and Abramson,[74] who found that the cataphoretic behavior of quartz particles, when coated with gelatin, is typical of gelatin.

In 1934 Dauphine [52] reported the intimate association of *proteins* with *cellulose* and *pectic materials* in the membranes of *Aucuba japonica.* Microchemical tests for these three substances were made in untreated membranes, after prolonged treatment with sodium hypochlorite. The ruthenium red reactions of pectic materials and the $H_2SO_4 + I_2KI$ reactions of cellulose were not affected by the hypochlorite treatment. On the other hand, protein reactions were obtained only in untreated material, in both cytoplasm and cell membranes. In the fresh material the protein reactions in the cytoplasm and the membrane differed only in intensity—they were fainter in the cytoplasm. In the triangular areas at the corners of the cells there were accumulations of granular material which, likewise, became brown in the xanthoproteic reaction. These observations of granular cellular inclusions (Guilliermond) of protein reactions of cell-wall particles (Bonner and Heyn and Dauphine) and of the functioning of protective colloids (Loeb, Reinders and Bendien, and Freundlich and Abramson) are reminiscent of the earlier observations of Strasburger,[254] who saw "protein granules" (*microsomes*) in the cell protoplasm which were used up in the formation of the cell membrane, and of Wiesner [268] who found that the "protein granules" (*dermatosomes*) in the protoplasm "in some mysterious manner" were changed into cellulose after they were deposited in the cell membrane.

In this same year Farr and Eckerson [62] reported the presence of ellipsoidal cellulose particles ($1.1 \times 1.5\mu$) in the cytoplasm of young cotton fibers, of sporangiophores of *Aspergillus,* and of cells of *Spirogyra, Oedogonium,* and *Valonia.* These particles are separate or joined end-to-end in single rows (Figs. 8a, b, c). The double refraction of the particles is clearly visible in the fresh mounts (Fig. 9a), both separate and in the chains. The application of iodine, in making the $H_2SO_4 + I_2KI$ test for cellulose, resulted in a reaction with the cytoplasmic protein on their surfaces, giving a yellow-brown color. At this stage of the reaction one might conclude, as had Strasburger and Wiesner, that the particles are protein. This protein precipitation upon the particle surface was prevented, however, by pretreatment of the microscopic mount with 10 per cent KOH for 5 minutes or 4 per cent NaOH for 10 to 15 minutes: [p. 190] "After washing, a drop of iodine solution is allowed to run under the cover glass; then a drop of the sulfuric acid is added at the edge of the cover glass. In most of the fibers the chains of particles are now a clear cellulose blue, standing out sharply from the almost colorless background of the cell" (Figs. 9b, c). The re-

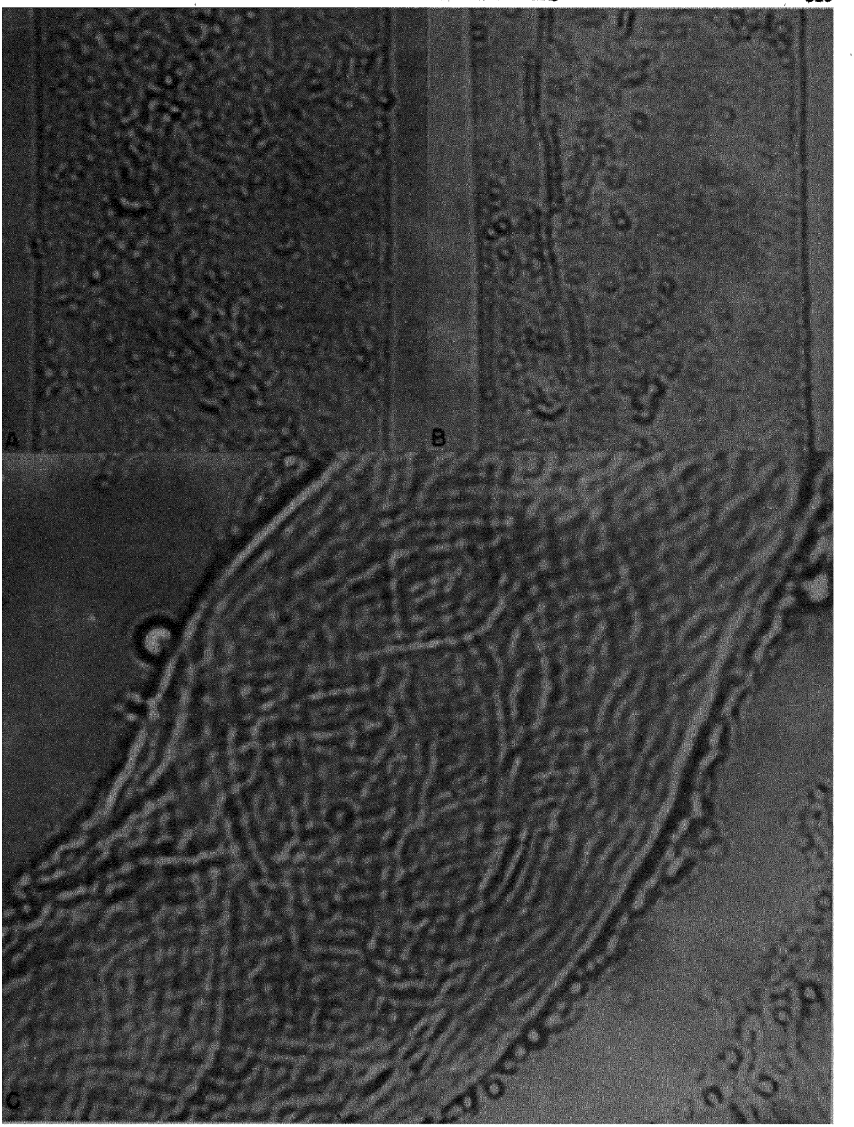


FIGURE 8. A. Separate cellulose particles in the protoplasm of a young cotton fiber (X 725). B. Early stage of fibril formation in a young cotton fiber (X 725). C. Cell membrane of the cotton fiber after deposition of cellulose fibrils in spiral arrangement in the fiber wall (X 725).

fractive indices determined by the immersion method are 1.565 lengthwise and 1.530 crosswise.

Even in the young, undried cotton fibers considerable difficulty is experienced in removing a coating of material from the cellulose particle surface which stains with ruthenium red. Concerning more mature fibers, the authors state,[62, p. 191] "When a nearly mature fiber from an unopened boll is put into a 4 per cent solution of sodium hydroxide, the separation of the fiber into fibrils and of the fibril into particles does not occur simultaneously. First there is the separation of the fiber into fibrils by

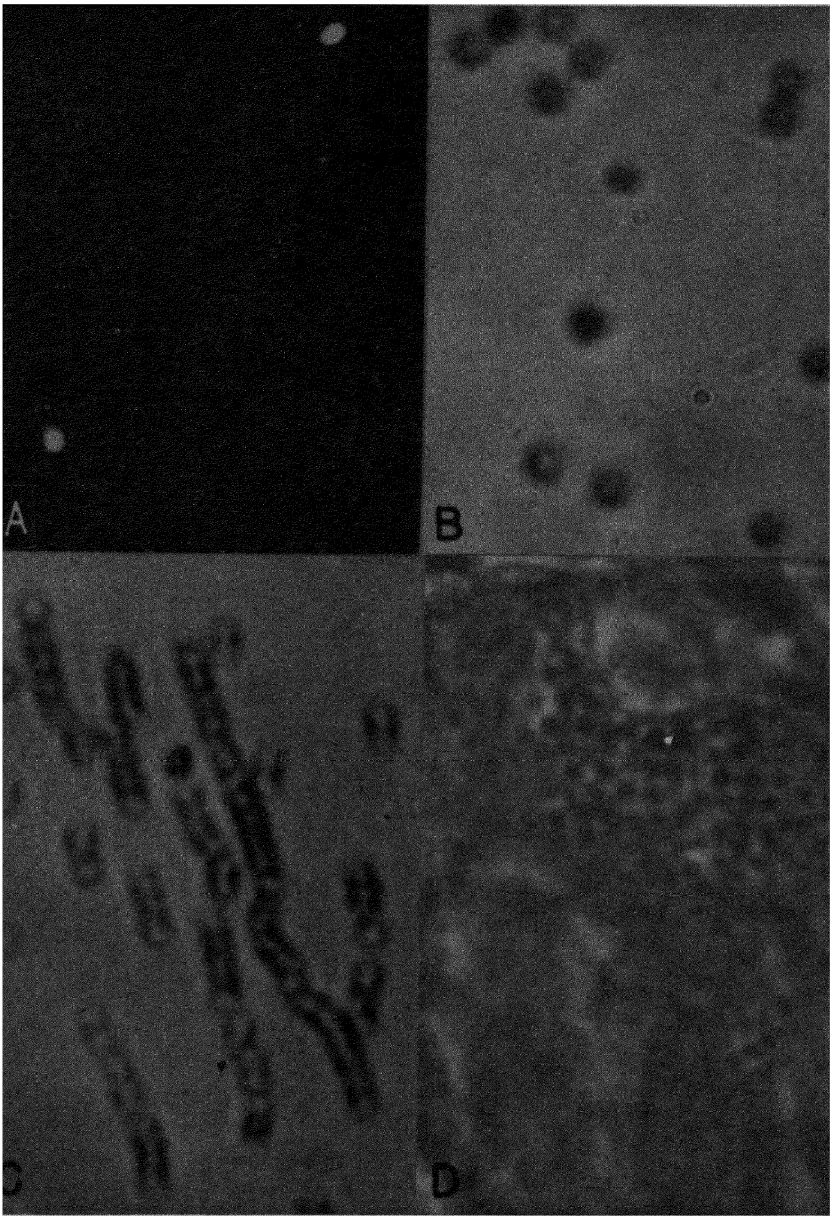


FIGURE 9. A. Cellulose particles are doubly refractive in polarized light (X 2000). B. Separate cellulose particles from protoplasm of a living cell swell and turn blue in sulfuric acid and iodine (X 1400). C. Cellulose fibrils from mature cell membrane show swelling and blue coloration of particles in sulfuric acid and iodine (X 1400). D. Mature cotton fiber treated with phosphoric acid and iodine. Swelling of cementing material and swelling and blue coloration of cellulose particles produce "honeycomb" appearance when particles are not obscured (X 1400).

partial solution of the pectic substance surrounding them. Then, an hour or more later, the fibril begins to separate into particles. A dry fiber from an opened boll requires much longer treatment to bring about the same separation."

Concerning the identification of the chemical nature of this film of material upon the particle surface, the authors state,[62, p. 191] "Although no chemical analyses of the galacturonic acid content were made, the solubility, non-double refraction, and the staining with ruthenium red are properties like those of the pectic substance in apple fruits and other tissues which have been analyzed. . . . It is not the water-soluble pectin, but whether it is pectic acid, a pectate, or protopectin is not known." Continued study of this material from many types of plant cells led to the statements in 1937 by Farr,[57, p. 988] "Its hydrophilic properties are, in general, those described by botanists of the last century as characteristic of 'pectic' materials. It is very unlikely that in final chemical analysis all of its basic constituents will meet the requirements of the modern chemical concept of pectic substance."; and:[58, p. 229] "Many of its properties are similar to those of the viscous colloidal phase of the cytoplasm in which the particles are suspended. It may be nothing more than an adhering layer of this mixture whose chemical properties themselves are not understood." The accuracy of this conception has been made clear in a more recent investigation by Farr,[61] in which the stages of formation of the cellulose particles in living cotton fibers and certain marine algae have been observed and described.

In the original paper the behavior of the cellulose particles during cell-membrane formation is described as follows:[62, p. 192] "From direct observation we may conclude, therefore, that the pectic-coated cellulose particles distinguishable in the living cytoplasm may exist singly or in chains; that toward the center of the lumen the single and short chain types of arrangement predominate; that as the wall is approached the chains become longer; and that a single chain of closely appressed particles forms a single fibril of the fiber wall." The spiral arrangement of these fibrils in the wall of the mature cotton fiber is shown in Fig. 8c.

Later studies by Farr and Eckerson[63] led to a method of "purification" of comparatively large quantities of cellulose particles, without evidence of degradation, by treatment with hydrochloric acid. Material was thus made available for study of additional particle properties. Before the end of the same year confirmation of the cellulose nature of these particles had been obtained by Farr and Sisson[64] by means of x-ray diffraction analyses (Fig. 10). After continued study a method was devised by Sisson,[237] which led to the identification by the same technique, of the crystalline cellulose in young cotton fibers at ages corresponding to those reported earlier by Farr and Eckerson.[62] The method involved at least a partial removal of the amorphous material from the surfaces of the particles used in the x-ray diffraction analysis. In discussing this work later Sisson[240, p. 196] states, "Comparative x-ray and microscopic studies indicate that the crystalline pattern of cellulose arises from the cellulose particle, which shows double refraction in polarized light. Furthermore, the size of the particle ($1.1 \times 1.5\mu$) is small enough to give the Debye-Scherrer patterns observed. It is unnecessary to assume crystalline units smaller than the particle, since powder patterns are obtained whenever the crystals of the diffraction sample become smaller than about 5μ.

"There is also reason to believe that the amorphous pattern present in the x-ray diagram of cellulose fibers may be partly accounted for by the non-doubly refractive material which may be observed on the surface of the particles. This assumption is based largely upon purification studies. . . . Extraction of these young fibers with organic solvents removes a fraction which gives an x-ray pattern characteristic of a waxy material; extraction with ammonium oxalate, a fraction which gives a pattern characteristic of a pectic material; and extraction with hot dilute alkali, a fraction which gives an amorphous pattern. After these extractions the residue gives only the crystalline diffraction pattern of cellulose, upon which there is superimposed a

weak amorphous pattern. Microscopic examination shows this residue to contain cellulose particles covered with a thin coat of non-doubly refractive material, the other non-cellulosic materials originally present in the young fiber having been removed by the purification process. . . . These preliminary studies, therefore, would indicate that the amorphous portion of the cellulose membrane diffraction diagram may arise from two more or less overlapping sources. One source is that of

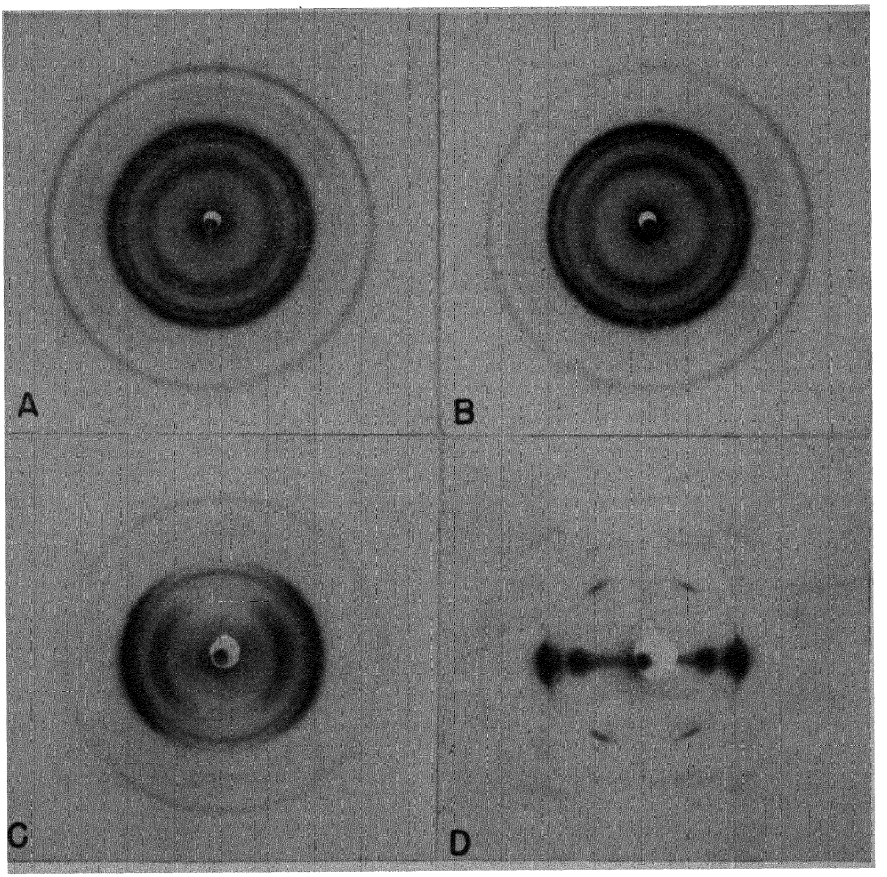


FIGURE 10. X-ray diagrams of: A. ground cotton; B. *cellulose particles* prepared from cotton fibers by treatment with HCl (sp. gr. 1.19); C. paralleled cotton fibers; D. paralleled ramie fibers. All diffraction lines are those of *native* cellulose. The purified cellulose in the form of cellulose particles is sharper than that of ground cotton due to the removal of amorphous materials and the ramie shows a more perfect alignment of fibrils with respect to the fiber axis than cotton.

the non-cellulosic materials which often may be removed by the usual purification processes without destroying the membrane structure. The other is that of an as yet unidentified material which is more intimately associated with the crystalline cellulose, and which apparently plays the role of a cementing material which holds the particles together."

Further evidence of the chemical complexity of this inter-particle material has been shown by Compton and Haver.[45, 46] In more recent articles, describing the formation of cellulose particles in the chloroplasts of *Halicystis* and in the colorless

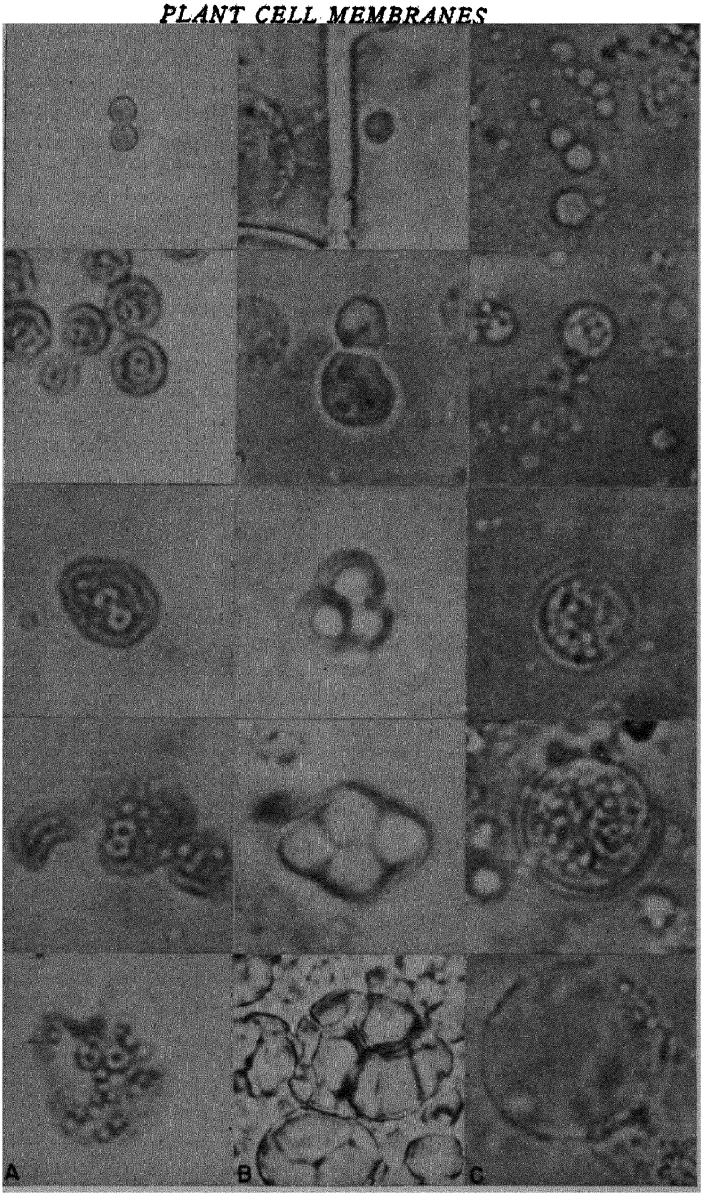


FIGURE 11. Column A. Stages in development of the chloroplast of *Halicystis* showing cellulose ring and cellulose particle formation (X 770). Column B. Stages in starch grain formation in the chloroplasts of the cotton plant (X 770). Column C. Cellulose particle formation in the colorless plastids of the cotton fiber by a process of successive ring formation and fragmentation essentially similar to the mechanism of cellulose formation in *Halicystis* (X 770).

plastids of the cotton fiber, Farr [60, 61] has presented additional evidence not only as to its complexity in any given membrane but also as to the high degree of variability in various types of plant-cell membranes. In the chloroplast of *Halicystis* * the

* A complete report of the studies of cellulose particle formation in *Halicystis* is to be published in a forthcoming issue of the Contributions from Boyce Thompson Institute.

formation of cellulose particles takes place through the fragmentation of cellulose rings of varying diameter but equal thickness. These rings are produced successively in the outer regions of the plastid plasma during the normal growth of the chloroplast. The membrane of the chloroplast finally breaks, and the entire plastid contents, including the cellulose particles and the green plasma, are deposited on the inner surface of the cell membrane in the course of the formation of a new membrane lamella. The various stages of cellulose ring formation and ring fragmentation to form cellulose particles can be followed microscopically, although the green chlorophyll in the plastid plasma tends to obscure such minute structural details (Fig. 11a). The cellulose rings and cellulose particles in the chloroplast of *Halicystis* react microchemically as mercerized cellulose, as indicated earlier by Sisson [238] in connection with cellulose particles in *Halicystis* membranes.

The mechanism of *native* cellulose particle formation in colorless plastids (Fig. 12a) of the cotton fiber is essentially similar to the mechanism of *mercerized* cellulose formation in the chloroplast of *Halicystis* [61, p. 191] "In the developing cotton fiber the cellulose-forming plastid is rendered nearly invisible by the similarity in its refractive index and that of the cytoplasm which surrounds it. Measurements show that both are between 1.33 and 1.34. While the cellulose particles with refractive indices of 1.565 (lengthwise) and 1.530 (crosswise) are held within the plastid membrane, they are very effectively obscured. When they are freed by the bursting of the membrane . . . they attain a degree of visibility which made possible their identification in 1934. Their complete development within a specialized organ such as a plastid accounts fully for their sudden appearance, full formed, in the cytoplasm, prior to the assumption of their final role in the formation of the cell membrane" (Figs. 12b, c; Fig. 11c; Figs. 8a, b, c).

Among the many immediate problems created by these new findings none was more important than the determination of the method of cellulose formation in plant cells in which both starch and cellulose are formed. In the cotton fiber cellulose alone is produced. A study of leaf, stem, and boll-wall cells of the cotton plant has revealed that while starch is formed in chloroplasts, usually lying near to the surface of the protoplast, cellulose is formed in colorless plastids which are located in the interior regions of the protoplast (Fig. 12d). It can be seen that these two closely related carbohydrates probably came into existence simultaneously, in separate plastids, in the same cell. A comparison of the stages of cellulose ring and cellulose particle formation in the chloroplasts of *Halicystis* and in the colorless plastid of the cotton fiber is shown in Figs. 11a and c, and these in turn can be compared with the stages of starch formation in Fig. 11b.

These observations of cellulose-particle formation in the living cells of *Halicystis* and the cotton fiber leave unquestionable the growing evidence of the chemical complexity of the inter-particle material. In *Halicystis* the entire plastid content apparently enters into the formation of the membrane lamellae. Along with any cytoplasmic material which may be carried with it, the plastid plasma with all of its components, including the chlorophyll, makes up the *cementing material* of the newly deposited membrane layers. Frequently entire green chloroplasts are found imbedded in the wall (Fig. 4c), a phenomenon reported first by Correns [46] in *Valonia* and found repeatedly in both *Halicystis* and *Valonia* membranes by the author. Their presence in the membranes was not understood until the relation of the chloroplasts in these cells to cell-membrane formation had been revealed. The absence of information always invites speculation, and many interesting suggestions have been forthcoming. Sisson [240, p. 195] states, "Since many non-cellulosic materials are so closely associated with cellulose during its formation in the cytoplasm, it is possible that there may be a sort of mixed crystallization of the cellulose with non-cellulosic materials at the surface of the particle. This is suggested by comparative x-ray and microscopic studies of a number of cellulose membranes, which indicate that when-

ever a large amount of non-doubly refractive material may be observed on the surface of the cellulose particle, the equatorial diffraction lines are always broad and unresolved. If the surface of the cellulose particle, on the other hand, is observed to be comparatively free from non-doubly refractive material, then the diffracting lines are sharp and more clearly resolved. One would not expect the surface of the cellulose particle, molecularly speaking, to be as smooth as the surface of crystals which are built up of small molecules held together by secondary valences. It is not unreasonable to assume that the long primary-valence cellulose chains would protrude at the surface and especially at the end of the particle to give a fringe structure.

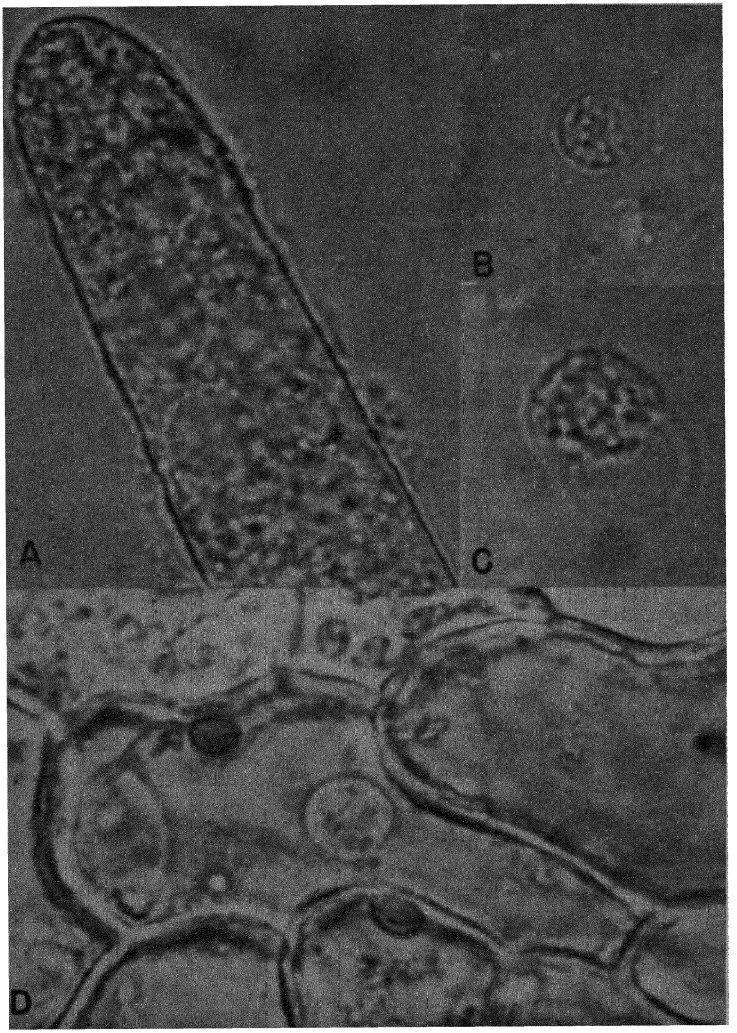


FIGURE 12. A. Young cotton fiber showing barely visible cellulose-forming colorless plastids (X 1155). B. Cellulose-forming plastid removed from cotton fiber to improve visibility (X 1095). C. Reaction of cellulose-forming plastid to sulfuric acid and iodine showing plastid plasma escaping from plastid membrane and blue coloration of cellulose particles in plastid plasma (X 1095). D. Single cell from a cotton leaf showing a chloroplast, a nucleus, and a cellulose-forming plastid in focus (X 1095)

Such a structure would not only greatly increase the effective particle surface, but also permit a more permanent end-to-end union of the particles to form fibrils and allow the polar hydroxyl groups of the protruding chains to become closely associated with other hydrophilic cytoplasmic constituents during particle formation." Likewise Compton,[44, p. 418] in commenting on the presence of pectic material in the cotton-fiber membrane, states: "The importance of the proven existence of at least one polyuronide in cotton fibers lies in the possibility that the intercrystalline, amorphous, fiber phase is composed of various molecular modifications intermediate between those of these substances and the true cellulose molecule."

In the light of our present knowledge of the early stages of cellulose particle formation, these and many other constructive suggestions are transferred from a state of postulation to one of experimentation. Cellulose-forming plastids can be made to give up their cellulose rings at the many stages of development from the liquid, through the gel, to the final crystalline state (Fig. 13); fractionations of plastid plasma can be attempted by electrophoretic techniques such as those developed by Tiselius;[259] comparisons of these fractions with fractions of peripheral cytoplasm immediately before and after deposition in a newly formed wall lamella might lead to definite knowledge of the contribution to cell wall constituents of the basic cytoplasm of the cell which surrounds the plastids. And, in the course of these investigations we might learn more of the important and unknown properties of the many types of microscopically visible structures which are distributed throughout the continuous colloidal ground substance in which the various activities peculiar to the life processes of a cell take place.

Microscopic observations and both chemical and physical analyses of cell-membrane formation and structure have accumulated in large numbers during the past decade. Sponsler[246] observed in *Rhizoclonium,* whose cells measure $100 \times 20\mu$, that the cross wall starts as a ring-like projection on the inside of the cell wall and grows by the addition of material on its inner edge, gradually building the ring into a solid disc. Concentric layers 0.1μ thick are said to be deposited on the inner edge of the ring at the rate of 1 layer per minute. Ritter[210] reviewed his extensive work with the disintegration of wood fibers into fibrils, fusiform bodies, and spherical units and reported,[p. 60] "The ash residue from wood forms a continuous skeleton consisting of long, slender, tube-like filaments resembling the fibrils and having innumerable constrictions. The constrictions make the tubes appear to consist of hollow beads, each of which is attached to its two adjoining neighbors end to end and partitioned into variously shaped compartments." The author believes that a large proportion of the ash-forming substance probably exists where water can reside, in the spaces around the structural units. Brown[34] found that the ash residue resulting from the burning of wood fibers retains the fiber skeleton, even though it constitutes less than 1 per cent of the total fiber. Astbury, Marwick, and Bernal[12] and van Iterson[187] made important contributions to the knowledge of the wall structure of *Valonia.* Hess[125] reported that cellulose membranes are made up mostly of a crystalline substance which diffracts x-rays, surrounded by a thin sheath of another substance whose reactions to fluids determine the permeability. Morphological differentiation was shown in dark-field, in swollen material. The crystalline substance was identified as cellulose, but the "Hautsubstanz," present in very small quantities, had been neither isolated nor identified. He concluded that the recognition of these two fiber components is important for the understanding of the characteristics of fiber structure as well as for reactions to reagents. Hess, Kiessig, Wergin and Engel[127] corroborated Sisson's[287] identifications of crystalline cellulose in young cotton fibers in which unextracted and chloroform-extracted cotton fibers did not show the cellulose diagram. It appears first after extraction with hot or cold water. An unknown substance which is removed by water extraction evidently masks the x-ray diagram. Goncharov and Burvasser[88] report that along with pure cellulose there are very thin

films surrounding every structural element of the fiber. Their percentage in the fiber is very small, but they greatly affect its properties. Lüdtke [159] commented that many errors arise in the interpretations of the structure of plant fibers by placing too much weight upon the results of a single method, and that x-ray diffraction and

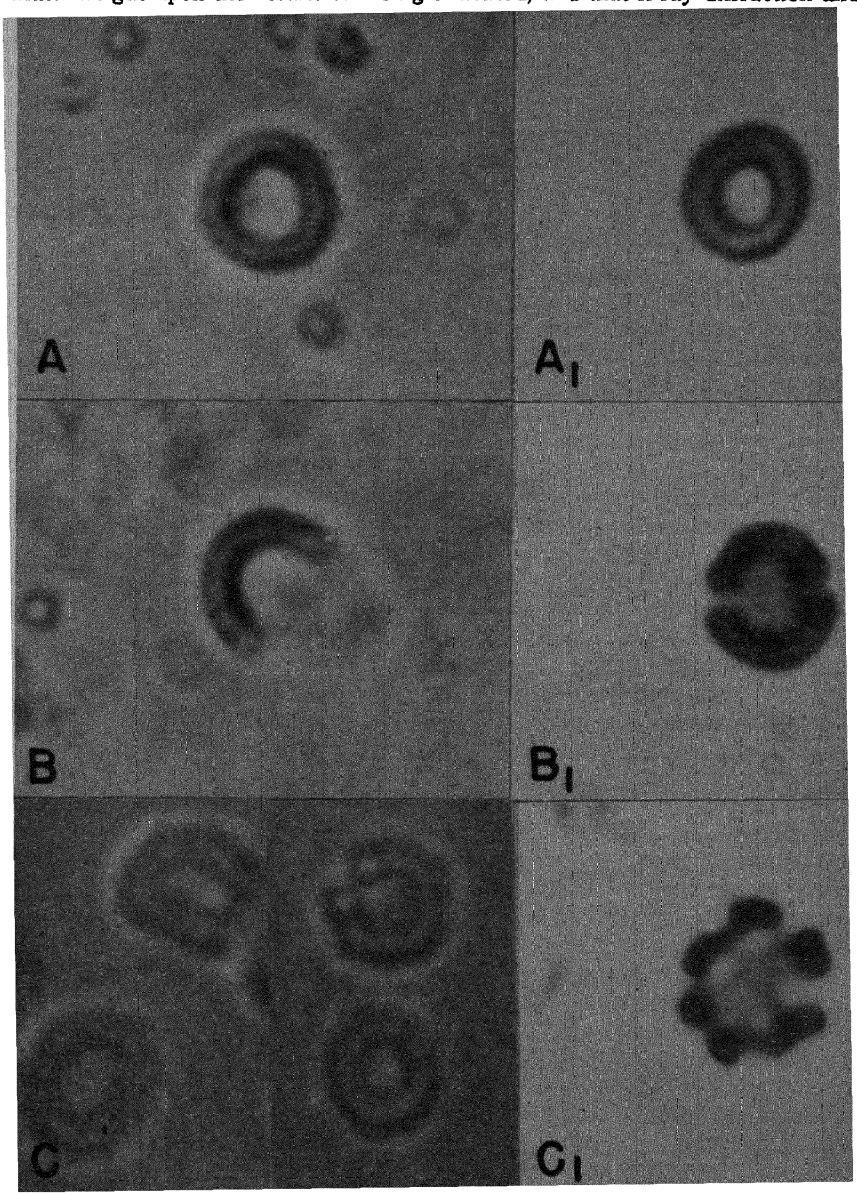


FIGURE 13. A. Cellulose ring can be removed from chloroplast of *Halicystis* without breaking. A_1. Unbroken ring swells in sulfuric acid and iodine. B. Later stage of ring development breaks easily during removal from chloroplast. B_1. Swelling of broken fragments in sulfuric acid and iodine. C. Cellulose particles beginning to show in cellulose rings. C_1. Treatment with sulfuric acid and iodine swells and separates them (X 1380—→3200).

microscopic analyses should be supplemented by microchemical tests in order to give a true picture of the identity and position of the components of the cell wall. Staudinger and Reinecke [251] reported a study of the degree of polymerization of celluloses in old fabrics. All cellulose with a value less than 200, as determined by viscosity measurements, had lost its characteristic property of forming fibers, and although degraded was not chemically changed. Farr [59] reported that the cellulose component of the cell membrane of the cotton fiber, which is in the form of diminutive cellulose particles, does not dissolve in the standard solution of cuprammonium hydroxide specified by the American Chemical Society to produce the viscosities commonly attributed to them. The fiber is transformed by the cuprammonium hydroxide into a swollen, viscous mass of cementing material in which the cellulose particles are dispersed. These results were corroborated by Sisson [239] by means of dispersion, electrokinetic and coagulation studies, and by Compton [42] through viscosity and slit ultramicroscopic determinations. Wergin [268] made visible the existing homogeneities of the cell wall by swelling it with cuprammonium hydroxide, and thus identified the fibrils which, in turn, separated from a string-of-pearl arrangement into small (0.25μ) cellulose particles. Roberts and Haring [211] showed the presence of a row of cellulose particles in the cross walls of the protenema of *Polytrichum commune*. Kerr [141] reported that the wall of the cotton fiber is a continuous matrix of cellulose, some portions of which are denser than others. The lamellated appearance of the swollen wall is considered to be due to alternations of dense zones of cellulose with zones in which the cellulose is much less dense, larger quantities being deposited during the day than at night, because of temperature changes. Anderson and Moore [11] reported that cotton fibers grown under continuous artificial light have no growth rings in their walls. Collenchymatous cell walls in the same plants show characteristic lamellations. Barrows,[20] in a study of the cross-sectional structure of cotton fibers grown in both continuous and intermittent light, found no relation between lamellate structure and the daily periodicities. Lamellae are always present, but vary so greatly in number from fiber to fiber on the same seed and from base to tip on the same fiber that no evidence of daily "growth rings" could be found. Bailey and Kerr [16] reported that the cellulose matrix of the swollen secondary wall of cotton fibers and of certain lignified cell membranes is an extremely heterogeneous but firmly coherent structure, the finer details of which grade down to the limits of microscopic visibility. I. W. Bailey [15] found that both primary and secondary walls of typical cells of the higher plants are composed of a porous but firmly coherent matrix of anisotropic cellulose, whose finer structural details likewise grade down to the limits of microscopic visibility, *i.e.*, 0.1μ or less. Lignin and other non-cellulosic constituents may be deposited in the elongated, intercommunicating interstices of the cellulose, thus resulting in two continuous, interpenetrating systems. A. J. Bailey and Brown [14] found the fibrils of wood fibers to be made up of single rows of cellulose particles approximately 1 micron in size and held together by an amorphous matrix. Anderson and Kerr [10] reported two spiral systems of cellulose making an angle of approximately 70° with the long axis of the fiber in the walls of 2-day-old cotton fibers; the cellulose in the primary wall was seen to form an open meshwork of fine thread-like strands that frequently anastomosed; the secondary lamellae are composed of dense and less dense areas of cellulose, the layers being formed of systems of spirally wound branching threads; in 5-day-old fibers *plastids* containing starch were seen throughout the cytoplasm; in 9-day fibers they had increased markedly in size and prominence. Hess,[126] as a result of chemical, microscopic, and x-ray studies, reported that between the fibrillae in a fiber wall lies an easily affected substance which appears in the microscope as a gelatinous material. When treated with swelling agents the fibrils break down into particles which are not exactly spherical, but are elongated in the direction of the fibril axis. In dark-field they measured 0.2μ in diameter. Heuser and Green [130] treated wax-free cotton fibers with

0.5 per cent and 2 per cent ammonium oxalate in the air and in a nitrogen atmosphere at 70 to 80° C for 600 to 3000 hours. In all cases the fibers emerged from the treatment physically undamaged. The existence of a *cementing substance* in the fiber as a fundamental constituent was rendered improbable and the cuprammonium viscosity of a solution of the fibers was found to be a property of the entire fiber and not of a problematical cementing material alone. Whistler, Martin, and Harris [266] found that cotton freed of pectic substance with alkali showed no significant change in tensile strength or in viscosity. Conversely, cuprammonium solution to which pectic material from various sources had been added substantiated this conclusion. Whistler, Martin, and Harris [265] described a new method for measurement of uronic acids in cellulose membranes. Some of the recent reviews of various phases of plant-cell membrane analysis which have appeared are by Norman,[190] Heuser,[129] Preston,[202] Stiles,[253] Nickerson,[187] and Kratky.[151] Studies dealing with the relation of cell-membrane structure to cell elongation in the living plant have been published by Ruge,[212] Heyn,[131] Preston,[201] and Farr and Sisson.[65]

The most comprehensive microscopic study of the formation and structure of plant-cell membranes which has appeared during the present decade is that of Wieler.[267] In many tissues from a large number of plants from all parts of the plant kingdom, he has followed the important function performed by the outer portion of the protoplast in organizing the secondary lamellae of the cell membrane. In the fresh tissues this viscous film is quite conspicuous, and Wieler expresses surprise that it has remained unnoticed in the researches made during the last years by Frey-Wyssling, Herzog, Hess, Meyer, Wergin, and others. The cellulose component of the membrane is made up of cellulose particles (1.25μ in diameter) whose relation to the other membrane components he compares to the relation of the cells in a honeycomb to the comb itself. This is an excellent analogy to be used in large numbers of cells where fibrils are not formed and is likewise descriptive of the appearance of the fibrillar wall of a cotton fiber when particles are pushed apart by the swollen cementing material (Fig. 9d). Wieler's techniques have included certain staining and swelling treatments not used by Farr and Eckerson and, from his description, it is clear that some of these are useful in bringing out more clearly finer structural details. In all cases in which the honeycomb structure was found, he obtained a positive cellulose reaction with $H_2SO_4 + I_2KI$ for the contents of the small cells (cellulose particles), while he found that the surrounding substance consisted of other chemical substance concerning the nature of which, he states, nothing can yet be said.

Finer structural details of cell membrane than have hitherto been observed are now being shown in studies made with the electron microscope. In the greater number of images there are shown extremely fine, clearly defined fibrils far below the limits of microscopic resolution. The plates which have been published by Ruska and Kretschmer,[218] Sears and Kregel,[234] and Barnes and Burton [19] are representative of the group, and the reproductions are characterized by a greater clearness than in many of the others. In addition to the fine fibrils there is shown globular material, apparently of a different composition, ranging in shape from spherical to ellipsoidal in both the swollen and unswollen states. From magnification and descriptions these appear to be approximately 1μ in size, with the exception of a more or less linear group shown by Barnes and Burton, which are approximately half that size. The finely fibrillar material is designated as cellulose by most of the authors, although it is referred to as "skin substance" extracted from wood pulp by Sears and Kregel.[234] The globular material has no such specific designation. In view of the heterogeneous chemical and physical nature of the cell membranes which form the subject matter for such studies, it is probable that careful separation and identification of all the separable constituents before examination will be helpful in utilizing this new technique to its fullest capacity. From the data thus far presented,

it is questionable whether, or to what extent, the images obtained with the electron microscope represent structures of cellulose as such.

Discussion

A review of the accumulated information concerning the formation and structure of plant-cell membranes reveals the fact that data concerning the microscopic structure of the mature membrane began to accumulate even before the time of recognition of the cell as the unit of structure in plant and animal tissues. Lyngbye reported the presence of "minutissime punctata" in the membrane of *Gastridium ovale,* now classified under the name of *Halicystis ovalis.* Valentin, using cells of different types and of smaller dimensions, was able to see something of the relationship of the protoplast of the cell to the cell membrane. Through his studies and those of many of his contemporaries it soon became evident that the cell membranes were increasing in thickness by the deposition of material from the protoplasm upon their inner surfaces. This observation led to the question of what kind of material is thus deposited from the protoplasm; Valentin, searching for an answer to this question, found granules which, in many types of cells, joined together end-to-end to form fibrils, and which were then moved from the protoplast into the developing cell wall. These earliest workers were not concerned with the chemical composition of these structural units. Their viewpoint was as that of an observer whose attention is held by the process of construction of a wall from bricks and mortar, without particular interest in the exact composition of the materials themselves. The brilliant researches of Prout, Payen, Fremy, and others furnished a basis for distinguishing between *proteins, carbohydrates,* and *fats* in the cell and served in turn to differentiate the carbohydrates into *sugars, starch,* and *cellulose.*

Confronted with the facts that sugar, starch, and cellulose are products of living protoplasm, the next problem in this field was to find *where,* in the living protoplast, these substances come into existence and, if possible, how they are used by the cell during its period of vital activity. In the course of these investigations it was found that sugar is usually in solution in the protoplast, and that cellulose and starch represent visible forms of carbohydrate in the cell. As early as 1837 von Mohl found that the granules of starch were formed in the *chloroplasts,* and the subsequent studies of Sachs, Schimper, Meyer, and others revealed the fact that the necessary conditions for starch formation exist only in these special morphological structures in living cells, *chloroplasts* and *leucoplasts.*

The starch granules themselves were found to develop by deposition of materials from the plastid plasma around small, starch-forming centers. This early linking of *starch* formation with a microscopically visible, morphological process has served to protect it, in some measure, from theoretical attacks upon its molecular structure, based entirely upon considerations of certain properties of mature starch grains. If one is free to assume that a substance "crystallizes" or "polymerizes" directly from the cytoplasmic substratum of the cell, laws which are believed to govern such crystallizations or polymerizations, as they take place in the test tube, can be theoretically applied. In connection with starch formation, this freedom has not been permitted, for the theorist has been confronted with facts of the visible aspects of its organized synthesis which can be adequately explained by no known laws of molecular behavior.

Cell membranes, until recently, have not had similar protection. Although nineteenth-century researches, which have been previously reviewed, furnished adequate evidence that microscopically visible granules formed in the protoplasm are used in the construction of plant-cell membranes, and that these units, along with other materials of protoplasmic origin, maintain their identity in the mature cell membranes, the cellulose nature of these protoplasmic granules was not known until they were identified microchemically by Farr and Eckerson in 1934. Even then it

was not known *where* they are formed in the living protoplast. While this lack of information existed, their method of origin was a matter for speculation. It was conceivable that they might crystallize or polymerize from the basic cytoplasm directly. Their existence in the form of visible particles and fibrils in the living protoplasm, however, excluded the possibility of membrane formation by deposition of molecular layers upon the inner surface of the cell membrane. The recent discovery of cellulose-forming plastids, in which the stages of development of cellulose particles are distinctly visible, now affords to cellulose the same degree of protection from theoretical consideration that has been afforded to starch. The gradual separation of cellulose-forming material from the plastid plasma, the gradual solidification of this material into rings of uniform thickness, the fragmentation of the rings into particles of remarkably uniform size, the active Brownian movement of these particles in the plastid plasma, the direct deposition of the entire plastid content in the cell membrane, as in *Halicystis*, and the intermediate stage of fibril formation preceding deposition, as in cotton, are processes even more elaborate and complex than the visible aspects of starch grain formation. And, as though Nature were not sufficiently satisfied with the complexity of successive ring formation and disintegration in the process of cellulose formation, we have found more recently, in the chloroplast of *Valonia*,* a different type of behavior (Fig. 14). Following the formation of a single closed cellulose ring (a), a cellulose fibril begins to form within the young plastid (b). Its various stages of development are somewhat obscured by the green plastid plasma with which it is surrounded (c, d), but if the plastid membrane is removed and the plasma partially washed away with distilled water, the original cellulose ring and the coiled cellulose fibril are more distinct. Two plastids thus prepared are shown in (e). After removal of the plastid membrane from mature or nearly mature chloroplasts, the fibrils quickly straighten and, with the usual manipulations involved in the dissections, tend to separate readily into cellulose particles (f). How frequently such fragmentation occurs in the living cell is not known. It would seem to be comparatively rare, however, from the examination of the inner surface of a fresh membrane. There the fibrils, uncoiling from the plastids, can be seen in the act of aligning themselves with previously deposited fibrils in the developing wall lamella (g, h). The precision with which this alignment takes place is evidenced both by the microscopic appearance and the x-ray diffraction pattern of the mature cell membrane (j). It is probable that this direct formation of the fibril within the plastid and its deposition, without disintegration, in the membrane, accounts for the high degree of crystallinity in the mature membrane, as shown by x-ray diffraction analysis. The organization of fibrils from separate particles, as in cotton fibers, can thus permit the incorporation of colloidal materials upon the surfaces of the particles, and less perfect alignment of the particles than is found in the fibril of *Valonia*, which, once formed in the plastid, apparently remains intact.

The formation of cellulose in the living cell has been found to take place, therefore, in a specialized cytoplasmic organ, a cellulose-forming chloroplast or colorless plastid. This formation has no visible aspects in common with the process of starch formation. As far as we know at present, the production of these two closely related carbohydrates always takes place in separate plastids, even though their formation goes on simultaneously in the same cell. Future research may furnish more conclusive information in this connection. A true understanding of the structure of the mature cell membrane must take into account these developmental factors which permitted it to function, originally, as a part of a living organic unit. Bensley,[28] in commenting upon the dilemma of the application of the concept of molecular pattern to the problem of protoplasmic structure states,[p. 390] "Protoplasm on the contrary respires, excretes, performs complicated chemical operations, uses or liberates energy

* A detailed report of cellulose membrane formation in *Valonia* is to be published in a forthcoming issue of the Contributions from Boyce Thompson Institute.

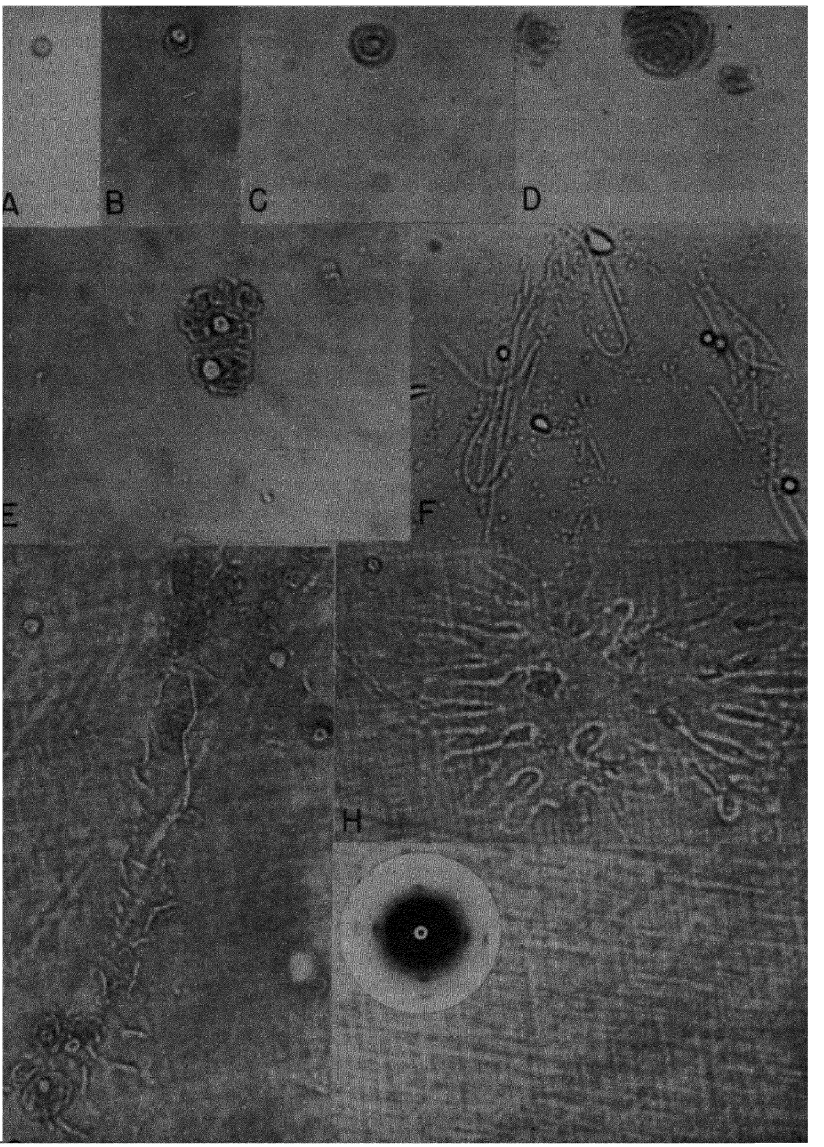


FIGURE 14. A-D. Cellulose ring and cellulose fibril formation in the developing chloroplast of *Valonia ventricosa*. E. Removal of plastid membranes from two plastids and washing away green plasma make rings and fibrils more clearly visible. F. After removal of plastid membranes, mature fibrils straighten and readily break into cellulose particles. G, H. Fibrils uncoil from plastids on inner surface of membrane and align themselves with fibrils in developing membrane. J. Orientation of fibrils in mature membrane as shown by microscope and x-ray diffraction.

and reproduces its own substance in kind. This metabolism can be imitated in part *in vitro* . . . but in the cell is characterized by speed, orderliness, and rhythm not to be found in a random mixture of chemical substances in solution." The production of physically and chemically complex structures such as starch grains and cell membranes are clear evidence of this coördinated behavior.

As shown in Fig. 6 the new theory of the construction of cell membranes from "fringed micells" has had a gradual evolution from the original concept of Nägeli that cell membranes are composed of submicroscopic crystalline micellae surrounded by films of water. This concept of membrane structure, shown diagrammatically by Frey in Fig. 6a, is shown in greater detail by Seifriz (Fig. 6b), in order to incorporate knowledge of molecular structure contributed by x-ray diffraction analysis. The possibility of disorderly as well as orderly arrangement of the micells is brought out in the illustration according to Mark shown in Figure 6c. The configurations of both orderly and disorderly arrangements with respect to cell membranes are shown in Fig. 6d, after Frey. These conceptions of cell-membrane structure necessitate no consideration of chemical inhomogeneity in the membrane and only such physical inhomogeneity as can be explained upon the basis of variations in micellar arrangement. It will be remembered, however, that even Nägeli observed that the *micellae* were comparatively stable and that changes in them do not account for the plastic properties of swollen membranes. For this reason he postulated that they are separated by water films which could increase or decrease in thickness while, in the meantime, the *micellae* themselves did not swell. This stability of the crystalline phase of the membrane has continued to be in evidence as measurements with newer techniques have been developed. Successive treatments of the membranes with chemical reagents may reduce them to a fine white powder which still gives the x-ray diffraction pattern of native cellulose. Upon the basis of *visible* cellulose crystallites it is seen that the coating of protoplasmic material, acquired by the particles before deposition in the membrane lamellae, has been removed more or less completely by the chemical reagents and the cellulose particle itself has shown in many instances no evidence of reaction with them. The conception of *invisible* structure found a theoretical basis for the explanation of these regions of greater and less sensitivity in the membrane in the work of Carothers dealing with chain molecules of unequal length (Fig. 6e). Adapted to cell-membrane structures, these appear as the "fringed micells," as shown by Kratky (Figs. 6f, g) and find their final representation, in two dimensions, in both the organized and disorganized states as shown by Freund and Mark in Fig. 6h.

This conception of cell-membrane structure has abandoned the original idea of discrete *micellae* surrounded by water. It is now stated by Berkeley [24] that between the dense areas *amorphous cellulose* surrounds the *fringe molecules.* The properties of amorphous cellulose, its microchemical reactions, etc. are not described. Membrane structure and composition of this type presumes, however, a clean-cut separation of crystalline and amorphous cellulose from all other protoplasmic constituents in the process of formation of the membrane lamellae, a consideration not substantiated in the observations of cell membrane formation in *Valonia, Halicystis,* the cotton fiber, and many other types of cells. These assumptions of a fringed micellar membrane structure have been evolved from chemical and physical studies of mature membranes. Membrane formation from cellulose particles held together with colloidal protoplasmic materials has, as a background, the entire process of the formation of visible structural components from the time of the beginning of their synthesis in the cellulose-forming plastid of the living cell. It is of interest, therefore, to compare the similarity between the diagrammatic representation of the fringed micellar structure as shown in Fig. 6h and the photomicrograph of the structure of the cotton fiber as shown in Figs. 8c and 15a, b, c, d. From the physical viewpoint the difference seems to be largely in size of the structures concerned—the crystallites and the intercrystalline material in the one case being microscopically visible, and in the other microscopically invisible.

The microscopic problems involved in observing the formation of cellulose particles in a colorless plastid of a cotton fiber, in following these cellulose particles through their subsequent alignment to form fibrils and the fibril deposition on the inner surface of the cell membrane, are those which are encountered in any attempt

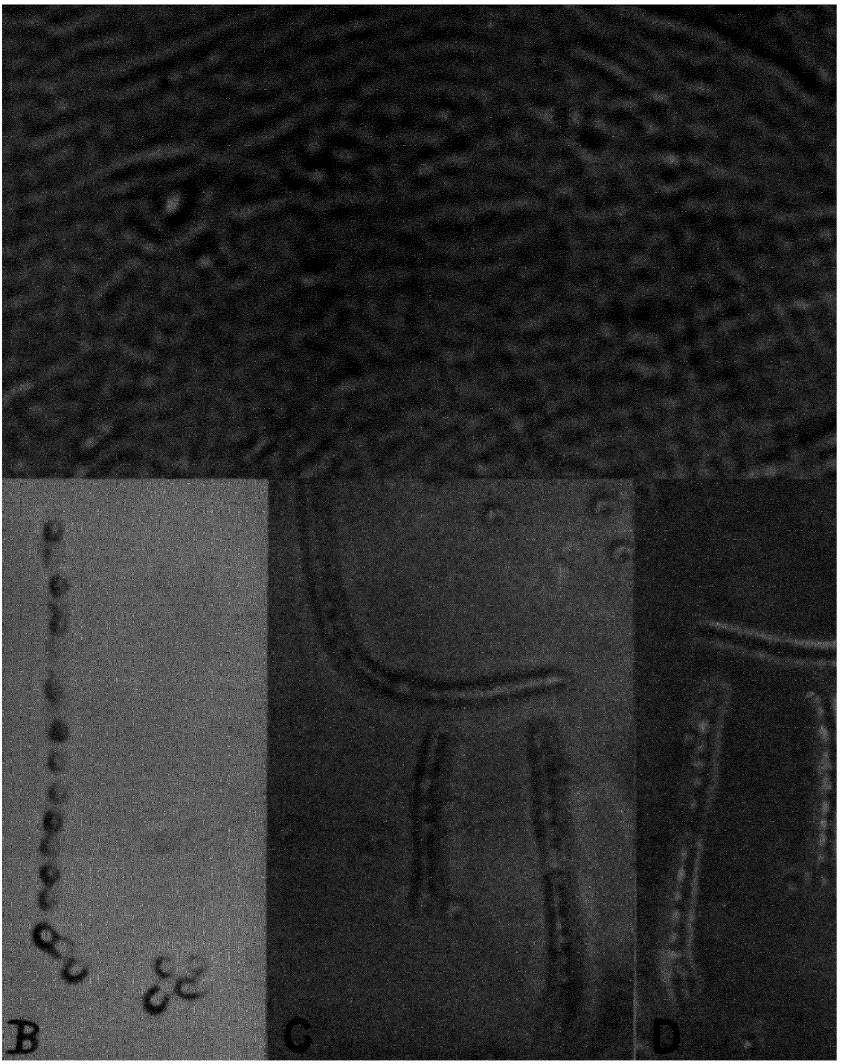


FIGURE 15. Top: Membrane of mature cotton fiber disintegrates into fibrils and the fibrils, in turn, into cellulose particles. B. Disintegration of cellulose ring and piece of cellulose fibril from chloroplast of *Valonia* into cellulose particles. C. Single fibrils and cellulose particles from membrane of *Valonia* show swollen amorphous material and blue cellulose particles after treatment with H_3PO_4 and I_2KI. D. The field shown in (C) with oblique illumination shows gel-like amorphous material in which fibril is imbedded more clearly (X 1380).

to observe a colorless, more or less translucent object imbedded in a matrix of very similar refractive index. Microscopists of the last century who worked with fresh material, lights of low intensity, and "high-dry" lenses with adequate resolving power were better equipped for such observations than many modern laboratories in which microscopic equipment, including the illuminating units, has been designed for use with killed, heavily stained tissues. With regard to the observation of the

individual particles and their subsequent behavior in cell-membrane formation Farr [58, p. 229] stated in 1937, "The degree of magnification needed to bring them into clear relief is near to one thousand diameters. Within the short focal distances involved, there is still a wide variation in the sharpness of focus of the different particles in any given field. Apparent differences in size often result from these variations in the sharpness and are not real. The approximate length of a single particle as it exists in the living cell is 1.5 microns. This measurement is believed to be very nearly correct since it is possible to check it with the average length per particle in chains of varying numbers of particles arranged end to end." "The length of the short axis is not so easily determined and may be verified only through comparison with the long axis, since, in the cotton fiber, the associated particles are always in single rows, end to end, and not side by side. The value which seems to be most nearly accurate for the length of the shorter axis is 1.1 microns. It must be pointed out that all measurements of a structure of this size and composition under the necessary optical conditions is difficult and is, at best, an approximation." Failure to see cellulose particles of this size at low magnifications is then to be expected.[182] Their confusion with crystalline granules which appear in drying solutions of cuprammonium hydroxide, however, could not take place with an observer familiar with the refractive indices, solubilities, and other properties of cellulose particles.[182, p. 324]

The refractive indices of cellulose particles taken from the living cytoplasm of the cotton fiber by the immersion method were found by Farr and Eckerson to be 1.565 lengthwise and 1.530 crosswise. Concerning these measurements Anderson and Kerr state: [10, p. 58] "The refractive indices of native cellulose have been measured in various cell walls by a number of different workers.[9] Frey-Wyssling determined the refractive indices of the cellulose in cotton fiber and found them to be $n_v = 1.596$ and $n_\alpha = 1.534$. These values are in excellent agreement with those obtained by other investigators on other native celluloses. They are not in agreement with the values reported by Farr and Eckerson [7] for ellipsoidal particles observed in the protoplasm of cotton hairs."

Results of other workers in refractive-index measurements of native cellulose are tabulated below for convenience in comparison with the values obtained by Frey-Wyssling and Farr and Eckerson: *

Author	Material	n_v	n_α
Schlotman	Cotton fibers (greenhouse)	1.565	
Preston	*Adamsonia digitata*	1.564	1.531
"	*Agave perfoliata*	1.559	1.536
"	*Yucca gloniosa*	1.554	1.537
"	Cotton fibers	1.578	1.532
Shaw	Water lily	1.568	1.534
Farr and Eckerson	Cotton fibers	1.565	1.530
Chamot and Mason		1.59	1.53
Frey-Wyssling		1.596	1.534

These values show that the n_v index is almost identical for all authors. Comparisons of methods used and comparative states of material at the time of measurement show that n_v indices measured under similar conditions agree closely, *e.g.*, Schlotman, Shaw, Farr and Eckerson. The high values for n_v, as shown by Frey-Wyssling and Chamot and Mason, can be interpreted as due to the fact that both authors made the measurements upon entire fibers in which the orientation of fibrils is not parallel either to the long axis of the single fibril or the long axis of the single cellulose particle. These and other questions concerning the existence of cellulose

* From a more detailed comparison of values and methods to be published later by Farr and Eckerson in the Contributions from Boyce Thompson Institute.

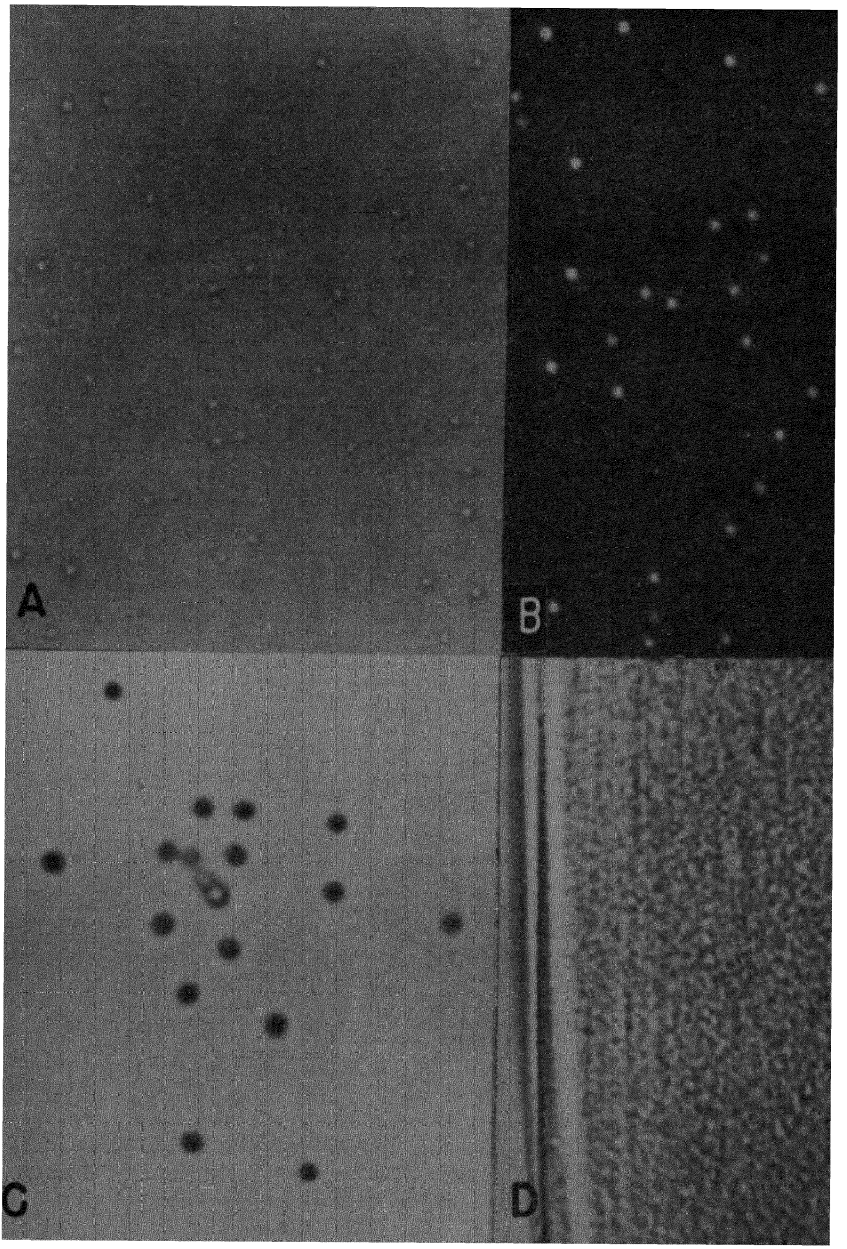


FIGURE 16. A, B. Dilute dispersion of cellulose xanthate shows cellulose particles in ordinary and oblique illumination (X 840). C. After removal from the dispersion shown above the particles show the typical cellulose reaction in H_2SO_4 and I_2KI (X 920). D. Viscose rayon filament shows granular structure (X 460).

particles in the protoplasm of living cells and in cell membranes have been considered in other publications from our laboratory.[59, 20]

Another group of papers describing the behavior of cotton-fiber membrane constituents in cuprammonium hydroxide and during xanthation [59, 42, 289, 48] showed, by means of microscopic studies in ordinary and polarized light and in dark-field, microchemical determinations, optical rotations, x-ray diffraction analyses, and observations of electrokinetic and coagulation behavior, that in these two industrial processes the cellulose particles undergo no visible physical change, but that the intercrystalline continuous fiber phase is considerably altered by the reagents employed (Fig. 16). Farr, using the standard solution of cuprammonium hydroxide specified by the American Chemical Society, containing 30 ± 2 g of copper, 165 ± 2 g of ammonia, and sucrose in the proportion of 10 g per liter, found microscopically that the fiber is transformed, in this solution, into a swollen mass of cementing substance in which the cellulose particles are dispersed. All other results by the various techniques employed by Farr, Sisson, and Compton corroborated this general conclusion. In an attempt to trace the source of the viscosity-producing fraction or fractions of the cementing material Farr states: [59, p. 89] "The one fraction of the cementing material of the cotton fiber which we had been able to separate and identify in sufficient quantities for such a study was the component which is extracted with 0.5 per cent ammonium oxalate. Upon the basis of uronic acid anhydride determinations according to the method of Dickson, Otterson, and Link this extract was identified as pectic material." [p. 91] "When 0.625 g of this powdered pectic substance was added to 100 cc of the A.C.S. standard solution of cuprammonium hydroxide, the viscosity of the cuprammonium solution was increased from 3 min, 43 sec, to 11 min, 24 sec. The addition of 1.25 g to 100 cc of the solution increased the viscosity to 26 min, 10 sec. Table IV shows the result of adding increasingly larger amounts of pectic material to the point when over ten hours were required for the mixture to flow through the viscosity tube. Any quantity of this pectic extract over 5 g per 100 cc produced a gel so stiff that it would not flow." [p 93] "The mixtures of pectic material and cuprammonium hydroxide were distinguished by their sparkling clarity, and their greatly reduced tendency to 'separate' as compared with the mixtures containing either raw cotton or both extracted pectic material and separated cellulose particles."

It was shown in Farr's data that if 1.87 grams of this powdered material were added to 100 cc of cuprammonium hydroxide, a viscosity approximately comparable to that of 0.25 g of cotton linters in the same quantity of cuprammonium hydroxide was obtained. In discussing these facts Farr states: [59, p. 91] "It must be emphasized, therefore, that in studying directly the viscosity-producing power of this extracted pectic substance we were using only one of the possible viscosity-producing components of the cementing material and that its original viscosity-producing power had been markedly reduced by extraction and repeated purification."

In commenting upon his observation that quantitative examination of variously treated plant fibers dispersed in cuprammonium solutions, using a slit ultramicroscope, reveals the presence of approximately the theoretical number of cellulose particles ($1.1 \times 1.5\mu$), Compton [42, pp. 66-67] states: "The two general methods employed for the gradual chemical disintegration of plant fibers are: first, the action of dilute acids, and second, the action of mild oxidizing agents. The disintegration of plant fibers in this manner is accompanied by a lowering of the cuprammonium viscosity (Table I) which has been previously interpreted as being due to a decrease in the chain length of the cellulose molecules. Based on the dual structure of plant fibers, however, a new interpretation is now placed on this behavior. Since the crystalline phase, or cellulose particles, remains largely intact by these treatments, it follows that the intercrystalline "cementing" material is the fiber phase most easily extracted or modified. The conclusion may also be drawn that this unstable, reactive phase of native plant fibers is largely responsible for the high viscosity when dis-

lose reaction of the particles. These microchemical distinctions are difficult to make but important to master, if a true understanding of the cell membrane is to be obtained. Observations of the structure of plant-cell membranes which have accumulated for more than a century have dealt, in many instances, with cells in which cellulose fibrils are conspicuous structural units. Descriptions of cell walls in which cellulose fibrils are not present, such as *Halicystis,* are not frequently found, although the earliest reports of the fine structure of cell membranes were upon cells of this type. The first reports of fibrils in membranes of various plant cells were based upon the fact that the microscopists *saw* them in the fresh, untreated tissues, commented upon their uniformity in diameter, and even followed their formation from the alignment of granules in single rows in the protoplasm of living cells.[262, 226, 198] With such techniques they were likewise able to distinguish membranes in which fibrils do not occur.[160] Wieler's recent emphasis upon the "honeycomb" structure of membranes will undoubtedly serve to broaden the current conceptions of plant-cell membrane structure in general. Likewise, the detailed observations of the finely fibrillar nature of the swollen intercrystalline membrane substance may help to throw some light upon the reports of the appearance of *anastomosing fibrils* grading down to and beyond the limits of microscopic visibility. In consideration of the fibrillar structure of cell membranes it is important that the cellulose fibrils, which can be seen to form within the plastids of living cells or from uniform-sized granules produced in such plastids, be distinguished as such from those finely fibrillar formations which make their appearance in advanced stages of membrane swelling. Little is known of the chemical composition of these fibrillar materials. Baier and Wilson [18] offer a possible clue in their descriptions of the formation of fibrous pectate pulps. Based upon these findings and observations of the high mineral content of the intercrystalline fiber material,[210] Marquette [165] has recently suggested that the finely fibrillar images of cotton and wood fibers as shown by the electron microscope may represent "ash-patterns" of the intercrystalline material, when high temperatures are involved.

Summary

1. One of the earliest known descriptions of the cellular structure of plant tissues is to be found in Hooke's "Micrographia," published in 1665. An illustration of a cross-section of cork tissue accompanies the description.

2. During the eighteenth century improvements in the construction of microscopes and the development of organic-chemical techniques laid the foundation for the important discoveries of the nineteenth century in the structure and composition of plants and animals.

3. The nature of the process by which plants produce complex organic substances from simple inorganic materials was a clearly defined, biological problem at the beginning of the nineteenth century. The dependence of both the plant and animal kingdoms upon these original syntheses was recognized but not understood.

4. The identification of the cell as the unit of both plant and animal structure in 1838 shifted the search for the seat of this synthesis from the plant as a whole to the single cells of its tissues.

5. The *chloroplast* in the protoplasm of the plant cell, not found in animal cells, was found to be the structure in which carbon dioxide and water are united to form sugar. This process, named "photosynthesis" in 1898, is the first step in the production of all simple and complex organic materials.

6. The classification of organic substances into *carbohydrates, proteins,* and *fats* by Prout made possible the identification of these substances in the plant cell before 1850 and marked the beginning of the science of *microchemistry*—the microscopic observation of chemical reactions.

7. The further classification of carbohydrates into sugar, starch, and cellulose

persed in cuprammonium solution. This interpretation is somewhat similar to that of Joyner,[16] who postulated the existence of a cellulose A with high and a cellulose B with low viscosity in native plant fibers."

Sisson, in commenting upon his observations upon the dispersion, electrokinetic and coagulation behavior of cotton fibers in cuprammonium hydroxide solution, states: [239, p. 121] "The most common visible unit of dispersion appeared to be small uniform-sized particles of cellulose. After complete fiber disintegration the apparent number and nature of the particles did not change even after storage of the solution in an ice box for two weeks." [p. 125] "In the fiber, the indications are that the cementing material [4] may form a true lyophilic colloid which may swell indefinitely with the dispersion medium. Cellulose particles on the other hand would appear to be lyophilic to the extent that cuprammonium solution enters into the cellulose particle to form a swelling compound, and lyophobic to the extent that during his process the particle itself is not measurably altered in size and shape."

These data indicated that the cuprammonium viscosity of cellulosic materials was not dependent upon any one of the many fiber components alone, but rather was dependent upon the colloidal system (cellulose particles, intercrystalline material) present in native fibers. In commenting upon the observations which led to the conclusions, Farr states: [59, p. 91] "These miscellaneous examples are cited in order to emphasize further the need for more knowledge of the viscosity-producing power of the non-cellulosic cell membrane constituents under a wide range of experimental conditions."

Further research in these fields should again lead to improvements in the techniques of separating, identifying, and maintaining the colloidal properties of the membrane components from many types of plant cell membranes and to a better understanding of the parts played by these components in the viscous states of cell membranes in many types of reagents. It could be expected that many of the anomalies of the viscosities of colloidal solutions would be clarified by a concerted attack with microscopic, chemical, x-ray diffraction, electron-diffraction, electrokinetic, and other techniques whose values have been established in these fields of colloid-chemical research. The parts played by the dispersed cellulose particles, the colloidal nature of the cuprammonium solution itself, and the swollen cementing material might thus be found to fall within known fields of colloidal behavior.

This experimental work with one extracted fraction of the non-cellulosic fiber material has been taken by Howells,[184] Heuser and Green,[180] and Hock and Harris [182] to mean that we * believe that the viscosity produced by cell membranes of cotton fibers in cuprammonium hydroxide is due *entirely* to the pectic material in the membrane. Their attempts to find dispersed particles of cellulose in cuprammonium hydroxide and to produce viscosities in cuprammonium hydroxide with any extracted membrane material other than cellulose have failed. These points will be settled only by further experimentation by other workers. One of the important preliminary studies for such an experiment is that of learning to recognize the cellulose particle in the living protoplasm of the cotton fiber and in disintegrated mature-cell membranes. While this technique is not exactly simple to employ, it is much less difficult than the identification of the crystallites in the midst of the swollen, blue matrix of cementing material in cuprammonium hydroxide solution.

Wieler's [267] recent comments concerning swelling reactions should also prove to be helpful in this connection. In describing the phenomena encountered in the study of swollen cotton-fiber membranes, he emphasizes the confusion produced by the mixture of fibrillar materials in the microscopic field. It was only after more careful manipulation of the reagents used in producing the swollen mass that he was able to trace the source of the finely fibrillar material to the inter-particle substance in the membrane and to distinguish its microchemical reactions from that of the cellu-

* Farr and coworkers.

lose reaction of the particles. These microchemical distinctions are difficult to make but important to master, if a true understanding of the cell membrane is to be obtained. Observations of the structure of plant-cell membranes which have accumulated for more than a century have dealt, in many instances, with cells in which cellulose fibrils are conspicuous structural units. Descriptions of cell walls in which cellulose fibrils are not present, such as *Halicystis,* are not frequently found, although the earliest reports of the fine structure of cell membranes were upon cells of this type. The first reports of fibrils in membranes of various plant cells were based upon the fact that the microscopists *saw* them in the fresh, untreated tissues, commented upon their uniformity in diameter, and even followed their formation from the alignment of granules in single rows in the protoplasm of living cells.[262, 226, 198] With such techniques they were likewise able to distinguish membranes in which fibrils do not occur.[160] Wieler's recent emphasis upon the "honeycomb" structure of membranes will undoubtedly serve to broaden the current conceptions of plant-cell membrane structure in general. Likewise, the detailed observations of the finely fibrillar nature of the swollen intercrystalline membrane substance may help to throw some light upon the reports of the appearance of *anastomosing fibrils* grading down to and beyond the limits of microscopic visibility. In consideration of the fibrillar structure of cell membranes it is important that the cellulose fibrils, which can be seen to form within the plastids of living cells or from uniform-sized granules produced in such plastids, be distinguished as such from those finely fibrillar formations which make their appearance in advanced stages of membrane swelling. Little is known of the chemical composition of these fibrillar materials. Baier and Wilson [13] offer a possible clue in their descriptions of the formation of fibrous pectate pulps. Based upon these findings and observations of the high mineral content of the intercrystalline fiber material,[210] Marquette [165] has recently suggested that the finely fibrillar images of cotton and wood fibers as shown by the electron microscope may represent "ash-patterns" of the intercrystalline material, when high temperatures are involved.

Summary

1. One of the earliest known descriptions of the cellular structure of plant tissues is to be found in Hooke's "Micrographia," published in 1665. An illustration of a cross-section of cork tissue accompanies the description.

2. During the eighteenth century improvements in the construction of microscopes and the development of organic-chemical techniques laid the foundation for the important discoveries of the nineteenth century in the structure and composition of plants and animals.

3. The nature of the process by which plants produce complex organic substances from simple inorganic materials was a clearly defined, biological problem at the beginning of the nineteenth century. The dependence of both the plant and animal kingdoms upon these original syntheses was recognized but not understood.

4. The identification of the cell as the unit of both plant and animal structure in 1838 shifted the search for the seat of this synthesis from the plant as a whole to the single cells of its tissues.

5. The *chloroplast* in the protoplasm of the plant cell, not found in animal cells, was found to be the structure in which carbon dioxide and water are united to form sugar. This process, named "photosynthesis" in 1898, is the first step in the production of all simple and complex organic materials.

6. The classification of organic substances into *carbohydrates, proteins,* and *fats* by Prout made possible the identification of these substances in the plant cell before 1850 and marked the beginning of the science of *microchemistry*—the microscopic observation of chemical reactions.

7. The further classification of carbohydrates into sugar, starch, and cellulose

permitted a more detailed study of these substances in living cells. *Sugar* and *starch* were found to be produced in chloroplasts and colorless plastids. In the earlier microchemical work *cellulose* was identified only in plant-cell membranes.

8. Observations of the fine structure of plant-cell membranes were made as early as 1819 by Lyngbye who reported "minutissime punctata" in the membrane of *Halicystis*. A drawing of this membrane accompanies his description and shows the arrangement of granules in the membrane similar to the arrangement which appears in a modern photomicrograph of the same membrane.

9. Fibrillar structures were observed in fresh plant-cell membranes before 1840, and their origin in the protoplasm from the alignment of minute particles in single rows was described.

10. These early observations showed that the plant-cell membrane increases in thickness by the deposition of visible granular or fibrillar structures from the protoplasm on the inner surface of the original limiting membrane of the cell.

11. Nägeli postulated his *micellar hypothesis* in 1858 to explain all the physical aspects of vegetative and reproductive phenomena upon the basis of submicroscopic unit structures—micellae [The micellae concerned in hereditary phenomena were specifically designated as "ids"]. All visible structures in the plant-cell membrane smaller than the fibril and all of its microchemical differentiations were ignored in the formulation of this micellar hypothesis. The membrane was described as consisting of submicroscopic crystalline micellae surrounded by films of water which could increase and decrease in thickness during alternate swelling and desiccation. Upon the basis of observations in polarized light, the micellae were found to be unchanged during the process of swelling.

12. These observations were at variance with the contemporary reports of physical and chemical heterogeneity of the plant-cell membrane materials made by Payen, Fremy, and others.

13. Strasburger suggested the name *microsomes* for the granular structures which appear in the protoplasm and are used up in the formation of the cell membrane. He identified them microchemically as protein granules.

14. Wiesner suggested the name *dermatosomes* for these structures and stated that although they are evidently protein while free in the protoplasm, they change, in some mysterious manner, into cellulose after they are deposited in the cell wall.

15. Important advances in the knowledge of the orientation of crystalline cellulose in plant-cell membranes were made by Ambronn between 1890 and 1925 through careful measurements in polarized light.

16. The crystalline nature of the cellulose component of plant-cell membranes was confirmed by x-ray diffraction analyses in 1913, and Ambronn's conclusions concerning the orientation in various types of cell membranes were rapidly corroborated.

17. By means of x-ray diffraction technique the first approximate measurements of the unit cell of cellulose were made by Sponsler in 1926.

18. X-ray diffraction analyses and measurements of viscosities of gelatinized cell membranes have been used in the development of a new theory based originally upon the conception of chemical and physical homogeneity of the cell membrane, but revised recently to represent chemical and physical inhomogeneity. It is now designated as the *fringed micellar hypothesis*. The gradual evolution of this hypothesis from the micellar hypothesis of Nägeli is shown in a series of diagrammatic sketches from the work of various authors.

19. The identification of crystalline cellulose in living protoplasm of cells from all parts of the plant kingdom was made by Farr and Eckerson in 1934. The unit crystallites were named *cellulose particles* and their identity with Lyngbye's "minutissime punctata," Valentin's *granules*, Strasburger's *microsomes*, and Wiesner's *dermatosomes* was established upon the basis of their location and general behavior

in the protoplasm during the formation of new lamellae and their presence in the mature membrane of the plant cell.

20. The presence of intercrystalline material in the cell membrane was shown by Farr, Eckerson, Sisson, and Compton to account for many of the physical and chemical properties of the cell membrane. Its chemical complexity and the possibility that it contains many protoplasmic constituents of the cell from which it is formed were early recognized.

21. The existence of cellulose crystallites of slightly smaller dimensions has been shown by Lüdtke, Hess, Wergin, Lieser, and others. The existence of cellulose particles of approximately the same size as those described by Farr and Eckerson has been shown by de Mosenthal, A. J. Bailey, Roberts and Haring, and Wieler.

22. The existence of visible cellulose crystallites in the protoplasm of living cells and in cell membranes has been denied by I. W. Bailey, Anderson and Kerr, Harris and his associates, and Berkeley. The physical and chemical behavior of cell membrane constituents in cuprammonium hydroxide as described by Farr, Compton, and Sisson has been denied by Harris and associates, Heuser and Green, and Howells.

23. The two polysaccharides, *starch* and *cellulose,* are formed simultaneously in separate plastids in the same cell. In certain cells the entire plastid plasma, including the chlorophyll, is deposited in the formation of a new lamella of the cell membrane.

24. The organization of the membrane lamella from cellulose crystallites and intercrystalline material of plastid and cytoplasmic origin takes place in the outer regions of the protoplast. Wieler has recently emphasized the importance of this function of the living protoplast which apparently has been overlooked by Frey-Wyssling, Wergin, Lüdtke, and others.

25. Current explanations of the physical and chemical properties of plant-cell membranes are made, in general, upon the basis of either submicroscopic "fringed micells," postulating continuous cellulose chain structure interspersed with more or less homogeneous, amorphous material, or upon the basis of microscopically visible cellulose particles or fibrils and a chemically heterogeneous cementing material which is composed of many of the protoplasmic constituents of the cell in which the membrane materials were formed.

26. The early identification of starch grains in chloroplasts linked their formation inextricably with biological processes of unknown physical and chemical nature and has served to protect them, in some measure, from attempts to explain their method of formation and chemical and physical properties upon a simple molecular basis.

27. The recent discovery of cellulose-forming plastids, in which the stages of development of cellulose particles are distinctly visible, now affords to cellulose a similar footing in observable phenomena, providing a brake to theoretical overspeculation.

28. The chemical and physical heterogeneity of plant-cell membranes indicates the necessity for improvements in the techniques of separation and identification of their intimately associated constituents, and in the measurement of the unknown forces which are involved in the synthesis and distribution of membrane-forming substances in the living plant cell.

References

1. Agardh, Jac. Geo., "De cellula vegetabili fibrillis tenuissimis contexta," 1852. 11 pp. Typis Berlingianis, Lundae.
2. Allen, C. E., "On the origin and nature of the middle lamella." *Bot. Gaz.,* **32**, 1-34 (1901).
3. Ambronn, H., "Anleitung zur Benützung des Polarizations-mikroskops bei histologischen Untersuchungen," 59 pp., J. H. Robolsky, Leipzig, 1892.
4. —, "Ueber das Zusammenwirken von Stäbchendoppelbrechung und Eigendoppelbrechung." I. *Zeitschr. Chem. u. Indus. Kolloide,* **18**, 90-97 (1916). II. *Ibid.,* **18**, 273-281 (1916); *Ibid.,* **20**, 173-185 (1917).

5. —, "Ueber Gleitflächen in Zellulosefasern," *Kolloid Zeitschr. Erg.*, (Zsigmondy Festschrift), **36**, 119-131 (1925).
6. Ambronn, H., and Frey, A., "Die Submikroskopische Struktur der Zellmembranen," *Jahrb. Wiss. Bot.*, **65**, 195-223 (1925-1926).
7. —, "Das Polarisationsmikroskop." 194 pp., Akademische Verlagsgesellschaft, Leipzig, 1926.
8. Anderson, D. B., "Über die Struktur der Kollenchymazellwand auf Grund mikrochemischer Untersuchungen," *Sitzungsber. Akad. Wiss. Wien*, **136**, 429-440 (1927).
9. —, "Some recent work on the structure of the plant cell wall," *Ohio J. Sci.*, **28**, 299-314 (1928).
10. Anderson, D. B., and Kerr, T., "Growth and structure of cotton fiber," *Ind. Eng. Chem.*, **30**, 48-54 (1938).
11. —, and Moore, J. H., "The influence of constant light and temperature on the structure of the walls of cotton fibers and collenchymatous cells," *Am. J. Bot.*, **24**, 503-507 (1937).
12. Astbury, W. T., Marwick, T. C., and Bernal, J. D., "X-ray analysis of the structure of the wall of *Valonia ventricosa*. I." *Proc. Roy. Soc. London B*, **109**, 443-450 (1932).
13. Baier, W. E., and Wilson, C. W., "Citrus pectates, properties, manufacture and uses," *Ind. Eng. Chem.*, **33**, 287-291 (1941).
14. Bailey, A. J., and Brown, R. M., "Diameter variation in cellulose fibrils," *Ind. Eng. Chem.*, **32**, 57-63 (1940).
15. Bailey, I. W., "Cell wall structure of higher plants," *Ind. Eng. Chem.*, **30**, 40-47 (1938).
16. Bailey, I. W., and Kerr, T., "The visible structure of the secondary wall and its significance in physical and chemical investigations of tracheary cells and fibers," *J. Arnold Arbor.*, **16**, 273-300 (1935).
17. Balls, W. L., "The existence of daily growth-rings in the cell wall of cotton hairs," *Proc. Roy. Soc. London B*, **90**, 542-555 (1919).
18. Barnes, C. R., "So-called 'assimilation,'" *Bot. Centralbl.*, **76**, 255-259 (1898).
19. Barnes, R. B., and Burton, C. J., "The electron microscope," *Am. Dyestuff Rept.*, **31**, 254-262 (1942).
20. Barrows, F. L., "Lamellate structure of cellulose membranes in cotton fibers," *Contrib. Boyce Thompson Inst.*, **11**, 161-179 (1940).
21. Behrens, H., "Anleitung zur mikrochemischen Analyse," Die wichtigsten Faserstoffe, 118 pp. L. Voss, Hamburg, 1908.
22. Behrens, J. W., "The microscope in botany. A guide for the microscopical investigation of vegetable substances," Trans. and ed. by A. B. Hervey and R. H. Ward, 466 pp., S. E. Cassino & Co., Boston, 1885.
23. Bensley, R. R., "Chemical structure of cytoplasm," *Science*, **96**, 390-393 (1942).
24. Berkeley, E. E., "Shrinkage and cell wall structure of cotton fibers," *Am. J. Bot.*, **29**, 416-423 (1942).
25. Berthold, G., "Ueber das Vorkommen von Protoplasma in interzellular Räumen," *Ber. Deutsch. Bot. Ges.*, **2**, 20 (1884).
26. Boehm, R. M., "The wood that improves on nature's best," Abstracts Div. Cellulose Chem., A. C. S., Chicago, Ill. (Sept., 1933).
27. Bonner, J., and Heyn, A. N. J., "Cell elongation and the electrical properties of the cell wall," *Protoplasma*, **24**, 466-469 (1935).
28. Borelli, P., "De vero telescopii inventore," Appendix Centuria observationum microscopicarum, 1665.
29. Bowman, F. H., "The structure of the cotton fiber and its relation to technical applications," 470 pp., Macmillan & Co., London, 1908.
30. Bracconot, H., "Recherches sur un novel acide universellement répandu dans tous les vegetaux," *Ann. Chim. Phys.*, **28**, 173-180 (1825).
31. —, "Mémoire sur le principe gélatineux des fruits precédé de quelques expériences sur le jus de groseille," *Ann. Chim. Phys.*, **47**, 266-273 (1831).
32. Bragg, W. L., "X-rays and crystal structure," London, 1915.
33. Branfoot, M. H., "A critical and historical study of the pectic substances of plants," Special Rept. No. 33, Dept. Sci. & Indus. Res., London, 1929.
34. Brown, F., "The silicious skeleton of tracheids and fibers," *Bull. Torrey Bot. Club*, **47**, 407-424 (1920).
35. Butler, M. R., "Comparison of the chemical composition of some marine algae," *Plant Physiol.*, **6**, 295-305 (1931).
36. Butterworth, J., "Cotton, and its treatment in the various processes of opening, carding, and spinning; being a full report of four papers read under the auspices of the Educational Department of the Cooperative Society, to which is added directions for the examination of cotton fibre, yarns, etc., under the microscope," 71 pp., Hirst & Rennie, King Street, Oldham, 1881.
37. Campbell, W. B., "Hydration and beating of cellulose pulp," Abstracts Div. Cellulose Chem., A. C. S., Washington, D. C., 1933. (Apr. 26-31.)
38. Carothers, W. H., "Polymerization," *Chem. Rev.*, **8**, 353-426 (1931).
39. Chardonnet, Count Hilaire de, "Artificial silk," British Patent No. 6,045 (1885).
40. Clark, G. L., "Cellulose as it is completely revealed by X-rays," *Ind. Eng. Chem.*, **22**, 474-487 (1930).
41. Clifford, A. T., and Cameron, F. K. "Dark-field study of fibers," *Ind. Eng. Chem.*, **26**, 1209-1213 (1934).
42. Compton, J., "On the behavior of plant fibers dispersed in cuprammonium hydroxide solution," *Contrib. Boyce Thompson Inst.*, **10**, 57-70 (1938).
43. —, "Structural relation of rayon to natural cellulosic fibers. Study of the viscose process," *Ind. Eng. Chem.*, **31**, 1250-1259 (1939).
44. —, "Studies on the developing cotton fiber. III," *Contrib. Boyce Thompson Inst.*, **11**, 403-420 (1941).
45. Compton, J., and Haver, F. E., Jr., "Studies on the developing cotton fiber. I." *Contrib. Boyce Thompson Inst.*, **11**, 105-118 (1939).
46. Correns, C., "Zur Kenntnis der inneren Struktur der vegetabilischen Zellmembranen," *Jahrb. Wiss. Bot.*, **23**, 254-338 (1892).
47. Cross, C. F., and Bevan, E. J., "Cotton cellulose. Structure and constitution," *J. Soc. Dyers Col.*, **34**, 215-222 (1918).
48. —, "Researches on cellulose," London, 1918.
49. Cross, C. F., and C. Dorée, "Researches on cellulose. IV." London, 1922.

50. Crum, W., "On the manner in which cotton unites with colouring matter. Third Memoir," *J. Chem. Soc.*, **16**, 404-414 (1863).
51. Czapek, F., "Biochemie der Pflanzen," Jena, 1913.
52. Dauphine, A., "Mise en évidence et localisation des substances proteiques dans la membrane vegetale," *Bull. Soc. Bot. France*, **81**, 328-331 (1934).
53. Debye, P., and Scherrer, P., "Interferenzen an regellos orientierten Teilchen im Röntgenlicht. III." *Physik. Zeitschr.*, **18**, 291-302 (1917).
54. Delage, Y., and Goldsmith, M., "The theories of evolution," New York, 1913.
55. Dippel, L., "Die intercellular Substanz und deren Entstehung," 50 pp., Rotterdam, 1857.
56. Ewart, A. J., "On the physics and physiology of protoplasmic streaming in plants," Oxford, 1903.
57. Farr, W. K., "Certain colloidal reactions of cellulose membranes," *J. Phys. Chem.*, **41**, 987-995 (1937).
58. —, "Structure of the cotton fiber," *J. App. Phys.*, **8**, 228-232 (1937).
59. —, "Behavior of the cell membrane of the cotton fiber in cuprammonium hydroxide solution," *Contrib. Boyce Thompson Inst.*, **10**, 71-112 (1938).
60. —, "Formation of cellulose particles in *Halicystis* sp." *Am. J. Bot.*, **26**, 1s (1939).
61. —, "Formation of microscopic cellulose particles in colorless plastids of the cotton fiber," *Contrib. Boyce Thompson Inst.*, **12**, 181-194 (1941).
62. —, and Eckerson, S. H., "Formation of cellulose membranes by microscopic particles of uniform size in linear arrangement," *Contrib. Boyce Thompson Inst.*, **6**, 189-203 (1934).
63. —, "Separation of cellulose particles in membranes of cotton fibers by treatment with hydrochloric acid," *Contrib. Boyce Thompson Inst.*, **6**, 309-313 (1934).
64. —, and Sisson, W. A., "X-ray diffraction patterns of cellulose particles and interpretations of cellulose diffraction data," *Contrib. Boyce Thompson Inst.*, **6**, 315-321 (1934).
65. —, "Observations on the membranes of epidermal cells of the *Avena* coleoptile," *Contrib. Boyce Thompson Inst.*, **10**, 127-137 (1939).
66. Fischer, H., "Bau und Eigenschaften pflanzlicher Kolloide. Grundlinien zu einer 'natürlichen' Kolloidtheorie," *Beitr. Biol. Pflanzen*, **15**, 327-356 (1927).
67. Flatters, A., "The cotton plant: its development and structure," 92 pp., Sherratt & Hughes, London, 1906.
68. Fort, M., "Breakdown of cotton fabrics under beetling treatment," *J. Soc. Dyers Col.*, **34**, 9-10 (1918).
69. Fremy, E., "Recherches sur les matières gelatineuses des végétaux," *Compt. Rend. Acad. Sci.*, **24**, 1046-1050 (1847).
70. —, "Mémoire sur la maturation des fruits," *Ann. Chim. Phys.*, **24**, 5-11 (1848).
71. —, "Chimie appliquée a la physiologie végétale. Recherches chimiques sur la composition des cellules végétales," *Compt. Rend. Acad. Sci.*, **48**, 202-212 (1859).
72. —, "Remarques à l'occasion de la précédente communication de M. Payen," *Compt. Rend. Acad. Sci.*, **48**, 319-325 (1859).
73. Freund, E. H., and Mark, H., "Internal structure of synthetic fibers and its influence on their physical properties," *Rayon Textile Monthly*, **23**, 59-61 (1942).
74. Freundlich, H., and Abramson, H. A., "Ueber die kataphoretische Wanderungsgeschwindigkeit gröberer Teilchen in Solen und Gelen. II." *Zeitschr. Physik. Chem.*, **133**, 51-68 (1928).
75. Freundlich, H., and Hauser, E. A., "Zur Kolloidchemie der Kautschukmilchsöfte," (Zsigmondy Festschrift), *Kolloid. Zeitschr. Erg.*, **36**, 15-36 (1925).
76. Freundlich, H., Stapelfeldt, F., and Zacher, H. "Quantitative Untersuchungen am Vanadinpentoxydsol," *Zeitschr. Physik. Chem.*, **114**, 161-207 (1924).
77. Frey, A., "Die heutige Stand der Micellartheorie," *Ber. Deutsch. Bot. Ges.*, **44**, 564-570 (1926).
78. —, "Die submikroskopische Struktur der Zellmembranen," *Jahrb. Wiss. Bot.*, **65**, 195-223 (1926).
79. —, "Der submikroskopische Feinbau der Zellmembranen." *Naturwiss.*, **15**, 760-765 (1927).
80. Frey-Wyssling, A., "Der Aufbau der pflanzlichen Zellwände," *Protoplasma*, **25**, 261-300 (1936).
81. —, "Fine structure of the plant cell wall," *Nature*, **142**, 899-904 (1938).
82. Gaidukov, N., "Ueber die Anwendung des Ultramikroskopes nach Seidentopf und des Mikrospektralphotometers nach Engelmann in der Textil- und Farbstoff-Industrie," *Z. Angew. Chem.*, **21**, 393-400 (1908).
83. Gaucher, L., "Etude générale de la membrane cellulaire chez les végétaux," 229 pp., Paris, 1904.
84. Gerngross, O., Hermann, K., and Lindemann, R., "Ueber die reversible Sol-Gel Umwandlung, die 'Kristallisation' der Gelatine und den Feinbau elastischer Gele insbesondere von Gelatine und Kautschuk im Lichte der Röntgenoptik," *Kolloid. Z.*, **60**, 276-288 (1932).
85. Gibson, W. H., [with Leo Spencer and Robert McCall], "The viscosity of solutions of cellulose," *J. Chem. Soc.*, **117**, 479-493 (1920).
86. Gilson, E., "La cristallisation de la cellulose et la composition chimique de la membrane cellulaire végétale," *Cellule*, **9**, 397-440 (1893).
87. Gleisberg, W., "Der gegenwärtige Stand der Membranforschung," *Bot. Centralbl. Beih., Abt. I*, **38**, 217-265 (1921).
88. Goncharov, S. V., and Burvasser, F. G., "Microstructure of cotton fibers," *Mem. Inst. Chem. Tech. Acad. Sci. Ukrain, S.S.R.* No. 9, 77-97 (1938).
89. Grafe, V., "Abderhalden's Biochem. Handlexikon," II, 1911.
90. —, "Chemie der Pflanzenzelle," Berlin, 1922.
91. —, "Zur physiologie der Pflanzenphosphatide," *Biochem. Z.*, **159**, 444-448 (1925).
92. —, "Über die Phosphatide aus *Daucus carota*," *Planta Arch. Wiss. Bot.*, **2**, 429-437 (1926).
93. Grafe, V., and Magistris, H., "Zur Chemie und Physiologie der Pflanzenphosphatide," *Biochem. Z.*, **162**, 366-398 (1925).
94. Graham, T., "Elements of chemistry: including the applications of the science in the arts," London, 1850.
95. Grew, N., "The anatomy of plants," London, 1682.
96. Guilliermond, A., "Récherches ultramicroscopiques sur les cellules végétales," *Rev. Gen. Bot.*, **42**, 193-204, 272-282 (1930).
97. Guth, E., and Rogovin, S., "Über die mechanischen Eigenschaften von Fäden und Filmen, die aus Zellulose und ihren Derivaten hergestellt wurden," *Akad. Wiss. Wien, Math. Naturwiss., Abt. II.*, **145**, 531-547 (1936).
98. Guthowsky, M., "Chemische Untersuchungen der Leinstengel," Just's "Bot. Jahresber.," Vol. I: p. 81-83 (1885).

99. Hall, A. G., "The physical and chemical properties of cellulose in relation to technical treatment of artificial silk," *J. Textile Inst.*, **21**, 7-8 (1930).
100. Hansteen-Cranner, B., "Über das Verhalten der Kulturpflanzen zu den Bodensalzen. III. Beiträge zur Biochemie und Physiologie der Zellwand lebender Zellen," *Jahrb. Wiss. Bot.*, **53**, 536-598 (1914).
101. —, "Beiträge zur Biochemie und Physiologie der Zellwand und der plasmolischen Grenzschichten," *Ber. Deutsch. Bot. Ges.*, **37**, 380-391 (1919).
102. —, "Zur Biochemie und Physiologie der Grenzschichten lebender Pflanzenzellen," Kristiana, 1922.
103. —, "Untersuchungen über die Frage ob in der Zellwand lösliche Phosphotide vorkommen," *Planta Arch. Wiss. Bot.*, **2**, 438-445 (1926).
104. Hanusek, T. F., "Microscopy of technical products," John Wiley and Sons, New York 1907.
105. Harrison, W., "Investigations on the physical and chemical properties of cotton," *Trans. Nat. Assoc. Cotton Mfgs.*, **101**, 200-249 (1916).
106. —, "Investigations on textile fibers," *Proc. Roy. Soc. London A*, **94**, 460-469 (1918).
107. Hartig, T., "Das Mikroskope," p. 586-596. 1859.
108. Harting, P., "On the artificial production of some of the principal organic calcareous formations," *Quart. J. Microscopic Sci.*, (N.S.) **12**, 118-123 (1872).
109. Haworth, W. N., and Machemer, H., "Polysaccharides. Part X. Molecular structure of cellulose," *J. Chem. Soc. (Part II)*, 2270-2277 (1932).
110. Hengstenberg, J., and Mark, H., "Über Form und Grösse der Mizelle von Zellulose und Kautschuk," *Z. Krist.*, **69**, 271-284 (1928).
111. Hermann, K., Gerngross, O., and Abitz, W., "Zur röntgenographischen Strukturerforschung des Gelatinemicells," *Z. Physik. Chem. Abt. B.*, **10**, 371-395 (1930).
112. Herschel, J. F. W., "On certain motions produced in fluid conductors when transmitting the electric current," *Phil. Trans. Roy. Soc., London*, **114**, 162-196 (1824).
113. Herzog, A., "Ueber die Bastzellen aus dem Hypokotyl der Flachspflanze," *Z. Farben- u. Textil-Indus.*, **3**, 377-382, 402-409 (1904).
114. —, "Zur Kenntnis der Doppelbrechung der Baumwollfasern," *Z. Chem. u. Indus. Kolloide*, **5**, (1909).
115. —, "Ueber das mikroskopische Verhalten der Baumwolle in Kupferoxydammoniak," *Kunststoffe*, **1**, 401-404, 424-427 (1911).
116. —, "Mikroskopische Studien über Baumwolle," *Chem. Zeit.*, **38**, 111-116, 1089-1091, 1097-1100 (1914).
117. —, "Zur Untersuchung der Faserstoffe im polarisierten Licht." *Textile Forsch.*, **2**, 52 (1920).
118. Herzog, R. O., "Ueber den Feinbau der Faserstoffe," *Naturwiss.*, **12**, 955-960 (1924).
119. —, "Possibilities in the constitution of cellulose," *Cellulosechem.*, **6**, 39-40 (1925).
120. —, "Zur Erkenntnis der Cellulose-Faser," *Ber. Deutsch. Chem. Ges.*, **68**, 1254-1262 (1925).
121. —, "The nature of the structure of cellulose and its significance in chemical transformations," *J. Phys. Chem.*, **30**, 457-469 (1926).
122. —, "Recent progress in the knowledge of fibrous materials," *Papierfabr.*, **24**, 137-142 (1926).
123. —, "X-ray investigations on cellulose," *Paper Trade J.*, **83**, 51-55 (1926).
124. Herzog, R. O., and Jancke, W., "X-ray spectroscopic investigations on cellulose," *Z. Physik*, **3**, 196-198 (1920).
125. Hess, K., "Morphologie und Chemismus bei den organischen hochmolekularen Naturstoffen," *Naturwiss.*, **22**, 469-476 (1934).
126. —, "Ueber neue Ergebnisse im Aufbau der pflanzlichen Zellwand," *Papierfabr.*, **4**, 25-28 (1939).
127. Hess, K., Kiessig, H., Wergin, W., and Engel, W., "Zur Kenntnis der Bildung von Zellulose in der Zellwand," *Ber. Deutsch. Chem. Ges.*, **72**, 642-652 (1939).
128. Hess, Kurt, Max Lüdtke, and Herbert Rein, "Über Cellulose aus jungen Trieben und altem Kernholz. V. Mitt. Über Begleitstoffe der Cellulose," Justus Liebig's *Ann. Chem.*, **466**, 27-72 (1928).
129. Heuser, E., "The nature of cellulose. An historical review," *Am. Dyestuff Reporter*, **25**, 55-60, 80-83, 116-120, 135-137, 315-318, 338-339 (1936).
130. Heuser, E., and Green, J. W., "Behavior of cotton fiber with ammonium oxalate and cuprammonium solution," *Ind. Eng. Chem.*, **33**, 868-872 (1941).
131. Heyn, A. N. J., "Further investigations on the mechanism of cell elongation and the properties of the cell wall in connection with elongation." *Protoplasma*, **19**, 78-96 (1933).
132. Hock, C. W., and Harris, M., "Microscopic examination of cotton fibers in cuprammonium hydroxide solutions," *Textile Res.*, **10**, 323-333 (1940).
133. Hooke, R., "Micrographia," London, 1665.
134. Howells, T. A. "A study of wood pulp-cuprammonium hydroxide system," *Paper Trade J.*, **113**, 88-93 (1942).
135. Hungerbühler, J., "Zur Kenntnis der Zusammensetzung nicht ausgereifter Kartoffelknollen," *Landw. Versuchs-Stationen*, **32**, 381-388 (1886).
136. Ingen-Housz, J., "Experiments on vegetables, discovering their great power of purifying common air in the sunshine and injuring in the shade and night," London, 1779.
137. Iterson, G. van, Jr., "Structure of the wall of *Valonia*," *Nature*, **138**, 364-365 (1936).
138. —, "A few observations on the hairs of the stamens of *Tradescantia virginica*." *Protoplasma*, **27**, 190-211 (1937).
139. Johnson, D. S., "The crystallization of cellulose," *Bot. Gaz.*, **20**, 16-22 (1895).
140. Kabsch, W., "Untersuchungen über die chemische Beschaffenheit der Pflanzengewebe," *Pringsh. Jahrb. Wiss. Bot.*, **3**, 357-399 (1863).
141. Kerr, T., "The structure of the growth rings in the secondary wall of the cotton hair," *Protoplasma*, **27**, 229-241 (1937).
142. Khouvine, Y., "Cellulose et bactéries. Décomposition et synthèse," *Actualitiés Scientifiques et Industrielles*, **164**, 1-44 (1934).
143. Kirchof, M., "Formation du sucre dans les graines céréales convertes, en malt, et dans la farine infusée dans l'eau bouillante," *J. de Pharm.*, **II**, 250-258 (1816).
144. Kisser, J., "Der heutige Stand botanisch-mikrotechnicher Schneidemethoden," *Biol. Gen.*, **4**, 131-180 (1928).
145. Klebs, G. A., "Ueber die Fortpflanzungsphysiologie der niederen Organismen," Jena, 1896.
146. Knecht, E., and Hall, W., "Further investigations on the constituents of raw cotton," *J. Soc. Dyers & Col.*, **34**, 220-225 (1918).
147. Krabbe, G., "Ein Beitrag zur Kenntnis der Struktur und des Wachstums vegetabilischer Zellhäute," *Jahrb. Wiss. Bot.*, **18**, 346-423 (1887).
148. Krasser, F., "Untersuchungen über das Vorkommen von Eiweiss in der pflanzlichen Zellhaut, nebst Bemerkungen über den mikrochemischen Nachweis der Eiweisskörper," *Sitzungsber. Akad. Wiss. Wien.*, **93**, 118-155 (1886).
149. Kratky, O., "Über die mizellaren Aufbau und die Deformationsvorgänge bei Faserstoffen," *Kolloid. Z.*, **70**, 14-19 (1935).

150. —, "Über den mizellaren Aufbau und die Deformationsvorgänge bei Faserstoffentheorie der Netzstrukturen," *Kolloid. Z.*, **84**, 149-168 (1938).
151. —, "Micellar structure of cellulose and its derivatives," *Angew. Chem.*, **53**, 153-162 (1940).
152. Kützing, T. F., "Phycologia generalis," 1843.
153. Laue, M., "Eine quantitative Prüfung der Theorie für die Interferenzerscheinungen bei Röntgenstrahlen," *Ann. Phys.*, **41**, 989-1002 (1913).
154. Leeuwenhoek, A. von, "Brieven," Delft, 1663.
155. Lester, J. H., "The hygroscopic nature of cotton," *J. Soc. Chem. Indus.*, **21**, 388-389 (1902).
156. Levine, B. S., "The importance of a consideration of the fiber proteins in the process of bleaching cotton," *Science* (*N. S.*), **41**, 543-545 (1915).
157. Link, H. F., "Icones anatomico. Botanicae ad illustranda elementa philosophiae botanicae," *Cum Tabulis Lithographicis*, **VIII**, (1837).
158. Loeb, J., "The influence of electrolytes on the cataphoretic charge of colloidal particles and the stability of their suspensions," *J. Gen. Physiol.*, **5**, 395-413 (1923).
159. Lüdtke, M., "Über einige neue Methoden der Faserforschung," Faserforschung, **1**, 43-58 (1932).
160. Lyngbye, H. C., "Teutamen hydrophytologiae danicae," 1819.
161. Malpighii, M., "Anatome plantarum," 1675.
162. Mangin, L., "Sur la constitution de la membrane des végétaux," *Compt. Rend. Acad. Sci.*, **109**, 579-582 (1889).
163. —, "Etude historique et critique sur la présence des composés pectiques dans les tissues des végétaux," *J. Bot.*, **5**, 400-440 (1891).
164. Mark, H., and Meyer, K. H., "Über die Kristallstruktur der Cellulose," *Z. Physik. Chem. Abt. B*, **36**, 232-236 (1937).
165. Marquette, W., "Review on the structure of cellulose fibers," *Rayon Textile Monthly*, **23**, 62-63 (1942).
166. Matthews, J. M., "The Textile Fibers: Their Physical, Microscopic and Chemical Properties," 592 pp., New York, 1913.
167. Meyen, F. J. F., "Neues System der Pflanzenphysiologie," 1837-1839.
168. Meyer, A., "Ueber die Struktur der Stärkekörner," *Bot. Zeit.*, **39**, 841-852, 856-864 (1881).
169. —, "Über die Assimilationsproducte der Laubblätter angiospermer Pflanzen," *Bot. Zeit.*, **43**, 417-423, 433-440, 449-458, 465-472, 481-491, 497-507 (1885).
170. —, "Bildung der Stärkekörner in den Laubblättern aus Zuckerarten, Mannit, und Glycerin," *Bot. Zeit.*, **44**, 81-88, 105-113, 129-137, 145-151 (1886).
171. —, "Untersuchungen über die Stärkekörnerwesen und Lebensgeschichte der Stärkekörner der hoheren Pflanzen," 318 pp., Jena, 1895.
172. Meyer, K. H., and Mark, H., "Der Aufbau der Hochpolymeren Organischen Naturstoffe auf grund molikular-morphologischer Betrachtungen," 264 pp., Akademische Verlagsgesellschaft, Leipzig, 1930.
173. Minaev, V. A., Verbal report of microchemical studies of cotton carried on in 1915 and reported at first all-union conference of laboratory workers of the textile industry in Moscow in June, 1930.
174. Minajeff, W., "Ueber das erhöhte Anfärben der mercerisierten Baumwolle und dessen Ursachen. Microskopische Untersuchung," *Z. Farben-Indus.*, **6**, 233-237 (1907).
175. Mohl, H. von, "Erläuterung und Verteidigung meiner Ausicht von der Pflanzen-Substanz," 1836.
176. —, "Vermischte Schriften," 1845.
177. —, "Principles of the anatomy and physiology of the vegetable cell," Trans. by Arthur Humfrey. 158 pp., John Van Voorst, London, 1852.
178. Molisch, H., "Grundriss einer Histochemie der pflanzlichen Genussmittel.," Jena, 1891.
179. —, "Mikrochemie der Pflanzen," Gustav Fischer, Jena, 1913.
180. Mosenthal, H. de, "Observations on cotton and nitrated cotton," *J. Soc. Chem. Ind.*, **23**, 292-298 (1904).
181. Müller, H. C., "Entstehung von Kalkoxalat in Zellmembranen," Leipziger Diss., 1890.
182. Nägeli, C. von, "Die Stärkekörner. Morphologische, physiologische, chemisch-physicalische und systematisch-botanische Monographie," 623 pp. Friedrich Schulthess, Zurich, 1858.
183. —, "Die Anwendung des Polarizationsmikroskops auf die Untersuchung der organischen Elementaartheile," *Beitr. Wiss. Bot.*, **3**, 1-126 (1863).
184. —, "Über den inneren Bau der vegetabilische Zellmembranen," *Sitzungsber. Bayerische Akad. Wiss. München*, **1**, 282-326 (1864).
185. Nageli, C. von, and Schwendener, S., "Das Mikroskop," 416 pp., 1877.
186. Neale, S. M., "The modification of cotton cellulose by swelling and by natural degradation," *Melliand Textile Monthly*, **4**, 507-508 (1932).
187. Nickerson, R. F., "Cotton fibers, constitution, structure and mechanical properties," *Rubber Chem. Tech.*, **14**, 273-288 (1941).
188. Nishikawa, S., and Ono, S., "The transmission of X-rays through fibrous, lamellar and granular substances," *Proc. Tokyo Math.-Phys. Soc.*, **7**, 131-138 (1913).
189. Noll, F., "Untersuchungen über die Wachstum der Zellmembran," 1887.
190. Norman, A. G., "The biochemistry of cellulose," Oxford Univ. Press, 1937.
191. Ord, W. M., "The influence of colloids upon crystalline form and cohesion," *Quart. J. Microsc. Sci.*, (1872).
192. Ost, H., "Die Viscosität der Celluloselösungen," *Z. Angew. Chem.*, **24**, 1892-1896 (1911).
193. Palladin, V. I., "Kohlenhydrate als Oxidationsproducte der Eiweissstoffe," *Ber. Deutsch. Bot. Ges.*, **7**, 126-130 (1889).
194. Payen, A., "Mémoire sur les applications theoriques et practiques des propriétés du tissue élémentaire des végétaux. *Compt. Rend. Acad. Sci.*, **8**, 59-61 (1839).
195. —, "Mémoire sur les developements des végétaux," Paris, 1842.
196. —, "Note sur la composition immediate de l'epiderme et de la cuticle epidermique des végétaux," *Compt. Rend. Acad. Sci.*, **42**, 1193-1195 (1856).
197. —, "Amidon et cellulose: observations sur des analogies remarquables et des differences characteristiques entre ces deux principes immediats," *Compt. Rend. Acad. Sci.*, **48**, 67-75 (1859).
198. —, "Observations sur les tissues végétaux, nouveau caractère distinctif entre la cellulose et l'amidon. *Compt. Rend. Acad. Sci.*, **48**, 319-325 (1859).
199. Pfeffer, W., "Studien zur Energetik der Pflanzen," *Abhandl. K. Sächs. Ges. Wiss. Leipzig*, **18**, 149-276 (1893).
200. Pfitzer, E., "Ueber die Einlargerung von Kalkoxalat-Krystallen in die pflanzliche Zellhaut," *Flora*, **55**, 130-137 (1872).
201. Preston, R. D., "The structure of the walls of parenchyma in *Avena* coleoptiles," *Proc. Roy. Soc. London B*, **125**, 372-386 (1938).
202. —, "The molecular chain structure of cellulose and its botanical significance," *Biol. Rev. Cambridge Phil. Soc.*, **14**, 281-313 (1939).
203. Priestley, J., "Experiments and observations on different kinds of air," 342 pp., 1774.

204. Prout, W., "On the ultimate composition of simple elementary substances; with some preliminary remarks on the analysis of organized bodies in general," *Proc. Roy. Soc. London,* **23**, 355-388 (1827).
205. Rainey, G., "The mode of formation of shells of animals, of bone, and of several other structures, by molecular coalescence, demonstrable in certain artificially formed products," *Quart. J. Microsc. Sci.,* (Jan., 1858).
206. Reinders, W., and Bendien, W. M., "Influence of lyophilic colloids on the stability and the electric charge of lyophobic sols," *Rec. Trav. Chim.,* **47**, 977-988 (1928).
207. Reiss, R., "Ueber die Natur der Reservecellulose und über ihre auflösungsweise bei der Keimung der Samen," *Ber. Deutsch. Bot. Ges.,* **7**, 322-329 (1889).
208. Reissek, S., "Die Fasergewebe des Leines, des Hanfes, der Nessel und Baumwolle," *Denkschrift. Akad. Wiss. Wien. Math. Naturwiss. Kl.,* **4**, 127-178 (1852).
209. Remec, B., "Über die specifische Doppelbrechung der Pflanzenfasern," *Sitzungsber. Akad. Wiss. Wien. Math. Naturwiss. Kl.,* **110**, 364-387 (1901).
210. Ritter, G. J., "The microstructure of cellulose." *Rayon Melliand Textile Monthly,* **16**, 58-60, 98-100 (1935).
211. Roberts, E. A., and Haring, I. M., "The water relations of the cell walls of certain mosses as a determining factor in their distribution," *Ann. Bryolog.,* **4**, 97-114 (1937).
212. Ruge, U., "Zur Charakteristik einer für die Physiologie der Zellstrekung wichtigen intermicellar Substanz pflanzlicher Membranen," *Biochem. Z.,* **295**, 29-43 (1937).
213. Ruska, H., and Kretschmer, M., "Über mikroskopische Untersuchungen über den Abbau von Cellulosefasern," *Kolloid. Z.,* **93**, 163-166 (1940).
214. Russel-Wells, B., "Chemistry of the red and brown algae," *Nature,* **133**, 651 (1934).
215. Russow, E., "Ueber den Zusammenhang der Protoplasmakörper benachbarten Zellen," *Sitzungsber. Dorpat. Naturforscher. Ges.,* **5**, 305-325 (1882).
216. Sachs, J., "Über den Einfluss des Lichtes auf die Bildung des Amylums in den Chlorophyllkörnern," *Bot. Zeit.,* **20**, 365-373 (1862).
217. —, "Über die Auflösung und Wiederbildung des Amylums in den Chlorophyllkörnern bei wechselnder Beleuchtung," *Bot. Zeit.,* **22**, 289-294 (1864).
218. —, "Ein Beitrag zur Kenntnis der Ernahrungstätigkeit der Bjätter," *Arb. Bot. Inst. Wurzburg,* **3**, 1-33 (1884).
219. Saussure, N. T. de, "Récherches chimiques sur la végétation," Paris, 1804.
220. Sauter, E., "Beiträge zur Röntgenographie und Morphologie der Cellulose," *Z. Physik. Chem. Abt. B,* **35**, 83-116 (1937).
221. Schacht, H., "Beiträge zur Anatomie und Physiologie der Gewächse," 1854.
222. Scherrer, P., "Kolloidchemie. Ein Lehrbuch von Richard Zsigmondy," 429 pp., Leipzig, 1920.
223. Schimper, A. F. W., "Untersuchungen über die Entstehung der Stärkekörner," *Bot. Zeit.,* **38**, 881-902 (1880).
224. —, "Untersuchungen über das Wachstum der Stärkekörner," *Bot. Zeit.,* **39**, 185-194, 201-211, 217-227 (1881).
225. —, "Über die Bildung und Wanderung der Kohlenhydrate in den Laubblättern," *Bot. Zeit.,* **43**, 737-743, 753-763, 769-787 (1885).
226. Schleiden, M. J., "Beiträge zur Phytogenesis," Müller's "Arch. Anat. u. Physiol." (Part II), 1838.
227. Schmitz, F., "Die Chromatophoren der Algen," 180 pp., Bonn, 1882.
228. Schorger, A. W., "The gelatinization of lignocellulose." *Ind. Eng. Chem.,* **16**, 141-144 (1924).
229. Schulze, E., "Zur Chemie der pflanzlichen Zellmembranen," Hoppe-Seyler's *Z. Physiol. Chem.,* **16**, 387-438 (1892).
230. Schulze, E., Steiger, E., and Maxwell, W., "Zur Chemie der Pflanzenzellmembranen," Hoppe-Seyler's *Z. Physiol. Chem.,* **14**, 227-273 (1890).
231. Schunck, E., "On some constituents of the cotton fiber," *Mem. Manchester Lit. & Phil. Soc.,* **4**, 95-128 (1871).
232. Schwann, T., "Mikroskopische Untersuchungen über die Übereinstimmung in der Struktur und dem Wachstum der Thiere und Pflanzen," 1839. [Prelim. statement. Froriep's notizen, No. **91**, 103, 112 (1838).]
233. Schweizer, E., "Das Kupferoxyd-Ammoniak, ein Auflösungs-mittel für die Pflanzenfaser," *J. prakt. Chem.,* **72**, 109-111 (1857).
234. Sears, G. R., and Kregel, E. A., "Application of the electron microscope to problems of the pulp, paper and paperboard industry," *Paper Trade J.,* **114**, 43-49 (1942).
235. Seifriz, W., "The origin, composition, and structure of cellulose in the living plant," *Protoplasma,* **21**, 129-159 (1934).
236. Sharp, L. W., "An introduction to cytology," 2nd ed., 581 pp., McGraw-Hill Book Co., Inc., New York, 1926.
237. Sisson, W. A., "Identification of crystalline cellulose in young cotton fibers by X-ray diffraction analysis," Contrib. Boyce Thompson Inst. **8**, 389-400 (1937).
238. —, "The existence of mercerized cellulose and its orientation in *Halicystis* as indicated by X-ray diffraction analysis," *Science,* **87**, 350 (1938).
239. —, "Some observations upon the dispersion, electrokinetic and coagulation behavior of cotton fibers in cuprammonium hydroxide solution," Contrib. Boyce Thompson Inst., **10**, 113-126 (1938).
240. —, "X-ray studies regarding the structure and behavior of native cellulose membranes," *Chem. Rev.,* **26**, 187-201 (1940).
241. Spierer, C., "Un nouvel ultra-microscope," *Arch. Sci. Phys. Natur.,* **8**, 121-131 (1926).
242. Sponsler, O. L., "X-ray diffraction patterns from plant fibers," *J. Gen. Physiol.,* **9**, 221-233 (1925).
243. —, "Molecular structure-of plant fibers determined by X-rays," *J. Gen. Physiol.,* **9**, 677-695 (1926).
244. —, "The molecule in biological structures as determined by X-ray methods," *Quart. Rev. Biol.,* **8**, 1-30 (1933).
245. —, "Rate of carbohydrate condensation on the cell wall," Abstracts Div. Cellulose Chem. A.C.S., Chicago, Ill. Symposium on the synthesis and structure of the cell wall of wood and the constitution and industrial application of lignin. 1933.
246. —, "Rate of carbohydrate condensation on a cell wall," *J. Am. Chem. Soc.,* **56**, 1599-1601 (1934).
247. Sponsler, O. L., and Dore, W. H., "The structure of ramie cellulose as derived from X-ray data," "Colloid Symposium Monographs," Vol. **4**, pp. 174-202, Reinhold Publishing Corp., 1926.
248. —, "X-ray studies on the structure of compounds of biochemical interest," *Ann. Rev. Biochem.,* **5**, 63-80 (1936).
249. Staudinger, H., "Die Chemie der hochmolekularen organischen Stoffe im Sinne der Kekuléschen Strukturlehre," *Angew. Chem.,* **42**, 67-73 (1929).
250. —, "Ueber hochpolymere Verbindungen. Ueber die Konstitution der Cellulose," *Cellulose Chem.,* **15**, 53-59 (1934).
251. Staudinger, H., and Reinecke, F., "Über die polymerisationsgrad von Zellulosen in alten Geweben," *Melliand Textilber.,* **20**, 109-110 (1939).

252. Staudinger, H., and Singer, R., "Über den Kristallbau hochmolekularer Verbindungen," *Zeit. Kristallograph, Abt. A*, **70**, 193-210 (1929).
253. Stiles, W., "The constitution of plant cell membranes," *Trans. Faraday Soc.*, **33**, 923-927 (1937).
254. Strasburger, E., "Ueber den Bau und das Wachstum der Zellhäute," 264 pp., Gustav Fischer, Jena, 1882.
255. —, "Ueber das Wachstum vegetabilischer Zellhäute," *Histol. Beitr.*, **2**, 1-186 (1889).
256. Terletzki, P., "Ueber den Zusammenhang des Protoplasmas benachbarter Zellen und über Vorkommen von Protoplasma in Zwischenzellräumen," *Ber. Deutsch. Bot. Gesell.*, **2**, 169-171 (1884).
257. Thiessen, P. A., "Von neuen Ergebnissen und Fragestellungen der Kolloidforschung," *Angew. Chem.*, **51**, 318-320 (1938).
258. —, "Physical structure of coal, cellulose fiber, and wood as shown by Spierer lens," *Ind. Eng. Chem.*, **24**, 1032-1041 (1932).
259. Tiselius, A., "A new apparatus for electrophoretic analysis of colloidal mixtures," *Trans. Faraday Soc.*, **33**, 524-531 (1937).
260. Tröndle, A., "Über die Kopulation und Keimung von *Spirogyra*," *Bot. Zeit.*, **65**, 187-216 (1907).
261. Tschirch, A., "Angewandte Pflanzenanatomie, I." 548 pp., Urban & Schwarzenberg, Wien, 1889.
262. Valentin, G., "Repertorium für Anatomie und Physiologie," Berlin, 1837.
263. Wergin, W., "Über den Aufbau pflanzlicher Zellwände. V. Mitteilung: Untersuchungen über die Baueinheiten mit Hilfe der Quellungsanalyse," *Protoplasma*, **32**, 116-139 (1939).
264. Wheeler, E., "The manufacture of artificial silk and its relation to colloid chemistry." "Fifth report on colloid chemistry and its general and industrial applications," *Brt. Assoc. Adv. Sci. Rept.*, pp. 50-71, 1923.
265. Whistler, R. L., Martin, A. R., and Harris, M., "Determination of uronic acids in cellulosic materials," *Textile Res.*, **10**, 109-119 (1940).
266. —, "Pectic substance in cotton and its relation to the properties of the fiber," *J. Res. Nat. Bur. Stand.*, **24**, 555-565 (1940).
267. Wieler, A., "Die Bedeutung der Innenhaut für die Zelle und die Struktur der sekundaren Verdickungsschichten. *Protoplasma*, **34**, 202-236 (1940).
268. Wiesner, J., "Untersuchungen über die Organization der vegetabilischen Zellhaut," *Sitzungsber. Akad. Wiss. Wien. Math. Naturwiss. Kl. Abt. I*, **93**, 17-80 (1886).
269. Wigand, A., Schr. gesell. z. Beförd d. ges. Naturwiss. zu Marburg, 1856.
270. Wisselingh, C. van, "Sur la paroi des cellules suberuses," *Arch. Néerl. Sci. Ex. Nat.*, **22**, 253-298 (1888).
271. —, "Die Zellmembran," 264 pp., Gebrüder Borntraeger, Berlin, 1924. ("Handbuch der Pflanzenanatomie," v. 3/2.)

Recent Advances in Starch Chemistry

R. M. HIXON AND R. E. RUNDLE

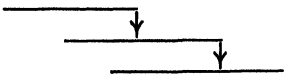
Department of Chemistry, Iowa State College, Ames, Iowa

Basic to the study of starch is a knowledge of the nature and degree of its heterogeneity. That starch is not homogeneous was claimed as early as 1819 by de Saussure on the basis of differences in solubility of component parts of the granule. The conflicting systems of nomenclature which resulted from the numerous subsequent attempts at fractionation are reported in the more extensive reviews by Alsberg,[1] Hanes,[2] Radley,[3] Samec[4] and Walton.[5] The difficulties were due largely to lack of methods for characterizing the component fractions. There were no criteria for complete fractionation and no methods of interpreting the results, even of excellent fractionations such as that of Baldwin.[6] The use of unfractionated or partially fractionated starch in the numerous studies by methylation techniques confused the issue until the heterogeneous character of these materials was clearly recognized.

Recently K. Meyer[7] has given new significance to work on the fractionation of starch by hot-water extraction and retrogradation. He has given a clear differentiation of the components on a structural basis and shown that starch consists of branched and unbranched molecules. The branched component has a much greater molecular weight than the unbranched, and constitutes the major portion of the starch. Excellent summaries of these studies have been published by K. Meyer[8,9] in the last two years. He has used the older terms "amylose" and "amylopectin" to designate the two components in starch. Retention of the terms will retain some of the confusion associated with their use in the older literature. Criticism might also be raised to retention of "pectin" in the nomenclature. Nevertheless, the fact that definite structural interpretations have now been given to the terms makes it seem

advisable to continue that usage rather than to introduce new terminology. Accordingly, in this summary the term "amylose" will be used for the unbranched component in starch, and "amylopectin" for the branched component.

The concept of branched molecules is not new. It was suggested as possible for xylan [10] in 1934 and for glycogen [11] in 1937 to explain the presence of monomethyl xylose and dimethyl glucose after hydrolysis of the methylated polysaccharides. The laminated formulation



was also used in the latter publication to illustrate such a branched structure, the arrowhead representing the reducing end of a chain of glucoses linked to unspecified positions in the other chains. At the time, however, there was still reservation as to complete methylation of the polysaccharide. There was also reservation as to whether the linkage between chains was due to an ordinary covalent bond or to a "coordinated" or "hydroxyl" bond.

The present evidence for the structure of starch has also come from a number of other sources. Staudinger and Eilers [12] furnished experimental evidence that the large molecules of starch exist as definite molecular entities rather than indefinite micelles composed of many small molecules held together by secondary valences. In this work, partially degraded starch was acetylated; the acetylated starch was in part regenerated and in part converted to methylated starch. Measurements of molecular size in several solvents gave comparable values for the original starch, and the acetylated, the regenerated, and the methylated starch. Following this work, Staudinger and Husemann [13] pointed out the value of a branched structure as an explanation for the discrepancies between the physical measurements and end-group assay in determining molecular size. In 1940, Freudenberg and Boppel [14] identified the dimethyl glucose obtained in the hydrolysis of methylated starch as 2, 3-dimethyl glucose, and concluded that α-1,6-glucosidic linkages must be present in addition to the customary α-1,4-glucosidic linkages.

The Amylose and Amylopectin Components of Starch

The differentiation between branched and unbranched molecules in starch has been confirmed by physical methods.[15] It has been shown that the amylose component (unbranched molecules) is oriented easily by flow while the amylopectin component (branched molecules) is oriented with difficulty, if at all. While viscosity results have not been published, sufficient data are at hand to state that these studies confirm the linear and non-linear character of the two components.[62] The difference is also shown in a striking manner by the filming properties of the two fractions and by the characters of their acetyl derivatives; the fibrous, cotton-like character of amylose acetate as compared to the very brittle quality of amylopectin acetate is apparent even to the non-technical observer.

The molecular species responsible for the heterogeneity of starch may vary in size and in degree of branching. Several properties of starch and starch fractions indicate that, so far as branching is concerned, only two fractions are present. The amylose component is adsorbed preferentially on cellulose; [16] it is precipitated in crystalline form by Schoch's butanol procedure; [17] it forms a complex with iodine in water solutions which is quite different from the complex formed by amylopectin.[18]

The reaction with iodine has been made the basis of a quantitative determination of the amylose component. The iodine activity is measured potentiometrically.[18] When iodine is added to an amylopectin solution, the iodine activity rises rapidly.

Upon adding iodine to an amylose solution, the iodine activity of the solution remains constant as long as complex formation takes place, and then rises with additional iodine as it does for water alone. Potentiometric titration curves of this type are shown in Figs. 1 and 2 for various starches, as well as for purified amyloses and amylopectins. If the quantity of iodine bound by Kerr's crystalline amylose[33] is taken as a standard, it is possible to estimate the amylose content of the different starches. The variation is considerable, as shown in Fig. 1 and summarized in Table 1:

Table 1. Amylose Contents of Starches Determined by Iodine Titration*

Starch	Amylose (%)	Starch	Amylose (%)
Waxy rice	0	Corn	21*
Waxy sorghum	0	Potato	22
Waxy corn	0	Popcorn	23
Waxy barley	0	Wheat	24
Tapioca	17	Sago	27
Rice	17	Lily bulb	34
Banana	20.5		

Upon examining any of the above-mentioned properties of starch, no evidence is found for components of intermediate properties, and it appears that none are present. It has been pointed out that amylopectin molecules must be quite branched in all portions of the molecule, since a long unbranched portion should act like an amylose molecule, and should be found in the amylopectin fraction by iodine titration.

It can be concluded, too, that amylose does not consist of long straight chains linked together with branched points. It has been demonstrated, at least for Meyer's amylose,[19] that reducing and non-reducing end-group determinations lead to about the same molecular weight, and that physical determinations of the molecular weight are in substantial agreement with end-group determinations This could be true only if the amylose were unbranched. Its behavior with β-amylase tends to confirm this point (*vide infra*). Amylose is completely, or essentially completely, converted to maltose by β-amylase, whereas amylopectin is not. The interpretation offered is that the enzyme starts at the non-reducing end of a chain, and continues to split off maltose units until it meets a branch-point, when it is stopped. If this interpretation is accepted, amylose cannot consist of long straight chains connected by branched points, since then a single branch point would suffice to protect a long-chain molecule from enzyme action.

If, then, it is agreed that starch consists of two components, one unbranched and one branched, the heterogeneity of each component can be examined. The above evidence will be taken to indicate that there is essentially no branching in amylose. Amylose can thus be heterogeneous only with respect to molecular size.

The amyloses from different starch sources are not identical in their behavior with iodine; for example, the iodine activity necessary for complex formation is distinctly different (Fig. 2). In general it has been found that the lower the iodine activity required for complex formation, the longer the starch chain. The results of the iodine titration indicate that the heterogeneity of chain lengths within a given amylose is fairly small, or at least less than that between amyloses from different sources.[18] It is not to be inferred, however, that the amylose from any starch source is truly homogeneous as to chain length.

The differences indicated by the iodine titration for the chain lengths of the amyloses are supported by viscosity measurements.[62] The results have not been previously published and are shown graphically in Fig. 3. No effort has been made

* Recently Dr. T. J. Schoch found that the removal of free fatty acid from cornstarch increased the value to about 28%. Similar increases can be expected for those starches which contain free fatty acid.

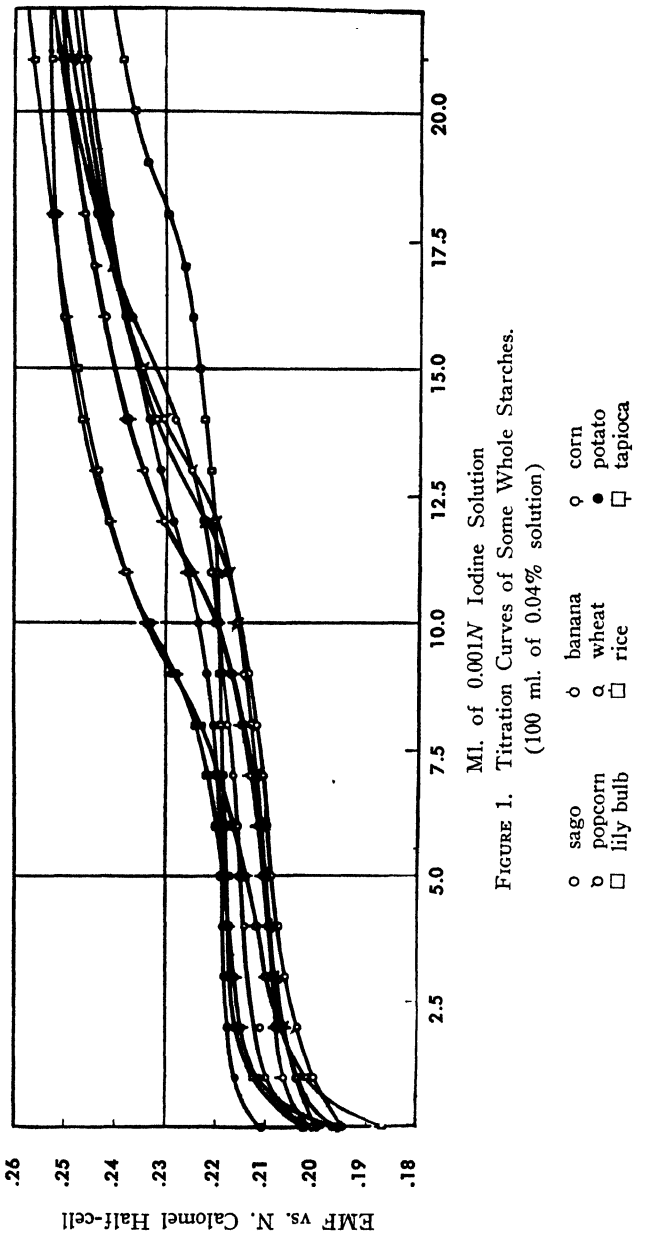


FIGURE 1. Titration Curves of Some Whole Starches. (100 ml. of 0.04% solution)

to calculate the relative molecular weights, but for comparative purposes it may be stated that the particular amylodextrin fraction shown contained about 60 glucose units (calculated from reducing value measurements). The results indicate that molecular weights of the amyloses from various sources will exceed the range of 10,000-50,000 indicated by Meyer.[8] The discrepancy is probably due to the fact that his extraction procedure gave only the shorter molecules from corn amylose, which is the smallest of all the amyloses.

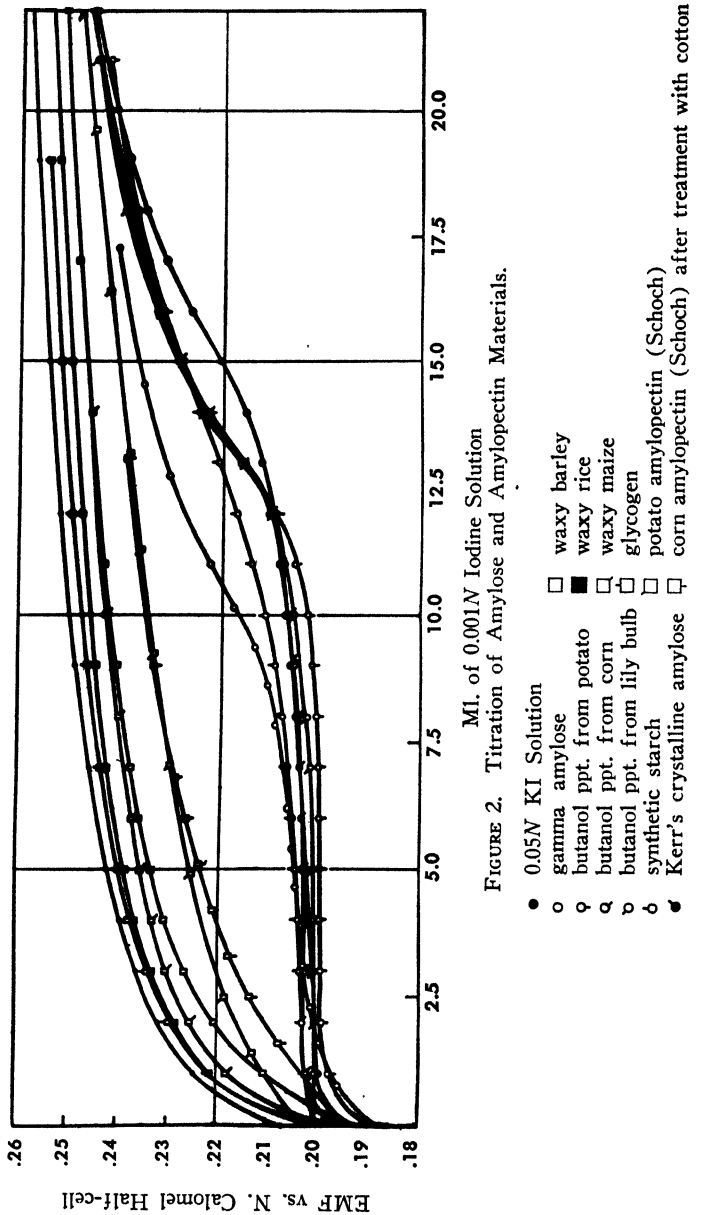


FIGURE 2. Titration of Amylose and Amylopectin Materials.

The amylopectins could be even more heterogeneous than the amylose components, since they could vary in degree of branching as well as in molecular size. From the iodine titration curves, it can be concluded that the branches of the amylopectins are shorter than any of the amylose chains, since a long unbranched portion of an amylopectin molecule should titrate like amylose. Further, it will be noted that the titration curve for glycogen (Fig. 2) falls at a higher potential than for any of the starches. In this case, short branches have been indicated from methylation data,

since there is one end-group for every 12-18 glucose units. Where data are available for comparison, it seems true that the most highly branched amylopectins have the least affinity for iodine. On this basis it would be concluded (Fig. 2) that the branches of waxy starch molecules (no amylose component) are shorter than those of the amylopectins of the common starches.

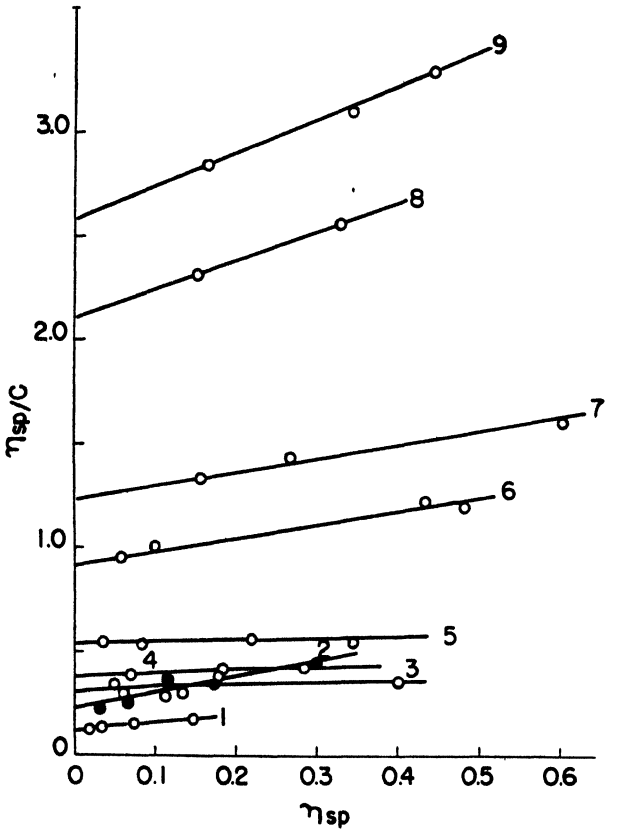


FIGURE 3. Relationship between the Staudinger function, η_{sp}/C, and η_{sp} for some amyloses. The curves are numbered as follows: 1, amylodextrin fraction; 2, synthetic starch; 3, corn amylose by Meyer's method; 4, mixture of 62.5% crystalline amylose and 37.5% amylodextrin fraction; 5, corn "crystalline amylose" by Kerr's method; 6, corn butanol precipitate; 7, lily butanol precipitate; 8, tapioca butanol precipitate; and 9, potato butanol precipitate. The solid circles are for synthetic starch, curve 2.

While the evidence for the amylopectin component is not as complete as that for amylose, all available data indicate that branching is controlled with considerable precision by the enzyme system during synthesis of starch or glycogen. On the basis of degradation with β-amylase, Meyer [20] has come to a similar conclusion and has attempted to work out a pattern of branching in certain amylopectins.

Molecularly Dispersed Starch

Many of the important recent advances in starch chemistry have come about through the use of molecularly dispersed starch. The apologetic way in which many chemists still speak of starch solutions resulted from an earlier and erroneous

interpretation of methylation results. The case for molecularly dispersed starch will therefore be presented here.

Physical measurements[21] gave molecular weights of a far different magnitude than the repeating units of approximately 25-30 glucoses found by end-group assay. The assumption had been drawn that starch in solution consisted of aggregates or micelles of many smaller molecules held together by secondary valence forces, although such a conclusion was carefully avoided by some workers.[10, 11] When the importance of the hydrogen bond to chemistry was demonstrated, it was realized that in carbohydrates the situation was ideal for hydrogen-bond formation, and this type of secondary valence force was assumed to be responsible for the difference between chemical and physical molecular weights. As has been pointed out,[18] branching of starch chains is a far better explanation.

There is good evidence that starch can be molecularly dispersed under careful conditions in water. It no longer seems questionable that this is essentially the case with such solvents as ethylene diamine, hydrazine hydrate, and chloral hydrate. Formic acid has also been suggested as a solvent for starch.[22] In this case esterification takes place, the final analysis approximating one formyl group per glucose unit in the starch.[23]

The viscosity and osmotic pressure of Meyer's[9] amylose in several solvents has been shown to correspond to a molecule of about 200 glucose units. This figure is in agreement with the amount of tetramethyl glucose produced by methylation, and the amount of silver ion required to oxidize the aldehyde group on the other end of the molecule.

It would, however, be absurd to minimize the importance of secondary forces, particularly hydrogen bonding, on the properties of starch.[24] For example, in the well-known process of retrogradation, parts of the starch molecules in solution tend to arrange themselves into crystallites of definite structure, no doubt through the agency of hydrogen bonds. The large aggregates thus formed are insoluble. It is significant that retrograded starch produces the same type of x-ray diffraction pattern as does the original granule.

Starch molecularly dispersed in water is not stable at pH ranges below 7.0, except in very dilute solutions. However, the process of retrogradation is slow, and starch dispersed in alkali or by heat remains molecularly dispersed long enough for study, even in water solution. Such solutions are analogous to supersaturated solutions and, as they are unstable, care must be taken in carrying out and interpreting such reactions as enzyme digestibility, where molecularly dispersed starch solutions are necessary for quantitative results. That the degree of association of starch in solution can be made negligible during the course of such reactions, even in water solution, is clearly indicated by the work of Meyer on the β-amylase digestion of starch and its components (*vide infra*).

Crystalline Nature of Starch Granules and Retrograded Starch

In discussing the crystalline nature of starch, it should be emphasized that the term "crystalline" is used with certain limitations. The fact that an x-ray diffraction pattern is obtained is proof that molecular orientation is present. Even in the starch granule it is presumed that definite orientation of the molecules exists only in certain regions; such crystalline regions are termed "crystallites." This is similar to the situation existing in cellulose. Since these crystalline regions are probably small, it is possible for any single molecule to exist partly inside and partly outside of such crystalline regions; a long amylose chain, or different branches of an amylopectin molecule, might even be included within several such regions. Between crystalline regions, the orientation of the molecules is not complete; such non-oriented molecules would constitute amorphous regions in the granule. The crystallites are doubtless tied together by a network of this amorphous material.

It would appear that the branches of the amylopectin molecules are partly included in the crystallites, while the actual points of branching must remain in the amorphous condition. This follows from the observations that waxy starch granules (no unbranched component) are crystalline and that their x-ray diffraction pattern is like that of retrograded amylose.

Results of X-ray Diffraction. X-ray diffraction has not as yet contributed as much to the knowledge of the structure of starch as it has to the structure of cellulose, due in large measure to the fact that only powder diffraction patterns have been obtained from starch. Katz [25, 26] was able to show, however, that granular starches produce two main types of patterns, the "A" patterns of the cereal starches and the "B" patterns of the tuber starches, with a few showing patterns intermediate in character. He also showed that low temperature caused starches of all sources to retrograde, producing diffraction patterns of the "B" or tuber type, while "A" or cereal type patterns are produced by starch retrograded at temperatures above 50° C.

Bear and French [27] were able to index both "A" and "B" diffraction patterns and to determine the unit cell of the starch structure. Their success was based on two new factors: First, they improved the quality of the diffraction patterns by recognizing starch as a hydrated substance and keeping it thoroughly wet during diffraction. Secondly, they demonstrated through a temperature series of diffraction patterns, that there is a continuous change from the "A" to the "B" modification. All intermediate patterns were obtained by varying the temperature of retrogradation. This led to a helpful correlation between the indices of the two crystalline modifications.

Bear and French report that both "A" and "B" modifications probably have triclinic structures of space group C_1^1, although within their limits of error the "B" modification is orthogonal. The dimensions of the two units are given in Table 2.

Table 2. Dimensions of Starch Unit Cell *

(a_o, b_o, c_o are the unit cell dimensions in Å units; α_o, β_o, γ_o are the angles in degrees between the crystallographic axes.

	a_o	b_o	c_o	α_o	β_o	γ_o	Vol. (Å³)
"A" Modification, ave.	15.4	8.87	6.18	87.0	86.9	92.8	843
"B" Modification, ave.	16.1	9.11	6.34	90.0	90.0	90.0	930

The volume of the "B" unit is larger than that of the "A." This is in accord with the observation that the "B" modification is the more highly hydrated. Four glucose units are calculated to occupy a unit cell for both modifications. This size unit would require that the starch chains be essentially extended. For granular and retrograded starch it would rule out the helical chains suggested by Hanes [2] and Freudenberg [28] (see below).

Results of Optical Studies. Frey-Wyssling [29] has made a study of the optical properties of starch granules, and has interpreted her results in terms of structure. Starch granules are spherocrystals with the long axes of the starch molecules oriented radially with respect to the granule. The granules are quite birefringent; the dark cross with its center at the hilum is a well-known characteristic of all granules. Frey-Wyssling has found the spherocrystals to be optically positive; that is, light with its electric vector radial to the granule is retarded more than light with its electric vector normal to the radius of the granule. It is quite possible that a helical chain would have its greatest polarizability normal to the helix axis, while from Silberstein's theory (see note on p. 677) the polarizability of an extended starch chain could doubtless be along the chain. If these polarizabilities are accepted, the observed sign of birefringence would require extended starch chains,

* Recent film and fiber diffraction diagrams prepared from amylose in "B" modification indicate cell dimensions are $a_o = 16.0$, $b_o = 10.6$, $c_o = 9.2$ Å; 10.6 is the fiber spacing. The gross aspects of the crystalline structure are discussed in an article to appear in J. Am. Chem. Soc. (1943).

in agreement with results from x-ray diffraction. "A" and "B" type granules are identical in these qualitative aspects.

The Helical Configuration of Starch and the Starch-Iodine Complex

Hanes [2] has noted in certain reactions with enzymes, notably with the enzyme produced by *B. macerans* where six-membered cycloamyloses are formed, that glucose residues six apart down the chain appear to be quite close to each other. He proposed, therefore, a helical starch structure with six glucose residues per turn in the helix. He also found that dextrins of more than six glucose residues intensify the color of iodine, while those below this figure do not. This was interpreted to mean that a full turn in the helix is necessary for the formation of the starch-iodine complex.

Freudenberg [28] built models of helical starch chains, and assuming a boat-shaped glucose ring, found that such helices should have the hydroxyls directed outward, leaving a hydrocarbon lining to the center of the helix. On this basis he would explain the blue color of the starch-iodine complex by placing iodine in the center of the helix, in an essentially hydrocarbon medium where iodine is known to give a blue color, rather than the red color which it produces in water or other polar solvents.

Neither Hanes nor Freudenberg has offered further evidence for their proposed configuration of the starch chain and the iodine complex. Recently evidence has been offered which leaves little doubt that the essential points of their proposals are correct when applied to starch in certain modifications, although it is clear that granular and retrograded starch do not possess a helical configuration. Since this confirmatory evidence has not been reviewed elsewhere, a rather complete treatment will be given here.

Katz [31] found that if starch was gelatinized and then precipitated by alcohol rather than by retrogradation, the precipitated material gave a different x-ray diffraction pattern, corresponding to a new crystalline modification of starch. He named it the "V" modification, from "Verkleisterung"—gelatinized starch. This is a misnomer since it is produced by crystallization, the gelatinized starch having no crystalline structure.

In an article by Bear,[32] attention was called to the fact that the "V" pattern bears some resemblance to the x-ray diffraction pattern given by the starch-iodine complex. He suggested that these patterns might result from helical structures such as those proposed by Hanes and by Freudenberg.

Schoch [17] has fractionated starch by means of butanol. The "amylose" fraction precipitated by the butanol has the form of small rosettes, very birefringent and definitely crystalline in nature. These rosettes, when dried, produce a "V" type diffraction pattern of far better quality than any heretofore obtainable.

Kerr and Severson [33] have applied Schoch's butanol precipitation to the components of starch extractable by hot water (straight-chain molecules of shorter chain length). The resulting precipitate, which they named "crystalline amylose," appears indeed to consist of very minute single amylose crystals. These too produce an excellent "V" pattern upon drying, and one closely allied to it when still wet with butanol and water.

Final demonstration that the "V" configuration of starch and the configuration of the starch-iodine complex are related is provided by the treatment of "V" starch with iodine vapor.[34] Any starch, wet or dry, showing a good "V" diffraction pattern will absorb iodine vapor rapidly and in quantity, whereas dry granular or retrograded starch cannot be induced to absorb iodine vapor to any appreciable extent even under the most vigorous treatment. The ability to absorb iodine appears to be related to the quality of the "V" diffraction pattern the starch is capable of producing; Schoch's and Kerr's material are particularly good iodine absorbers. This,

Another point of interest arising from dichroism of flow was that amylose-iodine solutions could be made to show dichroism of flow quite readily while amylopectin-iodine solutions did not produce this phenomenon. Amylopectin has a higher molecular weight than amylose, and failure to orient during flow must be attributed to a more nearly spherical shape. This is in agreement with a straight-chain structure for amylose and a branched structure for amylopectin.

The optical properties of both Schoch's and Kerr's material have been published by Rundle and French.[31] Since the optical properties of Kerr's material are more simple to interpret, and since Schoch's butanol precipitate appears to be a complex multiple crystal of similar properties, only the former will be discussed.

Kerr's crystalline amylose appears to consist of minute single crystals, at least pseudo-uniaxial in character. The flat platelets (Fig. 5) are optically negative; that is, the index of refraction is greatest for light with its electric vector in the plane of the platelet.

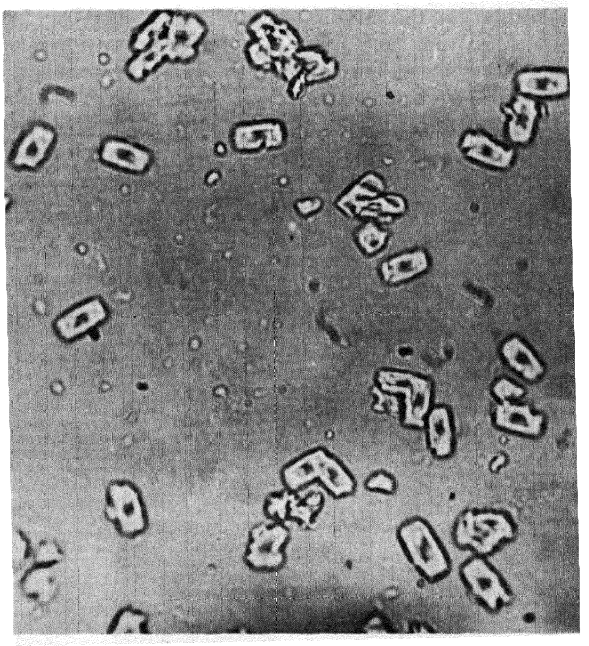


FIGURE 5. Crystalline amylose (Kerr, 1942).

When the small crystals are stained with iodine, they become very dichroic, but apparently remain uniaxial. Light with its electric vector normal to the plane of the platelet is very strongly absorbed, while light with its electric vector in the plane of the platelet is very weakly absorbed. From the dichroism of the iodine molecule discussed above, this would mean that the iodine molecules are arranged with their long axes normal to the plane of the platelet. If an extended-chain model is taken for the starch-iodine complex, the birefringence of the crystals would indicate that the chains lie in the plane of the platelet.* The dichroism would then mean that the long axes of the iodine molecules are normal to the long axis of the starch molecule. This is just the reverse of what was found in the dichroism of flow experiment.

* According to Silberstein's theory [30] the direction of greatest polarizability will be that direction in which neighboring induced dipoles will tend to reinforce each other. Light with its electric vector along the direction of greatest polarizability is retarded most.

On the other hand, the helical model provides a very satisfactory explanation of the optical properties and dichroism of flow. If the greatest polarizability of a helical starch chain is normal to the chain, then the axes of the starch helices are normal to the plane of the "crystalline amylose" platelet, and parallel to the long axes of the iodine molecule, as required by the dichroism of flow of starch-iodine solutions. The helical model also explains the uniaxial, optically negative character of the platelets which is very hard to visualize for any extended-chain model.

We see, therefore, that x-ray diffraction, dichroism of flow, the birefringence of Kerr's crystalline amylose, and dichroism of this amylose after staining with iodine all confirm closest packing of helices for "V" starch, and all are in accord with the model for the starch-iodine complex given in Fig. 4.

Swelling of the Starch Granule

The literature on the mechanism of swelling of starch granules has been reviewed in detail by Alsberg[1] and by Bear and Samsa.[36] A detailed analysis of the structural changes responsible for the swelling is not as yet possible.

The evidence indicates that there is no membrane or separate layer surrounding the starch granule. The unswollen granule seems to be pervious to most water-soluble electrolytes as well as to water. The forces (probably largely hydrogen bonds) within the crystallites are broken in the solution process, and enormous tangential forces are created within the granule which are released by its swelling. The insoluble tangled starch molecules, as they move outward in the swelling process, create the sac or membrane observed under the microscope. Water flows through

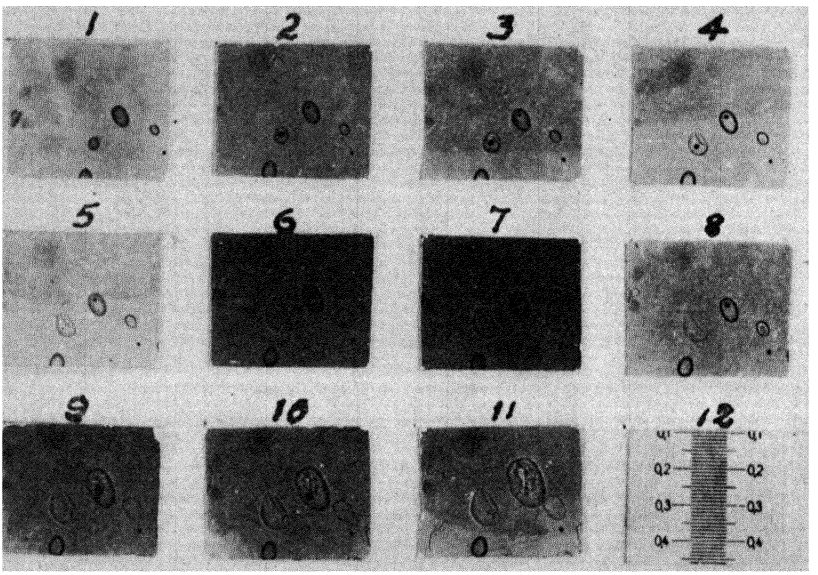


FIGURE 6. Swelling of potato starch granules in 2*M* calcium nitrate. The granule in lower center field shows maximum size of the vapor bubble (black spot) in No. 3; partial collapse of the granule is shown in No. 4. The granule in upper center field continues to swell and vapor bubble reaches maximum size in No. 8 and No. 9. Note shift in position of bubble in granule from No. 8 to No. 10 as the granule weakens and collapses to release the partial vacuum. Swelling of the granule continues after the vapor bubble disappears.

On the other hand, the helical model provides a very satisfactory explanation of the optical properties and dichroism of flow. If the greatest polarizability of a helical starch chain is normal to the chain, then the axes of the starch helices are normal to the plane of the "crystalline amylose" platelet, and parallel to the long axes of the iodine molecule, as required by the dichroism of flow of starch-iodine solutions. The helical model also explains the uniaxial, optically negative character of the platelets which is very hard to visualize for any extended-chain model.

We see, therefore, that x-ray diffraction, dichroism of flow, the birefringence of Kerr's crystalline amylose, and dichroism of this amylose after staining with iodine all confirm closest packing of helices for "V" starch, and all are in accord with the model for the starch-iodine complex given in Fig. 4.

Swelling of the Starch Granule

The literature on the mechanism of swelling of starch granules has been reviewed in detail by Alsberg[1] and by Bear and Samsa.[36] A detailed analysis of the structural changes responsible for the swelling is not as yet possible.

The evidence indicates that there is no membrane or separate layer surrounding the starch granule. The unswollen granule seems to be pervious to most water-soluble electrolytes as well as to water. The forces (probably largely hydrogen bonds) within the crystallites are broken in the solution process, and enormous tangential forces are created within the granule which are released by its swelling. The insoluble tangled starch molecules, as they move outward in the swelling process, create the sac or membrane observed under the microscope. Water flows through

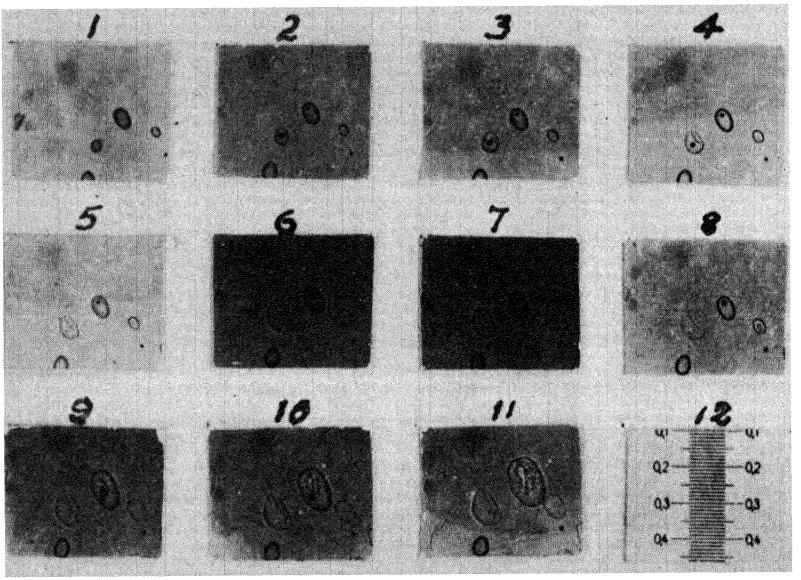


FIGURE 6. Swelling of potato starch granules in $2M$ calcium nitrate. The granule in lower center field shows maximum size of the vapor bubble (black spot) in No. 3; partial collapse of the granule is shown in No. 4. The granule in upper center field continues to swell and vapor bubble reaches maximum size in No. 8 and No. 9. Note shift in position of bubble in granule from No. 8 to No. 10 as the granule weakens and collapses to release the partial vacuum. Swelling of the granule continues after the vapor bubble disappears.

this membrane because af the reduced pressure created within the granule by the increase in volume. Conditions can be selected such that the increase in volume is more rapid than the flow of water through the periphery so that pressures within the granule are low enough for vapor bubbles to form. As the walls of the granule become less rigid because of expansion and hydration, one side of the granule will collapse, and the vapor bubble will disappear completely. A photographic record of granules passing through these stages is shown in Fig. 6.

It seems improbable that these forces are of osmotic origin. As stated by Bear, "It is difficult to believe that the relatively dilute solutions of starch of high molecular weight could very successfully compete with the concentrated calcium nitrate (2 molar) to secure water rapidly enough for the vigorous expansion observed. . . . One is almost forced to conclude that the striking swelling manifestations of starch granules are a result of an active tendency of starch chains to coil or otherwise contract, thus shortening in a radial direction and expanding tangentially as they are placed in the granule. Only in such fashion is it very easy to visualize the remarkable increases in cross-sectional area." This statement is qualified by the authors. The change from a radially extended molecule to a coiled molecule gives a mechanical picture which is possible but not essential to the phenomenon.

Starch Suspensions, Pastes and Gels

Starch products are generally used commercially as suspensions. Such systems contain swollen granules or fragments of granules which are clearly visible under a microscope. It is very apparent that the physical properties of such a system will depend primarily on the volume and physical state of the suspended phase (*i.e.*, the granules) rather than upon the composition of the granules. Samec[4, 37] has published several summaries of the colloidal literature as it relates to starch.

The complexity of the problem can be appreciated by reviewing the factors which will affect the swelling phenomenon. It is apparent that for the different starches under a given set of conditions, swelling will be influenced by the following factors: the original size of the granule; the relative percentages of amylose and amylopectin; the size and homogeneity of the amylose and amylopectin molecules; the degree to which crystallization has taken place in the granule and the type of crystallites ("A" or "B" pattern); and whether the structure of the granule has been previously influenced by chemical or physical treatment. The recent review of the effect of dry-grinding on the physical properties of wheat starch by Lampitt, Fuller and Goldenberg[38] might be cited as an illustration of the complications in the practical field.

A study of the individual and combined effect of these factors remains to be made; only a few of them can be illustrated in a general way at the present time. For example, ordinary and waxy cornstarches gelatinize at the same temperature, but the range for gelatinization of the waxy starch is much narrower. The crystal patterns are almost the same, and an equal temperature potential would be required to rupture the crystal forces. Since waxy starch contains only amylopectin, the crystallite regions (and accordingly the temperature range of gelatinization) would be expected to be more homogeneous than for ordinary cornstarch with 21 per cent of amylose. The tuber starches, in the "B" configuration, all gelatinize at lower temperatures than the cereal starches in the "A" configuration. Data obtained by Sair and Fetzer[39] indicate that the more easily gelatinized "B" starches can be partially changed to the less easily gelatinized "A" starches while still in granule form, and that the temperature range of gelatinization changes accordingly.

In the cereal starches, and particularly in retrograded starches, the crystal forces may be so great that they cannot be broken by boiling water alone. Agents which affect hydrogen bonding, such as alkali, will assist in peptizing such aggregates. Mullen and Pacsu[40] have shown that pyridine-water solutions of suitable concentra-

tion practically obliterate the differences of gelatinization temperature shown by the starches.

The work of K. Meyer [7] shows that at least a portion of the amylose chains are leached out of the granule at temperatures as low as at the beginning of gelatinization. The less soluble amylopectin would remain behind to form the granule residue, as explained above under the discussion of the mechanism of swelling. Upon cooling, the amylose chains and portions of the amylopectin molecules would orient and crystallize. The cooled paste would then become a magma of granule residues interlaced by crystallite regions. The cloudiness of cereal starch pastes is due to such crystallite areas. If these regions were sufficiently numerous the cooled paste would have the properties of a true solid as contrasted to a viscous sol. The measurement of elasticity or rigidity on such systems often gives greater differentiation than viscosity measurements. The contrast between cornstarch and waxy cornstarch is particularly striking.

While the role of amylose is very clear in explaining the differences between the waxy and ordinary cornstarch, there is danger in attempting to explain starch properties on the basis of composition only. For example, the waxy starches can be used for many of the purposes formerly supplied by tapioca,[41] "Minute Tapioca" being a particular example. Yet tapioca starch has about 17 per cent amylose, only slightly less than ordinary corn or potato starch (22 per cent). As stated above, the properties of starch pastes must be considered as due mostly to the volume and state of the dispersed phase. From the foregoing discussion, it is evident that composition, while not the only factor to be considered, is important, as it influences the tendency of the granule to swell, rupture, or recrystallize on cooling.

Industrial control of starch products is based largely on measurement of flow properties, the fluidity funnel being the commonest type of instrument. A recent analysis of the different methods of expressing viscosity as applied to starch has been published from this laboratory.[42] The use of hot rather than cold viscosities is recommended to avoid the complicated changes in physical state introduced by cooling. The accumulated evidence indicates that such measurements can be made with a high degree of precision in spite of their empirical character and that accurate control of the conditions for preparing the paste is more important than the type of instrument used in its measurement.[43, 44] As indicated above, starch pastes of sufficiently high concentration become solid in character upon cooling, whereupon the measurement of elasticity or rigidity is recommended.[45]

Enzymatic Action as Related to Starch Structure

Detailed reviews on enzymatic degradation and synthesis of starch and glycogen have been given by Samec,[46] Hanes [2] and Cori.[47] The older literature is complicated by the fact that mixtures of enzymes were frequently used, as well as by inadequate technics for evaluating the digestion products. Likewise, the use of poorly dispersed starch for the enzyme digestions led to mistaken conclusions. The following statement is limited to three types of enzymes which have been of particular importance to the problems of starch structure.

β-Amylase. This enzyme is presumed to attack at the non-reducing end, or ends, of the starch molecule and cleave off maltose groups in succession until a linkage other than the 1,4-glucosidic bond is encountered.[2, 48] Thus straight chains like those of amylose and amylodextrin are almost completely converted to maltose. Branched molecules, *e.g.*, glycogen or amylopectin, are digested only up to the points of branching. The residual "limit dextrin," according to Meyer, Wertheim, and Bernfeld,[49] possesses as many end groups as there were in the original molecule. After the α-1, 6 linkages have been hydrolyzed by an α-glucosidase, β-amylase action again proceeds.[19] The results of β-amylase digestion were used extensively by Meyer in arriving at the proposed structures for glycogen [50] and for starch molecules.[9] He

emphasizes that the starch fractions must be molecularly dispersed by solution in alkali and an excess of enzyme added immediately after neutralization to provide for complete digestion before retrogradation sets in. Soybeans provide an excellent source of pure β-amylase, since they contain relatively small amounts of other amylases.[51]

Schardinger enzyme. An enzyme from *Bacillus macerans* acts on starch to produce about 55 per cent of the Schardinger dextrins.[52] The two chief reaction products, the α- and β-dextrins, are ordinarily formed in a ratio of about 3 to 1, but if a preserving agent such as trichloroethylene or benzene is added, the product consists almost entirely of the β-dextrin. The molecular weights of the Schardinger dextrins have been accurately determined by x-ray diffraction and crystal density measurements. The β-dextrin (or "cycloheptaamylose") contains seven glucose residues per molecule bonded into a ring by 1,4-glucosidic linkages; the α-dextrin ("cyclohexaamylose") is similar in structure but contains six glucose units.[53] The lower molecular weight values previously reported by Freudenberg and Jacobi [54] were based on cryoscopic measurements.

The enzyme apparently cleaves groups of six or seven glucose units at a time from the end of the molecule and simultaneously forms them into a non-reducing ring. It follows that molecules which do not present a chain of at least 6 or 7 unbranched glucose residues would not be attacked by the Schardinger enzyme. *B. macerans* digests starch to about the same extent as does β-amylase, presumably because both begin at the non-reducing ends of the molecule and are halted by the branched points.

Phosphorylase. Cori [55] has synthesized both starch and glycogen by the action of phosphorylase on glucose-1-phosphate. Synthetic starch produces the same x-ray diffraction pattern as native starch [56, 57] but differs from it in consisting chiefly of the straight-chain rather than the branched type of molecule.* Evidence for the latter statement was obtained by Hanes,[58] who found that synthetic starch is completely converted to maltose by β-amylase, by Hassid and McCready [59] using methylation techniques, and by Bates, French and Rundle [18] using the potentiometric iodine titration described above. The iodine titration curve as well as the specific viscosity in ethylene diamine (Fig. 3) indicate that the chain length of synthetic starch, as usually prepared, is considerably less than that of amylose from native starches. In fact, Haworth,[60] by methylation and end-group assay, arrived at a value of 80-90 glucose units for the chain length. Hassid and McCready [59] failed to detect any tetramethyl glucose in an end group assay on synthetic starch, but a small percentage could have been overlooked due to the small sample analyzed.

Another enzyme often applied to starch is α-amylase. The mechanism of its action has not been clarified sufficiently for it to be of much value in structural studies on starch. Hanes [2] proposed a helical configuration for the starch molecule having six glucose units per turn, partly on the basis of his belief that α-amylase produced dextrins six glucoses in length from starch. Meyer [61] has recently studied the action of α-amylase on amylose and concluded that this enzyme breaks the chain at random; the longer chains are hydrolyzed most rapidly and a difficultly hydrolyzed trisaccharide remains. He believes that branching does not stop α-amylase as it does β-amylase.

SUMMARY

Starch chemistry has undergone revolutionary changes in the last few years, due in large part to the realization that starch consists partly of branched and partly of

* The reader is referred to a current review by K. Meyer ("Advances in Enzymology III," p. 109, New York, Interscience Publishers, 1943) for a discussion of the synthesis of branched chain starches and glycogen. See also Rundle, Daasch, and French, *J. Am. Chem. Soc.*, **66**, 130 (1944).

straight-chain molecules. Methods of separating the branched from the unbranched molecules in a fairly clean-cut manner have been devised and techniques for investigating them are rapidly being developed.

This review attempts to bring together recent pertinent literature on the different phases of starch chemistry, supplementing it with related material from this laboratory. Several advances of major importance, which have not been included in previous reviews of this type, are discussed, *e.g.*, evidence for the helical configuration assumed by amylose molecules in complex-formation with iodine and with butanol and the alcohols; a rapid method for determining the relative proportion of unbranched material in starch; and determination of the correct molecular weights for the Schardinger dextrins.

The advances are too recent for proper evaluation and there is a tendency to place exaggerated emphasis on certain phases of the problem. It can be said, however, that for the first time it is becoming possible to interpret and predict the behavior of starch to a limited extent on the basis of its structural composition. Rapid progress in this field is to be expected in the future.

References

1. Alsberg, *Plant Physiology*, **13**, 295-330 (1938).
2. Hanes, *New Phytologist*, **36**, 101, 189 (1937).
3. Radley, "Starch and Its Derivatives," pp. 7-20, New York, Van Nostrand, 1940.
4. Samec, "Kolloidchemie der Starke," pp. 1-48, Steinkopff, Dresden and Leipzig, 1927.
5. Walton, "A Comprehensive Survey of Starch Chemistry," Part 2, pp. 51-64, New York, Chemical Catalog Co., (Reinhold Publishing Corp.) 1928.
6. Baldwin, *J. Am. Chem. Soc.*, **52**, 2907 (1930).
7. Meyer, Bernfeld and Wolff, *Helv. Chim. Acta*, **23**, 854 (1940).
8. Meyer, "Natural and Synthetic High Polymers," pp. 387-416, New York, Interscience Publishers, 1942.
9. Meyer, in Kraemer: "Advances in Colloid Science," pp. 144-182, New York, Interscience Publishers, 1941.
10. Haworth, Hirst and Oliver, *J. Chem. Soc.*, **1934**, 1917.
11. Haworth and Isherwood, *J. Chem. Soc.*, **1937**, 577.
12. Staudinger and Eilers, *Ber.*, **69**, 819 (1936).
13. Staudinger and Husemann, *Ber.*, **70**, 1451 (1937).
14. Freudenberg and Boppel, *Ber.*, **73B**, 609 (1940).
15. Rundle and Baldwin, *J. Am. Chem. Soc.*, **65**, 554 (1943).
16. Pacsu and Mullen, *J. Am. Chem. Soc.*, **63**, 1168 (1941).
17. Schoch, *Cereal Chem.*, **18**, 127 (1941); *J. Am. Chem. Soc.*, **64**, 2957 (1942).
18. Bates, French and Rundle, *J. Am. Chem. Soc.*, **65**, 142 (1943).
19. Meyer, *Naturwiss.*, **28**, 397 (1940).
20. Meyer and Bernfeld, *Helv. Chim. Acta*, **23**, 875 (1940).
21. Higginbotham and Richardson, *J. Soc. Chem. Ind.*, **57**, 234 (1938).
22. Alexander, *J. Soc. Chem. Ind.*, **55**, 206 (1936).
23. Gottlieb, Caldwell and Hixon, *J. Am. Chem. Soc.*, **62**, 3342 (1940).
24. Bawn, Hirst and Young, *Trans. Faraday Soc.*, **36**, 880 (1940).
25. Katz and van Itallie, *Z. physik. Chem.*, **A150**, 90 (1930).
26. Katz and Derksen, *Z. physik. Chem.*, **A165**, 228 (1933).
27. Bear and French, *J. Am. Chem. Soc.*, **63**, 2298 (1941).
28. Freudenberg, Schaaf, Dumpert and Ploetz, *Naturwiss.*, **27**, 850 (1939).
29. Frey-Wyssling, *Naturwiss.*, **28**, 78 (1940).
30. Silberstein, *Phil. Mag.*, **33**, 92, 521 (1917).
31. Katz and Derksen, *Z. physik. Chem.*, **A167**, 129 (1933).
32. Bear, *J. Am. Chem. Soc.*, **64**, 1388 (1942).
33. Kerr and Severson, Paper presented at a joint program of the Divisions of Agr. and Food Chemistry and Sugar Chemistry and Technology at the Memphis meeting of the American Chemical Society, April 23, 1942.
34. Rundle and French, *J. Am. Soc.*, **65**, 558 (1943); Baldwin, R. R., Bear, R. S., and Rundle R. E., *J. Am. Chem. Soc.*, **66**, 111 (1944).
35. F. H. Spedding, *private communication.*
36. Bear and Samsa, *Ind. Eng. Chem.*, **35**, 721 (1943).
37. Samec, *Kolloid-Beihefte*, **51**, 378-444 (1940); also Vol. IV of this series.
38. Lampitt, Fuller and Goldenberg, *J. Soc. Chem. Ind.*, **60**, 47 (1941).
39. Sair and Fetzer, *Ind. Eng. Chem.*, in press.
40. Mullen and Pacsu, *Ind. Eng. Chem.*, **34**, 807 (1942).
41. Hixon and Sprague, *Ind. Eng. Chem.*, **34**, 959 (1942).
42. Brimhall and Hixon, *Cereal Chem.*, **19**, 425 (1942).
43. Gallay and Bell, *Can. J. Research*, **14B**, 360 (1936).
44. Farrow, Lowe and Neale, *J. Text. Inst.*, **19**, T18 (1928).
45. Brimhall and Hixon, *Ind. Eng. Chem., Anal. Ed.*, **11**, 358 (1939).
46. Samec and Blinc, *Kolloid-Beihefte*, **49**, 75 (1939).
47. Cori, *Quant. Biol.*, **7**, 260 (1939).
48. Myrback, *Biochem. Z.*, **297**, 179 (1938).
49. Meyer, Wertheim and Bernfeld, *Helv. Chim. Acta*, **24**, 212 (1941).
50. Meyer and Fuld, *Helv. Chim. Acta*, **24**, 375 (1941).
51. Newton and Naylor, *Cereal Chem.*, **16**, 71 (1939).
52. McClenahan, Tilden and Hudson, *J. Am. Chem. Soc.*, **64**, 2139 (1942).
53. French and Rundle, *J. Am. Chem. Soc.*, **64**, 1651 (1942).
54. Freudenberg and Jacobi, *Ann.*, **518**, 102 (1935).

55. Cori, Cori, and Schmidt, *J. Biol. Chem.*, **129**, 629 (1939).
56. Astbury, Bell and Hanes, *Nature*, **146**, 558 (1940).
57. Bear and Cori, *J. Biol. Chem.*, **140**, 111 (1941).
58. Hanes, *Proc. Roy. Soc., London*, **128B**, 421 (1940); **129**, 174 (1941).
59. Hassid and McCready, *J. Am. Chem. Soc.*, **63**, 2171 (1941).
60. Haworth, Heath and Peat, *J. Chem. Soc.*, **1942**, 55.
61. Meyer and Bernfeld, *Helv. Chim. Acta*, **24**, 359E (1941).
62. Foster and Hixon, *J. Am. Chem. Soc.*, **65**, 618 (1943).

Enzymes and the Biological Action of Vitamins

A. E. Axelrod and C. A. Elvehjem

Department of Biochemistry, University of Wisconsin, Madison, Wisconsin

Introduction

The concept of vitamins as "bio-catalysts" or as component parts of "bio-catalysts" is now generally accepted. Such a viewpoint is a necessary corollary of the fact that only trace amounts of vitamins are required in the diet in order to maintain the normal functioning of the organism. It is quite apparent, therefore, that vitamins by their very nature must serve as "catalysts" for certain vital processes. We propose in this review to summarize the evidence which substantiates the thesis that "enzyme catalysis" is the most rational explanation of the biological action of vitamins.

Before entering into a discussion of such evidence, it may be well to review briefly some of the earlier concepts of the function of vitamins. The first recognition of their biological action undoubtedly was the discovery that certain foodstuffs exerted profoundly beneficial effects upon various pathological states. Thus, the observation that an abundance of liver was a remedy for night blindness, or that green vegetables or fresh-fruit juices gave relief in scurvy were early examples of such biological activity. However, very little thought was given by the first observers to the mechanism by which these responses were effected.

Nevertheless, the pioneer workers in the field of vitamins did propose a number of theories to account for the symptoms which they observed in their deficient animals. It was very natural for anyone working in that period, *i.e.*, the close of the last century, to ascribe these symptoms to the presence of pathogenic microörganisms, since the work of Pasteur and others seemed to indicate that every disease must have a positive, causative agent. However, Eijkman's [59] experimental evidence led him to abandon the theories that beriberi might be due to the presence of pathogenic organisms in rice, to lack of mechanical stimulation of the intestine, or to insufficiency of total food. He believed that the excessive metabolism of starch resulted in toxic agents which produced the symptoms of beriberi, and he suggested that the silver skin of the rice contained a substance which counteracted these toxic products. Later Eijkman [60] was led to agree with Grijn's conception of beriberi, namely, that there is present in rice polishings a substance indispensable to health.

The early concept that the symptoms of a deficiency disease were due to toxic agents in certain constituents of the diet was also proposed by Hart, *et al.*[108] to explain their results with calves fed rations which were balanced from restricted sources. Later, these authors were able to show that the symptoms observed by them in their early experiments were due to a lack of vitamin A in the ration. Another example of such toxicity theories is given by the comparatively recent studies on the so-called "egg-white-toxicity" syndrome observed when animals are fed raw egg-white. It was generally held by workers in this field that the egg-white contained a positive toxic factor, but Parsons, Lease and Kelly [184] early concluded that the injury involved an interrelation between a positive toxicity and a relative absence of a protective factor.

It is now recognized through the work of György *et al.*[101] that the egg-white combines with biotin in the intestinal tract, and that the syndrome of egg-white toxicity is actually due to biotin deficiency.

In summary, it is now generally agreed that pathological conditions due to faulty nutrition are intrinsically caused by the *lack* of one or more necessary food constituents.

The study of tissues from vitamin-deficient animals constituted another approach toward the elucidation of the mechanism of vitamin action. A large body of literature has accumulated on this subject, and much is now known concerning the effect of various vitamin deficiencies upon the intimate structures of certain tissues. However valuable such studies may be, it is an apparent fact that such histological observations are to a great extent descriptive in nature. The fundamental changes underlying these structural modifications will be elucidated only when our knowledge of the metabolic disturbances resulting from the vitamin deficiency is much more complete.

A possible functional relationship between vitamins and hormones has been offered as an attractive hypothesis for the explanation of vitamin action. Unfortunately our information on this subject is far too meager to allow any clear-cut establishment of such relationships. However, it is most logical to assume that such relationships will be demonstrated when more is known about the specific action of both hormones and vitamins. There are many close analogies between vitamins and hormones; and if, with Green,[99] we are willing to accept the consideration that hormones are potential enzymes or prosthetic groups of enzymes, the relationship between vitamins and hormones becomes even more plausible.*

As has been indicated, the problem of the biological action of vitamins has been approached from a variety of viewpoints. It is our belief, however, that the most fundamental approach has resulted from the recent identification of some of the vitamins as components of the prosthetic groups of certain enzymes. The role of the vitamins in many of the metabolic processes of the body has thus been brought to the fore. There is ample evidence to favor the view that the biological action of vitamins is intimately related to this ability to serve as precursors for the prosthetic groups of enzymes. The acceptance of such a relationship has constituted one of the most significant steps toward our ultimate understanding of the biological action of vitamins.

Thiamine

The existence of a relationship between thiamine and carbohydrate metabolism was first recognized by Funk,[84] who suggested that the severity of thiamine deficiency symptoms could be correlated with the amount of carbohydrate in the diet. Evans and Lepkovsky [74] were the first to show definitely that rats can be maintained with greatly reduced thiamine intakes when the ration contains high levels of fat. This subject has been reviewed by Salmon and Goodman.[198] Using a more improved ration, Stirn *et al.*[215] have shown that the thiamine deficiency syndrome in the rat may be alleviated by the isocaloric replacement of the carbohydrate component of the thiamine-free ration by natural fats or synthetic fatty acid esters. On the other hand, Westenbrink [250] was unable to demonstrate a reduced thiamine requirement in pigeons fed high-fat diets. Arnold and Elvehjem [10] have shown that the nutritional requirement for thiamine in the case of rats, chicks and dogs may be accurately expressed in terms of the carbohydrate content of the ration. The "sparing action" of fat on the thiamine

* "Hormones and vitamins are considered together because they both represent substances whose presence in minute amounts is essential to the normality of development and function in plants and animals . . . their peculiar actions may, in some cases at least, prove to be due to their modification of biocatalysts or surfaces or to the production of changes in the permeability of tissues or membranes. Such mechanisms would permit minute amounts of materials to exert a profound influence on metabolism and function." J. Alexander, "Colloid Chemistry," 4th edition (D. Van Nostrand Co., 1937), p. 360 of Chapter XX.

requirement and the correlation between thiamine requirement and carbohydrate intake suggest that, although thiamine is not required for the catabolism of fat, it is intimately bound up with some fundamental stage in the degradation of carbohydrate. These nutritional studies have been of value in suggesting that the biological action of thiamine is related to its role in carbohydrate metabolism. However, the elucidation of the more precise function of thiamine has been derived from *in vitro* studies with surviving tissues.

From such studies definite indications that thiamine is related to carbohydrate metabolism came from the observations of Gavrilescu and Peters,[87] who, working with pigeons, showed that the respiration of brain tissue in the presence of glucose was significantly lowered in cases of thiamine deficiency. This observation gained in significance when it was demonstrated later that the *in vitro* addition of a few micrograms of thiamine markedly stimulated the oxygen uptake of polyneuritic pigeon brain in the presence of lactic acid. These workers concluded that thiamine functioned in the lactic acid oxidase system of pigeon brain. This viewpoint was abandoned, however, following the observations of Birch and Mann [39] that the coenzyme for lactic acid oxidation did not contain thiamine, and those of Meikeljohn and co-workers [171] that an increased oxygen uptake caused by the addition of thiamine to polyneuritic brain tissue was not accompanied by an increased removal of lactic acid. When pyruvate was employed as a substrate, the addition of thiamine stimulated both oxygen uptake and pyruvate removal.[187] The experimental work of Elvehjem and co-workers [201, 202, 150] verified and extended these results to other tissues, such as liver, kidney, and heart. A decrease in the respiration of liver from thiamine-deficient animals was observed even when the complicating factor of inanition was removed by forced feeding. It was proposed that thiamine was essential in all the tissues studied and that polyneuritis is a disease involving the derangement of the carbohydrate metabolism of the entire organism, rather than of the nervous tissue specifically.

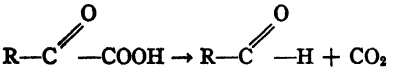
There is thus ample evidence that thiamine is related to pyruvate metabolism *in vitro*. It was very desirable to have the *in vivo* confirmation of such results. This was not long in coming. The lactic acid content of polyneuritic tissues is higher than that in normal tissues, especially after exercise.[121, 79] Light muscular work by thiamine-deficient humans [158] is followed by an increase in blood pyruvate. This rise in pyruvate may be maintained or still further increased at the end of half an hour's rest. After the intravenous injection of pure thiamine, "resting levels" are found following a half hour's rest after light exercise. In thiamine-deficient subjects there is, therefore, a delayed recovery from the high blood level pyruvate values which result from exercise. Results similar to the above were obtained with lactic acid by Inawashiro and Hayasaka [115] for patients with beriberi and dogs with thiamine deficiency. Fisher [79] also found a delayed removal of lactic acid from the heart and muscles of deficient pigeons after exercise. Sherman and Elvehjem [203] showed that polyneuritic chicks were unable to remove injected pyruvate from the blood as rapidly as did normal birds. Increased amounts of pyruvate were found in body fluids of human cases of fulminating, and in a number of cases of subacute beriberi.[189] A marked effect of thiamine upon the reduction of the pyruvate level was observed. Increased blood pyruvate values in thiamine-deficient rats and pigeons [232] and in rats, rabbits and humans [157] have been noted. A rapid increase in pyruvic acid content of urine occurs in rats on a low-thiamine diet.[204] The administration of thiamine brings this level back to normal within 24 hours. The isocaloric substitution of fat for sucrose in the diets fed to thiamine-deficient rats resulted in only a partial return to normal of urinary pyruvate values. Banerji and Harris [24] observed that thiamine-deficient rats excreted large amounts of bisulfite-binding substances in the urine; the quantity of these was proportional to the degree of the deficiency. Harper and Deuel [105] observed that the excretion of pyruvate in the urine increases as the deple-

tion of thiamine reserves proceeds. Urinary pyruvate is reduced by optimal thiamine administration. At variance with the data of Shils *et al.*,[204] Banerji[28] found that the ingestion of high-fat diets by rats deprived of thiamine resulted in a diminished excretion of bisulfite-binding substances in the urine.

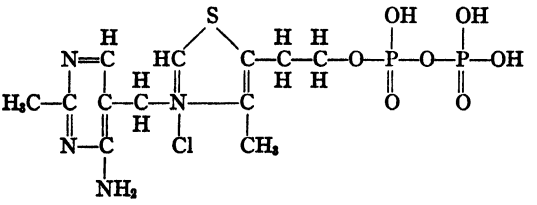
The above evidence is conclusive in its demonstration that thiamine is concerned with carbohydrate metabolism, and more specifically with the metabolism of pyruvic acid, an intermediate product in the breakdown of carbohydrate.

Such was the status of our knowledge of the biological action of thiamine when in 1937 Lohmann and Schuster[155] demonstrated in a now classic paper that cocarboxylase is the pyrophosphoric acid ester of thiamine. Our present understanding of the biological activity of thiamine stems largely from this observation.

In 1911 Neuberg and Hildesheimer[177] reported the presence in yeast of an enzyme called carboxylase which catalyzed the decarboxylation of α-keto acids according to the equation:



where R is a simple aliphatic radical. In 1932, Auhagen[11] demonstrated that the enzyme carboxylase consisted of two components, a protein and a thermostable part called cocarboxylase. Both factors were required for the complete activity of the enzyme. Previous to this observation of Auhagen's, Kinnersley and Peters[120] in 1928 claimed that there was a form of thiamine precipitable by lead acetate. The structure of cocarboxylase was given by Lohmann and Schuster as:



It is interesting to note that Simola[208] in 1932 observed that the cocarboxylase content of tissues from rats suffering from a complicated vitamin B deficiency was significantly lowered.

Very recently Green and co-workers[95] isolated carboxylase from top brewers' yeast in highly purified and stable form. The enzyme was found to be a diphosphothiamine-magnesium-protein. Their best preparations contained 0.46 per cent diphosphothiamine and 0.13 per cent magnesium. Green *et al.*[98] have obtained a crude pyruvic acid and α-ketoglutaric acid carboxylase from various animal tissues. This enzyme was found to consist of (1) protein, (2) a divalent cation (manganese or magnesium), and (3) diphosphothiamine. The authors suggest that this enzyme system is similar to, but apparently not identical with, that prepared from bacteria by Silverman and Werkman.[207]

The recognition of the relationship between thiamine and cocarboxylase has been the impetus for a considerable amount of work which has been conducted mainly along the following lines: (1) the relative distribution of thiamine and cocarboxylase in various cellular types; (2) the comparative biological activity of thiamine and cocarboxylase; (3) the mechanism of the enzymatic phosphorylation of thiamine to cocarboxylase; (4) the mechanism of the action of cocarboxylase; (5) the cellular reactions in which cocarboxylase is involved.

(1) **The relative distribution of thiamine and cocarboxylase in various cellular types.** The available evidence favors the view that cocarboxylase is the significant form of thiamine in animal tissues and yeast. Ochoa and Peters [180] have shown that in some tissues, such as brain and liver, cocarboxylase appears to be predominant, while in skeletal and heart muscle there is relatively more thiamine present and in amounts which sometimes approach that of cocarboxylase. Ochoa [178a] believes that the vitamin in pigeon breast muscle not present as cocarboxylase is very likely the monophosphoric acid ester. Westenbrink *et al.*[254] found only traces of free thiamine in muscle. This apparent discrepancy may be due to the presence in muscle of thiamine orthophosphate which is determined by the enzymatic method (Ochoa and Peters) along with free thiamine, and by the thiochrome method (Westenbrink *et al.*) together with cocarboxylase. Westenbrink and Goudsmit,[251] using a modification of the thiochrome method for the determination of both thiamine and cocarboxylase, have reached similar conclusions. In confirmation of the early work of Simola,[208] both groups of workers [178a, 254] have observed that the cocarboxylase content of tissues is much reduced in thiamine deficiency. The administration of thiamine led to a marked synthesis of cocarboxylase in the liver and kidney of avitaminous animals. Approximately 90 per cent of the total thiamine in blood is in the phosphorylated form, and Goodhart and Sinclair [90] have demonstrated that all the blood cocarboxylase is found within the blood cells. Evidence was presented for a freely diffusible form of thiamine in the blood. In addition, the authors believe that there probably exists a form of thiamine in combination with protein that is not cocarboxylase and is present in serum. Hennessy and Cerecedo [110] confirm the observation that in animal tissues there is predominantly cocarboxylase. Houston and Kon [118] have demonstrated that thiamine is present in combination with protein in milk. In normal cerebrospinal fluid thiamine is in the free, unphosphorylated form.[209] In plant tissues, thiamine appears mainly in the free state.[224] However, Horowitz and Heegaard [112] have extracted a carboxylase from pea roots which contains cocarboxylase.

(2) **The comparative biological activity of thiamine and cocarboxylase.** The predominance of the phosphorylated form of thiamine in animal tissues argued strongly for the viewpoint that cocarboxylase is the active form. Lohmann and Schuster [155] proposed that such is actually the case. However, contrary to the original observation of Lohmann and Schuster, Peters [185] observed that the catatorulin effect (increased oxygen uptake upon addition of thiamine) of cocarboxylase with avitaminous pigeon's brain tissue is much less than that of thiamine itself. This observation was most surprising in view of the fact that thiamine was phosphorylated rapidly under the conditions of the catatorulin test. The explanation for the anomalous results observed by Peters in the catatorulin test has recently been given by Banga *et al.*[25] Employing finely ground suspensions of brain tissue, these workers have been able to obtain preparations which show no catatorulin effect with free thiamine, whereas a good effect is observed with cocarboxylase. It is now evident that the previous failure of cocarboxylase to give good catatorulin effects with brain brei was due to the lack of penetration of the phosphorylated vitamin to the active enzyme centers, as had been previously suggested by Peters. Brain preparations which react to free thiamine were shown to synthesize cocarboxylase in amounts which accounted for the catatorulin effect observed. Thus, the hypothesis of Lohmann and Schuster [155] that the active form of thiamine in animal tissues is the pyrophosphate is now considered to be proved.

In bacteria the evidence is clear that cocarboxylase rather than free thiamine is effective in the metabolism of pyruvic acid. Working with acetone-dried preparations of *B. delbrückii* washed free from cocarboxylase, Lipmann [144] observed that the oxidation of pyruvic acid was catalyzed by cocarboxylase and not by free thiamine.

Barron and Lyman [84] confirmed Lipmann's studies by experiments with a variety of bacterial cells and animal tissues. They concluded that cocarboxylase acts as a catalyst for both the oxidation and dismutation of pyruvic acid.

In yeast, Lohmann and Schuster [155] showed that cocarboxylase and not free thiamine is the active constituent of the carboxylase system. This observation has since been confirmed by a number of workers. [218, 225, 151, 152, 95] Ochoa and Peters [180] have reported that thiamine as well as the monophosphoric acid ester stimulates the decarboxylation of pyruvic acid by alkaline-washed yeast in the presence of cocarboxylase. The action is due to the pyrimidine half of the molecule, containing the amino group. Lipton and Elvehjem [158] have questioned the physiological significance of this activation by free thiamine. Peters [186] has stated that this "activation" effect is an artifact. The explanations offered by various workers for this stimulatory effect of thiamine upon cocarboxylase activity will be discussed in a later section of this paper.

(3) **The mechanism of the enzymatic phosphorylation of thiamine to cocarboxylase.** The enzymatic phosphorylation of thiamine to cocarboxylase has been demonstrated in a large variety of bacterial, yeast and animal cells. [223, 73, 155, 186, 206, 84, 151, 152, 91, 178, 247] Lipschitz *et al.* [151] demonstrated that an aetiozymase preparation (alkaline-washed dried brewers' yeast) was able to convert thiamine into cocarboxylase in the presence of hexosediphosphate and boiled tissue extracts. A search for the thermostable constituents present in the boiled tissue extract which were necessary for this phosphorylation revealed that the synthesis of cocarboxylase could be readily achieved if cozymase and acetaldehyde were substituted for the boiled tissue extract. This synthesis was completely blocked by .003*M* sodium iodoacetate, and it was inhibited to the extent of about 50 per cent by the addition of .04*M* sodium fluoride. The authors were led to the hypothesis that the synthesis occurs as a result of the dismutation between triose phosphate and acetaldehyde, coupled with the esterification of inorganic phosphate. [153] It was also noted that the phosphorylation of thiamine could be accomplished through the dephosphorylation of phosphopyruvic acid resulting from added phosphoglyceric acid. Cozymase was not required for the synthesis under the latter conditions. Thus, two mechanisms for the phosphorylation of thiamine were indicated. Evidence was also presented to show that thiamine is phosphorylated by means of the phosphate-transferring reactions of the Embden-Meyerhof scheme with the intervention of the adenylic acid system. Ochoa and Peters [180] have criticized the results of Lipton and Elvehjem on the basis of their observation that free thiamine exerts a powerful stimulatory effect upon the activity of small amounts of cocarboxylase. This divergence in opinion has been satisfactorily explained by Lipton and Elvehjem. [154] It was demonstrated that the activation of cocarboxylase by thiamine is dependent upon the type of yeast employed. Brewers' yeast shows little activation, but bakers' yeast exhibits a strong activating effect. These observations were in accord with those of Lipmann [147] and Westenbrink and van Dorp. [253] Moreover, Weil-Malherbe [247] succeeded in preparing a soluble yeast protein, containing the apocarboxylase, which was activated by cocarboxylase and magnesium. This system was not stimulated by thiamine. Westenbrink *et al.* [252] have detected the presence of a powerful phosphatase in the yeast, capable of hydrolyzing the phosphoric acid esters of thiamine, and explained the stimulation of the carboxylase system by thiamine as being due to the inhibition of the action of this phosphatase by free thiamine. Lipton and Elvehjem [154] offer an alternate explanation for the thiamine activation observed in their samples of bakers' yeast. They suggest that there is a heat-labile material in bakers' yeast which can adsorb cocarboxylase without the production of an active enzyme. The addition of excess thiamine saturates this material and thus permits the adsorption of the cocarboxylase upon the active apoenzyme. In confirmation of Lipton and Elvehjem, [153] Weil-Malherbe [247] was also able to show that the phosphorylation of thiamine was performed by phosphate trans-

fer from adenosine triphosphate to thiamine. This phosphate transfer is apparently effected only at the surface of the carboxylase protein, since the synthesis stops as soon as the carboxylase protein becomes fully saturated. Phosphopyruvic acid could act as a phosphate donator for the synthesis of cocarboxylase in the presence of catalytic amounts of adenylic acid or adenylpyrophosphate. Thiamine monophosphate was found to be slightly inferior to thiamine as a precursor of cocarboxylase. This was interpreted as meaning that the monophosphate is not an intermediary of the synthesis, but must first be hydrolyzed to free thiamine.

The data of Lipton and Elvehjem, Weil-Malherbe, and Lohmann and Schuster all indicate that the synthesis of cocarboxylase from thiamine requires the intermedia tion of adenosine triphosphate. It is then to be expected that the phosphorylation of thiamine by inorganic phosphate is intimately connected with simultaneous dismutations or oxidations involving the catalysis by coenzymes. This viewpoint is in accord with the data of Ochoa [178] who, in a study of the enzymic synthesis of cocarboxylase in animal tissues, concluded that the phosphorylation of thiamine is enhanced when energy from intact respiration is available. Ochoa's failure to obtain increased synthesis with liver preparations on addition of adenosine triphosphate, phosphoglyceric acid, or both together, argues against a direct transfer of phosphate from either compound to thiamine.

(4) **The mechanism of the action of cocarboxylase.** The identification of cocarboxylase as thiamine pyrophosphate suggested two explanations for the mechanism of the catalytic activity of cocarboxylase: (1) a "Langenbeck cycle" involving the amino group in the pyrimidine ring, or (2) a reversible oxidation and reduction of the double bond adjoining the quaternary nitrogen in the thiazole nucleus.

Langenbeck [141] has shown that primary amines are capable of decarboxylating α-ketocarboxylic acids. The reaction is depicted in the following manner:

$$X{-}NH_2 + R{-}CO{-}COOH \rightarrow R{-}\underset{\displaystyle NX}{\overset{\displaystyle \|}{C}}{-}COOH + H_2O$$

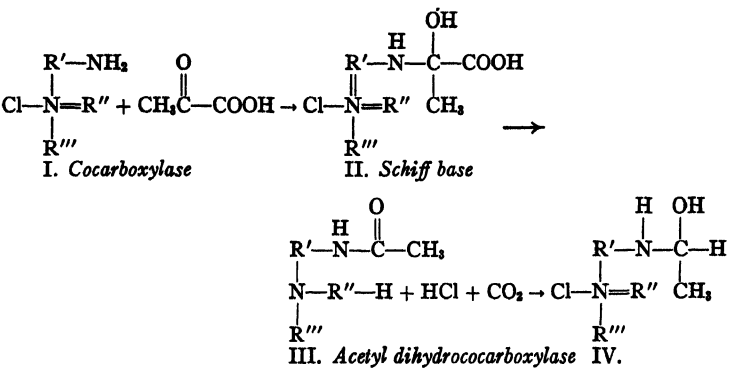
$$R{-}C{-}COOH \rightarrow RCH{=}NX + CO_2$$

$$RCH{=}NX + H_2O \rightarrow R.CHO + X{-}NH_2$$

From these observations, Langenbeck predicted that a primary amino group would be found at the active center of the carboxylase molecule. The elucidation of the structure of cocarboxylase appeared to confirm this hypothesis. Further support for Langenbeck's hypothesis was derived from the work of Bergel and Todd [87] and of Cline,[46] showing that hydroxythiamine has no biological activity. Cline [47] observed that replacement by alkyl groups of one or both hydrogen atoms of the amino group in the pyrimidine moiety results in a very marked reduction in the vitamin potency.

Stern and Melnick [214] examined the Langenbeck hypothesis in great detail and their evidence suggested the view that pyruvic acid is not decarboxylated via the "Langenbeck cycle." Their evidence, in brief, consisted of the observations that thiamine is not a typical primary amine and that, in the test system of Langenbeck, thiamine was found to be completely inactive as a carboxylase model. Weil-Malherbe [248] has suggested in support of the Langenbeck cycle that the properties of the amino group may be profoundly altered in the combination of cocarboxylase with the protein. Barron [81] prefers to believe that the combination of cocarboxylase with pyruvate may take place at the amino group and that the pyruvate thus activated may react with catalysts for its oxidation, reduction, dismutation or condensation. Weil-Malherbe [249] has proposed the following hypothesis for the catalytic activity of cocarboxylase. His theory is founded on the fact that cocarboxylase possesses a free amino group and, in addition, contains a pentavalent nitrogen atom which may be capable of alternate oxidation and reduction. Briefly, the initial steps in the decar-

boxylation of α-ketocarboxylic acids involve intramolecular oxidation-reduction of the Schiff base primarily formed:



I. *Cocarboxylase* II. *Schiff base*

III. *Acetyl dihydrococarboxylase* IV.

In animal tissues, III would be oxidized to acetyl diphosphothiamine, whereas in yeast a second intramolecular oxidation-reduction would occur, resulting in the Schiff base of acetaldehyde (IV). Evidence that cocarboxylase does not act as a sluggish reversible oxidation-reduction system argues against this theoretical proposal of Weil-Malherbe.

Lipmann and Perlmann [149] have shown that thiamine may be reduced by $Na_2S_2O_4$ or $Pt—H_2$. Cocarboxylase is also reduced by $Na_2S_2O_4$.[35] As pointed out by Barron,[28] it is quite possible that during the reduction process with $Na_2S_2O_4$ the thiamine molecule is split into pyrimidine and thiazole. The reduction of thiamine and cocarboxylase was confirmed by Stern and Melnick.[214] It was suggested by Lipmann [145] that thiamine acts as a sluggish reversible oxidation-reduction system. Lipmann [143] postulated a two-step reduction, and contended that an intermediary semiquinoid radical was formed in the reaction. There is, however, no evidence at the present which proves that cocarboxylase undergoes a cycle of oxidation and reduction during its catalytic activity. In fact, Barron and Lyman [35a] have shown that thiamine becomes, with phosporlyation, more resistant to the action of reducing and oxidizing agents.

It must be concluded that the mechanism of the catalytic activity of cocarboxylase remains unknown.

(5) **Reactions catalyzed by cocarboxylase.** It is apparent that a complete understanding of the reactions catalyzed by cocarboxylase is dependent to a great extent upon a detailed knowledge of the metabolic pathways of pyruvic acid. Pyruvic acid is an extraordinarily reactive compound. It is the key compound in most forms of carbohydrate breakdown, in transamination, in the formation of ketone bodies, and in many other biological reactions.

In yeast, cocarboxylase functions simply in the decarboxylation of α-ketocarboxylic acids to carbon dioxide and the corresponding aldehyde. In animal and in many bacterial species this reaction apparently does not occur. The reactions catalyzed by cocarboxylase in these cells are considerably more complex.

Lipmann [145] has studied in great detail the oxidative decarboxylation of pyruvic acid to acetic acid by *B. delbrückii*. He has found that at least five components are required for the catalysis of pyruvic acid dehydrogenation with oxygen as the hydrogen acceptor: (1) cocarboxylase; (2) flavin-adenine dinucleotide; (3) Mg^{++}, Mn^{++}, or Co^{++}; (4) protein; and (5) inorganic phosphate. The anaerobic dismutation of pyruvic acid in these bacteria involves riboflavin as the coupling link. The dismutation is a "fermentation-like" reaction with pyruvic acid-pyruvic dehydrogenase as the reductant, pyruvic acid-lactic dehydrogenase as the oxidant, riboflavin acting as the hydrogen carrier between the two systems. Cocarboxylase, therefore, also func-

tions in this dismutation reaction. Barron and Lyman [84] have demonstrated that cocarboxylase acts as a catalyst for both the dismutation and oxidation of pyruvic acid by a number of bacterial and animal tissues. The anaerobic dismutation of pyruvic acid carried out by certain bacteria depends on the presence of thiamine.[111] The oxidation of acetate by propionic acid bacteria is also catalyzed by thiamine.[194] Smyth,[210] employing thiamine-deficient staphylococci, has given evidence in support of the proposal of Krebs and Eggleston [132] that thiamine catalyzes the synthesis of oxalacetate from pyruvate and CO_2. Krampitz and Werkman [131] have shown that cocarboxylase is not a coenzyme for the enzyme which brings about the decarboxylation of oxalacetate to pyruvate and CO_2 in *Micrococcus lysodeikticus*. These results appear to be at variance with those reported by Smyth.

Barron [81] has reported that cocarboxylase accelerates the following reactions in animal tissues:

(1) Carbohydrate synthesis from pyruvate where the initial reaction seems to be: pyruvate + fumarate + H_3PO_4 → phosphopyruvate + $4CO_2$ + $2H_2O$.
(2) Citrate synthesis from pyruvate and oxalacetate.*
(3) Synthesis of succinate.
(4) Synthesis of acetoacetate.
(5) Oxidation of α-ketoglutarate to succinate.

The multiplicity of the reactions catalyzed by cocarboxylase is offered as evidence for the view that cocarboxylase is a catalyst not only for the oxidation and decarboxylation of pyruvate, but also for many other reactions where pyruvate is one of the reacting substances. Krebs and Eggleston [132] propose that cocarboxylase is the catalyst for the Wood and Werkman [259] reaction: $CO_2 + CH_3COCOOH \rightarrow COOH{\cdot}CH_2{\cdot}COCOOH$. Their evidence for this proposal is based on a study of the rate of metabolism of pyruvate by liver suspensions under conditions where the "carboxylation" reaction was assumed to be the chief metabolic reaction for the removal of pyruvate. Under these conditions the rate of removal of pyruvate was greatly reduced in polyneuritic pigeon liver and was restored to normal by the *in vitro* addition of small quantities of thiamine. Banga, Ochoa and Peters [26] state that the following substances are components of the pyruvate oxidation system of brain and, probably, of other animal tissues: (1) cocarboxylase; (2) inorganic phosphate; (3) C_4 dicarboxylic acids; (4) "adenine nucleotide"; (5) Mg^{++} or Mn^{++}; and probably (6) pyridine nucleotide.

Lipmann,[140] working with *B. delbrückii*, has found that the dehydrogenation of pyruvic acid is coupled with the phosphorylation of adenylic acid. If such a process were to occur in animal tissues, it would afford another source of phosphorlyation energy. Ochoa [179] has been able to demonstrate such a coupling between the oxidation of pyruvic acid in brain tissue with phosphorylation processes. He obtained evidence in support of the view that oxidation of pyruvate is connected with phosphorylation of adenylic acid to adenosine polyphosphate which transfers its labile phosphate groups to either hexose monophosphate or glucose, with formation of hexose disphosphate. Both Lipmann's and Ochoa's work have identified the oxidation of pyruvate with the esterification of inorganic phosphate. Thus, the role of cocarboxylase as a catalyst in phosphorylation processes has become evident.

Miscellaneous Reactions in Which Cocarboxylase May Function Catalytically. The emphasis hitherto has been laid upon the functional role of cocarboxylase in pyruvic acid metabolism. However, other metabolic reactions catalyzed by cocarboxylase deserve our attention.

Quastel *et al.*[195] have shown that a choline ester is formed when brain cortex slices are allowed to respire in a medium containing eserine. Stedman and Stedman [212] have subsequently isolated acetyl choline from minced ox brain. It is of in-

* These results verify the conclusions of Sober, Lipton and Elvehjem [211] that cocarboxylase is essential for the synthesis of endogenous citric acid from its precursors.

terest to note that the rate of formation of acetyl choline was greatest when substances such as glucose, sodium lactate and sodium pyruvate were present in the medium under aerobic conditions. There was no accumulation of acetyl choline under anaerobic conditions. With our present knowledge of the pyruvate oxidase system of brain it becomes easy to visualize the necessity for the formation of an energy-rich acetyl intermediate (perhaps acetyl phosphate from the oxidation of pyruvate) for the acetylation of choline. In fact, Mann and Quastel [159] have demonstrated that the rate of synthesis of acetyl choline with pyruvate as substrate is greatly diminished in polyneuritic brain tissue. The addition of thiamine to the polyneuritic brain restored the synthesis to normal. Hence, the action of thiamine on acetyl choline synthesis is linked with pyruvate metabolism, and acetyl choline synthesis may represent but another outlet for the energy derived from the oxidation of pyruvate.

Many of the cardiac and gastrointestinal symptoms of thiamine deficiency may be relieved temporarily by the administration of choline esters. Thus, it would appear as though thiamine might be an inhibitor of choline esterase, and a deficiency of thiamine would result in an increased choline esterase activity, which in turn would lead to a lowered acetyl choline concentration. Such an interpretation may be useful in explaining the observation of McHenry [105] that a thiamine deficiency had a curative effect upon the fatty livers of rats kept on a low-choline diet. It has been demonstrated that thiamine exerts an inhibitory action upon choline esterase *in vitro*.[89, 42, 218, 88] * The ability of thiamine to augment the activity of acetyl choline in numerous test systems has been observed.[40, 172, 5, 1] It is not yet clear whether thiamine acts directly upon acetyl choline or functions indirectly by its ability to inhibit choline esterase, the enzyme which hydrolyzes acetyl choline to choline and acetic acid. Support for the inhibitory action of thiamine upon choline esterase has been given by *in vivo* experiments.[8, 9] The results of Zeller and Birkhäuser [260] are not of a confirmatory nature.

Thiamine is liberated when cholinergic nerves are stimulated electrically.[173] Since choline is liberated as the acetyl compound, Kuhn *et al.*[139] have proposed that thiamine may also be liberated as the acetyl compound, and they have studied the effect of acetyl aneurin on isolated rat intestine. Acetyl aneurin, in contrast to thiamine, had an activity similar to that of acetyl choline, an observation which seemed to strengthen the validity of their hypothesis. If the nerve impulses were mediated through acetyl aneurin, one would expect that the mediator, acetyl aneurin, would be capable of rapid removal from the site of action. Since acetyl aneurin is hydrolyzed very slowly, Glick and Antopol [89] do not believe it likely that it is such a mediator.

Edlbacher and Becker [58] have observed that the liver histidase and arginase activities are greater in thiamine-deficient than in normal rats. Zeller *et al.*[261] reported that the oxidative deamination of histamine was strongly inhibited by thiamine. In accordance with this observation, the authors observed an increased histaminase activity in thiamine-deficient rats. The physiological significance of these results is obscure.

The hypothesis that thiamine is required for the synthesis of fat from carbohydrate has been advanced.[165-168] The data of Longenecker *et al.*[156] are in agreement with such a concept. It is likely that cocarboxylase may play an important role in fat synthesis.[256] However, the precise mechanism of such a catalysis is not clear at present.

Nicotinic Acid

The discovery of the biological activity of nicotinic acid and its relationship to the pyridine nucleotides constitutes a most absorbing chapter in the field of biochemistry.

* Cocarboxylase was less effective than thiamine in its inhibitory action.[89] Byer and Harpuder [42] conclude that thiamine has no effect upon choline esterase activity.

The story of this vitamin affords an excellent illustration of the merging of the field of nutritional research with that of respiratory enzymes. Thus, it is undoubtedly true that the identification of the anti-pellagra factor was aided greatly by studies on the constitution of the pyridine nucleotides.

Huber [114] first prepared nicotinic acid in 1867 by the oxidation of nicotine. The isolation of nicotinic acid from biological materials was achieved by Suzuki *et al.*[221] in 1912 from rice polishings, and by Funk [82, 88] in 1913 from yeast concentrates which possessed antineuritic activity. Nicotinic acid itself displayed no activity in curing pigeon beriberi, but the results of Funk's experiments caused him to suggest that it might play a definite role in the vitamin B complex. In the light of our present knowledge, such a suggestion appears to have been based on incomplete experimental evidence. However, this was the first instance in which a definite vitamin function was ascribed to nicotinic acid. Williams [257] in 1917 investigated the antineuritic potency of nicotinic acid, but found that it did not cause any permanent improvement in polyneuritic fowl. Continuing the work on the vitamin nature of the pyridine derivatives, Szymanska and Funk [222] observed that nicotinic acid amide had an appetite-stimulating and weight-preserving action on pigeon beriberi.

With the exception of the above-mentioned work, very little interest was shown in the possible role of pyridine derivatives in living systems until the work of Warburg and von Euler in 1935. Warburg and Christian [240] characterized nicotinamide as one of the hydrolysis products from the coenzyme which they had isolated from blood, and which now is known as Coenzyme II. The accurate characterization of Coenzyme I quickly followed that of Coenzyme II and it, too, was found to contain nicotinic acid amide.[241, 242, 65] Kuhn and Vetter [137] in the same year isolated nicotinic acid amide from heart muscle. Thus, a new impetus was given to the application of nicotinic acid in the field of nutrition. Euler and Malmberg,[68] using a ration very similar to the Sherman-Bourquin diet supplemented with a thiamine concentrate and riboflavin, found no growth response with nicotinic acid and its amide, although the rats receiving nicotinic acid lived longer. Funk and Funk [85] reported that rats and pigeons on certain rations showed a larger food intake and better growth when given nicotinic acid, and especially the amide. Frost and Elvehjem [81] observed a growth stimulus from nicotinic acid when fed with adenylic acid to rats on a diet deficient in factor W.

The significance of nicotinic acid in the nutrition of microörganisms was also recognized at about this time.[124, 174, 130, 76]

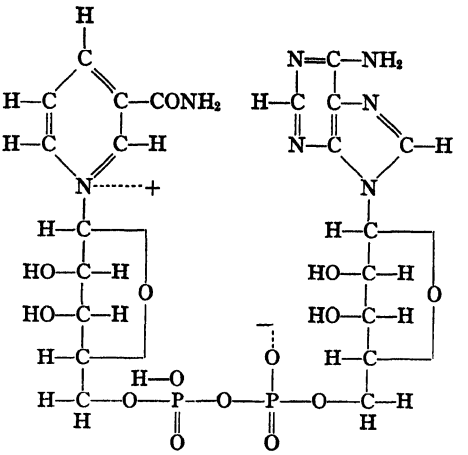
The earlier attempts to demonstrate the role of nicotinic acid in the animal economy cannot be considered as definite proof of the vitamin nature of this compound. The first conclusive evidence that nicotinic acid functions as a vitamin in higher animals was furnished by the demonstration that nicotinic acid or its amide was the "anti-blacktongue factor" in dogs.[61] The efficacy of nicotinic acid in human pellagra is now well known; Chick *et al.*[48] and Harris [107] have demonstrated its role in the nutrition of the pig and monkey.

It is clear from the foregoing that the isolation of substances important in biological oxidations greatly influenced the nutritional research on nicotinic acid.

The fact that nicotinic acid amide is a constituent of the pyridine nucleotides led to the natural assumption that the function of nicotinic acid is related to these nucleotides (Coenzymes I and II). Before discussing the evidence for such a relationship it may be well to review briefly our knowledge of the functional role of the pyridine nucleotides in the metabolic reactions of the organism. This subject has been adequately discussed in a number of review papers.[36, 21, 190, 68, 288, 77, 52, 163, 93, 188]

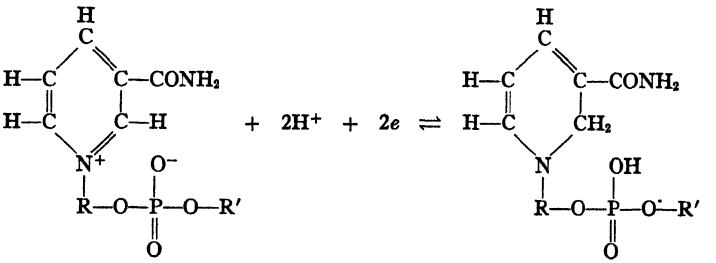
Diphosphopyridine nucleotide (cozymase, Coenzyme I) is found in a large variety of plant and bacterial species and is present in practically all animal tissues. Triphosphopyridine nucleotide (Coenzyme II) is also widely distributed in cells. It is safe to make the generalization that most metabolizing tissues contain these nucleo-

tides in varying amounts. The structure of diphosphopyridine nucleotide has been established by Warburg and Euler and their co-workers [246, 242, 71] and is given as:



Triphosphopyridine nucleotide has a similar structure, except that three phosphoric acid residues are present in the molecule instead of two.

The function of the pyridine nucleotides in enzyme reactions was originally elucidated for triphosphopyridine nucleotide by Warburg, Christian and Griese [246] in a reaction by which hexose monophosphate (Robison ester) was oxidized by molecular oxygen to phosphohexonic acid. They were able to show that triphosphopyridine nucleotide constituted a reversible oxidation-reduction system, and that the nicotinic acid amide moiety was the functional group in the process. The reversible oxidation-reduction reaction may be represented as follows:



The reduction occurs at the carbon-nitrogen linkage in the pyridine ring. Both the diphosphopyridine nucleotide and the triphosphopyridine nucleotide undergo the same type of oxidation-reduction reaction which is characteristic of their function whenever either of these compounds enters into a cycle of biological oxidation.

A number of metabolites are able to reduce diphosphopyridine nucleotide reversibly in the presence of their specific dehydrogenases. Thus it has been shown that alcohol, aldehyde, lactate, malate, triose, 1:3-diphosphoglyceraldehyde, glucose, α-glycerophosphate, β-hydroxybutyrate, glutamate and formate all require diphosphopyridine nucleotide for their oxidation.[94] Glucose and glutamate can apparently use either diphosphopyridine nucleotide or triphosphopyridine nucleotide. The reduction of the latter in the presence of specific enzymes is also effected by hexose monophosphate (Robison ester) and probably by its breakdown products, as well as by isocitric acid.[190] Since most of the reactions in which diphosphopyridine nucleotide and triphosphopyridine nucleotide are involved are reversible, it is possible for

one metabolite to be oxidized by the oxidized form of another metabolite of higher potential. Thus, certain dehydrogenase systems can be "linked" with others by means of the pyridine nucleotides.[51, 175, 97, 245] As an example, the hydroxybutyrate-acetoacetate and oxaloacetate-malate systems may be "linked" by diphosphopyridine nucleotide in the following manner:

$$\beta\text{-hydroxybutyrate} + \text{diphosphopyridine nucleotide} \rightarrow \text{acetoacetate} + \text{reduced diphosphopyridine nucleotide}$$
$$\text{reduced diphosphopyridine nucleotide} + \text{oxaloacetate} \rightarrow \text{diphosphopyridine nucleotide} + \text{malate.}$$

This type of reaction makes it conceivable that all the metabolites which can react with the pyridine nucleotides are in constant dynamic equilibrium with one another. The fundamental role of diphosphopyridine nucleotide in alcoholic fermentation and in glycolysis in animal tissues is discussed by Baumann and Stare[86] and Ball.[21] In addition to its action as a hydrogen carrier, diphosphopyridine nucleotide may also function in the transfer of phosphate. Thus, Ohlmeyer and Ochoa[182] have shown that diphosphopyridine nucleotide is an intermediate in the transfer of phosphate from phosphopyruvic acid to glucose.

The possible function of the pyridine nucleotides in protein metabolism as well as carbohydrate metabolism is discussed by Ball.[21] The role of the pyridine nucleotides in protein metabolism is made probable by the fact that the nucleotides have been shown to be active in the oxidation of glutamic acid.[66] The interrelationships involved indicate that protein and carbohydrate metabolism may be linked through the pyridine nucleotides.

It is apparent that, by virtue of their ability to function as reversible oxidation-reduction systems, the latter play a central role in the metabolism of carbohydrates and probably of proteins. The biological importance of nicotinic acid amide is emphasized by the fact that it serves as the functional group of the pyridine nucleotides.

The *in vitro* demonstration of the relationship between the pyridine nucleotides and nicotinic acid amide has stimulated considerable interest in an effort to establish the existence of this relationship *in vivo*.

Euler and co-workers were the first to undertake a systematic investigation of this problem. A number of papers have appeared from the Euler laboratory purporting to show that both nicotinic acid and the pyridine nucleotides are dietary essentials for the rat.[63, 64, 67, 69, 70] Their choice of the rat as the experimental animal in which to demonstrate the *in vivo* relationship between nicotinic acid and the pyridine nucleotides was an unfortunate one since it is very difficult to produce in this species an uncomplicated deficiency of the coenzymes or of any of their possible precursors. Recent evidence[50, 205] indicates that the normal rat is fully capable of synthesizing nicotinic acid and is, therefore, not dependent upon a dietary source of this compound. Recently von Euler and co-workers[72] found that the carcasses of rats fed diets low in the vitamin B complex tends to be lower in nicotinic acid amide and cozymase than those of normal rats. In the opinion of the reviewers, the experimental work presented by the Euler school does not offer conclusive evidence of an *in vivo* relationship between nicotinic acid and pyridine nucleotides. However, the existence of such a relationship in the rat has been indicated by Katzenelbogen *et al.*[116] These workers observed marked decreases in the diphosphopyridine nucleotide content of the liver and kidney cortex of hyperthyroid rats fed rations low in nicotinic acid. Supplementing such rations with excessive amounts of nicotinic acid restored the diphosphopyridine nucleotide content of these tissues to their normal values.

In other species, such as the dog, pig and human, considerable evidence has accumulated to prove definitely that the dietary intake of nicotinic acid markedly influences the coenzyme content of tissues. This demonstration has been made possible

by the fact that definite nicotinic acid deficiencies can be produced in the above-mentioned species.

Several workers have observed an increase in the coenzyme content of blood following the administration of nicotinic acid to human subjects.[125, 284, 127, 14] The *in vitro* synthesis of the pyridine nucleotides from nicotinic acid by defibrinated human blood has also been demonstrated.[127, 236, 18] Vilter *et al.*[236] question the ability of the erythrocytes to synthesize the coenzymes, and suggest the tentative hypothesis that nucleated cells are essential for the synthesis. Kohn and Klein [128] present strong evidence to substantiate their earlier claim that human erythrocytes are capable of synthesizing the pyridine nucleotides from nicotinic acid *in vitro*.

Vilter *et al.*, [284, 285] using *Bacillus influenzae* in a microbiological method, have reported decreases in the amount of factor V (Coenzymes I and II and possibly other related substances) in the whole blood of pellagrins. On the other hand, Kohn and Bernheim,[126] using *Haemophilus parainfluenzae*, did not find any significant decreases in the factor V of the erythrocytes from the blood of pellagrins. The results of Axelrod *et al.*,[18] who employed a yeast fermentation method specific for coenzyme I,[12] are in better agreement with those of Kohn and Bernheim. The conflicting results obtained by the various workers in this field can be attributed to the fact that different methods have been used both for measuring the pyridine nucleotide content and for expressing the units of concentration of these coenzymes.[235] The choice of patient would also undoubtedly affect the results obtained. Ballif *et al.*[22] found no reduction in the nicotinic acid amide content of the blood of ten pellagrins.

Axelrod *et al.*[18] have studied the diphosphopyridine nucleotide content of striated muscle in a series of normal subjects and patients with varying degrees of pellagra. The decrease in diphosphopyridine nucleotide content which was observed in the pellagrous patients could be correlated with the severity of the deficiency. The administration of nicotinic acid to the pellagrins led to marked increases in the diphosphopyridine nucleotide content of muscle.

The coenzyme content of certain tissues of the dog has been correlated with the dietary intake of nicotinic acid. Decreases in the coenzyme content of the liver and striated muscle of dogs suffering from a nicotinic acid deficiency have been reported.[129, 49, 188, 15] In addition, Pittman and Fraser [188] observed a decrease in the coenzyme content of the cardiac muscle. No significant changes were observed in the blood. Similar results in the pig have been reported by Axelrod *et al.*[15]

The ability of a number of microörganisms to synthesize the pyridine nucleotides from nicotinic acid has been demonstrated.[124, 76, 54, 123]

The above data clearly demonstrate that several different species are capable of synthesizing the pyridine nucleotides from nicotinic acid. This, plus the fact that a nicotinic acid deficiency results in a decreased synthesis of the coenzymes, argues strongly for the fact that at least one of the functions of nicotinic acid is to serve as a precursor for the pyridine nucleotides in the organism. However, we must not overlook the possibility that nicotinic acid may function in a manner entirely unrelated to its role as a constituent of the pyridine nucleotides. A number of lines of evidence are available which would indicate such a multiple function of nicotinic acid.

Dorfman *et al.*[55] have grown dysentery bacilli on a medium containing suboptimum amounts of nicotinic acid, and have observed that both the rate of reduction of methylene blue and the respiration rate of these bacterial cells are greatly increased, not only by the addition of diphosphopyridine nucleotide but also by the addition of nicotinamide or nicotinic acid. Dorfman *et al.*[58] have further found that nicotinamide is more active in stimulating the growth of the dysentery bacillus than an equivalent amount of either diphosphopyridine nucleotide or triphosphopyridine nucleotide. These results were taken as evidence to indicate that the function of nicotinic acid amide was not based entirely on its ability to serve as a constituent of the coenzyme molecule. However, alternate explanations could be based upon: (1) the ability of

the cell to synthesize the coenzymes from nicotinic acid, and (2) the varied permeability of the cell to nicotinamide and to the pyridine nucleotides. In a later paper, Saunders *et al.*[199] have presented evidence which appears to invalidate the objections which have been raised to the conclusions derived from their earlier work. The authors have suggested that nicotinamide may function as a constituent of a nicotinamide ironporphyrin which is capable of serving as a respiratory catalyst.[82] The results of McIlwain support the contention that nicotinamide may function as a respiratory catalyst distinct from the pyridine nucleotides.[170] He observed that the growth of Proteus promoted by diphosphopyridine nucleotide is much more strongly inhibited by pyridine-3-sulfonic acid than that promoted by nicotinamide. Similar results were obtained with Staphylococcus.

Another type of evidence is available which would suggest that a nicotinic acid deficiency cannot be explained solely in terms of a deficiency of the pyridine nucleotides. Thus, in the nicotinic acid-deficient dog, Kohn *et al.*[129] could not correlate the appearance of deficiency symptoms with the coenzyme content of various tissues. Furthermore, these authors could find no consistent relationship between the coenzyme content and the oxidative metabolism of isolated tissues. These experimental data prompted the suggestion that nicotinic acid enters the metabolism in essential ways as yet unknown. From an experimental standpoint any attempt to establish such correlations must admittedly be beset with many difficulties, and more work is necessary in this field before any definite conclusions concerning the role of nicotinic acid can be made. In their work with humans, Axelrod *et al.*[18] have determined that the antipellagric value of a compound is not necessarily associated with its ability to affect the diphosphopyridine nucleotide content of tissues. Thus, treatment of pellagrins with pyrazinemonocarboxylic acid and coramine resulted in a definite clinical improvement, which was not accompanied by consistent changes in the diphosphopyridine nucleotide content of erythrocytes or muscle. These compounds were also unable to effect an *in vitro* synthesis of diphosphopyridine nucleotide by defibrinated blood.

Dann and Handler [49] have reported that in certain organs of the dog a considerable proportion of the total nicotinic acid is present in the "free" form, *i.e.*, not as a component of the coenzymes. In blacktongue, the coenzyme level falls, but the "free" nicotinic acid is not considerably affected. This would suggest an important role for this "free" fraction of the total nicotinic acid distinct from its relationship with the pyridine nucleotides.

Mann and Quastel [160] have proposed still another mechanism for the biological activity of nicotinamide. They have observed that the rapid inactivation of diphosphopyridine nucleotide by brain and liver tissue is inhibited by nicotinamide, and they draw attention to the possibility that the therapeutic effects of nicotinamide may be due partially to its protective effects on diphosphopyridine nucleotide breakdown. To date, no experimental verification of this suggestion is at hand. The amounts of nicotinic acid amide necessary to prevent the breakdown of diphosphopyridine nucleotide *in vitro* appear to be unphysiological. However, the nucleotidase activity *in vivo* may be localized to portions of the tissue where the nicotinic acid is in fairly high concentration. This latter consideration would support the concept of Mann and Quastel.

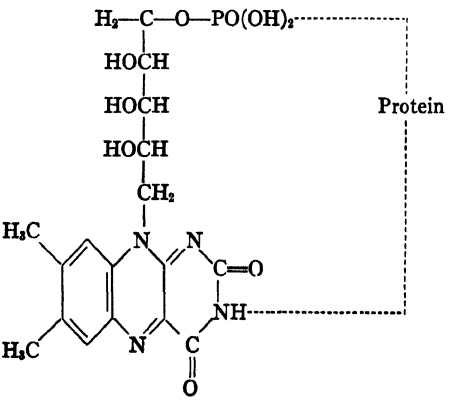
In summary, although the role of nicotinic acid is undoubtedly related to its ability to serve as a precursor of the pyridine nucleotides, it should be borne in mind that other functions for nicotinic acid remain a distinct possibility.

Riboflavin

In 1932 Warburg and Christian [239] isolated from bottom yeast a yellow conjugated protein which they called the "yellow enzyme." At the same time Banga and Szent-Györgyi [27] prepared a concentrate of "respiration-coferment" from the heart

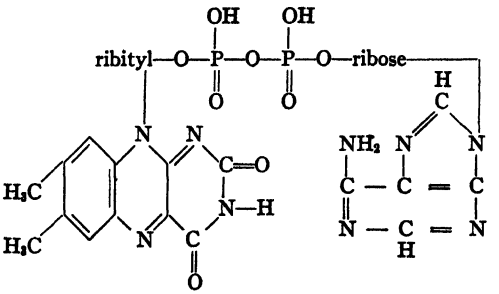
muscle of swine. The golden-yellow pigment present in this concentrate was named "cytoflav." The absorption spectrum of cytoflav revealed its structural similarity to Warburg's yellow enzyme,[140] and Kuhn and Rudy [135] showed cytoflav to be the phosphoric acid ester of riboflavin. Historically speaking, these discoveries appear to be the first steps in a long series of researches which have been necessary to establish our present concept of the close relationship between vitamins and oxidative enzymes. The story of riboflavin is also unusual, in that its isolation from natural sources and the recognition of its vitamin-like character paralleled very closely its identification as a constituent of the prosthetic group of an oxidative enzyme. The discovery of Warburg's "yellow enzyme" initiated the study of flavoproteins, and the past decade has witnessed the isolation of a large number of respiratory enzymes which contain riboflavin as a constituent of the prosthetic moiety. A resumé of this family of flavoprotein enzymes is in order, since the biological activity of riboflavin is intimately related to the enzymatic activity of the flavoproteins.

The Warburg-Christian "yellow enzyme," isolated in 1932 from bottom yeast, was the first flavoprotein to be described. The purification of this substance has been accomplished.[226, 228, 229, 255, 184] The yellow enzyme has been reversibly resolved into the protein group and the prosthetic component by Theorell [226, 228, 229] and by Warburg and Christian.[244] The prosthetic group, *i.e.*, the pigment portion, was found to exist as the phosphoric acid ester of riboflavin.[227] The synthesis of riboflavin-5′-phosphoric acid ester has enabled Kuhn *et al.*[186] to reconstruct the original yellow enzyme. It was postulated that the protein component of the enzyme is bound to the phosphoric acid radical and to the free imide group in position 3 of the riboflavin-5′-phosphoric acid ester.[231, 183] The combination of protein and prosthetic group has been given as:



Further indications that the protein and prosthetic group may be linked through the isoalloxazine ring are summarized by Ball.[20]

Research in the field of the flavoproteins was markedly stimulated by the observations of Straub [216] and of Warburg and Christian [243] that the prosthetic group of the *d*-amino acid oxidase was a riboflavin compound not identical with any hitherto recognized. Warburg and Christian [244] identified the prosthetic group as a riboflavin-adenine dinucleotide composed of one adenine, one riboflavin, two pentose and two phosphoric acid molecules. The compound yields adenosine-5-phosphoric acid and riboflavin phosphate on acid hydrolysis,[2] making it very probable that riboflavin-adenine dinucleotide is constructed through a union of flavin phosphoric acid with adenosine-5-phosphoric acid as indicated:



Riboflavin-adenine dinucleotide

Purification of the protein component of *d*-amino acid oxidase is described by Negelein and Brömel.[176]

The discovery of riboflavin-adenine dinucleotide led to the isolation of a number of flavoproteins containing riboflavin-adenine dinucleotide as their prosthetic group. In some cases the riboflavin-adenine dinucleotide is associated with other groupings of unknown constitution. It should be noted that the prosthetic group of cytochrome C reductase[103] is alloxazine mononucleotide. To date, the following respiratory enzymes may be classified as flavoproteins containing riboflavin-adenine dinucleotide as their prosthetic group:

(1) *d*-amino acid oxidase[243, 244]
(2) Straub's heart muscle flavoprotein[217]
(3) Milk flavoprotein *[19, 48]
(4) Pyruvic acid oxidase system of *B. delbrückii*[144]
(5) Yeast fumarate hydrase[78]
(6) Liver aldehyde oxidase *[92]
(7) Yeast flavoprotein of Haas[102]

The biological activity of the flavoprotein isolated from top yeast by Green *et al.*[96] has not been characterized as yet. The prosthetic component of this flavoprotein consists of riboflavin-adenine dinucleotide plus an unknown group.

The most important biological role of the flavoproteins appears to be that of "respiratory mediators," *i.e.*, compounds which catalyze the transfer of electrons between sluggish oxidation-reduction systems. As "respiratory mediators" the flavoproteins function in the transfer of electrons from the pyridine nucleotides to the cytochrome system. The cytochrome C reductase of Haas, Horecker and Hogness is an outstanding example of a flavoprotein which functions in this manner. This flavoprotein catalyzes the reduction of cytochrome C by reduced triphosphopyridine nucleotide at a very rapid rate, and probably represents the physiological mechanism for the oxidation of reduced triphosphopyridine nucleotide. The milk flavoprotein of Ball, and Corran and Green; Straub's heart-muscle flavoprotein; the yeast flavoprotein of Haas; and the diaphorase preparations of Adler *et al.*[4] also catalyze the oxidation of the reduced pyridine nucleotides. The physiological oxidants of these latter flavoproteins have not, however, been definitely determined, although it has been frequently suggested that these flavoproteins link the pyridine nucleotide systems to the cytochrome system.

The important role of flavoproteins as links between two sluggishly reversible systems was overlooked for a considerable time because the "yellow enzyme" of Warburg and Christian was autoxidizable. Thus, although the cytochrome system was absent from Warburg's original enzyme system which catalyzed the oxidation of hexosemonophosphate, the reduced "yellow enzyme" was oxidized by oxygen. How-

* The prosthetic groups of the flavoproteins marked with an asterisk contain unknown components in addition to riboflavin-adenine dinucleotide.

ever, doubt was cast upon the physiological role of flavoproteins as mediators between the pyridine nucleotides and oxygen by Theorell.[28a] He showed that the rate of oxidation of reduced yellow enzyme at oxygen tensions existing in animal tissues was so slow as to make it doubtful that that was the physiological pathway. Cytochrome C catalyzed the oxidation of reduced yellow enzyme, and Theorell suggested that the cytochrome system was the physiological oxidant of the "yellow enzyme." The cytochrome system has been shown to be involved in the oxidation of hexosemonophosphate by certain bacteria.[29] It is recognized that biological oxidations involving oxygen are inhibited by cyanide to at least 90 per cent. In the light of our present knowledge, it is quite probable that Warburg's original "yellow enzyme" is an artifact, and that flavoproteins analogous either to cytochrome reductase or to flavoproteins in which riboflavin-adenine dinucleotide is the prosthetic group, may be the true electron mediators of the cell.

The enzymatic activities of the remaining flavoproteins are summarized in Table 1.

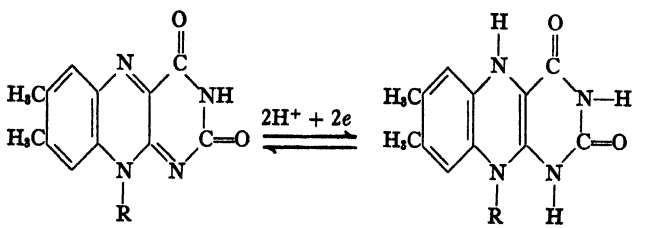
As is the case with the pyridine nucleotide dehydrogenases, the reductants of the flavoproteins of the above table are quite clearly defined. However, the information relative to the physiological pathways through which the reduced flavoproteins are oxidized is far too meager. In reconstructed enzyme systems these flavoproteins, with the exception of the yeast fumarate hydrase, are capable of being oxidized directly by oxygen. It is questionable whether this represents the true physiological mechanism.

Table 1

Enzyme	Reaction Catalyzed
d-Amino acid oxidase	Oxidative deamination of the *d*-amino acids
Milk flavoprotein	Oxidation of certain purines, aldehydes and dihydrocoenzyme I
Bacterial flavoprotein	Oxidative decarboxylation of pyruvic acid in *B. delbrückii*
Yeast fumarate hydrase	Reduction of fumaric acid by reduced dyestuffs
Liver aldehyde oxidase	Oxidation of aldehydes

Barron [29] has advanced objections to the concept of a direct transfer of electrons from the flavoproteins to molecular oxygen, and in support of his viewpoint [31] quotes evidence indicating that the oxidation of *d*-amino acids and pyruvic acid by bacteria possessing the cytochrome system is completely inhibited by cyanide.[38, 88] The intermediation of the cytochrome system in the oxidation of these substrates is indicated.

In summarizing the biological activity of the flavoproteins, it may be stated that they function as electron mediators in many biological oxidations by virtue of their ability to undergo a cycle of oxidation and reduction. The isoalloxazine ring is considered to be the functional group in these electroactive oxidation-reduction systems as indicated:



Oxidized form *Reduced form*

The elucidation of the true nature of the physiological oxidants of many of the flavoproteins remains a fruitful field for investigation. It is apparent that the flavo-

proteins are capable of playing fundamental roles in carbohydrate, protein, and purine metabolism.

Reconstructed enzyme systems have been generally employed to demonstrate the biological activity of the flavoproteins. However, results from *in vivo* studies have substantiated the physiological importance of these compounds. The direct approach has been employed by Urban and Peugnet,[233] who have been able to demonstrate that cytochrome C and the yellow ferment transfer electrons during individual twitches of a frog muscle. A decreased respiration of rat-liver tissue in riboflavin deficiency has been reported by Groen and Schuyl.[100] Hastings *et al.*,[109] although unable to confirm the results of Groen and Schuyl with regard to the respiration of liver, did observe a significant increase in the respiration of rat diaphragm in riboflavin deficiency. Adams[8] reports that the oxygen consumption of the skin of vitamin G-deficient rats is considerably lower than that of normal rats. Kik *et al.*[117] were unable to find any significant changes in the digestive efficiency of either the proteolytic or amylolytic enzymes in vitamin G deficiency in the rat, although they did observe a 20 per cent decrease in blood serum phosphatase.

A more clear-cut demonstration of the *in vivo* relationship between riboflavin and respiratory enzymes has been furnished by studies of individual enzyme systems or their components in riboflavin deficiency. Thus, Ochoa and Rossiter[181] have reported a decrease in the total riboflavin-adenine dinucleotide content of rat heart and liver in riboflavin avitaminosis. The dinucleotide content was rapidly restored to its normal level by the administration of riboflavin to the deficient rats. The values for the total riboflavin-adenine dinucleotide content of normal rat liver and kidney given by Ochoa and Rossiter[181] check well with the figures of Axelrod *et al.*[17] for the riboflavin content of these tissues if it is assumed that all the riboflavin is present as the dinucleotide. The decrease in total riboflavin-adenine dinucleotide content of liver observed by Ochoa and Rossiter is of the same order as that observed by Axelrod *et al.* for the riboflavin decrease in this tissue. A decrease in the riboflavin-adenine dinucleotide content could conceivably affect the concentrations of a number of flavoproteins, and such changes have been observed. Thus, Axelrod *et al.*[17] and Rossiter[196] have shown that a riboflavin deficiency in the rat results in a lowering of the *d*-amino acid oxidase content of various tissues. The former authors have indicated that other members of the vitamin B complex may facilitate the synthesis of this enzyme from riboflavin. The protein constituent of *d*-amino oxidase was apparently not affected in this deficiency.[196] In a similar study, Axelrod and Elvehjem[18] have reported that the xanthine oxidase activity of rat liver is greatly diminished in riboflavin deficiency. Although no direct evidence was presented to prove the point, the authors are of the opinion that the prosthetic group of this enzyme was diminished in the deficiency.

It has been suggested that succinic dehydrogenase is a flavoprotein.[190, 191] The final proof that riboflavin enters into the structure of any of the components of the succinoxidase system must await the isolation in pure state of such components. However, the observations of Axelrod *et al.*[16] that the dietary intake of riboflavin affects the succinoxidase content of rat-liver tissue constitute *prima facie* evidence in support of the flavoprotein nature of this enzyme.

Another illustration of the physiological relationship between riboflavin and riboflavin-adenine dinucleotide has been furnished by Klein and Kohn,[122] who demonstrated the synthesis of riboflavin-adenine dinucleotide from riboflavin by human blood cells both *in vitro* and *in vivo*.

Verzar and his school have been active proponents of the theory that the adrenal cortical hormones are catalysts for the phosphorylation of riboflavin and other members of the vitamin B complex prior to their incorporation as coenzymes of certain enzyme systems. This theory presents many attractive possibilities. However, the

bulk of our present evidence renders such a thesis untenable. An excellent summary of this controversial subject is given by Clark.[45]

Enzymatic Relationships of Other Vitamins

The discussion of the biological activities of thiamine, nicotinic acid and riboflavin largely completes our present knowledge of the interrelationships between vitamins and enzyme systems. However, suggestions have been proposed with regard to the enzymatic activities of the remaining vitamins, and these will be discussed briefly in this section.

Vitamin A. Green [99] reported that vitamin A deficiency brings about a pronounced fall in the esterase content of blood serum. However, he considered this to be a secondary change. Sure *et al.*[219] verified Green's results and observed further a decrease in liver esterase and lipase. Sure and DeWitt [220] reported slight decreases in the respiration of cerebrum and heart in vitamin A-deficient rats. No correlation between the severity of symptoms and extent of respiration was noted. Sister M. V. Ruddy [197] has reported that liver tissue of vitamin A-deficient rats has an abnormally low oxygen consumption. Feeding of vitamin A restored the metabolism to normal. Wald [237] has shown that visual purple, an enzyme which catalyzes the process of photoreception, is a conjugated protein with vitamin A as the prosthetic group.

Vitamin D. The high concentration of plasma phosphatase in vitamin D deficiency has long been recognized. This effect may be a secondary one due to the loss of phosphatase from the rachitic bones. Presnell [193] reported a decrease in the oxygen uptake of skin from rachitic rats.

Vitamin E. Friedman and Mattill [80] observed an increased oxygen uptake in muscle tissues during vitamin E deficiency.

Pyridoxine. The chemical similarity of pyridoxine to the pyridine nucleotides suggests an enzymatic role for this vitamin. It is significant that, in yeast, pyridoxine occurs in combination with protein.[138] However, no specific enzymatic function for pyridoxine has yet been determined. The relationship of pyridoxine to fat metabolism has been indicated by Halliday [104] and Engel.[62] A possible role of pyridoxine in protein metabolism has been suggested by McHenry and Gavin.[169]

Biotin. The catalytic function of biotin in fat metabolism is suggested by the observation that the feeding of biotin results in fatty livers characterized by a high content of cholesterol.[86]

A factor essential for the respiration of root nodule bacteria without concomitant growth has been designated as coenzyme R by Allison *et al.*[6] Coenzyme R is now known to be identical with biotin. Burk *et al.*[41] have demonstrated the ability of biotin to affect both the fermentation and respiration of yeast. A high Meyerhof oxidation quotient was reported for biotin-deficient yeast.

Pantothenic Acid. As is the case with biotin, the role of pantothenic acid in fat metabolism has been indicated by its ability to affect the concentration of liver fat.[62]

Pratt and Williams [192] have reported the stimulative effect of pantothenic acid upon the respiration of yeast, upon the fermentation by dialyzed maceration juice from yeast, and on the respiration of certain vegetable tissues.

Experiments suggestive of the possible role of pantothenic acid in porphyrin metabolism and in the regulation of water metabolism have been reported.[75, 164, 44]

***p*-Aminobenzoic Acid.** The chromotrichial activity of *p*-aminobenzoic acid [7, 161] and the antagonism between hydroquinone and *p*-aminobenzoic acid with respect to graying of hair, have indicated a possible relationship between the metabolism of phenolic substances and *p*-aminobenzoic acid. The observations that it has a pronounced influence on tyrosinase activity [258] and is capable of modifying enzymatic melanin formation [162] are in accord with such a relationship. The experiments of Lipmann [148] may have some bearing on this question.

Ascorbic Acid. The relatively high concentrations of ascorbic acid in living tissue and its intense reducing action early suggested the theory that this compound serves primarily as a hydrogen transport agent. Numerous attempts have been made to demonstrate the "carrier" function of ascorbic acid in animal tissues with but little success. King [119] concludes "that the evidence is more negative than positive in regard to the postulated respiratory role of ascorbic acid in animal tissues." Objections to the "carrier" role of ascorbic acid have also been raised by Barron.[80] Stronger evidence is available in support of the theory that ascorbic acid serves as a respiratory catalyst in certain plants.[119] However, the application of the more recent information relative to the biocatalysts necessary for the complete transfer of hydrogen (*i.e.*, flavoproteins) may help considerably toward closing the gaps in our present knowledge of the physiological function of ascorbic acid. It should also be borne in mind that the combination of acidity and reducing action which characterize the chemical behavior of ascorbic acid may determine its role as that of a regulating and protective agent toward other enzyme systems. The assumption of a protective action for ascorbic acid may serve to rationalize the observations of Harrer and King [106] that the concentrations of liver esterase and succinic dehydrogenase are diminished in scurvy. The inhibiting or activating effect of ascorbic acid on such enzymes as cathepsin, papain, amylase, arginase, catalase, urease, tyrosinase and nuclease may conceivably be attributed to the non-specific acidity and reducing capacity of ascorbic acid.[118]

Other functions for ascorbic acid have been proposed. Lemberg and associates [142] have demonstrated the part played by ascorbic acid in catalyzing the formation of bile pigments from oxyhemoglobin. Ecker and co-workers [57, 56] have shown a definite correlation between the complement titer of both guinea pig and human serum and the ascorbic acid content of such sera. Ascorbic acid is believed to function in maintaining the complement in a reduced, active state.

Sealock and co-workers [200] have proposed a functional role of ascorbic acid in the metabolism of phenylalanine and tyrosine.

Summary

A large body of evidence has been accumulated in support of the thesis that the biological activities of thiamine, nicotinic acid and riboflavin are due to their abilities to serve as components of enzymatic systems. The experimental proof has been derived from both *in vivo* and *in vitro* researches. A similar relationship has been postulated to explain the physiological role of other vitamins, but, in these cases, clear-cut experimental evidence for such functions has not been forthcoming.

The knowledge that the functional role of the vitamins is closely correlated with their enzymatic relationships is but the first step toward the complete elucidation of the biological action of vitamins. Some of the questions which still remain unanswered are included in the following:

(1) What individual metabolic reactions are affected in a given vitamin deficiency? Thus, although it is recognized that the cocarboxylase, the pyridine nucleotide and the flavoprotein concentrations of certain tissues are materially affected in some vitamin deficiencies, the information relative to the derangements in the specific metabolic reactions catalyzed by these enzyme systems is quite unsatisfactory. Answers to this question have been sought for most vigorously in the case of thiamine deficiency. The first attempts in this direction are being made in studies of riboflavin and nicotinic acid deficiencies.

(2) How are the over-all metabolic reactions of the organism affected by changes in the activities of their component parts? An answer to this question will involve the *in vivo* corroboration of the *in vitro* studies carried out with reconstructed enzyme systems. To take a specific example, it remains to be demonstrated that the decrease in the xanthine oxidase system, which has been observed in riboflavin

deficiency, in any way affects the ability of the animal to metabolize purines. A similar situation exists in other vitamin deficiencies where studies on carbohydrate, fat and protein metabolism have not been exhaustively pursued.

(3) Is there any positive correlation between the extent of the metabolic changes and the severity of the deficiency syndrome? Attempts to answer this question will promote the establishment of a more quantitative relationship between the enzymatic derangements and the observed pathological changes. Studies of this type may also be of value in any attempted correlation between a specific pathological event and a disturbance in a given enzymatic system.

References

1. Abderhalden, E. and Abderhalden, R., *Pflügers Arch. ges. Physiol.*, **240**, 388 (1938).
2. Abraham, E. P., *Biochem. J.*, **33**, 543 (1939).
3. Adams, P. D., *J. Biol. Chem.*, **116**, 641 (1936).
4. Adler, E., v. Euler, H., Günther, G. and Plass, M., *Skand. Archiv. Physiol.*, **82**, 61 (1939).
5. Agid, R., Beauvallet, M., and Minz, B., *Compt. rend. soc. biol.*, **126**, 982 (1937).
6. Allison, F. E., Hoover, S. R. and Burk, D., *Science*, **78**, 217 (1933).
7. Ansbacher, S., *Science*, **93**, 164 (1941).
8. Antopol, W., Glaubach, S. and Glick, D., *Proc. Soc. Exp. Biol. Med.*, **42**, 280 (1939).
9. Antopol, W., Tuchman, L. and Schifrin, A., *Proc. Soc. Exp. Biol. Med.*, **36**, 46 (1937).
10. Arnold, A. and Elvehjem, C. A., *Am. J. Physiol.*, **126**, 289 (1939).
11. Auhagen, E., *Z. physiol. Chem.*, **204**, 149 (1932).
12. Axelrod, A. E. and Elvehjem, C. A., *J. Biol. Chem.*, **131**, 77 (1939).
13. Axelrod, A. E. and Elvehjem, C. A., *J. Biol. Chem.*, **140**, 725 (1941).
14. Axelrod, A. E., Gordon, E. S. and Elvehjem, C. A., *Am. J. Med. Sci.*, **199**, 697 (1940).
15. Axelrod, A. E., Madden, R. J. and Elvehjem, C. A., *J. Biol. Chem.*, **131**, 85 (1939).
16. Axelrod, A. E., Potter, V. R. and Elvehjem, C. A., *J. Biol. Chem.*, **142**, 85 (1942).
17. Axelrod, A. E., Sober, H. A. and Elvehjem, C. A., *J. Biol. Chem.*, **134**, 749 (1940).
18. Axelrod, A. E., Spies, T. D. and Elvehjem, C. A., *J. Biol. Chem.*, **138**, 667 (1941).
19. Ball, E. G., *J. Biol. Chem.*, **128**, 51 (1938).
20. Ball, E. G., *Cold Spring Harbor Symposia on Quant. Biol.*, **7**, 100 (1939).
21. Ball, E. G., *Bull. Johns Hopkins Hosp.*, **65**, 253 (1939).
22. Ballif, L., Lwoff, A., Querido, A. and Ornstein, I., *Compt. rend. soc. biol.*, **131**, 903 (1939).
23. Banerji, G. G., *Biochem. J.*, **34**, 1329 (1940).
24. Banerji, G. G. and Harris, L. J., *Biochem. J.*, **33**, 1346 (1939).
25. Banga, I., Ochoa, S. and Peters, R. A., *Biochem. J.*, **33**, 1109 (1939).
26. Banga, I., Ochoa, S. and Peters, R. A., *Biochem. J.*, **33**, 1980 (1939).
27. Banga, I. and Szent-Györgyi, A., *Biochem. Zeit.*, **246**, 203 (1932).
28. Barron, E. S. G., *Physiol. Rev.*, **19**, 184 (1939).
29. Barron, E. S. G., *Cold Spring Harbor Symposia Quant. Biol.*, **7**, 108 (1939).
30. Barron, E. S. G., *Cold Spring Harbor Symposia Quant. Biol.*, **7**, 145 (1939).
31. Barron, E. S. G., *Ann. Review Biochem.*, **10**, 1 (1941).
32. Barron, E. S. G., Unpublished experiments.
33. Barron, E. S. G. and Friedemann, T. E., *J. Biol. Chem.*, **137**, 593 (1941).
34. Barron, E. S. G. and Lyman, C. M., *J. Biol. Chem.*, **127**, 143 (1939).
35. Barron, E. S. G. and Lyman, C. M., Cited in Barron, E. S. G., *Physiol. Rev.*, **19**, 184 (1939).

35a. Barron, E. S. G. and Lyman, C. M., Cited in Barron (31).

36. Baumann, C. A. and Stare, F. J., *Physiol. Rev.*, **19**, 353 (1939).
37. Bergel, F. and Todd, A. R., *J. Chem. Soc.*, **140**, 1504 (1937).
38. Bernheim, F., Bernheim, M. L. C. and Webster, M. D., *J. Biol. Chem.*, **110**, 165 (1935).
39. Birch, T. W. and Mann, P. J. G., *Biochem. J.*, **28**, 622 (1934).
40. Brecht, K., *Pflügers Arch. ges. Physiol.*, **243**, 714 (1940).
41. Burk, D., Winzler, R. J. and du Vigneaud, V., *Proc. Am. Soc. Biol. Chem.*, xxi (1941).
42. Byer, J. and Harpuder, K., *J. Pharmacol.*, **70**, 328 (1940).
43. Chick, H., Macrae, T. F., Martin, A. J. P. and Martin, C. J., *Biochem. J.*, **32**, 10 (1938).
44. Chick, H., Macrae, T. F. and Worden, A. N., *Biochem. J.*, **34**, 580 (1940).
45. Clark, W. G., *Endocrinology*, **28**, 545 (1941).
46. Cline, J. K., Cited in Williams, R. R. in Ruzicka, L. and Stepp, W., Ergebnisse Vitamin und Hormonforsch. (Leipzig, Vol. **1**, p. 254 [1938]).
47. Cline, J. K., Cited in Williams, R. R. and Spies, T. D., "Thiamine and Its Use in Medicine," p. 175, New York, Macmillan Medical Monographs, (The Macmillan Co.) (1938).
48. Corran, H. S. and Green, D. E., *Biochem. J.*, **32**, 2231 (1938).
49. Dann, W. J. and Handler, P., *J. Nutr.*, **22**, 409 (1941).
50. Dann, W. J. and Kohn, H. I., *J. Biol. Chem.*, **136**, 435 (1940).
51. Dewan, J. G. and Green, D. E., *Biochem. J.*, **31**, 1074 (1937).
52. Dixon, M., in Nord, F. F. and Weidenhagen, R., *Ergebnisse Enzymforsch.*, **8**, 217 (1939).
53. Dorfman, A., Koser, S. A., Horwitt, M. K., Berkman, S. and Saunders, F., *Proc. Soc. Exp. Biol. Med.*, **43**, 434 (1940).
54. Dorfman, A., Koser, S. A., Reames, H. R., Swingle, K. F. and Saunders, F., *J. Infectious Diseases*, **65**, 163 (1939).
55. Dorfman, A., Koser, S. A. and Saunders, F., *Science*, **90**, 544 (1939).
56. Ecker, E. E., Pillemer, L., Guffitts, J. J. and Schwartz, W. P., *J. Am. Med. Assoc.*, **112**, 1449 (1939).
57. Ecker, E. E., Pillemer, L., Wertheimer, D. and Gradis, H., *J. Immunol.*, **34**, 19 (1938).
58. Edlbacher, S. and Becker, M., *Z. physiol. Chem.*, **265**, 72 (1940).
59. Eijkman, C., *Arch. path. Anat.*, **149**, 187 (1897).
60. Eijkman, C., *Arch. Hyg.*, **58**, 150 (1906).
61. Elvehjem, C. A., Madden, R. J., Strong, F. M. and Woolley, D. W., *J. Biol. Chem.*, **123**, 137 (1938).
62. Engel, R. W., *Proc. Am. Soc. Biol. Chem.*, xxxvii (1941).
63. Euler, H. v., *Ergebnisse Physiol.*, **38**, 1 (1936).
64. Euler, H. v., *Angew. Chem.*, **50**, 831 (1937).

65. Euler, H. v., Albers, H. and Schlenk, F., *Z. physiol. Chem.*, **237**, 1 (1935).
66. Euler, H. v., Adler, E., Gunther, G. and Das, N. B., *Z. physiol. Chem.*, **254**, 61 (1938).
67. Euler, H. v., Heiwinkel, H., Malmberg, M., Robeznieks, I. and Schlenk, F., *Arkiv. för Kemi, Mineralogi, och Geologi*, 12A, no. 25 (1937).
68. Euler, H. v., and Malmberg, M., *Biochem. Zeit.*, **284**, 455 (1936).
69. Euler, H. v., Malmberg, M., Heiwinkel, H. and Schlenk, F., *Arkiv. för Kemi, Mineralogi, och Geologi*, 12B, no. 39 (1937).
70. Euler, H. v., Malmberg, M., Robeznieks, I. and Schlenk, F., *Naturwissenschaften*, **26**, 45 (1938).
71. Euler, H. v., and Schlenk, F., *Z. physiol. Chem.*, **246**, 64 (1936).
72. Euler, H. v., Schlenk, F., Melzer, L. and Högberg, B., *Z. physiol. Chem.*, **258**, 212 (1939).
73. Euler, H. v., and Vestin, R., *Naturwissenschaften*, **25**, 416 (1937).
74. Evans, H. M. and Lepkovsky, S., *Science*, **68**, 298 (1928).
75. Figge, F. H. J. and Atkinson, W. B., *Proc. Soc. Exp. Biol. Med.*, **48**, 112 (1941).
76. Fildes, P., *Brit. J. Exper. Path.*, **19**, 239 (1938).
77. Fischer, F. G., in Nord, F. F. and Weidenhagen, R., *Ergebnisse Enzymforsch.*, **8**, 185 (1939).
78. Fischer, F. G., Roedig, A. and Raüch, K., *Naturwissenschaften*, **27**, 197 (1939).
79. Fisher, R. B., *Biochem. J.*, **25**, 1410 (1931).
80. Friedman, I. and Mattill, H. A., *Am. J. Physiol.*, **131**, 595 (1941).
81. Frost, D. V. and Elvehjem, C. A., *J. Biol. Chem.*, **121**, 255 (1937).
82. Funk, C., *Brit. Med. J.*, **1**, 814 (1913).
83. Funk, C., *J. Physiol.*, **46**, 173 (1913).
84. Funk, C., *Z. physiol. Chem.*, **89**, 278 (1914).
85. Funk, C. and Funk, I. C., *Proc. Am. Soc. Biol. Chem.*, **119**, 35 (1937).
86. Gavin, G. and McHenry, E. W., *J. Biol. Chem.*, **141**, 619 (1941).
87. Gavrilescu, N. and Peters, R. A., *Biochem. J.*, **25**, 1397 (1931).
88. Glick, D. and Antopol, W., *J. Pharmacol.*, **65**, 389 (1939).
89. Glick, D. and Antopol, W., *Proc. Soc. Exp. Biol. Med.*, **42**, 396 (1939).
90. Goodhart, R. S. and Sinclair, H. M., *Biochem. J.*, **33**, 1099 (1939).
91. Goodhart, R. S. and Sinclair, H. M., *J. Physiol.*, **95**, 57P (1939).
92. Gordon, A. H., Green, D. E. and Subrahmanyan, V., *Biochem. J.*, **34**, 764 (1940).
93. Green, D. E., "Mechanism of Biological Oxidations," p. 36, Cambridge (Cambridge University Press), 1940.
94. Green, D. E., "Advances in Enzymology," Vol. 1, p. 180, New York, Interscience Publishers, Inc., 1941.
95. Green, D. E., Herbert, D. and Subrahmanyan, V., *J. Biol. Chem.*, **138**, 327 (1941).
96. Green, D. E., Knox, W. E. and Stumpf, P. K., *J. Biol. Chem.*, **138**, 775 (1941).
97. Green, D. E., Needham, D. M. and Dewan, J. G., *Biochem. J.*, **31**, 2327 (1937).
98. Green, D. E., Westerfeld, W. W., Vennesland, B. and Knox, W. E., *J. Biol. Chem.*, **140**, 683 (1941).
99. Green, H. N., *Biochem. J.*, **28**, 16 (1934).
100. Groen, J. and Schuyl, J. W., *Arch. néerland. physiol.*, **23**, 271 (1938).
101. György, P., Rose, C. S., Eaken, R. E., Snell, E. E. and Williams, R. J., *Science*, **93**, 477 (1941).
102. Haas, E., *Biochem. Zeit.*, **298**, 378 (1938).
103. Haas, E., Horecker, B. L. and Hogness, T. R., *J. Biol. Chem.*, **136**, 747 (1940).
104. Halliday, N., *J. Nutr.*, **16**, 285 (1938).
105. Harper, H. A. and Deuel, H. J., Jr., *J. Biol. Chem.*, **137**, 233 (1941).
106. Harrer, C. J. and King, C. G., *J. Biol. Chem.*, **138**, 111 (1941).
107. Harris, L. J., *Biochem. J.*, **31**, 1414 (1937).
108. Hart, E. B., McCollum, E. V., Steenbock, H. and Humphrey, G. C., *Wis. Agric. Exp. Sta. Res. Bull.*, **17**, (1911).
109. Hastings, A. B., Muus, J. and Bessey, O. A., *J. Biol. Chem.*, **129**, 295 (1939).
110. Hennessy, D. J. and Cerecedo, L. R., *J. Am. Chem. Soc.*, **61**, 179 (1939).
111. Hills, G. M., *Biochem. J.*, **32**, 383 (1938).
112. Horowitz, N. H. and Heegaard, E., *J. Biol. Chem.*, **137**, 475 (1941).
113. Houston, J. and Kon, S. K., *Nature*, **143**, 558 (1939).
114. Huber, C., *Ann.*, **141**, 271 (1867).
115. Inawashiro, R. and Hayasaka, E., *Tohoku J. Exp. Med.*, **12**, 1 (1928).
116. Katzenelbogen, E., Axelrod, A. E. and Elvehjem, C. A., *J. Biol. Chem.*, **141**, 611 (1941).
117. Kik, M. C., Sure, B. and Buchanan, K. S., *Am. J. Dig. Diseases and Nutr.*, **3**, 490 (1936).
118. King, C. G., *J. Am. Med. Assoc.*, **111**, 1098 (1938).
119. King, C. G., *Cold Spring Harbor Symposia Quant. Biol.*, **7**, 137 (1939).
120. Kinnersley, H. W. and Peters, R. A., *Biochem. J.*, **22**, 419 (1928).
121. Kinnersley, H. W. and Peters, R. A., *Biochem. J.*, **24**, 711 (1930).
122. Klein, J. R. and Kohn, H. I., *J. Biol. Chem.*, **136**, 177 (1940).
123. Kligler, I. J. and Grossowicz, N., *Nature*, **146**, 652 (1940).
124. Knight, B. C. J. G., *Biochem. J.*, **31**, 731 (1937).
125. Kohn, H. I., *Biochem. J.*, **32**, 2075 (1938).
126. Kohn, H. I. and Bernheim, F., *J. Clin. Investigation*, **18**, 585 (1939).
127. Kohn, H. I. and Klein, J. R., *J. Biol. Chem.*, **130**, 1 (1939).
128. Kohn, H. I. and Klein, J. R., *J. Biol. Chem.*, **135**, 685 (1940).
129. Kohn, H. I., Klein, J. R. and Dann, W. J., *Biochem. J.*, **33**, 1432 (1939).
130. Koser, S. A., Dorfman, A. and Saunders, F., *Proc. Soc. Exp. Biol. Med.*, **38**, 311 (1938).
131. Krampitz, L. O. and Werkman, C. H., *Biochem. J.*, **35**, 595 (1941).
132. Krebs, H. A. and Eggleston, L. V., *Biochem. J.*, **34**, 1383 (1940).
133. Kuhn, R. and Boulanger, P., *Ber.*, **69**, 1557 (1936).
134. Kuhn, R. and Desnuelle, P., *Ber.*, **70**, 1907 (1937).
135. Kuhn, R. and Rudy, H., *Z. physiol. Chem.*, **239**, 47 (1936).
136. Kuhn, R., Rudy, H. and Weygand, F., *Ber.*, **69**, 2034 (1936).
137. Kuhn, R. and Vetter, H., *Ber.*, **68**, 2374 (1935).
138. Kuhn, R., Wendt, G. and Westphal, O., *Ber.*, **72**, 310 (1939).
139. Kuhn, R., Wieland, T. and Huebschmann, H., *Z. physiol. Chem.* **259**, 48 (1939).
140. Laki, K., *Biochem. Zeit.*, **266**, 202 (1933).
141. Langenbeck, W., in Nord, F. F. and Weidenhagen, R., *Ergebnisse Enzymforsch.*, **2**, 314 (1933).
142. Lemberg, R., Legge, J. W. and Lockwood, W. H., *Biochem. J.*, **33**, 754 (1939).
143. Lipmann, F., *Nature*, **140**, 849 (1937).
144. Lipmann, F., *Enzymologia*, **4**, 65 (1937).
145. Lipmann, F., *Cold Spring Harbor Symposia Quant. Biol.*, **7**, 248 (1939).
146. Lipmann, F., *Nature*, **143**, 281 (1939).
147. Lipmann, F., *Enzymologia*, **7**, 142 (1939).
148. Lipmann, F., *J. Biol. Chem.*, **139**, 977 (1941).
149. Lipmann, F. and Perlmann, G., *J. Am. Chem. Soc.*, **60**, 2574 (1938).

150. Lipschitz, M. A., Potter, V. R. and Elvehjem, C. A., *J. Biol. Chem.*, **123**, 267 (1938).
151. Lipschitz, M. A., Potter, V. R. and Elvehjem, C. A., *Biochem. J.*, **32**, 474 (1938).
152. Lipschitz, M. A., Potter, V. R. and Elvehjem, C. A., *J. Biol. Chem.*, **124**, 147 (1938).
153. Lipton, M. A. and Elvehjem, C. A., *Cold Spring Harbor Symposia Quant. Biol.*, **7**, 184 (1939).
154. Lipton, M. A. and Elvehjem, C. A., *J. Biol. Chem.*, **136**, 637 (1940).
155. Lohmann, K. and Schuster, P., *Biochem. Zeit.*, **294**, 183 (1937).
156. Longenecker, H. E., Gavin, G. and McHenry, E. W., *J. Biol. Chem.*, **134**, 693 (1940).
157. Lu, G. D., *Biochem. J.*, **33**, 774 (1939).
158. Lu, G. D. and Platt, B. S., *Biochem. J.*, **33**, 1538 (1939).
159. Mann, P. J. G. and Quastel, J. H., *Nature*, **145**, 856 (1940).
160. Mann, P. J. G. and Quastel, J. H., *Biochem. J.*, **35**, 502 (1941).
161. Martin, G. J. and Ansbacher, S., *J. Biol. Chem.*, **138**, 441 (1941).
162. Martin, G. J., Wisansky, W. A. and Ansbacher, S., *Proc. Soc. Exp. Biol. Med.*, **47**, 26 (1941).
163. Martius, C., in Nord, F. F. and Weidenhagen, R., *Ergebnisse Enzymforsch.*, **8**, 247 (1939).
164. McElroy, L. W., Salomon, K., Figge, F. H. J. and Cowgill, G. R., *Science*, **94**, 466 (1941).
165. McHenry, E. W., *J. Physiol.*, **89**, 287 (1937).
166. McHenry, E. W., *Science*, **86**, 200 (1937).
167. McHenry, E. W. and Gavin, G., *J. Biol. Chem.*, **125**, 653 (1938).
168. McHenry, E. W. and Gavin, G., *J. Biol. Chem.*, **128**, 45 (1939).
169. McHenry, E. W. and Gavin, G., *J. Biol. Chem.*, **138**, 471 (1941).
170. McIlwain, H., *Brit. J. Exp. Path.*, **21**, 136 (1940).
171. Meikeljohn, A. P., Passmore, R. and Peters, R. A., *Biochem. J.*, **26**, 1872 (1932).
172. Minz, B., *Compt. rend. acad. sci.*, Paris, **205**, 576 (1937).
173. Minz, B., *Compt. rend. soc. biol.*, **127**, 1251 (1938).
174. Mueller, J. H., *J. Biol. Chem.*, **120**, 219 (1937).
175. Needham, D. M. and Lu, G. D., *Biochem. J.*, **32**, 2040 (1938).
176. Negelein, E. and Brömel, H., *Biochem. Zeit.*, **300**, 225 (1939).
177. Neuberg, C. and Hildesheimer, A., *Biochem. Zeit.*, **31**, 170 (1911).
178. Ochoa, S., *Biochem. J.*, **33**, 1262 (1939).
178a. Ochoa, S., Cited in 25.
179. Ochoa, S., *J. Biol. Chem.*, **138**, 751 (1941).
180. Ochoa, S. and Peters, R. A., *Biochem. J.*, **32**, 1501 (1938).
181. Ochoa, S. and Rossiter, R. J., *Biochem. J.*, **33**, 2008 (1939).
182. Ohlmeyer, P. and Ochoa, S., *Biochem. Zeit.*, **293**, 338 (1938).
183. Oppenheimer, C. and Stern, K. C., "Biological Oxidations," p. 213, New York (Nordemann Publishing Co.), 1939.
184. Parsons, H. T., Lease, J. G. and Kelly, E., *J. Biol. Chem.*, **100**, lxxvii (1933).
185. Peters, R. A., *Biochem. J.*, **31**, 2240 (1937).
186. Peters, R. A., *Nature*, **146**, 387 (1940).
187. Peters, R. A. and Thompson, R. H. S., *Biochem. J.*, **28**, 916 (1934).
188. Pittman, M. and Fraser, H. F., *Pub. Health Rep., U.S.P.H.S.*, **55**, 915 (1940).
189. Platt, B. S. and Lu, G. D., *Biochem. J.*, **33**, 1525 (1939).
190. Potter, V. R., *Medicine*, **19**, 441 (1940).
191. Potter, V. R., *J. Biol. Chem.*, **141**, 775 (1941).
192. Pratt, E. F. and Williams, R. J., *J. Gen. Physiol.*, **22**, 637 (1939).
193. Presnell, A. K., *J. Biol. Chem.*, **121**, 5 (1937).
194. Quastel, J. H. and Webley, D. M., *Biochem. J.*, **35**, 192 (1941).
195. Quastel, J. H., Tennenbaum, M. and Wheatley, A. H. M., *Biochem. J.*, **30**, 1668 (1936).
196. Rossiter, R. J., *J. Biol. Chem.*, **135**, 431 (1940).
197. Ruddy, Sister M. V., *Studies Inst. Divi Thomae*, **2**, 165 (1939).
198. Salmon, W. D. and Goodman, J. G., *J. Nutr.*, **13**, 477 (1937).
199. Saunders, F., Dorfman, A. and Koser, S. A., *J. Biol. Chem.*, **138**, 69 (1941).
200. Sealock, R. R., Perkinson, J. D. Jr., and Basinski, D. H., *J. Biol. Chem.*, **140**, 153 (1941).
201. Sherman, W. C. and Elvehjem, C. A., *Biochem. J.*, **30**, 785 (1936).
202. Sherman, W. C. and Elvehjem, C. A., *Am. J. Physiol.*, **117**, 142 (1936).
203. Sherman, W. C. and Elvehjem, C. A., *J. Nutr.*, **12**, 321 (1936).
204. Shils, M. E., Day, H. G. and McCollum, E. V., *J. Biol. Chem.*, **139**, 145 (1941).
205. Shourie, K. L. and Swaminathan, M., *Indian J. Med. Res.*, **27**, 679 (1940).
206. Silverman, M. and Werkman, C. H., *Proc. Soc. Exp. Biol. Med.*, **40**, 369 (1939).
207. Silverman, M. and Werkman, C. H., *J. Biol. Chem.*, **138**, 35 (1941).
208. Simola, P. E., *Biochem. Zeit.*, **254**, 229 (1937).
209. Sinclair, H. M., *Biochem. J.*, **33**, 1816 (1939).
210. Smyth, D. H., *Biochem. J.*, **34**, 1598 (1940).
211. Sober, H. A., Lipton, M. A. and Elvehjem, C. A., *J. Biol. Chem.*, **134**, 605 (1940).
212. Stedman, E. and Stedman, E., *Biochem. J.*, **31**, 817 (1937).
213. Stern, K. G. and Hofer, J. W., *Enzymologia*, **3**, 82 (1937).
214. Stern, K. G. and Melnick, J. L., *J. Biol. Chem.*, **131**, 597 (1939).
215. Stirn, F. E., Arnold, A. and Elvehjem, C. A., *J. Nutr.*, **17**, 485 (1939).
216. Straub, F. B., *Nature*, **141**, 603 (1938).
217. Straub, F. B., *Biochem. J.*, **33**, 787 (1939).
218. Süllmann, H. and Birkhauser, H., *Schweiz. med. Wochschr.*, **69**, 648 (1939).
219. Sure, B., Kik, M. C., Buchanan, K. C., Harrelson, R. T., Jr., and Theis, R. M., *Arkansas Agric. Expt. Sta. Bull.*, **373**, (1939).
220. Sure, B. and DeWitt, J. B., *J. Biol. Chem.*, **126**, 287 (1938).
221. Suzuki, U., Shimamura, T. and Odake, S., *Biochem. Zeit.*, **43**, 89 (1912).
222. Szymanska, R. M. and Funk, C., *Chem. Zelle Gewebe*, **13**, 44 (1926).
223. Tauber, H., *Enzymologia*, **2**, 171 (1937).
224. Tauber, H., *Proc. Soc. Exp. Biol. Med.*, **37**, 541 (1937).
225. Tauber, H., *J. Biol. Chem.*, **125**, 191 (1938).
226. Theorell, H., *Biochem. Zeit.*, **272**, 155 (1934).
227. Theorell, H., *Biochem. Zeit.*, **275**, 37 (1934).
228. Theorell, H., *Biochem. Zeit.*, **275**, 344 (1935).
229. Theorell, H., *Biochem. Zeit.*, **278**, 263 (1935).
230. Theorell, H., *Biochem. Zeit.*, **288**, 317 (1936).
231. Theorell, H., *Biochem. Zeit.*, **290**, 293 (1937).
232. Thompson, K. H. S. and Johnson, R. E., *Biochem. J.*, **29**, 694 (1935).
233. Urban, F. and Peugnet, H. B., *Proc. Roy. Soc. London*, **B125**, 93 (1938).
234. Vilter, R. W., Vilter, S. P. and Spies, T. D., *J. Am. Med. Assoc.*, **112**, 420 (1939).
235. Vilter, S P., Koch, M. B. and Spies, T. D., *J. Lab. Clin. Med.*, **26**, 31 (1940).
236. Vilter, S. P., Vilter, R. W. and Spies, T. D., *Nature*, **144**, 943 (1939).

237. Wald, G., *J. Gen. Physiol.*, **18**, 905 (1935).
238. Warburg, O., in Nord, F. F. and Weidenhagen, R., *Ergebnisse Enzymforsch.*, **7**, 210 (1938).
239. Warburg, O. and Christian, W., *Biochem. Zeit.*, **254**, 438 (1932).
240. Warburg, O. and Christian, W., *Biochem. Zeit.*, **275**, 464 (1935).
241. Warburg, O. and Christian, W., *Biochem. Zeit.*, **285**, 156 (1936).
242. Warburg, O. and Christian, W., *Biochem. Zeit.*, **287**, 291 (1936).
243. Warburg, O. and Christian, W., *Biochem. Zeit.*, **295**, 261 (1938).
244. Warburg, O. and Christian, W., *Biochem. Zeit.*, **298**, 150 (1938).
245. Warburg, O. and Christian, W., *Biochem. Zeit.*, **303**, 40 (1939).
246. Warburg, O., Christian, W. and Griese, A., *Biochem. Zeit.*, **282**, 157 (1935).
247. Weil-Malherbe, H., *Biochem. J.*, **33**, 1997 (1939).
248. Weil-Malherbe, H., *Biochem. J.*, **34**, 980 (1940).
249. Weil-Malherbe, H., *Nature*, **145**, 106 (1940).
250. Westenbrink, H. G. K., *Nutrition Abstracts Rev.*, **5**, 944 (1936).
251. Westenbrink, H. G. K. and Goudsmit, J., *Enzymologia*, **5**, 307 (1938).
252. Westenbrink, H. G. K., van Dorp, D. A., Gruber, N. and Veldman, H., *Enzymologia*, **9**, 73 (1940).
253. Westenbrink, H. G. K. and van Dorp, D. A., *Nature*, **145**, 465 (1940).
254. Westenbrink, H. G. K., Veldman, H., van Dorp, D. A. and Gruber, N., *Enzymologia*, **9**, 91 (1940).
255. Weygand, F. and Birkofer, L., *Z. physiol. Chem.*, **261**, 172 (1939).
256. Whipple, D. V. and Church, C. F., *J. Biol. Chem.*, **109**, 107 (1936).
257. Williams, R. R., *J. Biol. Chem.*, **29**, 495 (1917).
258. Wisansky, W. A., Martin, G. J. and Ansbacher, S., *J. Am. Chem. Soc.*, **63**, 1771 (1941).
259. Wood, H. G., and Werkman, C. H., *J. Biol. Chem.*, **135**, 789 (1940).
260. Zeller, E. A. and Birkhauser, H., *Helv. Chim. Acta*, **23**, 1457 (1940).
261. Zeller, E. A., Schär, B. and Staehlin, S., *Helv. Chim. Acta*, **22**, 837 (1939).

Minerals and Vitamins in Applied Nutrition

LELA E. BOOHER

General Mills, Inc., Minneapolis, Minn.

Man's dependence upon an array of minerals and vitamins in his food supply becomes more intelligible in the light of his evolution in, and from, a material environment wherein a wide range of inorganic matter has unfailingly persisted and wherein widespread synthesis of vitamins had previously developed in the plant kingdom. The preeminence often accorded the proteins as fundamental constituents of all living matter should not detract attention from the fact that the vitamins are indispensable for certain structural developments and functional processes associated with life, or from the fact that the chemical reactions which characterize living matter can proceed only in aqueous solutions of appropriate kinds and concentrations of electrolytes. The minerals, the vitamins, and the hormones regulate or control most, if not all, the vital cellular activities within man's body.

MINERALS IN APPLIED NUTRITION

The elementary composition of the human body has frequently been compared with that of the material environment from which it evolved, namely, the earth's crust, the atmosphere, and the sea. This comparison has revealed both similarities and dissimilarities of composition without having revealed any direct chain of continuity between the two. However, a distinct continuity between the material composition of sea water and that of the body fluids of man was suggested by Bunge in 1889 and by Quinton in 1897, a thesis which has been further developed by Macallum.[1] Without doubt, life began in the sea and in the course of evolutionary development some forms made the advance into fresh water or onto land, there to continue and modify in new environments their progression of evolutionary developments. Macallum develops the concept that whereas proteins are universal constituents of living things, they are not uniform in composition in the wide range of cellular life, and so cannot be regarded as chemical endowments from a remote past. On the other hand, certain inorganic constituents always associated with living matter are present in living cells and their media in so nearly constant proportions (although not in the same proportions in the cells as in their media) that they would

appear to be of paleogenetic character. The inorganic constituents presumed to be in this category include chloride, sulfate, phosphate, carbonate, sodium, potassium, calcium, and magnesium. Macallum advances the view that "the blood plasma of vertebrates and invertebrates with a closed circulatory system is, in its inorganic salts, but a reproduction of the sea water of the remote geological period in which the prototypic representatives of such animal forms first made their appearance." The tissues of ancient oceanic forms of life, it is suggested, had throughout a long period accommodated themselves to the composition of the sea water of that period, so that when, in the course of evolutionary processes, the circulatory system became closed, the composition of the sea water of that time was, with slight modification, reproduced in the vascular fluid and transmitted to descendent forms living in similar or different habitats. Macallum presents several types of data and deduces arguments and reflections in support of this view of the paleogenetic character of certain inorganic constituents of living cells and their media, the original exposition of which readers of this chapter may find of unusual interest.

The vascular fluid is separated from other body fluids by membranes which exhibit varying and highly selective permeabilities. There exist very definite physicochemical equilibria between the constituents of the vascular fluid and those of other body fluids. An inorganic composition of vascular fluid which is fixed within narrow physiologic limits, therefore, presupposes inorganic compositions of other body fluids which likewise vary only within respective narrow limits. Substantial and sustained departures from these narrowly circumscribed compositions of the body fluids are associated with morbid states of the body.

The materials required for the building and maintenance of the inorganic structures of the body are derived from food and water. Against insufficiencies of supplies of essential inorganic nutrients the body can make only limited compensatory adjustments. Excess supplies of one or more inorganic nutrients can ordinarily be disposed of by the body through one or more of the specialized channels of excretion. Intelligent application of the science of nutrition necessitates a thorough understanding of the fundamental principles of physiology and a knowledge of nutritive requirements in terms of common food items.

Sodium, Potassium and Chlorine. A loss of about 10 per cent of the total body water results in impairment of physical well-being, and a loss of about 20 per cent results in death. However, it is not the loss of body weight *per se* but rather the alteration in the concentrations of electrolytes in the body fluids which is the prime factor involved in these adverse phenomena.

On a fat-free basis, the water content of the body shows successive changes with age. Starting in early embryonic life there is a progressive decrease in the proportion of body water throughout the period of growth. From the period of attainment of maturity through middle life, the water content of the body remains almost stationary, but increases, at least in the muscle mass, in old age. The muscle mass of the infant constitutes about 25 per cent of the body weight, that of the adult usually somewhat over 40 per cent.

The changes in water content of the body with age are more readily discerned in the light of the successive changes which take place in the proportion of intercellular to intracellular fluid in the soft tissues, particularly in the muscles. Examples of the analytical procedures and calculations employed in estimating the relative water contents of the intercellular and intracellular fluids of muscle, together with references to previous work on this subject, can be found in a recent report of Lowry and Hastings.[2] The boundaries separating the intercellular fluid from the intracellular fluids are ordinarily considered to be impervious to inorganic cations. The osmotic pressure of the intercellular fluid is due principally to sodium and chloride ions, that of the intracellular fluids principally to potassium and phosphatic ions. The concentrations of the respective major cations and anions in the two fluids are

normally almost constant throughout life except for transitory changes which occur during periods of normal physiologic adjustment. Changes in the sodium, chlorine, potassium and phosphorus contents of the tissue, therefore, reflect definite changes in the proportions of the intercellular and intracellular fluids. During growth, expansion of the muscle mass is accompanied by corresponding decrease in the volume of intercellular or interstitial fluid, which, because it has a lower content of solids than the muscle fibers which it bathes, results in a lower percentage of total tissue water. With attainment of full development and until old age, the distribution of body water remains essentially unchanged. With the advance of old age, the size of the muscles usually diminishes, as a result of either atrophy of disuse or actual loss of muscle fiber, and is attended by an increase in the ratio of interstitial to intracellular fluid.[3] The suggestion is made that the somewhat lowered basal metabolic rate so frequently observed in aged individuals may be due to this increase in volume of interstitial fluid relative to the mass of active cytoplasm. The gross body weight or body surface in such case would, therefore, be associated with a smaller proportion of metabolically active tissue than in the case of a middle-aged adult of identical body weight or body surface. These same alterations in the proportions of interstitial and intracellular fluids in the aged may also not be without functional consequence on muscular capacity and efficiency, since the greater distance between capillaries and muscle fibers would diminish the rate of interchange of metabolites.

McCay and Crowell[4] have reported that animals maintained on a diet adequate except for calorie intake are, by comparison with animals on adequate diets without restriction of calories, especially long-lived and youthful into the period of advanced age. The skeletal muscles of animals retarded in growth through limitation of calorie intake were shown[5] to possess the muscle fiber composition and the relatively higher proportions of interstitial fluid characteristic of chronologically younger animals. It is apparent then that histochemical studies of tissues may reveal very interesting phenomena which remain puzzles when insight is limited to grosser studies of tissue composition.

Gamble[6] has estimated that the body water of an adult is distributed among vascular fluid, interstitial fluid, and intracellular fluid at roughly 5, 15, and 50 per cent, respectively, of the total body weight.

From studies of isolated living tissues, including blood (circulating tissue), it can be shown that the selective permeability of the boundaries between cells and intercellular fluids is the equivalent of semi-permeable membranes and that the ratio of concentration of permeating ions within, to the concentration without the cells conforms to the theory of Donnan equilibria. Studies of isolated tissues furnish abundant evidence that the boundaries between cells and the fluid surrounding them are impervious to cations. For the most part, these boundaries would seem to manifest the same characteristics in the intact tissues of the body, but there are obvious exceptions. During starvation, or following ingestion of acid, potassium ions do cross these boundaries, since both sodium and potassium ions may then be found in the urine in similar amounts. Likewise, potassium ions must cross in the other direction, else muscle fibers could not grow nor could the loss of potassium resulting from exercise be restored by ingestion of potassium.

The osmotic pressure relationships of the tissue fluids in which the inorganic ions play the leading role are interestingly manifested under various abnormal circumstances. The consumption of diets deficient in protein will, if sufficiently prolonged, lead to edema of the tissues. Some investigators insist that edema will not develop under these circumstances unless there is simultaneously an adequate or unrestricted intake of water, a condition which is not likely to be lacking in the realm of human experience.

As a result of the low protein intake, the concentration of proteins in the blood

plasma is reduced, and this is accompanied by progressive accumulation of interstitial fluid of normal osmotic pressure. Edema resulting from such diets was observed endemically in Central Europe during World War I and was commonly referred to as "war edema," "famine edema," or "hunger edema." The diets causing this condition were poor in many nutritional respects. However, attention is focused upon the low-protein aspect of the diets, because reduction of protein concentrations of the blood plasma of animals by mechanical means and the low-protein content of the blood plasma common in nephrosis, accompanied by marked albuminuria, are also associated with edema. The explanation of this phenomenon rests on the basis that reduction of the proteins of the plasma lowers the osmotic pressure of the colloids in the capillaries to a degree incompatible with the normal return of the transudates to the blood vessels. Fluid transudate, therefore, accumulates in the tissues. War edema, or edema caused by mechanical depletion of plasma proteins, can be alleviated by liberal feeding of protein.

The excessive accumulation of fluid in the tissues, commonly called edema, may be any one of three types, namely, hypertonic, hypotonic, or isotonic, depending upon whether the osmotic pressure of the tissue fluids has been increased, decreased, or unaltered. Any factor which changes the osmolar concentration in one compartment of the body fluids causes adjustments to take place in the others, since osmotic pressure is equalized throughout the body.

It is commonly known that if sea-water, which corresponds to about a 3 per cent NaCl solution, or other water of high NaCl content, is used as human drinking water, death will eventually ensue. The high ingestion of NaCl relative to the volume of water leads to diminished secretion of urine and increased osmotic pressure of the body fluids (hypertonic edema) beyond the limits of physiologic tolerance. The converse of this situation is exemplified by copious ingestion of ordinary drinking water following profuse sweating which has been attended with large salt as well as large water loss through the skin. The result is water intoxication (hypotonic edema of the tissues), or in more popular terms "heat cramps" or "miner's cramps." Persons who are subjected to environments which induce copious sweating are advised to replenish their salt and water losses by drinking water to which an appropriate amount of salt has been added, or to ingest salt tablets along with ordinary drinking water. In intensely warm environments, the body does gradually make some degree of adjustment to restrict NaCl losses. This is accomplished by secretion of sweat of lower NaCl concentration. In an environment of high temperature, the individual unaccustomed to very warm environments, therefore, is the more likely to suffer from heat cramps, if the drinking water is not supplemented with ingestion of salt.

A high intake of NaCl accompanied by unrestricted intake of water, such as man occasionally indulges in through consumption of highly salted foods, may also result in temporary edema. Similar results have been observed following consumption of fairly large quantities of isotonic salt solution (0.9 per cent sodium chloride solution). In the first instance, if the water intake has been low, the ready absorption of the NaCl into the vascular system initiates a desire for water which is usually very promptly gratified. In both cases, absorbed NaCl, along with water, rapidly leaves the blood stream and increases the volume of interstitial fluids (isotonic edema). Once dissipated, it is withdrawn from the tissues and eliminated only gradually. By contrast, ingested KCl is normally followed by its very prompt excretion by the kidneys. The interstitial fluid is incapable of storing KCl and the body cells, unless depleted of potassium previously, are impervious to potassium ions. Hence the concentration of KCl in the blood plasma is rapidly built up, with consequent rapid excretion of KCl by the kidneys. As this course of events suggests, potassium salts act as diuretics, and large doses are not well tolerated by the body.

Consumption of excessive quantities of ordinary drinking water by persons other-

wise following their customary habits of living can result in adverse effects. The excessive intake of water causes NaCl to be swept out of the body, and water, if taken in faster than it is eliminated, accumulates in the tissues (as hypotonic fluid), and such symptoms as headache, weakness, nausea, or even more adverse effects may result. These several illustrations serve to show the very distinct and essential function of the kidneys in maintaining physiologic balance among the inorganic constituents of the tissues within the narrow limits compatible with life.

It is obvious from these few illustrations, to which many more could be added, that the osmotic equilibria of and water distribution in the tissues are *largely* controlled by sodium, potassium, chloride, and phosphatic ions in consequence of their being the major ions in the body fluids.

Carnivorous animals, because of their relatively intense muscular activity, require diets rich in potassium. These are provided them by flesh foods, while their lesser need for sodium is provided chiefly by consumption of the body fluids of their prey. Herbivora, in consuming plant food, obtain a diet much richer in potassium than in sodium, but "salt licks" are coveted by them to such extent that, in the wild state, they will risk capture and loss of life to satisfy their desire for salt (NaCl). Man, as a consumer of a combination of foods from plant and animal sources, supplements his diet with relatively large intakes of NaCl, probably to a far greater degree than any normal need could possibly impose. The exercise of individual preference in the amount of salt consumed by man has, so far as we know, created no health hazards; it does promote good digestion and enjoyment of food and thus contributes to the utilization of certain important foods which, were they not seasoned with NaCl, might be refused. The average American adult consumes from 10 to 15 grams of NaCl per day, in addition to ingestion of such quantities of sodium and chlorine as exist naturally in the foods consumed. The average diet also carries an abundance of potassium, a common element in normal food items.

Life was adapted from the first to an environment where abundance of sodium, chlorine, and potassium enabled evolutionary forces to utilize these elements in ionic forms to stabilize fundamental life processes. By reason of the inherent necessity of retaining a composition of body fluids closely related to that of a sea-water of a remote past, natural selection, which rejects every false try and unsuccessful experiment, must have played a large role in the preservation of those non-aquatic forms of animal life which succeeded in establishing satisfaction of their primary needs for these elements. What the whole human race has already instinctively and justifiably applied as an established food habit is a problem of applied nutrition already solved.

Magnesium, Calcium and Phosphorus. Magnesium is found almost universally in living tissue. It is a normal constituent of all animal tissues. In plants, it forms complex ions with protein;[7] it is a fundamental constituent of chlorophyll; it is associated with many weak organic acids such as malic, tartaric, citric and oxalic in plant juices; and is widely distributed in the plant kingdom, especially in the pericarp of cereal grains as a magnesium-calcium salt of inositol phosphoric acid (phytin, so-called).

Magnesium is a constant constituent of bone (about 0.03 per cent of the bone ash), and may exist there as a phosphate, a mixture of phosphate and carbonate, or possibly as a complex salt with calcium phosphate. Whether the magnesium enters into the structural pattern of the deposited mineral of bone, or exists there as an admixture, is not known. Magnesium ions are definitely known to act as catalytic agents for certain enzymatic reactions associated with muscle metabolism, specifically energy transformations in the tissues.

Although they are closely related chemically to calcium, one of the most interesting properties of magnesium ions in living tissue is their counterpoise effect, or physiologic antagonism to calcium ions. McCollum and his associates[8, 9] have de-

scribed the spectacular syndrome of magnesium deficiency in animals which seemingly is related to an imbalance of magnesium and calcium ions in the body tissues. The adverse effects of diets deficient in magnesium are intensified by liberal ingestion of calcium, and *vice versa*. The presence of excessive quantities of magnesium in the tissues, which can be induced by injections of magnesium salts in suitable amounts, will result in anesthesia in man. In animals it has been shown[10] that, depending upon the quantity of magnesium salts injected, anesthesia or even death may ensue. The anesthesia induced in the two species by injection of magnesium salts can be offset by injection of soluble calcium salts. These interrelations of magnesium and calcium ions in living tissues all appear to be related to the irritability of the tissues.

The normal irritability of muscle is controlled and maintained by an interrelation of inorganic ions, including potassium, sodium, calcium, magnesium and hydrogen ions, in the extracellular body fluids. This relationship has been expressed[11] thus:

$$\text{Irritability} \propto \frac{[Na^+] + [K^+]}{[Cu^{++}] + [Mg^{++}] + [H^+]}$$

Hence sufficient alteration in the concentration of any one of these interrelated inorganic ions would impose change upon the irritability of muscle. A high concentration of magnesium ions would decrease muscle irritability, and a low concentration would increase it. These properties of magnesium have been demonstrated experimentally. The complete story of the interrelations of inorganic ions in the control of muscle irritability has not yet been revealed, as one may deduce from consideration of the above pseudo-mathematical expression and the results observed on muscle irritability by imposing a high concentration of calcium ions upon a high concentration of magnesium ions.

Hyperirritability of the neuromuscular system, resulting in tetany characterized by laryngeal spasm, spasm of the muscles of the wrists and ankles, and generalized convulsions, has often been observed by clinicians in such conditions as hypoparathyroidism, rickets, and severe osteomalacia, which are the result of low concentrations of calcium ions in the extracellular fluids, a condition which is revealed in a low calcium concentration in the blood serum. Factors resulting in sufficiently high pH values of the extracellular body fluids (also revealed in high pH values of the blood plasma) can also induce tetany. Some of the adverse affects of ingestion of large amounts of NaCl or KCl, particularly when accompanied by highly restricted water intake, are logically to be expected in view of the ionic interrelationships shown above.

Magnesium salts, when orally ingested, exhibit cathartic and diuretic effects. Following administration of hypertonic solutions of $MgSO_4$ (Epsom salt), for example, water is drawn into the intestine, thus reducing the degree of hypertonicity of the salt solution. Some of the magnesium salt will be absorbed as it moves down the intestinal tract; but as neither magnesium nor sulfate is readily absorbed, catharsis is manifested. In general, the readiness or efficiency of absorption of inorganic ions from the intestines follows the order of the Hofmeister series. Magnesium citrate has a less dehydrating effect than $MgSO_4$ and therefore exhibits milder cathartic effects.

Sherman[12] estimates from surveys of 150 freely chosen dietaries that the average American diet contains 0.34 gram (range 0.17 to 0.53 gram) of magnesium per 3000-calorie intake of ordinary foods. A deficiency of magnesium in the diet, while theoretically possible, is not of practical concern in applied nutrition because of the widespread occurrence of magnesium in common foods and the apparent adequacy of ordinary diets to supply physiologic needs. Balance experiments show wide varia-

tions in the retention of magnesium, and the physiologic requirements cannot be said to have a satisfactory basis to date.

Over 98 per cent of all the calcium in the body is found in the skeleton and teeth, but the small amount in the soft tissues is, nevertheless, of great physiologic importance. The phosphorus content of the bones and teeth accounts for about 80 per cent, that of the muscles for about 10 per cent and that of nervous tissue for about 1 per cent of the total body phosphorus. The phosphorus in the soft tissues occurs in a wide variety of forms, including phosphoproteins and phospholipids, as well as phosphate in ionic form and in combination with numerous organic substances.

In bone, calcium and phosphorus are laid down in an organic matrix, along with a certain proportion of magnesium. Bone also contains smaller proportions of other inorganic elements, notably sodium, potassium, and chlorine. The arrangement and composition of the inorganic matter of bone is apparently that of the apatite minerals, corresponding to $n Ca_3(PO_4)_2 \cdot CaCO_3$, in which n is not less than 2 nor greater than 3. This structure is very similar to the apatite mineral dahllite.[13] As one would expect, therefore, the ratio of calcium to phosphorus in bone ash is practically invariant, regardless of the degree of mineralization of bone. However, in elderly persons the bones do appear to change somewhat in ash composition. The bones of aged persons usually become somewhat brittle and are found to contain larger proportions of $CaCO_3$ than do those of younger persons. This situation is commonly reflected by an increased ratio of bicarbonate to phosphate in the plasma or serum as well.

Fluorine, when deposited in bone, replaces part of the carbonate normally present. Sodium and chlorine are present in the body fluid of bone. As the bones become mineralized to greater extents, their water, sodium, and chlorine contents diminish. Under special circumstances, radium, lead, barium, beryllium, and strontium may also occur in bone, depending largely on the quantities ingested.

Bones are not "fixed" body structures. They are vascular organs and can yield or deposit mineral substance as the occasion demands. Their net-work of bone trabeculae has been shown[14] to increase and decrease quite readily. However, this situation may be altered in aged individuals whose bone trabeculae have been greatly thinned or may have disappeared altogether or been replaced by cysts.[15] In such cases it would seem that mineral deposits lost from the trabeculae might not be capable of restoration, and it is inconceivable that it could be so if the matrix of the trabeculae was disorganized or had disappeared. The teeth, likewise, are not immune to changes in mineral composition, since radium, which presumably can replace some of the calcium in tooth structures, has been found in the dentine of persons exposed to radium poisoning. It is generally conceded, however, that once the teeth have been fully developed, they alter their structures much more slowly than the skeleton. For the promotion of structurally strong teeth, therefore, proper nutrition during their formative period would seem to be unusually important, and this period begins before birth. Calcium, phosphorus, and vitamins A, D, and C have been pointed out as dietary factors particularly important in building good tooth structures.

The difficulties of assessing the absolute and relative amounts of phosphorus in different forms in various parts of the body are very great in view of the fact that phosphorus is present in so many types of compounds, often very closely related chemically and some of them quite labile. The oxidation-reduction reactions in muscle, and the chemistry of muscle contraction, for which glucose and its derivatives supply the energy, proceed by a series of intermediate reactions wherein phosphorylations and dephosphorylations of the metabolites are essential parts of the processes. Recent contributions to our knowledge of muscle physiology form one of the brilliant chapters in biological science.* The use of radioactive phosphorus in

* See paper by O. Meyerhof in this volume. J. A.

compounds introduced into the body, and subsequent study of its transfer to various phosphorus-containing compounds, has also opened up a new field of research in phosphorus metabolism.

Nucleoproteins occur in every cell and cell nucleus in the body. Inorganic phosphate is a constituent of both intracellular fluids (excepting perhaps in the red blood cells) and intercellular fluids and of the body excretions and many of the body secretions as well.

In the discussion of the vitamins, riboflavin, nicotinic acid and thiamine, phosphorus will be seen to form an essential part of several coenzymes. The importance of phosphorus in the nutrition of the body can hardly be over-emphasized, for it is associated with both numerous and intricate reactions of life processes.

Rapid mineralization of the skeleton normally begins two or three months before birth, and under favorable circumstances continues at increasingly rapid rates until about the age of puberty, after which the rate of mineral deposit gradually falls off until, with attainment of full development, the body's needs for calcium and phosphorus are limited to maintenance requirements.

Numerous investigations have been conducted in an attempt to define the calcium and phosphorus requirements of persons of different ages. Even among supposedly normal persons, the needs of any age group show wide variations from individual to individual. It is, therefore, possible to define so-called requirements only in terms of rough approximations. In addition to inherent differences among individuals on the same dietary regimen, many other factors make for wide variations in average calcium and phosphorus requirements of persons in different age groups.

As mentioned earlier, the readiness with which inorganic ions are absorbed from the intestinal tract follows in general the order of the Hofmeister series. Calcium and phosphorus, therefore, are by their very nature less readily absorbed than are, for example, sodium, potassium, and chlorine. Many dietary factors and physiologic conditions affect the degree of utilization of calcium or phosphorus, or both. The presence of oxalates or large amounts of fatty acids in the intestinal tract causes precipitation of insoluble calcium salts and thus renders calcium unavailable for absorption. Large amounts of indigestible residue in the food are also believed to be unfavorable for the absorption of calcium and phosphorus. An excess of calcium or of phosphorus in the food causes corresponding excretions of these elements through the bowel. The presence of vitamins D and C in the diet reacts favorably toward the utilization of calcium and phosphorus. The degree to which the body stores approach being filled with calcium and phosphorus will affect the degree of retention of these elements. Abnormal physiologic conditions, such as acidosis, endocrine imbalance, and diarrhea may impose unusual restrictions on the use of dietary calcium and phosphorus for nutritional purposes, and may in some cases result in removal of large quantities of these inorganic nutrients from skeletal reserves.

It is apparent that the physiologic needs for calcium and phosphorus are not easily established even for normal individuals. In a recent review of the subject, Macy[16] summarizes the situation thus: "Neither the most satisfactory level of calcium intake nor the optimal retention of calcium at any physiologic age or stage of man's development is known."

It is, however, quite obvious that the infant or child requires a liberal intake of both calcium and phosphorus because normal development involves an increase not only in the amount, but also in the percentages of body calcium and phosphorus during growth. Adequate amounts of vitamin D are essential for their utilization. The calcium and phosphorus in milk are exceptionally well utilized and are present in milk in a ratio (Ca/P = 1.3) that corresponds roughly to the ratio in which they are retained by normally developing infants and children. A lower ratio of Ca/P in the diet is satisfactory for the average adult whose needs for calcium, once the skeleton has become fully developed, are less than for phosphorus, which is required

not only for skeletal maintenance but also in much larger quantities than calcium for servicing the needs of large masses of muscle and other soft tissues wherein phosphates play a highly important role.

Most of the efforts directed toward estimating calcium and phosphorus needs of the human body have been based on studies of the relationships between the amounts of these ingested and excreted by groups of individuals in various age groups. Very few studies have been made on the calcium or phosphorus requirements of elderly persons. Obviously excretion of a smaller amount of an inorganic nutrient than that ingested indicates bodily retention of the nutrient; a quantitative balance between the two indicates lack of retention, but no bodily loss of the nutrient; while an excretion exceeding the intake indicates a negative balance or loss of the inorganic nutrient from quantities previously acquired by the body.

The average daily requirement of calcium based on 97 experiments on 24 adults and calculated by arithmetic proportions to a uniform basis of a 70-kg adult was reported by Sherman [17] to be 0.45 g (range 0.27 to 0.82 g). Similarly, on the basis of 95 experiments on 21 adults, Sherman [18] has reported the average phosphorus requirement for 70-kg adults to be 0.88 g, with a range of values from 0.52 to 1.20 g per day. Sherman has until recently recommended an excess of 50 per cent over these average minimal requirements for calcium and phosphorus as a daily allowance, but he now considers that twice the minimal requirement is a better standard in the case of calcium. This increase in recommended daily allowance of calcium is based upon the favorable effects observed in physiologic performance of animals fed on diets providing higher levels of calcium than had previously been considered entirely adequate. According to his interpretation of the observations made by himself and his co-workers, and in the light of his interpretation of researches recorded in the literature, he recommends about 0.9 g of calcium and 1.3 g of phosphorus per 70-kg adult per day as a general guide for planning food selections which will provide good allowances of these nutrients.

Leitch,[19] as a result of critical evaluation of 400 calcium balance studies on women, as recorded in the literature, places the adult maintenance requirement for calcium at 0.55 g per day. She states that there is no adequate basis to show what the extra daily allowance of calcium for adults ought to be. The statistical method of interpretation of the available data as presented in the publications of Leitch is particularly worthy of mention, since it represents a critical approach and a logical means of bringing the numerous reports in the literature to bear on the objective of defining the adult maintenance requirements for calcium.

It is certainly true beyond any doubt that the average adult in this country ingests less than 50 per cent excess of calcium over the reputed average daily maintenance requirements, not to mention 100 per cent excess. It might be expected then that evidences of widespread deficiency of calcium among adults of this country would be observed. Actually, medical reports on such a state of affairs are decidedly lacking. Whether this is because such a state of affairs does not exist, or because the means of detecting less than very gross states of calcium deficiency are inadequate, is a question which at present has no answer.

Many of the calcium balance studies on adults have been of very short duration. The previous nutritional history as regards the calcium intake of the subjects studied has also been passed over or not recorded in the published reports in most cases. If a subject had a previous history of low calcium intake shortly before initiation of a short-term study of calcium balance, the results would not be a measure of maintenance requirement but a measure of this plus what was used to replenish some of the previous losses. In view of the large variations (of the order of 400 per cent) recorded for the daily maintenance requirements of adults, it is obvious that a recommended daily allowance of the order of 0.9 g per day would represent, for a considerable proportion of the subjects thus far studied, a several-fold excess over that

amount which would prevent any calcium loss from the body. It has been generally recognized that substantial excesses of calcium in the diet may unfavorably influence the assimilation of iron, and recent research reports [20] add additional evidence in support of this thesis. On a physiologic basis, it would be difficult to understand how any benefits *specifically* referable to calcium could be presumed to derive from larger intakes of calcium by adults than would support a balance between intake and excretion of this nutrient.

As a practical guide, it can be stated that when children receive about 1½ pints to a quart of milk daily, and average adults a pint of milk daily, with each group consuming such kinds and quantities of other common foods as are required to provide a well-balanced diet, the calcium and phosphorus needs of the body are satisfactorily met.

Infants fed on breast milk or on a suitable cow's milk formula in accordance with the best accepted practices, have no need for additional amounts of calcium or phosphorus. The breast-fed infant utilizes larger proportions of the calcium and phosphorus ingested than does the infant fed on cow's milk, but the total intake of calcium and phosphorus by the latter is such that the skeleton is the more rapidly mineralized. This does not imply that the infant fed on cow's milk is necessarily better-nourished, for there are many other differences between breast milk and modified cow's milk formula in which breast milk is obviously superior, for example, curd-type and ascorbic acid value.

The daily dietary allowances for calcium and phosphorus for persons of different ages and sexes, as well as the allowances for other nutrients, as recommended by the Food and Nutrition Board of the National Research Council,[21] should be consulted by those who are especially interested in contemporary thought along these lines. As stated in the introduction to this report, the term "Recommended Allowances" rather than "Standards" was adopted by the committee to avoid any implication of finality, thus indicating that the recommendations are offered tentatively and until such time as more exact data can be obtained. This document represents the combined judgment of many nutrition authorities and, as such, is a contribution to be highly valued. It is to be hoped that it will be a challenge to research workers in many fields to subject present knowledge of nutritive requirements to closer scrutiny and to devote renewed research effort to the problem of improving our knowledge of the nutritive requirements of man.

The special dietary needs of women during reproductive periods is a highly important subject in the field of applied nutrition, which it is not possible to elaborate on here. A proper application of nutrition during these periods necessitates adequate provision of the kinds and amounts of foods which will meet the ordinary nutritive demands of the maternal tissues and at the same time provide adequate structural materials for building a completely new organism and supplying it with the nutritive essentials required. Recommended daily allowances of specific nutrients for women during reproductive periods are included in the summary prepared by the Food and Nutrition Board.[21]

It might not be amiss to mention here that a large part of the science of nutrition as of today is based upon animal experimentation. Some of this new knowledge is directly applicable to the nutrition of man; some of it is not. In addition, it is well to keep in mind that the overall objective with regard to the feeding of domestic animals used by man as converters of plant food, is not the same objective as desired in the nutrition of man. It is generally accredited that high rates of growth and development, body size, and body weight in practices of animal husbandry are indications of dietary superiority, but the objective in animal husbandry is one of sheer exploitation with but a single purpose in view. We may well question whether the height and dimensions of man are reliable measures of his nutritive status. Certainly we have no evidence at hand to indicate that a relatively small man is, in any fundamentally important aspect of life, inferior to the six-foot tall, athletic type. Further-

more, there are indications that a certain degree of abstemiousness with regard to food consumption is highly favorable to longevity;[4] certainly mortality statistics of the life insurance companies bear witness that over-weight in the human race is distinctly unfavorable to a long life span. We are now in an era where great enthusiasm attaches to the protective food elements and their possibilities in promoting the welfare of mankind. There seems almost to be a general impression that "the more vitamins and minerals you consume, the better," an impression which has no scientific basis. This over-enthusiasm is already beginning to be subject to the scrutiny of more thorough-going researches, and a tone of more moderation with regard to vitamin and mineral consumption is almost certain to be not too far off. Nutrition is a young, a very young science, and it has opened up so many fields for the prevention of obvious malnutrition that the temptation, fanned by a wave of popular sentiment, to extrapolate beyond reasonable limits will, perhaps, only wear itself out as time and further researches bring more substantial evidence and more seasoned judgment to take its place.

Iron and Copper. Iron exists in largest functioning quantities in the body as a constituent of blood and muscle hemoglobin. Iron is also a constituent of the tissue cytochromes (concerned with certain steps in the process of hydrogen transport from tissue metabolities to molecular oxygen), and of the chromatin material of all cell nuclei.

Hemoglobin is synthesized by the body, a process in which traces of copper acting in the capacity of catalytic agents are required. Aside from this function, little is known of the physiologic need for copper, although it is found in traces to be quite widespread in the body tissues. Anemia in adults, which results from dietary deficiency of iron, seemingly responds to administration of iron without supplementary feeding of copper. This may be due to the fact that the iron salts used in the treatment of the anemia are sufficiently contaminated with copper or that copper is adequately provided for by ordinary diets. In animals, an anemia requiring supplemental copper in addition to iron has been produced as a result of feeding a diet composed exclusively of cow's milk,[22] so it is possible that an analogous situation could develop in infants too long restricted to a diet composed exclusively of this same food.

Iron, unlike the other inorganic nutrients discussed here, and also unlike the vitamins, is not used up nor excreted in appreciable amounts from the body once it has been absorbed. Iron is conserved by the body with great tenacity and is used over and over again. Brownish-yellow masses, believed to be a colloidal form of ferric hydroxide, so-called hemosiderin, have been noted especially in the liver, spleen, and kidneys, and are supposedly the endogenous sources of iron which the body conserves for reuse.

The infant at birth possesses a reserve supply of iron. Until rather recently, it was supposed that this reserve iron was stored in the livers of new born infants.[23] It is now believed that this reserve source of iron exists in the blood which in the newborn carries around 20 mg of hemoglobin per 100 ml in comparison with around 15 mg of hemoglobin per 100 ml of blood in adult men. Adult women usually have about 10 per cent less hemoglobin per 100 ml of blood than adult men. However, women also have correspondingly fewer red blood cells, so that the concentration of hemoglobin in the red blood cells of the two sexes is approximately the same.

Some concern has developed in recent years over the fact that quite a number of otherwise normal persons have low values for blood hemoglobin, in terms of standards established as norms some years ago. However, it would seem that this situation might be due to other causes than a dietary deficiency of iron, and might even exist in individuals possessing large stores of iron in their bodies. The body seems to regulate its hemoglobin formation to meet its needs for this oxygen-carrying compound. It is well known that the hemoglobin level and red blood cell count of the blood are increased as a natural physiologic consequence of living at a high altitude.

The body compensates for the lower oxygen content of the atmosphere by increasing its facilities for greater aeration. Thus there is a physiologic adjustment in response to need under changed conditions. The mechanical age in which we live has imposed a life of considerably lessened physical activity on many persons than was the case a generation or so ago. Lessened needs for high concentrations of hemoglobin may have brought about a physiologic response of lessened production of hemoglobin. The normal adult may find that the simplest means of calling forth a physiologic response to increased production of blood hemoglobin is to engage in active and regular exercise, in other words, to create a need for higher concentrations of blood hemoglobin.

During growth, there is a need for liberal intake of iron to supply the requirements for hemoglobin formation in an expanding blood volume, a condition accelerated during puberty when growth is normally very rapid. Women, because of periodic blood losses, and children because of expansion of blood volume, require larger intakes of iron than men.

The availability of iron salts for nutritional purposes in large measure depends upon their solubilities. The absorption of iron occurs primarily in the uppermost part of the intestinal tract adjacent to the pylorus. With neutralization of the acidity of the gastric juice in this region, the facilities for absorption of iron are diminished. The availability of elemental iron (ferrum reductum) would seem to depend upon the amount that goes into solution as a result of action of the hydrochloric acid of the gastric juice. It is generally assumed that any ferric ions are reduced to the ferrous state prior to absorption. Large amounts of phosphate in the gastrointestinal tract are considered to be unfavorable for the absorption of iron by reason of the formation of insoluble iron phosphate. The iron in ordinary foods varies in degree of availability, and for this reason a good margin of safety is advisable in recommendations of dietary supplies. Egg yolk, meats, whole grain and enriched cereal products, and green leafy vegetables are the most important sources of iron in the average diet. Milk is an outstandingly poor source of iron. For further discussion of the subject of iron in nutrition, readers are referred to a recent article by Heath.[24] In conclusion, it is well to mention that anemia results from a variety of causes other than a dietary deficiency of iron, and discussions of these, therefore, do not properly belong in this brief summary of iron in relation to applied nutrition.

Other Minerals. Although iodine occurs in minute traces elsewhere in the body, particular significance attaches to the fact that a large proportion of the body iodine is found in the thyroid gland. Iodine is a constituent of thyroxin, a component of the secretion of the thyroid gland, which in large measure and in conjunction with other internal secretions of the body, controls the rate of energy metabolism of the body. Thyroxin was first isolated from thyroid tissues by Kendall[25] in 1914, and has since been synthesized.[26] Kolnitz and Remington[27] have reported as a result of analyses of the thyroid glands of 150 persons that the average iodine content of the thyroid gland is 8.8 mg.

When the amounts of iodine supplied by food and drinking water are insufficient, the thyroid gland becomes enlarged, resulting in what is called simple goiter. In regions where the soil and water are iodine-deficient, simple goiter has been found to be endemic unless preventive measures are instituted. Iodine of the soil may be derived either from geologic deposition of iodides or from sea-salt, which is relatively rich in iodide. Sea-spray is blown variable distances inland by the wind and may eventually be deposited on the soil as residue from evaporation of the spray, or carried down by rain to the soil and soil water. Regions where geologic deposits of iodine are scant and which are too remote from the seashore to gain iodine from sea-spray have iodine-poor soils and water. In this country, the soil and water in the region of the Great Lakes and large areas in the northwest are distinctly iodine-poor.

Sea foods, including lobsters, oysters, crabs, shrimp, and marine fish are com-

paratively rich in iodine. Garden vegetables grown on iodine-rich soils rank next as natural sources of iodine. The use of iodized salt commonly containing one part of sodium (or potassium) iodide to 5000 parts of sodium chloride is now quite widely and successfully used as a preventive against iodine deficiency. Although wide acceptance of iodized salt has not led to any adverse effects, the Council on Foods of the American Medical Association has recommended [28] that individuals over 30 years of age, who have a swelling at the base of the neck, consult a competent physician before using iodized salt. This advice is issued as a caution to those who may have an adenomatous type of goiter which might develop into toxic goiter following ingestion of iodides.

Sulfur is a nutritionally essential element and is a constituent of every living cell. The body tissues contain small amounts of inorganic sulfates. In addition, sulfur is found to occur in body tissues in combination with carbohydrate, with lipids (sulfatides of brain and nerve tissues), and as an integral part of the sulfur-containing amino acids, methionine and cystine, of thiamine (vitamin B_1), of the melanin pigments of the hair and skin, of glutathione (a coenzyme in cellular oxidation-reduction systems) and of the protein insulin, contained in the internal secretion of the pancreas and concerned with the metabolism of glucose.

Although small percentages of inorganic sulfate are contained in foods and frequently in drinking water, by far the greater intake of sulfur is provided in the form of sulfur-containing proteins. By utilizing inorganic sulfates in the soil, plants are able to synthesize the sulfur-containing amino-acids required by man. Rose and his co-workers [29] have shown that methionine is an essential amino acid, whereas cystine is not. Apparently methionine alone can supply all the nutritional needs for sulfur-containing amino acids; nevertheless, cystine added to a diet containing a low level of methionine will supplement the growth-promoting properties of methionine. It would seem that cystine could take the place of methionine for some but not all of the functions for which sulfur-containing amino acids are essential.

In applied nutrition, the sulfur requirements are not a problem at all, provided the protein intake is adequate in supplying the kinds and quantities of amino acids that are essential. The sulfur requirements of the body are satisfied when the requirements for sulfur-containing amino acids are satisfied.

Traces of many other inorganic elements have been found in human tissues than those just discussed, some of them would seem to be essential for nutrition and others perhaps present only as casual contaminants. However, it does not *now* appear that many of the trace elements present problems in applied nutrition. This is not a certainty in view of the meager amount of information available on their functional properties, distribution in foods, and quantities that may be required for normal human nutrition. For further information with regard to the trace elements in nutrition, together with a comprehensive list of references, the reader is referred to a recent summary by Shils and McCollum.[30]

Maintenance of Physiologic Electroneutrality

Obviously all of the ionic constituents in the body contribute to the maintenance of its electroneutrality. The most prominent ionic constituents of the body include sodium, potassium, chloride, phosphatic ions, bicarbonate ions, magnesium, calcium, hydrogen, and sulfate. In addition to these, the small percentages of the so-called trace elements contribute in small degree and in proportion as they occur in ionic form. In view of the uneven distribution of the various ions in the body organs and fluids, it is apparent that electroneutrality of the different parts of the body is maintained by different mechanisms or combinations of electropositive and electronegative ions.

Acid- and Base-forming Foods. The normal acidity of the blood plasma, which serves as an indicator of the acid-base status of the body tissues, is about pH 7.40

(range about pH 7.35 to 7.43). This slight alkalinity is maintained by an intricate system of buffers and through appropriate exercise of excretory functions. Except perhaps in certain diseased states of the body, the acid-base relationships of ordinary dietaries do not affect the acid-base relationships of the body tissues. The pH of the body fluids is maintained with great tenacity, largely through the medium of the lungs which give off CO_2 (the anhydride of H_2CO_3), through excretions from the intestines, and through the medium of urinary excretions, which may be acid, neutral, or alkaline in reaction. The inorganic composition of the food is the principal determinant of the kinds and amounts of inorganic material excreted and of the paths by which elimination from the body is accomplished. In general, the less soluble forms of inorganic matter are the ones most prominent in excretions from the intestines, which, in contrast to the kidney, can excrete inorganic matter in the solid phase. The kidney has the capacity to form NH_4^+ at the expense of urea, thus making it possible to neutralize large excesses of acidogenic ions for excretion without the necessity of sacrificing mineral cations. The ammonia formation is thus relatively large following ingestion of acid or acid-producing foods, and may be small or negligible in alkaline urines.

As eaten, many fruits have an acid reaction due in most cases to potassium acid salts and free organic acids. Whether such foods are potentially acidic or potentially basic in metabolism depends upon whether or not the organic acid radicals are oxidized by the body. A rough estimate of the acid-forming or base-forming values of various foods involves the amounts and proportions of inorganic elements contained in them. Such estimates have been made by Sherman and Gettler [81] for a considerable number of ordinary food items. In their calculations, the amounts of chlorine, phosphorus, and sulfur were expressed as the equivalent volume of normal acids; the amounts of calcium, magnesium, potassium, and sodium were expressed as the equivalent volume of normal alkali; and the excess of acid or base was obtained by difference. These calculations were based upon the assumption that the sulfur in the foods was completely oxidized; the phosphorus was calculated as the equivalent of a dibasic acid; and no account was taken of such organic acids as might not be oxidized in the body. Nevertheless, for most foods the values obtained do furnish a rough approximation of the acid or alkali values and have shown a fairly good relationship to the effects of the foods on urinary acidity and ammonia excretion. On the basis of such analyses and calculations, meat, eggs, and cereals are shown to be predominately acid-forming foods, while fruits, vegetables, and milk are alkaline-forming in metabolism.

Some years ago, enthusiasm ran high in favor of consumption of foods which would yield an alkaline ash. It was argued that since milk was the natural food for the young of all higher animals and yielded an alkaline ash, a predominately alkaline-forming diet must be desirable, especially for infants and children. Added to that was the argument that the kidneys already were required to excrete an excess of inorganic anions, and acid-ash food put a further strain upon the mechanisms of excretion. However, it has never been possible to demonstrate experimentally or to observe clinically that the acid- or alkaline-ash value of diets composed of ordinary foods is either beneficial or harmful. The question of acid-base relations in the freely-chosen diets of normal individuals no longer commands interest. This summary of the present attitude on the acid-base relation in food is not, of course, to be extended indiscriminately to all individuals, for it is recognized that in certain types of illness, special and individual consideration of this question may be of considerable importance.

Vitamins in Applied Nutrition

Vitamins are essential food nutrients. In general, they are regulators for biochemical changes that underlie vital cellular physiology and/or specific tissue mor-

phology. As parts of biologic catalytic systems, they are concerned with energy transformations in the tissues and/or with the orderly building, maintenance and normal functioning of specific tissue structures. The widespread syntheses of the organic compounds, collectively called vitamins, in the plant kingdom have enabled man to be steadily nurtured in part by them and to evolve with a singular dependence upon them.

The B Vitamins. In Fig. 1 (p. 723) are depicted the chemical relationships of thiamine, riboflavin, and nicotinamide in the roles of reversible hydrogen carriers in catalytic dehydrogenating systems. In the chemical combinations indicated, these vitamins take part in many substrate-enzyme-coenzyme systems of fundamental importance in energy transformations in living tissue.

The yeast enzyme, carboxylase, which catalyzes the decarboxylation of pyruvic acid, was found[32] to require activation by a thiamine-containing coenzyme (cocarboxylase), which is separable from yeast by washing with alkaline phosphate. By combining the washed yeast and the washing fluid, the capacity to decarboxylate pyruvic acid was restored. It was furthermore noted that the activity of the enzyme-coenzyme system was increased by addition of magnesium salts ($MgCl_2$). Lohmann and Schuster[33] isolated the coenzyme and found it to be a pyrophosphoric ester of thiamine (thiamine pyrophosphate or diphospho-thiamine). As a result of acid hydrolysis or on subjection to the action of kidney phosphatase (alkaline phosphatase), one molecule of phosphoric acid is readily split off, the resulting thiamine monophosphate being without cocarboxylase activity. The second molecule of phosphoric acid is removed by prolonged incubation with kidney phosphatase, the resulting thiamine being without cocarboxylase activity. Alkaline hydrolysis of the thiamine pyrophosphate results in splitting off both molecules of phosphoric acid as inorganic pyrophosphate. The synthesis of thiamine pyrophosphate has been achieved both by enzymatic[34, 35] and by chemical means.[36, 37]

It was earlier considered that thiamine pyrophosphate might act as a hydrogen carrier by virtue of alternate oxidation and reduction of the quaternary thiazole nitrogen atom. However, reduction of thiamine or of thiamine pyrophosphate with hydrogen activated with platinum or palladium has yielded biologically inactive products.[38] It now appears more likely that the compound oscillates between its reduced state, represented by the usual structure for the vitamin itself and an oxidized state, a disulfide combination, involving two molecules of the vitamin.[39] Its reduced and oxidized states, therefore, are analogous to reduced and oxidized glutathione. Thiamine-disulfide is biologically active.

Kubowitz and Lüttgens[40] have shown that carboxylase consists of three components; protein (one gram molecular weight), thiamine pyrophosphate (one gram molecular weight), and magnesium (one gram atomic weight). In weak acid solution, pH 5 to 6, the three components are very firmly united. At pH 8.1, the enzyme is dissociated, but under suitable conditions, can be reunited with restoration of 85 per cent of its original carboxylase activity.

All the animal tissues examined (brain, skeletal muscle, heart, and liver), contain thiamine, very largely in the form of thiamine pyrophosphate.[41] In thiamine-deficient animals, the concentration of thiamine pyrophosphate in the tissues is markedly diminished; this is accompanied by an accumulation of pyruvate in the tissues, as first noted by Peters.[42] The administration of thiamine to thiamine-deficient animals brings about a prompt increase in the concentration of thiamine pyrophosphate in the tissues.

There are indications that thiamine pyrophosphate will not readily cross the cell boundaries of body tissues.[43] Thiamine absorbed into the blood stream is largely taken up by the liver and kidney, converted into, and stored as thiamine pyrophosphate.[44] These same tissues can also dephosphorylate the coenzyme[45, 46] and thus supply free thiamine to the blood stream, whence it can be carried to other body

tissues or be excreted in the urine as circumstances dictate. By such a mechanism, free thiamine can penetrate the cell boundaries and be slowly rephosphorylated within the cells.[41]

In animal tissues, thiamine pyrophosphate is primarily concerned with decarboxylation of α-keto acids. In contrast to yeast, which decarboxylates pyruvic acid to yield acetaldehyde and carbon dioxide, in animal tissues the decarboxylation is usually accompanied by dehydrogenation. Whether both the decarboxylation and the dehydrogenation of pyruvic acid in animal tissues are to be ascribed to thiamine pyrophosphate, or whether dehydrogenation is carried out through the medium of a second vitamin-containing coenzyme (diphosphopyridine nucleotide) has not been settled, but it would appear that the second mechanism is the more probable, so far as can be judged at present.

Riboflavin in the role of coenzyme acts by reversibly accepting or releasing two atoms of hydrogen, a property also shown by the free flavin and observable in the reversible shift between the yellow color of the oxidized form and the colorless product representing the dihydro form of the vitamin.[47, 48, 49] The redox potential of an equimolar mixture of riboflavin and dihydro-riboflavin is −0.21 volt (pH 7.0, and referred to a normal hydrogen electrode). In the role of coenzyme, riboflavin participates in many oxidation-reduction reactions in animal tissues, its part being played in each case by virtue of its capacity to transport hydrogen. These reactions are catalyzed by specific enzymes. Nicotinic acid amide, another vitamin of the B-group, is also known to act in several enzyme systems concerned in carbohydrate and protein metabolism of animal tissues.[50, 51] And again the function of the vitamin is that of a hydrogen carrier, although this may not be the only means by which it acts in biologic processes. The coenzyme combination of this vitamin, diphosphopyridine nucleotide, forms an internal salt, the positive charge on the N of the pyridine structure balancing a negative charge on one of the phosphate residues, as can be seen from the structure of this coenzyme. The isoelectric point of this coenzyme is at pH 3. The hydrogenation sets in at the N = C double bond in the pyridine nucleus, while the nitrogen becomes trivalent; one hydrogen becomes ionized. With reduction of the pyridine nucleotides the ultraviolet absorption band at 260 mμ, referable to the adenine structures, is observed to sink slightly and a strong band appears with a maximum of 340 mμ. Furthermore, the dihydro derivatives of the pyridine nucleotides fluoresce white in ultraviolet light. These characteristics make it possible to measure the course of the reduction in systems comprised of substrate-enzyme-pyridine nucleotide.[50]

In seeking an over-all picture of the role of B-vitamins in chemical reactions of tissues, it should be pointed out that the energy released from oxidations of foodstuffs is not all immediately set free. Energy is made available to the cell for a variety of syntheses for which a source of energy is required. Oxidative energy, for example, is transformed into so-called "phosphate bond energy" in such forms as adenosine triphosphate and creatine phosphate which constitute the immediate source of mechanical energy of muscle, and of course food energy in excess of the body's needs is largely stored as body fat. The Warburg school conceives the chemical entities, sketched in Fig. 1 with B-vitamin nuclei in their structures, as *active groups* of specific proteins in combination with which they comprise distinct enzymes. This conception regards the large array of dehydrogenases as conjugated proteins analogous to phosphoproteins, nucleo-proteins, etc. According to this concept cocarboxylase or thiamine pyrophosphate, for example, would be regarded as the active or prosthetic group of a specific protein which catalyzes the decarboxylation and carboxylation of aliphatic α-keto-acids such as pyruvic acid; the protein, prosthetic group, and magnesium together comprising the complete enzyme. The structure of thiamine pyrophosphate is shown in Fig. 2. However, in the light of recent developments it would seem more appropriate to consider chemical combina-

tions in which the B vitamins act in many cellular catalytic reactions as *distinct parts* of substrate-enzyme-coenzyme systems. The structures of the coenzymes, flavin-adenine-dinucleotide and diphosphopyridine nucleotide are shown in Figs. 3 and 4 respectively. Nicotinamide also constitutes a part of at least one other coenzyme, triphosphopyridine nucleotide, which differs from the structure shown in Fig. 4 by inclusion of a third phosphoric acid residue. In addition to their role as hydrogen

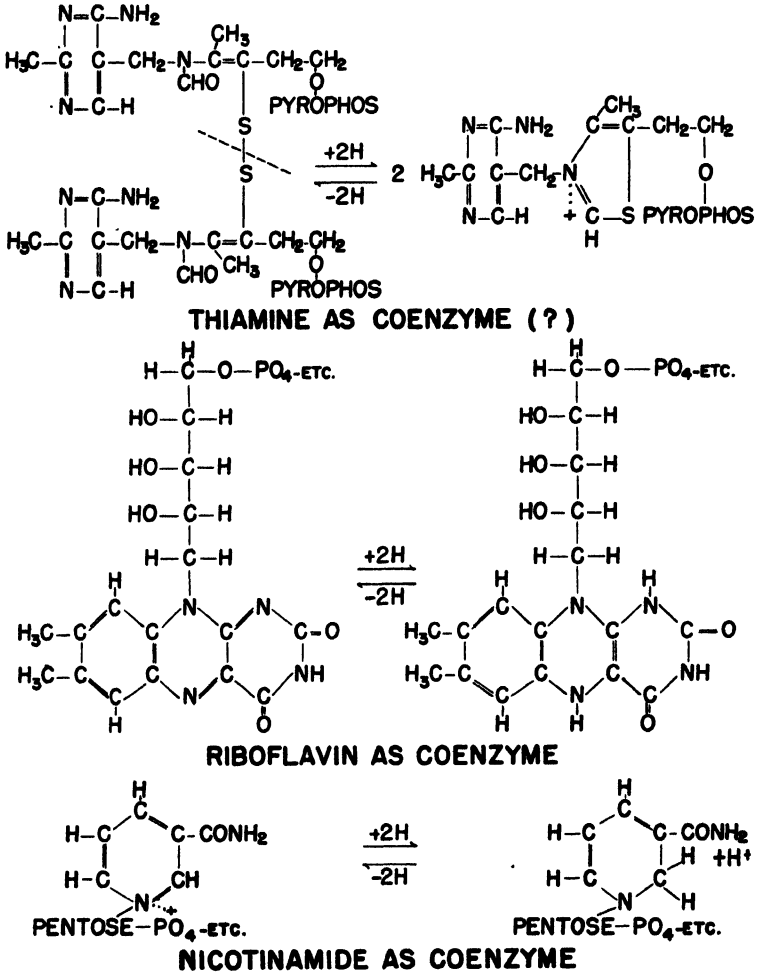


FIGURE 1. Thiamine, riboflavin and nicotinamide in the roles of coenzymes (hydrogen carriers).

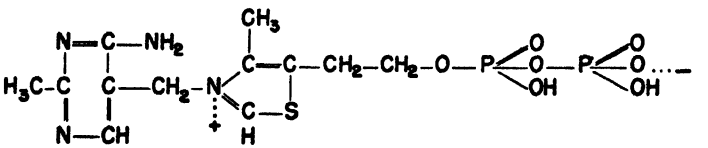


FIGURE 2. Thiamine pyrophosphate (cocarboxylase or diphosphothiamine).

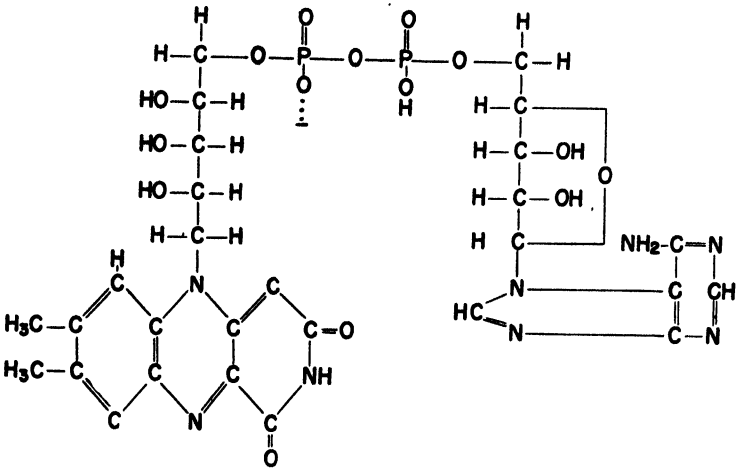


FIGURE 3. Flavin-adenine-dinucleotide.

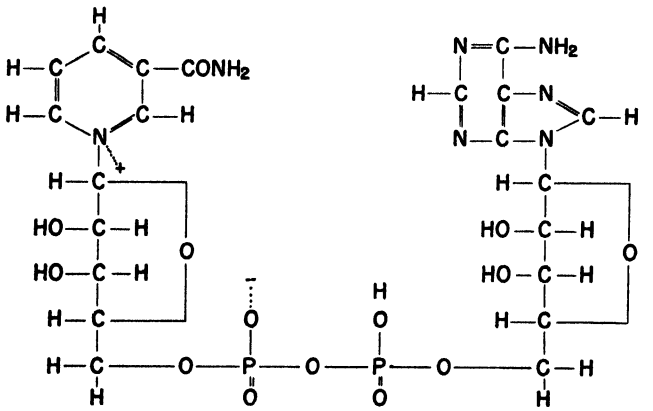


FIGURE 4. Diphosphopyridine nucleotide (coenzyme I).

carriers, the pyridine nucleotides are also known to take part in some of the phosphorylation reactions of tissues.[51]

Certain dehydrogenations conducted *in vitro* with enzymes and substrates, which in biological environments require the presence of the coenzyme triphosphopyridine nucleotide, *can* proceed without an intermediary coenzyme to transfer the hydrogen from substrate to acceptor.[52] So the hydrogen-carrying coenzyme is necessary only when certain acceptors are used. And by the same token it may be concluded that nicotinamide is essential for dehydrogenations in living cells because of the particular hydrogen acceptors available to receive the hydrogen from the oxidized substrate. This may presage an observation that not all living organisms incapable of synthesizing nicotinamide will be found to require nicotinamide as a preformed nutrient. It would also seem that this observation is opposed to the concept that coenzymes, such as the pyridine nucleotides, are actually prosthetic groups of dehydrogenases, since these dehydrogenases have now been observed to catalyze certain dehydrogenations in the absence of coenzyme, which they would be unable to do if the coenzyme structures were essential for the activity of the enzymes. The simplest picture of

the substrate-enzyme-coenzyme relationships of such dehydrogenating systems as involve the B vitamins as essential parts of hydrogen-carrying coenzymes may be visualized as follows: The substrate forms a highly dissociable compound with a very specific group in the enzyme whereby the substrate is "activated," that is, made more tractable to giving up some of its hydrogen. The coenzyme also combines with the enzyme, likewise forming a highly dissociable compound. Two hydrogen atoms then pass from the combined substrate to the combined coenzyme. The dehydrogenated substrate and the hydrogenated coenzyme then leave the enzyme, the former usually to be further dehydrogenated and the latter to be reoxidized by an appropriate redox system. The enzyme meanwhile is free to combine with further molecules of substrate and coenzyme, and the process is repeated. After a one-step oxidation of a given substrate by a hydrogen-carrying coenzyme in the presence of the enzyme specific for the reaction, the resulting oxidation product can act as a substrate in another substrate-enzyme-coenzyme system, and so on in succession to complete oxidation, the energy of food being thus released through a series of catalytic oxidations. Simultaneously the hydrogen taken up by the coenzymes is transported from one cellular redox system to another, until finally it can be transferred to molecular oxygen brought to the tissues by oxygenated blood. This hydrogen transport system is catalyzed by specific enzymes.

Pyruvic acid, which appears early in the oxidative breakdown of carbohydrates in living tissues, is not only an intermediate in the metabolism of carbohydrate, but also of protein and probably of fat as well. The flavin-containing coenzyme, flavin-adenine-dinucleotide, together with a specific enzyme, can catalyze the oxidative deamination of *d*-amino acids, that of *d*-alanine yielding pyruvic acid.[58] It is generally assumed that the biologically *l*-amino acids *in vivo* also form pyruvate in the course of their metabolism. Pyridoxine has recently been reported to be associated with the utilization of unsaturated fatty acids, [54, 55, 56] and to be essential to animals for the production of fat from protein,[57] the latter involving, it is assumed, preliminary synthesis of carbohydrate or carbohydrate derivatives such as pyruvic acid, which can subsequently be converted into fat. The conversion to fat also requires a supply of thiamine. It would appear, therefore, that pyruvic acid occupies a central field position in most energy transformations involved in living tissues.

In Fig. 5 is shown the so-called "citric acid cycle" of Krebs,[58] which is quite widely accepted as representing the main path of carbohydrate oxidation in various body tissues.* The cycle, as represented, rests on excellent experimental support and represents a cyclic chain of chemical transformations summarizing the main points of current knowledge concerning pyruvate oxidation in various animal tissues.[58] When this scheme first appeared, the product fed into the cycle was termed "triose" or "carbohydrate derivative" and later was replaced by pyruvate, subject to further inquiry. Since pyruvate has been shown to react in preference to other substances, and since such other substances as may react here can be oxidized to pyruvate in muscle tissue, we may visualize the whole process as hinged upon pyruvate oxidation. Pyruvate then feeds into this cycle where, in the presence of oxygen, it condenses with oxalo-acetate to form citrate, or possibly to form *cis*-aconitate, which is broken down by a series of enzymatic reactions to oxaloacetic acid. The complete cycle effects a total oxidation of pyruvic acid to carbon dioxide and water. Considering the chain of reactions as depicted by the succession shown by reading in a clockwise direction, it is clear that the cycle involves the transfer of six pairs of hydrogen atoms; thus the complete cycle effects a total oxidation of pyruvic acid to water and carbon dioxide.

In isolated systems and in work with various tissue preparations it has been found that the coenzymes, flavin-adenine-dinucleotide, diphosphopyridine nucleotide and triphosphopyridine nucleotide (the first containing riboflavin and the latter two nico-

* See Addendum at the end of this paper.

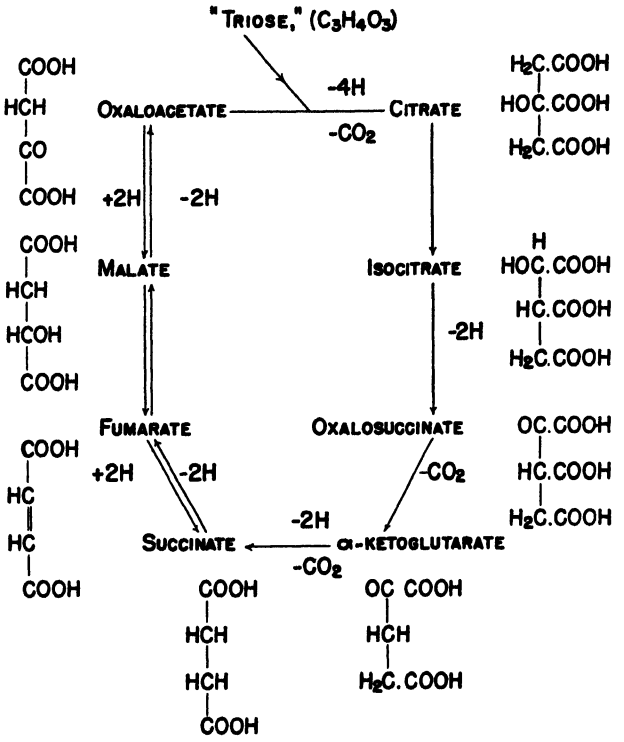


FIGURE 5. Kreb's citric acid cycle.

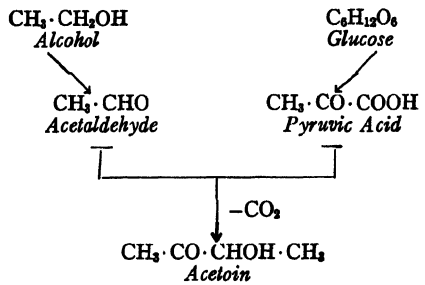
tinamide structures) participate in reactions shown as parts of the so-called Krebs cycle. And, as mentioned earlier, thiamine participates in controlling the rate of pyruvate utilization by the tissue. Although the chemistry of the B-vitamins as they relate to the intimate energy metabolism taking place in tissue cells is not entirely in the category of a system of successive, well established chemical reactions which will elucidate all of the intricate detail involved, great progress has nevertheless been made in a better understanding of cellular energy transformations.

The specific functions of the more recently discovered members of the B-group of vitamins are only beginning to be investigated. In view of the close interlocking and interdependence of the reactions catalyzed by the enzyme-coenzyme systems of which the B-vitamins are constituent parts, it is easy to comprehend why the signs of respective deficiencies of these in man are *not always* clearly circumscribed. Probably in the ordinary course of events most clinical manifestations ascribed to a deficiency of any one of the group of B-vitamins represent a multiple-deficiency state. In the final analysis, it appears that the deficiency of one B-vitamin may block off some of the chain of successive reactions in which other vitamins of this group act as coenzymes, and that restoration of the specific shortage might operate as both a primary and secondary corrective mechanism. The *terminal* pathology of the various B-vitamin deficiency states seems to be fundamentally the same and dependent upon disturbed oxidation in the tissues. Physical weakness or tiredness, as one would expect with disturbed tissue respiration, is evident in any B-vitamin deficiency state.

The induced B-vitamin deficiency states in man and animals, under controlled conditions, may be more an illusion than a reality when the deficiency of one may

conceivably block the normal functioning of another. Often fatigability and a very red mouth and tongue are the only evidence indicative of some B-vitamin deficiency state, rapidly brought on as a result of disease which interferes with normal assimilation of one or more of these vitamins. In such cases, the very red mouth and tongue return to normal tints fairly promptly when the missing vitamin or vitamins are supplied so that the body can use the oxygen of arterial blood; otherwise the degree of oxygen saturation in venous blood approaches that of arterial blood, a fact which explains the redness of the tissues mentioned.

The frequent occurrence of varying degrees of thiamine and nicotinic acid deficiency states in chronic alcohol addicts has been more or less popularized in the public mind. The explanation most commonly offered is that the food consumed by the chronic alcohol addict is usually inadequate in quantity and decidedly deficient in protective food values, prominently so in the B-vitamins. The metabolism of alcohol is not known in its entirety, but according to recent information,[59] the first two steps are presumably as follows:



The enzyme, alcohol dehydrogenase, which is concerned with the first step in oxidation of alcohol (to acetaldehyde) is found almost exclusively in the liver. This statement is based on alcohol perfusion experiments and studies on liver poisoning. In this first step, one of the coenzymes containing nicotinamide is believed to act as the hydrogen carrier. The second step, which involves the condensation of acetaldehyde and pyruvic acid, requires thiamine pyrophosphate.[59] Therefore, the chronic alcohol addict is burning his candle, so to speak, at both ends—his diet is deficient in thiamine and nicotinic acid values; the principal source of his calories (alcoholic beverages) carry little or none of these two vitamins, and yet both of these vitamins are just as essential for alcohol metabolism apparently as for the metabolism of carbohydrates from food. It might be of interest to point out here too that the ingestion of alcohol increases the blood sugar level (mobilized from glycogen in the liver if necessary) as a corollary to the mode of its metabolism; and its oxidation or disappearance, therefore, is delayed if the glycogen stores in the liver are relatively near exhaustion. If the glycogen stores of the liver have been exhausted, a more rapid disappearance of alcohol is favored when it is ingested along with some readily available form of carbohydrate such as sugars.

The respective signs of thiamine, riboflavin, and nicotinic acid deficiencies and the pathology of the body tissues of man and other animals suffering from a deficiency of one or more of these vitamins will not be discussed further here except to mention that they are but the manifestations or terminal results of default in the biochemical reactions catalyzed by the respective coenzymes depicted in Fig. 1.

The B-vitamins seem to be universally important wherever energy transformation in living matter takes place or is provided for by nature. The vitamins in dormant seeds include, in most cases at least, all those mentioned in the previous discussion. As new plants spring from seeds, these vitamins are undoubtedly a prime factor in the energy transformations involved. It is also apparent that plants must synthe-

size these vitamins in parts most intimately concerned in their respiration, namely, the leaves, and from this synthesis presumably the seeds derive their vitamins. In animals those tissues, where oxidative processes are most in evidence, are found to have the highest relative concentrations of B-vitamins, in descending order the liver, heart, kidney, and muscles. The muscle of swine, by comparison with that of other species, is remarkably rich in thiamine. One could easily surmise, and there is some evidence to the point, that this may be associated with the tremendous capacity of swine to transform carbohydrate into fat.

Since thiamine, riboflavin, and nicotinic acid (actually its amide) are concerned with cellular oxidations; are not capable of being stored in the body to any large extent; are known to be destroyed in part in the tissues; are capable of being excreted by the kidney (as vitamins or products derived from vitamins); and are intermeshed with one another in cell respiration, it is obvious that the body's requirements for these vitamins are distinctly related to the sum total of energy metabolism, are required day by day, and probably are required in definite proportion to one another. According to very recent work [60] 0.45 mg of thiamine per 1000 calories of diet contributed by protein, fat, and carbohydrates in proportions conventional for the majority of American diets, appears to be approximately the level of intake required to prevent all but perhaps very mild degrees of biochemical defects in pyruvic acid metabolism. The Food and Nutrition Board of the National Research Council [21] has listed tentative daily allowances of thiamine for persons of seventeen age-sex-activity groups. These allowances, weighted on the basis of the 1940 census of composition of our population, reduce to 1.6 mg of thiamine per capita per day. Computations based on governmental surveys of dietary habits and thiamine analyses of the foods as eaten[61] indicate that for the few years prior to 1940 the average American diet provided only about 0.9 mg of thiamine per capita per day (or about 56 per cent of the recommended daily per capita allowance). Although concrete evidence such as that just summarized was lacking until very recently, evidence of various kinds pointed to the fact that enrichment of white flour with thiamine would be in the interest of better nutrition for the American people. This idea has now been extended to include thiamine, nicotinic acid, riboflavin and iron. Without going into the history of progress along the way, it can be said that developments have now reached the stage where it is expected that before the end of the year 1943 all family white flour will be enriched to carry 2.0 mg of thiamine per pound and all commercial white bread will be made from enriched flour or will carry the equivalent enrichment from other sources. When this program becomes effective, the average peace-time American diet will provide about 1.46 mg (or 91% of the per capita daily allowance) of thiamine which is contributed by various food classes as follows: meat, fish and poultry 16 per cent, dairy products 13 per cent, flour and cereals 53 per cent, vegetables 12 per cent, and fruits 6 per cent. No precise statement can be offered with regard to the magnitude to which unequal distribution may effect individuals. However, since flour and cereals, which contribute in largest measure to the thiamine value of the diet are among the least expensive types of foods, any shortages due to unequal distribution of foods should not particularly affect persons of relatively limited income.

In view of the relation of riboflavin to tissue oxidation, it would seem that, as in the case of thiamine, the daily requirement for riboflavin must be proportionate to energy metabolism. Earlier reports, [62, 63, 64] upon which the tentatively recommended daily allowances for riboflavin were largely based, indicated that the riboflavin requirement of man might be a little less than 0.06 mg per kilogram of body weight. More recent and more thoroughgoing work however has provided good evidence that 0.5 mg of riboflavin per 1000 calories contributed by an ordinary mixed diet provides for satisfactory tissue stores of this vitamin.[65] The average daily per capita allowance of calories for our present population amounts to 2800, and

on this basis the riboflavin allowance would be 1.4 mg per capita per day. This figure represents the best estimate available at present. For the few years prior to 1940 the average American diet, corrected for riboflavin losses in food preparation,[66] provided about 1.4 mg (or 100 per cent of the estimated allowance) of riboflavin per capita per day. With enrichment of family white flour to the proposed level of 1.2 mg per pound of flour and of commercial white bread, either through the use of enriched flour or by equivalent enrichment from other sources, the average peace-time American diet will provide about 1.9 mg (or 34 per cent excess over the estimated per capita need) of riboflavin. The riboflavin contributions of various food groups will then be as follows: dairy products 42 per cent, flour and cereals 30 per cent, meat, fish and poultry 16 per cent, vegetables 9 per cent and fruits 3 per cent. The extra allotment of riboflavin can probably be justified in times of certain food shortages, but it is not an adequate substitute for application of the newer knowledge of nutrition, for a progressive agricultural program which will insure adequate amounts of the basic foods in American dietaries, nor for solution of the economic ills of the country.

The human requirements for nicotinic acid are not well established. The tentative daily allowances recommended by the Food and Nutrition Board,[21] when computed for the seventeen age-sex-activity groups specified and weighted by the number of persons in each group on the basis of the 1940 census of population amounts to 16 mg per capita. The average per capita *intake* of nicotinic acid was until recently estimated[66] at approximately 77 per cent of the recommended allowance. With enrichment of all commercial bread and family white flour at the levels now in process of adoption, the full per capita quota will be met with a reasonable margin of excess. The proposed standard for nicotinic acid-enrichment of white flour is 16 mg per pound of flour. The common foods which contribute most significantly to the nicotinic acid intake of the American people include lean meats, and enriched white flour and bread. Normally these two classes of foods will provide about 94 per cent of the per capita daily allowance of nicotinic acid. Milk, fatty meats, cornmeal, fruits, and vegetables contain small amounts. In view of the endemic occurrence of pellagra in certain sections of this country, it is apparent that there has been and still is, in lesser degree, an uneven distribution of nicotinic acid-containing foods. Solution of this problem of applied nutrition calls for expansion of nutrition education and probably some economic readjustments.

Many other more recently discovered members of the B-vitamin group will undoubtedly prove to be of significance in human nutrition. Some of them, for example, pantothenic acid and biotin, are apparently concerned with tissue respiration. Pantothenic acid, (α, γ-dihydroxy-β, β-dimethylbutyryl-β-alanide) has been isolated[67] and synthesized.[68] It is apparently a constituent of all living tissues, and in consequence may be expected to have broad physiologic importance. By the same token, the chances of pantothenic acid deficiency states in man are reduced. Although a dietary deficiency of this vitamin has been observed by numerous investigators to be associated with symmetrical depigmentation of the hair or fur of animals, from which it may be inferred that it has some effect on the metabolism of melanin pigments, the mechanism of its action related to this property is not clear. Before applying this knowledge, it is well to note that pyridoxine, another vitamin of the B-group already mentioned, and with which pantothenic acid is often associated in nature is reported to promote or accelerate the graying of hair.[69]

Biotin was isolated[70, 71] and synthesized[72] in 1943. The structural formula for this vitamin is shown in Fig. 6. Three independent lines of research involving a nutrient essential for yeast growth, a nutrient essential for the growth of *Rhizobium*, a nitrogen-fixing bacterium,[73] and a substance capable of inhibiting in animals the deleterious effects of feeding raw egg white,[74] finally culminated in the knowledge that the three biologic effects were attributable to the same chemical entity, currently

referred to as biotin. The story of the convergence of these three parallel lines of investigation, which finally merged into the discovery that one substance accounted for all three biologic phenomena, is already familiar to students of nutrition and has been adequately presented [75] elsewhere. Biotin is widely distributed in the plant kingdom and also in animal tissues. Its significance in human nutrition is not known, but, from preliminary indications, it may play an important role as a catalyst in cellular respiration.

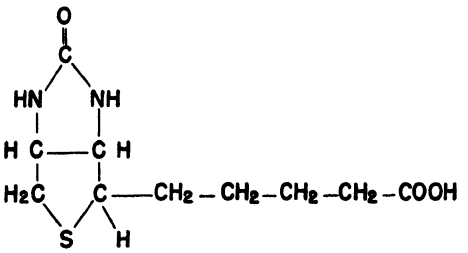


FIGURE 6. Biotin.

Inositol, para-aminobenzoic and folic acid (a vitamin-like substance found in green leaves), may also be considered members of the expanding family of B-vitamins. All are widely distributed in nature but their fundamental significance in human nutrition is only beginning to be investigated. Several of the more recently discovered members of B-vitamin group are synthesized by the bacterial flora in the gastro-intestinal tracts of animals. This phenomenon is very prominently featured in animals with spacious rumen.

Many of the B-vitamins stimulate the growth of microörganisms, which in turn, and as a result of growth, produce further quantities of B-vitamins. Ruminants probably are provided with a considerable proportion of their requirements for certain B-vitamins as a result of vitamin syntheses in their gastro-intestinal tracts. Under ordinary circumstances, other animals probably derive a part of their requirements for B-vitamins as a result of bacterial syntheses followed by subsequent absorption of these vitamins from the gastro-intestinal tract and in part from plant and animal tissues used as food and in which certain members of the group of B-vitamins appear to be almost universally present. This may be the case with man also but the question of which B-vitamins and to what extent man obtains them by reason of the growth of intestinal flora is a problem for the future. Those B-vitamins for which fairly well-established human requirements or tentative estimates of requirements have been mentioned in this chapter, however, must be given practical considerations in food selection, since otherwise specific vitamin-deficiency diseases will develop.

Ascorbic Acid. All intercellular substances of bone, cartilage, dentine and fibrous connective tissues have a common type of protein-substructure known as collagen, for the setting, fibrillation, or jelling of which ascorbic acid is a necessary agent.[76] The fundamental function of ascorbic acid is therefore concerned with the formation and maintenance of intercellular materials.

Ascorbic acid is not a dietary requirement of all species of animals nor even of all mammals, some of which can synthesize this vitamin. It is a dietary requirement of primates, including man and monkeys, and of guinea-pigs.

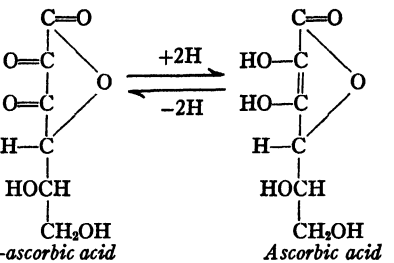
In ascorbic acid deficiency states in the young, there develops retarded or defective production of intercellular matrices in bone, dentine, and connective tissues; with prolongation of the deficiency, all deposition of intercellular substances ceases and there sets in a slow resorption of matrices previously laid down. In scorbutic states, generally, the organic matrices of the supporting tissue structures become dis-

organized and are slowly resorbed, in consequence of which the minerals previously deposited in the teeth and bones are also resorbed.[76] This explains in brief the marked osteoporosis associated with severe scurvy.

The normal and complete repair of wounds is dependent upon the action of ascorbic acid by virtue of the latter's essential function in the formation of normal intercellular materials.

The mechanism by which ascorbic acid promotes the formation and maintenance of normal intercellular material is unknown.

The most obvious chemical property of ascorbic acid is its capacity for reversible oxidation and reduction, which can be depicted as follows:



Dehydro-ascorbic acid *Ascorbic acid*

Glutathione, a substance already mentioned as being concerned in tissue respiration, possesses the capacity of reducing the oxidized form of ascorbic acid. It is more than likely that the reversibility of the oxidized and reduced forms of ascorbic acid which impart the property of a hydrogen transporter to this vitamin is of major importance in the physiology of the body tissues.

The principal sources of ascorbic acid in human dietaries include the citrus fruits, tomatoes, and green leafy vegetables. In quantities commonly consumed, potatoes can, if properly prepared, provide a large share of the dietary requirements for ascorbic acid.

There is some question as to whether or not the body derives any special benefit from saturation of the tissues with this vitamin over and above maintenance of some lesser concentration. Tissue saturation with ascorbic acid is demonstrated by a prompt and precipitous rise in the ascorbic acid level in the blood plasma following the administration of large doses (on the order of 8 mg or more per kg of body weight) of ascorbic acid, the rise in ascorbic acid level in the plasma being accompanied by urinary excretion of a large part of the ingested vitamin. For the maintenance of tissue saturation with ascorbic acid, infants under six months old require daily on the order of 2.0 mg of ascorbic acid per kg of body weight,[77] while older children[77] and adults[21] require about 1.0 mg per kg of body weight. It should be noted here that the normal requirements for this vitamin are proportional to body weight, an observation which is in accord with the function of this vitamin in maintenance of normal tissue structures.

The Fat-soluble Vitamins. Vitamin A or some carotenoid precursor which has the β-ionone ring and conjugated side chain structure characteristic of Vitamin A is essential for normal action of the retina in dim light and probably in other aspects of visual processes. It has been known at least since 1500 B.C. that eating the livers of animals which are now known generally to be rich sources of vitamin A was an effective means of restoring certain types of impaired dark adaptation. The recent work of Wald and his associates[78] has offered considerable new advance in understanding the chemical basis of dark adaptation. The visual cycle as it relates to the role of vitamin A in dark adaptation and as depicted by Wald is shown in Fig. 7.*

* See paper by George Wald in this volume. J. A.

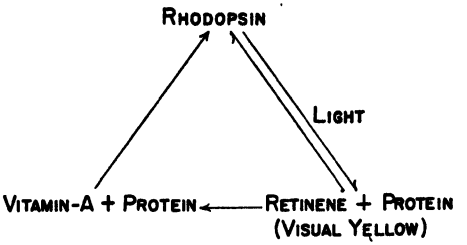


FIGURE 7. Rhodopsin cycle in visual adaptation.

The sensations of light are due to nerve impulses initiated by changes in light sensitive-materials in the retina. When light reaches the visual purple pigment or, more strictly speaking, the rhodopsin in retinal structures, of the mammalian species, some of the rhodopsin is broken down in the first stage to another pigment known as retinene and a protein and in a second stage to vitamin A and protein. The protein constituent of visual purple appears to be a globulin and to be joined to phospholipid to form a phospho-protein.[79] For a given intensity of light a corresponding equilibrium in the concentrations of materials shown as parts of this cycle becomes established. If then the light is extinguished, a new equilibrium gradually establishes itself with a greater proportion of rhodopsin being regenerated. In vitamin A deficiency attended with nutritional night-blindness, the equilibrium established for the dark-adapted eye is such that vision in dim light is impaired, and the light threshold is raised. Some years ago Fridericia and Holm[80] observed that, whereas the retinas of light-adapted eyes could be readily detached from the choroid layer underneath, in dark-adapting eyes the retina adhered closely to the choroid layer which is known to contain numerous blood capillaries. It would appear that the level of vitamin A in blood plasma was the determining factor in the presence or absence of nutritional night-blindness. Under ordinary circumstances the level of vitamin A in blood plasma appears to remain fairly constant so long as the vitamin A supply in the food is adequate, or so long as substantial vitamin A reserves are present in the liver. Lacking an adequate supply in the food or appreciable reserves in the liver, nutritional night-blindness results and progresses as the state of vitamin A deficiency is prolonged.

Vitamin A is also essential for the normal development and maintenance of epithelial structures which cover the free surfaces of the body (cutaneous, mucous, serous, and including the glands) and for promotion of the normal functioning of these structures. The so-called enamel organ, an epithelial organ which plays a leading role in the formation of organized tooth structure, atrophies and undergoes morphologic changes as a result of vitamin A-deficiency. A characteristic paralysis has been noted for many years to be an accompaniment of a dietary deficiency of vitamin A in young experimental animals. This has recently been shown to result from the fact that following deprivation of Vitamin A, the central nervous system and other soft tissues continue to grow at approximately normal rates after the growth of bone has been greatly slowed up. With complete cessation of growth, this disproportion of body parts remains. The brain and spinal cord and nerve roots suffer mechanical damage as a result of being overcrowded in the cranial cavity, the tract of the spinal cord, and intervetebral foramina respectively. The nervous lesions in vitamin A-deficiency are thus explained as being wholly mechanical in origin. All the functions of vitamin A, and the pathologic lesions attendant upon a deficiency of vitamin A mentioned in this paragraph, have been systematically reviewed and discussed recently by Wolbach and Bessey,[76] who have also contributed in very large measure to present knowledge of this subject.

Vitamin A is abundant in common foods. All dark green leaves, all deep yellow-

colored vegetables (carrots, sweet potato, pumpkin) and fruits (apricots, persimmons, muskmelons and peaches), fish-liver oils and livers of all well-fed animals are rich sources of Vitamin A by virtue of containing either Vitamin A or its esters or by being rich in vitamin A-active carotenoids (alpha, beta and gamma carotene and cryptoxanthin). Milk and eggs are good sources of Vitamin A and many other foods contain small amounts of this vitamin. Probably it could be stated with more than a modicum of truth that a dietary insufficiency of this vitamin can result, under ordinary conditions which prevail in this country, only through ignorance of the simplest facts about food values.

During World War I, evidence of rather widespread vitamin A-deficiency leading to blindness among Danish children was reported,[81] and was attributed to the fact that butter and whole milk were so scarce and expensive as to make it impossible for those with low or moderate incomes to provide their children with these foods. The incidence of the disease (xerophthalmia) was checked by the German blockade, which prevented export of butter and by the fact that the butter thus made available was rationed.

Booher *et al.*[82] have reported that from 8 to 17 micrograms of vitamin A per kilogram of body weight per day are required to prevent impaired dark adaptation in adults, which finding has been confirmed by Wagner [83] and others. Lewis and Haig [84] have reported essentially similar daily vitamin A requirements (per kilogram of body weight) for the prevention of impaired dark adaptation in infants 1½ to 10½ months old. The minimum daily intake of carotene required for the prevention of impaired dark adaptation in adults, when supplied in the form of commercial carotene (85 per cent beta- and 15 per cent alpha-carotene) dissolved in cottonseed oil varied between the limits of about 26 and 62 micrograms per kilogram of body weight.[82] At least part of the relatively large differences between minimum vitamin A requirements in terms of vitamin A from cod-liver oil and those in terms of carotene (one microgram of vitamin A is the biologic equivalent of about two micrograms of β-carotene) are probably due to the relatively poorer absorption of the carotene.[85] In terms of biologic equivalents, β-carotene was only about 50 to 60 per cent as effective as the vitamin A in cod-liver oil in the maintenance of normal dark adaptation. The carotene in spinach or in green peas was somewhat better utilized than carotene itself.[86] From observations of Guilbert *et al.*[87] on cattle, sheep, and swine, in conjunction with those on man, it would appear that the daily vitamin A (and/or carotene) requirements on a body-weight basis were possibly of the same order for all mammals.

Vitamin E is reported [88, 89, 90] to exert a sparing effect upon vitamin A, and more so on carotene, resulting in improved utilization of vitamin A or carotene for growth and vitamin A storage. The antioxidant property of vitamin E has been known for some time, and it is assumed that this property manifests itself in the gastro-intestinal tract to check the destruction of vitamin A and carotene, thus exerting indirectly a sparing effect on these. It has also been found [91] that previously accumulated stores of vitamin A in the livers of animals are more slowly depleted when the animals are transferred to a vitamin A-deficient diet if vitamin E is provided than if the animals are deprived of vitamin E. Whether or not this sparing effect of vitamin E on the utilization of accumulated body-stores of vitamin A is the result of the antioxidant properties of vitamin E has not been determined.

The vitamins E (alpha, beta, and gamma tocopherol) occur chiefly in vegetable oils (wheat-germ oil, cottonseed oil, and other seed oils), in olive oil and in green leaves. A deficiency of vitamin E in the diets of animals has been shown to result in reproductive failure and in severe and generalized degeneration of the skeletal muscles (so-called nutritional muscular dystrophy), which phenomena has been the subject of intensive studies in recent years. The pathology of the vitamin E-deficiency state has been summarized in the article of Wolbach and Bessey,[76] who have

included in their review a comprehensive list of references to original published reports on this subject.

The necessity for vitamin E in the human dietary has not been established. If it is an essential nutrient for man, the needs are not likely to present any major problem in the practical applications of nutrition, since this vitamin is very widely distributed in common foods.

Vitamin K, sometimes referred to as the anti-hemorrhagic vitamin, occurs in nature in at least two forms, designated as K_1 (in green leafy vegetables) and K_2 (produced by bacteria). The first evidence pointing to the existence of vitamin K was reported by Dam,[92] who observed the development of extensive subcutaneous intramuscular hemorrhages and a delay in blood-clotting time in young chicks as a result of feeding them a special fat-free diet.

Vitamin K is essential for maintaining a normal prothrombin level in blood.[93] Marked lowering of the prothrombin level of the blood (to about 20 per cent of its normal value) is attended with delayed blood-clotting time. The mechanism by which vitamin K functions to maintain normal prothrombin levels in the blood is not known. The so-called hemorrhagic disease occasionally observed in new-born infants is tractable to vitamin K therapy. Some clinicians have recommended administration or injection of vitamin K for the new-born infant or for the mothers shortly prior to the birth of the infant as a precautionary measure. After the first few days of infant life, vitamin K is normally produced by bacterial action in the intestinal tract, from which in the presence of bile, it can be absorbed and utilized for the maintenance of normal prothrombin levels in the blood.

The naturally occurring vitamins K are fat-soluble and require the presence of bile salts for their absorption from the intestinal tract. Individuals with biliary obstructions may develop K-avitaminosis, notwithstanding the fact that their food may have contained adequate amounts of Vitamin K.[94] As indicated in this discussion, there is little need for concern about adequate provision for vitamin K for the normal individual. K-avitaminosis is largely a problem of the clinician who has had made available to him the necessary materia medica with which to combat it.

Several synthetic products, possessing vitamin K-activity, have been developed, and some are being used for medicinal purposes. The most biologically active synthetic compound is 2-methyl-1, 4-naphthoquinone,[95] which possesses about three times the potency of natural vitamin K, (2-methyl-3phytyl-1, 4-naphthoquinone).[96]

The term "vitamin D" is applied to substances of a sterol nature, which possess antirachitic activity. Vitamin D, as such, occurs in significant amounts in very few foods other than fish-liver and body oils. At least ten different sterols, known as provitamins D, can be activated to vitamins D by irradiation with ultraviolet light or by bombardment with cathode rays.[97] The conversion of provitamin D to vitamin D is not a simple process, since many intermediate compounds are formed and the resulting Vitamin D is also capable of transformation into other products. Thus, from a mixture of substances, there is just one that possesses vitamin D activity and the others are closely related contaminants, some of them highly undesirable or toxic. Slight alterations in the conditions of activation will cause changes in the relative proportions of the products formed.

The provitamin D most prominent in plants or more specifically in molds and yeast is ergosterol. The provitamin D most prominent in man and domestic animals, where it exists largely in the skin, is 7-dehydro-cholesterol. Activated ergosterol is also known as Vitamin D_2, calciferol, or viosterol. The most prominent D-vitamin in fish-liver oils is Vitamin D_3, which is identical with activated 7-dehydro-cholesterol and is, therefore, the Vitamin D formed in the skin as a result of its exposure to ultraviolet rays from the sun and sky.

Vitamin D is essential for normal growth sequences in bone and for normal mineralization of bone. The deleterious effects of a deficiency of vitamin D on

teeth (defective calcification of the enamel and dentine and atrophy of the developmental organs of these structures) are most marked during the period of tooth formation.[76]

Lack of vitamin D alone will produce rickets in infants. Satisfactory mineralization of the skeleton of infants, however, requires adequate supplies of calcium, phosphorus and vitamin D. Osteoporosis, the adult counterpart of rickets, probably results from a deficiency of vitamin D together with a relative deficiency of calcium or phosphorus in the diet, or a decidedly low intake of one or the other or both of these inorganic nutrients.[76]

There is no evidence which indicates a local effect of vitamin D at the zone of deposition of calcium and phosphorus in the bone matrix. Vitamin D functions chiefly, it is believed, by increasing the efficiency of absorption of calcium and phosphorus from the digestive tract. By so doing, it promotes the maintenance of normal concentrations of calcium and phosphate ions in the blood plasma so that calcification of the skeleton can take place. There is some evidence which indicates that the means by which vitamin D increases the efficiency of absorption of calcium and phosphorus is through lowering of the pH of the gastro-intestinal contents.

Vitamin D values of foods and medicinal preparations are usually expressed in terms of units. One International Unit or one U. S. Pharmacopeia Unit is equivalent to about 0.025 microgram of vitamin D_2. The primary vitamin D reference standard is a solution of irradiated ergosterol (made under specified conditions) in olive oil. The solution contains one milligram of the activated ergosterol in 10 ml of olive oil. A unit of vitamin D is one milligram of this solution. Secondary standards are prepared on the basis of biologic assays wherein the criteria of equivalence of primary to secondary standard is estimated by the quantities of the two required to promote a definite degree of healing in rachitic animals.

Infants receiving adequate quantities of milk require about 135 units of vitamin D daily in order to be fully protected from rickets, but they require from 350 to 400 units of vitamin D to support a satisfactory rate of growth and good mineralization of the skeleton.[98] Daily intakes of the same general order of magnitude are required throughout childhood, and similar quantities are advocated for women during periods of pregnancy and lactation.[21]

The administration of excessive amounts of vitamin D (of the order of 1600-1800 units per day) to infants, causes a retardation of growth,[98] an observation which has substantiation in the histologic changes observed to develop as an accompaniment of excessive intakes of vitamin D in animals.[76]

There would appear to be no important difference in effectiveness for the human species between Vitamin D_2 and Vitamin D_3.[98] The vitamin D requirement of average adults has not been established. A part of the adult's requirement is provided as a consequence of exposure to sunshine and sky shine, but it is probably to the best interests of the average adult to take vitamin D in some form, up to quantities of 350 or 400 units daily if their lives are largely spent indoors.

Except for appreciable quantities of vitamin D in herring, salmon, tuna, and sardines, all of which are classed as fatty fish, foods are unimportant sources of vitamin D. The quantities of vitamin D in butter, milk, cream, cheese, and egg yolks produced under ordinary conditions, are small and variable. They depend upon the degree to which the animal or fowl producing the food has been effectively exposed to ultraviolet light or fed foods rich in Vitamin D. Animals and fowls are not efficient in their transference of Vitamin D from their bodies to such products as milk and eggs. The practical alternative for enhancing the vitamin D values of foods is to activate the foods themselves by irradiation with ultraviolet light or to add concentrates of vitamin D to them. Both of these means have been applied to preparation of Vitamin D milk. In the judgment of many nutritionists, the addition of vitamin D to all fluid milk would go a long way toward eliminating the possi-

bility of vitamin D deficiencies, since this vitamin would then be more generally distributed in the medium which furnishes the best sources of those nutrients (calcium and phosphorus) in conjunction with which vitamin D specifically functions.

The cause of demineralization of the skeleton sometimes observed in elderly persons, and known as senile osteoporosis, may be due in part at least to chronically poor food habits and especially to inadequate intakes of calcium, phosphorus, ascorbic acid and vitamin D. It is not certain but that development of this condition may have other underlying causes, such as endocrine readjustments, which accompany other phases of ageing. However, satisfactory intakes of those nutrients known to promote the maintenance of good skeletal structures are certainly to be recommended to aged persons as well as to persons of younger years.

From this short summary of the role of minerals and vitamins in applied nutrition, it is abundantly evident that we are beginning to understand some of the fundamentals of nutritive processes. We can also begin to understand some of the fundamental relationships between ourselves and the great expanse of ocean from whence we came and the plant life from whence we derive sustenance. The science of nutrition, as it gradually develops, has pointed distinctly more toward a better life than toward a longer life. Lest we hold too much hope of an increased life span, I quote here from J. H. Cohausen (Hermippus Redivivus): "Of all the wild chimeras which have in all ages haunted the minds of fantastic dreamers, none has taken so great a hold on its votaries as the search for the 'Elixir Vitae' or means of prolonging life beyond its allotted span." Enthusiasts have believed that "by lengthening life, they would confer a priceless boon on the human race; forgetting that it is not the length of the day which makes us love the summer, but the brightness of the sun, the beauty of the flowers, the singing of the birds."

Addendum

Subsequent to preparation of this chapter, an article dealing with the metabolic oxidation of sugar and fat and the relation of these to the citric acid cycle shown in Fig. 5 has been released [Breusch, F. L., Citric Acid Cycle: "Sugar and Fat-breakdown in Tissue Metabolism," *Science*, **97**, 490-492 (1943)]. In view of the fundamental implications contained in this article, it is being cited here for the benefit of readers especially interested in the chemistry of tissue oxidations since there is introduced in this article a possibility of some important departures and extensions of earlier concepts of these reactions.

References

1. Macallum, A. B., "The Paleochemistry of the Body Fluids and Tissues," *Physiol. Rev.*, **6**, 316-357 (1926).
2. Lowry, O. H., and Hastings, A. B., "Histochemical Changes Associated with Aging. I. Methods and Calculation," *J. Biol. Chem.*, **143**, 257-269 (1942).
3. Lowry, O. H., Hastings, A. B., Hull, T. Z., and Brown, A. N., "Histochemical Changes Associated with Aging. II. Skeletal and Cardiac Muscle in the Rat," *J. Biol. Chem.*, **143**, 271-280 (1942).
4. McCay, C. M., and Crowell, M. F., "Prolonging the Life Span," *Scient. Mo.*, **39**, 405-414 (1934).
5. Lowry, O. H., McCay, C. M., Hastings, A. B., and Brown, A. N., "Histochemical Changes Associated with Aging. III. The Effects of Retardation of Growth on Skeletal Muscle," *J. Biol. Chem.*, **143**, 281-284 (1942).
6. Gamble, J. L., Ross, G. S., and Tisdall, F. F., "The Metabolism of Fixed Base During Fasting," *J. Biol. Chem.*, **57**, 633-695 (1923).
7. Schmidt, C. L. A., and Greenberg, D. M., "Occurrence, Transport and Regulation of Calcium, Magnesium and Phosphorus in the Animal Organism," *Physiol. Rev.*, **15**, 297-434 (1935).
8. McCollum, E. V., and Orent, E. R., "Effect on the Rat of Deprivation of Magnesium," *J. Biol. Chem.*, **92** (Proceedings), XXX (1931).
9. Klein, H., Orent, E. R., and McCollum, E. V., "The Effects of Magnesium Deficiency on the Teeth and Their Supporting Structures in Rats," *Am. J. Physiol.*, **112**, 256-262 (1935).
10. Meltzer, S. J., and Auer, J., "The Antagonistic Action of Calcium Upon the Inhibitory Effect of Magnesium," *Am. J. Physiol.*, **21**, 400-419 (1908).
11. Shohl, A. T., "Mineral Metabolism," Chapter 6, Am. Chem. Soc. Monograph Series, Reinhold Publishing Corp., New York, 1939.
12. Sherman, H. C., "Chemistry of Food and Nutrition," 6th Ed., Chapter XII, Macmillan Co., New York, 1941.
13. Roseberry, H. H., Hastings, A. B., and Morse, J. K., "X-ray Analysis of Bone and Teeth," *J. Biol. Chem.*, **90**, 395-408 (1931).
14. Bauer, W., Aub, J. C., and Albright, F., "Bone Trabeculae as a Readily Available Reserve Supply of Calcium, *J. Exptl. Med.*, **49**, 145-161 (1929).

15. Private Communication from G. H. Hansmann, Pathologist, Columbia Hospital, Milwaukee, Wisconsin.
16. Macy, I. G., "Principal Mineral Elements in Nutrition"; Handbook of Nutrition, VI, *J. Am. Med. Assn.*, **120**, 34-42 (1943).
17. Sherman, H. C., "Calcium Requirement of Maintenance in Man," *J. Biol. Chem.*, **44**, 21-27 (1920).
18. Sherman, H. C., "Phosphorus Requirement of Maintenance in Man," *J. Biol. Chem.*, **41**, 173-179 (1920).
19. Leitch, I., "The Determination of the Calcium Required by Man," *Nutr. Abs. Rev.*, **6**, 553-578 (1937); **8**, 1-2 (1938).
20. Fuhr, I., and Steenbock, H., "The Effect of Dietary Calcium and Phosphorus and Vitamin D on the Utilization of Iron," *J. Biol. Chem.*, **147**, 59-64 (1943); **147**, 65-69 (1943); **147**, 71-74 (1943).
21. National Research Council, Reprint and Circular Series No. 115: "Recommended Dietary Allowances," Publication Office, National Research Council, Washington, D. C.
22. Hart, E. B., Steenbock, H., Waddell, J., and Elvehjem, C. A., "Iron in Nutrition. VII. Copper as a Supplement to Iron for Hemoglobin Building in the Rat," *J. Biol. Chem.*, **77**, 797-812 (1928).
23. Stearns, G., and McKinley, J. B., "The Conservation of Blood Iron During the Period of Physiological Hemoglobin Destruction in Early Infancy," *J. Nutrition*, **13**, 143-156 (1937).
24. Heath, C. W., "Iron in Nutrition"; Handbook of Nutrition, *J. Am. Med. Assn.*, **120**, 366-370 (1942).
25. Kendall, E. C., "The Isolation in the Crystalline Form of the Compound Containing Iodine Which Occurs in the Thyroid; Its Chemical Nature and Physiological Activity," *J. Am. Med. Assn.*, **64**, 2042-2043 (1915).
26. Harington, C. R., and Barger, G., "Chemistry of Thyroxin, III. Constitution and Synthesis," *Biochem. J.*, **21**, 169-183 (1927).
27. v. Kolnitz, H., and Remington, R. E., "Chemical Studies on 150 Normal Thyroids from Charleston, S. C.," *Endocrinology*, **17**, 563-568 (1933).
28. Council on Foods, *J. Am. Med. Assn.*, **111**, 156-157 (1938).
29. Rose, W. C., "The Nutritive Significance of the Amino Acids and Certain Related Compounds," *Science*, **86**, 298-300 (1937); *Physiol. Rev.*, **18**, 109-136 (1938).
30. Shils, M. E., and McCollum, E. V., "The Trace Elements in Nutrition"; Handbook of Nutrition IX, *J. Am. Med. Assn.*, **120**, 609-619 (1942).
31. Sherman, H. C., and Gettler, A. O., "The Balance of Acid-forming and Base-forming Elements in Food in Its Relation to Ammonia Metabolism," *J. Biol. Chem.*, **11**, 323-338 (1912).
32. Auhagen, E., "Über Co-Carboxylase. Reinigungsversuche und Vorkommen in tierischen Organen," *Biochem. Z.*, **258**, 330-339 (1933).
33. Lohmann, K., and Schuster, P., "Untersuchungen über die Co-Carboxylase," *Biochem. Z.*, **294**, 188-214 (1937).
34. Tauber, J., "The Interaction of Vitamin B_1 in Enzymic Reactions," *J. Biol. Chem.*, **123**, 499-506 (1938).
35. Kinnersley, H. W., and Peters, R. A., "Note Upon the Preparation of Crude Co-Carboxylase from Vitamin B_1 by Yeast," *Biochem. J.*, **32**, 697-698 (1938).
36. Weijlard, J., and Tauber, H., "Synthesis, Isolation and Identification of Co-carboxylase," *J. Am. Chem. Soc.*, **60**, 2263-2264 (1938).
37. Stern, K. G., and Hofer, J. W., "Synthesis of Co-carboxylase from Vitamin B_1," *Enzymologia*, **3**, 82-95 (1937).
38. Lipmann, F., "Hydrogenation of Vitamin B_1," *Nature*, **138**, 1097-1098 (1936).
39. Zima, O., and Williams, R. R., "Über ein antineuritisch wirksames Oxydationsprodukt des Aneurins," *Ber. deutsch. Chem. Gesellsch.*, **73**, 941-949 (1940).
40. Kubowitz, F., and Lüttgens, W., "Zusammensetzung, Spaltung und Resynthese der Carboxylase," *Biochem. Z.*, **307**, 170-172 (1941).
41. Ochoa, S., and Peters, R. A., "Vitamin B_1 and Co-carboxylase in Animal Tissues," *Biochem. J.*, **32**, 1501-1505 (1938).
42. Peters, R. A., "The Biochemical Lesion in Vitamin B_1 Deficiency, Application of Modern Biochemical Analysis in Its Diagnosis," *Lancet*, **230**, 1161-1164 (1936).
43. Banga, I., Ochoa, S., and Peters, R. A., "Pyruvate Oxidation in Brain. VI. The Active Form of Vitamin B_1 and the Role of C_4 Dicarboxylic Acids," *Biochem. J.*, **33**, 1109-1121 (1939).
44. Westenbrink, H. G. K., and Goudsmit, J., "Investigations on the Aneurin and Cocarboxylase Content of Animal Tissues, Estimated by the Thiochrome Method," *Enzymologia*, **5**, 307-320 (1938).
45. Tauber, H., "The Interaction of Vitamin B_1 in Enzymic Reactions," *J. Biol. Chem.*, **123**, 499-506 (1938).
46. Ochoa, S., "Enzymic Synthesis of Cocarboxylase in Animal Tissues," *Biochem. J.*, **33**, 1262-1270 (1939).
47. Warburg, O., and Christian, W., "Uber ein neues Oxydationsferment und sein Absorptionsspektrum," *Biochem. Z.*, **254**, 438-458 (1932).
48. Kuhn, R., György, P., and Wagner-Jauregg, T., "Uber ein neue Klasse von Naturfarbsloffen," *Ber. deutsch. Chem. Gesellsch.*, **66**, 317, 576, 1034, 1577 (1933).
49. Booher, L. E., "Chemical Aspects of Riboflavin," The Vitamins, A Symposium, Ch. XIII, Am. Med. Assn., Chicago, 1939.
50. Warburg, O. and Christian, W., "Pyridin, der Wasserstoff-ubertragende Bestandteil von Garungsfermenten (Pyridin-Nucleotide)," *Biochem. Z.*, **287**, 291-328 (1936).
51. Rosenberg, H. R., "Chemistry and Physiology of the Vitamins," pp. 219-250, Interscience Publishers, Inc., New York, 1942.
52. Dixon, M., and Zerfas, L. G., "The Role of Coenzymes in Dehydrogenase Systems," *Biochem. J.*, **34**, 371-391 (1940).
53. Elliott, K. A. C., "Intermediary Metabolites and Respiratory Catalysis," *Physiol. Rev.*, **21**, 267-306 (1941).
54. Birch, T. W., "The Relation between Vitamin B_6 and the Unsaturated Fatty Acid Factor," *J. Biol. Chem.*, **124**, 775-793 (1938).
55. Schneider, H. A., Ascham, J. K., Platz, B. R., and Steenbock, H., "The Anti-acrodynic Properties of Certain Foods," *J. Nutrition*, **18**, 99-104 (1939).
56. Schneider, H., Steenbock, H., and Platz, B. R., "Essential Fatty Acids, Vitamin B_6 and Other Factors in the Cure of Rat Acrodynia," *J. Biol. Chem.*, **132**, 539-551 (1940).
57. McHenry, E. W., and Gavin, G., "The B Vitamins and Fat Metabolism, IV. The Synthesis of Fat from Protein," *J. Biol. Chem.*, **138**, 471-475 (1941).
58. Krebs, H. A., and Johnson, W. A., "Citric Acid in Intermediate Metabolism in Animal Tissues," *Enzymologia*, **4**, 148-156 (1937).
59. Westerfeld, W. W., Stolz, E., and Berg, R. L., "The Role of Pyruvate in the Metabolism of Ethyl Alcohol," *J. Biol. Chem.*, **144**, 657-665 (1942).

60. Williams, R. D., Mason, H. L., and Wilder, R. M., "The Minimum Daily Requirement of Thiamine of Man," *J. Nutrition*, **25**, 71-97 (1943).
61. Lane, R. L., Johnson, E., and Williams, R. R., "Studies of the Average American Diet. I. Thiamine Content," *J. Nutrition*, **23**, 613-624 (1942).
62. Strong, F. M., Feeney, R. E., Moore, B., and Parsons, H. T., "The Riboflavin Content of Blood and Urine," *J. Biol. Chem.*, **137**, 363-372 (1941).
63. Sebrell, W. H., Butler, R. E., Wooley, J. G., and Isbell, H., "Human Riboflavin Requirement Estimated by Urinary Excretion of Subjects on Controlled Intake," *Pub. Health Rep.*, **56**, 510-520 (1941).
64. Sebrell, W. H., "Human Riboflavin Deficiency (Ariboflavinosis)," The Biological Action of Vitamins, A Symposium, University of Chicago Press, Chicago, 1942.
65. Williams, R. D., Mason, H. L., Cusick, P. L., and Wilder, R. M., "Observations on Induced Riboflavin Deficiency and the Riboflavin Requirement of Man," *J. Nutrition*, **25**, 361-377 (1943).
66. Williams, R. R. (private communication), material in press.
67. Williams, R. J., Truesdail, J. H., Weinstock, H. H., Rohrmann, E., Lyman, C. M., and McBurney, C. H., "Pantothenic Acid. II. Its Concentration and Purification from Liver," *J. Am. Chem. Soc.*, **60**, 2719-2723 (1938).
68. Williams, R. J., "Pantothenic Acid—A Vitamin," *Science*, **89**, 486 (1939).
69. Morgan, A. F., Cook, B. B., and Davidson, H. G., "Vitamin B_2 Deficiencies as Affected by Dietary Carbohydrate," *J. Nutrition*, **15**, 27-43 (1938).
70. duVigneaud, V., Hofmann, K., Melville, D. B., and György, P., "Isolation of Biotin (Vitamin H) from Liver," *J. Biol. Chem.*, **140**, 643-651 (1941).
71. Kögl, F., and Tönnis, B., "Über das Bios-Problem, Darstellung von krystallisierten Biotin aus Eigelb," *Z. physiol. Chem.*, **242**, 43-73 (1936).
72. Harris, S. A., Wolf, D. E., Mozingo, R., and Folkers, K., "Synthetic Biotin," *Science*, **97**, 447-448 (1943).
73. Allison, F. E., Hoover, S. R., and Burk, D., "A Respiration Coenzyme," *Science*, **78**, 217-218 (1933).
74. Boas, M. A., "The Effect of Desiccation upon the Nutritive Properties of Egg-white," *Biochem. J.*, **21**, 712-724 (1927).
75. duVigneaud, V., "Biotin," The Biological Action of the Vitamins. A Symposium. University of Chicago Press, Chicago, 1942.
76. Wolbach, S. B., and Bessey, O. A., "Tissue Changes in Vitamin Deficiencies," *Physiol. Rev.*, **22**, 233-289 (1942).
77. Booher, L. E. (unpublished data).
78. Wald, G., and Clark, A. B., "Visual Adaptation and Chemistry of the Rods," *J. Gen. Physiol.*, **21**, 93-105 (1938).
79. Broda, E. E., "The Role of Phospholipin in Visual Purple Solutions," *Biochem. J.*, **35**, 960-964 (1941).
80. Fridericia, L. S., and Holm, E., "Experimental Contribution to the Study of the Relation Between Night-blindness and Malnutrition," *Am. J. Physiol.*, **73**, 63-78 (1925).
81. Bloch, C. E., "Blindness and Other Diseases in Children Arising from Deficient Nutrition (Lack of Fat-Soluble Vitamin A Factor)," *Am. J. Dis. Children*, **27**, 139-148 (1924); **42**, 263-278 (1931).
82. Booher, L. E., Callison, E. C. and Hewston, E. M., "An Experimental Determination of the Minimum Vitamin A Requirements of Normal Adults," *J. Nutrition*, **17**, 317-331 (1939).
83. Wagner, K. H., "Der Tagesbedarf des Menschen an Vitamin A und β-carotin," *Z. Klin. Medicin*, **137**, 659-663 (1940).
84. Lewis, J. M., and Haig, C., "Vitamin A Requirements in Infancy as Determined by Dark Adaptation," *J. Pediat.*, **15**, 812-823 (1939).
85. Wald, G., Carroll, W. R., and Sciarra, D., "The Human Excretion of Carotenoids and Vitamin A," *Science*, **94**, 95-96 (1941).
86. Booher, L. E., and Callison, E. C., "The Minimum Vitamin A Requirements of Normal Adults. II. The Utilization of Carotene as Affected by Certain Dietary Factors and Variations in Light Exposure," *J. Nutrition*, **18**, 459-471 (1939).
87. Guilbert, H. R., Miller, R. F. and Hughes, E. H., "The Minimum Vitamin A and Carotene Requirement of Cattle, Sheep and Swine," *J. Nutrition*, **13**, 543-564 (1937).
88. Moore, T., "The Effect of Vitamin E Deficiency on the Vitamin A Reserves of the Rat," *Biochem. J.*, **34**, 1321-1328 (1940).
89. Davies, A. W., and Moore, T., "Interaction of Vitamins A and E," *Nature*, **147**, 794-796 (1941).
90. Hickman, K., Harris, P. L., and Woodside, M. R., "Interrelationships of Vitamins A and E," *Nature*, **150**, 91-92 (1942).
91. Quackenbush, F. W., Cox, R. P., and Steenbock, H., "Tocopherol and the Stability of Carotene," *J. Biol. Chem.*, **145**, 169-177 (1942).
92. Dam, H., "Über ein Cholesterinsynthese in Tierkörper," *Biochem. Z.*, **220**, 158-163 (1930).
93. Dam, H., Schønheyder, F., and Tage-Hansen, E., "Studies on the Mode of Action of Vitamin K," *Biochem. J.*, **30**, 1075-1079 (1936).
94. Bollman, J. L., Butt, H. R., and Snell, A. M., "The Influence of the Liver on the Utilization of Vitamin K," *J. Am. Med. Assn.*, **115**, 1087-1091 (1940).
95. Ansbacher, S., and Fernholz, E., "Simple Compounds with Vitamin K Activity," *J. Am. Chem. Soc.*, **61**, 1924-1925 (1939).
96. Fieser, L. F., "Synthesis of Vitamin K," *J. Am. Chem. Soc.*, **61**, 3467-3475 (1939).
97. Bills, C. E., "The Chemistry of Vitamin D," Chap. XXXIII, in "The Vitamins," Am. Med. Assn., Chicago, 1939.
98. Jeans, P. C., "Vitamin D Milks, with Clinical Discussion," *J. Am. Med. Assn.*, **106**, 2066-2069, 2150-2159; also Jeans, P. C., and Stearns, G., "The Human Requirement of Vitamin D," Chap. XXVI, in The Vitamins, Am. Med. Assn., Chicago, 1939.

The Mechanism of Hormone Action

OLIVER KAMM AND D. K. KITCHEN

Research Department, Parke, Davis & Co., Detroit, Mich.

Introduction: Chemical Activators, Vitamins, Hormones and Enzymes

The biological catalysts that control the intricate chemical processes which occur in the animal body are, in the order of increasing complexity, the chemical activators, the vitamins, the hormones, and the enzymes. This chapter will deal in an elementary way with the mechanism of action of the hormones; the mechanism of vitamin and enzyme actions has been treated elsewhere.

All forms of life, including single-celled organisms, are in part under enzymatic control, and even bacteria require certain vitamins and growth factors, especially para-aminobenzoic acid; * but only highly organized forms of life—mammalia, aves, reptilia, amphibia, pisces—require hormones. The vertebrates are listed above in reverse order to their state of development; all possess a full complement of endocrine glands [1]—specialized organs for the manufacture, storage, and distribution of hormones.

The term *hormone*, which is derived from a Greek word which literally means "to arouse" or "to stir up," has suffered abuse in recent years and has been loosely applied to a wide variety of chemical substances. This may be readily appreciated by a review of Starling's original definition [2] which was intended to include principles elaborated by cells at one site but which produce a physiologic response in certain cells or structures remotely situated in the body. In its broadest interpretation, such catabolites as carbon dioxide and histamine could be considered as hormones.[3]

In his review article, *Chemical Regulation and the Hormone Concept,* Huxley [4] has proposed a classification of active principles which affect growth into (a) local activators and (b) distance activators or hormones. The latter group then is further classified into (1) diffusion hormones and (2) circulating hormones. Goldschmidt [5] went even further, and referred to the hypothetical masculinizing and feminizing substances which have been postulated to account for the facts of sex in insects, as "intracellular hormones." "These would thus differ from the ordinary chemo-differentiators of vertebrate embryology in being produced in every cell of the body according to its chromosome constitution, but would resemble them fundamentally in being produced by the same cells on which they exert their specific effect. When, however, the characters concerned are non-sexual, there are of course parallels between vertebrates, insects and other organisms such as plants, for in all these there may exist sharply marked mosaic patches of tissue produced in a heterozygote by the accidental elimination of a chromosome carrying a dominant gene." [4]

Endocrinologists, however, have limited the term *hormone* to an active principle produced by a specialized tissue, an endocrine or internally secreting gland, which delivers its secretion directly to the blood stream for transportation to other parts of the body where its specific physiologic effects are produced. (This does not imply that hormones do not also diffuse through the tissues.) There is no sound reason why a substance which stimulates plant growth or vegetative reparative processes

* This statement is not intended to imply that para-aminobenzoic acid is a vitamin. Its status, in this respect, is uncertain.

should be regarded as a hormone; such substances either should be termed plant stimulants or plant activators. They are related in a way to compounds occurring in the animal body which will be referred to subsequently as chemical activators.

Chemical activators are compounds of low molecular weight and of simple chemical structure and, in general, are formed at the site of reaction. The primary difficulty in their detection and isolation is due to their occurrence in the tissues in extremely minute amounts. The ones best understood are acetylcholine and adrenine; the latter, referred to also as sympathin, may possibly be identical with the most simple of the hormones, adrenalin. These are referred to subsequently in the section dealing with chemical activators.

The vitamins also are compounds of relatively simple structure; some of them have been synthesized and we can anticipate that ultimately all of them will be identified and synthesized. Fourteen vitamins are now known [6, 7] and it is believed that two or more still remain to be discovered.

The vitamins occur widely in nature and are obtained by animals in a haphazard manner with their food. They are not destroyed by the digestive processes and are readily absorbed. They are essential to the proper functioning of the animal body, but fortunately an excess does little or no harm; indeed, in some cases ten to a hundred times the amount of a vitamin necessary for natural functional maintenance has proved to be harmless; in fact, beneficial and specific drug effects have been ascribed to such dosage.[8, 9, 10, 11]

The hormones, on the other hand, are more specific in their mode of action; they are produced within the body itself in exactly the amounts required for normal function and an excess may be as deleterious as a deficiency. Examples of both types of dysfunctions are seen regularly by the endocrinologist in patients displaying hyperthyroidism or hypothyroidism, and hyperinsulinism or hypoinsulinism, while endocrine imbalance secondary to pituitary failure is the basis of the most common of all dyscrinisms. This quantitative distinction between the physiological actions of hormones and vitamins ultimately should throw some light on their respective reaction mechanisms; perhaps the former are primary in type, whereas dysfunctions due to the latter are caused by the absence of complementary factors.

Although much is known about the chemical structure of vitamins and hormones, we have no adequate explanation as to why such structures produce specific physiologic effects and why very slight chemical changes in the active molecules should lead in many cases to complete loss of activity; the lock-and-key analogy may be correct but it discloses no mechanism. There is no question, however, that as our knowledge accumulates, a clearer picture of the mechanism of action will emerge. Most of the references to hormone mechanism in the endocrine literature relate to the physiologic effects produced by the hormones or studies of their metabolic products; but the basic question of *how* the hormones function remains unknown. At this stage, therefore, a certain amount of speculation may be in order, but it should not take the place of continuous laboratory work.

Enzymes are even more complex than either vitamins or hormones; in addition, they are more heat-labile. They appear to act by bringing substances into molecular contact under conditions where chemical changes occur readily. Evidence is available which indicates that some of the vitamins may function as parts of enzyme systems. This is particularly true of some of the B vitamins, but it is probable that the other vitamins and some of the hormones also may be serving primarily as constituents of enzyme systems.

The Endocrine System

The following glands are usually regarded as endocrine structures: anterior pituitary, posterior pituitary, pineal, thyroid, parathyroid, thymus, pancreas, adrenal cortex, adrenal medulla, and the sex glands.

The status of two of the glands in the above list, the pineal and thymus, is uncertain from the standpoint of incretory function, although anatomical alterations in these glands are sometimes noted in such pathological conditions as hypergenitalism and osseous maldevelopment. It must be recognized that no convincing evidence has as yet been produced to ascribe unquestionable endocrine function to these glands. Comparatively, it will be noted that the female possesses an additional gland to those either actual or comparable in the male. This is seen in the case of the ovary, which secretes estrogenic compounds but which in addition contains the corpus luteum, a component part of the ovary which elaborates the hormone progesterone and gives the female gonad a dual performance. In the male the testis is recognized as the origin of the male sex hormones, although a lesser amount of androgens is secreted also by the adrenal cortex. Some workers believe that the prostate may prove to be a gland of internal secretion; however, it remains for further research to present unquestioned evidence upon which such a decision can be based.

We must not confuse the closely related estrogens, androgens and cortical steroids * with multiple hormones such as occur in the anterior and posterior pituitary glands. The latter differ from one another both in chemical and physiologic properties whereas the individuals in a given group of steroidal hormones possess qualitatively the same physiological properties.

In addition to the hormones, or rather groups of hormones, listed above it is recognized that certain active principles (chemical activators) are produced in the upper regions of the intestinal tract. We refer specifically to the principle, *secretin,* which stimulates the flow of pancreatic juice, *cholocystokinin,* the so-called gall-bladder hormone and *enterogastrin,* which influences the flow of gastric juices. These products truly are chemical messengers. They are secreted in one part of the body and control the functions of structures in distant parts of the body; in fact, the word *hormone* was coined by Starling specifically for the product *secretin.* It will be noted, however, that these three activators do not satisfy the full definition for *hormone* as accepted by endocrinologists and as presented above.

An interesting peculiarity of hormone formation may be seen at times in the ability of certain specialized tissues to take over production at certain critical periods in life. This may be illustrated by the characteristic secretion of ovarian hormones by the placenta. During the non-pregnant state, the bulk of estrogenic secretion is maintained by the ovary, a small amount originating in the adrenal cortex. Progesterone is also rhythmically elaborated from the corpus luteum in the non-gravid state. At approximately the end of the fourth month of pregnancy the chorionic elements of the placenta take over the production of estrogens and progesterone, both of which are excreted (the latter in the form of its reduction products) in enormous amounts throughout the remainder of gestation. It is also well known that certain tumor tissues originating within endocrine glands have the ability to secrete hormones in excessive amounts, depending upon the type of neoplasm involved. Such aberrations of growth may take place in either sex at any period of life and during the non-pregnant or pregnant state.

Although there is room for argument regarding the mechanism of hormone action, scientists agree that the enzymes exert their effects through catalytic means. That may be true also of the vitamins.[12] Apparently they act upon a substrate by

* The term *hormomer* is proposed for the multiple active principles which are produced by a given endocrine structure and which are closely related chemically and physiologically. Thus, the six active steroids which have been isolated from suprarenal cortex extracts are hormomers.

A related term *vitamers* has been found useful by workers in the vitamin field. Vitamins D_1, D_2, and D_3 and likewise K_1 and K_2 are vitamers, whereas the various B vitamins, B_1, B_2, etc., are really independent vitamins, each of which corresponds to a series of vitamers, such as B_1 vitamers, B_2 vitamers, biotin vitamers, etc.

forming a temporary adsorption compound, thus producing specific strains upon electronic areas, provided that they correspond to specific electronic areas in the catalyst itself. As a result, specific chemical reactions take place; the adsorption compound then resolves itself (perhaps due to the action of the end-products produced), and the enzyme is regenerated and enabled to exert its catalytic effect upon other molecules. There is considerable evidence that hormones act in a related manner.

An ideal catalyst, even though present only in traces, assists in the chemical conversion of a large amount of material without itself undergoing permanent changes. From this standpoint, and by strict definition, the hormones are *not* ideal catalysts; they undergo chemical changes in the body and are either destroyed or excreted at times in modified form. Extensive research on the metabolism of the hormones is under way in many laboratories; such work is of importance, since undoubtedly it will contribute something to the solution of the problem of how the hormones produce their effects.

Inasmuch as an excessive elaboration of hormones may produce deleterious effects, it is readily apparent that a natural mechanism of hormone destruction is imperative; were such a process not available, one might readily visualize cumulative effects which would in a short time result in impairment of physiologic processes as well as in mutilation of the endocrine system to a point where endocrine salvage would be highly improbable.

A few of the hormones, such as adrenalin, the steroidal sex hormones, and thyroxin have been synthesized,* but the more complex members, such as the peptide and protein types (posterior pituitary, insulin, anterior pituitary and parathyroid hormones) are beyond the scope of synthetic methods now available. There is always the possibility, however, that physiologic activity may reside in an integral part of a complex molecule—the prosthetic group—and that a simpler structure possessing the identical physiologic activity of the hormone might be producible by synthesis.

At the present stage of our knowledge of the endocrine system and its hormones and hormomers, we are likely to overlook the Herculean labors † which have led from one mountain peak to another and have given a glimpse of the foreboding and towering ranges still beckoning to the workers in this field. Extensive and painstaking investigation was required even to prove that certain glands actually *are* endocrine structures. In some instances a detoxifying action was attributed to a gland, and years of research were necessary to prove that the gland actually produces active principles (hormones) which find their way into the circulating blood for transport to distant parts of the body. For years the adrenal cortex and the parathyroids were believed to possess detoxifying properties; instead of contributing a constructive factor to the blood stream they were assumed to destroy toxic factors in the blood as it passed through the glands. An excellent review of that early work will be found in Barker's monumental treatise.[18]

After the demonstration that a given gland produces a hormone, and therefore is an endocrine gland, the questions to be asked are: (1) What physiologic effects are produced by the hormone? (2) Upon which tissues do the hormones act?

* The term *synthesis* is used in its broadest sense. Actually the syntheses involved in the production of the steroidal hormones and hormomers are recognized as "partial syntheses." They have been produced by degradation (and resynthesis) of cholesterol and other complex steroids which themselves have not yet yielded to synthesis. On the other hand one estrogen, equilenin, has been produced by total synthesis (Bachmann, W. E., *et al., J. Am. Chem. Soc.* **62**, 824 (1940).

† The Sixth Labor, the Cleansing of the Augean Stables, is not far removed from the endocrinologist's labor in collecting and isolating hormomer and related compounds from the urine of animals.

(3) What is the mechanism of the actions? The difficulty of answering these questions may be illustrated by referring to the parathyroid glands as an example.

The Parathyroid Glands

After early goiter operations (thyroidectomy) the symptom-complex known as tetany was at times encountered, and finally it was discovered that tetany was due not to the removal of the thyroid gland itself but to the simultaneous and accidental removal of four minute glands (the parathyroids) which are closely associated with thyroid tissue and usually embedded therein. The symptoms of tetany were associated with the action of toxins, and it was logical to assume that the parathyroids might exert their effect by destroying such toxins. Subsequently, however, it was found that tetany was accompanied by a low blood calcium level and that the administration of calcium salts temporarily relieved the tetany. Finally, the work of Hanson [14] and shortly thereafter of Collip [15] demonstrated that parathyroid glands could be extracted by a special technique (strong acid digestion) to yield an active principle which, when administered to parathyroidectomized animals, raised the blood calcium and relieved the tetany.[16]

The work referred to answers at least partially the first of the questions propounded above: What physiological effects are produced by the hormone? We must still determine *where* and *how* the hormone acts; naturally various hypotheses are set up, tested, and discarded or retained on the basis of experiment. Possibly the hormone facilitates absorption of calcium from the intestinal tract; perhaps it prevents calcium loss from the kidney; perhaps it releases calcium from the main calcium depots of the body—the bones; these are just a few of the hypotheses that might be, and actually were, presented. The last-mentioned is accepted at the present time. The next question is: How does the hormone act? This question regarding mechanism is more complex—it may involve, as will be seen, a series of questions. In addition, Nature may have provided *several* mechanisms. In the case of calcium metabolism it is known that vitamin D is involved also. At this stage workers speculated whether vitamin D was merely an intermediate from which the parathyroid hormone is synthesized in the body, as iodine is utilized in the synthesis of the thyroid hormone.

It is known now that the parathyroid hormone and vitamin D have independent and opposite functions although both relate to calcium metabolism. Under normal conditions both calcium deposition and calcium absorption from bone are going on at all times. Such dual (dynamic) control over essential physiological functions is common in the endocrine field and will be referred to again.

But *how* does the hormone exert its effect in maintaining a constant calcium level in the blood? The generally accepted view is that it acts indirectly; its action is on phosphorus metabolism, and the changes in the calcium level are dependent on changes in the metabolism of phosphates.[16] Thus one must ask next, How is the phosphate level controlled? And if the answer is that the kidney is involved, the next question regarding mechanism must be, How does the parathyroid hormone affect the tubules or other parts of the kidneys in producing the final result? The observed facts have been well presented by Albright.[17] "The theory to which I adhere is based on the hypothesis that the hormone affects phosphates in the circulating body fluids in such a way that their excretion in the urine is increased. This would explain the immediate hyperphosphaturia on administration of parathyroid extract and the resulting hypophosphatemia. Furthermore, because of the lowered level of serum phosphorus, it is contended that the serum would be less saturated with respect to calcium phosphate and there would be an increased tendency for calcium phosphate to enter the serum from the gastrointestinal tract or from the bone. This would lead to hypercalcemia, and finally the hypercalcemia would lead to hypercalciuria." [18]

The Anterior Pituitary

"The most important endocrine gland is the anterior pituitary. It occupies that exalted position because it controls the other endocrine glands. In popular language it has been called 'the leader of the endocrine orchestra.'"[19]

Five anterior pituitary hormones are generally recognized. They are (1) the growth (somatotropic) hormone; (2) the gonadotropic hormone; (3) the thyrotropic hormone; (4) the corticotropic (adrenotropic) hormone; and (5) the luteotropic (lactation) hormone. The growth hormone[20, 21] seems to be a specific principle essential to general body growth (skeletal, soft tissues and viscera). It can be considered as a general cellular stimulant. The other hormones listed are specific stimulants for individual endocrine glands, as their names indicate.

Although claims have been advanced in favor of the existence of two gonadotropic hormones, a follicle-stimulating and a luteinizing hormone,[22, 23, 24, 25] the existence of the latter may be questioned. The effect on the corpus luteum may well be due to the presence of the growth and lactogenic hormones; indeed the recent work of Astwood[26] has shown that the principle previously designated as the anterior pituitary lactogenic hormone is a luteotropic factor. Its function is to stimulate and maintain the corpus luteum; its effect in causing lactation in animals under certain conditions may be merely a corollary of its luteotropic effect. Although no adequate proof has been submitted in favor of the existence of an anterior pituitary parathyroid-stimulating hormone, a posterior pituitary-stimulating hormone and an adrenal medulla-stimulating hormone, there is no reason why the "endocrine leader" should direct five members of the "orchestra" and not the others. The work of Houssay[27] has definitely shown that the anterior pituitary exerts a striking effect on carbohydrate metabolism. He showed that although the surgical removal of the pancreas alone will cause diabetes, if the anterior lobe of the pituitary is also removed the animal has no glycosuria and stays in reasonable health. It is clear, therefore, that the anterior pituitary secretes a hormone with an opposite effect on sugar metabolism to that of insulin, and that we have here another instance of dual control of a metabolic process.

Mechanism of Action of Anterior Pituitary Hormones

We are concerned primarily not with the number of these multiple hormones or their physiological effects, but rather with the mechanism whereby a single gland can produce profound effects on every part of the body. It may appear strange that a gland with only two types of secreting cells[28] can produce five (and perhaps eight or more) separate hormones. That fact, however, need not disturb us, since there seems to be no end to the multitudinous reactions carried on in the liver with the aid of only one type of cell.[29] Possibly the growth hormone is produced in one type of cell and all the glandulo-tropic hormones in the other.

A principle that might be accepted as a working hypothesis is this: When multiple hormones are produced by an individual gland they will prove to be closely related in chemical structure. One would not expect a given manufacturing plant to produce threshing machines and microscopes with the same equipment. In the case of the posterior pituitary gland where multiple hormones were first demonstrated, it was found that the two principles were both sulfur-containing peptide-like compounds of approximately the same molecular weight.[30] Similarly, all the anterior pituitary hormones are protein-like in nature. That they are released directly into the circulation as the blood passes through the pituitary is more than an assumption, since Friedgood recently has supplied objective evidence.[31] Once in the circulation, hormones pass successively through the various endocrine glands, each of which has the power of attracting one of them. Thus the thyroid retains the thyrotropic hormone; the gonad, the gonad-stimulating hormones; the adrenal cortex the adrenotropic hormone; etc. There is only one reasonable mechanism by

which such a result can be explained, namely, adsorption of the protein-like hormone on a protein constituent in the gland. This, after all, is analogous to chromatographic adsorption; similar methods are being used in biochemical laboratories to separate the active principles from crude extracts containing multiple hormones and multiple B vitamins.

The endocrine gland can be considered as a manufacturing plant and the glandulotropic hormone as the specialized machinery that is needed for the manufacture of the true hormone. Although the next step in the mechanism needs clarification, it obviously is catalytic, the adsorption compound acting as an enzyme to convert suitable intermediate compounds into the final hormone. In the case of the steroidal hormones (male, female, and adrenal cortical) the intermediate either is cholesterol or simpler intermediates produced in the liver. The evidence in favor of the cholesterol theory of hormonogenesis is almost overpowering. There are 256 possible optical isomers of cholesterol, only one of which occurs in the body. It is observed, however, that in every instance the steroidal hormone produced by a gland corresponds in structure to the one specific cholesterol found in the animal body.

One other step in the hormone mechanism must be mentioned here. After the true hormone has been synthesized and is poured into the blood stream, it in turn checks the activity of the pituitary and thus prevents the development of a vicious cycle. Hormones introduced extraneously also produce such effects; indeed, that is the basis for some of their clinical uses.

Adrenalin and Chemical Activators

The mechanism of the action of adrenalin is better understood than that of any of the other hormones because it is simple; its effect is distinctly adrenergic. Adrenalin acts directly on the tissue cells as a chemical activator, and such well known pharmacological effects as vasoconstriction, acceleration of the heart, dilatation of the pupil, inhibition of tracheal and bronchial musculature, etc. are thus more readily understood.

Further consideration of the manner in which hormones exert their effects should logically be prefaced by a presentation of the known facts regarding autonomic function,[82] and such knowledge is a prerequisite for some of our speculations to be deduced from our fragmentary information. Definite evidence has been accumulated indicating that the cytoneural junction is the scene of elaboration of chemical substances which affect cellular reactivity following a nervous impulse. It has long been recognized that injections of adrenalin produce effects in most instances identical with those elicited by electrical stimulation of the sympathetic nerves. It is now postulated that this effect of sympathetic stimulation is due to the formation of either adrenalin or a chemical substance closely related to adrenalin at the effector terminal of the preganglionic neuron. Inasmuch as this compound has never been isolated in crystalline form, and since adrenalin may exhibit both exitor and inhibitor effects, this substance has been designated by Cannon [88] by the convenient term *sympathin,* reserving the terms sympathin E and sympathin I for the effects which either are exitatory or inhibitory in nature.

Since functionally the sympathetic and para-sympathetic nervous systems are antagonistic, it would seem reasonable to postulate the elaboration of a pharmacodynamically opposite substance from the distal terminals of the first and second para-sympathetic neurons. Such a substance actually has been localized after parasympathetic nerve stimulation and identified chemically as acetylcholine. Investigations conducted by Dale, Loewi, and others over a period of years lend credence to such a concept.[88]

An explanation of the transitory effect of these chemical substances following a single nerve impulse has been shown, in the case of acetylcholine, to be due to the hydrolytic effect of an ever-present enzyme known as cholinesterase; no similar

enzyme has been demonstrated which is responsible for immediately nullifying the sympathin effect. Green and Richter [34] have emphasized the readiness with which adrenalin is auto-oxidizable *in vitro*. On the basis of this fact, they point out that the assumption is often made that such a process is responsible for the destruction of adrenalin *in vivo*. They draw attention to the fact, however, that the tissues contain an inhibitor which is very efficient in protecting adrenalin from auto-oxidation, citing the work of Wiltshire,[35] Welch,[36] and Heard and Welch.[37]

Green and Richter have suggested a mechanism by which the sympathin effect might be nullified in a manner similar to the destruction of acetylcholine by the choline esterase system. They present evidence to show that adrenalin is rapidly oxidized by cytochrome and by the indophenol oxidase-cytochrome system. The resulting oxidation product is known to be adrenochrome, which is an active carrier of oxygen but is not adrenergic. Since this indophenol oxidase-cytochrome system is known to be present in most animal tissues, they suggest that the property of inactivation from this system might account for the transitory sympathin effect. These workers also point out that the sympathomimetic drugs which are most rapidly destroyed *in vivo* prove to be the most rapidly oxidized by the indophenol oxidase-cytochrome system.

Dale [38] has suggested that chemically these effects of sympathetic and parasympathetic nerve stimulation should be designated either as adrenergic or cholinergic, depending upon whether the sympathetic or para-sympathetic system is being stimulated; thus, pilocarpine is distinctly cholinergic in effect, whereas ephedrine obviously is adrenergic. Parker [39] has also contributed to our knowledge of these neurohumeral substances.

The Sex Hormones

While the above information has enabled investigators to explain some of the effects produced by hormone-like substances, the mechanisms of most of the known end-points of hormonal influence are still obscure. Actual change in tissue morphology is one of the most important functions of the steroidal hormones. Hyperplasia of accessory sex organs in both sexes is a well known process which occurs concomitantly with the utilization of estrogenic or androgenic hormones.* How can the reasons for such tissue growth be postulated? Shall the presence be assumed of highly specific chemical stimulators, perhaps intracellular in location, which are activated by these steroids and which account for cellular mitosis? Such a viewpoint at this time must be considered purely hypothetical.

On the other hand, how can the actual inhibition of cellular growth and tissue regression be explained—a phenomenon which may be easily demonstrated in the average test object following the administration of some of these same steroids? Pituitary inhibition following large injections of estrogenic substance may be cited as an example. The presence of intracellular growth factors which operate locally and by diffusion has been suggested, and they have been erroneously referred to as hormones. It might logically be assumed that the causes of cellular growth or changes in tissue morphology are intimately connected with the cause of life itself. These changes in anatomical structure are the most obvious and important end-points of estrogenic utilization, and these well-known effects apply primarily to the utero-vaginal tract and certain elements of the breast.

From the above facts, it is obvious that there is no explanation for the exact mechanism responsible for this growth; but in connection with known facts concerning autonomic function, striking evidence is available which seems to clarify the mode of action of estrogenic substances on some portions of the vascular tree. Reynolds,[40] working with rabbits in a well controlled experiment, has shown that

* Cancer of the prostate is now being successfully treated by administration of estrogens (estrone, stilbestrol) or the removal of androgens (by castration). J. A.

the endometrial and myometrial hyperemia which follows injections of estrogenic hormone is due to an increase in tissue acetylcholine. In untreated controls the acetylcholine content was found to be very low, and he arrived at the conclusion that these estrogenic substances are strictly cholinergic in nature, accounting for the vasodilatation and increase in muscular activity. A later study by Reynolds and Foster,[41] however, is somewhat confusing, for by the same experimental approach they showed that injections of stilbestrol, although highly estrogenic, did not significantly alter the acetylcholine content of the uterus even when high dosages were used. Inasmuch as stilbestrol is a member of a different chemical group, the stilbenes, it is possible that only estrogens possessing a steroidal structure will exhibit cholinergic gradients.

Water retention in some tissues, particularly the skin, following estrogens may be accounted for by an increase in capillary permeability due to this cholinergic effect of the estrogens; but changes in calcium metabolism following estrogens are not readily explained. The increased calcium deposition in cancellous * bone following the administration of estrogens in some experimental animals is well known, but the cause for this derangement still remains unknown.

The mode of action involved in the inhibition or stimulation of the anterior lobe of the pituitary by estrogens likewise is unexplainable; these phenomena may be associated with cholinergic stimulation of some of the secreting cells of the gland, which would require a highly specialized mechanism.

Progesterone physiologically is responsible for secretory changes in the cells of the endometrial glands, and perhaps this effect is due to cholinergic stimulation. Although such has not been proved, it is assumed that progesterone is cholinergic and thereby secretory to the endometrial glands.

Steroids possessing androgenic activity are similar to the estrogens in that their end-points are characterized primarily by their effects on tissue growth; these male sex hormones are responsible for growth and physiologic development of the secondary sex characteristics, and hyperemia of these organs may be explained on a similar basis to that offered for the estrogens, namely, cholinergic effects.

Hormones of the Adrenal Cortex

The hormones of the cortex exert their effects in three spheres: first, the salt and water balance of the body, second, the carbohydrate metabolism, and third, the sexual "make-up," particularly the secondary sex attributes.

Of the 24 steroids[42, 43, 44, 45] which are unique constituents of the adrenal cortex, six are known to have the basic cortical action, namely that of maintaining life in an adrenalectomized animal. They are corticosterone, 11-dehydrocorticosterone, 17-hydroxycorticosterone, and 17-hydroxy-desoxycorticosterone. Of the remaining 16, three have masculinizing activity of varying degree; adrenosterone, 11-hydroxy-isoandrosterone and 17-hydroxy-progesterone. It should also be mentioned that the adrenal cortex has yielded progesterone and estrone.

The actual mechanics of how the cortex hormones influence the salt and water balance is obscure, but the bulk of the evidence indicates that they may act by influencing the specific reabsorption of sodium from the glomerular filtrate as it traverses the ascending loop of the kidney nephron. In the absence of the cortex hormones, sodium and water are thus lost to the body and the vicious cycle of dehydration and circulatory collapse is initiated. That the mere withdrawal of sodium and water is not the really basic effect is evidenced by the fact that clinical improvement can be effected without change in the salt or water content of the body. The other changes in the blood and tissues, such as the blood levels of potassium, chloride, etc., can be traced to the dehydration. There has been a great deal of speculation concerning the effect of the hormones on capillary permeability, but the careful work

* Of a reticular, spongy, or lattice-like structure.

of Muntwyler [46, 47] and his associates demonstrates that the Donnan equilibrium for semi-permeable membranes is maintained in the deficient animal (dog), making any large change in the permeability characteristics of the cellular membranes appear improbable. That the mere loss of salt itself is not the lethal factor is evidenced by the fact that interrenalectomy in the shark is fatal, even though it is maintained in sea water.

The cortex hormones have a profound effect on carbohydrate metabolism, but here again nothing is really known as to how they act. They are concerned primarily with the maintenance of the liver glycogen reserve, that is, gluconeogenesis, and even in this there is great variability in the different species, the carbohydrate disturbance being much more pronounced in the rat than in the dog. Influence on carbohydrate metabolism appears to be related to the structural feature of an oxygen atom on the eleven carbon atom of the steroid skeleton. The muscular fatigue which is a cardinal feature of adrenal insufficiency is perhaps related to the faulty carbohydrate metabolism, but it may be equally referable to the disturbance in the salt and water balance. In fact the cortex hormones may be the key to the problem of the relationship of carbohydrate metabolism to the salt and water balance.

In the sexual sphere the adrenal cortex probably acts in conjunction with the sex glands. An overactivity of the cortex can cause either masculinization or feminization. Presumably these end effects are obtained by mechanisms common to the sex hormones themselves, and represent merely an excessive secretion of sex active compounds from the cortex.

The Action of Insulin

Many workers have offered individual concepts to explain the physiologic state which occurs on removal of the pancreas and the symptoms seen in diabetes mellitus. A concise review of the subject has been set forth by Cameron.[48] The two principal theories to explain this state of hypo-insulinism are known as (a) the non-utilization theory and (b) the overproduction theory. Adherents to the former believe that in the absence of insulin the capacity of tissues to metabolize glucose is nearly or completely lost. There is a resulting increased protein and defective fat metabolism from this inability to convert glucose into its end products. This failure of glucose metabolism results in the characteristic hyperglycemia and glycosuria. Another theory explains that hyperglycemia occurs as a result of over-production of glucose in the liver from noncarbohydrate sources such as protein and fat. Hyperglycemia and glycosuria, according to such a concept, are not due to failure of utilization, but to glucose over-production. It may be seen from the foregoing that an explanation of insulin action will vary somewhat, depending on which of the two theories is accepted, but in general there seems to be adequate proof that the following functions may be attributed to insulin. Insulin evidently increases the *rate* at which glucose is converted to glycogen in the muscles and other tissues. Although insulin is probably not specific in this regard, it has been shown by Gemmill [49] to act *in vitro* in muscle tissue cultures.

Insulin prevents loss of liver glycogen probably because of its inhibitory effect on hepatic glycogenolysis caused by adrenal hormones. Insulin also exerts a certain amount of regulation of oxidation in tissues, that is, it plays the role of arbitrator as to relative amounts of carbohydrate, fat and protein oxidized. Insulin almost certainly decreases the rate of gluconeogenesis from protein. It also prevents formation of ketone bodies from fats. The latter two actions occur in the liver. Insulin both anatomically and functionally prevents and repairs damage to islets caused by anterior pituitary extracts.[50] There is no clear-cut evidence that insulin influences oxidative mechanisms either directly on metabolites or indirectly on enzyme systems involved in oxidation.

The relative importance in carbohydrate metabolism of the functions of insulin

and of non-pancreatic factors such as the influence of the pituitary, adrenal and thyroid is highly controversial.

The Posterior Pituitary and Water Metabolism

Several of the hormones exert a profound effect on metabolism. Thus the thyroid hormone and the hormomer, thyroxin, affect protein metabolism, insulin controls sugar metabolism, and the parathyroid hormone calcium and phosphate metabolism. Even the utilization of water is under hormonic control—it is the main function of one of the hormones of the posterior lobe of the pituitary gland. The amount of water that we imbibe may be under our own control, but what happens to it after absorption—whether it is promptly eliminated or whether it is retained and stored in certain tissues—is partly under the control of the antidiuretic hormone of the posterior pituitary.

As has been indicated in several of the preceding sections, various processes occurring in the body are under the control of *several* different hormones. Enzymes usually control equilibrium reactions; hormones on the other hand, are absolute monarchs—they control processes which are irreversible. Nature, therefore, has not trusted these basic processes of life each to one individual hormone; it has instituted a system of checks and balances—usually several hormones are concerned in controlling a given mechanism; even vitamins have a part in this democratic system.

It has been shown above that the parathyroid, in cooperation with vitamin D, controls calcium metabolism and that the pancreas and the anterior pituitary are concerned (and that both are major factors) in the control of carbohydrate metabolism; however, even this dual control is not adequate—some powers have been relegated also to the adrenal cortex. One of the functions, perhaps the most important, of the adrenal cortex is to assist in the control of carbohydrate metabolism. It is not surprising, therefore, that the mechanism of action in relation to sugar metabolism still is confused. Similarly, water metabolism is too important a matter to be left to the control of one gland; in addition to the posterior pituitary, one of the hormones of the anterior pituitary, and incidentally one not yet effectively extracted and concentrated, is concerned with water metabolism. Moreover, certain hormoners of the adrenal cortex influence water metabolism profoundly, and secondary effects are exerted by the thyroid and the pancreas. Any gland (or hormone) concerned with the metabolism of large molecules and the production from them of compounds of low molecular weight which exert a considerable effect on osmotic pressure in the tissues, will affect water metabolism.

A striking demonstration of the dual control of water metabolism of the posterior and anterior pituitary glands can be made by surgical means. If the posterior pituitary is removed, a condition of diabetes insipidus is produced; an excessive loss of water occurs and the animal suffers from extreme thirst and consumes an enormous amount of water. A condition corresponding to this is met occasionally in medical practice—not necessarily as a result of surgery—although fortunately it is rare; only approximately a thousand patients in our 120 million population suffer in an extreme way from such a defect. Although the condition is very inconvenient, it is not fatal or even dangerous. Fortunately, it is controlled definitely by administration either of posterior lobe extracts or of the antidiuretic hormone "Pitressin." A striking phenomenon in connection with diabetes insipidus is the fact that the kidney effectively conserves (reabsorbs) needed constituents and the urine excreted resembles distilled water. The insipid taste of distilled water in contrast to that of ordinary drinking water ("hard" water) is responsible for the medical term, diabetes insipidus. Unfortunately no reference can be located to the observation of the physician who first tasted the urine of such patients and found it, in contrast to that from those suffering from diabetes mellitus, to be insipid. Diabetes mellitus, on the other hand, is characterized by the elimination of an appreciable amount of sugar. Such urine

tastes like honey; the presence of sugar was noted in Claude Bernard's laboratory, where it was observed that flies were attracted by urine excreted by pancreatectomized dogs. However, nearly three centuries ago Thomas Willis observed such urine to be "wonderfully sweet, as if imbued with honey or sugar." [51]

After diabetes insipidus has been produced in an experimental animal by surgical removal of the posterior lobe, a second operation—the removal of the anterior lobe—corrects the condition. The experiment [52, 53] demonstrates that the anterior lobe *also* affects water metabolism and probably produces a hormone which effectively balances the water-metabolizing hormone of the posterior lobe.

We are concerned here with the mechanism whereby the posterior lobe hormone produces its effect. In this case, fortunately, definite theories and facts can be presented.[54]

The posterior pituitary is the first gland in which the presence of multiple hormones was demonstrated [55]—a pressor-antidiuretic hormone and an oxytocic hormone. Although some claims have been made to the effect that the pressor effect is due to the presence of a separate factor,[56] all the work done in Kamm's laboratory * indicates that a single hormone is involved. This is not an unusual situation; even from as simple a hormone as adrenalin we note various physiologic responses—effects on blood pressure, on the bronchioles, on carbohydrate metabolism, etc. Why should the antidiuretic hormone of the posterior pituitary be an exception to this basic principle?

The possibility that man has descended from sea-living animals is supported by the fact that the composition of his blood is related chemically to that of sea water. Moreover, it is an axiom of physiology that the composition of the blood is constant—that constancy is maintained partly by the glomeruli and tubules of the kidney. Enormous quantities of water are metabolized—possibly 40,000 cc per day—and yet man needs to drink and obtain in his food only approximately a liter of water per day. The explanation is that water is reabsorbed by the tubules of the kidney and is utilized again and again. An estimate of the quantities of water involved in life processes in an average case is illustrated in Table 1.[57]

Table 1. Intermediate Water Exchange (Human Subject)

Secretion		Amount (cc.)
Saliva		1500-(5000?)
Gastric juice		2000- 3000
Pancreatic juice		500- 800
Bile		300- 500
Succus entericus		2000- 3000
Kidney reabsorption		10000- ?
Average water intake (including food)		1700 cc
Average Excretion	Urine	800
	Feces	200
	Respiratory Tract and Skin	700
	Total	1700 cc

The *facts* concerned in water metabolism are important and obvious, but it seems desirable to seek the mechanism whereby the antidiuretic-pressor hormone of the posterior pituitary produces its striking effect. If that can be found, a clew may be available also as to how other endocrine principles exert their effect on water metabolism. It does not follow, however, that the reaction mechanisms are identical; in fact Nature's safeguard lies not only in multiple control of the process but

* The work is being assembled for publication.

in varying mechanisms. The posterior pituitary theory of mechanism is presented herewith.

It is well known that body fluids are rich in sodium and poor in potassium. The latter ion is essential for the heart mechanism, but an excess is fatal. On the other hand, body cells are rich in potassium and low in sodium; for example, if the potassium in the red blood cells were released even partially, death would result. Fortunately, Nature has guarded against such an emergency by providing the red cell (and other cells also) with a specialized (impermeable) membrane which prevents sodium ions from diffusing into the cell and potassium ions from diffusing out.[58]

Many theories have been advanced to explain the effect of posterior pituitary extracts upon water metabolism, since it must be intimately associated with salt metabolism. Such theories, or rather hypotheses, cannot be reviewed here; it is sufficient to state that the most common one is to ascribe the effect to a kidney mechanism. If the kidney controls the sodium and potassium levels of the blood (and of the anions as well), then the blood and lymph levels are controlled by the same mechanism, since the capillaries are truly semi-permeable membranes such as the chemist deals with in his laboratory, and osmotic relations will furnish the rest of the picture. Such a theory, however, overlooks the specialized membrane possessed by most of the body cells.

A more logical mechanism appears to be the following: The antidiuretic hormone acts upon the specialized membranes of the body cells (those of the muscles especially) and modifies them *slightly* in the direction of true dialyzing membranes. The effect may be due mainly to changing the type of emulsion or colloid present in the cell membrane. This effect is mediated through a nervous control—that is why the red cells, freely floating in the blood, are not influenced. As a result of the change in the distribution of electrolytes, a change in water concentration will follow. A relatively minute change in salt content of the cell can effect an enormous change in water transport. The effect of the posterior pituitary antidiuretic hormone, therefore, is to hydrate tissue cells. The anterior pituitary probably exerts a similar effect, but in the opposite direction. The net result following a dose of vasopressin ("Pitressin") is to cause a flow of water to the tissue cells (provided an excess of water is available for transportation) and to cause increased salt (especially potassium) excretion by the kidney. Long before this theory of reaction mechanism was presented, Stehle[59] had shown that large doses of posterior lobe extract increased greatly the excretion of potassium, thus substantiating the theory.

Conclusion

In the chapter entitled "Catalysis as a Biological Factor," the statement is made that "though much is known as to the *effects* produced by hormones, little is known as to the basic *mechanism* whereby these effects are produced." Possibly the survey presented in the present chapter substantiates that conclusion. It is believed, however, in spite of the lack of progress that has been made up to this time, that research in the field of hormone mechanisms offers a fruitful range of work. At a recent Hormone Symposium,[60] practically nothing was mentioned concerning this aspect of endocrinology. It is predicted, however, that in the future, hormone symposia will be held in which most of the papers will contribute to this important field. Possibly, as Jeans[61] points out, from the nature of things science can never attain to a knowledge of the realities of the world. He states, "Our sense-organs form a sort of screen on which nature's lantern is forever projecting pictures; we can study these pictures, but can never pass behind the screen to see how the lantern works. And as the sequence of pictures thrown on the screen is all that we can ever know or study, a set of laws that links the pictures together in a perfect systematic order is the most that science can provide, or can properly be asked to provide."

As has been pointed out, an investigation of the mechanism of hormone action aims to answer the question How?; perhaps we can never answer the question Why? Perhaps we can never look behind the screen and obtain a glimpse of the lantern, but still there is no reason to be discouraged. Great discoveries await workers in this field and at times, no doubt, enthusiastic investigators will even attempt to invade the field that Jeans tells us is beyond the screen. They should not be discouraged.

References

1. de Beer, "The Comparative Anatomy, Histology and Development of the Pituitary Body," 1926.
2. Starling, E. H., *Proc. Roy. Soc. Med.*, **7**, 29 (1914).
3. Gley, E., Quoted by S. Vincent in "Secretion and the Ductless Glands," 1920.
4. Huxley, J. S., *Biological Rev.* (Cambridge Phil. Soc.), **10**, 427 (1935).
5. Goldschmidt, R., "Physiologische Theorie der Vererbung," 1927.
6. Pfiffner, J. J. *et al.*, *Science*, **97**, 404 (1943).
7. Hogan and Kamm, *Science*, **97**, 353 (1943).
8. Holmes, H. N., and Alexander, W., *Science*, **96**, 497 (1942).
9. Maisel and Sompkin, *J. Allergy*, **13**, 397 (1942).
10. Klassen, K. P., and Curtis, J. M., *Arch. Int. Med.*, **71**, 78 (1943).
11. Queries & Minor Notes, *J. Am. Med. Assn.*, **122**, 839 (1943).
12. Bernheim, F., *J. Lab. Clin. Med.*, **26**, No. 1, 20-30 (1940).
13. "Endocrinology and Metabolism," 1922.
14. Hanson, A. M., *Mil. Surgeon*, **52**, 434 (1923); **54**, 76 (1924); **54**, 218 (1924); **55**, 701 (1924).
15. Collip, J. B., *Am. J. Physiol.*, **72**, 182 (1925); *J. Biol. Chem.*, **63**, 395 (1925).
16. Hanson, A. M., *J. Am. Med. Assn.*, **90**, 747 (1928).
17. Albright, F., and Ellsworth, R., *J. Clin. Invest.*, **7**, 183 (1929).
18. "Glandular Physiology and Therapy," American Medical Assn., p. 445, 1942.
19. Langdon-Brown, W., *Practitioner*, **127**, 614 (1931).
20. Smith, P. E., "Glandular Physiology and Therapy," p. 6, 1942.
21. Evans, H. M., *Ibid.*, p. 19, 1942.
22. Fevold, Hisaw and Leonard, *Am. J. Physiol.*, **97**, 291 (1931).
23. Wallen-Lawrence, Z., *J. Pharmacol.*, **51**, 263 (1934).
24. Evans, H. M., *et al.*, *Univ. Calif. Pub. Anat.*, **1**, 255 (1936).
25. "Sex and Internal Secretions," p. 966, 1939.
26. Astwood, E. B., *Endocrinol.*, **28**, 309 (1941).
27. Houssay, B. A., *N. E. J. Med.*, **214**, 961 (1936).
28. Sevringhaus, A. E., *Physiol. Rev.*, **17**, 556 (1937).
29. Friedgood, H. B., Argument presented at A.A.A.S. Hormone Symposium, July, 1943. Referred to with his permission.
30. Kamm, O., *Science*, **67**, 199 (1928).
31. Personal communication, A.A.A.S. Hormone Symposium, July, 1943.
32. Werner, A. A., "Endocrinology, Clinical Application and Treatment," 2nd Ed., 1942.
33. Cannon, W. B., *Science*, **78**, 43 (1933). See also paper by W. B. Cannon in this volume.
34. Green, D. E., and Richter, D., *Biochem. J.*, **31**, 596 (1937).
35. Wiltshire, M. O. P., *J. Physiol.*, **72**, 88 (1931).
36. Welch, A. De M., *Am. J. Physiol.*, **108**, 360 (1934).
37. Heard, R. D. H., and Welch, A. De M., *Biochem. J.*, **29**, 998 (1935).
38. Dale, H., *Edin. Med. J.* (*N.S.*), **45**, 461 (1938).
39. Parker, G. H., *Science*, **81**, 279-83 (1933); **91**, 455 (1940).
40. Reynolds, S. R. M., *J. Physiol.*, **95**, 258 (1939).
41. Reynolds, S. R. M., and Foster, F. I., *Am. J. Physiol.*, **128**, 147 (1939-1940).
42. Reichstein, T., "Vitamins and Hormones," pp. 345-413, Vol. I, Acad. Press, New York, 1943.
43. Wintersteiner, O., and Pfiffner, J. J., *J. Biol. Chem.*, **116**, 291 (1936).
44. Kendall, E. C., *Ann. Rev. Biochem.*, **10**, 285 (1941); *Endocrinology*, **30**, 853 (1942).
45. Pfiffner, J. J., and Kamm, O., *Ann. Rev. Biochem.*, **11**, 283 (1942).
46. Muntwyler, E., Mellors, R. C., and Mautz, F. R., *J. Biol. Chem.*, **134**, 345 (1940).
47. Muntwyler, E., Mellors, R. C., Mautz, F. R., and Mangun, G. H., *J. Biol. Chem.*, **134**, 367 (1940).
48. Cameron, A. T., "Recent Advances in Endocrinology," 2nd Ed., Blakiston, 1935.
49. Gemmill, C. L., *Bull. Johns Hopkins Hosp.*, **66**, 232 (1940).
50. Houssay, B. A. and Biasotti, A., *Comp. rend. soc. biol.*, **105**, 121 (1930); Arch. ges. Physiol. (Pfluger's) **227**, 239 (1931).
51. Mirsky and Nelson, *Arch. Int. Med.*, **71**, 827 (June, 1943).
52. Richter, C. P., *Am. J. Physiol.*, **110**, 439 (1934).
53. Fisher, Ingram and Ranson, *Arch. Neurol. Psychiatry*, **34**, 124 (1935).
54. Kamm, O., Unpublished work, presented in a presidential address before the Michigan Academy of Science, 1929.
55. Kamm, Aldrich, Grote, Rowe and Bugbee, *J. Am. Chem. Soc.*, **50**, 573-601 (1928).
56. Heller, H., *J. Physiol.*, **96**, 337 (1939).
57. Rowntree, L. G., *Physiol. Rev.*, **2**, 116 (1922).
58. Van Slyke, D. D., "Factors Affecting the Distribution of Electrolytes, Water and Gasses in the Animal Body," Lippincott, 1926.
59. Stehle, R. L., *Am. J. Physiol.*, **79**, 289 (1926-7).
60. A.A.A.S. Hormone Symposium, Gibson Island, July, 1943.
61. *Proc. Roy. Soc. London*, **131**, 237 (1943).

The Molecular Organization of Visual Processes *

GEORGE WALD

Biological Laboratories, Harvard University, Cambridge, Mass.

The purpose of this paper is to discuss briefly certain macromolecular and microanatomical relations which govern the photoreceptor processes in vision. The further chemistry and phylogeny of these processes have recently been reviewed elsewhere (Hecht, 1942; Wald, 1942, 1943); detailed references to the literature will be found in these papers.

The Anatomy of Photoreception

The eye contains three tissues, each of which makes specific contributions to visual metabolism. Proceeding from the interior outward these are (1) the retina proper, consisting in order of a layer of ganglion cells from which spring the fibers of the optic nerve, a layer of communicating neurones, and the photoreceptor layer; (2) the single-layered pigment epithelium; and (3) the choroid, one of the most highly vascular tissues in the body (Fig. 1 A).

Most vertebrate retinas contain two kinds of light receptors: rods, ordinarily concerned with vision in dim light, and in man conveying only neutral gray sensations; and cones, the organs of vision in bright light, detail and color vision. Both cells bear at their extremities dense structures, their so-called outer limbs, to the shape of which they owe their names and to whose photosensitive pigments they owe their reactivity to light.

Outside the class of mammals, few retinas have a blood supply of their own. Even in mammalian retinas the circulation is sparse and in certain areas entirely absent, as in the human fovea, the region of most acute vision. The retina depends quite generally, therefore, upon the choroidal circulation for the supply and removal of metabolites; and these must pass in transit through the intervening pigment epithelium.

The latter is a true retinal layer, derived embryonically from the proximal wall of the ectodermal optic cup, just as the retina proper is developed from its distal portion. Normally both tissues lie in intimate contact, though joined only at the rim of the pupil. Franz Boll (1877) in the paper which announced his discovery of rhodopsin, the photosensitive pigment of the rods, expressed the conviction that the rods and the cells of the pigment epithelium constitute a physiological unit. In the frog, as in other cold-blooded animals, light causes the granules of epithelial pigment to stream into long protoplasmic filaments which project among the outer limbs of the rods and cones; in darkness the granules retire to the basal portions of the epithelial cells. In the light also the retina proper adheres firmly to the pigment epithelium, while in the dark these tissues lie in loose contact. The pigment epithelium of the frog contains golden oil droplets, in the pigment of which Boll thought he had found the precursor of rhodopsin. This shrewd guess narrowly missed its mark; it is now clear that this pigment, the carotenoid xanthophyll or lutein ($C_{40}H_{54}(OH)_2$), does not participate in the rhodopsin system (Wald, 1935-36a, b). Shortly afterward, however, Kühne (1878) was led by observations mentioned below to ascribe to the pigment epithelium a central role in the synthesis of rhodopsin.

* The original researches described in this paper have been supported in part by the Milton Fund of Harvard University, and the Josiah Macy, Jr. Foundation of New York.

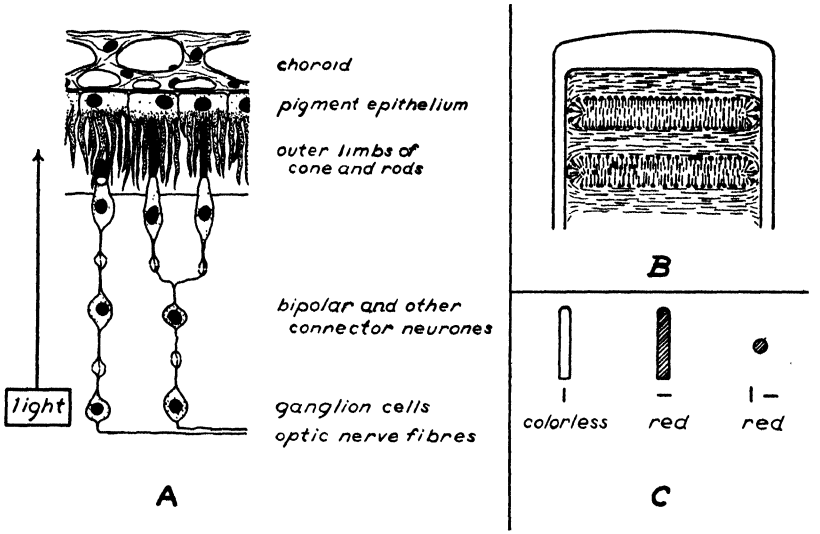


FIGURE 1—A. Arrangement of cell layers in the interior of the eye. The rods and cones, connector neurones, ganglion cells and optic fibres constitute the retina proper. The pigment epithelium, though anatomically distinct, is a true retinal layer.

B. Microstructure of the outer limb of a rod. Within a sheath of neurokeratin are alternate layers of protein and lipids. Rod-shaped protein molecules are oriented transversely; long-chain lipid molecules are principally oriented longitudinally. Many of the latter have their polar groups (shown as small black circles) directed toward or bound to the protein layers. The thickness of both types of layers is here greatly exaggerated; it is actually below the limits of visual resolution (about 200 mμ) and probably of the order of $\frac{1}{25}$-$\frac{1}{50}$ the total width of the rod. (Modified from Schmidt, 1938.)

C. Dichroism of rhodopsin in the dark adapted rod. In light polarized in a plane parallel to the long axis of the rod it is colorless; in the perpendicular plane it is red. In cross-section the rod appears red in all planes of polarization, showing that in this dimension the molecules of rhodopsin are oriented at random. (Modified from Schmidt 1938.)

The outer limb of a rod or cone is a complex structure. It is completely enclosed in a sheath of relatively inert protein identified by Kühne (1878b, p. 40) as neurokeratin. Its interior normally looks entirely homogeneous even under high magnification. But when rods are bathed in water or acid or alkaline solutions, the outer limbs lengthen and attenuate, and simultaneously are divided by cross-striations into the appearance of cylinders filled with discs. The basis of this behavior is a microstructure of alternate layers of protein and of a mixture of lipids—mainly phosphatides—resembling that found in nerve myelin. The swelling of these layers in certain media makes this construction visible. The molecules of both protein and lipids are oriented to a degree, and the optical properties of the outer limb in polarized light have their source largely in this factor (Fig. 1 B) (Schmidt, 1938).

The outer limb is positively birefringent, its single optic axis lying parallel to its long dimension. This property is the resultant of a number of separable components: (a) Strong positive intrinsic birefringence (*Eigendoppelbrechung*) originating in the lipid layers, in which long carbon chains appear to be oriented parallel to the length of the rod; (b) weak positive intrinsic birefringence due to the protein layers, which appear to contain rod-shaped molecules oriented perpendicularly to the length of the rod; and (c) negative structure birefringence (*Formdoppelbrechung*) due to the alternation of these layers which differ in refractive index; that of the

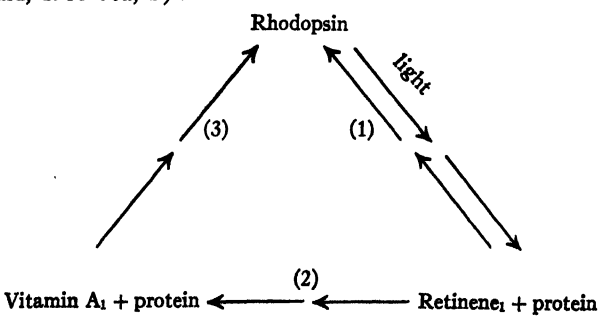
protein is extraordinarily high, about 1.60, while that of the lipids is probably comparable with nerve myelin, about 1.50. It is the transverse orientation of its protein molecules that causes the rod to lengthen and constrict when treated with acids, alkalies or heat. On such treatment most protein fibrils, the molecules of which lie longitudinally, thicken and contract.

It is important to note that the rod, viewed end-on, is isotropic. Its molecules apparently are oriented at random in the cross-sectional plane. The way in which photosensitive pigments enter into this structure is not wholly clear. The rod photopigment rhodopsin, however, is known to be oriented within the rod, since it is dichroic in polarized light. Dark adapted rods lying perpendicular to the plane of polarization look pink, those parallel to this plane look colorless. In cross-section the rod looks pink always, showing that in this plane rhodopsin shares with the total structure of the outer limb a random orientation (Fig. 1 C). Light travelling down the long axis of the rod—its normal direction of penetration in the eye—encounters only isotropic structures. It is for this reason that plane-polarized light has almost no special properties as a visual stimulus.

Retinal Photopigments

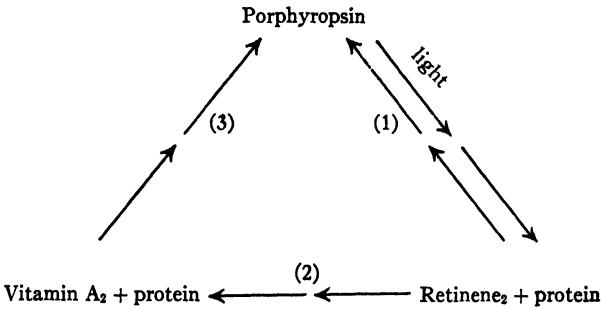
Three retinal photopigments are now known: rhodopsin and porphyropsin in rods, and iodopsin in certain cones. The rod pigments—and probably iodopsin also —represent a conjunction of protein and lipid elements reminiscent of the total structure of the outer limb. They are conjugated proteins which owe their color to the possession of carotenoid prosthetic groups—groups belonging to the class of fat-soluble animal and plant pigments which includes the vitamins A. This composition lends the photopigments a dual character. Like all proteins they have colloidal dimensions, intricate and highly perishable structures, and some tendency toward solution in water. Their carotenoid residues, however, are highly water-repellent, and beyond markedly affecting their solubility and optical properties, involve them in a series of special external relations—the dietary supply and general metabolism of the carotenoids and vitamins A, none of which can be synthesized *de novo* in the vertebrate body.

Rhodopsin is a rose-colored pigment found in the rods of marine fishes, certain amphibia, birds and mammals (Wald, 1942, 1943). When exposed to light it bleaches to orange or yellow products (Kühne's "visual yellow") which yield all their color to neutral fat solvents. The pigment so extracted is the yellow carotenoid retinene$_1$. Following irradiation of dark-adapted retinas, the initial color due to retinene$_1$ or its immediate precursors slowly fades, a visible indication of the conversion of these substances to the colorless vitamin A_1 ($C_{20}H_{29}OH$). This is an ordinary thermal reaction. It predominates even in darkness, though here in addition a little rhodopsin is also regenerated. In isolated retinas vitamin A_1 and protein are the final products of bleaching. But in the intact eye these substances in turn are resynthesized to rhodopsin. The system as a whole therefore constitutes a cycle of the form (Wald, 1935-36a, b):



Isolation of the retina cuts the cycle at (3). Extraction of rhodopsin into solution virtually eliminates also reactions (1) and (2), leaving only the bleaching of rhodopsin to a mixture of retinene$_1$ and colorless protein.*

Porphyropsin is a purple photopigment found in the rods of fresh-water fishes, cyclostomes and certain amphibia. It is linked genetically throughout the vertebrate phylum with the capacity for fresh-water existence, particularly with a fresh-water embryonic history (Wald, 1942, 1943). Its general chemical behavior and the form of its retinal cycle are identical with those of rhodopsin. The absorption spectrum of each component of the porphyropsin system, however, is displaced 20-30 mμ toward the red from that of its analog in the rhodopsin cycle. The basis of this difference is the presence in the porphyropsin system of new carotenoids, called at present retinene$_2$ and vitamin A$_2$. It therefore has the composition (Wald, 1937a, 1938-39b):



Just as in the rhodopsin system, isolation of the retina cuts this cycle at (3); and extraction of porphyropsin eliminates also reactions (1) and (2).

The structure of vitamin A$_2$ is not known. Within the class of carotenoids, however, a shift of spectrum 20-30 mμ toward the red is known to accompany the addition to the polyene chain of one conjugated double bond. Such an added double bond appears to be the sole difference between the porphyropsin and rhodopsin systems. This small change explains not only the constant spectral displacement which involves all members of both cycles, but their extraordinary parallelism in behavior, since this change should only slightly affect their chemical properties.

Little is yet known of iodopsin (Wald, 1937b; Chase, 1938). Its idiosyncracies in solubility and its optical properties suggest close chemical relationship to the photopigments of the rods. Its spectrum is shifted some 40 mμ toward the red from that of porphyropsin, and by analogy one might suppose that it contains a still longer conjugated polyene group. In human night-blindness, due to lack of vitamin A, the threshold of cone vision rises with that of the rods, and returns simultaneously to normal on administration of vitamin A$_1$ (Haig, Hecht and Patek, 1938; Wald, Jeghers and Arminio, 1938). It is probable, therefore, that the cone photopigments are closely related to rhodopsin in structure, and are similarly derived from vitamin A or a comparable substance.

It is well known that the sensitivity of human vision in various regions of the spectrum is displaced toward the red as one passes from dim to bright light, *i.e.*, from rod to cone function. This is the Purkinje shift. Since the *shape* of the spectral sensitivity curve is almost identical in dim and bright light, Hecht and Williams (1922-23) suggested that its difference in *position* in the spectrum might have a purely structural explanation: that cones, like rods, may contain rhodopsin, but that, possibly because of the higher refractility of the outer limbs of the cones, its spectrum

* Under certain conditions a small residue of reaction (1) is temporarily retained in solutions of rhodopsin (Ewald and Kühne, 1878, p. 262; Hecht, Chase, Shlaer and Haig, 1936).

in these organs is shifted toward the red in accordance with Kundt's rule. Hecht later discarded this view. It is now clear that rods and cones possess chemically distinct photopigments; and in the only retina in which this matter has been explored —that of the chicken—the optical properties of rhodopsin and iodopsin in solution account directly for the Purkinje shift (Wald, 1937b).

Form and Function

Within the outer limbs of the rods and cones the photopigments must form part of the protein layers, though possibly they are oriented so that their carotenoid groups project into the lipoidal interstices. Such an arrangement would aid the passage into and out of the retina of the vitamins A, a process prominent in the operation of the rhodopsin cycle. It has been observed that frog retinas light-adapted *in vivo* lose about ⅔-¾ of their vitamin A_1. Originally this datum was taken to mean that vitamin A_1 is destroyed in the visual processes (Wald, 1935-36a). It appears instead, however, to be direct evidence that, as the concentration of retinal vitamin A rises during light adaptation, most of it diffuses out into contiguous tissues and the circulation. Conversely, during dark adaptation the retina recaptures vitamin A_1 by binding it in non-diffusible form in rhodopsin (Wald, 1935-36b).

These migrations make an important contribution to visual physiology, since they connect the otherwise closed rhodopsin cycle with the general level and metabolism of vitamin A_1 throughout the body. They appear to find direct histological expression in the conditions which govern the appearance of Kolmer's droplets—lipoidal structures found under certain circumstances clustered about the outer limbs of the rods (Johnson and Detwiler, 1942). The droplets are prominent in histological preparations from dark-adapted animals, but are sparse in those from light-adapted eyes. They occur in large numbers in situations in which the retinene and vitamin A content of the retina is high: (a) dark-adapted tissues in which the destruction of rhodopsin by the histological fixative has liberated retinene$_1$; or (b) retinas light-adapted at 0° C, in which the thermal reactions which lead to the production and emigration of vitamin A_1 have been inhibited by cold. Conversely, Kolmer's droplets are scarce in situations which have permitted the diffusion of vitamin A_1 from the retina or prevented its return: (a) light-adaptation at room temperature; or (b) dark-adaptation in the cold, which by inhibiting the synthesis of rhodopsin has prevented the recapture by the retina of vitamin A_1. It seems from these experiments that the coagulation of retinal protein by the histological fixative sweats retinene$_1$ and vitamin A_1 out of the structure of the outer limb to form the Kolmer droplets on its surface.

It was noted above that an isolated retina, though it can reverse the bleaching of rhodopsin to retinene$_1$ and protein, is unable to synthesize rhodopsin from vitamin A_1. Kühne (1878a, p. 8) made the remarkable observation that the retina regains this capacity when laid back smoothly upon a pigment epithelium. The reaction requires a living pigment epithelium, though not necessarily a living retina (Ewald and Kühne, 1878, p. 261). One might suppose this need for contact with the pigment epithelium to reside in the retina's dependence upon it for the transfer of material from the choroidal circulation. But the circulation is not required in this synthesis, since it proceeds in a frog eye completely removed from the body, though otherwise intact. Furthermore, in the human condition of retinal detachment, portions of retina which have risen from the underlying pigment epithelium are blind, though they possess an intact circulation. Sight may be restored to them by an operation introduced by Gonin in which the elevated sections of retina are laid back upon and attached at points to the pigmented tissue (cf. Duke-Elder, 1940). What it is that the pigment epithelium contributes to the synthesis of rhodopsin is not known; but it must be clear from the foregoing discussion that contact and diffusion phenomena play an important part in this process.

The synthesis of retinal photopigments in darkness is the basis of the sensory phenomenon of dark-adaptation, *i.e.*, the rise of visual sensitivity to a constant maximal level. The existence of this level shows that there are fixed limits to the quantities of photopigment synthesized in rods or cones. Judging from the constancy of threshold of the completely dark-adapted eye, these limits are maintained with considerable precision. What determines the maximum concentrations of photopigment in the retina? Not normally exhaustion in the supply of vitamins A, since small quantities of vitamin are always found in the dark-adapted tissue, and further supplies are continuously available from the circulation. The synthesis of photopigments must be limited ultimately by the supply of their specific proteins. It is for this reason that—contrary to a widespread belief—the sensitivity of the normal eye cannot be improved by dosage with vitamin A.

This consideration raises a neglected possibility in the etiology of night-blindness. There has been much interest recently in the transitory night-blindness associated with vitamin A deficiency, in which the synthesis of rhodopsin—and perhaps also of cone photopigment—fails, due apparently to simple reduction in the retinal availability of vitamin A_1. Congenital forms of night-blindness exist also, however, which are not accompanied by obvious anatomical lesions, and are not responsive to vitamin A therapy. Possibly such conditions involve the lack of specific photopigment proteins. A puzzling type of night-blindness recently observed in experimental human vitamin A deficiency, and highly refractory to administration of vitamin A, may also be of this nature (cf. Hecht and Mandelbaum, 1940; Steven and Wald, 1941).

It is not sufficiently appreciated that photopigment must form a considerable portion of the total structure of the rod outer limb. A frog retina of 3 mg dry weight may contain 1.25 microgms. of vitamin A_1 (Wald, 1934-35). On the conservative assumption that in the synthesis of rhodopsin one mol (286 g) of vitamin A_1 combines with about 18,000 grams of protein—a minimal molar unit of protein structure —this retina, when dark-adapted, contains about 0.08 mg of rhodopsin. Rod outer limbs probably account for less than ⅙ the weight of the retina, or about 0.60 mg. On this basis, therefore, rhodopsin constitutes about 13 per cent of the outer limb. This is almost certainly a minimal estimate. It raises the possibility that the protein of the rod outer limb is largely, or even predominantly, rhodopsin-protein, and is an assurance in any case that this substance plays a considerable structural as well as chemical role in the retina.*

How is the bleaching of photopigments in the rods and cones transmuted into a neural excitation? It is difficult to suppose that liberated carotenoid groups act as chemical irritants comparable, for example, with the neurohumor acetylcholine, for few substances are so bland as the known carotenoids. In a curious series of experiments Kühne (1878c. p. 468) attempted to test directly the possible excitatory effects of unbleached and bleached rhodopsin solutions upon frog nerve-muscle preparations, with negative results. He also crushed retinas before and after bleaching, with the photoreceptor layer pressed against the tongue, and found no taste stimulation. The realization that the photopigments form a substantial portion of the oriented structure of the outer limb places this matter in a new light. In this case the bleaching of photopigment represents directly an appreciable disruption of the organization of the outer limb such as may constitute its excitation in every real sense.

* It follows that deficiency of rhodopsin-protein, suggested in the preceding paragraph to be one basis of night-blindness, amounts in effect to the structural disintegration of the red outer limb. We may already have a histological description of this process. Johnson (cited in Detwiler, 1943) reports that in moderately vitamin A-deficient rats, red outer segments degenerate and may entirely disappear. After several months of adequate feeding they regenerate to an approximately normal condition. These results offer a close parallel to the observation that in certain cases of experimental human vitamin A-deficiency, night-blindness is repaired only after 6 weeks to several months of a complete diet.

The alternation of oriented protein and lipid layers in the outer limb resembles closely that found in the axon sheath of nerve; it is found also in the stroma of the red blood cell, and may be very generally characteristic of protoplasmic surfaces (cf. Schmidt, 1938; Schmitt, 1936). It was remarked above that the photopigments may be oriented in the protein layers of the outer limb so that their carotenoid groups project into the lipoidal interstices. In this event the multiple interfaces between protein and lipid are actually composed largely of photopigment. If, as seems likely, surface structure plays as large a part in the excitation and recovery of rods and cones as it does in nerve, this consideration provides a ready explanation of the role of the photopigments in the visual process.

The Photopigments in Solution

Kühne attempted without success to extract rhodopsin from the retina with a wide array of organic and aqueous solvents. Finally, recognizing that the rod outer limb is very similar in composition to the red cell stroma and the myelin sheath of nerve, upon both of which bile salts exercise a strong lytic action, he tried this reagent upon the retina (1878b, p. 42). On exposure to bile salts solution the outer limbs shot "like rockets" from the retinal surface, lengthened and coiled "like worms," then spilled forth the lamellae of which they are composed "like coins shot out of a pipe." In the course of this disintegration rhodopsin went into solution. Until recently 5 per cent aqueous bile salts (sodium glycocholate, taurocholate, desoxycholate, or a mixture of these salts) remained its only known extractant and solvent.

Within recent years a number of new solvents have been found for the visual photopigments. Tansley (1931) introduced the use of the glucosides saponin and digitonin. The latter in 2 per cent aqueous solution is probably the most satisfactory solvent available for rhodopsin, and has proved equally effective for the extraction and solution of porphyropsin and iodopsin. Hosoya (1933) reported the extraction of rhodopsin with aqueous solutions of the plant glucosides panaxtoxin and kalotoxin, with 0.1-0.2 per cent sodium oleate, and with 2-3 per cent sodium salicylate. Several years later (1936) he added two snake venoms to this list; their efficacy is probably related to Lythgoe's observation (1938) that lysolecithin, a modified phosphatide formed from lecithin by cobra venom, and a powerful hemolytic agent, liberates rhodopsin from the rod in a concentration of $\frac{1}{10,000}$. Lythgoe (1938) also reported that rhodopsin is "liberated" by sodium palmityl sulfonate; but Smith (Smith and Pickels, 1940-41) found this pigment to be destroyed instantaneously by the closely related sodium dodecyl sulfonate.

All these substances are detergents—compounds characterized by the possession of polar and large hydrocarbon groups. Their solvent action upon the retinal photopigments is an example of their general capacity to carry hydrophobic molecules into aqueous solution, a property called hydrotropy by Neuberg (1916). The action of bile salts in this regard has been studied in detail and may serve as a model for the group (Wieland and Sorge, 1916; Rheinbolt, 1927; Verzár and Kúthy, 1929). The bile acids form water-soluble addition complexes of large size with a great variety of organic molecules. The binding involves their hydrocarbon rather than their polar groups; in general the larger the non-polar portion of a molecule, the more molecules of bile acid combine with it.

The complexes so formed are stable in composition, often beautifully crystallizable, and may be of very large size. So for example stearic acid, $C_{17}H_{35}COOH$, combines with 8 molecules of desoxycholic acid to form a water-soluble, crystalline complex of molecular weight 3420. Detergent molecules also combine with themselves to form large colloidal complexes. Digitonin on standing in solution aggregates to the point of precipitation; such solutions must be freshened before use by bringing to a boil. In this case also stable complexes of definite composition may be formed. Smith and Pickels (1940) have shown that in the ultracentrifuge digitonin

sediments with a sharp boundary, indicating particles of uniform size, of apparent molecular weight at least 75,000, and containing at least 60 molecules of digitonin per micelle.

This situation has confused initial attempts to measure the molecular weight of frog rhodopsin in digitonin and bile salts solutions. Diffusion measurements have indicated that this is some sub-multiple of about 810,000 grams (Hecht, Chase and Shlaer, 1937); ultracentrifuge studies have yielded apparent weights of about 270,000 (Hecht and Pickels, 1938) and about 100,000 (Lythgoe, 1938). Rhodopsin in such solutions, however, is not an individual but rather the nucleus of a complex with detergent molecules, to which rhodopsin itself may contribute only a small fraction of the mass. The measurements cited concern entire rhodopsin-detergent micelles, and offer no indication of the dimensions of the rhodopsin molecule alone. Our information of the latter has advanced little beyond the observation of Ewald and Kühne (1878, p. 454) that rhodopsin in bile salts solution does not diffuse through a parchment membrane.

Kühne (1895) made the technically important discovery that rhodopsin remains soluble and otherwise apparently unharmed following treatment of the retina with 4 per cent alum solution. Alum coagulates hemoglobin and other retinal proteins which otherwise enter the extracts, and in this way rhodopsin is singled out for extraction in a relatively pure condition. The value of this procedure can hardly be exaggerated, since rhodopsin is never available in more than minimal quantities and is scarcely approachable with the ordinary, wasteful methods of protein fractionation. Rhodopsin also survives the action on the retina of 4 per cent formaldehyde, which may be employed in its preparation in the same manner as alum. Either reagent can also be used in preparing porphyropsin; but alum treatment either destroys iodopsin or renders it insoluble (Wald, 1937b).*

The color and photosensitivity of rhodopsin are not destroyed by the death and subsequent coagulation and autolysis of the retina, nor in the course of intense rotting of the incubated tissue. Under these circumstances however rhodopsin may lose its solubility. This can be preserved by pickling the retina in 10 per cent brine (Ayres, 1878). Brine also has been reported to restore the solubility of rhodopsin when, as occasionally happens, it is lowered by alum treatment (Kühne, 1895).

These observations furnished the basis for Kühne's final directions for preparing rhodopsin solutions (1895). They have since been found equally effective with porphyropsin, and except for the alum treatment, with iodopsin. After removal of the anterior portions of the eye and the vitreous humor, the entire fundus is hardened in 4 per cent alum. The retina is dissected out, laid in water for one hour, then in 10 per cent brine for several hours. Rhodopsin is extracted with at least 4 per cent bile salts. To prevent putrefaction the final solution may be saturated with sodium chloride. For subsequent purification of such solutions, advantage may be taken of the fact that rhodopsin is precipitated from bile salts solution by saturation with magnesium sulfate or half-saturation with ammonium sulfate; the bile salts come down with the photopigment. Minor variations of all these procedures are still in use. The most notable innovations since Kühne have been the replacement of bile salts by digitonin, the use of the centrifuge to replace tedious and wasteful filtrations, and most important of all, the use of buffers to regulate the pH of the final solutions.

For the rest, the properties of the retinal photopigments are those common to proteins. Many of them, including several already mentioned, were described in the

* Glycerol also does not injure rhodopsin, and Ewald and Kühne (1878, p. 454) describe a mixture of 1 part rhodopsin in bile salts solution with 3 parts glycerol of which they remark the striking depth of color and high photosensitivity. Saito (1933) also has studied the behavior of rhodopsin in mixtures of 10-100 per cent glycerol with water. The capacity of this pigment to withstand admixture with glycerol has recently been rediscovered and used to good advantage in a study of its properties at −73° C (Broda and Goodeve, 1941-42).

case of rhodopsin by Ewald and Kühne, though the conclusion that this pigment is a protein was drawn only recently (Wald, 1935-36a). The primary data which identify these substances as proteins involve their destruction in ways characteristic of protein denaturation.

Rhodopsin and porphyropsin are rapidly destroyed by exposure to a variety of typical protein denaturants: alcohols, acetone, chloroform, heavy metal ions, and mineral acids and alkalies. Rhodopsin is destroyed by warming above 52° C (Ewald and Kühne, 1878, p. 440), porphyropsin above 62° C (Wald, 1938-39b). This process is accelerated in acid or alkaline solution; and in desiccated retinas rhodopsin survives heating even to 100° C (Ewald and Kühne). I have calculated from the latter authors' crude measurements of the velocity of heat destruction of rhodopsin at various temperatures an apparent activation energy of about 75,000 cal per mol in the retina and about 40,000 cal per mol in aqueous solution. Lythgoe and Quilliam (1938) have recently obtained very nearly the latter value in a excellent reinvestigation of the heat destruction of rhodopsin in solution. All these observations are typical of the heat denaturation of proteins.

The differences which Ewald and Kühne observed in the responses of rhodopsin to heat in the retina and in solution suggest some profound alteration in its structure as a result of extraction. This possibility should be re-examined with modern procedures. Ewald and Kühne (1878, p. 392) also found rhodopsin obtained from rabbit retinas to be much more sensitive to heat than that of frogs, and concluded that the rhodopsins of various animals, like their hemoglobins, may in general differ chemically. The parallel may now be recognized as exact, since such specific differences, as in the case of hemoglobin, do not involve the prosthetic group but are confined to the protein portion of the rhodopsin molecule.

Like other proteins, rhodopsin is amphoteric. Solutions prepared from frog retinas have an isoelectric point, determined by the moving-boundary method, at about pH 4.5 (Wald and Raymont, cited in Wald, 1938-39b). This observation has been confirmed by the microelectrophoretic measurements of Broda and Victor (1940) on frog rhodopsin adsorbed on quartz particles, which show an isoelectric point at pH 4.47. Broda and Victor found the isoelectric point of such preparations to rise to 4.57 on irradiation, indicating that the bleaching of rhodopsin makes the molecule slightly more alkaline. They recall that the isoelectric point of globin is 0.7 pH unit higher than that of hemoglobin, and suggest that their comparable observation on rhodopsin is consistent with the view that in bleaching this molecule is cleaved as described above.

References

Ayres, W. C., *Unters. physiol. Inst. Univ. Heidelberg*, **2**, 444 (1878).
Boll, F., *Arch. Anat. Physiol. Abt.*, **4** (1877).
Broda, E. E., and Goodeve, C. F., *Proc. Roy. Soc. London*, **B130**, 217 (1941-42).
Broda, E. E., and Victor, E., *Biochem. J.*, **34**, 1501 (1940).
Chase, A. M., *Science*, **87**, 238 (1938).
Detwiler, S. R., "Vertebrate photoreceptors," Macmillan Co., New York, 1943.
Duke-Elder, W. S., "Text-book of ophthalmology," Vol. iii, p. 2864, Henry Kimpton, London, 1940.
Ewald, A., and Kühne, W., *Unters. physiol. Inst. Univ. Heidelberg*, **1**, I-IV (1878).
Haig, C., Hecht, S., and Patek, A. J., Jr., *Science*, **87**, 534 (1938).
Hecht, S., *Ann. Rev. Biochem.*, **11**, 465 (1942).
Hecht, S., Chase, A. M., and Shlaer, S., *Science*, **85**, 567 (1937).
Hecht, S., Chase, A. M., Shlaer S., and Haig, C., *Science*, **84**, 331 (1936).
Hecht, S. and Mandelbaum, J., *Am. J. Physiol.*, **130**, 651 (1940).
Hecht, S. and Pickels, E. G., *Proc. Nat. Acad. Sci.*, **24**, 172 (1938).
Hecht, S. and Williams, R. E., *J. Gen. Physiol.*, **5**, 1 (1922-23).
Hosoya, Y., *Arch. di. Sci. biol.*, **18**, 183 (1933).
Hosoya, Y., *Jap. J. Med. Sci.*, **4**, 122 (1936).
Johnson, M. L. and Detwiler, S. R., *J. Exp. Zool.*, **89**, 233 (1942).
Kühne, W., *Unters. physiol. Inst. Univ. Heidelberg*, **1**, (a) 1, (b) 15, (c) 455 (1878).
Kühne, W., *Z. Biol.*, **32**, 21 (1895).
Lythgoe, R. J., *Proc. Physical Soc.*, **50**, 321 (1938).
Lythgoe, R. J., and Quilliam, J. P., *J. Physiol.*, **93**, 24 (1938).
Neuberg, C., *Biochem. Z.*, **76**, 107 (1916).
Rheinbolt, H., *Ann. Chemie*, **451**, 256 (1927).
Saito, K., *Jap. J. Med. Sci.*, **2**, 284 (1933).
Schmidt, W. J., *Kolloid-Z.*, **85**, 137 (1938).

Schmitt, F. O., *Cold Spring Harbor Symp.*, **4**, 7 (1936).
Smith, E. L., and Pickels, E. G., *Proc. Nat. Acad. Sci.*, **26**, 272 (1940).
Smith, E. L., and Pickels, E. G., *J. Gen. Physiol.*, **24**, 753 (1940-41).
Steven, D., and Wald, G., *J. Nutrition*, **21**, 461 (1941).
Tansley, K., *J. Physiol.*, **71**, 442 (1931).
Verzár, F., and Küthy, A., *Biochem. Z.*, **210**, 265 (1929).
Wald, G., *J. Gen. Physiol.*, **18**, 905 (1934-35).
Wald, G., *J. Gen. Physiol.*, **19**, (a) 351, (b) 781 (1935-36).
Wald, G., *Nature*, **139**, 1017, 1937 (a).
Wald, G., *Nature*, **140**, 545, 1937 (b).
Wald, G., *J. Gen. Physiol.*, **22**, (a) 391, (b) 775 (1938-39).
Wald, G., *Biol. Symp.*, **7**, 43 (1942).
Wald, G., *Vitamins and Hormones*, **1**, 195 (1943).
Wald, G., Jeghers, H. and Arminio, J., *Am. J. Physiol.*, **123**, 732 (1938).
Wieland, H., and Sorge, H., *Z. physiol. Chem.*, **97**, 1 (1916).

Chemical Pacemakers and Physiological Rhythms

HUDSON HOAGLAND

The Physiological Laboratories, Clark University, Worcester, Mass.

INTRODUCTION: BIOELECTRIC POTENTIALS AND METABOLISM

Living cells are continuous converters of energy, taking in foodstuffs and usually oxygen and excreting carbon dioxide and other metabolites. Within each cell are many specific oxidative and hydrolytic enzymes which promote in stepwise sequences the degradation of foodstuffs coming to the cell from its immediate environment of lymph or other aqueous media. Membranes separating the cell's contents from its environment are selectively permeable as a result of their peculiar colloidal properties. Because of the steady-state flow of substances across the cell walls, the latter becomes the source of interesting electrical properties. The interiors of most living cells are at electrical potentials different from their exteriors. While a few plant cells have been found in which the resting inside potential is electrically positive to the outside, nearly all living cells that have been examined are electrically negative inside with respect to their external membranes. As a consequence, injured regions of most cells, since they represent internal portions, are likewise electrically negative to uninjured (external) regions. The potentials so measured are spoken of as demarcation or injury potentials. In muscle and in nerve cells, for example, they are of the order of 60 millivolts.

Studies of large water-inhabiting plant cells by Osterhout [42, 43] and his colleagues indicate that the resting potentials of these cells are primarily due to diffusion potentials of potassium. The ionic constitution of the cell sap, as determined by direct chemical analysis after its extraction, is found to be strikingly different from the normal surrounding medium bathing the cells. Differences of concentration of diffusible substances across membranes impermeable to certain osmotically active constituents might result from a true thermodynamic equilibrium, but this is not the case for living cells (for a discussion of the Donnan equilibrium and its applications, see Bolam [4]).

The Donnan principle requires for ionic activities the relation

$$a_{K_i} + a_{K_o} = a_{Na_i} + a_{Na_o} = a_{H_i} + a_{H_o} = a_{Cl_o} + a_{Cl_i}$$

where a is activity and the subscripts i and o refer to the inside and the outside of the membrane. Analysis, however, makes it clear that these conditions are not even approximated for living plant cells. Since the solutions inside and outside have roughly equal total ionic strengths, it is permissible to use ionic concentrations instead of activities for comparative purposes. In the case of the marine plant cell, *Valonia*,

for example, the pH of the sap is 5.8 and that of the sea water in which it lives is 8.0, and we have [41]

$$\frac{K_i}{K_o} = 41.60; \frac{Na_i}{Na_o} = 0.18; \frac{H_i}{H_o} = 158.49; \frac{Cl_o}{Cl_i} = 0.97.$$

The inequalities of these values are conspicuous, and prove that the cells are not in true equilibrium with their environments. This type of disequilibrium has been demonstrated in recent years for many plant and animal cells, and is in fact the general rule.

One way to account for the concentration differences of Na and K inside and outside of the cell is to assume that K and Na enter because the thermodynamic potential of KOH and NaOH is greater outside than it is inside, that is

$$(K_o)\ (OH_o) > (K_i)\ (OH_i)$$

The protoplasm is found continuously to produce acids (HA) as metabolic products which tend to accumulate in the sap and thus lower the right-hand product of this expression by the neutralization of OH_i. KOH and NaOH then enter, transforming HA to its salts. These salts must then diffuse out of the sap. All this requires energy, which is produced by the chain of metabolic reactions in protoplasm, forming as a primary end product the acid CO_2 (*i.e.*, H_2CO_3). Other acids incidentally produced by certain plants, such as malic, oxalic, citric, and tartaric, also may serve the same purpose; only when the cellular metabolism is inhibited do the ratios of diffusible electrolytes approach a Donnan equilibrium. Only dead cells are in true equilibrium with their environments.

While this hypothesis has been supported by a variety of evidence from Osterhout's laboratory, D. R. Hoagland and his collaborators have obtained experimental evidence opposed to the view that K concentration in cells depends on the (K) (OH) gradient. These workers have shown that electrolyte concentration in a variety of forms of plant cells is directly dependent upon the carbohydrate metabolism of the cell, but the mere production of CO_2 in the sap does not appear to bring about the concentration. For a review of this subject see Hoagland.[28a] For the purposes of our subsequent discussion of pacemakers, the intimate mechanism underlying electrolyte gradients across cell membranes is not important. The significant thing is the existence of well established quantitative relations between metabolism, electrolyte gradients, and bioelectric potentials. Blinks,[8a] in a recent review of this matter, has written:

"In conclusion we may state that metabolic processes appear to be connected with permeability and electrical phenomena, not only in the formation of organic ion gradients, and of membranes responsible for the manifestation of potentials, resistances and capacities, but for their maintenance, and their alteration under various influences. The nature of these alterations, reflecting the nature of the cell surface itself, is not yet evident, although acidity changes are suggested in some cases."

The mobility of potassium is much greater than that of other substances normally traversing the cell membranes. It thus produces correspondingly large diffusion potentials. Its mobility in protoplasm is about 85 times as great as is that of chloride and 40 times as great as sodium. NH_4, Li, Mg, and Ca have mobilities lower than Na. The order of mobilities is very different from that in water, thus indicating that the protoplasmic surfaces are non-aqueous.[42, 43]

A mechanism similar to those involved in plant cells may also account for resting potentials in other tissues, although caution must be exercised in these considerations. Cowan [6] has shown the importance of potassium in determining potentials in nerve fibers. For a discussion of the role of potassium in muscle the reader is referred to a recent review by Fenn.[11]

Cells may be locally depolarized by chemical, electrical and mechanical stimuli, and such local depolarization may initiate a propagated wave of electrochemical change called an action current. The action current sweeps over the cell and may, at one-way synaptic junctions between neurones, be propagated from one cell to another. The propagated region of activity in nerve and muscle is, in general, electrically negative to the inactive region. For a brief interval following a propagated disturbance the tissue is completely refractory to further stimulation. Metabolic recovery of excitability takes place in nerve and muscle cells in a few thousandths of a second, and during this interval the excitability returns to its normal resting level, as can be determined by testing with stimulating electrical shocks of varying strengths applied at intervals after an initial effective stimulus.

It is not the purpose of this paper to discuss the subject of action currents, and the reader is referred to standard physiological texts for further information on this subject. The points we wish to make thus far indicate that cellular metabolism maintains cell structure and function by a dynamic steady state. The cell membranes, constituting an essential part of this structure, are electrically polarized as a consequence of the energy of cell metabolism. The membrane's permeability is modifiable by stimuli. An action current, once started, sweeps over the cell's protoplasm by a mechanism whereby the electrically active region, as defined above, serves as an electrical stimulus (cathode) to adjacent inactive regions thus causing them in turn to become electrically active. Such propagated action currents arriving at muscles may, through special intermediate mechanisms, stimulate them to contract. Similar propagated disturbances arriving at the central nervous system over afferent nerve fibers inform us about our environment.

Nerve impulses have been defined as a conducted tendency to excite, and action potentials are their readily determinable signs. As a corollary to the fact that the nerve fiber is briefly refractory after the passage of an impulse, the action potentials can travel over a fiber only in quanta. Even the most intensive stimulus to sense organs can, as Adrian[1] was the first to show, only increase the frequency of the number of impulses in a *single* nerve fiber. Owing to statistical variations in the thresholds of sense cells,[24] more sense organs and hence more sensory fibers in a nerve trunk are brought into play as the intensity of stimulation increases. Thus with increasing strength of stimulus the frequency of afferent nerve impulses alone increases. The magnitude of the action potential along a given fiber is independent of the initiating stimulus, since it depends only on the local polarized conditions of the fibers, which in turn are determined by their local metabolic steady-states. The electrical activity along the tissue is thus produced by the local release of stored potential energy. Amplitude increases in action potentials recorded from a nerve trunk, with increasing strength of stimulation to sense organs, results from the increased number of active fibers under the lead electrodes, and the shunting effect of inactive tissue.

We thus see that rhythmicity of nerve impulse discharge is characteristically set up by stimulating sensory cells connected to the nerve fibers. In the presence of a constant physical stimulus the sensory cells may show a progressive failure to initiate nerve impulses, so that the frequency declines in time. The rate of this "adaptation" or "accommodation" to a constant stimulus varies considerably from one type of sense organ to another. Some sense organs, such as touch spots on the skin, adapt so rapidly to a constant stimulus as only to set up one or two impulses per nerve fiber, while other kinds of sense cells, such as muscle and joint sense organs, discharge impulses repetitively for many minutes with little decline in frequency.[1] The mechanism of accommodation of excitable tissue has received much attention from physiologists.[19, 20, 28] In some tissues a region of injury or a constant electrical current produces no impulses after the initial burst when the injury was made or the

circuit closed. Other tissues may discharge impulses repetitively from similar regions, which thus act as pacemakers for the activity.

A morphological pacemaker is usually regarded as a region of an organ or tissue from which rhythmic activity arises. Thus the sinoauricular node of the vertebrate heart is spoken of as its pacemaker. Pacemaker conditions may be established experimentally in excitable tissues. The application of a drop of chloroform to the filamentous plant cell *Nitella* may, by locally reducing the resting potential, set up local circuits between the normal and treated tissue which start waves of action currents along the cell. These waves may be rhythmically repeated as soon as the tissue recovers from the refractory state set up by the previous impulse.[40] In like manner a drop of 0.05*M* KCl, by locally reversing the resting potential of *Nitella,* acts to produce a pacemaker locus. The ends of certain cut or damaged mammalian nerve fibers have been shown to initiate rhythmic discharges of nerve impulses.[2] Similar injury discharges have been described in certain sensory nerves of fishes.[26] Presumably local depolarization of the nerve membranes resulting from the injury acts as a region of constant stimulation, causing the repetitive discharge. In all these cases a steep and sharply localized concentration gradient of electrolytes between the damaged or treated region and the normal tissue is necessary to produce rhythmicity. Lillie's well-known iron wire nerve model[84, 85] has been made to discharge rhythmically by immersing the wire in nitric acid, 60-80 volumes per cent, and inserting one end of it in a snugly fitting glass tube, thus limiting diffusion of acid around the sheathed end. If a wave of activity is then initiated in the wire it is usually found that the sheathed tip acts subsequently as a pacemaker, setting up a repetitive discharge at a frequency limited by the recovery period of the unsheathed wire. After the first impulse the sheathed part of the wire does not become passivated with the usual bright oxide layer, and effervescence goes on continually in the sheathing tube. This is brought about by the limited diffusion of acid in the sheathed part, which, after the passage of one wave of activity, becomes locally so reduced in concentration (ca. 53 per cent) as to be unable to repassivate the sheathed tip. The continuously oxidizing region, anodal to the passivated wire by about 0.7 volt, thus initiates repetitive discharges.

The problem of pacemaker mechanisms is one of very wide interest. Sensory cells, when stimulated, set up excitatory states which function with respect to the nerve fibers as pacemakers for the repeated discharge of impulses. Thus the essentially *arhythmic* photochemical changes initiated in the retina by light develop rhythmic discharges of impulses over the optic nerve. Mechanoreceptors, such as Pacinian corpuscles in joints and tendons and muscle spindle receptors informing us of muscle tension, fire impulses repetitively at a declining frequency depending on the strength of stimulus and on the rate of accommodation of the ending.

In recent years it has become well established, following the work of the late Hans Berger and others, that most brain centers are normally engaged in rhythmic fluctuations of electrical activity. These fluctuating potentials in man can be picked up through skull and scalp by electrodes attached to the scalp. That they come from the cerebral cortex has been established beyond question. The most characteristic pattern obtained through the human skull from resting subjects with closed eyes is a regular sinusoidal rhythm at a frequency of about ten per second which is known as the alpha rhythm, and which is usually most clear over the back of the head external to the occipital cortex, known to be primarily concerned with vision. Opening the eyes breaks up this uniform rhythm because the brain cells producing it are electrically synchronized and pattern vision, projected from the retina to the occipital cortex, breaks up the synchronized "resting" activity of these cells.

Actually there has been detected a whole spectrum of electrical waves from the brain ranging in frequency from less than one per second to around fifty per second.

Certain frequency bands, however, are more conspicuous than others and their prominence may vary considerably from one normal individual to another. Variations in brain wave patterns have also been extensively studied in connection with a variety of neurological disorders. For a review of brain-wave work, especially from the clinical side, the reader is referred to the "Atlas of Electroencephalography" by Gibbs and Gibbs.[15]

We thus see that the central nervous interaction of groups of neurones involves complex pacemaker considerations. Certain neurones of the respiratory center fluctuate rhythmically and independently of afferent impulses from peripheral sense organs. This was clearly indicated by the experiments of Adrian and Buytendjik,[8] who demonstrated that the brain stem of the goldfish, when completely removed from the body, continues to manifest electrical potential fluctuations at a frequency characteristic of the normal gill movements. Subsequently work of Gesell (see especially reference 14) and his collaborators has demonstrated intrinsic rhythmicity of the mammalian respiratory centers. These rhythms, like heart rhythms, are modifiable by afferent nerve impulses, but they are basically a product of the center's metabolism—just as the myogenic heart rhythm is a result of the heart's own metabolism.

Chemical Pacemakers

One might well ask at this point, how can continuous steady-state chemical events produce electrical rhythmicity in tissues? A model originally proposed by Van der Pol [52] in connection with cardiac rhythmicity may be helpful in thinking about this matter, although the subsequent discussion in no way depends upon the applicability of this, or any other specific model, to living tissues.

Fig. 1, from a paper by Fessard,[12] shows the essential principles of this model,

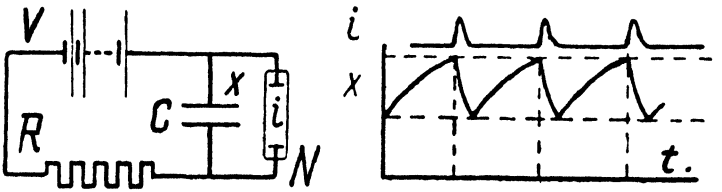


FIGURE 1. Figure from Fessard [12] showing at left a neon tube, N, in circuit with a condenser, C, battery, V, and resistance, R. The tube fires intermittently, as shown on the right; x is voltage across the condenser and i is current across the tube.

which consists of a neon tube connected in parallel with a condenser and in series with a battery and a resistance. Such a tube will flash rhythmically at frequencies determined by two factors. One of these is the impedance of the circuit as measured by the condenser capacity and the resistance. The other is the speed of the physicochemical events going on in the battery. If the impedance is constant the frequency will vary with what we might refer to as the speed of the oxidative "metabolism" of the battery. In this model time is required for the battery to load the condenser to its firing threshold across the tube. Other things being equal, the faster the battery's oxidations, the more rapidly the charge can be loaded and hence the greater the frequency of flashing.

This analogy to living tissues is suggestive, since living cell walls have a definite measurable impedance and are stimulated by critical voltages. Moreover, within the cell respiration may be considered analogous to the battery part of the model in determining recovery of the cell's excitability. As a byproduct of respiration, electrical polarization takes place across the membranes which, in some cells, may act as a stimulus when critical voltages are reached. The fact that a mouse's heart beats normally six times as fast as a man's heart does not, of course, imply that the mouse's

heart is metabolizing six times as fast. The absolute differences would presumably be due to differences in the electrical constants of the tissues. We might, however, expect that increasing or decreasing the respiration of heart muscle in the two cases might proportionally increase or decrease the frequency.

In the battery part of the model, oxidation is proceeding in a series of physico-chemical steps. Whichever one of these is slow with respect to the others may be conceived as acting as a sort of bottleneck or pacemaker. Changes in the flash frequency might thus be due to changes in the speed of one limiting step in a series. In cell respiration we know that foodstuffs are broken down in sequential steps yielding energy. If one of these steps is slower than the others it, too, may serve as chemical pacemaker. The chemical pacemaker is thus analogous to the idea of the industrial bottleneck which, unfortunately, we have heard so much about of late. Foodstuff molecules, on being dismantled of their hydrogen, yield energy which determines velocities of physiological events. The foodstuff molecule is degraded in the cell in a stepwise series of reaction, each step under the control of a specific enzyme system. A slow enzyme step in such a chain is analogous to a slow worker on a continuous assembly line engaged in assembling or dissembling a complex machine. His speed determines the final output. In like manner one enzyme-catalyzed step in a sequence may directly determine the speed of some complicated physiological process. We shall presently discuss a method of studying chemical bottlenecks of this type in some detail.

To summarize: A *morphological pacemaker* is a region of tissue which, after discharging, recovers its excitability more rapidly than other regions of the same cell system. Its short recovery period thus makes possible discharge from this region somewhat before adjacent regions have fully recovered their excitability. Thus in a spontaneously rhythmic system like the heart, such a region becomes the origin of the beat.

A *chemical pacemaker,* on the other hand, is the slowest physico-chemical event of a sequence determining the energy exchange, and hence the frequency of rhythms of the morphological pacemaker. It might thus, for example, be the slowest step in oxidative recovery of excitability taking place in the most rapidly recovering cells of the sinoauricular node of the vertebrate heart. Thus we see that a single, simple continuous reaction of a complex series, if it is relatively slow, may theoretically set the pace of a physiological rhythm.

Temperature and Enzyme Pacemakers of Physiological Rhythms

It should be possible to test the view that physiological rhythms are dependent on the velocities of simple chemical pacemakers, by studying these rhythms and also biological oxidations as a function of temperatures.

The Arrhenius equation may be written as follows:

$$v = ze^{-\mu/RT} \tag{1}$$

where v is chemical velocity, z is a constant, e is the base of natural logarithms, T is the absolute temperature, R is the gas constant, and μ is the critical thermal increment or energy of activation, *i.e.,* the amount of energy per mol over and above the average energy of the system that molecules must acquire before they can react. In the literature of physical chemistry, E is used instead of μ for the energy of activation. To avoid the implication that the μ values of biological processes are necessarily identical with activation energies as used by physical chemists, Crozier [10] originally introduced the symbol μ and referred to it as the "temperature characteristic" rather than as the energy of activation. Recent work, to be described below, suggests that E and μ are probably identical.

Taking logarithms on both sides of equation (1) we obtain

$$\log v = c - \mu/2.3RT \tag{2}$$

If experimental data conform to the Arrhenius equation, it is thus clear that a plot of log v vs $1/T$ should yield a straight line with ordinate intercept at c and negative slope of $\mu/2.3R$.

If frequency (f) of some physiological rhythm is directly proportional to the velocity of some underlying chemical pacemaker we may write

$$f = kv = kze^{-\mu/RT} = ae^{-\mu/RT} \tag{3}$$

and again taking logarithms

$$\log f = c' - \mu/2.3RT \tag{4}$$

Thus in comparing equations (2) and (4) it is clear that a plot of log f vs $1/T$ would give the same value of μ as if we plotted log v vs $1/T$ directly. In a log frequency plot the intercept on the ordinate axis would be at c' instead of at c. The two plots would give parallel lines of the same negative slope equal to $\mu/2.3R$, or $\mu/4.6$, since R is 1.99, or 2 calories per mol per degree.

It is thus clear that an interesting test of the master reaction principle might be had by comparing μ values for frequencies of physiological rhythms with those for directly measured biological oxidations both for cell systems as a whole and for isolated respiratory enzyme systems acting with their respective substrates.

The Arrhenius equation describes with precision a large number of homogeneous and heterogeneous reactions both in solution and in the gaseous state. The reader is referred in particular to Hinshelwood's monograph[23] for a discussion of the equation and the concept of the energy of activation. For our purposes we wish to see to what extent the equation may be used to describe physiological and biochemical processes with the view to examining μ values as possible indicators of specific chemical pacemakers.

The Arrhenius equation has been applied by a host of workers to many hundreds of biological rates and is found to describe the data in most cases. The mere fact of a fit to data of this very general relation is not especially interesting. What is more significant is the order of magnitudes of the μ values and the fact that they occur, not distributed at random, but grouped in certain well-defined modes.

Fig. 2 shows an Arrhenius equation plot of data from Navez[89] for the rate of

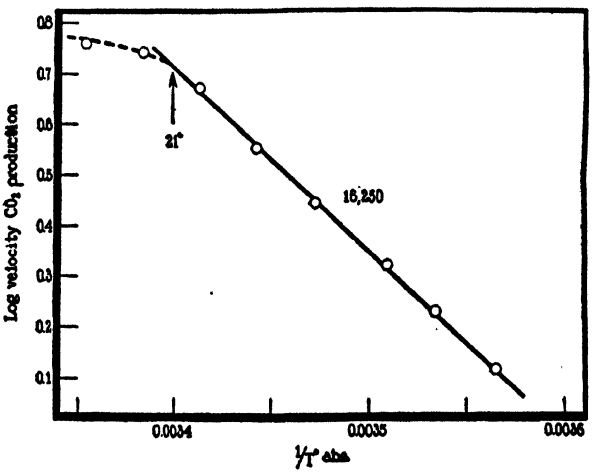


FIGURE 2. CO_2 production of seedlings of *Vicia faba* as a function of temperature, from Navez.[89]

production of CO_2 by seedlings of the plant *Vicia faba*. Above 21° C the relation in this particular case falls off. Note that because of the reciprocal method of plotting low temperatures are at the right, high temperatures at the left on the abscissa.

Fig. 3, on the other hand, shows a group of μ plots from a paper of Fries [18] on

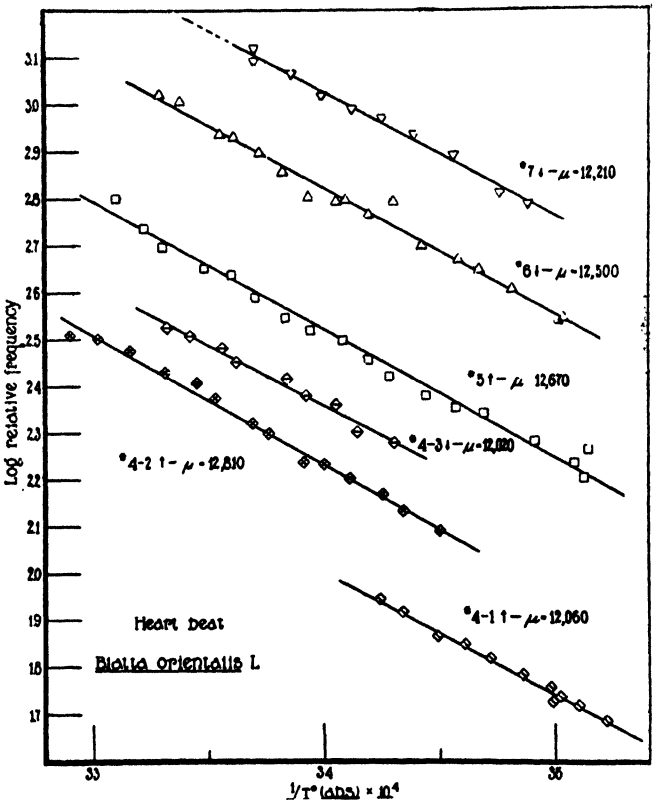


FIGURE 3. Frequencies of beating of cockroach hearts as a function of temperature. From Fries.[18]

the frequency of heartbeats of cockroaches as a function of their internal body temperatures. The μ averages 12,500 calories.

Fig. 4, from Crozier,[7] shows data obtained by three different workers at different

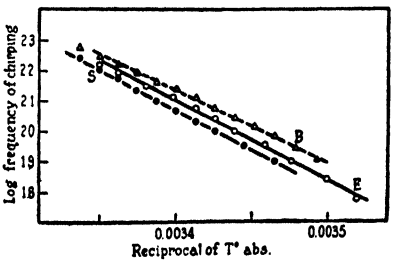


FIGURE 4. Data from three independent observers for the frequency of chirping of crickets as a function of temperature, from Crozier.[7]

times and places relating the frequency of chirping of crickets to temperature. The μ is 12,200 calories, a value frequently encountered in connection with rhythmic activity in insects.

Crozier [10] in 1925 made a frequency distribution plot of some 360 μ values for diverse biological rates. Fig. 5 from his paper shows clearly that the μ values are

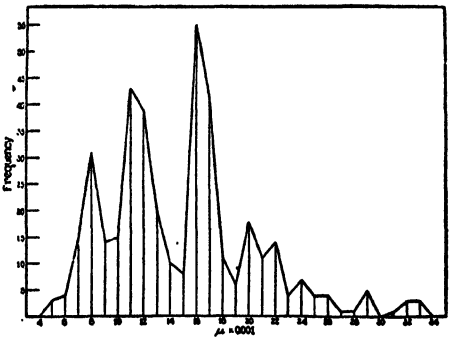


FIGURE 5. Frequency distribution plot of 360 μ values for diverse biological rates, from Crozier.[10]

not distributed at random but are grouped in certain modal values. Over 90 per cent of the μ values are those for physiological rhythms such as breathing movements, heartbeat frequencies, movements of appendages, cilia, etc. of "cold-blooded," *i.e.*, poikilothermous animals.

In 1936 [29] the writer published a figure for the distribution of μ values occurring in the literature for some 76 studies of biological oxidations. Fig. 6 is a reproduc-

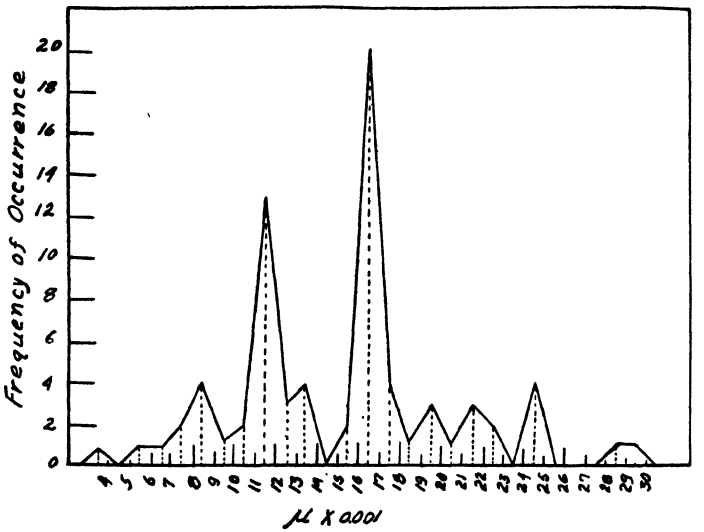


FIGURE 6. Frequency distribution plot of 74 μ values for diverse cellular respiration experiments, from Hoagland.[29]

tion of this figure. It is clear that these μ values group in essentially the same major modal categories as do those of the data for rhythmic processes earlier figured by Crozier. Chi square tests have demonstrated the statistical reality of these modes.[27] The most conspicuous μ groupings are at 16-17,000 calories, 11,000 calories and 8,000 calories.

To account for the often recurring μ values in different protoplasmic systems, Crozier originally suggested that the modal values may correspond to energies of activation of specific catalyzed links in sequential steps in the chemical chains constituting cellular metabolism. The wide recurrence of certain respiratory enzymes in highly diverse tissue systems is in conformity with this view. Catalysts promote reactions by so orienting molecules that particular bonds can be broken in a way

not usually possible except by a marked increase in kinetic energy produced at elevated temperatures. In a chain of catalyzed reactions we should expect the slowest link to act as chemical pacemaker. If in different cell systems the slowest link is now one and now another of a few enzyme systems common to many different tissues, a multimodal distribution of μ values such as has been obtained might be expected. Moreover, experimental conditions altering the ratio of velocity constants of the steps in the sequence might be found to shift the μ from one of the prominent characteristic group values to another. Such experimental shifts of μ from one of the 8, 11, or 16 thousand groups to another have been described.[27]

The oxidation of ferrous to ferric iron *in vitro* requires an energy of activation of 16,200 calories (for citations see Crozier [8]). In many cell respiratory processes it has been demonstrated by direct chemical methods that an enzyme system (cytochrome-cytochrome oxidase) involving iron is often the limiting factor,[58] and it is known that the catalytic iron of this system undergoes oxidation and reduction. This fact is at least consistent with the predominance of 16,000 calories among physiological μ values.

Most physiological rates are described by the Arrhenius equation, but some are not. For systems involving two or more concurrent chains of processes with different μ values, each process of which contributes appreciably to the measured rate, we should not expect a rectilinear relation. The plot should be concave upward.[9] When such concurrent processes are analyzed *separately* they are found to give rectilinear relations.[22]

Some Arrhenius plots show sharp breaks in the curves at critical temperatures. The significance of these breaks has been discussed elsewhere in a number of papers from Crozier's laboratory in the *Journal of General Physiology*. Their presence is consistent with the view that in a sequence of reactions, the same reaction may not be the slowest or master reaction over an extensive temperature range, since each step varies independently with temperature (*i.e.*, has its own μ value) of each other step. Thus one reaction may be slowest over part of an experimental

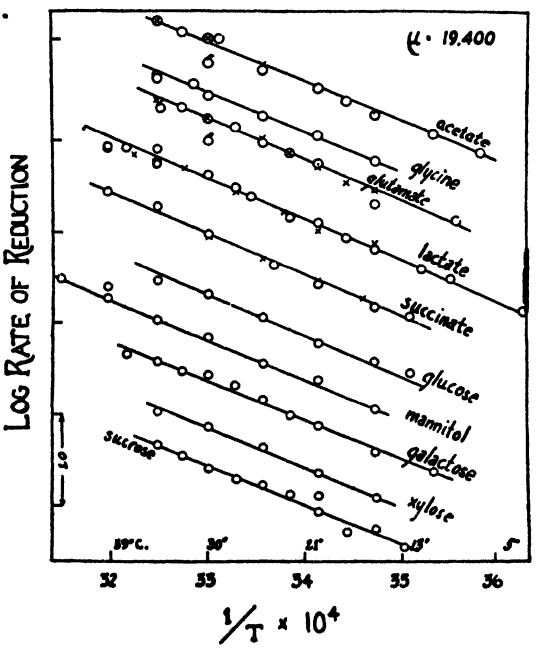


FIGURE 7. Log of rate of reduction (calculated from the time required for 75 per cent reduction of methylene blue) in presence of various substrates by a suspension of *Escherichia coli* plotted against 1/T. μ in the Arrhenius equation = 19,400. ○ refers to the series of determinations made with a first bacterial preparation; × to determinations made with second separate bacterial preparations. For lactate × refers to determinations made with a preparation consisting of autolyzed cells. From Gould and Sizer.[16]

temperature range and another slowest over another part, *i.e.*, the temperature rate curves may intersect at a critical temperature. The μ plot under these circumstances would be expected to show a break in slope at the critical temperature. These findings are entirely consistent with the master reaction principle, and indeed strengthen it. Burton [5] has criticized the master reaction principle in relation to μ values, but he has, in doing so, made certain assumptions not involved in the present discussion. For an answer to these criticisms see Hoagland.[30]

Sizer [47] and Gould and Sizer [16] have demonstrated that μ values of biological processes may be indicative of specific enzymes. Sizer [48] has shown that the hydrolysis of sucrose and of raffinose by yeast invertase gives a μ of about 12,000 calories which is independent of pH changes ranging between 3.2 and 7.9 units and is not altered by the presence of electrolytes. Gould and Sizer have shown that the decoloration of methylene blue by bacteria (b. coli) in the presence of ten widely different substrates yields a μ value of 19,400 calories for the dehydrogenations and is thus independent of the gross nature of the substrates. Fig. 7, taken from their paper, demonstrates this.

Earlier in this discussion, mention was made of electrical brain waves which may be recorded in man through skull and scalp. In 1936 the writer [29] studied the effect of internal body temperature on the alpha brain wave frequencies of a group of normal persons and a group of patients suffering from brain syphilis or general paresis. The body temperatures of the subjects were raised, in the cases of some of the patients, to over 106° F by passing high-frequency alternating currents through their well wrapped bodies.*

Fig. 8 shows alpha rhythms for a normal subject at each of a number of body

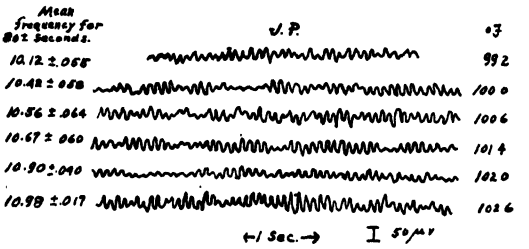


FIGURE 8. Sample of brain waves from a normal human subject showing strong alpha rhythms. Mean frequencies for each temperature for 80 seconds of recording (ca. 800 cycles) with probable errors shown at left. From Hoagland.[29]

temperatures. Fig. 9 is an Arrhenius equation plot of these data yielding a μ of 8,000 ± 200 calories. Fig. 10 shows a composite plot of all the data. The lower band of points ($\mu = 8{,}000$) represents data on three normal subjects, a multiple sclerosis patient and two patients who were least advanced in their paretic symptoms. The middle band of points corresponds to data from two decidedly more advanced paretic patients ($\mu = 11{,}000 \pm 300$ calories). The upper band of points represents data from two very advanced paretics ($\mu = 16{,}000 \pm 300$ calories). Each point is a mean of 80 seconds of record (about 800 alpha cycles).

It should be pointed out that the ordinate intercepts of these plots are here meaningless; the data are arbitrarily arranged on the figure in order to compare the slopes or μ values. The brain-wave frequencies of the advanced paretic patients are

* Large currents at frequencies of the order of a million cycles may be safely passed through the body without shocking or stimulating its tissues. To stimulate, a current must pass for an interval of time that varies hyperbolically with its magnitude. To stimulate, currents must move electrically charged particles definite distances and must hence act for appreciable times in a given direction. Oscillating a million times a second, they cannot stimulate.

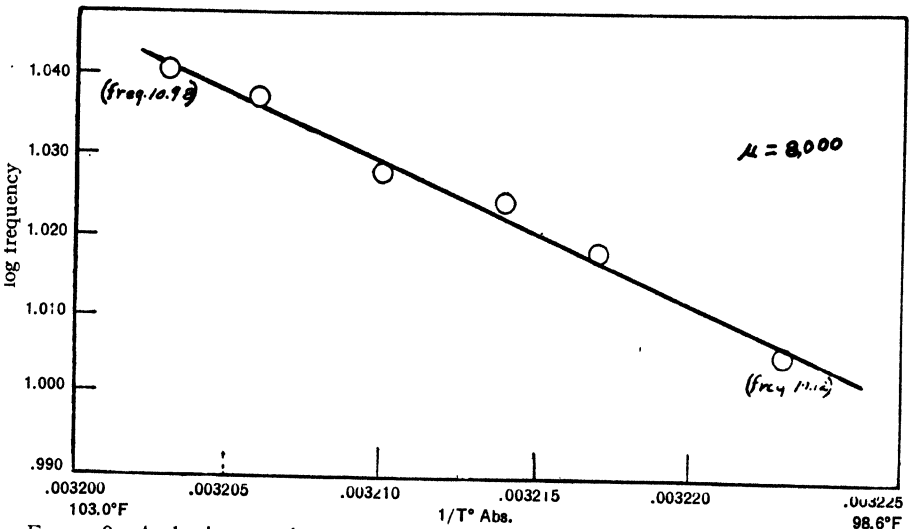


FIGURE 9. Arrhenius equation plot of alpha brain-wave frequencies from the same subject whose brain waves are shown in Fig. 8, from Hoagland.[29]

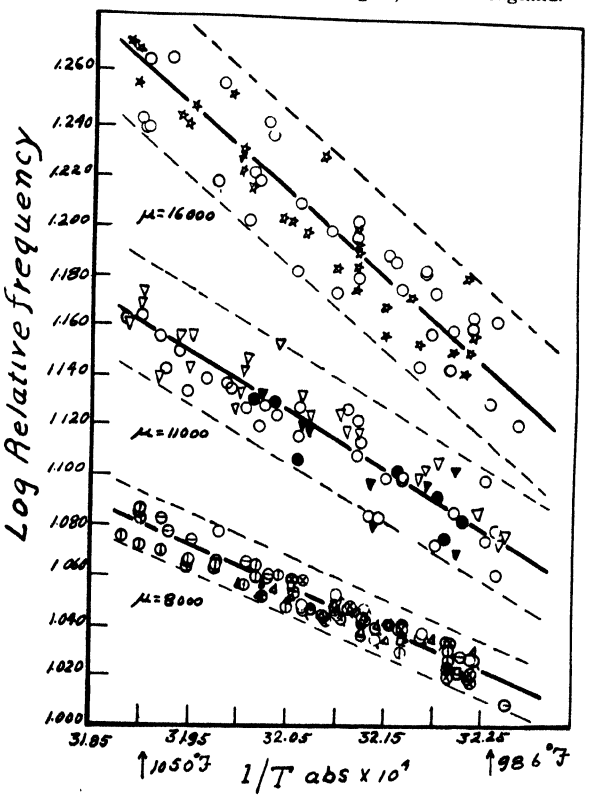


FIGURE 10. Composite plot of data showing a μ of 8,000 calories for three normal persons, a multiple-sclerosis patient and two very early general paretics. $\mu = 11{,}000$ for two more advanced paretics and 16,000 for two *very* advanced paretics. From Hoagland.[29]

not faster at normal body temperatures than are those of normals. They are, however, more rapidly *accelerated* with rising temperatures. The results indicate that these μ values are energies of activation of particular chemical pacemaker links in respiratory reaction chains in the cells producing the alpha rhythms. Evidently the advancing spirochete infection shifts the chemical pacemaker from one of these links to another.

In view of these findings a series of experiments was carried out by the writer and various collaborators to test further metabolic factors underlying central nervous rhythms; the following results were obtained:

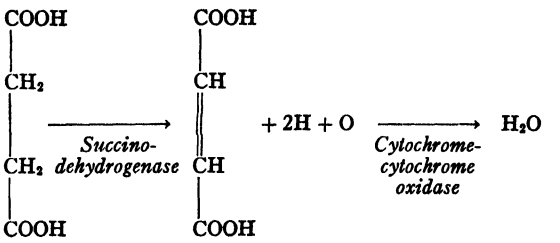
Lowering blood sugar by insulin below a critical level and thus lowering brain sugar causes the alpha frequencies in man and dogs to fall. The frequencies recover when sugar is injected.[31, 33] Brain O_2 consumption follows the curve of sugar after insulin administration, and the alpha frequency curve, within certain limits, parallels both these curves.[21] Dinitrophenol and thyroxin, well known metabolic stimulants, cause the frequencies to increase.[32, 45] Phentobarbital sodium, which inhibits brain respiration *in vitro,* slows the cortical frequencies.[33]

These findings are, of course, *not* to be interpreted as meaning that the *only* modifiers of brain-wave frequencies are necessarily changes in cell respiration. Afferent stimulation which modifies frequencies may do so by locally changing cell respiration or by modifying the electrical constants of the cells or their connections by direct permeability changes. Absolute differences in frequencies of different cell groups also are not to be regarded as due to corresponding absolute differences in rates of O_2 consumptions, but rather to differences in structure (impedance) of the cell walls or connecting fibers. These considerations in no way militate against the view that *relative* frequency changes under specific conditions are due to changes in rates of cell metabolism.

The prominence of the μ values of 8, 11 and 16 thousand calories in particular leads to the question as to what specific steps in cellular respiration they may correspond to. Reasons have been mentioned above for thinking that the last of these values may be association with an iron-catalyzed pacemaker step. This view is justified by the following considerations. Iron pigment, which occurs in all normal brains in a characteristics manner, is found to stain in advanced paretic brains and is distributed quite differently. The paretic brain iron occurs in colloidal spicules scattered about as if it had been precipitated in some way by the infection. In fact this phenomenon is used as a valuable postmortem test to tell just where the spirochete has been busy in the brain.[38] This suggests that iron, normally plentifully present in the cytochrome-cytochrome oxidase system in brain, has become inactivated catalytically. For evidence for this see reference.[17] A reduction of normal catalytic cytochrome brain iron might thus slow this particular link to a point where it would become the pacemaker step, thus shifting the respiratory measure to a μ of 16,000 calories which we have seen to be characteristic of the oxidation of ferrous to ferric iron.

It occurred to the writer that a check on all this might be had in the following way. If an oxidative process catalyzed by an enzyme system containing two or more components *acting in sequence* were to be studied *in vitro,* it should be possible to obtain different μ values for the process by partially poisoning different components of the system, thus making now one and now another step the slow one. Furthermore, if one of these components be an iron-containing enzyme, one of the two μ values might turn out to be 16,000 $\pm$ calories in terms of the above considerations. The objection might be raised that the enzyme-catalyzed processes *in vivo* need not necessarily correspond to such processes *in vitro.* This is true but the objection has been met at least in part by Sizer,[48] who has shown that the invertase inversion of sucrose by living yeast cells and by yeast cells killed in toluene yields identical μ values.

Such a desired enzyme system may be extracted from the beef heart by methods described by Stotz and Hastings.[49] The system has been shown by these workers to oxidize succinate to fumarate and to involve at least two components, succino-dehydrogenase and cytochrome-cytochrome oxidase. The reaction is followed by measuring the oxygen consumption manometrically in Warburg vessels. The reaction for our purposes may conveniently be written as follows:



In this reaction succinic acid is oxidized to fumaric acid, and oxygen activated by the oxidase system is consumed. The oxygen combines with hydrogen from the succinate activated by its dehydrogenase. The system is of special interest since it is almost universally present in tissues and since nearly all respiration proceeds through it.[50] Both succino-dehydrogenase and cytochrome-cytochrome oxidase are amply present in brain.[44] Torrès [51] has shown that 97.5 per cent of respiration of rat-brain cortex involves the cytochrome system which we know functions sequentially with succino-dehydrogenase.

Stotz and Hastings have shown (a) that the enzyme as prepared by them is free of fumarase, since the amount of fumarate formed agrees stoichiometrically with the oxygen consumed, indicating that none of the fumarate formed by the oxidation of succinate is hydrolyzed to malic acid. (b) The reaction follows first-order kinetics. The value of the velocity constant does not deviate from the mean significantly for the first 85 per cent of the reaction, thus indicating that there is no destruction of the enzyme, at least for the first 85 per cent of the reaction. (c) The addition of 6.0×10^{-6} *M* NaCN, or 2.0×10^{-5} *M* selenite per 0.5 ml enzyme solution stops the oxidation of succinate completely. (d) Oxidase activity can be studied independently, since cytochrome-cytochrome oxidase can oxidase *p*-phenylenediamine without a dehydrogenase and the dehydrogenase activity can be studied independently if a hydrogen acceptor, such as methylene blue, cresyl blue, or some other suitable dye requiring no activating catalyst, is substituted for oxygen and the cytochrome system. These studies show that the addition of 6.0×10^{-6} *M* NaCN stops the oxidase activity completely and does not diminish the dehydrogenase activity, while the addition of 2.0×10^{-5} *M* selenite stops the dehydrogenase activity completely and does not diminish the oxidase activity. Thus each component has a poison specific for it. The addition of less NaCN or less selenite partially and selectively inactivates one or the other of the components.

These observations show that the oxidation of succinate to fumarate by the beef heart extract proceeds in at least two steps. Two different enzymes are involved, which act sequentially. If the view is correct that in such a system the speed of the slower step determines that of the overall reaction (in this case the rate of oxygen consumption), and if μ values are associated with enzyme-catalyzed steps, then we may expect it to follow that (1) the complex enzyme system, as extracted from beef heart, should yield a μ value characteristic of the particular step which is the slower. (2) By partially poisoning the other (faster) component of the enzyme system, a different μ value corresponding to this now slower component should be obtained. (3) This shift in μ should come abruptly with increasing concentrations of the poison. (4) The μ value should shift back to the original, if ***both components*** are

poisoned such that the original slow step is again made the slower. (5) A μ value of about 16,000 should be obtained when the cytochrome-cytochrome oxidase component, which contains catalytically active iron, is poisoned sufficiently to make this step the slow step.

An investigation was undertaken by the writer, in collaboration with Dr. Hadidian, to determine whether these deductions could be verified experimentally. All of the five propositions have been tested and the results have been found to conform to the predictions.[17]

Fig. 11 shows reaction velocity curves at different temperatures for a typical

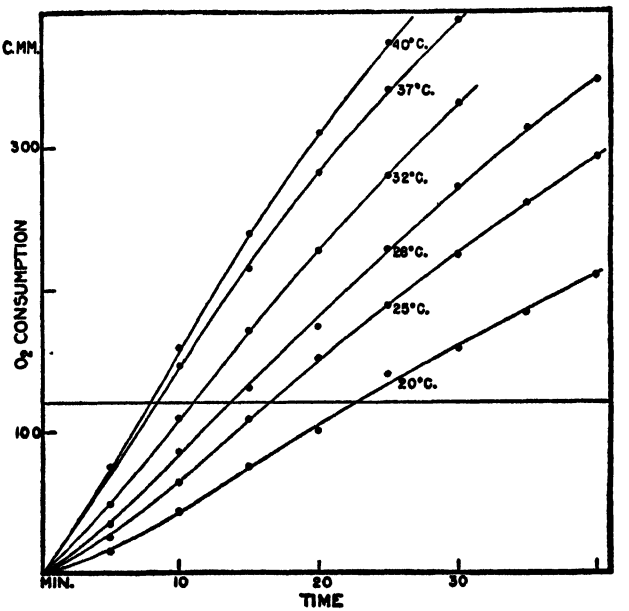


FIGURE 11. Reaction curves for a single experiment with unpoisoned enzyme at different temperatures. The horizontal line indicates the points at which velocities were measured. From Hadidian and Hoagland.[17]

experiment with the unpoisoned enzyme extract. The horizontal line indicates the points at which the slopes of the curves were determined as a measure of the velocities. Fig. 12 shows Arrhenius plots in a single experiment with varying concentrations of NaCN. The upper curve yields a μ of 11,000 calories for the normal system. Cyanide added, up to a critical amount, slows the rate at each temperature but does not change the μ, as may be seen by the constancy of slope of the upper two curves. Above this critical amount of cyanide the μ shifts to approximately 16,000 calories (*i.e.*, 15,800 calories). Fig. 13 shows Arrhenius plots of a number of similar experiments on normal and cyanide-poisoned enzyme systems, demonstrating that the critical shift from 11,000 calories, presumably the activation energy of the succinodehydrogenase step, to 16,000 calories for the cytochrome oxidase step occurs at a cyanide concentration somewhere between 1.2×10^{-7} and 1.8×10^{-7} *M*. In this figure individual experiments have been brought together by shifting the lines vertically to the abscissa. Each symbol represents a different experiment.

Fig. 14 shows what happens when one adds selenite, the dehydrogenase specific poison, to our enzyme system which has already been slowed with sufficient cyanide to yield a μ of 16,000 calories. In this case the μ shifts *back again* from 16,000

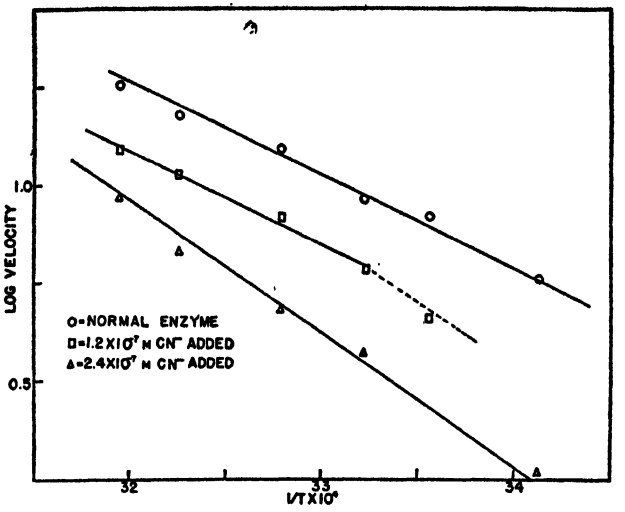


FIGURE 12. Arrhenius equation plot of a single experiment on normal and cyanide-poisoned enzyme. From Hadidian and Hoagland.[17] ○ : $\mu = 11{,}000$; □ : $\mu = 11{,}000$; Δ − $\mu = 15{,}800$.

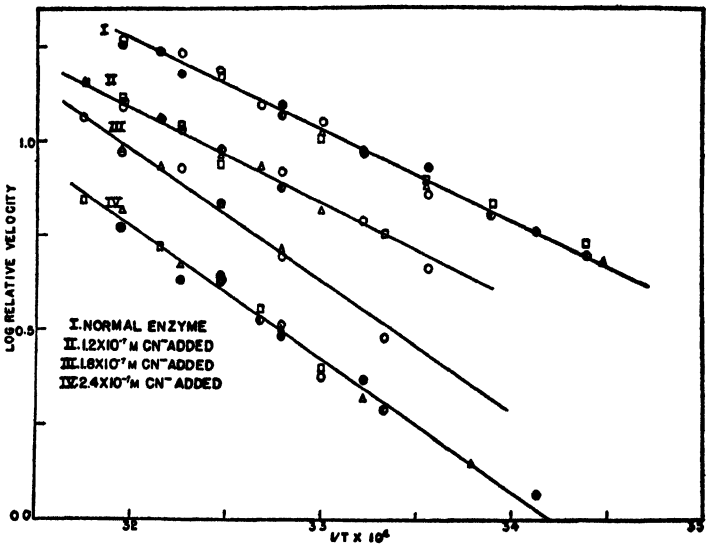


FIGURE 13. Arrhenius equation plot of a group of experiments on normal and cyanide-poisoned enzyme. Individual experiments have been brought together by shifting the lines along the ordinate. Each symbol represents a different experiment. From Hadidian and Hoagland.[17]

I : $\mu = 11{,}200$; II : $\mu = 11{,}400$; III : $\mu = 16{,}000$; IV : $\mu = 16{,}000$.

calories to that of the unpoisoned system (11,200). This clearly results from making the dehydrogenase step the relatively slower of the two.

Fig. 15 is a massed plot for the experiments showing the two different μ values, and it should be compared with Fig. 10 showing similar μ values for brain-wave frequencies in advancing general paresis. In Fig. 15 the mean heavy lines are drawn by eye through the data for the respective μ values as indicated; the lighter lines fanning out from these heavy lines are for $\pm$ 200 calories around the means.

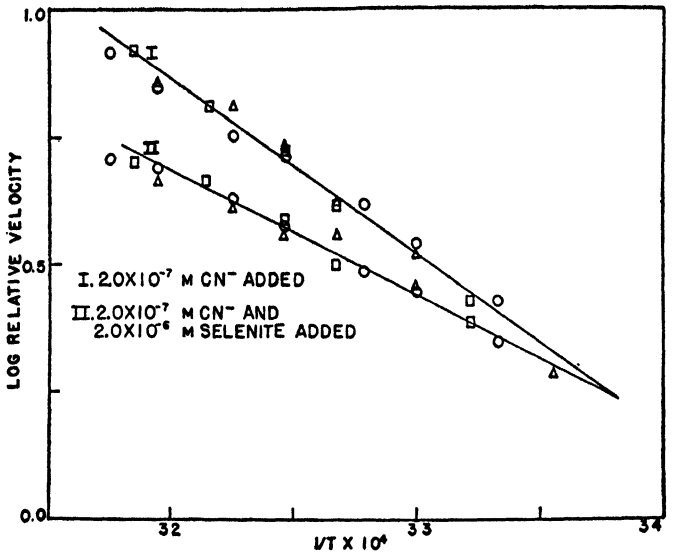


FIGURE 14. Arrhenius equation plot of a group of experiments on enzyme poisoned with cyanide and with cyanide plus selenite. From Hadidian and Hoagland.[17] I : $\mu = 16{,}000$; II : $\mu = 11{,}400$.

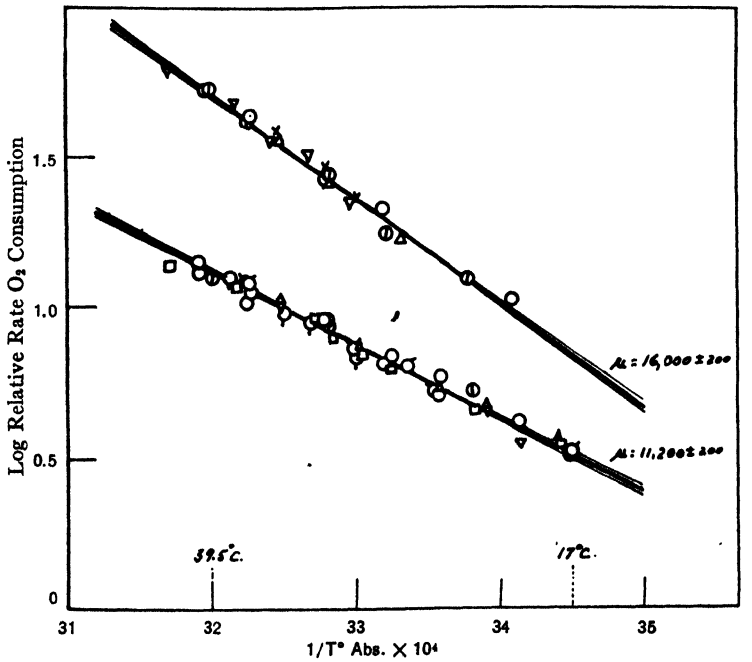


FIGURE 15. Composite massed plot for the enzyme experiments described in the text (compare with Fig. 10). Mean heavy lines are drawn by eye through the data; the lighter lines fanning out are for ± 200 calories around the means.

From these experiments it thus appears that the two frequently recurring μ values of approximately 11,000 calories and 16,000 calories for biological rates are due to the fact that succino-dehydrogenase and the cytochrome system respectively are frequently rate-limiting bottlenecks or master reactions. What reaction 8,000 calories, which among other things is the μ for normal alpha brain-wave frequencies, corresponds to, remains to be determined.

From the foregoing discussion one might be misled into taking the view that some sort of constant and invariable relation exists between μ value and enzyme specificity, not unlike the specificity existing between band spectra and substance specificity. Unfortunately no such clear-cut relation as this exists as the following experiments by Hadidian and Hoagland show. As was mentioned under item (d) above, the oxidation of *p*-phenylenediamine catalyzed by the beef-heart extract does not involve the dehydrogenase component but only the cytochrome system. Under these circumstances shall we get a μ value of 16,000 calories? The answer is that we do not. Fig. 16 shows an Arrhenius plot of two typical experiments yielding a μ

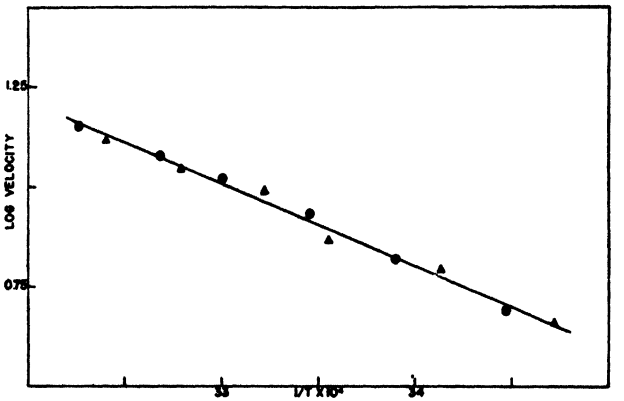


FIGURE 16. Arrhenius equation plot for oxidation of *p*-phenylenediamine. 1 x $10^{-4}M$ *p*-phenylenediamene and 0.5 ml enzyme. • = determinations made the day of preparation of enzyme. ▲ = determinations made the following day. $\mu = 9{,}500$ calories. From Hadidian and Hoagland.[18]

of 9,500 calories. Thus we see that with a substrate as different from the succinate as is *p*-phenylenediamine, a different energy of activation is involved, even though the same oxygen-activating enzyme system is presumably functioning in both cases.

We may poison the cytochrome system completely with cyanide so as to stop all oxygen uptake and then substitute for it an artificial oxygen carrier, such as the dye cresyl blue, which Stotz and Hastings [49] have shown to be especially efficient for this purpose. Fig. 17 shows oxygen consumption curves for the normal reaction and for the reaction in which cresyl blue has been substituted for the completely inhibited cytochrome oxidase system. In our dye-substituted system the dehydrogenase step is functioning, but in conjunction with cresyl blue and not with the cytochrome system. What μ values will we get by making one or the other of these steps slow ones? The simplest way to make one or the other of these components limiting, is to compare μ values for experiments in which, on the one hand, we have reduced the *amount* of the enzyme extract containing the dehydrogenase with, on the other hand, μ values obtained from experiments involving reduced amounts of added cresyl blue. Fig. 18 shows plots in which the dehydrogenase is thus made the limiting factor. In these experiments so much cresyl blue was

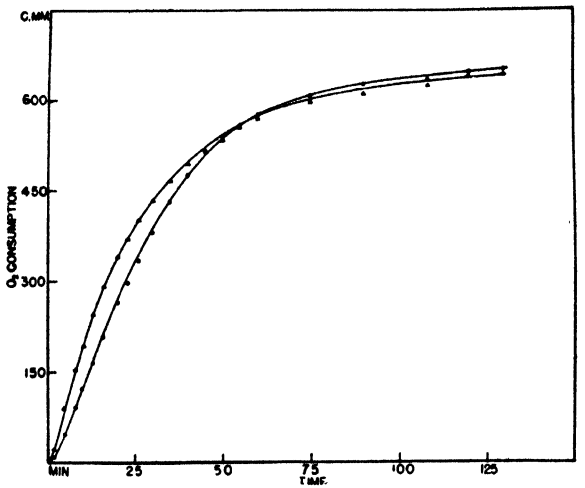


FIGURE 17. Oxygen consumption curves for the normal and the dye-substituted reactions. ● = normal: 6.0 x $10^{-5}M$ succinate, 0.5 ml enzyme. ▲ = dye-substituted: 5.0 x $10^{-6}M$ cresyl blue, 1.0 x $10^{-5}M$ cyanide, 6.0 x $10^{-5}M$ succinate, 0.5 ml enzyme. From Hadidian and Hoagland.[18]

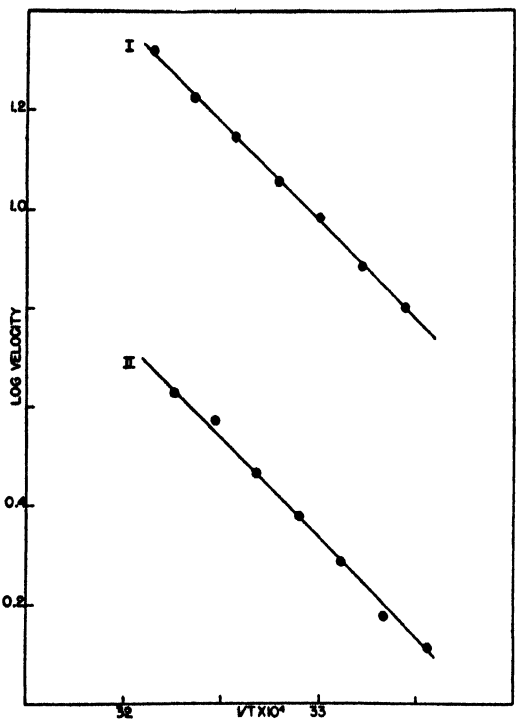


FIGURE 18. Curve I. Arrhenius plot for oxygen-consumption curves using dye-substituted enzyme as described in the legend to Fig. 17. Curve II. Arrhenius plot using one-tenth as much enzyme but the same amount of cresyl blue. From Hadidian and Hoagland.[18]

present that its further increase did not accelerate the reaction. The μ (curve I) is 18,500 calories and is presumably associated with succino-dehydrogenase which, when working normally with the cytochrome system, yields a μ of 11,200 calories. That 18,500 is the μ for the dehydrogenase working sequentially with cresyl blue is

further borne out by the fact that with one-tenth as much enzyme extract we obtain, in curve II, the same μ despite a great lowering of the reaction velocities. Fig. 19 shows a μ of 22,300 calories obtained by reducing the relative amount of cresyl blue in comparison to the enzyme concentration. This evidently is the activation energy of our artificial catalyst under the conditions of these experiments.

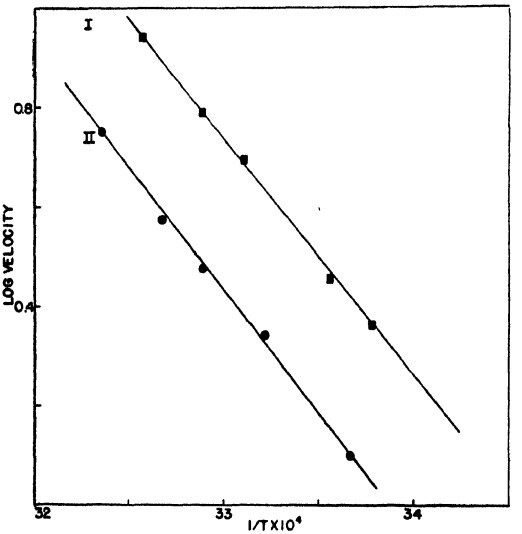


FIGURE 19. Arrhenius plots for experiments with low concentrations of cresyl blue, thus making it the limiting factor. Curve I. $6.0 \times 10^{-5}M$ succinate, $1.0 \times 10^{-5}M$ cyanide, $5.0 \times 10^{-7}M$ cresyl blue (one-tenth as much as in Fig. 18). $\mu = 22{,}000$. Curve II. $6.0 \times 10^{-5}M$ succinate, $1.0 \times 10^{-5}M$ cyanide, 0.5 ml enzyme, $2.5 \times 10^{-7}M$ cresyl blue. $\mu = 22{,}300$, *i.e.*, the μ is unchanged when the very small amount of cresyl blue is halved. From Hadidian and Hoagland.[18]

These results suggest that an energy of activation is associated, not with the enzyme as a simple chemical entity, but with the particular reaction step involving it as catalyst. Enzymes are, in general, large protein molecules which may be catalytically active at any one of several points, depending on experimental conditions. Thus the same chemical entity might, by functioning differently, yield correspondingly different μ values. It is, of course, also true that our succinate-fumarate enzyme system may contain other steps acting sequentially in addition to succino-dehydrogenase and cytochrome-cytochrome oxidase. There may be several cytochromes with their oxidases acting sequentially. The cytochrome-c reductases recently reported by T. R. Hogness at the University of Wisconsin symposium may also render this system far more complex than is indicated by our simple paradigm. From such a complex system we should expect a number of μ values to emerge with manipulation. It is thus especially interesting that in the simple, aqueously extracted enzyme we get 11,200 calories and 16,000 calories, clearly indicating respectively succino-dehydrogenase and the cytochrome system as they normally function *in vivo*.

The fact that we cannot speak of the activation energy of an enzyme as we can of the absorption spectrum of a substance need not militate against the ultimate usefulness of this concept. It is clear that throughout a vast array of cell systems the same substrates, *e.g.*, glucose, are broken down through identical steps by enzymes specific for each step. The μ value appears to be the only key at the present time to the chemical pacemaker, which, as in the case of general paresis in man, may shift under pathological conditions. As yet no work has been attempted to correlate the significance of this shift with definite symptomatology. Would it be possible to prevent the shift from 8 to 11 to 16 thousand calories by chemotherapy or other forms of treatment? Would prevention of such a shift of μ be correlated with amelioration of psychiatric symptoms? These and many other questions await investigation.

Time and Chemical Kinetics

Some years ago the writer became interested in the problem of physiological time in relation to the chemical pacemaker concept. We all possess a subjective appreciation of duration. Even when asleep our time sense continues to function, since some persons can waken themselves within a few minutes of a predesignated time. However, as this subjective or private time is variable, man has created for himself a public time scale based on the regular cyclic properties of the solar system.

Since clock time depends on relative motion, we might ask ourselves what is the source of motion determining our private or physiological time? One thing that might be considered would be molecular motions involved in steady-states maintained by continuous oxidations of foodstuffs by living tissues. Could, let us say, continuous oxidations in the brain constitute a sort of "chemical clock" from which our basic conceptions of duration are derived?

If psychophysiological time is determined by chemical kinetics, raising the internal body temperature should accelerate the reactions, thus making more physiological time pass in a given interval of clock time than would normally be the case. If, let us say, two minutes of subjective time were thus to pass in one minute of clock time we would say that time was dragging; on looking at the clock it would be slower than we think it should be.* Lowering the temperature, on the other hand, should have an opposite effect, making clock time seem to pass faster, since the slowed biochemical rates are measured as reciprocal clock times.

An opportunity to test this hypothesis presented itself in the winter of 1932 when my wife chanced to fall ill with influenza and had a temperature of nearly 104° F. I accordingly asked her to count to sixty at a speed she believed to be one per second without telling her why. As a trained musician she has a good sense of duration. This count was repeated thirty or forty times in the course of her illness; her speed of counting was measured with a stopwatch, and her temperature recorded each time. Not only did she unknowingly count faster at higher than at lower temperatures but her data conformed well to the Arrhenius equation relating chemical velocities to temperature. On the strength of this I repeated the experiment with several volunteer subjects given artificial fever by diathermy, and later I discovered some published data by François on the frequency of tapping a prescribed rhythm at different body temperatures. All of these data when grouped together not only conformed to the Arrhenius equation but yielded a temperature characteristic, or μ value, of 24,000 calories (see Fig. 20)—a value of the order of magnitude reported in several studies for biological oxidations.

These results, which were published in 1933,[25] are consistent with the view that our sense of duration, other things being equal, acts as if it were linearly proportional to the speed of some underlying chemical master reaction.

Now it is well known that some living organisms can recover after being subjected to the very low temperatures of the liquefied gases such as air, hydrogen and helium. For these organisms time, at these reduced temperatures, must pass at a greatly accelerated speed if we are right in regarding private time as a by-product of chemical kinetics. Since life and its concomitant ageing processes is a series of chemical events proceeding irreversibly toward death, these temperatures must essentially stop ageing processes. Such organisms are, for all intents and purposes, *projected forward into the future* at these low temperatures. A stay of a century at −270° C, the temperature of liquid helium, with subsequent warming and recovery, would, if it were possible to effect, be precisely equivalent to projecting the organism forward a century into the future—a sort of Wellsian time machine idea.

* It is interesting to note that in "Allgemeiner deutscher Bier-Komment," Chapter II, Part 3 ("Tempus"), p. 27, it is stated:

"Während der Kneipe wird die Zeit nach Bierminuten gerechnet. 5 Bierminuten 3 Zeitminuten." J. A.

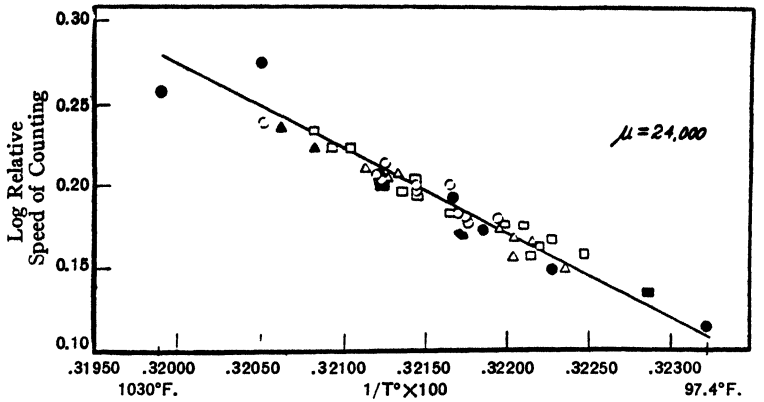


FIGURE 20. Temperature characteristic for the human time sense. Arrhenius plot for the effect of internal body temperature on the frequency of subjectively counting seconds and tapping proscribed rhythms for naive subjects. For discussion see text. From Hoagland.[25]

Professor B. J. Luyet, who has made a number of contributions to the field of low-temperature effects on organisms, has recently reviewed with Gehenio [87] the subject of life and death at low temperatures. Their review shows that some 120 studies have demonstrated that a variety of small organisms can withstand the temperature of the liquefied gases. These include some bacteria, protozoa, plant cells and a few metazoans. The procedure in general involves rapid chilling and warming so as to vitrify the organisms and not permit their water to crystallize. Crystallization of water takes time and colloidal solutions are slower to crystallize than others. The crystallization range according to Luyet extends for only some 30-40° below zero centigrade. Rapid passage through this range is, for most organisms, essential if they are to survive extremes of chilling. Moreover, the length of time spent at these low temperatures usually, as we would expect, has no effect on the percentage of revivable organisms. Organisms of more than a millimeter or so in size cannot survive this treatment, since heat cannot be conducted from and to their interiors fast enough on chilling and warming to prevent internal crystallization and death.

It recently occurred to me that while it is thus not possible to project a mammal into the future it might, nevertheless, be possible to project his offspring provided that one could vitrify and later recover fertile mammalian sperm. Such a result might be interesting in the future if social sanctions ultimately permitted its human application and its value in connection with the storing of sperm of prize animals to renew deteriorated stock generations after the donor's death is obvious.

On examining the literature we found that Luyet and Hodapp [6] had successfully vitrified and revived an appreciable number of frog sperm after removing some of their water content by plasmolysis with hypertonic sucrose solutions. They were, however, unable to revive any rat sperm. Shettles [46] in 1940, in the course of a study of the physiology of human sperm, had been able to vitrify and revive a few per cent of seminal human sperm immersed in capillary tubes in the liquefied gases and later rapidly warmed.

Dr. Pincus and I have confirmed Shettles' findings and extended them. We have been able to revive as many as 67 per cent of human sperm after exposure to liquid nitrogen at a temperature of −195° C.[88a, 88b]

We do not know how long sperm would remain viable at −195° C but we would expect them to do so indefinitely. We have kept human sperm samples stored in dry ice after vitrification in liquid nitrogen for four months, and found no decline in the

per cent motile on warming, as compared to samples kept at the low temperature for only a few minutes.

Human sperm withstand vitrification and revival better than do those from other mammals that we have examined. Our preliminary experiments with sperm from rat, mouse, guinea pig, rabbit and bull yielded completely negative results. We accordingly decided to investigate systematically rabbit sperm. We have studied sperm samples from the vas deferens of 27 rabbits and also 4 samples of ejaculated rabbit sperm. Luyet and Hodapp were unable to revive frog sperm smeared in films on mica and dipped in liquid air without first plasmolyzing them. After 3 minutes of plasmolysis in 2*M* sucrose solution, about 40 per cent of their frog sperm ceased motion, but the remainder survived immersion in liquid air if followed by very rapid warming. These revived sperm were motile as long afterwards as were controls.

We have found that this same procedure does not permit any revival of mammalian sperm just described. Positive results have been obtained by us only as follows: Occasionally we have had approximately 0.1 per cent revival of bull and of rabbit sperm after liquid-nitrogen immersion by pretreating them for several minutes with Ringer solution of from 1 to 2 times isotonicity containing double the normal amount of Ca. Very occasionally we have had a comparable yield of rabbit sperm after plasmolyzing for 5 to 10 minutes in a solution of 5 times the normal Ringer concentration. Our best, consistently reproducible, results on recovery of motile rabbit sperm after immersion at −195° C, have been obtained by first drying thin smears of otherwise untreated vas sperm nearly to completion on Cellophane in air. These, after the chilling, yield from 0.1 per cent to as much as 1.0 per cent motile sperm when rapidly warmed. Pretreatment of rabbit sperm for from 2 to 10 minutes with a solution made up of 0.01*N* butyric acid in 3-4 times the normal Ringer or serum concentration also yields positive results to the extent of about 0.1 per cent recoverable motile sperm after liquid nitrogen. All of the plasmolyzing solutions that are effective produce complete immobilization on their own account, but this is reversible when the sperm are put in isotonic Ringer serum or glucose phosphate.

We are continuing our experiments and hope that, as in the case of human sperm, we may be able to augment the viable rabbit sperm recovered from liquid nitrogen, from one per cent to a value available for practical artificial insemination. This would make possible further experiments which may be of practical value to animal husbandry in ultimately permitting the storage of the sperm of prize animals for the fertilization of offspring generations after the donor's death. It is startling to consider this aspect of physiological time in relation to human societies if social sanctions ever were to permit women to choose to have children by great men of past generations!

References

1. Adrian, E. D., "The Basis of Sensation," Christophers, London, 1928.
2. —, *Proc. Roy. Soc. London*, **106B**, 596 (1930).
3. —, and Buytendijk, F., *J. Physiol.* **72**, 377 (1931).
3a. Blinks, L. R., *Cold Spring Harbor Symp. on Quant. Biol.*, **8**, 204 (1940).
4. Bolam, T. R., "The Donnan Equilibrium," G. Bell & Sons, London, 1932.
5. Burton, A. C., *J. Cell. & Comp. Physiol.*, **9**, 1 (1936).
6. Cowan, S. L., *Proc. Roy. Soc. London*, **115B**, 216 (1934).
7. Crozier, W. J., *J. Gen. Physiol.*, **7**, 123 (1924-25).
8. —, *J. Gen. Physiol.*, **7**, 189 (1924-25).
9. —, *J. Gen. Physiol.*, **7**, 192 (1924-25).
10. —, *J. Gen. Physiol.*, **9**, 531 (1925-26).
11. Fenn, W. O., *Physiol. Rev.*, **20**, 377 (1940).
12. Fessard, A., *L'Année Psychologique*, **32**, 49 (1931).
13. Fries, E. F. B., *J. Gen. Physiol.* **10**, 227 (1925-26).
14. Gesell, R., *Ergebn. d. Physiol. biol. Chem. u. exp. Pharm.* **43**, 477 (1940).
15. Gibbs, F. A. and Gibbs, E. L., "Atlas of Electroencephalography," Privately published by F. A. Gibbs, Boston City Hospital, Boston, 1941.
16. Gould, B. S. and Sizer, I. W., *J. Biol. Chem.*, **124**, 269 (1938).
17. Hadidian, Z. and Hoagland, H., *J. Gen. Physiol.*, **23**, 81 (1939).
18. —, *J. Gen. Physiol.*, **24**, 339 (1941).
19. Hill, A. V., *Proc. Roy. Soc. London*, **111B**, 106 (1932).
20. —, *Proc. Roy. Soc. London*, **113B**, 345 (1933).

21. Himwich, H. E., Hadidian, Z., Fazekas, J. F., and Hoagland, H., *Am. J. Physiol.*, **125**, 678 (1939).
22. Hinshelwood, C. M., "The Kinetics of Chemical Change in Gaseous Systems," p. 45. Oxford University Press, 1929.
23. —, "Kinetics of Chemical Change," Oxford, 1940.
23a. Hoagland, D. R., *Cold Spring Harbor Symp. on Quant. Biol.*, **8**, 181 (1940).
24. Hoagland, H., *J. Gen. Psychol.*, **3**, 351 (1930).
25. —, *J. Gen. Psychol.*, **9**, 267 (1933).
26. —, *J. Gen. Physiol.*, **17**, 195 (1933-34).
27. —, *Cold Spring Harbor Symp. on Quant. Biol.*, **4**, 267 (1936).
28. —, *Cold Spring Harbor Symp. on Quant. Biol.*, **4**, 347 (1936).
29. —, *Am. J. Physiol.*, **116**, 604 (1936).
30. —, *J. Cell. & Comp. Physiol.*, **10**, 29 (1937).
31. —, Rubin, M. A. and Cameron, D. E., *Am. J. Physiol.*, **120**, 559 (1937).
32. —, —, —, *J. Neurophysiol.*, **2**, 170 (1939).
33. —, Himwich, H. E., Cambell, E. Fazekas, J. F. and Hadidian, Z., *J. Neurophysiol.*, **2**, 276 (1939).
33a. Hoagland, H., Pincus, G., *J. Gen. Physiol.* **25**, 337 (1942).
33b. Hoagland, H., *Scientific Monthly*, **56**, 56 (1943).
34. Lillie, R. S., "Protoplasmic Action and Nervous Action," Chicago University Press, 1923.
35. —, *J. Gen. Physiol.*, **13**, 1 (1929-30).
36. Luyet, B. J. and Hodapp, E. L., *Proc. Soc. Exper. Biol. & Med.*, **39**, 433 (1938).
37. — and Gehenio, P. M., "Life and Death at Low Temperatures," Biodynamica, Normandy, Mo., 1940.
38. Merritt, H. H., Moore, M. and Solomon, H. C., *Am. J. Syph.*, **17**, 387 (1933).
39. Navez, A. E., *J. Gen. Physiol.*, **12**, 641 (1929).
40. Osterhout, W. J. V., and Hill, S. E., *J. Gen. Physiol.*, **13**, 459 (1929-30).
41. —, *Biol. Bull.*, **6**, 369 (1931).
42. —, *Ergebnisse d. Physiol. u. exper. Pharm.*, **35**, 967 (1934).
43. —, *Physiol. Rev.*, **16**, 216 (1936).
44. Page, I. N., "Chemistry of the Brain," C. C. Thomas, Springfield, Ill., 1937.
45. Rubin, M. A., Cohen, L. H., and Hoagland, H., *Endocrinology*, **21**, 536 (1937).
46. Shettles, L. B., *Am. J. Physiol.*, **128**, 404 (1940).
47. Sizer, I. W., *J. Cell. & Comp. Physiol.*, **10**, 61 (1937).
48. —, *J. Gen. Physiol.*, **21**, 695 (1938).
49. Stotz, E., and Hastings, A. B., *J. Biol. Chem.*, **118**, 479 (1937).
50. Szent-Gyorgyi, A. V., "Perspectives in Biochemistry," p. 165 ff. Cambridge University Press, 1937.
51. Torrès, I., *Biochem. Z.* (Berlin), **280**, 114 (1935).
52. Van der Pol, B., and Van der Mark, J., *Arch. Neerland. d. Physiol.*, **14**, 418 (1929).
53. Warburg, O., Inst. Intnat. de Chimie Solvay Paris Gautier-Villard, p. 303 (1935).

The Colloid Chemistry of Purified Viruses *

MAX A. LAUFFER AND W. M. STANLEY

Department of Animal and Plant Pathology of The Rockefeller Institute for Medical Research, Princeton, N. J.

The vast field of virus research was opened in 1898 when Beijerinck,[10] after having apparently independently repeated Iwanowski's[80] discovery of 1892 that juice from plants diseased with tobacco mosaic remained infectious after passing through a porcelain filter, realized that the virulent agent differed from ordinary bacteria and described it as being a "contagium vivum fluidum." Since that time numerous virus diseases have been recognized. As is discussed in more detail in an earlier volume of this series,[86] one of the many quite diverse ideas which developed gradually throughout the subsequent years was that the agents responsible for these diseases are not fluids but discrete particles,[1, 21] usually of submicroscopic dimensions, of materials consisting largely of protein. This point of view was finally substantiated in 1935 when a crystalline protein possessing all the properties of tobacco mosaic virus was isolated by means of chemical precipitation from the juice of mosaic-diseased tobacco plants.[91] This protein has been shown to be in chemical combination with a nucleic acid resembling yeast nucleic acid and is, therefore, a nucleoprotein.[3, 57, 108] Its chemical composition and physical characteristics are quite definite and specific. It gives the usual protein color reactions, is precipitated by the usual protein-precipitating agents, has characteristic heat and pH stability ranges, and is denatured under certain definite conditions. Very dilute solutions give a

* Much of the material discussed in this chapter has been reviewed previously in several completely documented articles.[52, 95, 97, 98]

specific precipitin reaction with antiserum to the protein. The protein is insoluble at its isoelectric point and in 20 per cent ammonium sulfate solutions. It crystallizes in definite needle-shaped crystals in the paracrystalline or mesomorphic state, visible only with the microscope.[91] The nucleoprotein was found to possess an unusually high sedimentation rate,[22] and, based on this property, a method of isolation and purification by high-speed centrifugation was developed.[111] The latter means proved to be particularly useful in isolating viruses that are too unstable to undergo the more rigorous chemical treatment. By one or the other of these two techniques, the viruses of tobacco mosaic,[91] aucuba mosaic,[98] enation mosaic,[3] rib-grass mosaic,[24] tobacco ring spot,[104] latent mosaic of potato,[60] cucumber mosaics 3 and 4,[4, 86] tomato bushy stunt,[5, 99] tobacco necrosis,[79] alfalfa mosaic,[81] influenza,[16] foot-and-mouth disease,[31] Shope rabbit papilloma,[8] equine encephalomyelitis,[109] and mouse encephalomyelitis,[26] the elementary bodies of vaccinia,[20] and certain bacteriophages[72] have been isolated in the form of materials of high molecular weight consisting largely of protein and in some instances obtainable in the crystalline state.

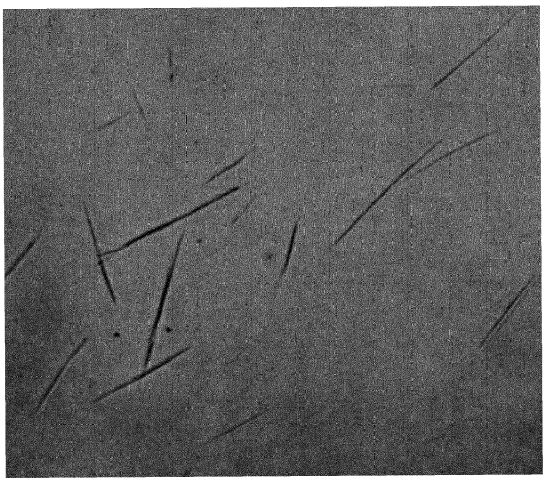


FIGURE 1. Paracrystals of tobacco mosaic virus protein. (X 675). (From Stanley.[95])

The question naturally arises whether these nucleoproteins are actually the viruses or merely materials contaminated by the virulent agents. The answer has been worked out in greatest detail in the case of tobacco mosaic virus, and all of the evidence at present available indicates that this virus nucleoprotein molecule is actually the active disease-causing entity. The virus protein isolated from many different batches of diseased Turkish tobacco plants and from other plant species infected with the virus possesses the same chemical, physical, biological, and immunological properties.[59] This protein is capable of infecting other plants, will multiply in them, and can be isolated from such diseased plants. There is no reason to believe that the protein is not pure, for its chemical, physical, biological, and immunological properties remain unchanged following fractionation of the protein by various procedures.[94] The preparations of the material are completely homogeneous with respect to electrochemical behavior.[22] It was found impossible to demonstrate the presence of an impurity in such preparations even by the sensitive precipitin and anaphylactic tests. It has never been found possible to separate the virus activity from the protein by any one of several procedures. The ultraviolet light absorption spectrum of the protein agrees essentially with the destruction spectrum of virus activity.[54] The pH stability range of the protein was found to coincide with that of virus activity.[95, 106] Partial or complete denaturation of a preparation by each of

several procedures was found to result in a corresponding loss of virus activity.[42] Finally, it is possible not only to inactivate and reactivate the virus protein, but also to demonstrate that the inactivation and reactivation are accompanied by simultaneous changes in the structure of the protein molecule.[82]

It was evident as early as 1916, when Allard found that tobacco mosaic virus passed through a Berkefeld filter but was held back by a Livingstone atmometer porous cup, that the tobacco mosaic virus consisted of particles of dimensions which fall within the colloidal range.[1] Subsequent measurements of the filtration,[21, 106] sedimentation,[9] and diffusion of tobacco mosaic virus in unpurified and in semi-purified preparations substantiated this conclusion completely. When plant juice containing tobacco mosaic virus was found to show stream double refraction, it was concluded that the particles of the virus or of something regularly associated with it are rod-shaped bodies.[105] Subsequent studies have shown that several, but by no means all, other viruses are also rod-shaped.[5, 50, 51] The interpretation of the properties of colloidal systems containing extremely anisometric particles presents a real challenge to the colloidal chemist, whereas the interpretation of the behavior of essentially spherical colloidal particles is a much simpler matter. Therefore, in considering the colloidal chemistry of viruses, it is natural to divide them into two general groups: rod-shaped viruses and essentially spherical viruses. Included in the former are tobacco mosaic virus and its many strains, the latent mosaic of potato virus, and two cucumber mosaic viruses. The latter group includes the viruses of tomato bushy stunt, tobacco necrosis, and alfalfa mosaic, and most of the viruses causing diseases of animals that have thus far been studied sufficiently to permit classification.

Rod-shaped Viruses

Tobacco mosaic virus has been studied more thoroughly than any other representative of the rod group. From a knowledge of its general colloid-chemical behavior, the properties of the other rod-like viruses can usually be inferred. Except when otherwise stated, therefore, the immediately ensuing discussion refers to tobacco mosaic virus. One of the most striking physical characteristics of this ma-

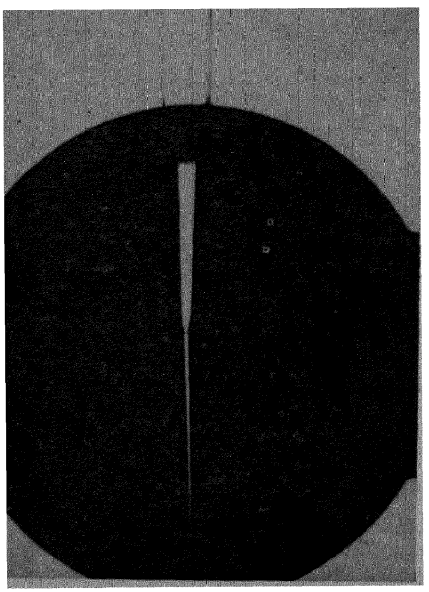


FIGURE 2. Doubly refracting stream of a tobacco mosaic virus solution photographed between crossed Polaroid plates. (From Lauffer and Stanley.[82])

terial is its ability to exhibit strong double refraction of flow. The phenomenon was first observed by Takahashi and Rawlins [105] with the juice of mosaic-diseased tobacco plants; and since the isolation of the virus in the form of a nucleoprotein, the phenomenon has been studied in many laboratories.[3, 7, 32, 41, 42, 51, 61, 65, 80, 105] Double refraction of flow could be due to the orientation of rod-like particles, to the orientation of plate-like particles, or to the photoelastic effect. Takahashi and Rawlins [105] eliminated the second possibility by showing, by a method described by Freundlich, that the optic axis of the entire stream of virus flowing from a capillary is in the direction of the flow. The third possibility was eliminated by the authors,[51] who showed that double refraction persists in a stream of virus for a considerable time after strains due to flow have ceased. According to the theory of Wiener, this double refraction of flow could be due to the orientation of either anisotropic rods (intrinsic double refraction) or isotropic rods of refractive index different from that of the dispersion medium (morphic double refraction). Intrinsic double refraction may be positive or negative; morphic rod double refraction must always be positive. The stream double refraction of tobacco mosaic virus is positive, and it was found to decrease markedly as the index of refraction of the solvent was allowed to approach that of the virus, indicating that at least a considerable fraction of the birefringence is morphic.[42] This conclusion is entirely consistent with the observation that dried films of the virus show very much less double refraction than wet films.[7] The double refraction of flow of virus solutions increases as the isoelectric point is approached from either side.[41] It is thought that this is caused by linear aggregation of the virus particles.

The Kerr electro-optical effect of tobacco mosaic and two other viruses has been studied in the authors' laboratory, and it was found that the electrical double refrac-

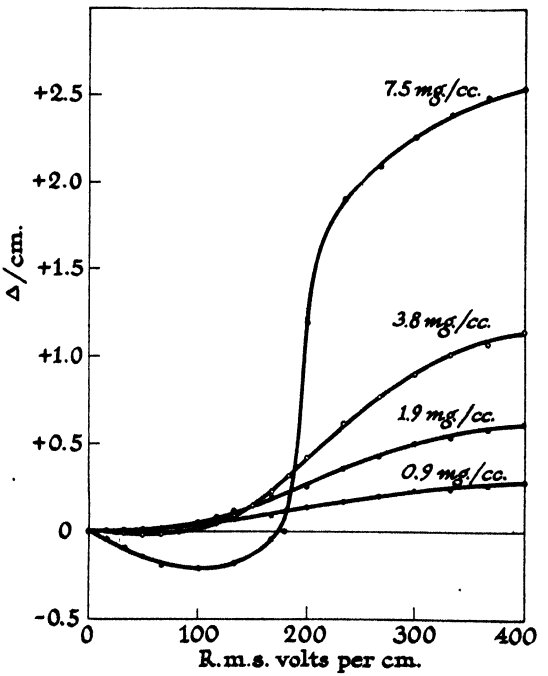


FIGURE 3. Double refraction, expressed as wave length displacements per centimeter, of various concentrations of tobacco mosaic virus at pH 6.4 plotted as a function of the strength of an alternating electric field of 60 cycles. (From Lauffer.[48])

tion in an alternating field of 60 cycles was negative for weak electrical fields but became positive for stronger fields.[48] The inversion point between positive and negative double refraction was found to vary with the concentration of the virus, its age, and the hydrogen ion concentration. In general, modifications of the virus solutions which cause an increase in the forces of attraction between particles were accompanied by an increase in the negative effect and by an inversion point at higher field strengths. Analogous inversions of the sign of electrical double refraction were observed by Norton [73] and by Mueller [69] in bentonite sols. In this case it was found not only that the field strength corresponding to the inversion point decreased upon dilution of the colloid, but also that it decreased when the frequency of the alternating field was increased. In this case, too, a correlation between particle interaction and the magnitude of the negative component was observed. Mueller and Sakmann,[69] therefore, suggested that this component could conceivably be due to the influence of the electrical field on the interaction, so as to cause an anisotropic distribution of particles. Further studies with bentonite carried out in Mueller's laboratory showed that the positive component of the double refraction fluctuated in intensity with twice the frequency of the electric field at low frequencies, but that the fluctuations disappeared entirely at high frequencies. This is an ideal relaxation phenomenon, and it shows that the positive electrical double refraction is probably caused by orientation of the bentonite particles. Professor Mueller has graciously

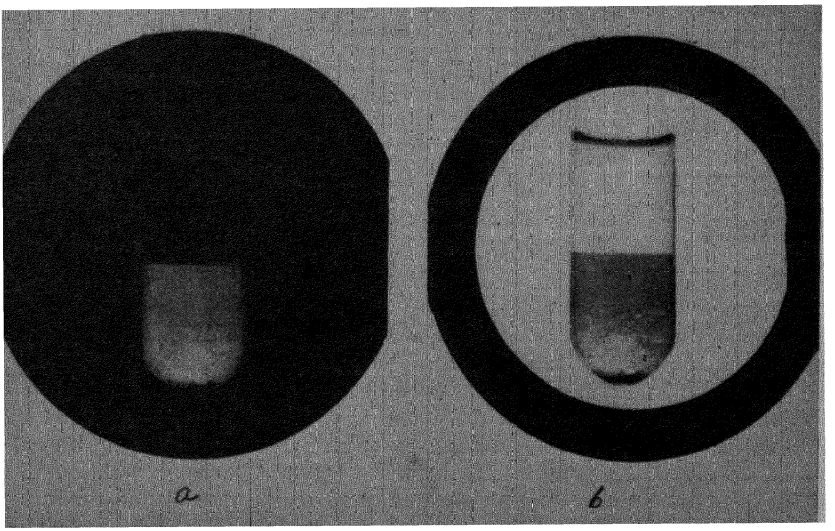


FIGURE 4. Layered solution of tobacco mosaic virus protein viewed through crossed Polaroid plates, a, and parallel plates, b. (From Lauffer and Stanley.[52])

permitted the authors to state that preliminary studies on tobacco mosaic virus made in his laboratories show that the electro-optical properties of the virus parallel those of bentonite in all the respects discussed above. Hence, it seems reasonable to suppose that the explanations advanced to account for the electro-optical properties of bentonite also apply to tobacco mosaic virus.

Bawden *et al.*[6, 7] observed that, in relatively concentrated solutions of tobacco mosaic virus, a liquid crystalline phase separates in the form of micro tactoids and settles to the bottom of the solutions. This behavior is analogous to that described previously by Zocher and Jacobsohn for vanadium pentoxide sols.[218] The pellets obtained by ultracentrifuging solutions of the virus are also liquid crystalline.[53] It

is evident that these liquid crystalline materials are composed of rather poorly defined mosaics of small volumes, in each of which all of the virus particles are oriented parallel to one another.[3, 51] Best [18] has found, further, that the virus separates from 0.4*M* solutions of $(NH_4)_2SO_4$ at pH 5 in the form of long, very thin fibers showing positive double refraction. Bernal and Fankuchen [11] observed that, when solutions of tobacco mosaic virus are evaporated under certain conditions, films of dry or semi-dry protein are obtained in which all the particles are lined up parallel. These films will swell in water and finally disperse, but in ammonium sulfate or in buffered solutions they swell to equilibrium positions which are functions of the ionic strength and the pH.

Attempts have been made by Levine [56] and by Langmuir [88] to account for these phenomena in terms of long-range interparticle forces. It was pointed out that it follows from the Debye-Huckel theory that the net effect of the electrical charges on colloidal particles, considered in conjunction with their necessarily present gegenions, is attraction at long distances and repulsion at short distances. In such a case, one would expect to have a potential energy minimum at some definite distance from each particle, as is illustrated in Fig. 5. If the depth of the potential trough is greater than *kT*, this position ought to determine the distance between particles in an equilibrium gel. Furthermore, although the theory has not been

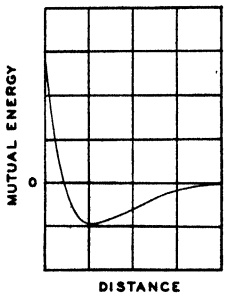


FIGURE 5. Potential energy plotted as a function of the distance between two charged colloidal particles surrounded by ion atmospheres. (From Bernal and Fankuchen.[11])

worked out in detail for such a case, the occurrence of a potential trough would provide a mechanism for maintaining a perfectly parallel orientation of rod-like particles.

The applicability of these theoretical considerations to the case of tobacco mosaic virus protein has been tested in a qualitative sense by Bernal and Fankuchen.[11] For the case of spherical particles, Levine showed that the depth of the potential trough should increase, and the distance of the potential energy minimum from the particle should decrease when the concentration of electrolyte is increased. Bernal and Fankuchen measured the equilibrium distances between virus particles in gels in equilibrium with ammonium sulfate solutions of various concentrations, using x-ray diffraction techniques. Their data are presented graphically in Fig. 6. The curve fitting the data has the shape predicted by Levine's theory, but some of the con-

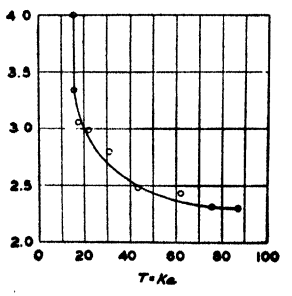


FIGURE 6. Lateral distance between parallel tobacco mosaic virus rods in an equilibrium gel plotted as a function of the $(NH_4)_2SO_4$ concentration. (From Bernal and Fankuchen.[11])

stants had to be assigned arbitrarily to make it fit. In any case, the data are in qualitative accord with the theory. A further qualitative test is afforded by the results of Best,[18] who determined the concentrations of virus at which the liquid crystalline and isotropic phases (see Fig. 4) can coexist. He found that in systems at pH 7 and 30° C the two phases coexist at virus concentrations between 1.4 and 2.2 per cent in the absence of electrolyte and at concentrations between 1.8 and 2.9 per cent in NaCl solutions more concentrated than 0.005*M*. These results would seem to indicate that the equilibrium distance in the oriented phase decreases with increase in electrolyte concentration—a result in qualitative accord with the Levine theory. In view of these results, it seems reasonable to conclude that one can establish a fair case for the existence of long-range intermolecular forces of the Levine-Langmuir type in tobacco mosaic virus systems.

All the properties thus far considered are, of course, a consequence of the anisometric nature of the virus particles. As was pointed out, the stream double refraction of the viruses is that characteristic of rod-like rather than plate-like particles. As will be seen later, the x-ray diffraction patterns of the rod-like crystals of the virus support the interpretation that the virus particles themselves are rod-like bodies and, as final proof, direct micrography with the electron microscope shows that the particles are rod-like.

Many attempts have been made to measure the size of tobacco mosaic virus particles. The most frequently used method is that of ultracentrifugation. The molecular weight, *M*, of a particle may be expressed as a function of the sedimentation velocity in unit centrifugal field, *S*, according to equation (1).

$$M^{2/3} = \frac{6\pi\eta_0 NS\left(3\frac{V}{4\pi N}\right)^{1/3}}{(1 - V\rho)\frac{f_0}{f}} \qquad (1)$$

η_0 is the viscosity of the solvent, N Avogadro's number, V the partial specific volume of the solute, ρ the density of the solution, and f_0/f the dissymmetry factor or the ratio of the frictional factor the particle would have if it were a sphere to that it actually does have. Bechhold and Schlesinger[9] were the first to study the rate of sedimentation of the virus in plant juice. Their results, which were based on virus activity measurements, are in good agreement with the results obtained later on the purified protein by Eriksson-Quensel and Svedberg,[22] by Wyckoff,[107, 108, 111] and by the authors.[44] The latter workers found that the protein isolated and purified by chemical means is quite heterodisperse, whereas that isolated and purified by means of differential centrifugation gives a single reasonably sharp boundary in the ultracentrifuge. Values for the sedimentation constant, S_{20}, as low as 174×10^{-13} and as high as 193×10^{-13} cm/sec/dyne/gram have been reported for the centrifugally isolated material. It has been shown recently that the sedimentation constant of the infection principle of the virus agrees within a probable error of 8% with that of the virus protein.[40] Because of the extremely anisometric character of the virus particles, many difficulties, the most important of which are discussed in the following paragraphs, are encountered in interpreting these sedimentation results.

The validity of the assumption, inherent in Stokes' law, that the sedimenting virus particles are independent of one another has been questioned, especially by Frampton.[23, 25] The spontaneous formation of liquid crystalline layers of the virus must be due to forces of attraction between particles in moderately concentrated solutions.[7, 11] The viscosity results of Stanley[96] and Lauffer[41] indicate that, in virus solutions as concentrated as 1 per cent or perhaps even somewhat less, one encounters interaction between particles. The diffusion results of Neurath and Saum[71] would indicate that such interaction was present in solution more concentrated than 0.5 per cent. The studies of Frampton and Saum,[25] who measured the viscosity of the

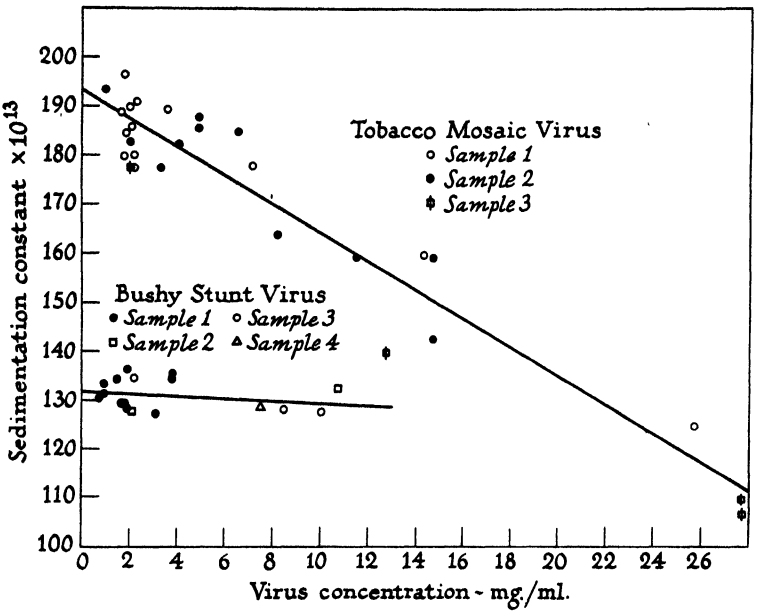


FIGURE 7. Sedimentation constant plotted as a function of tobacco mosaic virus concentration. (From Lauffer.[44])

virus with varying rates of shear, show that even in very dilute solutions the viscosity coefficient is not independent of the rate of shear; that is, anomalous viscosity is encountered. These results have led Frampton to conclude that the fundamental assumption of Stokes' law is violated by even very dilute solutions of the virus. A direct experimental approach to this problem was made in the authors' laboratory.[44] Sedimentation measurements were carried out on preparations of purified tobacco mosaic virus at various concentrations. As is shown in Fig. 7, there is a strong dependence of the sedimentation rate on virus concentration—a relationship which seems to hold even for dilute solutions. This behavior is analogous to that of the linear high polymers. If all the assumptions inherent in Stokes' law were fulfilled by tobacco mosaic virus solutions, such a relationship would not obtain. It is a matter of definition that at infinite dilution suspended particles must be mutually isolated. Therefore, by extrapolating the sedimentation-concentration data to zero concentration, a value of the sedimentation constant for a solution meeting the theoretical requirements can be obtained. The obvious way to extrapolate is simply to extend the straight line which best fits the experimental data. Unfortunately, however, an element of uncertainty attends any projection of this sort.

It seems very probable that the retardation of the sedimentation rate is at least in part a manifestation of the same fundamental attribute of the virus particles as that which is responsible for the anomalous viscosity in concentrated solutions, for the thixotropic nature of concentrated solutions, for the formation of birefringent phases, and for the retarded diffusion in concentrated solutions. As was discussed previously, the latter phenomena can be accounted for in terms of interaction of an electrostatic nature between virus particles. Only virus solutions more concentrated than 1.4 per cent will separate into two phases.[18] The retardation of the diffusion rate of the virus vanishes upon dilution.[71] The anomalous negative electro-optical behavior, which can be accounted for in terms of interaction, disappears upon dilution.[48] These results lead to the expected conclusion that this interaction varies in

some inverse manner with the distance between particles. Viscosity increments imparted by the virus to its solutions are, with a high degree of precision, directly proportional to the amounts of virus added for concentrations up to 0.1 per cent [24, 41, 52] (Fig. 8). Double refraction of flow is almost exactly proportional to the amount of virus for all concentrations less than 0.5 per cent [65, 80] (Fig. 9). Both of these behaviors are characteristic of ideal systems, and they support the conclusion that interaction plays a negligible role in virus solutions more dilute than 0.1 per cent. The only apparent obstacle to accepting this conclusion is the fact that even very dilute solutions exhibit anomalous viscosity.[28] However, this can be explained without reference to particle interaction, for Robinson,[80] by measuring simultaneously double refraction of flow, average orientation of particles, and resistance to flow, was able to account for the anomalous viscosity of very dilute tobacco mosaic virus solutions in a satisfactory manner as being due to the change in the average orientation of mutually independent rod-like particles caused by changing the flow velocity gradient (Fig. 10). Hence, it seems safe to conclude that tobacco mosaic virus solutions as dilute as 0.1 per cent do not depart much from the ideal condition and that the sedimentation constant of the virus in such solutions ought to be very nearly the same as that at infinite dilution. The method adopted above for extra-

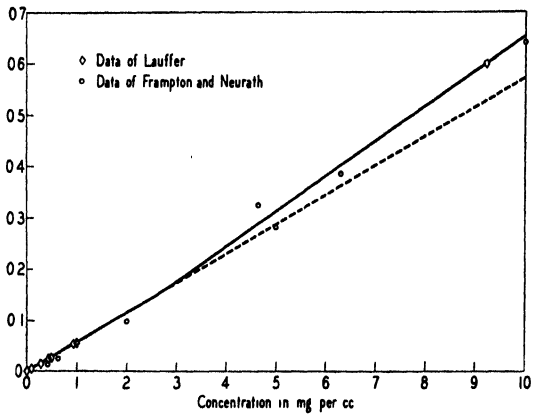


FIGURE 8. The relationship between the specific viscosity and the concentration of tobacco mosaic virus. (From Lauffer and Stanley.[52])

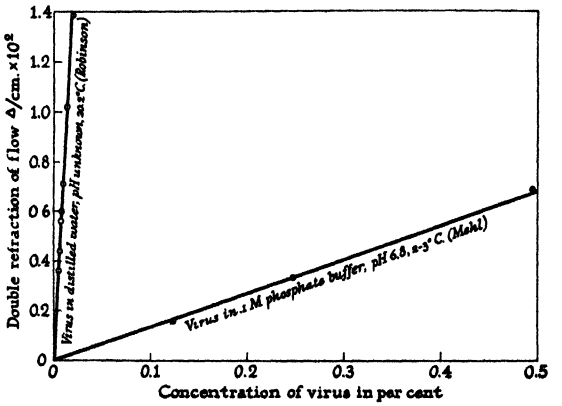


FIGURE 9. The relationship between virus concentration and stream double refraction, expressed as wave length displacements per centimeter, in a velocity gradient of 3.7 sec^{-1}. (From Lauffer and Stanley.[52])

polating the sedimentation-concentration data yields a value for the sedimentation constant at infinite dilution which differs from that for a concentration of 0.1 per cent by less than the error inherent in the measurement. Therefore, in view of the foregoing, it ought to be a reasonable estimate of the true constant at infinite dilution.

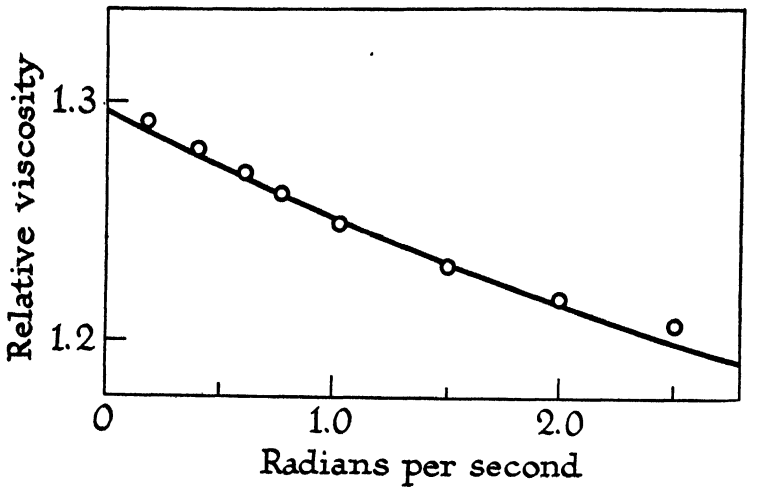


FIGURE 10. The relative viscosity of a 0.02% solution of tobacco mosaic virus at 14.4° C. plotted as a function of shear velocity. The points are experimentally found, and the curve calculated from measurements of stream double refraction and angles of isocline by means of an equation derived by Robinson. (From Lauffer and Stanley.[52])

A difficulty is encountered in interpreting the sedimentation data of asymmetrical viruses because the frictional coefficient of rod-like particles is very different from that of spheres. The usual method of overcoming this obstacle, the study of sedimentation equilibrium, was found not to yield satisfactory results because of technical difficulties.[22] From the ratio of length to thickness of particles it is possible, by making use of equation (2) (Herzog *et al.*,[29] Perrin [75]), to calculate a factor, commonly known as a dissymmetry constant, which is meant to correct for the deviation of the frictional coefficient of rod-like particles from the value given by the usual form of Stokes' law.

$$\frac{f}{f_0} = \frac{(a/b)^{2/3}}{\sqrt{1-(a/b)^2}} \ln \frac{1+\sqrt{1-(a/b)^2}}{\frac{a}{b}} \qquad (2$$

where a and b are the minor and major axes of an ellipsoid of revolution. Many attempts have been made to find a relationship between the viscosity of solutions of asymmetrical particles and the relative dimensions of the particles. Polson [77] has shown that an empirical equation differing only slightly from Kuhn's theoretical one,[37] when applied to a great many proteins, yields values for the ratio of particle length to thickness which are in excellent agreement with values obtained by an independent method. Lauffer [41] and Frampton and Neurath,[24] using Kuhn's *. [37]

* Simha [85] has more recently derived a somewhat different equation relating relative viscosity to particle dimensions. This equation seems to be the most satisfactory yet obtained,[45] but it was not available at the time the studies on tobacco mosaic virus referred to in the text were carried out.

equation (Eq. 3), estimated the ratio of length to thickness of the virus particles from viscosity data obtained in Ostwald viscometers and obtained a value of about 35:1.

$$\frac{\eta}{\eta_0} = 1 + 2.5G + \frac{G}{16}(a/b)^2 \tag{3}$$

where G is the volume of the dispersed material per cc of solution and $\frac{\eta}{\eta_0}$ is the relative viscosity of the solution. The dissymmetry constant was then calculated by equation (2) and from it, Wyckoff's[107] sedimentation data, and a partial specific volume of 0.73, the molecular weight was estimated by Lauffer,[41] using equation (1), to be about 43×10^6, corresponding to a particle 430 mμ long and 12.3 mμ in diameter.

It was pointed out that several limitations are inherent in this method. They include the necessity of assuming that hydration and the electro-viscous effect are negligible and that the rod-like particles are randomly oriented in the flowing stream in the viscometer. It was shown experimentally that the electro-viscous effect was of some importance, and it was stated that a considerable degree of orientation of particles should be expected in a capillary viscometer. It was, however, pointed out that these two errors tend to compensate each other, at least partially. Robinson[80] has emphasized the importance of the orientation of rod-like particles in respect to their contribution to the viscosity of their solutions and has devised a method of studying simultaneously the stream double refraction, the average orientation of particles (angle of isocline), and the viscosity of a solution of asymmetrical particles at various velocity gradients. By means of an approximate equation derived by him involving all these factors, he estimated the relative dimensions of the particles to be about 80 to 1. This method has been very useful in elucidating the nature of the anomalous viscosity of very dilute solutions of tobacco mosaic virus, but the value for the relative dimensions of the virus particles is in poorer agreement with the values derived from independent methods than those obtained by the less elegant method involving the use of Kuhn's equation. This may be due in part to aggregation in the virus preparations used by Robinson, as he suggested, especially since the preparations were obtained by chemical means.

By assuming that the frictional coefficients in sedimentation and in diffusion are the same, an equation expressing the molecular weight as a function of the sedimentation constant and the diffusion constant, D, can be derived (Eq. 4).

$$M = \frac{RTS}{D(1 - V\rho)} \tag{4}$$

where R is the gas constant and T is the absolute temperature. Using this equation, it is possible to calculate the size of particles of any shape from sedimentation and diffusion data. Neurath and Saum,[71] using the method of Lamm, estimated the diffusion constant of the virus to be about 3×10^{-8} cm^2 per sec. The molecular weight calculated from this value and Wyckoff's sedimentation constant of 174×10^{-13}, using equation (4), is about 60×10^6. From the sedimentation and diffusion data, Svedberg's dissymmetry factor may readily be calculated, and from it, employing the Herzog-Perrin equation (Eq. 2), the ratio of the dimensions of the rodlike particles may be estimated. This value is 58 for tobacco mosaic virus, corresponding to particles 675 mμ long and 11.6 mμ in diameter. Even if the seemingly reasonable assumption that the particles are essentially mutually independent in the dilute solutions used for the diffusion and sedimentation measurements is granted, a limitation to this method still remains. The assumption that the frictional resistances in diffusion and in sedimentation are the same is strictly valid only if

the orientation of the particles is random in both cases. It is entirely possible that the randomness is disturbed somewhat in the centrifuge, due to an orienting tendency caused by the radial gradient of the centrifugal field and not compensated for by thermal agitation because of the extremely large size of the virus particles.

Mehl [65] studied the double refraction of flow and the average orientation of particles of tobacco mosaic virus in an apparatus resembling somewhat a Couette viscometer. Following a method due to Boeder,[15] he estimated the rotational diffusion coefficient of the virus to be 25 at 0° C. By substituting this in an expression derived by Kuhn [87] relating the rotational diffusion coefficient to the reciprocal of the third power of the length of rod-like particles, he obtained a value of 610 mμ for the length of virus particles.

Ever since the agent was first recognized, tobacco mosaic virus has been the subject of filtration experiments conducted by numerous workers. A painstaking study of many strains of tobacco mosaic virus and of other plant viruses, some of which are symmetrical, was carried out by Thornberry,[106] using the carefully standardized method of Elford. He found that all the viruses which he studied, when diluted with solvents containing nutrient broth, would just pass through filters with an average pore diameter of 45 mμ. Tobacco mosaic virus purified by the modified lead acetate method of Stanley just passed a filter with a pore diameter of 34 mμ. If it is assumed that rod-shaped particles pass through the pores of the filters lengthwise and that the ratio of pore diameter to particle diameter obtained by Elford for symmetrical particles holds for the diameters of the rods, these results would indicate that the diameters of the rods in unpurified samples of tobacco mosaic virus and its related strains fall within the range of 15-22 mμ and that of the purified virus within the range of 11-17 mμ. These dimensions are thus seen to be in agreement with those determined by a variety of other methods. This is true for all the cases investigated by Thornberry, including symmetrical and asymmetrical viruses. Smith and MacClement [89] did not confirm his results in their earlier experiments, but in their more recent study they found that some tobacco mosaic virus would pass filters with pore diameters as low as 40 mμ. These workers found, moreover, that in an electrical field the tobacco mosaic virus will pass through filters with an average pore diameter of 13 mμ. One may conclude that the probability that the virus particles will squeeze through the pores of the filters is greatly increased by the orientation of the virus particles parallel to the pores by the electric field and by the tendency for the particles to migrate in an electric field.

Langmuir and Schaefer [89] found that they were able to adsorb a film of tobacco mosaic virus on a surface containing a monolayer of egg albumin. By an interference method, they measured the thickness of the film to be 12.5 mμ. They interpreted this to be a film one molecule thick of virus particles lying flat on the surface.

Bernal and Fankuchen [11] reported studies of the x-ray diffraction patterns of dilute solutions of tobacco mosaic virus, of liquid crystalline solutions, of films of the virus protein obtained by evaporation and designated as wet and dry gels, and of oriented virus crystals. Wyckoff and Corey [112] studied the diffraction pattern of crystals and of pellets obtained with the ultracentrifuge. The findings from these two laboratories agree on what Bernal and Fankuchen regard as the intramolecular structure of the material. Bernal and Fankuchen regard the unit which is usually thought to be the molecule as having an internal structure of great regularity, analogous to the structure of crystalline proteins. The patterns which indicate this structure are identical for the protein in all states studied, including solutions, and therefore must define the molecule rather than the crystal. This constancy of internal structure is evidence that the virus particles do not swell in solution. A lateral interspacing of 15.2 mμ is found in the dry gel studied by Bernal and Fankuchen. This spacing becomes wider and wider in wet gels and in liquid crystalline solutions

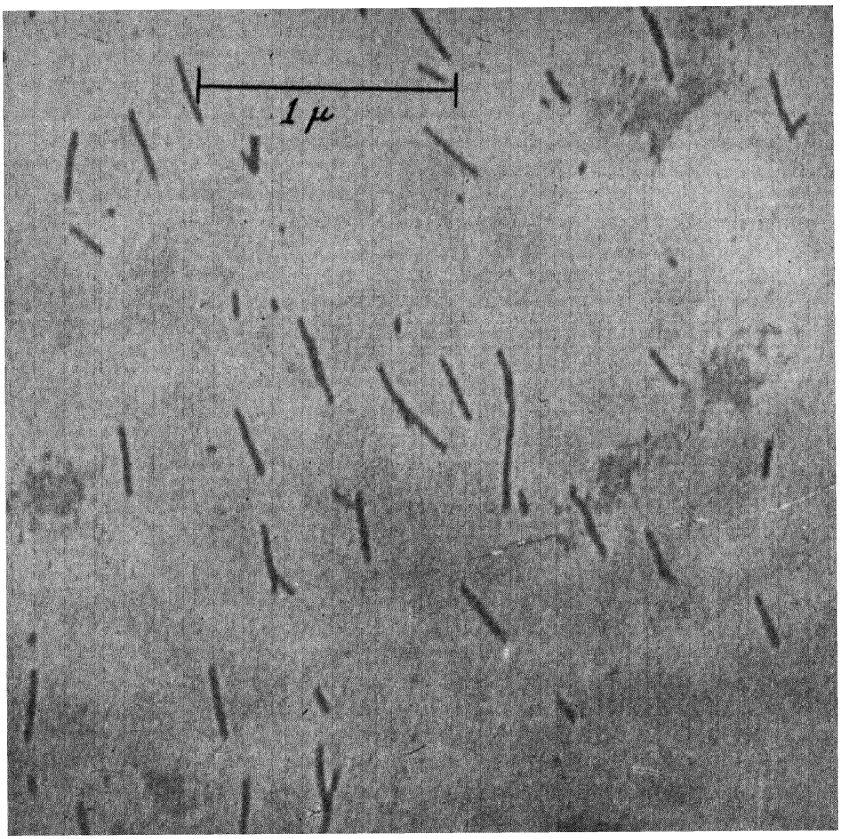


FIGURE 12. Electron micrographs of tobacco mosaic virus particles. (X 35,000). (From Stanley and Anderson.[100])

could be isolated from plants diseased with various strains.[91, 93] Recent studies by Knight and Stanley have shown that the nucleoproteins of at least several of these strains differ slightly, but significantly, in amino acid composition.[34, 36] The stream double refraction studies of the authors [51] and the x-ray diffraction studies of Bernal and Fankuchen [11] showed that, in general, the particles of these nucleoproteins had similar shapes. However, the sedimentation results from Wyckoff's [107, 108, 110] laboratory first suggested that there were minor differences in the sizes of the particles of the various strains. Recent studies have tended to substantiate that indication. Pfankuch, Kausche, and Stubbe [74] recently isolated several tobacco mosaic virus strains. A comparison of one of their strains, designated Tm 44 by them, with what they considered normal tobacco mosaic virus may serve as an example of how the protein particles can vary from strain to strain. Tm 44 was found to have a higher specific turbidity, twice the specific viscosity per unit concentration, a sedimentation constant greater by roughly 14 per cent, and an electrophoretic mobility at 0° in phosphate buffer at pH 6.9, $s = 0.1$, about 12 per cent less than that of the protein isolated from the normal strain of the virus. In general, these data may be interpreted to mean that the particles of Tm 44 are longer and heavier than those of the normal virus. Knight and Lauffer [35] found that the rib-grass strain of tobacco

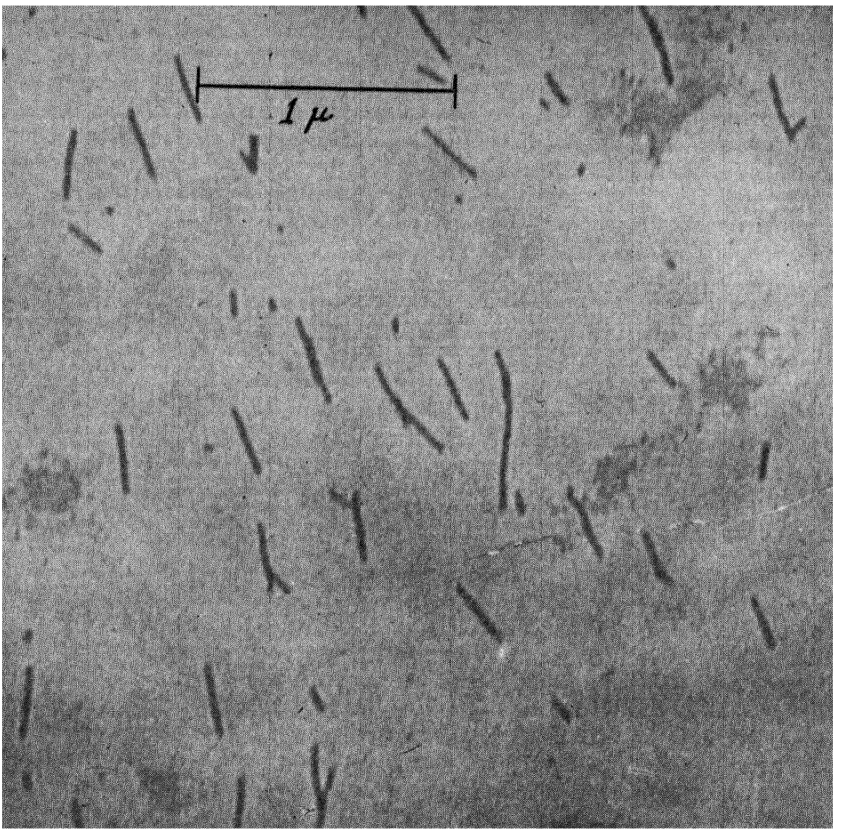


FIGURE 12. Electron micrographs of tobacco mosaic virus particles. (X 35,000). (From Stanley and Anderson.[100])

could be isolated from plants diseased with various strains.[91, 93] Recent studies by Knight and Stanley have shown that the nucleoproteins of at least several of these strains differ slightly, but significantly, in amino acid composition.[34, 36] The stream double refraction studies of the authors [51] and the x-ray diffraction studies of Bernal and Fankuchen [11] showed that, in general, the particles of these nucleoproteins had similar shapes. However, the sedimentation results from Wyckoff's [107, 108, 110] laboratory first suggested that there were minor differences in the sizes of the particles of the various strains. Recent studies have tended to substantiate that indication. Pfankuch, Kausche, and Stubbe [74] recently isolated several tobacco mosaic virus strains. A comparison of one of their strains, designated Tm 44 by them, with what they considered normal tobacco mosaic virus may serve as an example of how the protein particles can vary from strain to strain. Tm 44 was found to have a higher specific turbidity, twice the specific viscosity per unit concentration, a sedimentation constant greater by roughly 14 per cent, and an electrophoretic mobility at 0° in phosphate buffer at pH 6.9, $s = 0.1$, about 12 per cent less than that of the protein isolated from the normal strain of the virus. In general, these data may be interpreted to mean that the particles of Tm 44 are longer and heavier than those of the normal virus. Knight and Lauffer [35] found that the rib-grass strain of tobacco

mosaic virus migrates more slowly in the electrophoresis apparatus than the normal strain. Melchers [66] and co-workers studied a strain of tobacco mosaic virus isolated by them from tomato plants and called "tomato mosaic virus Dahlem 1940." This material was reported to have a slightly different electrophoretic mobility from tobacco mosaic virus but the same sedimentation constant. However, since they studied the sedimentation of the tomato virus at a concentration of 0.2 per cent and that of the tobacco virus at 0.3 per cent, and in view of the fact that the sedimentation constant of tobacco mosaic virus increases upon dilution, this result, if sufficiently accurate, really means that tomato mosaic virus has a slightly lower sedimentation constant than their normal strain of tobacco mosaic virus. Electron micrographs show that the particles of the tomato mosaic strain are actually only about two-thirds as long as those of the tobacco mosaic strain. Even the particles of the strain which Melchers and his associates consider to be normal tobacco mosaic virus are shown by their electron micrographs to be only 180-190 mμ in length—about two-thirds the value obtained in the other laboratories. If the possibility that this difference is due to uncertainty in the computation of enlargement factors is discarded, the possibility remains that the normal tobacco mosaic virus used by Melchers is really a strain different from the one or ones studied in the other two laboratories. The possibility that there are strains of tobacco mosaic virus which are differentiated by the length of the particles but not by the commonly applied biological tests is not an unreasonable one.

The recognition that strains of tobacco mosaic virus differ in the chemical composition of the nucleoproteins has led to attempts to produce new strains in the laboratory by chemical modification of the normal strains of the virus. Miller and Stanley prepared acetyl, phenylureido, carbobenzoxy, *p*-chlorobenzoyl, and benzenesulfonyl derivatives of the normal strain of the virus.[68] When the extent of substitution was not carried too far, these chemically modified viruses retained full infectivity. Anson and Stanley [2] oxidized the S-H groups of tobacco mosaic virus nucleoprotein with iodine beyond the S-S stage without destroying infectivity. However, all these chemically modified viruses reproduced in the host plant to form normal tobacco mosaic virus. Nevertheless, Miller and Stanley [68] found that biological properties of the derivatives were altered.

Electrophoresis experiments on tobacco mosaic virus were first carried out by Eriksson-Quensel and Svedberg [22] using the Tiselius electrophoresis method. They found the material to have an isoelectric point at pH 3.49 and a pH-mobility curve with a slope of 1.23 cm per sec/volt per cm/pH unit in acetate buffers with an ionic strength of 0.02. Although the chemically prepared virus examined was inhomogeneous with respect to the criterion of sedimentation rate, it was found to be homogeneous with respect to that of electrophoretic mobility. This introduces the general question of the homogeneity of tobacco mosaic virus preparations. Further evidence on this question may be gained from a consideration of x-ray diffraction data, centrifugation data, solubility data, and electron micrographs. Because of the degree of perfection of the intermolecular x-ray pattern of dried gels of tobacco mosaic virus,[11] it seems likely that the virus particles are of absolutely uniform thickness. The earlier sedimentation studies made in Svedberg's laboratory [22] on chemically isolated material indicated a considerable degree of inhomogeneity, but from later studies made on ultracentrifugally isolated virus Wyckoff [107, 108, 111] concluded that the virus particles were homogeneous. A more recent consideration of the sedimentation behavior carried out in the authors' laboratory [44] indicates that even ultracentrifugally isolated preparations are not strictly homogeneous with respect to this criterion. Since x-ray data show that the particles are of uniform thickness, the inhomogeneity must be the result of variability in the lengths of the particles. Indeed, electron micrographs show this to be the case.[100] Fig. 13 is a frequency

chart showing the distribution of particle lengths measured from an electron micrograph of tobacco mosaic virus. It may be seen that most of the particles are about 280 mμ long, but that there are some longer and some shorter ones.

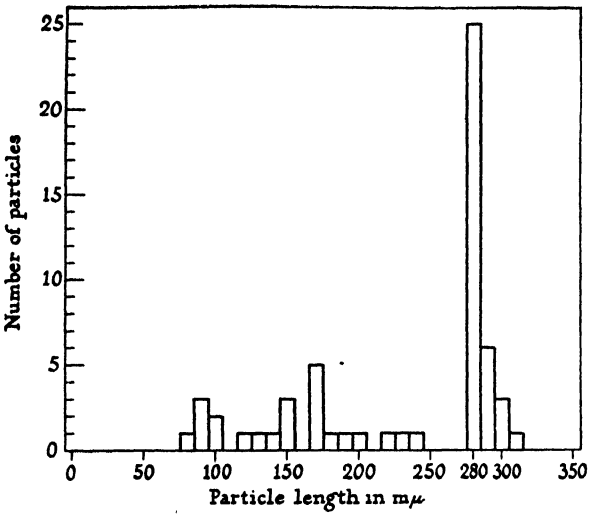


FIGURE 13. Distribution of lengths of particles in an ultracentrifugally prepared sample of tobacco mosaic virus. (From Stanley and Anderson.[100])

Another approach to this question of homogeneity is afforded by solubility studies. The early work of Stanley [91] showed that tobacco mosaic virus was insoluble in strong salt solutions and in solutions having acidities near the isoelectric point. Best [14] has shown that many electrolytes will precipitate the virus and that the relative precipitating powers of various electrolytes depend upon the valence and the character of the ions. More recently, Cohen found that the virus can be precipitated in paracrystalline form by small amounts of heparin and other hydrophilic colloids.[17] Loring [58] attempted a phase rule study of tobacco mosaic virus in ammonium sulfate solutions of constant composition. It is well known that, if a material consists of a single pure substance, it should have a constant solubility, independent of the amount of solid phase in contact with the solvent. Loring's data are shown in Fig. 14. It is plain that the solubility behavior of tobacco mosaic virus protein is not what one would expect of a homogeneous material.

The question of the homogeneity of tobacco mosaic virus is complicated by the tendency of the virus particles to aggregate upon ageing.[61] Wyckoff [107] first observed that, if a freshly prepared batch of tobacco mosaic virus, with a sedimentation constant of about 174×10^{-13}, is allowed to stand in contact with electrolytes for a few days, a second component with a sedimentation constant of about 200×10^{-13} is formed. The authors showed that this behavior could be interpreted quantitatively to mean that virus rods were aggregating end to end to form dimers.[61, 62] Electron micrographs have confirmed the correctness of this interpretation and have shown that this aggregation can be carried much further than the dimeric stage.[100]

The colloid chemical properties of tobacco mosaic virus nucleoprotein are changed completely when the protein is denatured. Denaturation can be brought about by the action of heat,[49] high pressure,[48] and chemical agents such as acids, alkalies, urea,[6, 47, 102] and sodium dodecyl sulfate.[90] Denaturation by heat, high pressure, and urea has been studied in most detail. In all these reactions, the macromolecular nucleoprotein is broken down by first-order processes into much smaller protein frag-

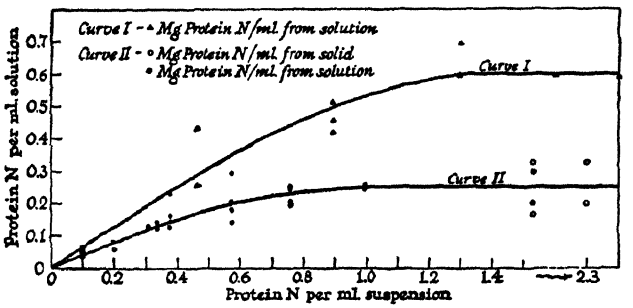


FIGURE 14. Solubility of tobacco mosaic virus in 0.1 *M* potassium phosphate and 11.18% ammonium sulfate buffer. Curve I at pH 5.6; curve II at pH 6.4. (From Loring.[58])

ments, with the loss of infectivity and with the separation of nucleic acid from the protein. The separation of nucleic acid by denaturation of tobacco mosaic virus was observed by Bawden and Pirie [8] and was recently studied by Cohen and Stanley.[19] The latter found by a combination of sedimentation, diffusion, and viscosity studies that the freshly isolated nucleic acid has an average particle weight of around 300,000 and is highly asymmetric. This decomposes spontaneously to form asymmetric particles of about one-fifth the weight of the original. These can in turn be converted into still smaller fragments by the action of cold alkali. The rate of the thermal denaturation was found to vary with the pH and, contrary to all simple theories, with the reciprocal of the initial virus concentration.[49, 78] The rate of denaturation in urea was also found to vary with respect to those two variables, in addition to electrolyte concentration and urea concentration.[47] Miller found that phenylureido, carbobenzoxy, *p*-chlorobenzoyl, and benzenesulfonyl derivatives of the virus denatured more slowly in urea than did normal and acetylated virus.[67] The denaturation in 6*M* urea was found to proceed faster in the cold than at room temperature and also faster at 45° C than at room temperature.[6, 47, 102] An explanation for this very unusual behavior has been advanced, and evidence has been obtained which tends to substantiate the theory.[47]

Tobacco mosaic virus, in common with the elementary bodies of vaccinia, with tomato bushy stunt virus, and with other viruses, may be inactivated by irradiation with ultraviolet light, x-ray, alpha rays, and gamma rays.[27, 55, 64] Surviving infectivity appears to be an exponential function of the amount of radiation. This result is generally interpreted to mean that a single hit is sufficient to cause the destruction of the infectivity of a virus particle. According to current concepts, an atom which is hit by a unit of suitable radiation is ionized. Thereupon, the molecule containing that atom becomes extremely reactive, and if that molecule is a virus particle the loss of infectivity may result. It has been observed that the amount of inactivation due to the irradiation of tobacco mosaic virus is greater the higher the net positive charge of the virus.[64] This result has been interpreted to mean that viruses may also become inactivated by secondary reactions following the ionization of solvent molecules.[64] From the amount of radiation, it is possible to calculate the density of ions or ion clusters produced. Then from the fraction of virus particles inactivated at a given ion density, it is possible to compute the size of the target a virus particle affords. Such a target size really represents the portion of a virus particle labile to radiation. The target sizes for most plant viruses seem to be somewhat though not greatly less than the known particle sizes.[27, 55]

Spherical Viruses

Although the colloidal properties of tobacco mosaic virus illustrate what one can expect, in general, for most of the rod-like viruses, they do not apply to all viruses.

The colloid-chemical behavior of the spherical viruses is much simpler and much more easily understood. Indeed, the answer to such important questions as whether or not virus particles are molecules in the strictest sense of the term will probably have to be sought from studies carried out with the spherical viruses. As indicated previously, many virus nucleoproteins with essentially spherical particles have been isolated, but of this whole group only the elementary bodies of vaccinia and the tomato bushy stunt virus have been subjected to detailed colloid-chemical study. Several very interesting colloid-chemical experiments have been carried out with the former.[25, 76, 84, 87, 88] However, this virus is somewhat more complex than most viruses and, hence, the tomato bushy stunt virus is probably more typical of the spherical viruses as a class. Tomato bushy stunt virus was obtained by Bawden and Pirie [5] and later by Stanley [99] in the form of a nucleoprotein which crystallizes from ammonium sulfate solutions in beautiful dodecahedra. Prisms and rhombohedra may be obtained when the crystallization is induced by certain hydrophilic colloids.[17, 18] Bushy stunt virus has a nucleic acid content of about 17 per cent, a value about three times that of tobacco mosaic virus. This is perhaps the most conspicuous chemical difference that has as yet been established between the two viruses.

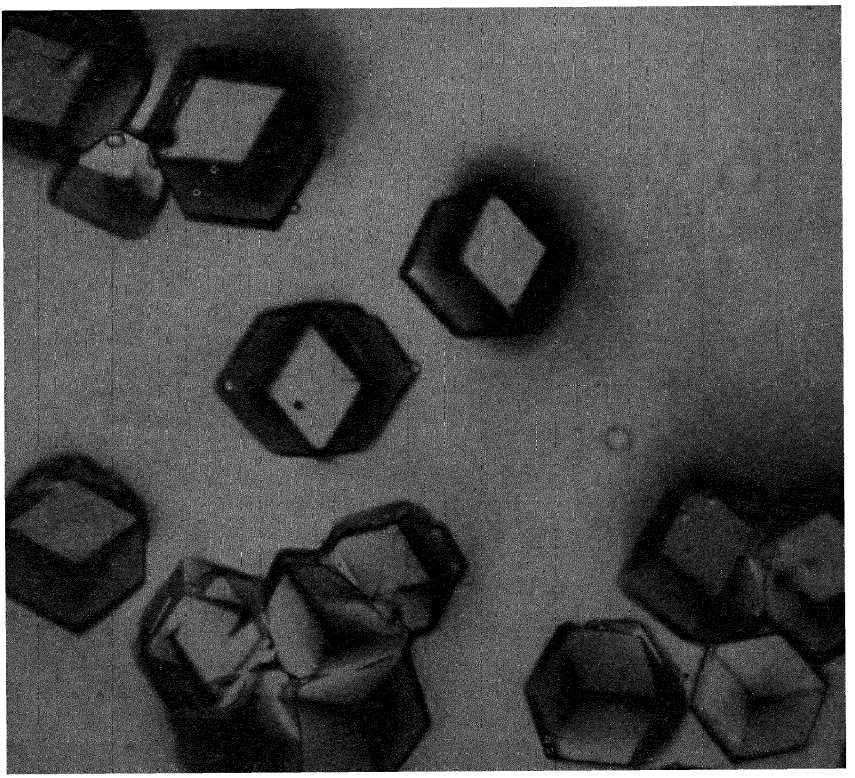


FIGURE 15. Crystals of bushy stunt virus protein. (From Stanley.[99])

Bushy stunt virus preparations do not show double refraction of flow; pellets of the virus obtained upon high-speed centrifugation are not liquid crystalline, and the virus itself crystallizes in the cubic system. All of these facts point to the conclusion that the particles of bushy stunt virus are spherical or nearly so. More exact information concerning the size and shape of the virus particles has been derived from

sedimentation equilibrium,[62] sedimentation velocity,[44, 53, 62] diffusion,[70] x-ray diffraction,[11, 12] and electron microscope studies.[101]

Sedimentation equilibrium studies were conducted by McFarlane and Kekwick.[62] An analysis of their published data shows that there is only a slight drift in the values calculated for the molecular weight at various distances through the equilibrium column. This is an indication that the particles of bushy stunt virus are reasonably homogeneous, although it is not a particularly sensitive criterion of that attribute. Taking the specific volume to be 0.739, they calculated the mean molecular weight from three equilibrium runs to be 7.6×10^6.

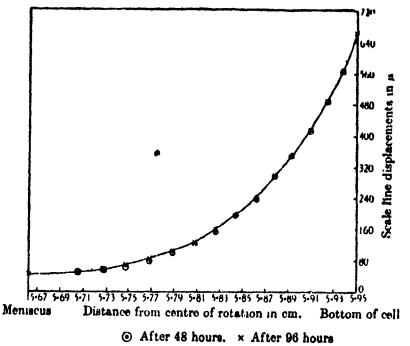


FIGURE 16. Sedimentation equilibrium experiment on bushy stunt virus. (From McFarlane and Kekwick.[62])

McFarlane and Kekwick [62] also made a series of sedimentation velocity experiments on bushy stunt virus dissolved in various buffers covering the pH range from 2.4 to 8.7. The sedimenting boundaries were found to be quite sharp, again indicating a reasonable degree of particle homogeneity. An average of eleven determinations on one preparation of virus gave a value of 146×10^{-13} for the sedimentation constant. On the assumption that the virus particles are unhydrated spheres, one can calculate a molecular weight of 8.8×10^6 from this value of the sedimentation constant. This calculation may be carried out by using equation (1), setting f_o/f equal to 1. The sedimentation velocity of this virus was also studied in the authors' laboratory.[44, 53] It was shown that the sedimentation rate, unlike that of tobacco mosaic virus, was essentially independent of virus concentration. An average value of 132×10^{-13}, with a standard deviation of about ± 3 per cent, was calculated for the sedimentation constant from almost fifty measurements made on several preparations of the virus. This value for the specific material being studied was confirmed in four other laboratories, including McFarlane's, and is, therefore, almost certainly correct within 1 or 2 per cent.* By assuming that the virus particles are unhydrated spheres, and by using equation (1) as previously described, one can calculate the molecular weight from a sedimentation constant of 132×10^{-13} to be 7.4×10^6. Both this value and that of 8.8×10^6, obtained by McFarlane and Kekwick, agree reasonably well with the molecular weight calculated from sedimentation equilibrium, 7.6×10^6.

The molecular weight of bushy stunt virus has also been evaluated from a combination of sedimentation with diffusion data, using equation (4). Diffusion studies were carried out by Neurath and Cooper,[70] using the Lamm Scale procedure. A value of 1.15×10^{-7} was obtained for the diffusion constant, corrected to water at 20° C. If this value for the diffusion constant is used in conjunction with the authors' sedimentation data, a value of 10.6×10^6 may be calculated for the molecular weight.

* The lower value for the sedimentation rate of bushy stunt virus has recently been confirmed by Ogston [*Biochem. J.*, **37**, 78 (1943)], who stated that McFarlane and Kekwick regard the higher value they obtained in their initial study as being in error due probably to unsuspected irregularities in temperature in the rotor.

The discrepancy between this value and the 7.4×10^6 obtained from sedimentation data alone could possibly mean that the assumption that the particles are unhydrated spheres is not valid. Had it been assumed either that the virus particles are hydrated spheres containing about 67 per cent water on a dry basis or that they are unhydrated rods or disks with asymmetries of 5:1, the value calculated from sedimentation data alone would have been in agreement with this value of 10.6×10^6. It is not possible, however, to rationalize this value with that obtained from sedimentation equilibrium without questioning the accuracy of one or the other.

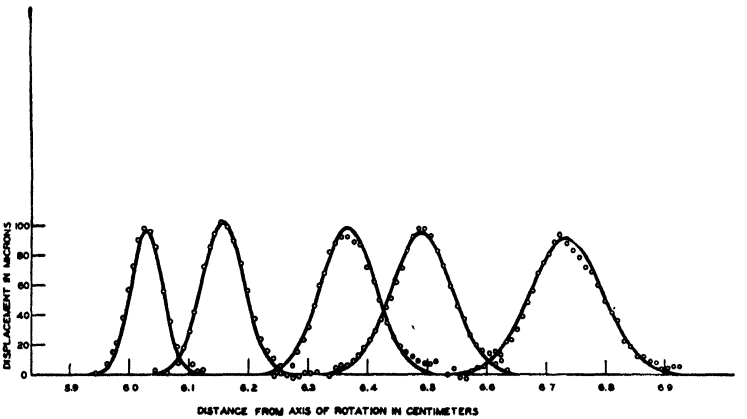


FIGURE 17. Theoretical vs. actual boundary spreading in sedimentation experiment on bushy stunt virus. Open circles are experimental points obtained by Lamm Scale method. Smooth curves are theoretical boundaries calculated from known diffusion constant. The times in minutes and scale cell distances in centimeters for the successive boundary curves, beginning at the left, are: 50, 2.1; 85, 3.1; 135, 4.1; 150, 4.6; 220, 5.6. (From Lauffer.[45])

The fact that the distribution of bushy stunt virus particles in sedimentation equilibrium studies is approximately of the theoretical form indicates that the virus solutions are reasonably homogeneous. A much more sensitive criterion of homogeneity may be afforded by sedimentation velocity experiments. In every such experiment, the sedimenting boundary becomes progressively more diffuse. This may be caused entirely by diffusion, or by diffusion and inhomogeneity. The spreading of a sedimenting bushy stunt virus boundary was measured very carefully by the Lamm Scale method.[45] Then the theoretical boundary spreading was calculated from the diffusion constant, on the assumption that the virus particles are all of the same size and shape. The results are illustrated in Fig. 17, where the open circles are the experimental data and the smooth curves the theoretical boundaries. The diffusion constant was taken to be 1.15×10^{-7}. The excellent agreement demonstrates that the observed boundary spreading can be accounted for completely by the known diffusion rate of the virus. This fact is entirely consistent with the supposition that the virus is absolutely homogeneous and that the virus particles are therefore molecules. The data demonstrate conclusively that, if the particles are not homogeneous, the standard deviation of the distribution of particle diameters must be less than 1 per cent of the mean diameter.

Electrophoresis experiments conducted by McFarlane and Kekwick[62] indicate that the virus is homogeneous with respect to electrochemical behavior. This evidence is also consistent with the idea that the virus particles are molecular in nature. The isoelectric point of the material was found to be at pH 4.2 and the

electrophoretic mobility was found to change 0.4 μ per second per pH unit in unit field.

The x-ray diffraction pattern of crystals of bushy stunt virus suspended in their mother liquid was studied by the powder method by Bernal, Fankuchen, and Riley.[11, 12] They observed two lines which correspond to spacings of 279 Å and 160 Å, respectively. The relationship between them is expressed by the ratio 3:1. Because the crystals are dodecahedra, the investigators assumed that the crystal lattice is body-centered, and they ascribed these two spacings to 110 and 112 planes. From this it can be computed that the edge of a unit cell would be 394 Å. The density of the wet crystals was taken to be 1.286. However, such a high value is not compatible with the known density of the virus particles and the amount of hydration indicated for the wet crystals. A value for the density, calculated from theory, of 1.18 would seem to be a much more reasonable figure. Since there are two molecules in a body-centered unit cell, one can calculate a value of 22×10^6 for the wet molecular weight of bushy stunt virus. Had the investigators assumed that the observed spacings corresponded to 100 and 111 planes of a simple cubic lattice, they would have calculated a unit cell of 279 Å on edge containing only one molecule. This would give a wet molecular weight of 15.6×10^6. These workers also observed that, as crystals of bushy stunt virus are dried, they shrink symmetrically to final dimensions about 80 per cent of those in the wet state. This shrinkage is paralleled by a corresponding decrease of the same magnitude in the crystal lattice. From this fact it can be calculated that a wet crystal of the virus contains 0.67 gram of water for each gram of dry protein. Hence, the dry molecular weight of the virus would be 9.35×10^6 or 13.2×10^6, depending upon whether the crystal lattice is assumed to be simple cubic or face-centered. The lower of these is somewhat too high to agree with the value of the molecular weight calculated from centrifugation studies alone, 7.5×10^6, and the higher is too high to agree with that calculated from a combination of sedimentation and diffusion data. This discrepancy can be met by assuming that a perfectly dry virus crystal can soak up about 15 per cent of its own volume of water before the crystal begins to swell, that is, that it can hold 15 per cent by volume of water zeolytically. On this assumption, the dry molecular weight calculated for a simple cubic cell would be 7.9×10^6, and that calculated for a face-centered lattice would be 11.2×10^6.

Electron micrographs of bushy stunt virus have been obtained.[101] The particles appear to be essentially spherical in shape, with a diameter of about 26 mμ. This value is in excellent agreement with that calculated from sedimentation velocity data on the assumption that the particles are unhydrated spheres. The filtration end point of this virus was found to be 40 mμ, indicating that the particles are 13-20 mμ in diameter.[89] Hence, these data also agree in general with the other studies on the size of the bushy stunt particle.

Diffusion, sedimentation velocity, sedimentation equilibrium, and electrophoresis experiments all indicate that bushy stunt virus nucleoprotein preparations possess a degree of physical homogeneity heretofore never demonstrated for any virus protein. The physical evidence for absolute homogeneity, which is demanded of the particles if they are to be regarded as molecules in the strictest sense, is as good as that for any of the proteins generally conceded to be in a molecular state. There is still an unresolved discrepancy in the data relative to the molecular weight of the virus, but the evidence indicates that the particles are unhydrated spheres with a molecular weight of about 7½ million, or either hydrated spheres or non-spherical bodies with a molecular weight of about 10½ million.

The colloid-chemical properties, particularly the size and shape, of two viruses have been considered in detail. Although few other viruses have been studied to the same extent as these two, considerable evidence is available defining at least the order of magnitude of the sizes of many other viruses. This evidence is sum-

COMPARATIVE SIZES OF VIRUSES

	Molecular weight X 10^{-6} (Particle weight X 6 06 X 10^{17})	Diam or length X width in $m\mu$
Red blood cells*	173 000 000	7 500
Bacillus prodigiosus*	173 000	750
Rickettsia*	11 100	300
Psittacosis*	8 500	275
Vaccinia*	4 300	225
Myxoma*	4 300	
Canary pox*	4 300	
Pleuro-pneumonia organism*	1 400	150
Pseudo rabies	1 400	
Ectromelia	1 400	
Herpes simplex	1 400	
Rabies fixe'	800	125
Borna disease	800	
Vesicular stomatitis	410	100
Staphylococcus bacteriophage†	300	90
Fowl plague	300	
C_{16} bacteriophage	173	75
Chicken tumor I*	142	70
Tobacco mosaic*	43	280 X 15
Cucumber mosaics 3 and 4*	43	
Gene (Muller's est of max size)*	33	125 X 20
Latent mosaic of potato*	26	430 X 9 8
Rabbit papilloma (Shope)*	25	40
Equine encephalitis	23	38
Megatherium bacteriophage	23	
Rift valley fever	11	30
Tomato bushy stunt*	8	26
Hemocyanin molecule (Busycon)*	6 7	22
Yellow fever	4 3	22
Tobacco ring spot*	3 4	19
Louping ill	2 8	19
Hemocyanin molecule (Octopus)*	2 8	20
Alfalfa mosaic*	2 1	16 5
Poliomyelitis	0 7	12
Influenza	0 6	11
Staphylococcus bacteriophage†	0 4	10
Foot-and-mouth disease	0 4	
Hemoglobin molecule (Horse)*	0 069	15 X 3
Egg albumin molecule*	0.040	9 X 3

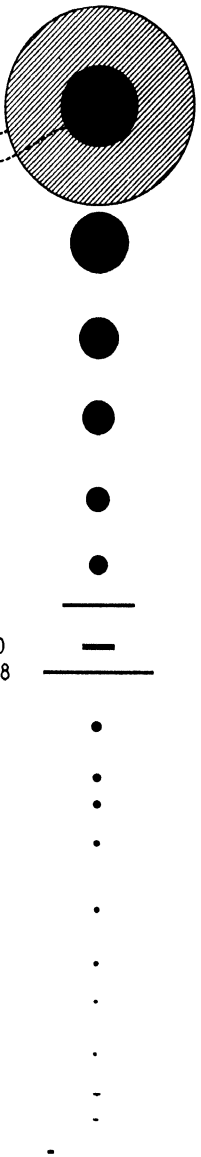


FIGURE 18. Comparative sizes of viruses and other biological materials. The figures for size have been arbitrarily selected from data available in the literature. Particles known to be asymmetric are so indicated and the estimated length and width and the known molecular weight in accordance with the asymmetry are given. In other cases where the particles are known or assumed to be spherical, the diameter and the molecular weight based on a sphere of density 1.3 are given. * = evidence regarding shape available. † = large size from filtration and sedimentation of concentrated solutions and small size from diffusion of dilute solutions. (From Stanley.[98])

marized in Fig. 18. It has been assembled and presented in somewhat greater detail by Markham, Smith and Lea.[68]

References

1. Allard, H. A., *J. Agr. Res.*, **6**, 649 (1916).
2. Anson, M. L., and Stanley, W. M., *J. Gen. Physiol.*, **24**, 679 (1941).
3. Bawden, F. C., and Pirie, N. W., *Proc. Roy. Soc. London*, **B 123**, 274 (1937).
4. Bawden, F. C., and Pirie, N. W., *Brit. J. Exp. Path.*, **18**, 275 (1937).
5. Bawden, F. C., and Pirie, N. W., *Nature*, **141**, 513 (1938); *Brit. J. Exp. Path.*, **19**, 251 (1938).
6. Bawden, F. C., and Pirie, N. W., *Biochem. J.*, **34**, 1258 (1940).
7. Bawden, F. C., Pirie, N. W., Bernal, J. D., and Fankuchen, I., *Nature*, **138**, 1051 (1936).
8. Beard, J. W., and Wyckoff, R. W. G., *Science*, **85**, 201 (1937).
9. Bechhold, H., and Schlesinger, M., *Phytopath. Z.*, **6**, 627 (1933).
10. Beijerinck, M. W., *Verhandel. Konink. Akad. Wetenschappen, Amsterdam*, **6**, n.s., **1** (1898).
11. Bernal, J. D., and Fankuchen, I., *Nature*, **139**, 923 (1937); *J. Gen. Physiol.*, **25**, 111, 147 (1941).
12. Bernal, J. D., Fankuchen, I., and Riley, D. P., *Nature*, **142**, 1075 (1938).
13. Best, R. J., *Nature*, **139**, 628; **140**, 547 (1937); *J. Austr. Inst. Agr. Sci.*, **5**, 94 (1939).
14. Best, R. J., *Austr. J. Exp. Biol. and Med. Sci.*, **18**, 307 (1940).
15. Boeder, P., *Z. Physik*, **75**, 258 (1932); *J. Rheol.*, **3**, 494 (1932).
16. Chambers, L. A., Henle, W., Lauffer, M. A., and Anderson, T. F., *J. Exp. Med.*, **77**, 265 (1943).
17. Cohen, S. S., *J. Biol. Chem.*, **144**, 353 (1942).
18. Cohen, S. S., *Proc. Soc. Exp. Biol. and Med.*, **51**, 104 (1942).
19. Cohen, S. S., and Stanley, W. M., *J. Biol. Chem.*, **144**, 589 (1942).
20. Craigie, J., *Brit. J. Exp. Path.*, **13**, 259 (1932).
21. Duggar, B. M., and Karrer [Armstrong], Joanne L., *Ann. Missouri Bot. Gard.*, **8**, 343 (1921); **10**, 191 (1923).
22. Eriksson-Quensel, I.-B., and Svedberg, T., *J. Am. Chem. Soc.*, **58**, 1863 (1936).
23. Frampton, V. L., *Phytopath.*, **29**, 495 (1939); *J. Biol. Chem.*, **129**, 233 (1939); *Science*, **90**, 305 (1939).
24. Frampton, V. L., and Neurath, H., *Science*, **87**, 468 (1938).
25. Frampton, V. L., and Saum, A. M., *Science*, **89**, 84 (1939).
26. Gard, S., and Pedersen, K. O., *Science*, **94**, 494 (1941).
27. Gowen, J. W., *Proc. Nat. Acad. Sci.*, **26**, 8 (1940).
28. Green, R. H., Anderson, T. F., and Smadel, J. E., *J. Exp. Med.*, **75**, 651 (1942).
29. Herzog, R. O., Illig, R., and Kudar, H., *Z. physik. Chem.*, **A 167**, 329 (1934).
30. Iwanowski, Dm., *Bull. Akad. Imp. Sci. St. Petersburg*, n.s., 3, **35**, 67 (1892).
31. Janssen, L. W., *Z. Hyg.*, **119**, 558 (1937); *Naturwissensch.*, **29**, 102 (1941).
32. Kausche, G. A., Guggisberg, H., and Wissler, A., *Naturwissensch.*, **27**, 303 (1939).
33. Kausche, G. A., Pfankuch, E., and Ruska, H., *Naturwissensch.*, **27**, 292 (1939).
34. Knight, C. A., *J. Biol. Chem.*, **145**, 11 (1942).
35. Knight, C. A., and Lauffer, M. A., *J. Biol. Chem.*, **144**, 411 (1942).
36. Knight, C. A., and Stanley, W. M., *J. Biol. Chem.*, **141**, 29, 39 (1941).
37. Kuhn, W., *Kolloid-Z.*, **62**, 269 (1933).
38. Langmuir, I., *J. Chem. Phys.*, **6**, 873 (1938).
39. Langmuir, I., and Schaefer, V. J., *J. Am. Chem. Soc.*, **59**, 1406 (1937).
40. Lauffer, M. A., Paper presented at 106th Meeting, American Chemical Society, Pittsburgh, Pa., Sept. (1943).
41. Lauffer, M. A., *Science*, **87**, 469 (1938); *J. Biol. Chem.*, **126**, 443 (1938).
42. Lauffer, M. A., *J. Phys. Chem.*, **42**, 935 (1938).
43. Lauffer, M. A., *J. Am. Chem. Soc.*, **61**, 2412 (1939).
44. Lauffer, M. A., *J. Phys. Chem.*, **44**, 1137 (1940).
45. Lauffer, M. A., *J. Biol. Chem.*, **143**, 99 (1942).
46. Lauffer, M. A., *Chem. Rev.*, **31**, 561 (1942).
47. Lauffer, M., *J. Am. Chem. Soc.*, **65**, 1793 (1943).
48. Lauffer, M. A., and Dow, R. B., *J. Biol. Chem.*, **140**, 509 (1941).
49. Lauffer, M. A., and Price, W. C., *J. Biol. Chem.*, **133**, 1 (1940).
50. Lauffer, M. A., and Ross, A. F., *J. Am. Chem. Soc.*, **62**, 3296 (1940).
51. Lauffer, M. A., and Stanley, W. M., *J. Biol. Chem.*, **123**, 507 (1938).
52. Lauffer, M. A., and Stanley, W. M., *Chem. Rev.*, **24**, 303 (1939); *Kolloid-Z.*, **91**, 62 (1940).
53. Lauffer, M. A., and Stanley, W. M., *J. Biol. Chem.*, **135**, 463 (1940).
54. Lavin, G. I., and Stanley, W. M., *J. Biol. Chem.*, **118**, 269 (1937).
55. Lea, D. E., and Smith, K. M., *Parasitology*, **32**, 405 (1940); **34**, 227 (1942).
56. Levine, S., *Proc. Roy. Soc.*, **A 170**, 145, 165 (1939).
57. Loring, H. S., *J. Biol. Chem.*, **123** [Sci. Proc. XXXII], lxxvi (1938); **130**, 251 (1939).
58. Loring, H. S., *J. Gen. Physiol.*, **23**, 719 (1940).
59. Loring, H. S., and Stanley, W. M., *J. Biol. Chem.*, **117**, 733 (1937).
60. Loring, H. S., and Wyckoff, R. W. G., *J. Biol. Chem.*, **121**, 255 (1937).
61. Loring, H. S., Lauffer, M. A., and Stanley, W. M., *Nature*, **142**, 841 (1938).
62. McFarlane, A. S., and Kekwick, R. A., *Biochem. J.*, **32**, 1607 (1938).
63. Markham, R., Smith, K. M., and Lea, D., *Parasitology*, **34**, 315 (1942).
64. Marshak, A., and Takahashi, W. N., *Proc. Nat. Acad. Sci.*, **28**, 211 (1942).
65. Mehl, J. W., *Cold Spring Harbor Symposia Quant. Biol.*, **6**, 218 (1938).
66. Melchers, G., Schramm, G., Trurnit, H., and Friedrich-Freksa, H., *Biol. Zentralbl.*, **60**, 524 (1940).
67. Miller, G. L., *J. Biol. Chem.* **146**, 339 (1942).
68. Miller, G. L., and Stanley, W. M., *J. Biol. Chem.*, **141**, 905 (1941); **146**, 331 (1942).
69. Mueller, H., *Phys. Rev.*, **55**, 508, 792 (1939); Mueller, H., and Sakmann, B. W., *Phys. Rev.*, **56**, 615 (1939); *J. Opt. Soc. Am.*, **32**, 309 (1942).
70. Neurath, H., and Cooper, G. R., *J. Biol. Chem.*, **135**, 455 (1940).
71. Neurath, H., and Saum, A. M., *J. Biol. Chem.*, **126**, 435 (1938).
72. Northrop, J. H., *Science*, **84**, 90 (1936); *Collecting Net*, **12**, 188 (1937); *J. Gen. Physiol.*, **21**, 335 (1938).
73. Norton, F. J., *Phys. Rev.*, **55**, 668 (1939).
74. Pfankuch, E., Kausche, G. A., and Stubbe, H., *Biochem. Z.*, **304**, 238 (1940).
75. Perrin, F., *J. Phys. Radium*, **7**, 1 (1936).
76. Pickels, E. G., and Smadel, J. E., *J. Exp. Med.*, **68**, 583 (1938).
77. Polson, A., *Kolloid-Z.*, **88**, 51 (1939).

78. Price, W. C., *Arch. ges. Virusforsch.*, **1**, 373 (1940).
79. Price, W. C., and Wyckoff, R. W. G., *Phytopath.*, **29**, 83 (1939).
80. Robinson, J. R., *Proc. Roy. Soc.*, **A 170**, 519 (1939).
81. Ross, A. F., *Phytopath.*, **31**, 394, 410 (1941).
82. Ross, A. F., and Stanley, W. M., *Proc. Soc. Exp. Biol. Med.*, **38**, 260 (1938); *J. Gen. Physiol.*, **22**, 165 (1938).
83. Ruska, H., Borries, B., and Ruska, E., *Arch. ges. Virusforsch.*, **1**, 155 (1939).
84. Shedlovsky, T., and Smadel, J. E., *J. Exp. Med.*, **72**, 511 (1940).
85. Simha, R., *J. Phys. Chem.*, **44**, 25 (1940).
86. Simon, C. E., in Jerome Alexander's "Colloid Chemistry" II, 525, New York (1928).
87. Smadel, J. E., Pickels, E. G., and Shedlovsky, T., *J. Exp. Med.*, **68**, 607 (1938).
88. Smadel, J. E., Pickels, E. G., Shedlovsky, T., and Rivers, T. M., *J. Exp. Med.*, **72**, 523 (1940).
89. Smith, K. M., and MacClement, W. D., *Proc. Roy. Soc.*, **B 125**, 295 (1938); *Parasitology*, **33**, 320 (1941).
90. Sreenivasaya, M., and Pirie, N. W., *Biochem. J.*, **32**, 1707 (1938).
91. Stanley, W. M., *Science*, **81**, 644 (1935); *Phytopath.*, **26**, 305 (1936).
92. Stanley, W. M., *Phytopath.*, **25**, 899 (1935).
93. Stanley, W. M., *J. Biol. Chem.*, **117**, 325 (1937).
94. Stanley, W. M., *J. Biol. Chem.*, **117**, 755 (1937).
95. Stanley, W. M., *Ergeb. Physiol.*, **39**, 294 (1937); *Handbuch d. Virusforschung*, p. 477, Vienna (1938).
96. Stanley, W. M., *J. Phys. Chem.*, **42**, 55 (1938).
97. Stanley, W. M., *Phys. Rev.*, **19**, 524 (1939).
98. Stanley, W. M., in "Virus Diseases," p. 35, The Messenger Lectures, Cornell Univ. Press, Ithaca, N. Y., 1943.
99. Stanley, W. M., *J. Biol. Chem.*, **135**, 437 (1940).
100. Stanley, W. M., and Anderson, T. F., *J. Biol. Chem.*, **139**, 325 (1941).
101. Stanley, W. M., and Anderson, T. F., *J. Biol. Chem.*, **146**, 25 (1942).
102. Stanley, W. M., and Lauffer, M. A., *Science*, **89**, 345 (1939).
103. Stanley, W. M., and Loring, H. S., *Cold Spring Harbor Symposia Quant. Biol.*, **6**, 341 (1938).
104. Stanley, W. M., and Wyckoff, R. W. G., *Science*, **85**, 181 (1937).
105. Takahashi, W. N., and Rawlins, T. A., *Proc. Soc. Exp. Biol. Med.*, **30**, 155 (1932); *Science*, **77**, 26, 284 (1933); **85**, 103 (1937).
106. Thornberry, H. H., *Phytopath.*, **25**, 36, 601, 938 (1935).
107. Wyckoff, R. W. G., *J. Biol. Chem.*, **121**, 219 (1937).
108. Wyckoff, R. W. G., *J. Biol. Chem.*, **122**, 239 (1937).
109. Wyckoff, R. W. G., *Proc. Soc. Exp. Biol. Med.*, **36**, 771 (1937).
110. Wyckoff, R. W. G., *J. Biol. Chem.*, **124**, 585 (1938).
111. Wyckoff, R. W. G., Biscoe, J., and Stanley, W. M., *J. Biol. Chem.*, **117**, 57 (1937).
112. Wyckoff, R. W. G., and Corey, R. B., *J. Biol. Chem.*, **116**, 51 (1936).
113. Zocher, H., and Jacobsohn, K., *Kolloid-Z.*, **41**, 220 (1927); *Kolloid Chem. Beih.*, **28**, 167 (1929).

The Gene—A Structure of Colloidal Dimensions

JEROME ALEXANDER

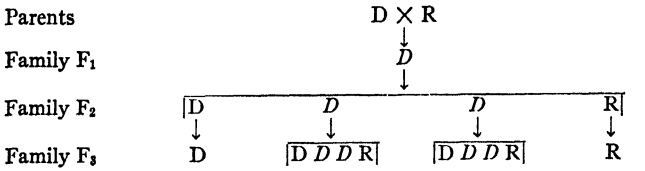
Genetics is a branch of biology dealing with heredity and its variations, and the mechanisms underlying these phenomena. The following historical summary and glossary have been prepared for the benefit of those users of this series who have not kept in touch with the fascinating developments in genetics, leading to demonstration of the fact that genes are specific structures of colloidal dimensions. The paper by J. Alexander and C. B. Bridges in Vol. II of this series, entitled "Some Physicochemical Aspects of Life, Mutation and Evolution," gives considerable genetic data, including (p. 36) a diagrammatic sketch of the chromosome cycle in cell division. The following is largely based on an outline in "Colloid Chemistry," by J. Alexander * from which most of the illustrations included here have been reproduced.

In 1865 Gregor Mendel, Abbott of Brünn (Bohemia), published results of the breeding experiments he had carried out in his cloister garden with the common pea (*Pisum sativum*); but it was not until 1900, eighteen years after his death, that the importance of his work was recognized by three distinguished botanists, Hugo de Vries, C. Correns, and E. Tschermak. Mendel crossed a tall with a dwarf pea, and focussed his attention on how these characteristics were inherited. All members of the first hybrid family (F_1) were tall.

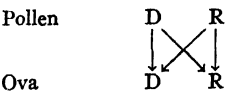
Mendel's First Law: Segregation. The second family (F_2) developing from the seeds of the F_1 hybrids, always gave an average ratio of three talls to one dwarf. The F_2 dwarfs always bred true to type, as did also one-third of the talls. The other two-thirds of the talls behaved like the original F_1 hybrids, *i.e.*, they gave in F_2 three

* 4th edition, D. Van Nostrand Co. N. Y., 1937.

talls to one dwarf. Mendel termed "tallness" a *dominant* character and "dwarfness" a *recessive* character because it reappeared in subsequent generations raised from the hybrid type seed. He also obtained similar results with several other pairs of contrasting character. The following diagram illustrates the results, R indicating plants with a recessive character, D dominants which breed true, and *D* dominants which carry the recessive type:

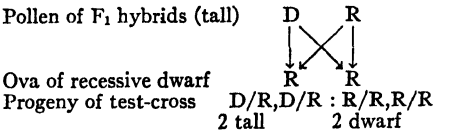


Mendel pointed out that this segregation of characters in a 3 to 1 ratio could be explained by assuming that they were determined by discrete representative producers (later called genes). Though each cell of the hybrids F_1 would develop under the joint control of a dominant and a recessive gene, in both ova and pollen these factors separate cleanly from each other (segregate), so that, according to the laws of chance, each ovum or pollen grain would get a pure dominant gene, D, or equally often a pure recessive gene, R. Thus with random or haphazard fertilizations, we would get



leading to an F_2 family D/D;(D/R,D/R);R/R.

As a test for this hypothesis, Mendel fertilized *dwarf* flowers with pollen from F_1 hybrids, obtaining, as he had expected, *equal* numbers of tall- and dwarf-producing seeds, as the following diagram indicates:



Mendel's Second Law: Independent Assortment. On crossing peas having *yellow and round* seeds with peas having *green and wrinkled seeds,* Mendel found that the F_1 seeds were all yellow and round, these characters being dominant over the recessive alternates green and wrinkled. The plants from these F_1 seeds, when self-fertilized, yielded seeds (and therefore progeny) of four types: yellow-round; yellow-wrinkled; green-round; green-wrinkled, in the ratios 9:3:3:1. Mendel explained these results by assuming that the distribution of the segregated genes of the pair green/yellow takes place independently of the round/wrinkled segregated pair. That is, there would be four equally numerous kinds of ova and four equally numerous kinds of pollen grains, which, uniting at random during fertilization, would give sixteen different combinations in the 9:3:3:1 ratio found in the visible types.

Linkage. This is a failure of random assortment, understandable on the assumption that the genes of the two pairs which fail to assort at random act as discrete units (gene blocks) in the chromosomes, microscopically resolvable bodies containing invisible genes arranged in linear order. Just after the resurrection of Mendel's work, the phenomenon of linkage was observed by Sir William Bateson and R. C. Punnett * in experiments with sweet peas, in which they found that the factor pairs

* Second Rept., Evolution Committee, Royal Soc., London, 1905.

(red vs. purple flower) and (long vs. round pollen grain) did *not* assort at random. When (red) and (long) entered a cross together, they tended to remain together in subsequent generations to a greater extent than expected under random assortment.

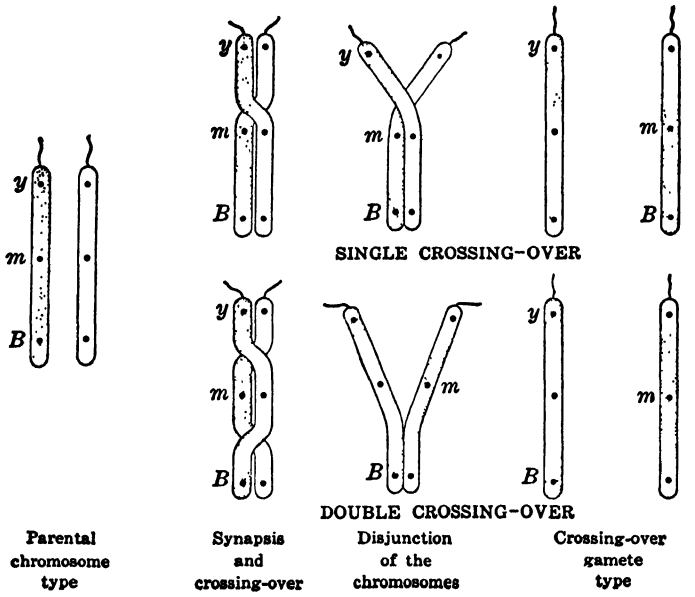


FIGURE 1.

Variations in Chromosomal Behavior. *Crossing-over*, the commonest of these, is illustrated in Fig. 1, and occurs when two chromosomes of a pair twist about each other in some stage of germ cell formation (meiosis). Less common are: *inversion*, which occurs when a block of genes breaks loose from a chromosome and returns to that chromosome "upside down"; *translocation*, which occurs when the loose gene-block joins another chromosome; *deletion* or *deficiency*, which occurs if the loose gene-block is lost. These phenomena are diagrammed in Fig. 2.

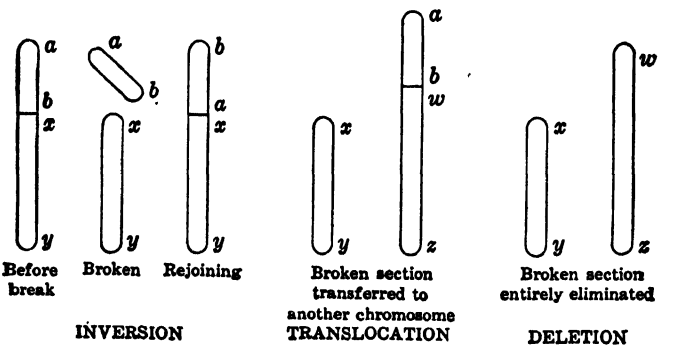


FIGURE 2.

Drosophila melanogaster. This tiny yeast-eating "fruit fly" has been largely used by T. H. Morgan (Nobel Laureate, 1935) and his school, as well as by many others, in speeding up genetic experiments. The fly matures in about 10 days and

lives about 30 days. About thirty or more prolific generations can be cheaply raised in a year, in glass bottles in the laboratory, independent of weather conditions. The breeding and crossing of strains is readily controlled, and there are many easily recognizable variations. A "gene map" prepared by C. B. Bridges is shown in Fig. 3,

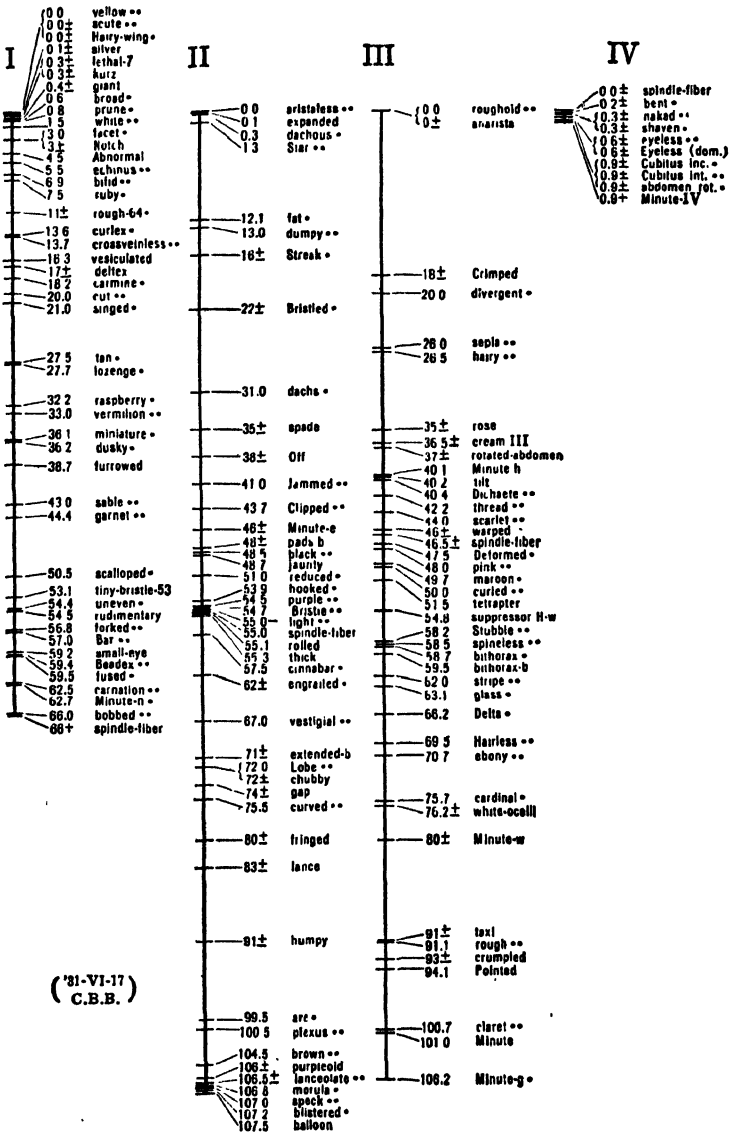


FIGURE 3. Genic map of the four chromosomes of *Drosophila melanogaster*, showing the linear order and distance apart of the genes (after Calvin B. Bridges).

which shows the relative positions of numerous genes dominating the characteristics indicated in the "shorthand" of geneticists. Some of these characters, *e.g.*, vestigial wing, are most striking, and are illustrated in Fig. 4. A large variety of plants and animals are used in genetic experiments, including microorganisms.

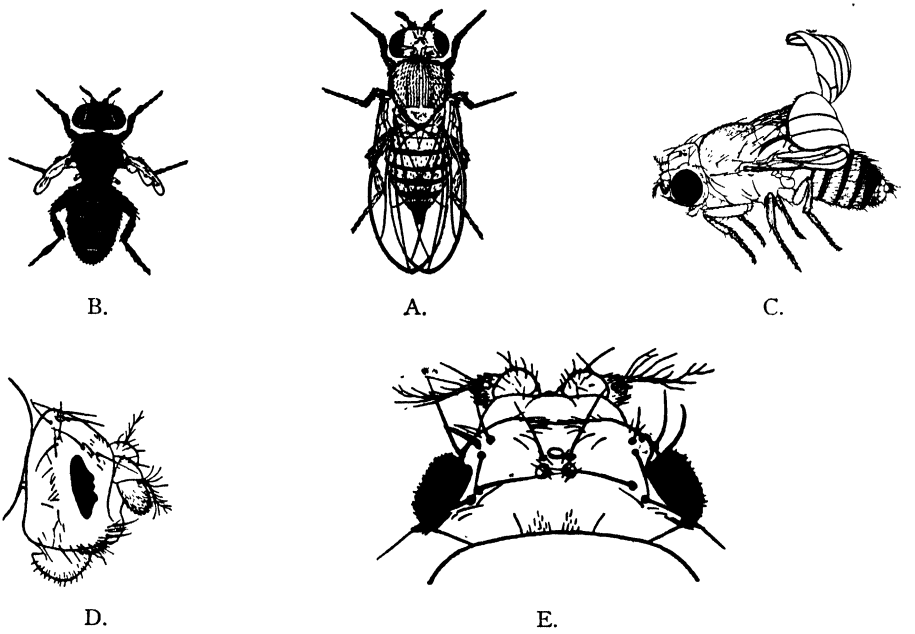


FIGURE 4. A—Normal wild-type female *Drosophila,* gray body color, red eyes, long wings, absence of speck at base of wings.
B—Male *Drosophila* having four linked, recessive characters, viz., black body color, purple eyes, vestigial wings, and speck at the base of the wings.
C—Fly showing the mutant character "Curly." (The wings curl up at the ends and are held somewhat apart.)
D—Eye of female fly homozygous for "Bar" (eye is bar-shaped).
E—Fly showing the mutant character "Lobe ^{2}". (The eyes are small and protruding.)

These illustrations are reproduced from "The Theory of the Gene" by Thomas Hunt Morgan, Nobel Laureate. (Yale Univ. Press, 1926)

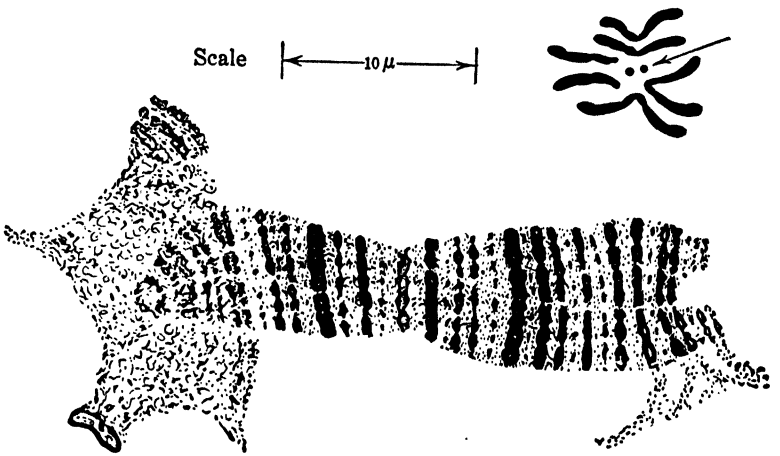


FIGURE 5. Salivary gland chromosome from *Drosophila melanogaster.*

An outstanding advance was made in 1927 by H. J. Muller, who demonstrated the "artificial transmutation of the gene" by x-rays * and this work has been greatly extended. Another great advance was due to the observation (1935) by T. S. Painter that the cells in the salivary gland of the tiny fruit fly grub have chromosomes which are from 175 to 200 times the size of ordinary chromosomes, and which, on staining, reveal a "genetic spectrum," a series of characteristic bands varying with and corre-

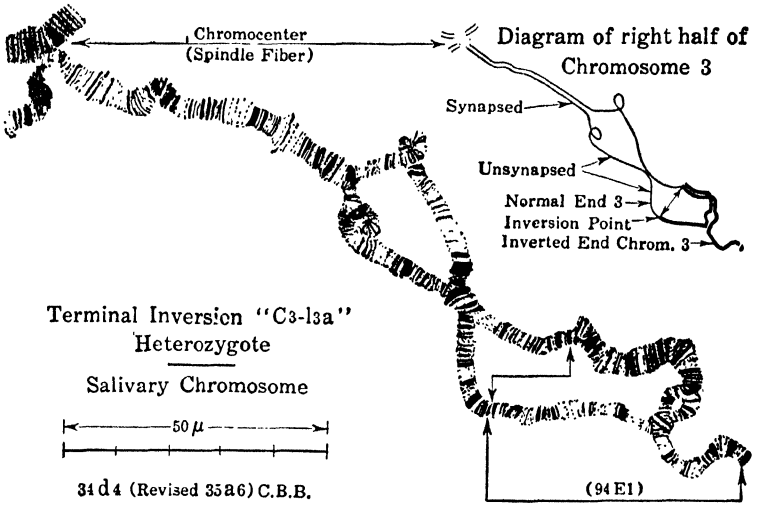


FIGURE 6.

sponding to the genetic constitution of the animal. The upper right-hand corner of Fig. 5 (scale of magnification indicated) shows the four Drosophila chromosomes as usually observed, the arrow indicating the two tiny chromosomes which, in the main part of the cut, are shown as they appear in the salivary glands, under the same magnification. In Fig. 6 we have a case of a salivary gland chromosome where com-

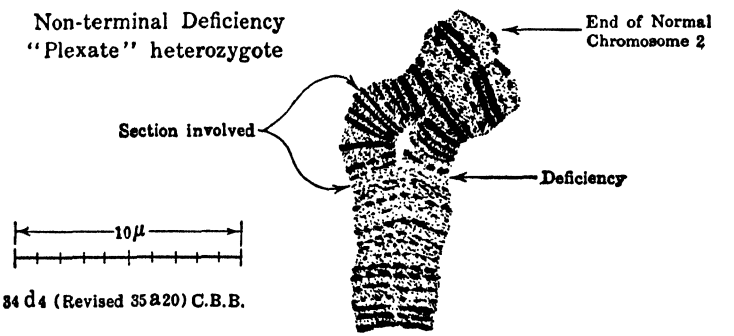


FIGURE 7.

plete synapsis has failed because of a terminal inversion. Fig. 7 shows a deficiency in a salivary-gland chromosome, lack of the gene-block causing an outward looping of the normal chromosome during conjugation. In Fig. 8 we have a portion of a salivary-gland chromosome, showing how the gene map finds its visual check in the

* *Science*, 66, 84-87 (1927).

"genetic spectrum." More complete concordances are illustrated in the elaborate chromosome maps published in the *Journal of Heredity* (1935 *et seq.*).

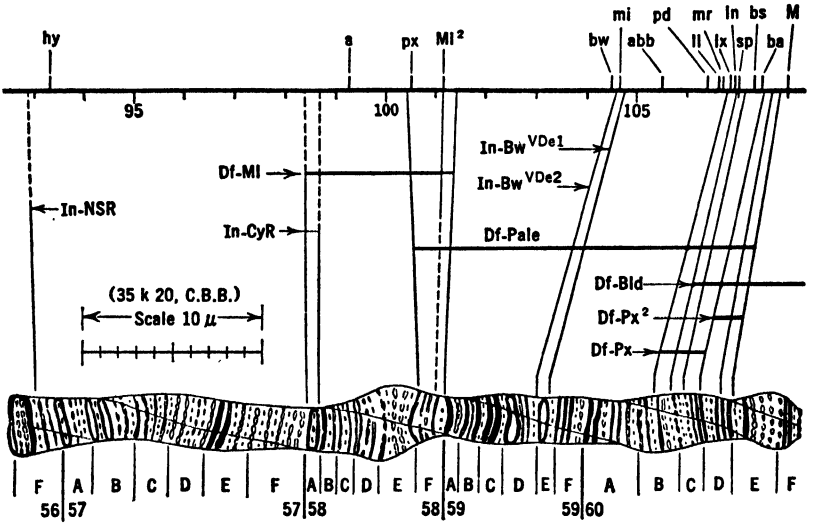
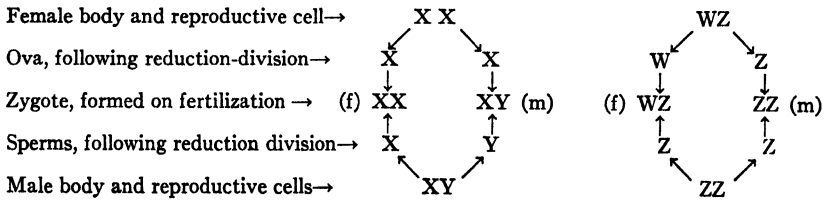


FIGURE 8. Section of the right end of second chromosome (salivary gland of *Drosophila melanogaster*) with corresponding portion of the chromosome map (after Calvin B. Bridges).

Sex Determination. Chromosomes of a predominant type are called *autosomes*, in contradistinction to the one or more pairs of "sex chromosomes" whose segregation determines the sex of each offspring. The basic sex chromosome, known as the X-chromosome, has no discovered mate in some species (XX-XO) types, while in other species it has a mate, the Y-chromosome (XX-XY types). Comparatively recently, the Y-chromosome was found in human sperm, placing man in a major XX-XY group in which the fertilized ovum that develops into a female has two X-chromosomes, whereas the male has one X and usually one Y. In man, reduction-division leaves the female gamete (ovum) with one X-chromosome, while the sperms are equally divided between X and Y, the sex of the human being determined by which kind of sperm chances to fertilize the ovum. In another group, including birds and moths, it is the female that has differening sex chromosomes (WZ), male cells being (ZZ) and yielding only Z sperms. The following diagrams indicate how these microscopic gene-carrying units dominate sex-determination:



The 48 chromosomes of man, 24 from each parent, are outlined in Fig. 9. In sex as in all other cellular developments, *all* the genes exert an influence, the net emergent results (only some of which may be easily demonstrable) depending upon the final results of the complex cellular catalyses, a summation expressed in the term *genic balance*. The biont (plant or animal) may die or may fail to reproduce if genic bal-

ance is too greatly disturbed. Thus when the chromosome pairs fail to separate in reduction-division (a phenomenon termed *non-disjunction*), abnormal numbers of chromosomes may enter a germ cell and thus enter the zygote. Some of the consequences with Drosophila are tabulated below, A representing a single set of autosomes; each parent contributes one autosome set.

Chromosomal Constitution	Sex of Individual
AAXY (normal diploid fly)	Male (normal)
AAXX " " "	Female (normal)
AAXXY	Female, giving abnormalities
AAXXX	"Superfemale," abnormal
AAXO	Male, sterile
AAYO	Dies,-sex?
AAAXXX (triploid fly)	Female, producing abnormalities
AAAXX	Intersex, sterile
AAAXXY	" "
AAAXY	"Super-male," sterile

In tracing demonstrable changes occurring in biological and technological phenomena, chemists naturally try to identify definite chemical compounds in every stage of each process; and this is often possible. But as has been pointed out repeatedly in this series, *e.g.*, on p. 11 of Vol. II, in a paper by J. Alexander and



FIGURE 9. The 48 human chromosomes, in outline (after H. M. Evans and O. Swezy, Mem. Univ. Cal., Vol. 9, No. 1, 1929).

C. B. Bridges, the line of demarcation between "chemical compounds" and "physical combinations" is purely a matter of convention and convenience. Nature is oblivious of textbook definitions. In his book on "The Nature of the Chemical Bond" (Cornell Univ. Press, 1939) Linus Pauling defines the chemical bond as follows (p. 3):

"We shall say that there is a chemical bond between two atoms or groups of atoms in case that the forces acting between them are such as to lead to the formation of an aggregate with sufficient stability to make it convenient for the chemist to consider it as an independent molecular species."

Where biological or technological "aggregates" lack the definiteness and stability entitling them to be called "chemical compounds," we must not close our eyes to the fact that they actually *are* aggregates, and that their existence, even though transitory, may be of great importance in the explanation and the outcome of a process.

It is hoped that the above brief outline will lead the reader to seek specialized books on genetics for further information. Before appending a brief glossary, we add the following quotation from a chapter on "The Genetics of Sex in Drosophila," by C. B. Bridges in "Sex and Internal Secretions," edited by Dr. Edgar Allen:

"Considered from the standpoint of its products, each gene is a chemical factory, devoted to the manufacture of a series of active chemicals and their delivery in sequence to the surrounding chromosomoplasm. Each gene in a given gene-string or chromosome has its own 'orders' to fill, and has the machinery for producing its particular set of manufactured chemicals. A chromosome is like a row of factories set upon the bank of a canal, all delivering their goods by means of the waters of the canal, as well as helping themselves to raw material brought by the canal."

Glossary of Some Terms Used by Geneticists *

Allelomorph (Allel, or Allele): The alternative form of a gene, having the same locus in homologous chromosomes. Thus in many organisms there is a gene which produces albinism. This is usually recessive to the gene for normal pigmentation, which is a dominant allelomorph of albinism.

Anaphase: The later stages of mitosis during which the daughter chromosomes diverge from the equatorial plate.

Atavism: Reappearance of a character after a lapse of one or more generations, as when a recessive character reappears after having been transmitted through several generations of heterozygous individuals.

Backcross: The mating of an F_1 hybrid to one of the parental races crossed to produce this hybrid.

Blastomere: One of the cluster of cells formed as a result of the first few divisions of the animal egg.

Centrosome: A minute body generally present in animal cells and in the cells of some lower plants, but rarely visible in flowering plants. It is the center of the *astral rays* (which appear during mitosis, and form the *aster*) which play an important part in cell division.

Character (short for "characteristic"): A term used, often rather vaguely, to designate any form, function, or feature of an organism. The Mendelian characters of genetics represent the end products of development in which a definite gene, or genes, has a decisive effect. "The constant and unchanging thing, therefore, is the factor (gene) itself rather than the character, and the unity which Mendel observed lies rather in this underlying factor than in the visible, and perhaps variable character which it produces." (Sinnott and Dunn).

Chiasma: "In the diplotene stage (in meiosis, this precedes diakinesis) the chromatids are associated in pairs in such a way that in one part of their length chromatids are associated but in another part each is associated with a different chromatid. The point of exchange is termed a chiasma." (Sansome and Phipp, 1932).

Chromatid: Chromosomes frequently bear the foreshadowing of one or more longitudinal divisions so that they appear to be made up of parallel threads. These threads are the chromatids.

Chromatin: That part of the nuclear substance which forms the most conspicuous part of the nuclear network and the chromosomes, and stains deeply with "nuclear" or basic dyes.

Chromocenter: Inert chromatin to which the chromosome elements are attached in the salivary glands of the Diptera.

Chromomere: One of the linearly arranged chromatin granules comprising the chromosome. (Identified by Belling and others with the gene, but by more recent workers considered an artifact produced by fixation of spirally coiled chromatids.)

Chromonema: A fine thread of chromatin material from which arises the spireme thread.

Chromosome Map: A diagram showing the relative positions of genes in linkage groups, determined by observation of cross-over frequencies, or by other means.

Chromosomes: Tiny, dark-staining bodies visible under the microscope in the nucleus of the cell at the time of cell-division. The number of chromosomes in any species is usually constant. The chromosomes carry the genes, linearly arranged.

Conjugation: Side-by-side association of homologous chromosomes as in the reduction division.

Crossing Over: The exchange of linkage relations between two characters due to an interchange of segments between the chromatids of two homologous chromosomes in the germ cells early in meiosis. This results in a change in the linkage relations of the characters as they appeared in the parents. See *Chiasma*.

Cytoplasm: The protoplasm of a cell in which the cell organs (nucleus, plastids) are embedded.

Cytoplasmic: Pertaining to the protoplasm outside of the nucleus; said of inheritance dependent on the cytoplasm or on objects contained in it.

Diakinesis: A stage of meiosis in which the homologous chromosomes are associated in pairs, just before reduction, when the pairs separate to complete the heterotypic mitosis in which the chromosome number is reduced.

Differentiation: The process whereby cells and tissues become specialized for specific functions during the process of growth and development.

* Based mainly on "A Glossary of Genetic Terms," published in the *Journal of Heredity*, **28**, pp. 71-80 (1937), assembled over twenty years "to make understandable to the not over-technical reader the peculiar lingo of genetics." Words understandable from the text of this bookk are for the most part omitted.

Diploid: Having two sets of chromosomes. Body-tissues of higher plants and animals are ordinarily diploid in chromosome constitution, in contrast with the haploid gametes. See *Somatic*.

Epigenesis: The concept of generation now accepted by biologists, which holds that the germ or embryo develops afresh in each generation through the interaction of genes and other protoplasmic components with the environment, in contrast with the theory once held that development was merely an unfolding of organs already microscopically present in the egg.

Epistasis: The suppression of the action of a gene or pair of genes not allelomorphic to the gene suppressed. Thus, the gene for wild pattern (agouti) in the rabbit is masked in the albino. Albinism in this case is said to be epistatic over the wild pattern, which is hypostatic to albinism.

Equatorial Plate: The figure formed by the metaphase chromosomes lying in the equator of the spindle during mitosis.

Eugenics: The application of knowledge of heredity to the improvement of the human race. "The study of agencies under social control that may improve or impair the racial qualities of future generations, either physically or mentally."

Fertilization: The name given to the events occurring after pollination or mating, which culminate in the fusion of a male gamete with an egg (ovum) and initiate the development of an embryo.

Gamete: A reproductive cell of either sex; *e.g.*, a sperm or an egg (ovum).

Gene: (1) The unit of inheritance, which is transmitted in the germ cells, and which, by interaction with the genic and cytoplasmic complex and the environment, controls the development of a character. The alternative units (allelomorphs), occupying the same locus in a chromosome, are held to be phases of the same gene. The genes are held to be linearly arranged in the chromosomes. (2) "The physical basis of heredity."

Genom: A chromosome set, as it is inherited as a unit from one parent.

Genotype: The entire genetic constitution, expressed and latent, of an organism, in contrast to *phenotype*.

Germ Plasm: The material basis of heredity, taken collectively. The sum of the gene-constitution of an organism.

Gynandromorph: An individual of which one part of the body is female in constitution, and another part is male.

Haploid: Single; referring to the reduced number of chromosomes, as in the mature germ cells of bisexual organisms. Cf. *diploid*.

Heterotypic division: One of the two meiotic divisions, in which the homologous chromosomes are separated to form daughter cells with the haploid chromosome number.

Heterozygote: An organism heterozygous (hybrid) with respect to any given pair of allelomorphs; that is, containing two different forms of the same gene.

Homeosynapsis: Synapsis between like chromosomes.

Homologous: Chromosomes occur in somatic cells in pairs; two members of a pair are spoken of as homologous chromosomes.

Homozygous: (1) (An organism) formed by germ cells alike with respect to a given gene. (2) An organism is said to be homozygous for a given character when all the germ cells transmit identical genes for that character. This is the most frequently used meaning of the term.

Identical twins: Two individuals developed from the same fertilized egg. They are extremely alike and are generally (but not always) found in the same chorion before birth. Monozygotic twins.

Interchange: An exchange of segments of non-homologous chromosomes.

Intersex: An organism displaying sexual characters and secondary sexual characters intermediate between male and female.

Leptotene: A term applied to the fine spireme-thread stage of the chromosomes before their union in synapsis.

Linkage: Association of characters in inheritance, due to the fact that the genes determining them are physically located in the same chromosome. Such a group of linked characters is termed a linkage group.

Locus: The position occupied by a gene on a chromosome, in a linkage group, or on a chromosome map.

Maturation Divisions: Meiosis.

Meiosis: The two divisions which precede the formation of gametes (or in plants the development of a gametophyte), in which the members of each chromosome pair separate, and the chromosome number in the resulting daughter cells is reduced to half the somatic number, as distinguished from ordinary cell division (mitosis).

Metaphase: That stage in cell division in which the chromosomes are arranged in an equatorial plate.

Mitosis: Cell division involving the formation (by self-duplication) of chromosomes, of spindle fibers, and the division of the chromosomes by a process of longitudinal splitting. Thus each daughter cell gets a full complement of the chromosomes which existed in the original cell before division. [Since the division of the cytoplasm between the daughter cells is not so precise, inequalities may occur in the distribution of the cytoplasm and any of its small randomly dispersed units].

Modifier or Modifying Gene: A gene which affects the expression of another gene.

Monoploid, Haploid: Also sometimes used in the sense of the basic haploid chromosome number in a ploydploid series. Thus in the genus *Rosa*, species are known in which the haploid chromosome numbers are 7, 14, 21, or 28 (the respective somatic chromosome numbers being 14, 28, 42 and 56). In this polyploid series the monoploid ("basic haploid") chromosome number is 7, but the haploid number in any given species of the series may be 7, 14, 21, etc., depending on whether the form is diploid, triploid, etc., in terms of the seven fundamental monoploid chromosomes.

Mosaic: An organism, part of which is made up of tissue genetically different from the remaining part. Mosaics are produced by mutations during development, or by chromosome non-disjunction, or by double fertilization of an egg, or in other ways.

Mutation: A sudden variation which is inherited, resulting in changes in the gene or genes concerned. The term is used rather indefinitely, to include "point mutations" (of a single gene) and chromosome deletions, rearrangements, duplications, chimeral changes, de Vriesian mutations, etc.

Mutual Translocation: Reciprocal transfer of parts of two non-homologous chromosomes. "Essentially crossing-over between non-homologous chromosomes."

Non-disjunction: (1) Failure of chromosome halves to separate in cell division, both halves going to the same daughter cell. (2) Failure of a pair of synapsed chromosomes to separate at the reduction division, both chromosomes of the pair going to the same gamete.

Nucleolus: A darkly staining body found in the nuclei of most cells. Its function is not clearly understood.

Nucleus: A highly refractive, deeply staining body of specialized protoplasm found within the cells of all organisms except bacteria. It contains the chromosomes and the karyolymph, the nucleolus, etc.

Ontogeny: The developmental history of an organism from egg (spore, bud, etc.) to adult individual.

Oöcyte: A female reproductive cell during oögenesis, up to the time of the second division. Oöcytes are designated as *primary* before the completion of the first division, then *secondary* thereafter. *Primary* have the diploid chromosome number and *secondary* the haploid chromosome number.

Pangen: A hypothetical particle (which might be likened to a "floating gene") suggested by Darwin as an explanation of inheritance and modification of characters in his pangenesis hypothesis.

Parthenogenesis: The development of a new individual from a female germ cell without fertilization.

Phenotype: The organism as exemplified by its expressed characters, as contrasted with its genetic constitution (the genotype).

Phylogeny: The evolutionary history of a species, genus, or type.

Plastid: One of the several kinds of protoplasmic bodies in cells, like the chlorophyll bodies in plant cells, which are centers of chemical activity.

Plastid Inheritance: Inheritance through plastids carried in the cytoplasm of the ovum. Variegation in several plants is known to be inherited in this manner.

Polar Body: A minute cell of which two are cast off and discarded from the animal ovum during maturation.

Prophase: The stage in cell division prior to the metaphase, during which the chromosomes appear in the nucleus, following the resting stage.

Protoplasm: "The physical basis of life"; a chemical compound or probably an emulsion of numerous compounds. It contains proteins which differ slightly in each species of organism. It contains carbon, hydrogen, oxygen, nitrogen and various salts, but is so complex as to defy structural analysis. [In a paper entitled "Some Chemical Aspects of Life, Food and Evolution" published in *Scientia* (Milan), Oct. 1933, J. Alexander stated: "Although isolated protoplasm may maintain activity for a short time under suitable conditions, it is incapable of self-reproduction and should be regarded rather as a highly specific milieu in which the real living self-reproductive units of cells exist and function. The concept of protoplasm as the ultimate 'living jelly' is a relict of antiquated text-books and should be definitely abandoned. The modern concept of protoplasm embraces the cytoplasm *with* its included nuclear and other particulate units." J. A.]

Reciprocal: Involving the same types of individuals, but with the sexes reversed; said of two crosses, in one of which the female possesses the same character as does the male in the other cross, and *vice versa*.
Recombination: An offspring which has a combination of linked characters different from that exhibited by the parents.
Reduction Division: The heterotypic division (*q.v.*).
Sex-Limited: Expression of a character in only one sex. Examples: milk production; horns in Rambouillet sheep; egg production.
Sibling (*Sib*): One of two or more children of the same parents, but not of the same birth.
Somatic: Referring to body tissues; having two sets of chromosomes, one set coming normally from the female parent and one from the male; as contrasted with mature germ cells having a single set of chromosomes.
Somatic Mutation: A mutation in a somatic cell, resulting in a chimeral individual with respect to the mutant character.
Spermatid: One of the two cells formed by the second division in spermatogenesis. By transformation in shape the spermatids become mature spermatozoa.
Spindle: A group of structures, observed in fixed cytological preparations, resembling threads, in the form of a spindle, formed during mitosis.
Sport: An abrupt deviation from type. An obsolescent term for mutation.
Synapsis: The conjugation or union in pairs of homologous chromosomes, respectively of maternal and paternal origin, to form bivalents; the primary step in meiosis.
Telophase: The closing phase of mitosis, during which the daughter nuclei are formed and the mother cell usually divides.
Test Cross: A backcross of a diheterozygote to a double recessive, to test linkage.
Tetraploid: An organism whose cells contain four haploid (monoploid) sets of chromosomes.
Trait: A term loosely used by geneticists as a synonym for "character."
Translocation: Attachment of a fragment of one chromosome to a non-homologous chromosome, resulting in a new arrangement of genes.
Triploid: An organism characterized by having three times the haploid (or monoploid) chromosome number.
Unit Character: A heredity trait that behaves as a unit in transmission, being capable of inheritance independently of other unit characters and determined by the allelomorphs of a single gene. Example: green and yellow seed color in peas.
Variation: In biology, the occurrence of differences among individuals of the same species.
Variety: In taxonomy, a division of a species; a group of individuals within a species which differ in some minor respect from the rest of the species. Agronomic varieties are the product of human selection, and differ in characters of economic importance.
Xenoplastic Graft: A heteroplastic graft that "takes" and continues to develop in its new site (most heteroplastic grafts are short-lived).
Zygote: The cell produced by the union of two cells (gametes) in reproduction; also the individual developing from such a combined cell.
Zygotene: A stage in meiosis, preceding synapsis (also called synezesis), during which the chromosome homologs begin to pair.

The Gene as a Chemical Unit

JACK SCHULTZ

The Lankenau Hospital Research Institute, Philadelphia, Pa.

The Genetic Technique and the Definition of the Gene

It is now a commonplace of biological thinking that the basic units of the living world are the genes. The experiments upon which this view is based owe their success to a special technique, peculiar to the field. This technique consists in the utilization of the internal variations of the biological system as indicators of the behavior of special components of that system. Mendel defined the individual units which we now call the genes in terms of their variants. And these inferred variants

of the ultimate units were recognized by the directly observable differences in the organism. Traced to their origins, these must in turn be the consequence of differences in the developmental processes of the standard type and its variant. For the definition of the gene as a hereditary unit, however, it was sufficient that the behavior of the hereditary system was explicable in terms of the interaction of these units. The nature of the changes in the postulated units, and how the observed differences could be traced to these changes, could be ignored. The result was the definition of the gene as a structural unit in the cell; and this definition was amply justified by the visible identification of these structural units as the elements composing the chromosomes.

With the increase of knowledge, a definition of the gene has evolved based upon its function in development. The early views of the nature of the gene, considering it essentially to be the archetype of intracellular catalyst, were founded on an argument by analogy. Minute in quantity compared to its effects and (so the argument went) not itself changed by the reaction it facilitated, the hereditary unit was compared to an enzyme (Troland 1917, Muller 1923, 1929, Alexander and Bridges 1928; see the review by Wright 1941); but an enzyme of a peculiar and unique type, an autocatalyst, which in addition to catalyzing other reactions, itself was the progenitor of its own kind (cf. Morgan 1926).* It is evident that the establishment of a comparison of this kind requires experiments in the domain which had been disregarded in the primary analysis of the hereditary mechanism. Now the nature of the developmental change must be studied, and the attempt to define a gene in terms of its function leads us into chemical embryology. A change in the sequence of developmental reactions is effected by the substitution of a mutant gene for the normal one; thus a specific series of synthetic reactions is dissociated from the welter going on in development. The precise determination of the differential reactions can then guide us towards a functional definition of the gene. As a result of many analyses of this kind, the picture has emerged of a network of reactions occurring in development (for references see Goldschmidt 1939, Wright 1941, Haldane 1942, and the chapters on physiological aspects of genetics in the Annual Reviews of Physiology). Each of these reactions is presumed to be governed by its specific catalyst; and in some cases it has been demonstrated that the synthesis of the catalyst itself is changed in the mutant type (Daneel 1941, Moewus 1941). Thus the early analogy with enzymatic processes is correct in principle: functionally the gene is concerned with the production of the enzymes of biosynthesis.

But for the precise definition of the gene as a chemical unit evidence of still another sort is required. The enzymes of biosynthesis must be traced to their source in the chromosomes. And the direct analysis of the chromosomes themselves provides the starting point here. The substances of the chromosomes must include the genes and possibly their primary products. Hence the development of techniques whereby the transformations of these substances can be followed within the cell, makes possible the critical experiments. These have as their aim the pursuit of synthesis in the individual cell from the initial reaction in the chromosome to the final effects in the cytoplasm. Analyses of this kind have only begun; they are essential in that they provide the topographical evidence needed to complete a functional definition of the gene.

From the point of view of biochemistry, it is evident that the genetic technique, with the implied analysis of the gene as a chemical unit, is basically an attempt at the study of synthetic processes in the intact organism. The degree to which these attempts are successful is reflected in the accuracy and the detail of the definition of the gene. In the following, we shall consider the genetic evidence for the existence of genes as separable units, more or less autonomous in function. It will be seen that

* Several papers in this volume deal with heritable changes in biocatalysts, as well as with *catalyst modification,* a notion advanced by J. Alexander (*Protoplasma,* 1931).

this involves consideration of their arrangement in the chromosomes, of the mechanism of the maintenance of this arrangement, of questions of gene change and the like. Having discussed these problems, we will proceed to the discussion of the chemical evidence associating special constituents of the chromosomes with genic activity. It will appear in the sequel that the evidence is accumulating for the concept that the genes are nucleoprotein in nature, and that their primary functional activities are concerned with a system of nucleoprotein syntheses which lie at the basis of the synthetic activities of the cell.

The Linear Order of the Genes

The first experimental demonstration that the specific hereditary units are arranged in linear order in the chromosomes came from the study of the phenomenon called crossing over (Morgan 1911, Sturtevant 1913). It is the separation from each other of linked characters in breeding experiments that is crucial. For one of the major criteria for the distinction between the several genes within a chromosome is the production, by hybrids for a number of variants, of gametes containing recombinations of the original characters. In other words, the individual units are separated by a process of exchange between the chromosomes. How far these units are the irreducible elements will be discussed presently. But it is the circumstance that this particular system of producing new combinations happens to be prevalent among animals and plants, that is at the basis of the elaboration of Mendelian heredity into the theory of the gene. There is no need here to review the experiments that have been carried out: their major result, the construction of the classical chromosome maps of Drosophila, is sufficiently well known (for an introductory account, see the text of Sturtevant and Beadle 1939).* It is one of the familiar triumphs of genetic and cytological analysis that the correspondence is complete between the linear sequence deduced from the breeding experiments and that observed in the microscopic analysis of the chromosomes themselves (Painter 1934, Bridges 1935).

The linear order of the elements of the chromosomes was first inferred from the fact of the division of the chromosomes, as a necessity if mitosis were to be equal. The same line of reasoning is in fact now used to argue for the necessity of a two dimensional arrangement to permit the synthesis of identical molecules. It is not so frequently realized that there are further implications of the argument. An arrangement in a linear series is also convenient for the maintenance of precise relations for functional purposes, assuming for example that the individual members are centers of synthesis. But does the linear series persist during the functional phases of the nucleus?

The synthetic work of the nucleus is performed during the interphases, when the chromosomes have undergone a series of maneuvers whose end is the formation of a functional nucleus. Within this, the unwound spirals of the metaphase chromosomes persist, with their nucleolar products; and upon its cytoplasmic surface, the nuclear membrane must retain part by part a definite relation to the chromosomes whose fused surfaces it represents. As one possibility, conceive that each chromosome produces a monolayer, patterned according to its own chemical design, which serves as a template for the cytoplasmic syntheses with which we are visibly concerned in differentiation. The implications of the linear series then become important; changes of the specific pattern due to changes of the linear series could have far reaching effects. The very definition of the gene as a unit becomes difficult. The test of such possibilities can be made however: rearrangements of the chromosomes are known, and their effects are relatively slight except in the immediate neighborhood of the point of rearrangement. Yet these effects must be understood, in order

* For the benefit of those unfamiliar with genetics and its technical terms, an epitome and brief glossary are given in the preceding paper of this Volume. J. A.

to set the limits of the concept of the gene as a unit. For the proper understanding of the data, it is advisable first to consider the differentiated regions of the chromosomes and their relation to the function of the nucleus.

The Differentiated Structures of the Chromosomes

The metaphase chromosomes of small cells show little external differentiation, except for an area which lacks the staining capacity shown by the remainder for basic stains (for a review see Geitler 1938). This "primary constriction" is the region which contains the "centromere," the chromosomal organelle associated with the mechanism for the separation of the daughter chromosomes at mitosis. The centromeres are correlated in their function with the central bodies of the cytoplasm; they are connected by fibers in the mitotic spindle, and are apparently homologous bodies (Schrader 1936, Darlington 1939, Pollister 1939). Various types of experiment agree in attributing to the centromere the active movement of the chromosomes to the poles of the spindle; fragments of chromosome lacking a centromere separate, may even be included in the daughter nuclei (McClintock 1938, Carlson 1938) but form no attachment to the spindle.

It may be remarked in passing that the spindle mechanism appears to be independent of the reproductive mechanisms of the chromosomes, the contrary also being true. Thus under the influence of such drugs as colchicine or acenapthene (for a discussion see Blakeslee 1941), the spindle is suppressed, but the chromosomes reproduce: no cell division takes place, and the reconstructed nucleus is polyploid. These effects are of interest in the study of the physiology of the centromere, in addition to their great use in polyploid study.

In the larger types of metaphase chromosome, more detail is apparent. A coiled thread, the chromonema, can be demonstrated: the extended string of genes is compressed here into a spiral (for discussions of the spiral structure of chromosomes see Geitler 1938, Darlington 1937, Huskins 1942, Kuwada 1940, White 1942). In the very largest chromosomes, there is evidence of still further differentiations: the thread of the large coiled spiral itself is contorted into what has been described as a minor spiral. This latter seems however more easily to be interpreted as the visible evidence in these chromosomes of the fundamental differentiation existing in the chromonema, into the chromomere and its interchromeric portion.

The chromomeres have been most extensively studied in the meiotic prophases, and in the giant chromosomes found in the salivary glands of the Diptera. They are most evident when the chromonema is maximally extended, exhibiting at this time a degree of diversity that led Belling (1928, 1933), to consider these ultimate chromomeres as the visible loci of the genes. He considered the interchromomeric substance to be non-specific, a mere connecting material between genes. This distinction, based originally on the reaction to basic dyes, is a fundamental one, and will be discussed in detail in later sections. It may be said at once however, that in their variety, their property of specific association with each other in the nuclei where pairing of homologues occurs, and their approximate number where they can be counted, the chromomeres are likely candidates for the role in which Belling cast them.

There is one especially striking differentiation of chromomeres, to which Heitz drew attention (review, 1935). Resting nuclei have long been known to contain deeply staining bodies, which Heitz proved were special parts of the chromosomes, designating them heterochromatic regions, after the similarly deeply staining sex chromosomes (heterochromosomes). Since that original definition, the scope of enquiry has widened. By analogy with inactive deeply staining nuclei, these regions were supposed to be genetically inert, and evidence of various kinds in genetic work supported the assumption. More recently, synthetic functions of a non-specific nature have been accorded these regions, and a succession of new attributes found

for them: special behavior at synapsis (Schultz 1936, Prokofyeva-Belgovskaia 1939, Schrader 1941); relation to nucleolus formation (Heitz 1931, McClintock 1934; see the review of Gates 1942); a differential reaction to low temperatures (Darlington and La Cour 1940, Callan 1942; but see also Wilson and Boothroyd 1942); and a differential reaction to X-radiation (Muller 1940, Prokofyeva-Belgovskaia and Khvostova 1939, Kaufmann 1939). These functional characteristics will be discussed below.

The distribution of heterochromatin in the chromosomes suggests a localization at centromeres and at the ends. The major heterochromatic masses are found flanking the centromere in the great majority of cases. But there are accumulating data which suggest a much more varied distribution of heterochromatic regions; indeed Kaufmann has made the suggestion that the dense chromomeres of resting nuclei (not Belling's ultimate chromomeres) may belong to this genre. In the salivary gland chromosomes of the Diptera, where the chromomeres are enormously enlarged, the heterochromatic regions generally have a vesiculated type of chromomere instead of the dense band which is set up as the paragon of excellence in the euchromatic regions. Bauer (1936) has suggested that these be named heterochromomeres; in that case however the high frequency with which such vesicles occur on occasion in Drosophila (cf. Bridges 1938) would lead to difficult decisions in classifying the regions in such an organism as Drosophila melanogaster. What the situation is in other organisms is still less clear, and one may guess that accidents of history play a large role in this distribution. Yet if this be the case, "accidents" which tend to concentrate heterochromatin around the centromere are especially frequent.

Position and Function

Against this brief outline of chromosome differentiation, the functional consequences of chromosome rearrangement can be examined concretely. When a chromosome rearrangement occurs as the result of exchange of segments, in addition to the major disturbance, the four genes at the point of rearrangement have exchanged neighbors (for an introductory discussion see Sturtevant and Beadle, *loc. cit.*). There is however a simpler situation, the tandem duplication, with a serial repetition of a chromosome segment, in which the number of exchanges of neighbor is reduced to one. These cases have proved particularly instructive; the discovery of the previously unsuspected relation between the position of genes in the chromosome and their function was made in the analysis of such a case.

The reversions to the wild type of a dominant sex-linked mutant in Drosophila melanogaster, Bar eye, were shown by Sturtevant (1925, 1928) to be associated with a process of crossing over whose results were unequal. One of the emergent chromosomes contained two representatives of the "mutant gene," the other contained none, and was wild type. The basis of this phenomenon became clear when the salivary gland chromosome analysis (Bridges 1936, Muller, Prokofyeva-Belgovskaia and Kossikov 1936) showed that the mutant was associated with a tandem duplication for a small section of chromosome. The apparent mutant gene could be due then to the duplication, or to the single change of neighbors, or to a simultaneous or coincidental occurrence of an intragenic change in the duplication. Further analysis was provided by the occurrence of a series of similar mutants in the progeny of X-ray experiments (Dobzhansky 1932, Dubinin and Volotov 1936, Volotov 1937, Sutton 1943). It appears from these studies that the mutant is in fact associated with rearrangements in which a specific locus in the sex chromosome is separated from its normal neighbor. The nature of the new neighbor appears to be of minor consequence. Thus the functional change depends upon a close association: the effect demands collaboration between a special pair of neighboring genes, localized at neighboring salivary gland chromosome bands.

However the original "position effect" did not concern the Bar mutant itself, but the derivatives produced by the asymmetrical crossing over. Sturtevant found that in these derivatives the departure from the normal phenotype was greater in the cases where the additional segments were accumulated in one chromosome as a series of tandem duplications, than it was in individuals with the same number of segments distributed between the members of the homologous pair. He was able in this case to perform the crucial experiment of recovering a segment from a chromosome in which the position effect had been manifested, and showing that its own intrinsic properties as a single tandem duplication, were unchanged. More recently, Rappoport (1940), in an elaborate and ingenious set of experiments, has extended the series to as many as eight segments arranged in a tandem order. He has thus been able to study the conditions under which the interaction between the successive tandem duplications, the original case of position effect, is manifest. It appears that this effect is at its maximum in the lower members of the series, the curve flattening out as the distances grow longer and other conditions of balance are operative.

Thus in the Bar case there is evidence for two kinds of position effect: one, the dependence of a gene on the presence of specific immediate neighbors for the maintenance of its effects in synthesis; and second an effect over a slightly greater distance, the coöperative function of similar genes close together. The two types should perhaps not be contrasted too sharply: both must depend ultimately upon competitions for available materials of synthesis, or accumulations of unusable reserves due to the new proximities. In any event, the question, concrete in this case, may assume the more general form: in how far can the genes be considered as legitimate separate units?

An essential datum for the answer is the distance along the chromosome over which the effects of changes in gene position are manifest. If this is a long distance, the view must be examined seriously that the pattern of the chromosome as a whole determines the specificity of its local components (Goldschmidt 1938). At the other extreme, where the distances are short the chromosome is merely the linear association of specific elements whose interrelationships depend rather upon the properties of the individual components than upon their modes of association in series.

The evidence already reviewed in the Bar case goes to show that the distances are short. Indeed the Bar mutant itself approaches the extreme: the distances are presumably the distances between genes. And in another case, the position effect due to a tandem duplication of the achaete * locus in Drosophila melanogaster (Demerec and Hoover 1939), separation of the responsible elements by a two band inversion has been sufficient to remove the position effect (Schultz 1941). Again the relation is one between adjacent elements.

The serial duplication of chromosome segments is apparently a rather frequent phenomenon. Thus detailed study of the morphology of the giant chromomers of the salivary gland chromosomes of the Diptera has shown, in addition to the remarkable pattern of differentiation, a high frequency of a definite type of double band (see Painter 1941a, Bauer 1936, Bridges 1935, 1938 and Metz 1941). More elaborate repetitions of pattern are also found. The interpretation has been offered (Bridges 1935, Offerman 1936) that these are duplications of the hereditary material that have become established in the course of evolution. The occurrence of tandem duplications under experimental conditions shows that one of the processes postulated in this interpretation still occurs. It follows that the analysis of the genetic conditions within such duplicated chromomeres should throw further light on the degree to which the genes in their present form are independent units. For the high frequency of these doublets, and the occurrence of numerous cases of association of

* The achaete fly lacks certain spines or hairs. J. A.

similar mutants in close linkage as a genetic parallel, suggests that local coöperative effects may be frequent. And those established in the chromosome might obey different rules than those newly arisen.

The most complete analysis thus far is that of the Star locus in Drosophila melanogaster. Lewis has been able to show here that the double band (a "capsule" formed by the association of its members) contains two separable genes, one located in its right, the other in its left member. Both have similar phenotypic effects. Deficiency or inactivation of the left component (Star) changes the component at the right (asteroid) in such a way that it no longer exerts its full function in maintaining dominance over a recessive allel. Adequate crossover tests show that the effect is a position effect, demonstrating coöperative action between the members of a doublet. The maintenance of complete functional efficiency at the asteroid locus demands the action of a normal allel at the Star locus. Serebrovsky (1938) has suggested that in duplicate genes of this type, a process of complementary loss is likely to take place; modification is such that each member of the pair retains those qualities that the other loses. The result would be genes which were non-allelic by the test of heterozygosis—each containing the dominant for the other—but which would occupy neighboring loci. If this is the case, Lewis' analysis of the Star complex shows that some of these losses concern primary reactions, at the site of the chromosome. The presence of both members is necessary for the normal function of the one whose primary synthetic product is inadequate.

The observed effects in other euchromatic regions may depend upon similar upsets of coöperative systems. Yet effects of other types may be involved; relatively few of the rearrangements involve breaks within "capsules" (Lewis), but most of them show genetic changes at the points of rearrangement (for a review of the earlier literature see Dobzhansky 1936; Demerec 1939). Hence the greatest number of changes in gene function associated with changes of position do not seem to be concerned with the interactions of duplicate genes.

Our discussion so far has dealt with the euchromatic rearrangements in Drosophila, and the effects observed there. The picture is different in other material. In maize for example, no clear cases of functional changes at points of rearrangement have been found, although they have been looked for carefully (Stadler 1941b). The existence of so great a discrepancy raises the question of the comparability of the experiments in the two organisms. The rearrangements compared are in both cases the produce of X-ray experiments, performed upon sperm or pollen grains in which the chromosomes are in a comparable state. It is possible that the dominant lethal types may differ, and that the missing position effects in maize are the aborted embryos so frequent after X-radiation of pollen. It would, however, be an odd occurrence that rearrangements should be induced by radiation with a high frequency in both forms, leaving an excess of surviving rearrangements in Drosophila with position effects, and none at all of that type in maize. The question is whether there is a difference in principle between the chromosome structures or the manner of gene function in Drosophila and maize.

Maize is of course a plant and Drosophila an animal; but a case of position effect has been found in Oenothera (Catcheside 1939), and the rearrangements studied in mice show no changes at the points of rearrangement (Snell). A possible clue (Muller, 1941; also J. Schultz) appears in the system of pairing in maize; in contrast to the high degree of specificity found in Drosophila, where intimate association occurs between chromomeres in the somatic prophases, even at meiosis in maize, pairing may be non-homologous (McClintock 1933). If position effects are the evidence of specific functional relations between genes, the pairing relations are the structural results of associated specificities; and the absence of position effects in maize may indicate a correlation between the structural and functional evidences of specificity. In maize it would appear that almost any neigh-

bor will serve either to work with or to pair with; in the Drosophilas, the individual gene is more limited.

The discussion thus far has been confined to position effects within the euchromatic regions. Heterochromatin shows somewhat different effects; the changes in euchromatic regions which have acquired neighbors of heterochromatin extend over a greater distance than is usual in euchromatic rearrangements (Schultz 1936a,c, Demerec 1940, 1941). The functional change in these heterochromatin rearrangements frequently differs in the different cells of the developing organism, with the result that the adults display variegation for the characters dependent upon genes in the region juxtaposed to the heterochromatin. These cases will be discussed later, where their relevance to the analysis of gene reproduction will be considered. Their importance here consists in the fact that they appear to show position effects extending over longer distances than those we have considered. However the study of the variegation patterns, and other evidence, shows that these long range effects are generally exerted over a series of cell divisions. Thus the initial effect here too is a short range one; even in these cases the larger pattern of the chromosomes has little to do with the events in the separate smaller regions.

The cursory examination we have given the data of the functional changes consequent upon chromosome rearrangement is sufficient to show that the concept of the gene as an autonomous unit in the chromosome is justified by the test they impose. The interrelations between genes in the linear series are of the special sort that might be expected from a system that has developed its evolutionary history as the chromosome must have. But in these interrelations, the gross pattern of the chromosome appears to play no role discernible from the present data. Having thus the evidence that the units exist as such, we may proceed to the analysis of their relation to the visible components of the chromosomes. A simple approach to this question lies in the consideration of the means by which the genes are maintained in linear order: are the connections directly between genes, or is there an especial intergenic material having this function?

The Mechanism of Chromosome Rearrangement and the Maintenance of the Genes in Linear Order

As has been said, the basic differentiation in the continuous chromosome fiber is between chromomere and connecting substance. Since the chromomeres have so distinctive an individuality, the temptation is great to identify the functionally distinguishable units of the chromosomes with them. In this case, the problems associated with the maintenance of linear order, and especially of the nature of the intergenic connections, concern the nature of the interchromomeric substance. But it is equally conceivable that the visible differences between chromomeres are effects and not causes. Evidently the analysis of the mechanism of rearrangement of genes, that is, the processes by which the connections between genes are broken and reëstablished should prove informative.

Exchange of Parts between Homologues: Crossing over

The mechanism of the normal exchange of parts of chromosomes that occurs during most meioses, crossing over, has been the center of as acute controversies as have developed in the history of genetics. The microscopic study of the critical stages is difficult, the genetic analysis of the relevant data complex. The cytological and genetic analysis early agreed in placing the occurrence of the event early in the meiotic prophases, possibly coincident with the chromosome split (see for example Muller 1916, Janssens 1909, 1924, Wilson 1928, for some of the earlier discussions). At this time four chromosome strands are in close association, with the exchange taking place between homologous genes in two of the four strands. These exchanges have been interpreted as the cause or the consequence of the cross

shaped configurations (chiasmata) seen at the later stages of the meiotic prophases; at present the former view holds the field (Darlington 1937). The rationale of crossing over has been considered in two ways. Of these one has supposed a simultaneous breakage at identical points, and a subsequent fusion in the new arrangement of the paired homologues (a recent statement in Darlington 1940). The implication here is that the chromosome is a continuum, as far as reproduction is concerned, and that a mechanism exists for coördinating the breakages so that they occur precisely between homologous genes. The contrasted point of view (Belling) derives crossing over from the mode of reproduction of the chromosome fiber. The chromosome is reproduced in two stages; the interchromomeric fibers are the products of a synthesis by the chromomeres. These are the basic self-perpetuating units. At meiosis, reproduction occurs while the paired strands are twisted about each other, the interchromomeric substances are formed between the new partners in a twisted configuration, and the resultant fiber is a crossover. At the time it was proposed the suggestion that the intergenic connections were the product of gene synthesis was a completely novel one. Its implications for the more general questions of chromosome behavior have only begun to be appreciated. Many of the issues raised at the time were special ones, connected with the calculation of crossover data, and with the cytology of chiasmata, and served to obscure the fundamental question of reproductive behavior.

At the present time, the breakage and reunion hypothesis of crossing over has been amplified by considerations of the stresses imposed upon the paired chromosomes as the result of their cycles of spiralization. In this way the attempt has been made to devise a mechanism for the precision of the breakage in both homologues (cf. Darlington 1937). The Belling hypothesis has not been re-evaluated in detail in recent years, nor have there been researches designed to test its basic assumptions. It may be said that many of the specific objections to it that were earlier urged on the basis of the genetic data, can easily be surmounted by a closer consideration of the opportunities for the postulated twists now afforded by the more recently studied cycle of spiralization.

Thus the study of crossing over affords an opportunity for the analysis of the mechanism of the maintenance of the genes in linear order, as it earlier gave the means of establishing the existence of this linear order in the chromosome. The promise is far from fulfillment, however; and yet, as we shall see, the problems already met here recur constantly in this study.

The Mechanical Breakage and Reattachment of Non-Homologous Chromosomes

While the exchange of homologous parts of chromosomes during crossing over is a regular feature of most meiotic cycles, the exchange of non-homologous portions, leading to chromosome rearrangement, is a much rarer event. Rearrangements are produced experimentally however, and much attention has been given to the analysis of their occurrence. For our present purpose, the problem is the same as that met in crossing over, with the difference that it has become more general, including all chromosomes and all stages of the cycle.

Chromosomes, like other objects, may be broken mechanically. In maize, McClintock (1938a, b; 1941a, b) has developed ingenious genetic and cytological techniques for detecting the occurrence of such breakages, and for the analysis of their characteristics. Plants containing special chromosome rearrangements provide sources of chromosomes with two centromeres. At the anaphases of the proper mitoses (in most cases the meiotic divisions) these dicentric chromosomes form bridges between the two daughter cells, bridges which are eventually broken as telophase is completed. The chromosomes at this time are under obvious tension; if the breakage were the result of the tension, it would follow that the portion broken must be the elastic component. This is known from microdissection experiments

to be the interchromomeric portion (Buck 1942). We are, however, in a less satisfactory position; for the actual breakage does not occur until the telophase, when the cell wall between the two daughter cells is formed, and may be more complex than the simple overstepping of the elastic limits.

Reattachment of the broken fragments occurs where other broken ends are present and available for reunion. Since there is no evidence at present that genes are lost at the point of reattachment, the reattachments must occur between genes. Since, however, we cannot say with certainty which of the chromosomal components was broken, it is equally difficult to say which attaches, the chromomere or the interchromomeric substance. A recent striking study by McClintock indicates that the reattachment occurs in the period of the telophasic reconstruction of the nucleus, and, in certain cases, that once this period has passed without reattachment, the broken end has lost the capacity to form attachments to other broken ends. Since this is the period of the chromosome cycle when the chromomema increase in length, and interchromomeric substance is forming, it is evident that a plausible case can be made for the importance of the interchromomeric substance in these mechanical breakages and reattachments. But the data do not as yet afford a clear discrimination, as has been seen.

Radiation-induced Chromosome Rearrangements

Our difficulties in the study of the mechanical breakages of chromosomes resulted from our failure to distinguish between tension and constriction as the immediate cause of breakage. The radiation-induced breaks present a different problem. Here the effective agents are powerful short-wave radiations, x-rays and the like. These cannot be useful in a simple discrimination between the substances in the chromosome that they affect: to a speeding electron all matter is an opportunity for energy transfer. Ultra-violet light is more delicate, and has given some information of use. But the problem is further complicated by biological considerations implicit in the techniques employed. These complications arise from the fact that the broken chromosomes observed in the experiments are usually the result of new reunions of chromosomes. In other words the nature of the initial excitation is inferred from the product of a complex series of chromosome syntheses. And the dissociation of the various steps in these processes has only begun.

The techniques for the detection of the chromosomal effects have depended upon the organism studied. Where convenient, as in Drosophila, genetic techniques are employed, with the advantage that the individual rearrangements are available for as detailed study as proves necessary. Under other circumstances, cytological examination of the chromosomes either in the abnormal mitoses following irradiation (e.g., Sax 1941, Marshak Carlson 1941, among others) or in the interphase nuclei of the heavily exploited salivary gland chromosomes (Bauer, Demerec and Kaufmann 1938, Bauer 1939, Kaufmann 1941) has been employed. With both techniques, the analytic procedure depends upon the type of chromosome irradiated and upon the number of mitoses between the irradiated and the observed chromosome. Thus when spermatozoa, which are a relatively homogeneous population, are irradiated, we deal with the state of the chromosomes in the sperm at the time of radiation, their metabolic state during the inactive period before fertilization, the processes which occur as the sperm head is transformed into the male pronucleus after fertilization. Furthermore, the statistics of the types of rearrangement discovered pose questions of their own: are the examined rearrangements a random sample of the events occurring in the irradiated sperm, etc.? (For recent discussions see Muller 1940, Fano 1941, Lea and Catcheside 1943, Giles 1943.)

In terms of the problem with which we are presently concerned, the delimitation of gene and intergenic bond, a primary question is the relative stabilities of the two components with respect to the impinging disturbance. If for example, the gene

were more sensitive, the problem of breakage and reattachment becomes one of re-establishment of linear connections between two neighboring genes either in the presence of the debris of an intermediate destroyed fellow, or of one which while maintaining its spatial relations has lost the power to reproduce. On the other hand were the intergenic bond the more sensitive, the problems are the conditions of its regeneration. And in both cases, such factors as the initial proximity of genes adjacent to the injured link, rates of regeneration as compared with rates of separation of the adjacent members, and chances that rearrangement instead of restitution of the original sequence will occur, all enter.

It should be pointed out that the assumption has been made that the initial excitation must result in the breakage of bonds. This excitation cannot however be equated to the breakage of chromosomes: indeed, for purposes of chromosome rearrangement, where contiguity is a condition, an excitation between two adjacent strands, serving to attach them, provides a possible mechanism. Actually in the history of the subject the dominant contrasted hypotheses were the "contact" and the "breakage first" hypothesis. According to the first a single excitation sufficed to attach two contiguous strands, whose subsequent maneuvers during separation produce the observed rearrangements. Apparent single breaks are on this hypothesis explained as the result of attachment of the sister strands of a newly split chromosome. It is to be noted that the assumption that a single excitation suffices to produce an attachment is purely gratuitous: *a priori,* were the hypothesis correct, this might equally well be a multiple hit phenomenon. On the "breakage first" hypothesis, independent single excitations give rise to independent breakages of the chromosomes. The broken chromosomes reattach—and that they should do so the mechanically broken chromosomes attest—and rearrangements are formed. Thus the minimum requirement for the production of a rearrangement is two hits.

According to the "target" hypothesis of radiation effects, the number of hits required to produce a given effect can be determined from experiments in which dosage, wave length are varied. To this end a large body of data have been accumulated, particularly in Drosophila and Tradescantia, forms then in which the chromosomes when treated are quite different in their size and in the details of their behavior (Bauer 1939, Kaufmann 1941, Fano 1941, Giles 1943, Lea and Catcheside 1943, Sax 1941, Muller 1940). The data show clearly that for x-rays, more than one excitation is required to produce rearrangements; neutrons, however, with a denser distribution of energy, give one hit curves. Spatial relations, movement of the chromosomes, and other factors of a similar nature have been shown to play a considerable role also. It is clear that whether or not the breaks are independent, the reattachments depend upon the availability of combining ends. And there seems also to be, according to Newcombe, some evidence for a direct effect in Tradescantia upon the restitution process, which complicates the analysis further.

For our present purpose, the point of chief interest is the nature of the initial disruption, and the demonstration that new connections are regenerated. Chromomeres are not usually lost at the points of rearrangement in Drosophila, where the analysis of the salivary gland chromosomes gives sufficient detail for conviction. Neither, as far as the evidence from a relatively elastic substance can take us, is there any interchromomeric substance lost. The picture then must be one of the destruction of intergenic connections and their subsequent regeneration. It appears that the conditions for this regeneration are optimal in the stages when the chromosomes are uncoiling to form a resting nucleus, when it may be presumed that the interchromomeric connections are growing. Thus when the condensed chromosomes of the Drosophila sperm are irradiated, breaks are stored, with the new attachments forming after fertilization, in the male pronucleus (see Muller 1940, also Kaufmann 1941). Similarly, a series of data from the different stages of the meiotic sequence in the eggs of Sciara (Bozeman 1943) show that the metaphase and

anaphase chromosomes likewise accumulate the disturbances, giving rise to many rearrangements. The earlier stages, with greater distances between the chromosomes, and opportunities for immediate restitution of the original sequences, are relatively low in the rearrangements produced upon irradiation. Apparently the most favorable circumstance for the production of rearrangements by x-rays is the subjection of a condensed chromosome for the initial effect, following which the optimum conditions for the regeneration of the new sequences are present.

The behavior of the chromosomes when irradiated with ultra-violet light is somewhat different than it is under the influence of the stronger short wave radiations. Both the genetic data in Zea mays (Stadler and Uber 1942) and the cytological data in Tradescantia (Swanson, 1942) show that the frequency of chromosome rearrangement is low, although the frequency of singly broken chromosomes is high. There is similar evidence in Drosophila, of a low frequency of rearrangements; no singly broken chromosomes are recovered however. Thus the changes appear to be different than those produced by x-rays, and a recent striking study of Swanson's shows a suppression of the x-ray effect under simultaneous ultra-violet irradiation. He interprets the result as evidence for a dual effect of the ultra-violet light, on the chromosome for breakage, and a stiffening (coagulation) of the surrounding sheath which prevents the chromosome movement necessary for rearrangements to occur.

The experiments upon the relative efficiency of the different wave lengths of the ultra-violet, designed to identify the sensitive component of the chromosome, have been carried out in Zea mays, where their genetic nature can be determined, as well as in the moss Sphaerocarpus (Knapp and Schrieber 1941) and in fungus spores (Hollaender and Emmons 1941), where the mutants found may or may not be chromosome fragmentations. In all three cases, a similar type of efficiency spectrum has been found, approximating the known absorption of the nucleic acids. These substances are extracted from the chromosomes, and as we shall see are responsible for the staining properties of the chromomeres. Hence the production in Zea of these single breaks, for such they appear to be, as the result of absorption by the nucleic acid component, indicates that effects upon the chromomeres disturb the intergenic bonds. Whether this is the result of inactivation of the terminal gene so that restitution can no longer occur, or whether the nucleic acid portion is itself intergenic, does not follow directly from the data as they stand.

Gene Reproduction

As has already been indicated, the study of the disruptions and reestablishments of connections in the chromosome rearrangements leads directly to the consideration of the nature of chromosome reproduction, and hence to gene reproduction. It is commonly considered the mechanism of gene reproduction is essentially a duplication: the mother gene affords a template upon which processes analogous to crystallization take place (cf. Alexander and Bridges 1928, Wrinch 1941) or more complex transformations (Delbrück 1941). One may question how far gene reproduction is a real duplication, that is, that there is a doubling of mass followed by the separation into two equal entities. The assumption that the gene has a constant size in different cells is based partly upon the idea that a given molecular species must have a constant size, and partly upon the equality of the daughter chromosomes at mitosis. Neither of these assumptions is binding. In the world of molecules, polymers of a given species are frequent, and particularly so in the proteins. Nor do the chromosomes maintain a constant size, or even a constant number of strands per chromosome in the different tissues of the organism (see Geitler 1941, Nebel 1941, Darlington 1937 and the recent paper of Biesele, Poyner and Painter 1942 for the discussion of various mechanisms and controversial points). It would seem that further evidence is needed before it can be assumed without question that gene reproduction is of the nature of duplication (but see Wright 1941).

It may be remarked here that various attempts have been made to estimate the size of the gene. These are estimates of the minimal size, based upon calculations of the number of genes which must be packed into a given volume (Morgan 1922, Muller 1926, 1935, Gowen and Gay 1933). The number of genes in Drosophila as estimated from considerations of the probability of the recurrence of mutations, agrees fairly well with the number of bands on the salivary gland chromosomes—around five thousand. The size, while it gives the order of magnitude as that of a protein macromolecule, is not a datum of sufficient accuracy to be useful for discrimination between different hypotheses of gene structure.

As has already been noted, the reproduction of the chromosomes and their separation into daughter nuclei are independent although correlated processes. It seems clear now that there is a general mechanism for the intranuclear reproduction of the chromosomes, in which all the stages of the mitotic process except the breakdown of the nuclear membrane and the distribution of the chromosomes into separate nuclei occur (see the review by Geitler, 1941). In some of these cases, usually found in differentiated highly active cells, where mitosis has been induced, the cells have been proved to be polyploid (for example, Berger 1941). In others, where mitosis does not occur, the high degree of polyploidy is inferred, either from volume relations, or from an increase in the number of heterochromatic chromosomes. In these endomitotic divisions, as Geitler has called them, the question may be raised whether all regions multiply at the same rate. Centromeres are evidently slower, and in the nurse cells of Drosophila it has been shown (Schultz, 1941c) that the heterochromatic regions even when removed from the centromere, lag behind the other regions. Thus although usually chromosome reproduction is adjusted so that everything comes out even, under special circumstances differences in the reproductive rates of the genes may be evident.

The origin of the giant salivary gland chromosomes has, as might be expected, been much discussed in these connections; the discussion was the stimulus for much of the development of the study of endomitosis. The original suggestion of Koltzoff (1934) and Bridges (1935) that these originate by repetitive splitting of the chromonemata without their separation was modified (Bridges 1938) to cover an increase in length over the despiralization originally postulated. Examination of the development of these chromosomes (Painter and Griffen 1937, Geitler 1938, Frolowa 1938) has shown them to be a hypertrophied four stranded structure, presumably arising by a process of growth within the original sheaths of the four strands known to be present at the final mitosis which these cells undergo prior to their hypertrophy. Whether simple splitting (duplication) or some more continuous process of synthesis at a surface occurs here is not yet clear (cf. the discussion of Painter 1941b, and of Metz 1941).

In endomitotic divisions both the chromomeres and the interchromomeric substances reproduce, and as far as can be judged, although some reserve must be maintained here, even in the more slowly reproducing heterochromatic regions, the two components keep in step. But there is another type of giant chromosome, the lampbrush chromosome so-called, found in the maturing eggs of certain fish and amphibia. These are elongate threads, quadripartite as synapsed homologues should be, with highly individualized brush-like loops (whence the name) extending out at the site of the chromomeres. Painter and Taylor (1942) have recently reported the migration of chromomeric (Feulgen-positive, see below) granules into the cytoplasm with their nucleolar products which will interest us later. One may interpret these findings as evidence for Belling's hypothesis of gene reproduction; for here evidently the chromomeres are reproducing without forming the connecting links. Their linear continuity is not maintained, and migrating as they do they are directly of use in cytoplasmic syntheses. The original four strands are maintained, and at the later stages of maturation appear as the appropriately small chromosomes of the

maturation spindle. Thus the intergenic material is restricted to the amount necessary to continue the chromosome structure into the next mitosis. The application, therefore, of Belling's dualistic hypothesis of chromosome reproduction makes intelligible the otherwise puzzling behavior of these chromosomes, interpreted by Painter and Taylor as a quite special case of "diffuse heterochromatin."

Thus the only scheme that covers these cytological data without special assumptions at present is Belling's hypothesis. It is too infrequently recognized that Belling himself based his case upon direct observation of the early stages of meiosis in the lilies: work which has not been repeated since, owing to the difficulties of observation at these stages. The clearest evidence would of course come from *in vivo* observations, and it is possible that these may be carried out using some of the more modern microscopic techniques.

Genetic techniques for the study of gene reproduction depend upon the labelling of one of the products of reproduction: the daughter cells must differ genetically, for a discrimination between mother and daughter to be possible. The frequently mutating loci in plants give numerous examples in which there are mutant spots; the plants are variegated due to a high mutation rate in the development. Demerec pointed out, in a study of the rose α gene in Delphinium, that the existence of single celled purple spots meant that the mutant change had occurred in one of the two daughter chromosomes. One might therefore infer that the mutation was associated with an error in gene reproduction—a false copy, as Haldane has called it. It is difficult to draw conclusions here, since the nature of the mutation process, attractive though the concept is, may be independent of the reproduction. There are two unknowns.

Analogous difficulties are encountered in the use of the "fractional" mutants among the radiation treatment progenies. For example, of the x-ray mutants induced in Drosophila sperm, one out of approximately seven appears in a mosaic individual (see Patterson 1933, Muller 1940). It has been suggested that the chromosome is two stranded in the sperm head, and only one of the two strands is hit in the sperm giving rise to the mosaic individual. The other possibility is an effect on the gene reproduction, which would provide a situation in which the products of reproduction are different. In maize, similar mosaics occur after pollen irradiation, both with x-rays and, with a much higher frequency, with ultra-violet light (Stadler 1941a, b). These differences might occur from effects on a two stranded structure in the pollen nucleus, the ultra-violet being more likely than the x-rays to break only one of the two strands. Or, again, the effect might be on the chromomere, delaying its reproduction, causing changes thereby in one of the two daughter strands with a high frequency in the ultraviolet series. As in the case of the unstable genes however, there are too many variables at the present stage for a useful discussion until more data are available.

We have already discussed the variegational translocations in Drosophila in connection with the extended radius of action of the heterochromatin position effect. This is due to repetition of the effect upon gene after gene at successive mitoses, until genes as many as fifty bands away from the point of breakage may be affected. The study of the pattern of variegation, where genes are involved whose changes are detectable in the same tissue, shows that the cells showing the change in the most distant gene always show a change in the genes more proximal to the break. Thus the more extreme changes are the descendants of those less extreme.

The occurrence of single cell spots in these cases permits us to apply Demerec's argument, and relate the changes to changes in reproduction. Fortunately, it will be recalled, we have independent cytological evidence in endomitotic divisions, that heterochromatic regions in Drosophila do have a peculiarly slow reproductive rate. And better still, the cytological examination of the salivary gland chromosomes shows a series of changes correlated with the phenotypic variegation (Schultz 1936,

Caspersson and Schultz 1938, Prokofyeva-Belgovskaia 1939). These changes occur in both the chromomeres and the interchromomeric regions; but they are graded. The first change is in the chromomere which becomes heavy and as will be seen, more densely laden with nucleic acid; more extreme changes are the gradual assumption by these originally euchromatic regions of the characteristics of heterochromatin. In the most extreme cases, where phenotypically there is a complete loss of function by the genes involved, the chromosome section can no longer be distinguished as such: it is effectively lost (Schultz 1941a). Thus the position effect of the heterochromatin may be due to an imposition of its slower rate of reproduction upon its euchromatic neighbors.

It seems likely that further study of these cases will permit a test of whether the chromomere or the interchromomeric section is the prime mover. At present it seems that the first changes are detectable in the chromomeres; but this may only be because we can see changes in them more easily. The fact that the structural changes characteristic of heterochromatin are seen in the more extreme cases, and could have been detected in the interchromomeres, does, however, reinforce the impression that the chromomere is the basic unit, and the interchromomere secondary.

The study of gene reproduction is still in its early stages, as can be seen from this discussion. Methods and material are only beginning to be available which will permit that correlation of genetic and cytological techniques necessary to establish the nature of the basic reproductive unit of the chromosome. As has been intimated, the Belling hypothesis offers a more attractive mechanism for the reproduction of a series of linearly connected units of specificity than does the duplicating split of a chromonema. How durable this attraction will be remains to be seen.

"Intragenic" Mutations

We have seen that both chromomere and interchromomere are associated in maintaining the structure of the chromosome, from the evidence of chromosome rearrangement. It is of interest therefore to turn to the changes which the individual gene can undergo without being lost. Instead of the unit of structure, we now examine the unit of function the gene mutation.

These distinctions are, however, limited ones. The position effects already reviewed show the high degree of integration between the intimate structure of the chromosome and the units of which it is composed. And the variegational rearrangements just discussed give evidence for a relation between the synthesis of new genes and the specific function of these genes. For, as we have seen, the series of changes in reproductive rate of the rearranged genes parallels a series of changes in function. These changes are similar to those found in "intragenic" mutants at these loci; the results of the two types of change are difficult, and in certain cases impossible to distinguish.

Thus the discussion of techniques for discrimination between structural changes and intragenic mutations is a serious one. Unfortunately, as Stadler (1932) has emphasized, there are no certain criteria. Both in Drosophila and in Zea where the most detailed studies have been carried out, it appears that extraordinarily similar effects are produced by changes that do and by those that do not have visible cytological effects (Stadler, McClintock 1938, 1941, Muller 1935). Even in the cases where a variety of changes at the same locus are available, it is not certain that the intragenic changes can be distinguished from minute rearrangements. The hope is brightest in this direction, however (Stadler 1942).

Even where no cytological changes is visible, however, there is no guarantee that rearrangement has not occurred. For a time it appeared that since ultraviolet light produced few rearrangements and many mutants, a means for discrimination

might exist here (Stadler 1941, Muller 1941). But the recent analysis of ultraviolet induced mutants in Drosophila (Mackenzie and Muller 1940, Slizynski 1942) shows that these also contain small rearrangements. Since these are difficult to detect even their absence is not significant. And more to the point is the question whether any *a priori* distinction is possible between intragenic change and the results of the coöperative disturbance of the effects of neighboring genes in the chromosomes—the doublets in Drosophila studied by Lewis come to mind. As the limiting case in any event, the intragenic change is a structural rearrangement of atoms either within a molecule, or binding molecular groups in polymeric linkages. The essence of the question reduces again to that we have already considered, the nature of the linkages between genes.

A curious aspect of the discussion of the possibility that all mutations involve structural changes has been the tendency to infer that the existence of specific separable units in the chromosome is incompatible with such a datum. We are not however as has already been apparent, faced with the paradox of Zeno; for the line in which the genes are arranged does have a local differentiation related to the specificities. And we have seen that there is a mechanism for getting a cohesive fiber out of a series of relatively independent units. Thus the fact that these units are functionally integrated, and that changes in these integrations may be the cause of mutation, does not imperil their existence as units. It should be emphasized however that it is by no means shown that intragenic mutations do not exist: it is only that there is no positive criterion for their identification.

It is evident that the study of the mutation process itself contains too many unknowns to serve, except very indirectly, as a means for the identification of the specific genic constituent of the chromosome, assuming for the moment that there is a dual system present. Since we cannot distinguish between the structural changes which involve both components of our dual system and the intragenic mutants which involve only the one, the production of mutants by radiation and the like gives information about stability but not much more. Only by a technique for the production of mutations in which it could be shown that the active agent was, for example, incorporated into a chromosomal component, does it seem likely that mutation study will give us the needed information.

In the meantime the study of the relative stabilities of different loci, and of different forms of the same locus (Timoféef-Ressovsky 1941, Muller 1941) has interest of its own. On various hypotheses concerning the nature of the gene or the intergenic bond, the rates of change from one allel to another might give useful information. The consideration of these stabilities has problems of a statistical nature. There is no *a priori* method of determining which should be the most stable condition of a particular gene, for this must depend upon the evolutionary history as well as the present state of the gene in question. This is partially the result of selection at the locus itself, by way of attainment of optimum functional efficiency; and it is also the result of selection of mutation rates in the genotype as a whole, a regulation of the overall production of the inviable mutant types in a population (see, for example Sturtevant 1937). It appears in Drosophila that the wild type allel is generally the least stable (Timoféef-Ressovsky, *loc. cit.*). The general picture is best interpreted both for the radiation-induced and the spontaneous mutants in terms of a disturbance of the gene followed by a return to stability, the most probable form assumed being the most stable discussion in (Schultz 1936b). There are departures from these rules in the cases of certain unstable genes, and a case of unique interest in Zea, in which instability of a recessive locus is conditioned by another locus in a different chromosome (Rhoades, 1941; and for a discussion of spontaneous mutation in maize see Stadler 1942).

In the disturbance of stability of a gene, there is one condition that must be fulfilled that is peculiar to intragenic mutations: the change must be of such a nature

as not to interfere with the capacity of the gene for self reproduction. This sets of course a limitation upon the possibilities for change, and increases the interest of the surveys that have been made both in maize and Drosophila of the types of mutational change that do occur "spontaneously." As has already been noted, each locus may exist in a variety of forms. The more delicate the tests for discriminating between allels, the greater the diversity of allels encountered (cf. Wright 1941, Stadler 1942, and for an earlier review Stern 1930). But the general rule appears to hold that all are concerned with the same physiological process, which in chemical terms means that a specific locus is concerned with a single type of synthesis.

Cases have been reported in which this rule appears to be contravened. In some of these it appears that nests of dissimilar genes are the answer. In other cases the answer is not so simple, and the conclusion that a single synthesis is affected seems more an article of faith. Thus the simultaneous effect at the white locus of Drosophila, upon the pigmentation of the eye and the shape of the seminal receptacle in the female certainly is a challenge to the discovery of some common denominator for these so apparently dissimilar characters (for a recent discussion see Dobzhansky and Holz 1943). Likewise, the connection between the intracellular effects observed in the losses of certain loci and the grosser morphological changes they occasion presents an interesting problem. Yet even in these cases it seems simpler for the present to regard the individual locus as concerned with a specific type of synthesis.

The implications of these analyses of multiple allels for the mutation study are evident. A mutational change at a locus still compatible with its continued self-reproduction is generally of such a nature that the locus continues in the same functional category. Genes do not apparently learn new jobs with facility; or if they do, the results are fatal to the organism.

Chemical Analysis: The Desoxyribonucleoproteins and Their Distribution in the Chromosomes

From the preceding, we see that the genetic and cytological analysis remove the question of the nature of the gene into the question as to which of the alternating elements of the chromosome contains the basic reproductive unit. The components are found to be so integrated that a disturbance in one results in an effect upon the other. Thus structural rearrangements give functional changes of genes; and the induction of intragenic mutations is hardly dissociable at present from an effect upon a bond between genes affecting the stability of the neighboring units.

The genetic technique, however, is not the only one available for the study of the gene even though it was necessary for the initial analysis. The chemical constitution of the chromosomes can be used to approach the problem by methods more closely allied to the conventional methods of biochemistry: the correlation of the properties of substances *in vivo* with their *in vitro* properties. As Mirsky has pointed out, the usual biochemical activity test for the gene would necessitate experiments in which an extract of chromosomes is used to make good the abnormal development in a genetically deficient individual; the potent extract coming from a non-deficient nucleus and that of a deficient nucleus proving ineffective. At the present time we are not in a position to carry out any such test. There is however a different line of evidence, provided by the unique occurrence in the nucleus of substances whose properties when correlated with those of similar substances in the cytoplasm, almost of themselves formulate rather specific hypotheses concerning the nature of the gene.

The nucleoprotein nature of the chromosomes has already entered these discussions: it is essentially responsible for their cytological staining properties. These have served to identify the components of the chromosomes associated with changes

of gene arrangement, and as we have seen the locus of a gene in the salivary gland chromosomes of Drosophila has been restricted to a band (chromomere) and its neighboring interband spaces. Fortunately as a result of the recent integration of chemical and cytochemical analysis, it is possible to speak with some confidence of the respective constitutions of these components of the chromosomes. They have been known for a long time, the chemistry of the nucleus having been one of the earliest growths in the histochemical field cultivated during the latter portion of the last century (see Gulick 1941, Mirsky 1943).

There are two main types of nucleoprotein, of which one is found only in the chromosomes. The distinction is most easily based upon the nature of the nucleic acid; recent data (Mirsky and Pollister 1942, 1943) support the inference that the proteins are characteristic as well. The desoxyribose nucleic acid of the chromosomes contains the unique pentose sugar, desoxyribose (Levene and Bass 1931), in contradistinction to the ribose found in the more widely distributed ribose nucleic acid. Associated with this difference is a difference in one of the component pyrimidines, the methylated form being found in the desoxyribose type. This nucleic acid, when prepared by gentle means, is an enormously long molecule, forming polymers of a molecular weight as high as a million (Signer, Caspersson and Hammarsten 1938, Hammarsten 1939, Pedersen 1940). In these long molecules the component nucleotides, phosphoric acid esters of purine or pyrimidine desoxyribose glycosides, are connected by their phosphoric acid groups. X-ray diffraction analysis of these molecules has been made, with the striking result that the distance between nucleotides corresponds almost exactly to the backbone spacing of the fully extended polypeptide chain of a protein (Astbury and Bell 1938).*

The proteins associated with the desoxyribose nucleic acids have constituted a challenge since the earliest work done with them by Miescher and Kossel. Even in their time the idea of the nucleus as a center of synthesis was prevalent, and to find the protein associated with a nuclear substance so absurdly simple as protamine was an extraordinary occurrence. Not only do they have a relatively low molecular weight but they lack completely the aromatic aminoacids which are the usual hallmark of protein specificity. Immunologically they show no reactivity unless coupled with other linkages. And yet in the sperm heads of teleost fishes, over ninety percent of the total substance present is nucleoprotein, with little evidence of other protein. It may be remarked parenthetically that a recent report describes another type of protein extract, without a clear demonstration that this does not contain nucleic acid (Stedman and Stedman 1943); until further evidence is presented it seems wise to hold to the earlier self-consistent account.

For the nucleus to contain the simplest of all proteins instead of a gallery of the most complex is a piquant situation. The genes are in the chromosomes, they control the course of development, but the substances important in the developmental sequence are complex, far more complex than the nucleoproteins analyzed. Then if these simpler substances play a part in the elaboration of the more complex ones, it must be as a pattern on which the precursors of the cytoplasmic complexities are moulded. These considerations led Kossel to the "protamine block" hypothesis of protein synthesis, according to which all proteins are built around a protamine core (see the more recent considerations of Edlbacher 1938, Block 1938). It will be seen presently that the process is more complex, and the story of protein synthesis has a leading character of nucleoprotein nature.

Quite recently, with the introduction of modern methods, the nucleoproteins of a large variety of nuclei have been extracted and studied (Mirsky and Pollister *loc. cit.*; see also Dounce 1943, Claude and Potter 1943). The proteins here are also basic proteins, the histones. More complex than the protamines, they contain

* See paper by W. T. Astbury in this Volume. J. A.

tyrosine, are larger molecular units and can form fibers. Of singular interest is the discovery of Mirsky and Pollister that these substances contain no tryptophane, although the different histones vary in other ways. In these experiments the nuclei and chromosomes have either been isolated mechanically, or a gentle salt extraction technique employed. The chromatin threads obtained mechanically are like the sperm head, over ninety percent nucleoprotein; moreover the substances removed by salt extraction agree in their properties with those obtained mechanically. In all these the nucleohistone is a protein nucleate, the two components being combined in an easily dissociable salt linkage.

We cannot as yet be sure that the nucleoproteins thus far isolated constitute the whole of the nucleoproteins of the nucleus. There is some indication (Dounce) that there may be some combined more closely, possibly with lipoids. For the question of the nature of the gene itself, the distribution of these nucleoproteins in the chromosomes themselves demands consideration.

An obvious attack is provided by the comparison of the results of extraction with the cytochemical analysis. We can correlate the localization of the substances with the functional differentiations of the chromosomes in so far as accurate chemical tests are available. The components of the chromosomes are especially favorable for cytochemical study. The strongly acidic character of nucleic acid is at the basis of the affinity of the chromosome for basic stains, and better still, Feulgen 1924 has developed a cytochemical test for the desoxyribose nucleic acids *in situ*. Most precise, in that it depends upon an intrinsic property of the substances is the analysis of the absorption spectra in the ultra-violet (Caspersson 1936, 1940c), where the nucleic acids have a striking absorption maximum at 2600 Å. This is due to their component purine and pyrimidine bases, and is quite distinct from the band at 2800 Å, present in proteins containing the aromatic amino acids. The proteins can also therefore be studied in this way, albeit with some difficulty due to their less intense absorption bands (Caspersson 1940a). They also combine with dyes, although here only a beginning has been made. Both the nucleic acids and the proteins can be digested by specific enzymes, and given the specificity of an enzyme, the breakdown of a structure by its activity can be used for the maintenance of that structure.

The major cytochemical work has been done so far with the especially favorable giant chromosomes of the salivary glands. Here it will be recalled that we are dealing with the chromosomes of a working nucleus, and it must therefore be expected that all the possible complexities will be encountered. We avoid the difficulties of the extractive techniques but now meet those occasioned by the presence of all the substances of interest at the same place, confusing possibly the accuracies of determination. Happily this is not quite the case; and for the analysis of the relation between gene and intergenic material, the chemical distinction between the differentiated portions of the chromosomes is easily made. The basic differentiation of cytology turns out to be a differentiation between segments of the chromosome relatively rich and poor in their nucleic acid content, thus a differentiation into protein and nucleoprotein segments.

The evidence for this derives from the comparison of the reactions of the segments in ultraviolet light, to the Feulgen reaction, on microincineration, to basic and acidic dyes, and to proteolytic enzymes and nucleases. It is found that the segments with a high ultraviolet absorption, shown quantitatively to be that of the nucleic acids (Caspersson 1936, 1940a) give the Feulgen reaction for the desoxyribose nucleic acids, combine with the basic dyes and show a high ash content on microincineration. The structure of these nucleic acid containing discs is somewhat disturbed on tryptic digestion, (Caspersson 1936), and nucleases remove the highly absorbing substances (Mazia and Jaeger, 1939; Mazia 1941; Frolowa 1941a).

Salt, alkali, and urea treatment (Calvin and Kodani 1941, Kodani 1942, Painter 1941b) also serve to break the linkages holding the nucleic acid to the protein component. Thus the differentiated discs must contain both nucleic acid and protein, and the intervening clear spaces contain chiefly protein.

The nature of these proteins has been the subject of study. There are apparently at least two distinguishable groups of protein present, which can be distinguished by their reaction to peptic digestion and to acid dyes. One of these, relatively resistant to pepsin, does not combine appreciably with such acid dyes as the triphenylmethane dye light green; this is the component that is easily digested by trypsin, and which forms the continuous, presumably skeletal component of the chromosomes. The other component, pepsin digestible, (Mazia 1941, Frolowa 1941b), forming combinations with the acid dye, is easily dissociated from its combination with the rest of the chromosome: a slight heat treatment before fixation often suffices. It occurs on both protein and nucleoprotein segments, and is especially heavy on the fainter nucleoprotein bands (Schultz 1941c).

The ultraviolet absorption spectrum of both types of protein is known from the studies of Caspersson (1940); as might be expected, both contain aromatic amino acids. Caspersson has carried his analysis farther, with results that at present do not quite accord with the determinations made by Mirsky and Pollister on their preparations of histones. The question at issue is the locus of the absorption maximum, deduced by Caspersson to lie somewhat shifted towards the long wave lengths in the more basic proteins. In the histones prepared by Mirsky and Pollister, direct measurements show the maximum at 2750 Å, somewhat shifted towards the shorter wave lengths as compared with the ordinary proteins. This is not surprising, in view of the absence in the histones of tryptophane, with its subsidiary absorption maximum at 2920 Å. It appears from the data as they stand that the absorption spectrum of the fibrous interband space could quite properly be considered that of a histone, as judged by the extraction data. The proper interpretation of the very real shift observed by Caspersson in the regions which contain aggregations of the sheath protein is still a matter of conjecture; one may suspect histones which have now incorporated tryptophane into the molecule (see below). In any event this protein is not the skeletal protein of the chromosomes and is rather to be regarded as a chromosomal product, of the nature of the nucleolus. To this we return later.

A direct comparison of extracted materials with cytochemical analyses in the salivary gland chromosomes is not as yet available. Treatment of these chromosomes with salt solutions similar to those used by Mirsky and Pollister, causes the dispersion of the nucleoprotein within the matrix of the chromosome. In what form they would be extracted, and how closely the extract would represent the combination actually present in the chromosome, can only be conjectured. But the comparisons already made between mechanically isolated chromosomes, and the salt extracts lead one to believe that the dissolution of the structure by the salt solution does not break up the original nucleoprotein of the chromosome.

Thus the chemical evidence leads us to the view that the chromosome is a protein fiber with nucleoprotein differentiation. This differentiation has dominated our cytogenetic discussion. In itself, almost as a trap for the unwary, it suggests the hypothesis that there is a differentiation into genic and intergenic material. As a caution against too much haste, the recent analysis of striated muscle should be recalled: here also the protein fiber has nucleoprotein striations (Caspersson and Thorell 1942) and the differentiation has presumably to do with the energetics of the contractile process.* Contraction occurs in the chromosome as well, and without the detailed cytogenetic evidence, the occurrence of this differentiation need not

* The electron microscope reveals a banded structure in collagen. See paper by A. F. Prebus in this Volume. J. A.

imply any relation to the genetic specificities. Since however the previous evidence, while not conclusive, was preponderantly favorable to the idea that the chromomere, or nucleoprotein, was actually the seat of specificity, the chemical analysis is most welcome.

If it is granted that there is an intergenic material, it should be noted that it is this, and not the gene, that must form the continuous fiber of the chromosome, the material responsible for the maintenance of structure. This consideration is quite independent of our previous discussions of Belling's dualistic hypothesis of chromosome reproduction; once the intergenic bonds are synthesized, they must form the continuous element. But if this is so, the results of the digestion experiments with enzymes prove that the intergenic material must be protein, and protein of the first type, characterized by its fibrous nature and relative simplicity. Interestingly, the absorption spectra of the interband spaces agree fairly well with the description of the ultraviolet absorption spectra of Mirsky and Pollister's histones. These are themselves capable of forming fibers, and it would seem not unlikely therefore (cf. Mazia 1941) that they are at the basis of the fibrous structure of the chromosomes.

In this way we obtain a picture of the structure of the giant salivary gland chromosomes. This can be extended by a consideration of the mitotic chromosomes and their changes during the mitotic cycle. Here the search is for the substances continuous through the cycle, which on that ground would have genetic value. It is evident as Astbury (1941) has pointed out, however, that there need be no basic substance, merely a set of conditions which serve to perpetuate the possibilities of synthesis. We are not reduced to this circumstance; while there are quantitative fluctuations nucleoproteins are always present in the chromosomes. These quantitative fluctuations are significant, for by their study correlations may be achieved with the changes that occur during the reproduction of the chromosomes.

Even in Miescher's work, the concept appears that during the ripening of the sperm there is a change in the nature of the proteins; he made comparisons of the isolated spermatocyte nuclei with those of the sperm head. Later Kossel and Schenk (1928) proposed that the more complex proteins were elaborated around a core of the simpler protamine, and during the ripening of the sperm there was a converse breakdown. This concept has recently been generalized by Caspersson (1940, 1941) and brought into line with cytological concepts. According to his point of view, at the close of every mitosis, when the chromosomes form the interphase nucleus, there is the elaboration of the specific proteins used in the function of that nucleus. At this time the synthesis of nucleoprotein is held in check; but with the onset of a new mitosis, prior to the visible splitting of the chromosomes, there is a synthesis of new nucleoprotein. The metaphase chromosomes contain, on this view, the substances of the nucleoprotein bands of the salivary gland chromosomes. The interband regions are regarded as telophasic products containing the more complex proteins. The metaphase chromosomes do not however have the properties of a coalesced nucleoprotein spiral. Microdissection experiments show them to be far too elastic: there must be elastic interband spaces in the metaphase chromosome also. Thus if there is, and it seems most likely that there is, a succession of changes in the type of protein present around the chromosomes, the alternation of nucleoprotein and protein segments remains permanent. The protein characteristics dominate the interphase chromosome, the nucleoprotein the mitotic chromosome.

From the evidence as it stands no conclusion can be drawn to aid our discrimination between the protein and the nucleoprotein as the primary unit of chromosome synthesis. Caspersson (1939) has followed the nucleic acid content of the chromosomes during the meiotic prophases of the grasshopper: the maximum synthesis occurs in the very early stages before the four strands of the synapsed homologues

are evident. However, we have no comparable data for the protein components. The importance of the point is considerable; for since, as we have seen, the intergenic substance must be protein on structural grounds, the demonstration of protein fibrous outgrowths from the chromomeres assumes crucial significance.

As an addendum to this discussion, the question of the capability of fiber formation by the interband proteins may be considered. Histones have, as has been said, this capacity, but the proteins do not. Now, of all the material examined, only the fish sperm contain the protamines; but we have no data for the interphase nuclei in these forms comparable to that available in the salivary gland chromosomes. Thus it may be that the structural integrity of the chromsomes in fish depends upon the nucleoprotein. Conversely, the possibility is not entirely excluded in the histone-containing forms that fine fibers of nucleoprotein extend from one chromomere to to the next, as a sort of residual template upon which the synthesis of histone took place. And if this nucleic acid continuity occurs at all, it may be basic to the chromosome structure in fish, holding the smaller protamine molecules in line.

The foregoing discussion has intentionally been restricted to the consideration of the chemical data as it is relevant to the identification of the genic component of the chromosomes. There are problems of intrinsic interest however in the chemical structure of the chromosomes. For in few places is the interrelation between the submicroscopic, molecular architecture and the visible structure of a biological system so obvious. The long rod molecules that are the nucleic acids, with their geometrical relations to the polypeptide chain, are manifest bases for the linear structure of the chromosome thread. The birefringence of chromosomes has been studied to gain evidence for the orientation of the molecules in the chromosome, and the role of the nucleic acids in this orientation has been much discussed (Schmidt 1937, Caspersson 1940b, 1941b). It has been tempting to consider the chromosomes as cables of polypeptide chains, with local differentiations imposed by the accumulation of basic groups to which the nucleic acids are attached in salt linkages (Koltzoff 1938, Wrinch 1936, Frey-Wissling 1938, and others). The intrinsic negative birefringence of the nucleic acids would then give rise to the observed birefringence of the sperm head for example, since they would serve to orient the protein in combination with them. Difficulties have arisen with these interpretations; the birefringence of the nucleoprotein bands of the salivary gland chromosomes is much lower than that of the sperm head, indeed its very existence has been called into question as an artefact (Caspersson 1940b).

Actually the consideration of the complex structure of the individual bands of the salivary gland chromosomes calls for caution in the interpretation of any observed orientation of molecules. Most of the bands which can easily be studied belong to the category of the doublets which have already been discussed. These are associated with each other as structures in which the pairing may be asymmetrical, and any intrinsic orientation may be masked. Another point that has been raised (Mazia 1941, Caspersson 1940b) concerns the degree of polymerization of the nucleoprotein in these giant chromosomes. The evidence indicates that lower polymers, which would have a lesser orienting effect, are present. This may be the result of various influences: the known effect of proteins on the depolymerization of thymonucleic acid (Greenstein 1941) could itself be enough, since the concentration of protein in the salivary gland chromosome is so high. The stricter orientation in sperm head and metaphase chromosome may be the result of the decrease in protein concentration in this type of chromosome. It seems not unlikely—a point of view that will be developed in detail elsewhere—that the orientation of the nucleoprotein molecules in the sperm head has nothing to do directly with the molecular architecture of the chromosomes, but is the expression of the long range forces found in tactoids. The sperm head may therefore be comparable to the micelles of the plant viruses studied by Bernal and Fankuchen (1941).

THE STABILITIES OF THE CHROMOSOMAL COMPONENTS AND THE NATURE OF THE MUTATION PROCESS

The chemical data of the preceding section have shown that the skeletal substance of the chromosome, the intergenic material if it exists as such, must be protein in nature. It does not immediately follow that the nucleoprotein discs are the genes. One might conceive a dual system in which the specific function of the nucleoprotein would be the maintenance of linear order, rather than the control of cellular synthesis. This does not seem likely, since we would now have to account for the remarkable variety of the nucleoprotein bands as the resultant of action by the protein segments. These would now have to be conceived as complex, and it is presumably to the sheath protein that one would appeal for this complexity. But this type of protein is also present in the nucleoprotein bands. Thus all the visible complexity and the specific variability called for are already present in the nucleoprotein segments. Thus as regards economy of assumption, the nucleoprotein hypothesis has its advantages.

A further test may be sought in a comparison of the conditions under which mutational changes occur genetically, with those inducing changes of composition or arrangement in the proteins as far as present information carries us. Current views consider changes in protein structure as due either to the differences in the folding of polypeptide chains, or to changes in the order and the variety of the constituent amino acids. Thus whether the gene is protein or nucleoprotein, a mutational change implies, as we have seen, that either the new type of folding or the new order or composition must be perpetuated. The recent studies of protein specificity, particularly those using the delicate serological techniques, have tended to emphasize the role of the folding of protein molecules in the establishment of the specific patterns (see Pauling 1940; cf. also Heidelberger 1943). But these are specificities of the second order, developed around the pattern of folding afforded by the hapten, assuming the correctness of the Pauling theory of antibody formation as an example. In a mutation however the original pattern is altered; and if we assume that the desoxyribose nucleoproteins are all like the clupein thymonucleate of Astbury and Bell 1938b, the protein in them is completely unfolded. Hence if a gene is nucleohistone in character, the pattern of folding will not be very helpful in providing mutations; changes in amino acid composition are needed. The interband regions however with their long range elasticity may well contain folded molecules, and may enter the picture here. If, however, as has frequently been suggested, gene reproduction is two-dimensional, taking place upon a surface, we must in any case look to properties of the unfolded molecule. But all these considerations are dependent on the metabolism of gene reproduction, which is not necessarily a simple surface reaction, but involves syntheses of the nucleic acids themselves, as we have already seen and will discuss further presently.

The agents that have so far proved successful in inducing changes in chromosomes or genes are chiefly different types of radiation, all of which produce changes in the folding of proteins and some of which change their composition. The effects in the ultraviolet region, to which reference has already been made, are of most interest to us. The chief ultraviolet sensitive component of the chromosomes, judging by the efficiency spectrum for genetic effects, is the nucleic acid. As we have seen, this in itself is only a criterion of relative stability. But it is of interest that chromoproteins, when irradiated, are denatured as the result of energy absorbed by the prosthetic group, which in the present case would correspond to the nucleic acid. Thus the absorption of energy by the nucleic acid moiety of the molecule would affect the protein also. But the mutated gene cannot be regarded as the denatured protein, if we are to take this analogy seriously; for the denatured chromoproteins are no longer in combination with their prosthetic groups. Thus a process of renaturation must take place (Mirsky 1941, for a recent discussion of reversible

denaturation) for the gene to maintain its nucleoprotein nature. And it is in this process of renaturation that the specific constitutional changes must occur that give rise to the changes in function by which the mutation is recognized.

It is evident that such an analogy requires more information than we have at present concerning the reversible denaturation of the nucleoproteins; information which is difficult to obtain due to the relative lack of tests of specificity in these proteins. One may easily conceive conditions under which the renatured protein would be not quite like its original state, either in folding, or even, with more difficulty, in composition. And on such a hypothesis the state of the protein when combined with the nucleic acid does not limit the stereochemistry of the mutation process.

Some support for the view that mutations may initially depend upon a process of denaturation is found in considering the effects of other agents upon mutation. The more usual chemical techniques for the irreversible denaturation of proteins—acid or alkaline treatment, urea, etc.—are in general too drastic to permit the survival of the treated cells. There are some data (Lobashov 1937) which indicate a slight effect of an alkaline intracellular pH (treatment of Drosophila with ammonia vapors) and not of acid (acetic acid fumes). These effects, although quite consistent with our speculations, are too uncertain to be considered a support. Similar slight effects have also been obtained with copper and iodine treatment (Law 1938, Sacharov 1938).

More interesting are the temperature effects (Muller 1942, Plough 1941, Timoféef-Ressovsky, Delbruck and Zimmer 1935). It has been shown that the effect of temperature on spontaneous mutation is considerable. Uncorrected for a difference in developmental rate, the Q_{10} is 2.9; with the appropriate correction for the shorter developmental rate at the higher temperature, Delbruck's calculation raises the figure to 5.7. This is a temperature coefficient which approaches fairly closely that found in the denaturation of proteins; since the mutation process is a local one, restricted to the individual locus, the correspondence is of some importance. A process analogous to protein denaturation in its energetics must be at the basis of mutation.

This said, we return to our question: denaturation of which component? And the answer is ambiguous here as it was in the purely genetic consideration of the problem. In the integrated linear series, a disturbance in either component may produce a genetic effect on the other. It appeared for a time that there might be a more specific answer forthcoming from a series of experiments (Gershenson) in which it was claimed that mutations were produced by the feeding of thymonucleic acid to Drosophila. Repetitions of these experiments have however yielded negative results (Muller 1941). The negative result is not surprising, since many questions of technique remain to be solved before the problems of mutation in relation to the metabolism of the chromosomes can be worked with effectively. For the present we may conclude that the stabilities of the chromosomal components probably correspond to the stabilities of the genes: not a contribution of decisive value, since these same stabilities are common to most biological systems.

Nuclear Function and the Nature of the Gene

Although the stabilities of the nucleoproteins are consistent with genic properties, they are not unique and so not sufficient for the analysis of the nature of the gene. But it is conceivable that the analysis of the function of the nucleoproteins in cellular metabolism might be instructive, particularly if it could be related to nuclear function. If we were to find nucleoproteins only in the chromosomes, hypotheses in which their role was considered as structural (Hammarsten 1924) or as "protective" for the active groups (Frey-Wyssling 1938) would be strengthened. But if it should turn out that nucleoproteins have roles of importance in cellular

syntheses other than those directly concerned with the chromosomes, if they can in other words be related to gene function, their relation to the gene would become closer, and together with the other evidence we have discussed, would make a strong case for the nucleoprotein nature of the gene. As it turns out, the nucleoproteins are now known to be intimately associated with the synthetic processes of the cell.

Here we are chiefly concerned with the second of the two great groups of the nucleic acids, the ribonucleic acids, and with the metabolism of the nucleus in the non-dividing, functional cell. For although we have considered the chromosomes of the interphase nuclei, we have not yet considered the composition and nature of their products in the interphase nucleus, the nucleoli. And for the consideration of the synthetic functions of the nucleus, these are of major importance. Since however the nucleoli turn out to be ribonucleoprotein in composition, our interest turns to these substances.

The ribonucleic acids differ from the desoxyribonucleic acids in the presence of *d*-ribose as the sugar, and uracil instead of thymine as one of the component pyrimidines. Their physical chemistry is only now beginning to be studied with the care and delicacy previously accorded the desoxyribose nucleic acids; some of the differences in capacity for the formation of high polymers are not so striking with improved methods of preparation (see Cohen and Stanley 1942; also Hammarsten, cited in Caspersson and Santesson 1943). Apparently, though perhaps to a lesser extent, the ribonucleic acid molecules form fibers as a result of their rod-like shape.

It was formerly believed that the ribonucleic acids were characteristic of plants, the desoxyribose type of animals. As we have seen, this is incorrect; all chromosomes contain desoxyribose nucleic acid. The error of interpretation was based on the source from which the preparations of the ribonucleic acids first came: the nuclei of yeasts are relatively small, and—the important point—the ribonucleic acids are the nucleic acids of the cytoplasm.

The evidence for this conclusion is largely cytochemical, although some separations for the analysis of cytoplasm have been carried out also (Behrens 1938). Since there is no specific cytochemical test for the ribonucleic acids, it is necessary to depend upon the concurrent use of a number of partial tests. Thus the ultraviolet absorption spectrum with a 2600 Å maximum, the absence of a Feulgen test, the presence of ash on microincineration (and more recently the development of a micro test for phosphorus (Norberg 1942)), solubility, reaction with basic dyes, all together permit a fair judgment of the presence of the ribonucleic acids. In addition, the enzyme ribonuclease, available in crystalline form (Kunitz 1940) splits the ribonucleic acids into products soluble in the usual histological fixatives. Thus the digestion followed by cytological examination has shown the disappearance of the basic stain and where it has been followed, (Brachet 1940, Schultz 1941c) of the ultraviolet absorption (Gersh). There is however still a certain ambiguity, since ribonuclease has been shown to act upon the chromosomes (Schultz) and there is evidence (Mazia, Schultz unpublished data) that the effect may be one on coagulated or fibrous proteins, causing them to go into solution once more. What the relations are here is still a matter for investigation. Nevertheless, the agreement of these various tests allows us to trace the ribonucleic acids with a moderate amount of surety.

In this way, tests have been made in a wide variety of cells, with the result that the ribose nucleoproteins seem to be as consistent accompaniments of cytoplasmic growth as the desoxyribose type is of the reproduction of the chromosomes. It appears from the comparisons that ribonucleoproteins are consistently present in considerable quantities in mitotic cells (Caspersson and Schultz 1939, 1940; Caspersson and Thorell 1941). The extreme case is to be found in the cytoplasm of the egg before the cleavage divisions have taken place: here a series of rapid mitoses must take place in a closed system such that the cytoplasm must supply quickly the

substances required for chromosome growth. In the sea urchin egg, Brachet (1937) has followed the changes in the ribonucleic acids during cleavage, and found that they decrease as the desoxyribose nucleic acid content rises. In the rather different systems in the meristematic tissues of plants (Caspersson and Schultz 1938) and the embryonic tissues of animals there is a similar ribonucleoprotein cytoplasm. But, and especially so in view of Brachet's data, a relation of the ribonucleoproteins to mitosis does not itself satisfy the requirement we have set up, of a relation of the nucleoproteins to syntheses of various types in the cytoplasm. They might simply be part of a reciprocal system for the multiplication of chromosomes, and have no necessary relation to the genic constituents. From the evidence already presented this would seem improbable, but conceivable. It can, however, be excluded.

Studies of cytoplasm in which other synthetic activities are occurring show that preliminary to the production of the particular synthetic product there is an accumulation of ribonucleic acids in the cytoplasm. For example, in the growth of a yeast cell it has been shown that the first activity when a nitrogen source is supplied, is the multiplication of the ribonucleoprotein system. Still better is the evidence in secretory cells. Here the formation of the protein secretory product, for example in the pancreas, is preceded by a synthesis of ribonucleoprotein in connection with nuclear activity (Caspersson, Landström-Hydén and Aquilonius 1941). Similar relations have been found in the studies of the nerve cell (Landström-Hydén 1942; but see Gersh and Bodian 1943). And other cases could be added. On this evidence, the conclusion seems fair that the ribonucleoproteins are concerned with protein synthesis at least, and very likely with synthesis in general. The exact *in vitro* implementation of these cytochemical facts has only begun: we know for example that many of the respiratory enzymes are nucleoprotein, of a special kind it is true, but dependent for their activity upon a nucleotide in combination with a protein. And some nucleoproteins have a phosphatase activity, of a kind which has already been suggested in other connections to have a use in the development of chemical theories of protein synthesis (Stern 1942).

Thus in a preliminary way the argument for the importance of the nucleoproteins in the cell can be based upon functional as well as structural grounds. It would be pleasant if the analysis of the intermediary metabolism of the development in gene differences—such studies as those of the anthocyanin pigments of plants, the growth factors and amino acid metabolism of Neurospora, the eye pigments of insects, to name a few representative cases—could be correlated with these discussions. Unfortunately, the analysis of these specific syntheses controlled by genes has not yet proceeded far enough: in none of these cases has the synthesis of the enzymes postulated to carry out the transformations observed been studied as yet. Thus they give us no present guide, since the primary chromosomal reactions must be prior to the synthesis of these enzymes; the enzymatic activities of isolated nuclei are rather restricted (Brachet 1939, Behrens 1939, Dounce 1943). But as has already been intimated, the primary chromosomal reactions are visible in the formation of ribonucleoprotein bodies, the nucleoli.

All of the early theories of nucleolar function associated these bodies in some way with the synthetic functions of the cell, either as reserves, by-products or precursors (cf. the review by Gates, Caspersson and Schultz 1940, Caspersson 1940, 1941). The cytological analysis shows that these bodies have their origin at special heterochromatic regions of the chromosomes (Heitz, McClintock, and others). Thus we have here the association, intriguing in view of the discussion of the cytoplasmic nucleic acids, between the massive desoxyribose nucleoprotein regions of the chromosomes, and the ribonucleoprotein products. This gives a functional basis for ascribing importance in the chromosomes to the nucleoproteins. For we have here the topographical evidence of a link between the nucleoproteins that are characteris-

tic of the chromosomes and those that are the characteristic cytoplasmic accompaniments of synthesis.

The topographical relation implies a reaction sequence of these related substances that is the converse of that found by Brachet in the cleavage of the sea urchin egg: thus nucleoprotein synthesis involves, as it should, both nucleus and cytoplasm. But by integrating this approach with the genetic technique, it is possible to show that the pattern of the cytoplasmic nucleoprotein synthesis is set by the nucleus. In the earlier literature there was much discussion of actual transfer of substances from nucleus to cytoplasm, the thought being that the substances of the chromosomes themselves were so transferred. The desoxyribose nucleoproteins are however restricted to the nucleus in all but exceptional cases. The mechanism is a different one, made evident in the accumulation of ribonucleic acids around the nuclear membrane (Caspersson and Schultz 1940, Landström-Hydén 1942, Caspersson 1940a) in cells where this synthesis is taking place. In some cases there is a visible migration of ribonucleoprotein nucleoli (Painter and Taylor 1942) associated with desoxyribose nucleic acid granules; in others the sequence is not so clear, and the interpretation (Caspersson) has been that precursors pass through the nuclear membrane forming the ribonucleic acids in the cytoplasm. The difference is probably the size of the particle migrating, since in the case described by Painter and Taylor there are many small nucleoli, whose origin has been ascribed to diffuse heterochromatin instead of the massive blocks that are the rule. What is implied is essentially the hypothesis that all nucleolar substance is the product of heterochromatin; and more than this, that the function of heterochromatin is the regulation and synthesis of the ribonucleoproteins of the cell.

The peculiarities of heterochromatin have already entered our discussion earlier, particularly in connection with gene reproduction. In the salivary gland nucleus of the Diptera, special "heterochromomeres" are associated with nucleolus formation. It should be noted here that their staining reactions and behavior in digestion experiments are not dissimilar to those of other chromomeres, except perhaps for a greater variability in nucleic acid content. They do differ in their mode of association; the different heterochromatic regions tend to associate with each other, being coalesced into a "chromocenter," in which the usual regularities of the interband regions of the chromosomes are not apparent. This apparently characteristic difference of synapsis may however be spurious, since rather similar pictures occur in the behavior of the tandem duplications already encountered. In fact an early hypothesis (Muller and Prokofyeva 1935) held that the heterochromatic regions themselves consisted of such duplications. Thus the indiscriminate synapsis would be the result not of a difference in principle between the behavior of the heterochromatic and the euchromatic regions, but rather of the extensive duplication of the former. It is not however easy to reach a decision at the moment, until we know more about the behavior of heterochromatin in chromosome reproduction. For with the slower reproductive rates shown by these regions in the nurse cell endomitoses already discussed, a consistent application of the hypothesis of interband prediction by the nucleoprotein bands leads to the inference that in a system like the salivary gland chromosome, in which separation does not take place and pairing is maintained, a difference in the interband structure might be one way in which the slower reproductive rate could manifest itself. Conversely, it might be argued that one reason for the slower reproductive rate is an inability of the heterochromatic regions to synthesize the fibrous interbands on time. Arguments have been raised however, on evolutionary grounds, for inferring that the gene contained in small duplications "degenerate" (Offerman 1935, Serebrovsky 1939, Muller)—meaning essentially that they lose their specific functions by mutation, and that the resulting mutant genes contain in the extreme case only the non-specific reproductive functions

common to all genes. Thus the heterochromatic regions would be considered as containing genes which, whatever their initial character, have now reached the lowest common denominator of the gene—self reproduction and the non-specific gene product. On this view the nucleoli, which for example are formed at all loci in the lampbrush chromosomes, need not necessarily be the products only of the heterochromatic regions. They may, as indeed McClintock has suggested, be formed all along the chromosomes, and be collected at the nucleolar organizer near the major heterochromatin. More information concerning the genetic properties of heterochromatin will help in clearing up these questions; their importance at the moment is in the clarity with which they present the immediate problems to be dissociated in the analysis of the function of the gene. If the heterochromatic regions and the nucleoli are really to be compared with gene and gene product, the composition and function of the nucleoli are data of first importance. Thus the effect of changes in heterochromatin—the genetic technique—upon the composition of the nucleoli becomes a useful approach.

We are not limited in our information concerning the nucleoli to the qualitative content of ribonucleic acids; the absorption spectra have given information concerning the protein as well. Moreover in the salivary gland chromosomes local aggregations of material similar to the nucleoli occur at many places in the chromosomes (minor heterochromatic regions presumably). Caspersson has compared the absorption spectra of these regions associated with heterochromatin, and finds them all to exhibit on analysis the characteristic shift of the protein band to the longer wave lengths that he has defined as indicative of the histones. Whether or not the comparison of these non-fibrous substances with Mirsky and Pollister's histones is correct, is questionable; they correspond in their properties (pepsin digestibility, staining, etc.) to the somewhat more acidic sheath proteins already discussed. It may be that they are histones into whose molecule tryptophane has been introduced; analysis of nucleolar protein by extractive methods would be desirable. The interest is considerable in the fact, which should be emphasized, that these absorption spectra occur characteristically in the presence of the ribonucleic acids. These may be proteins as distinctive in their combination with the ribonucleic acids as the histones and protamines are with the desoxyribonucleic acids.

In Drosophila, changes of the composition of the nucleoli have been shown to accompany changes in the proportion or the arrangement of the heterochromatic regions (Schultz, Caspersson and Aquilonius 1940). There are obvious changes of the relative amounts of nucleic acid and protein, and may even involve the composition of the protein making up the nucleoli. We have thus direct genetic evidence for changes in the genic function in heterochromatin affecting the ribonucleoprotein structure. There are as yet no analyses sufficiently detailed as to permit of an estimate of possible significant variations in the composition of the different minor aggregations of sheath material ("puffs" or minor nucleoli). Such analyses when they are forthcoming will provide direct evidence concerning specific differentiations in heterochromatin. Since the changes of composition in the nucleolus itself are so varied, the results should prove of interest. But already the contribution to our discussion is clear: the nucleoproteins important in cytoplasmic synthesis are shown, within the nucleus, to be influenced by genetic changes in the chromosomes.

The argument can be carried a step farther: having changed the conditions in the nucleus, is there evidence for a concomitant change in the cytoplasm, of the sort to be expected from the aggregations of ribonucleic acid around the nuclear membrane. Increases of heterochromatin in the nucleus should have an effect on the cytoplasmic nucleic acids and the processes they are concerned with. One case of the kind already exists: the heterochromatic Y chromosome, present as an additional chromosome in the female Drosophila, increases the concentration of cyto-

plasmic nucleic acids in the egg cytoplasm (Caspersson and Schultz 1938). The functional changes correlated with this cytoplasmic effect we have met before. They are the changes in gene reproduction effected by heterochromatic regions on their neighbors in translocations.

The severity of the effects in these variegational translocations is strongly affected by the heterochromatin present elsewhere in the nucleus: the addition of a Y chromosome induces, in many cases, a return to the normal condition both in the adult characters affected and in the correlated nucleoprotein changes of the salivary gland chromosomes (see Gowen and Gay 1934; Schultz 1936a, c; Dubinin and Heptner 1935; Prokofyeva-Belgovskaia 1939; Panshin 1938). Important for our present case is the discovery of Noujdin that a similar, though somewhat slighter effect can be mediated by the cytoplasm synthesized under the influence of the additional Y chromosome. Thus the presence of the additional cytoplasmic nucleic acid, or rather ribonucleoprotein, has an effect on the character of the embryonic mitoses in Drosophila: the effect of the Y chromosome in restoring to the normal condition gene reproduction near heterochromatin, is paralleled by its effect on the nucleoproteins, which, we must still infer following Brachet and the sea urchin, have their influence on the early cleavages. And it is their effects that are visible in the variegation suppression.

We have found, then, a system in which there is evidence of an interrelation between the ribonucleoproteins of the cytoplasm, the desoxyribose nucleoproteins of the nucleus and the degree of function of special genes at given loci. It will be remembered that the initial change was a change in the desoxyribose nucleoprotein on the translocated band. The analysis has not yet proceeded far enough to show whether sheath protein or skeletal protein of the chromosomes is primarily affected by this desoxyribose nucleoprotein change. The increase in nucleic acid is characteristic for non-functioning nuclei, and here obviously to the malfunctioning genes. An abnormality of sheath protein, an inhibition of the normal fibrous component which according to Belling's dualistic hypothesis must be synthesized by the nucleoprotein gene, would be to the point; similar to the discussion previously of the delayed reproduction of heterochromatin itself. The relation to the cytoplasmic nucleic acids is in accord with the idea of a reciprocity between cytoplasm and nucleus previously met in the general discussion of the function of the cytoplasmic nucleic acids. Here however, we have it influencing, by way of the reproduction of the genes, their specific function. In the variegated races, the initial effects are due presumably to the imposition upon the translocated genes of the slower rate of the heterochromatin; with the increased cytoplasmic ribonucleoprotein, the general rate is increased and the imposed delay avoided. It may be that similar relations are involved in the effects of supernumerary heterochromatic chromosomes upon polymitosis in the pollen grain of Sorghum, as Darlington and Thomas (1941) suggest, although as yet we are provided with no data concerning the cytoplasmic nucleoproteins in this case.

The considerations of the present section have been rewarding. Although the field is still in its infancy, the functional importance of the nucleoproteins in synthesis is demonstrated. If the nucleoproteins of the chromosomes are not the genes, they are so closely associated with them that they are not distinguishable by any simple means. But we have still to consider an argument of another type.

Nucleoproteins, Viruses and Genes *

We have preferred in our discussion thus far to avoid arguments from analogous self-reproducing systems; it seemed useful to determine whether the evidence of the chemical nature of the gene could be made convincing by the use of the data in the

* See "The Colloid Chemistry of Purified Viruses," by M. A. Lauffer and W. M. Stanley in this Volume. J. A.

field. But having reached a conclusion, it is of interest to test it by comparison with systems which, lacking the structural linear order of specificities in the chromosomes, nevertheless do have the property of self-reproduction. For the genes, the viruses provide the test case. And the demonstration that these are nucleoprotein in nature reassures us. In fact not only the viruses and genes but also the remaining self-reproducing element in biological systems, the plastids (Menke 1938) contain nucleoprotein. The basic reproductive units of living matter are the nucleoproteins.

Both genes and viruses, to return to the analogy, have the property of self-reproduction (Muller 1922). Similarly both mutate. These would appear to be a distinction in that genes influence specific syntheses in the cytoplasm. However the syndrome that accompanies the virus disease often mimics the effect of a gene mutation. The difference would appear to lie in the fact that the virus syndrome is due to the multiplication of the virus at the expense of the host; the mutant effect is due to the inability of the organism itself to provide the full complement of syntheses.

The point of importance in the present discussion is that these simplest self-reproducing units are composed of ribosenucleoprotein (Bawden and Pirie, Stanley 1941). Comparisons of the different strains show that the presumable mutations which result in their differentiation involved in some cases at least a change in their amino acid composition (Knight and Stanley 1941). We deal here with proteins more complex than the histones and protamines, more like (although at present more thoroughly analyzed than) the cytoplasmic ribonucleoproteins. Thus the parasitic self-reproducing bodies are ribonucleoprotein.

Apparently then the unique property of the genes from the present point of view, is that in addition to performing their own autosynthesis, they provide the substances important for other cellular syntheses, the ribonucleoproteins. Systems containing the desoxyribose nucleoproteins are the only ones which are not obligatory intracellular parasites. Thus the ribosenucleoprotein viruses may be regarded as competing in the cell for the products of nuclear synthesis. It is extraordinary to consider the simplest of all proteins setting the conditions for the production of the most complex. By this technique they have extended their environment: the simple nucleohistone has the capacity, by setting up hierarchies of synthesis, both to buffer itself against the outside world and to use it more efficiently. The primordial evolutionary step, one may guess, involved the invention of the reciprocal desoxyribose and ribonucleoprotein system.

References

Alexander, J. and Bridges, C. B., "Colloid Chemistry," Vol. 1, p. 9 (1928).
Astbury, W. T. and Bell, F. O., *Nature*, 141, 747 (1938).
—, —, *Cold Spring Harbor Symp. Quant. Biol.*, **8** (1938b).
—, Proc. 7th Int. Cong. Genetics, p. 49 (1941). See also paper by W. T. Astbury in this Volume.
Bauer, H., *Zool. Jahrb. (Physiol.)*, 56, 239 (1936).
—, *Chromosoma*, 1, 343 (1939).
—, Kaufmann, B. P., and Demerec, M., *Genetics*, 23, 610 (1938).
Behrens, M. *Z. physiol. chem.*, **253**, 185 (1938).
—, *Z. physiol. chem.*, **258**, 27 (1939).
Belling, J. *Univ. Calif. Publ. Botany*, **14**, 307 (1928).
—, *Genetics*, 18, 388 (1933).
Berger, C. A., *Cold Spring Harbor Symp. Quant. Biol*, **9** (1941).
Bernal, J. D., and Fankuchen, I., *J. Gen. Physiol.*, **25**, 111 (1941).
Biesele, J., Poyner, H., and Painter, T. S., Univ. Texas Publ. 4243 (1942).
Blakeslee, A. F., *Proc. 7th Int. Congr. Genetics*, 65 (1939).
Block, R. J., in "The Chemistry of Proteins and Amino Acids," C. L. A. Schmidt, editor, Springfield, Ill., C. C. Thomas Co.
Bozeman, M. L., *Genetics*, **28**, 71 (1943).
Bridges, C. B., *J. Hered.*, **26**, 60 (1935).
—, *J. Hered.*, **29**, 11 (1938).
—, *Science*, **83**, 210 (1936).
Brachet, J., *Arch. biol.*, **48**, 529 (1937).
—, *Arch. exp. Zellforsch.*, **22** (1939).
—, *Compt. Rend. Soc. Biol.*, **83**, 88 (1940).
Buck, J. B., *J. Hered.*, **33**, 3 (1942).
Callan, H. G., *Proc. Roy. Soc. London*, **B130**, 324 (1942).
Calvin, M., and Kodani, M., *Proc. Nat. Acad. Sci. U. S.*, **27**, 291 (1941).
Carlson, J. G., *Proc. Nat. Acad. Sci. U. S.*, **24**, 500 (1938).
—, *Cold Spring Harbor Symp. Quant. Biol.*, **9**, 104 (1941).
Catcheside, D. G., *J. Genetics*, **38**, 345 (1939).

Caspersson, T., *Skand. Arch. Physiol.*, **73**, Suppl. 8 (1936).
—, *Chromosoma*, **1**, 149 (1939).
—, *Chromosoma*, **1**, 562 (1940a).
—, *Chromosoma*, **1**, 605 (1940b).
—, *J. Roy. Microscop. Soc.*, **60**, 8 (1940c).
—, *Naturwiss.*, **29**, 33 (1941a).
—, *Chromosoma*, **2**, 247 (1941b).
—, and Brandt, K., *Protoplasma*, **35**, 507 (1941).
—, Landström-Hydén, H., and Aquilonius, L., *Chromosoma*, **2**, 111 (1941).
—, and Santesson, L., "Studies on Protein Metabolism in the Cells of Epithelial Tumors," P. A. Norstedt Sons, Stockholm, 1942.
—, and Schultz, J., *Nature*, **142**, 294 (1938).
—, —, *Nature*, **143**, 602 (1939).
—, —, *Proc. Nat. Acad. Sci. U. S.*, **26**, 507 (1940).
—, and Thorell, B., *Chromosoma*, **2**, 132 (1941).
—, —, *Acta Physiol. Scand.*, **4**, 97 (1942).
Claude, A. 1943, Biol. Symp. 10:111.
Claude, A., and Potter, J. S., *J. Exp. Med.*, **77**, 345 (1943).
Cohen, S. S., and Stanley, W. M., *J. Biol. Chem.*, **144**, 589 (1942).
Darlington, C. D., "Recent Advances in Cytology," P. Blakiston's Sons and Co., Philadelphia, 1937.
—, *J. Genetics*, **37**, 341 (1939).
—, *Biol. Rev. Cambridge Phil. Soc.*, **15** (3), 307 (1940).
—, *Nature*, **149**, 66 (1942).
—, and La Cour, L., *J. Genetics*, **40**, 185 (1940).
—, and Thomas, P. T., *Proc. Roy. Soc. London*, **B130**, 127 (1941).
Daneel, R., *Ergeb. biol.*, **18**, 55 (1941).
Demerec, M., *J. Genetics* (1929).
—, *Genetics*, **25**, 618 (1940).
—, *Proc. 7th Int. Genet. Cong.*, 99 (1941).
—, and Hoover, M. E., *Genetics*, **24**, 271 (1939).
Dobzhansky, Th., *Genetics*, **17**, 369 (1932).
—, *Biol. Reviews*, **11**, 364 (1936).
—, and Holz, A. M., *Genetics*, **28**, 295 (1943).
Dounce, A., *J. Biol. Chem.* (1943).
Dubinin, N. P., and Heptner, M. A., *J. Genetics*, **30**, 423 (1935).
—, and Volotov, E. N., *Bull. Biol. Med. Exp. U.R.S.S.*, **1**, 327 (1936).
Delbruck, M., *Cold Spring Harbor Symp. Quant. Biol.*, **9**, (1941).
Fano, U., *Ibid*, **9**, 113 (1941).
Feulgen, R, and Rossenbeck, H., *Z. Physiol. Chem.*, **135**, 203 (1924).
Frey-Wyssling, A., "Submicroscopische Morphologie des Protoplasmas und seiner Derivate," Protoplasma Monographs, **15**, 317 (1938).
Frolova, S. L., *Nature*, **141**, 1014 (1938).
—, *Compt. rend. acad. sci. U.R.S.S.*, **30**, 459 (1941).
—, *Ibid.*, **32**, 654 (1941).
Gates, R. R., *Bot. Rev.*, **8**, 337 (1942).
Geitler, L., "Chromosomenbau," Protoplasma Monographs, **14**, 22 (1938).
—, *Ergeb. Biologie*, **18**, 1 (1941).
Gersh, I. and Bodian, D. 1943, Biol. Symp. 10:163.
Giles, N. H., *Genetics*, **28**, 398 (1943).
Gowen, J. W., and Gay, E. H., *Genetics*, **18**, 1 (1933).
—, —, *Genetics*, **19**, 189 (1934).
Goldschmidt, R., "Physiological Genetics," McGraw-Hill Book Co., 1938.
Greenstein, J., and Jenrette, W., *Cold Spring Harbor Symp. Quant. Biol.*, **9**, (1941).
Gulick, A., *Bot. Rev.*, **7**, 433 (1941).
Haldane, J. B. S., "New Paths in Genetics," 1942.
Hammarsten, E., *Biochem. Zeitschr.*, **144**, 383 (1924).
—, *J. Mount Sinai Hosptl.*, **6**, 115 (1939).
Heidelberger, M., *Amer. Nat.*, (1943).
Heitz, E., *Planta*, **12**, 775 (1931).
—, *Z. indukt. abstamm. vererb.*, **70**, 402 (1935).
Hollaender, A., and Emmons, C. W., *Cold Spring Harbor Symp. Quant. Biol.*, **9**, 179 (1941).
Huskins, C. L., in "The Structure of Protoplasm," p. 109, Iowa State College Press, Ames, Iowa, 1942.
Janssens, F. A., *Cellule*, **25**, 387 (1909).
—, *Cellule*, **34**, 133 (1924).
Kaufmann, B. P., *Proc. Nat. Acad. Sci. U. S.*, **25**, 571 (1939).
—, *Cold Spring Harbor Symp. Quant. Biol.*, **9**, 82 (1941).
Knapp, E., and Schreiber, H., *Proc. 7th Int. Genetics Cong.*, 175 (1941).
Knight, C. A., and Stanley, W. M., *J. Biol. Chem.*, **141**, 39 (1941).
Kodani, M., *J. Hered.*, **33**, 115 (1942).
Koltzoff, N. K., *Science*, **80**, 312 (1934).
—, *Biol. Zhurn. (Moscow)*, **7**, 3 (1938).
Kossel, A., and Schenck, E., *Z. physiol. chem.*, **173**, 278 (1928).
—, "The Protamines and Histones," Longmans, Green & Co., New York, 1928.
Kunitz, M. 1940, J. Gen. Physiol. 24:15.
Kuwada, Y. 1940, Cytologia 10:213.
Landström-Hydén, H. 1942, Nordisk Medicin 13:144.
Lea, D. E. and Catcheside, D. G. 1943, J. Genetics 44:216.
Lewis, E. B. 1941 Proc. Nat. Acad. Sci. U. S.
—, 1942 Genetics 27:
Levene, P. A., and Bass, L. W., "Nucleic Acids," Chemical Catalog Co. New York, 1931.
Lobashov, M. E., *Genetica*, **19**, 200 (1937).
Mackenzie, K., and Muller, H. J., *Proc. Roy. Soc. London*, **B129**, 491 (1940).
Mazia, D. 1941, Cold Spring Harbor Symp. Quant. Biol. 9:40.
—, and Jaeger, L. 1939, Proc. Nat. Acad. Sci. U. S., 25:456.
McClintock, B. 1933, Z. zellf. 19:191.
—, 1934, Z. zellforsch. 21:294.
—, 1938a, Missouri Agri. Expt. Sta. Bull. 290:1.
—, 1938b, Genetics 23:315.
—, 1941a, Genetics 26:542.
—, 1941b, Cold Spring Harbor Symp. Quant. Biol. 9:72.

—, 1943, Proc. Nat. Acad. Sci. U. S.
Menke, W. 1938, Z. physiol. chem. 257:43.
Metz, C. W. 1941, Cold Spring Harbor Symp. Quant. Biol. 9
Miescher, Friedrich Die histochemischen und physiologischen Arbeiten. Leipzig 1897.
Mirsky, A. E. 1941, Cold Spring Harbor Symp. Quant. Biol. 9.
—, 1943, Advances in Enzymology 3:1.
—, and Pollister, A. W. 1942, Proc. Nat. Acad. Sci. U. S., 28:344.
—, and Pollister, A. W. 1943, Trans. N. Y. Acad. Sci. Ser. II, 5:190.
Morgan, T. H. 1911, J. Exp. Zool., 11:365.
—, 1926, Amer. Nat. 60:489.
—, 1922, Proc. Roy. Soc. (London) B, 94:162.
Muller, H. J. 1916, Amer. Nat. 50:193; 284; 350; 421.
—, 1923, Eugenics, Genetics and the Family, 1:106.
—, 1922, Amer. Nat., 56:32.
—, 1929, Proc. 4th Int. Congr. Plant Sci. 1:897.
—, 1932, Proc. 6th Int. Congr. Genet. 1:213.
—, 1935, J. Hered. 26:469.
—, 1935, Amer. Nat. 69:405.
—, 1940, J. Genet. 40:1.
—, 1941, Cold Spring Harbor Symp. Quant. Biol. 9.
—, 1942, Biol. Symp. 6.
—, and Prokofyeva-Belgovskaia, A. A. 1935, Compt. Rend. (Doklady) Acad. Sci. U.R.S.S. (N.S.) 1:658.
—, —, and Kossikov, K. V. 1936, Comp. Rend. Acad. Sci. (Doklady) (N.S.) 1:87.
—, —, 1935, Proc. Nat. Acad. Sci. U. S. 21:16.
Moewus, F. 1941, Ergeb. biol. 18:287.
Nebel, B. R. 1939, Bot. Rev. 5.563.
—, 1941, Cold Spring Harbor Symp. Quant. Biol. 9.
Newcombe, H. B. 1942, J. Genet. 43:237.
Norberg, B. 1942, Acta Physiol. Scand. 5, Suppl. 14.
Noujdin, N. I. 1936, Nature 137:319.
Offerman, C. A. 1936, J. Genet. 32:103.
Painter, T. S. 1934, Genetics 19:448.
—, 1939, Proc. 7th Int. Cong. Genetics :228.
—, 1941, Cold Spring Harbor Symp. Quant. Biol. 9.
—, and Griffen, A. B. 1937, Genetics 22:612.
—, and Taylor, A. N. 1942, Proc. Nat. Acad. Sci. U. S. 28:311.
Panshin, I. B. 1938, Biol. Zhurn. (Moscow) 7:837.
Pauling, L. 1940, J. Am. Chem. Soc. 62:2643. See also J. Alexander, *Protoplasma,* 1931.
Patterson, J. T. 1933, Genetics 18:32.
Plough, H. H., 1941 Cold Spring Harbor Symp. Quant. Biol. 9.
Pollister, A. W. 1939, Proc. Nat. Acad. Sci. U. S. 25:189.
Prokofyeva-Belgovskaia, A. A., 1939, Com. Rend. (Doklady) Acad. Sci. U.R.S.S. 22:270.
—, 1939, Bull. Acad. Sci. U.R.S.S. 362-370.
—, and Khvostova, V. V. 1939, Comp. Rend. Acad. Sci. U.R.S.S. 23:270.
Rappoport, J. A. 1940, J. Gen. Biol. (Russian) 1:235.
Rhoades, M. M. 1941, Cold Spring Harbor Symp. Quant. Biol. 9:138.
Sax, K. 1941, Cold Spring Harbor Symp. Quant. Biol. 9:93.
Serebrovsky, A. S. 1938, Comp. Rend. (Doklady) Acad. Sci. U.R.S.S. 19.77.
Schmidt, W. J. 1937, Die Doppelbrechung von Karyoplasma. Protoplasma Monographien 11 Berlin.
Schraded, F. 1936, Biol. Bull. 70.
Schultz, J. 1936a, quoted in Morgan, Bridges and Schultz, J., Yearbook Carnegie Inst. Wash., 35, 294.
—, 1936b. Biological Effects of Radiation, II:1209 McGraw-Hill Book Co.
—, 1936c, Proc. Nat. Acad. Sci. U. S. 22:27.
—, 1941a, Proc. 7th Int. Congr. Genetics :257.
—, 1941b, quoted in Morgan, T. H., Schultz, J. and Curry, V., Year Book Carnegie Inst. Wash. 40:282.
—, 1941c, Cold Spring Harbor Symp. Quant. Biol. 9.
—, Caspersson, T. and Aquilonius, L., 1940, Proc. Natl. Acad. Sci. U. S., 26:515.
Signer, R., Caspersson, T. and Hammarsten, E. 1938, Nature 141:122.
Slizynski, B. M. 1942, Proc. Roy. Soc. (Edinburgh) B 61:297.
Stadler, L. J. 1932, Proc. 6th Int. Genetics Congr. 1:274.
—, 1941a, Proc. 7th Int. Genetics Cong. :269.
—, 1941b, Cold Spring Harbor Symp. Quant. Biol. 9.
—, 1942, Spragg Memorial Lectures on Plant Breeding, Third Ser. :3.
—, and Uber, F. M. 1942, Genetics 27:84.
—, and Roman, H. 1943, Genetics 28:91.
Stanley, W. M. 1941, Cold Spring Harbor Symp. Quant. Biol. 9; see also paper by Lauffer and Stanley in this volume.
Stern, Curt, 1930, Multiple Allelie. Handbuch Vererbsw., 1, G. Gebr. Borntraeger, Berlin.
Stern, K. 1942, Growth Symposium.
Sturtevant, A. H. 1913, J. Exp. Zool. 14:43.
—, 1925, Genetics 10:117.
—, 1928, Genetics 13:401.
—, 1937, Quart. Rev. Biol. 12:464.
—, and Beadle, G. W. 1939, An Introduction to Genetics. W. B. Saunders Co., Philadelphia.
Svedberg, The, and Pedersen, K. O. 1940, The Ultracentrifuge, p. 443. Oxford University Press.
Swanson, C. P. 1942, Genetics 27:491.
Sutton, E. 1943, Genetics 28:
Timofeef-Ressovsky, N. W., Zimmer, K. G. and Delbruck, M. 1935, Nachr. Ges. Wiss Gottingen (Math. Phys. K., Biol) 1:234.
—, 1941, Proc. 7th Inter. Genet. Cong. :281.
Troland, L. T. 1917, Amer. Nat. 51:321.
Volotov, E. N. 1937, Biol. Zhurn. (Moscow) 6:907.
Wilson, E. B. 1928, The Cell in Development and Heredity, 3rd edit. Macmillan Co., New York.
Wilson, G. B. and Boothroyd, E. R., Canad. J. Research 19:400.
White, M. J. D. 1942, Cytology and Cell Physiology, p. 140. Oxford Univ. Press.
Wright, S. Physiol. Rev. 21:487.
Wrinch, D. 1936, Protoplasma 25.
—, 1941, Proc. 7th Int. Genet. Cong. :329.

Colloid Chemistry in Embryonic Development

L. G. BARTH

Department of Zoology, Columbia University, New York City

In a sense, of course, the colloid state is the basis of all living processes, since without it there can be no life. However, this is such a broad generalization that it is of little use, since in some *particular* processes, such as organization, the colloids may not be important. What we are trying to show here is that changes in colloidal units are not only important but fundamental in making organization of the egg possible.

Introduction

The development of an egg or of a mass of undifferentiated cells is a problem involving *quantitative* chemical and physical differences in the protoplasm, such as concentration, activity, rate of diffusion, etc. There is no indication from experimental work that *qualitatively* different chemical substances are responsible for the *early* stages of differentiation. Thus, generally speaking, all parts of an egg may be stimulated to form all structures, and therefore *all parts* must contain *all* the substances necessary for development. What, then, causes the parts of a developing system to form into different structures—into an organized embryo? Since homogeneous distribution of the contents of a system would render any development impossible, it must be to some extent heterogeneous; this heterogeneity must occur in the physical and chemical state of the substances in various parts of the developing system. In some parts a substance may be bound, so that it is not reactive, while in other parts it may be free. Or, a substance may be very finely dispersed and highly reactive, or very coarsely dispersed, and only slightly active.

An increase in chemical activity could be brought about by selective adsorption of prosthetic groups upon the large protein molecules forming an active catalyst area. Similarly, elution of prosthetic groups from a catalyst would reduce activity. This adsorption or elution would be controlled by the chemical milieu of the catalyst (Alexander, 1939). A shift of the prosthetic group from one protein to another would form an entirely new catalyst.

With parts differing only in rate of chemical activity, a developing system would become essentially a series of processes in steady state, involving *sources,* such as oxygen and substrates, and *sinks,* such as carbon dioxide and diffusible metabolites. The overall rate of these processes is governed by the activity of the enzymes which catalyze the various steps in the processes. Since these enzymes are proteins, or proteins plus prosthetic groups, the velocity of the reactions will depend upon their degree of dispersion and on the contact relations of enzyme and substrate. In the case of enzymes with prosthetic groups, the catalytic activity is also controlled by the degree of dissociation of the combined enzyme-prosthetic group, since it is only the combined product which exhibits activity.

Fertilization

The stimulus for development involves a change in the rate of chemical processes in the egg through some internal rearrangement. No special compounds have to be added to the system to produce this change, since the action of the sperm can be

duplicated by a large variety of physical and chemical agents. Therefore we must look for a release (or binding) of substances already present in the egg, or an increase (or decrease) in the chemical activity of substances through changes in degree of dispersion or in percentage of adsorption, or in the nature of catalytic areas. Release of substances could be brought about by removal of lipoids from the surface of particles, thus bringing them into contact with the aqueous phase of protoplasm. Changes in dispersion and adsorption could be effected by changes in pH, alteration of the salt balance, and by special organic compounds.

The various factors in the stimulation of the egg were crystallized recently by Heilbrunn (1940) into a general theory of stimulation. The essential factors involve (1) agents which release calcium ions from a bound form in the surface or cortex of the egg, and (2) the action of this increase in concentration of free calcium on the internal protoplasm.

That stimulation results in a release of calcium from the surface of the cell is shown by the formation of calcium oxalate crystals in the cell after treatment with various physical and chemical stimulating agents. Furthermore, an increase in the viscosity of interior protoplasm can be detected after stimulation, and this is consistent with the effect of increase in concentration of calcium. Finally, typical calcium-binding ions, such as citrate and oxalate, will prevent stimulation, presumably by precipitating the free calcium.

The origin of the free calcium upon stimulation is the cortex of the cell, which contains calcium in bound form. If this bound form of calcium is a lipo-protein complex, it becomes easy to explain why such a wide variety of agents act as stimulators. Any agent which acts upon either the protein fraction or the lipid fraction could release calcium and elicit a response, if the calcium is set free to act upon the internal protoplasm.

The reaction of the protoplasm of the egg after stimulation involves a number of changes which appear to be in part colloidal. In some cases a sudden increase in the rate of oxygen consumption occurs, indicating an activation of catalysts. According to recent findings (Ballentine, 1940), the changes in the rate of oxygen consumption are governed by change in the active concentration of the dehydrogenase of the egg. Such a change should be detectable optically, and indeed Runnstrom (1928) has observed a change in the degree of dispersion of the colloids in the cortex of the egg. More recently Moser (1940) has demonstrated that certain visible granules in the cortex of the egg dissolve (or at least become invisible) a few seconds after the sperm penetrates the egg.

The manner in which the normal stimulating agent in the sperm is introduced into the egg cell involves a number of colloidal phenomena. In the first place the egg liberates a substance, fertilizin (Lillie, 1919), which acts upon the surface of the sperm so that a reversible agglutination of the sperm heads occurs. Likewise, the sperm contains a substance, egg agglutinin, which produces a solvent action on the jelly and surface of the egg (Frank, 1939). These two substances are colloids, probably proteins, and will react with each other in a test-tube, with the result that both become inactivated. Thus a mixture of the two will agglutinate neither sperm nor eggs. The exact role of these substances is uncertain, but that they aid in fertilization is likely (Tyler, 1941).

Differentiation

The response of the egg after stimulation is a series of colloid chemical changes which lead to differentiation, or the formation of different structures from the parts of the egg. It seems quite clear now that the various parts of the egg do not contain any special chemical compounds which act as differentiating factors. Rather it becomes more and more certain that *all* parts of the egg contain *all* of the *basic* chemical constituents for *all* parts. The fact that all parts of the egg can be made

to form all structures must be the starting point of any attempt to interpret development on a colloidal basis. For colloids, by exhibiting a great range of dispersion and adsorption, can thus show a wide range of chemical activity, even when present in the same concentration throughout the egg. A respiratory enzyme present largely in the form of coarse granules would possess a rather low activity compared with the same amount of enzyme more highly dispersed. For example, the dehydrogenase activity of the fertilized sea-urchin egg is five times that of the unfertilized egg, yet there appears to be no change in the *quantity* of those compounds already present in the unfertilized egg, but only a change in their *activity*.

These changes in colloid state are important in other phenomena. The reducing activity of the egg is highest at the animal pole,* and lowest at the vegetal pole; the former normally develops into ectoderm while the latter forms endoderm and mesoderm. However, this difference is not caused by presence or absence of special substances, since the animal (ectoderm-forming) half of the egg can be converted into endoderm and mesoderm by treatment with lithium chloride. Similarly the vegetal half can form ectoderm. It is not reasonable to assume that lithium chloride forms some special substance necessary for the formation of mesoderm and endoderm in the animal half, but rather that it modifies the substances already present by altering their degree of dispersion or catalytic activity. The difference in reducing activity between the two halves of the egg clearly indicates a difference in chemical activity of the enzyme-substrate relationship in respiration. This difference could be caused by differences in the degree of dispersion of dehydrogenases, or in degree of dissociation of prosthetic groups from proteins, and this in turn could be modified by the action of lithium chloride.

In this connection Child (1941) has stressed in his theory of development the quantitative differences within the egg, which can be demonstrated by a variety of methods. These quantitative differences frequently are expressed as differences in the rate of metabolism which can be measured. Ultimately, of course, these give rise to differences in chemical constitution, which finally result in chemical differentiation of parts. Qualitative differences may also follow specific catalyst changes due to adsorption or elution of specific prosthetic groups.

The conversion of a mass of cells into various structures arranged in orderly fashion in space and time has been termed *organization,* and the factors responsible for the change are termed *organizers.* The organizer may be contained in a particular group of cells, and the structures then arrange themselves with respect to this locus. Since we encounter various levels of organization it is well to classify developing systems with regard to complexity of organization.

1. **Labile systems.** These groups of cells form a chemical system which is very susceptible to change. As a result, external agents of various sorts can act as organizers and bring about a complete rearrangement of structures. The coelenterates are striking examples of such a system. The egg of *Fucus* is another outstanding case where the center of organization can be completely changed by a variety of external agents (Whitaker, 1940).

2. **Semi-labile systems.** In these masses of cells the organizers are more stable, and external agents as a rule do not act as organizers. A more drastic action is necessary to change the organization in these forms. The sea-urchin egg is a good example.

* Eggs are for the most part spherical, and it has become conventional to distinguish two poles of a diameter as markers of one polarity of an egg. Thus the point at which the polar bodies of the egg form is termed the animal pole, while the opposite pole is the vegetal pole. The hemisphere from the equator of the egg to the animal pole is called the animal half; the remaining hemisphere the vegetal half. The protoplasm around the animal pole usually forms the outer layer of the embryo, the ectoderm, while that of the vegetal pole usually forms the mesoderm, giving rise to muscle and connective tissue and also the endoderm of the gut.

3. **Stable systems.** Some forms appear to be unaffected by external agents, which suggests that the organizers are very stable, and as yet no adequate means of reorganization has been found. Typical examples are the eggs of annelids and molluscs.

Labile systems. We shall use the stems of *Tubularia,* a marine coelenterate, to show the characteristics of this kind of organization. In the first place, we find that all parts of the system are capable of developing into an organized structure, so that no difference in chemical composition can account for organization proceeding at one locus as compared with no action at other loci. Yet, since quantitative differences in the rate of regeneration are observed along the length of the stem, there must be differences in the rate of chemical reactions which govern the rate of regeneration. This again leads us to the necessary assumption that the chemical compounds are present in different states of aggregation or adsorption resulting in differences in catalytic activity.

The second characteristic is the ease with which a structure is organized from a group of cells. For wherever a cut is made so that the cells are exposed to sea water, a hydranth will form. Now this exposure of the cells to sea water does two things: (1) it provides a better supply of oxygen, and (2) it allows excretions to be carried away faster. The effect of change in the supply of oxygen can be seen from studies showing the rate of regeneration as directly dependent upon oxygen tension. Thus it can be seen that the higher the oxygen tension the more rapidly regeneration or organization proceeds.

The effect of accumulation of excretions and carbon dioxide is to inhibit organization. Thus cells at a cut surface will tend to organize into a hydranth because of a relatively high oxygen/metabolite ratio. There is then a dual effect—a stimulation due to increased supply of oxygen, and a release of inhibition caused by a faster removal of metabolites.

The mechanism of the action of a relatively high oxygen/metabolite ratio upon cells at the cut surface is suggested by studies on the relation between oxygen tension, oxygen consumption and the rate of organization. First, the rate of oxygen consumption is highest at the apical end of the stem, and it is at this end where cells organize into a hydranth. Second, the oxygen tension determines the rate of oxygen consumption. Thus at the exposed surface of the stem, where the oxygen/metabolite ratio is highest, the oxygen consumption will also be highest. As we go deeper into the stem from the cut surface, the oxygen/metabolite ratio will be progressively lower and lower (which means that the oxygen consumption drops off), until we come to regions where it is too low for organization. Between this latter region and the cut surface there will be intermediate regions where various degrees of organization will be possible. It may be that these different degrees of organization are expressed as different structures, so that the mouth forms at the region of highest rate of oxygen consumption, the oral tentacles at a somewhat lower rate, the basal tentacles at a still lower rate and finally the base of the hydranth at the lowest rate.

Why these differences in organization due to differences in rate of oxygen consumption should arise follows from the study of dominance (Child, 1929). In studies of this phenomenon we find that when two groups of cells are competing with each other, only the higher metabolizing group is able to organize into a hydranth. This dominance of the more active group of cells over the less active appears to be a competition phenomenon, whereby more rapidly metabolizing cells remove materials from slower metabolizing parts. For, when the circulation between the two groups of cells is blocked, the slower-metabolizing group is able to organize into a hydranth. In this way the competition between two groups of cells can determine whether or not they are able to organize into structures (Barth, 1940).

This same phenomenon may be applied to the orderly arrangement of the parts of the hydranth. Given a sufficiently high oxygen/metabolite ratio for the organization of a hydranth, then, within this group of cells, part A will be metabolizing faster than part B. Therefore, B is not able to develop into the same structure as A, since A is removing materials from B. B then forms some other structure. Similarly, B is metabolizing faster than C, and so C under the competitive action of B forms a different structure. Generalizing, since the metabolism of $A>B>C>D$ . . . , the structures which develop will be such that $A \neq B \neq C \neq D$. . . . We have essentially a competition between the chemical reactions whose velocities are different.

Here the colloidal dispersion of enzymes operating in metabolism must play an important role for, as pointed out earlier, the rate of oxygen consumption varies along the stem. This could be explained if the activity of the enzymes simply decreased with time, since the younger regions respire at a higher rate than older regions. This difference in activity results in a gradient in metabolism such as outlined above, and forms a system which will organize into a hydranth. However, this difference in activity in enzymes is easily obliterated, since exposure to high oxygen/metabolite ratio can speed up metabolism locally and set a new locus for organization. We see that there exists a labile determination or organization in the stem, but that this labile organization is not so important as organization imposed by the oxygen/metabolite ratio. In other systems we will find that the internal organization is more fixed and that external factors do not change it so readily. However, in these semi-labile systems the internal organization appears to be the same as in the very labile system of *Tubularia*.

The egg of the marine alga, *Fucus*, presents a very labile system, since a rhizoid can form from any part of the cell. This again shows that the basic materials of the egg are uniformly distributed and that only the activity of these materials can be responsible for internal organization. The activity is very easily controlled by a variety of agents, so that temperature and pH gradients, ultraviolet light, and centrifugal force are able to direct the outgrowth of the rhizoid (Whitaker, 1940). These agents in some way increase the activity of an auxin, which causes elongation of the cellulose wall of the egg. However, it becomes difficult to explain the action of these diverse agents as directly affecting auxin, so that an indirect effect on the mechanism of auxin production becomes necessary. It will be noted that some of these agents, pH and ultraviolet light, would be expected to change the activity of enzymes producing auxin while others, like temperature, would have a direct effect on auxin activity.

Semi-labile systems. How does this concept of competing reactions apply to more complex systems, such as the egg of the sea urchin? As early as 1916, Child gave evidence to show that the factors bringing about organization were quantitative in nature. Later (Child, 1940) these ideas were confirmed and extended. Runnstrom (1933) similarly has advanced the theory of quantitative factors as responsible for differentiation. While the two ideas are not entirely in agreement, the essential part of both is (1) that no qualitative differences exist within the egg, and (2) that a quantitative difference of some kind governs organization.

These quantitative differences again must trace back to differences of activity and not to concentration of enzymes. For, in the whole egg, the animal portion is said to develop into ectoderm because of a predominance of animal factors, while the vegetal portion forms mesoderm and endoderm because of the presence of vegetal factors. However, the full complement of vegetal factors must be present in some state in the animal portion, because this latter may form a whole embryo under certain conditions. Similarly, the animal factors necessary for organization must be present in the vegetal portion, since it also can form an entire embryo. Obviously, then, the basic chemical composition of the two portions must be identical, although in normal development they form vastly different parts of the embryo. Since the

chemical composition is the same but the development is different, there remain only variations in colloidal or physiochemical state to explain the difference in behavior in the animal and vegetal portions of the egg. This could be a difference in the adsorption of free prosthetic groups of cellular catalysts.

What is the consequence of these quantitative differences? In the labile systems they were shown to be expressed as differences in rate of respiration. Child (1936) demonstrated that in echinoderm eggs the ability to reduce methylene blue is greater at the animal than at the vegetal pole. Since methylene blue is a hydrogen acceptor, it is a good measure of dehydrogenase activity, and therefore the evidence indicates that this activity is measurably higher in the animal portion of the egg. Given this difference in rate of reduction between the animal and vegetal parts of the egg, we can apply the concept of competing chemical reactions, as developed under labile systems, to the organization of the sea-urchin egg. This greater dehydrogenase activity of the animal portion during the formation of ectoderm removes from the sphere of action substances which may be ultilized by the vegetal portion. The lowered concentration of these substances in the vegetal portion then causes this part to form endoderm and mesoderm, possibly through selective adsorption, forming new catalysts. Experimental procedure shows that, when the animal portion is removed, the remaining vegetal portion forms an almost complete embryo. The isolated animal portion normally forms only ectoderm, although if treated with lithium chloride, which acts as a respiratory inhibitor (Lindahl, 1936), this isolated animal region may form an entire embryo (von Ubisch, 1925). Thus, by inhibiting the high dehydrogenase activity, the substances normally utilized in formation of ectoderm are diverted into endoderm and mesoderm.

This makes the comparison with labile systems almost complete. We have an internal organization which is basically a difference in the colloidal state of the dehydrogenase-substrate system. However, this system is semi-labile, to the extent that lithium chloride can change it. It is more fixed than a labile system, since the organization persists to some extent after the parts are separated. This is not true of the labile system, which becomes completely reorganized after separation by a changed oxygen/metabolite ratio.

What are some of the special characteristics of the semi-labile system? We have seen how lithium chloride acting on the isolated animal portion diverts some of the ectoderm into endoderm. Similarly, lithium chloride acting on the entire egg converts ectoderm into endoderm. On the other hand, potassium thiocyanate, applied before the egg is fertilized, produces embryos composed exclusively of ectoderm thus converting the normal endoderm into ectoderm. In this respect it is interesting to know that lithium and SCN are antagonistic with respect to hydration of proteins. Lithium decreases, while SCN increases hydration. Sulfate ions likewise decrease hydration, and when *absent* from sea water produce the same effect on the development of the egg as the SCN. Other substances altering organization would be expected to affect the colloidal structure of the proteins, for example, KCl and KI. These agents, by altering the activity of the dehydrogenase-substrate complex, could easily change the quantitative relationships between the parts of the egg and so change internal organization. The alteration could simply be an adsorption or elution of catalyst modifiers.

The amphibian organizer phenomena should be included in a discussion of semi-labile systems. For, while the ventral half of the egg (non-organizer) usually forms an unorganized mass of cells and the dorsal half (organizer) develops into a whole embryo, the ventral half may form an embryo if some of the organizer is added. Thus, as in the case of the sea urchin, the differences between the dorsal and ventral halves of the egg must be largely, if not completely, quantitative in nature.

If we limit ourselves to the organization of a neural tube, then certain similarities to previously discussed systems appear. In the first place, the organizing factors

must be relatively simple, for repeated attempts to find the naturally occurring stimulus for neural differentiation have led only to the discovery that several chemically unrelated compounds will stimulate: for example, certain sterols, phospholipids, fatty acids, acid and alkaline buffers, digitonin, crude protein extracts, and others. These diverse compounds must activate some internal organization in the ectoderm somewhat in the manner that agents are able to activate the egg in earlier development, *i.e.*, by modification of the catalysts through adsorption or elution. In normal development the roof of the archenteron supplies the stimulus, but, like the sperm cell in fertilization, the roof of the archenteron is not necessary. Apparently it simply creates a change in the ectoderm, releasing substances already present which form the basic structure of the neural tube.

Secondly, there appears to be some labile internal organization in the presumptive neural plate itself, such that it responds in a definite way to the stimulus of the organizer. Studies in progress are beginning to show that this internal organization is correlated with oxygen consumption. For the rate of oxygen consumption is highest at the animal pole and lowest at the vegetal pole. The presumptive neural plate is arranged with the brain at the animal pole and the spinal cord running vegetally. Thus it is fair to presume that the brain consumes oxygen at a higher rate than the spinal cord. When the stimulus (or organizer) is applied to this presumptive neural plate, the brain and spinal cord develop with respect to this internal organization, which may be based on differences in respiratory activity. This internal organization is semi-labile, as the part which becomes spinal cord will become brain if transplanted to the normal brain region. This is entirely consistent with the characteristics of labile and semi-labile systems.

Stable systems. These are characterized by the difficulty with which normal development can be altered. Cutting these systems and so exposing new surfaces rarely alters the internal organization, as contrasted with labile and semi-labile systems. New arrangements of the stable system likewise show that the parts are relatively unaltered and that the internal organization persists. Chemical treatment has little effect beyond injury.

In these forms it is not possible to decide whether qualitative differences comprise internal organization, so that particular chemical compounds determine the structures which various cells form; or whether the quantitative differences are so rigidly fixed to the structure of the egg that changes in concentration are difficult. In the first case the change of internal organization would be very difficult, since it would involve the synthesis of some new compound. However, if the latter case is true, more hope for changes in internal organization may be offered. For supposing that the actual concentration of enzyme-substrate was different in different parts of the cell, we might still hope to change this organization by means of inhibitors and stimulators of various sorts. These agents would be respiratory poisons and catalysts, or even salt solutions, as in the case of semi-labile systems.

In this connection it should be remembered that no system is absolutely stable, and recent years have modified greatly our concept of some of the classical stable systems. Their cells have been demonstrated to undergo changes in development (Rose, 1939) and in other cases two embryos have been obtained from one egg (Novikoff, 1939). Even the very early workers found that a small portion of the protoplasm of the egg was responsible for a large part of the embryo, indicating organizer phenomena.

The guiding principle throughout this discussion is that internal organization is largely delineated by quantitative differences in the activity of the enzymes of the system. The activity of these enzymes depends upon their colloidal state which is affected by external as well as internal factors. Salts, acids, and bases can bring about physical changes in the enzymes affecting their activities and changing internal organization of the egg. The sea urchin is a well studied example. Respiratory

catalysts and inhibitors increase or decrease the rate of reactions in which these enzymes serve as catalysts, thereby producing effects similar to the catalysts themselves and so changing internal organization.

Two consequences of these differing activities of enzymes are the competition of the parts of a system for substrates and the differential production of metabolites. In some forms it is clear that this competition phenomenon enables the low-metabolizing parts of a system to form certain structures, but not others. This gives rise to organization of the parts of a system into an orderly whole, with the control of organization residing in the part exhibiting the greatest over-all velocity of chemical reactions. For, once such a region is established, the competition phenomena will set up the regions exhibiting lower reaction rates.

The Maintainance of Differentiation

The foregoing has dealt with the initial steps in the process of differentiation and organization. However, the process very early must reach a stage at which it becomes relatively irreversible, *i.e.*, epidermal cells continue to produce epidermal cells even when placed in the environment of other types of cells. To explain this inheritance in the different cell types we must resort to a hypothesis such as that of Alexander and Bridges (1928) and of Alexander, (1931, 1932, 1937, 1939). The reproducible units are considered to be similar to viruses in that they catalyze their own formation. These autocatalytic catalysts are termed moleculobionts, and are to be sought for in the proteins of the cell. They vary from one type of cell to another, but arise in some way from the original egg proteins.

A possible mode of origin of these different catalysts from the original catalysts in the egg is by catalyst modification (Alexander, 1931, 1932, 1937, 1939), which may be brought about by a variety of means, *e.g.*, change in chemical structure, adsorption, or elution. In view of the foregoing discussion the following suggestion is made. We begin with a simple difference in the activity of the catalysts in different parts of the egg. This leads to a modification of the chemical milieu of those parts of the egg, caused by (1) changes in the concentration of metabolites brought about by differences in the rate of metabolism, and (2) changes in the concentration of substances which are used in metabolism. These concentration differences alter the catalysts present by selective adsorption of different molecules in various parts of the developing egg. Thus a number of new catalysts arise, which change the structure and activity of the cells, and result in what we call differentiation. Since these are auto-catalytic catalysts, they continue to build up, and daughter cells are formed which are similar to the parent cells.

The indications are that, once the primary cells are modified by selective adsorption, they become rather stable and it is not easy to change the cell type thereafter. However, such a change is simply a chemical problem of how further to modify the catalysts by adsorption of still more new molecules, or by elution of those molecules which have become adsorbed. It is quite possible that cell types can be changed back and forth with ease when the appropriate treatment is found. Our discussion would point to a treatment by substances which cause elution of adsorbed molecules, so that the modified catalyst could return to its original form, which then might be modified in a different way by another milieu.

Rate of Metabolism and Organization

Certain precautions must be taken in interpreting experiments dealing with the relation between the rate of metabolism and organization. In the first place, not all metabolism is concerned with organization. For example, respiration may be changed without affecting organization. Thus there is an active and a resting metabolism with respect to organization. Indications are that the active metabolism in some cases is only a fraction, say 10 per cent, of the total metabolism. In these

cases one must find which part of the total metabolism is concerned with organization. This can be done by applying to the system a wide variety of respiratory substrates, poisons and catalysts, and determining which part of the process is changed simultaneously with the organization.

In the second place, it is not the absolute rate of metabolism which is important, but rather the relative rate, or better the ratio between the rates of the different parts of the system. Thus temperature changes which may double or halve the absolute rate have little (however distinct and significant) effect on the internal organization. This again shows that the significant part of internal organization is the competition (either for substrates or by inhibitory action of metabolites) between regions of varying rates of metabolism. Changes in temperature bring about minor but very significant differences in organization, such as change in the relative size or number of structures.

Bibliography

Alexander, J., *Protoplasma*, 1931, *Biodynamica*, **54**, 1 (1939); "Colloid Chemistry," 4th Ed., D. Van Nostrand Co., New York, 1937; also paper in this volume, "Catalysis as a Biological Factor."
Alexander, J., and Bridges, C. B., "Colloid Chemistry," Vol. II., Chem. Cat. Co., (Reinhold Publishing Corp.), New York, 1928.
Barth, L. G., *Biol. Reviews*, **15**, 405 (1940).
Ballentine, R., *J. Cell Comp. Physiol.*, **15**, 217 (1940).
Child, C. M., "Patterns and Problems in Development," Univ. of Chicago Press, 1941.
Child, C. M., *Physiol. Zool.* **13**, 34 (1940).
Child, C. M., *Arch. Entw. mechan.*, **135**, 426 (1936).
Child, C. M., *Ibid.*, **117**, 21 (1929).
Child, C. M., *J. Morphol.*, **28** (1916).
Frank, J. A., *Biol. Bull.*, **76**, 190 (1939).
Heilbrunn, L. V., Publications American Association for the Advancement of Science, No. 14, p. **188** (1940).
Lillie, F. R., "Problems of Fertilization," Univ. of Chicago Press (1919).
Lindahl, P. E., *Acta Zool.*, **17**, 179 (1936).
Moser, F., *J. Exper. Zool.*, **80**, 447 (1940).
Novikoff, A., *Biol. Bull.*, **77** (1939).
Rose, S. M., *Biol. Bull.*, **77**, 216 (1939).
Runnstrom, J., *Arch. Entw. mechan.*, **129** (1933).
Runnstrom, J., *Acta Zool.*, **9**, 445 (1928).
Tyler, A., *Biol. Bull.*, **81**, 190 (1941).
Ubisch, L. von, *Zeitschr. Wiss. Zool.*, **124** (1925).
Wittaker, D. M., "Growth Supplement," **2** (1940).

The Physical States of Protoplasm Compatible with Life

B. J. Luyet

St. Louis University, St. Louis, Missouri

Physical Principles Concerning the Vitreous State, Vitrification and Devitrification

Relation between Velocity of Crystallization and Temperature. Crystallization is one of the many phenomena of nature which are governed by different laws at macroscopic and at molecular levels. The velocity of crystallization of a relatively large quantity of liquid *increases* at decreasing temperatures, while the velocity of growth of individual crystals *decreases* at decreasing temperatures (below the freezing point). The second part of this statement has been established by Tammann,[1] who demonstrated that the relation between the velocity of crystal growth and the temperature at the surface of a growing crystal—as opposed to the temperature in the bulk of the liquid—is of the type represented by the curve CG in Fig. 1. The difference between the macroscopic and the molecular phenomena is due to the fact

that, when crystals begin to form in a liquid, heat is generated and the temperature rises to the melting point, thus causing the growth of crystals to stop. Before a new growth period can start, the heat produced must be eliminated. So the factor which controls the velocity of bulk freezing is not the rapidity of crystal growth, but the rapidity of heat elimination. Consequently, low temperatures accelerate bulk freezing and retard crystal growth.

Relation between the Velocity of Formation of Crystalline Nuclei and Temperature. In the same series of studies, Tammann investigated the rate of formation of crystalline nuclei by counting the number of crystals which appear independently at several points in a supercooled liquid, during a given length of time, at given temperatures. He obtained curves of the type represented in Fig. 1, under NF. In these curves the range of temperatures extending from the freezing point downward consists of three zones: (1) a zone A where the velocity of formation of the nuclei is infinitely small, (2) a zone B where the velocity increases to a maximum and decreases again, (3) a zone C where the velocity once more reaches zero.

Various Types of Solidification. The resultant action of temperature on the two phenomena of crystal formation and crystal growth is as follows:

(a) In the zone A (Fig. 1) where the number of nuclei is low and crystal growth rapid, only a few, but large crystals will result.

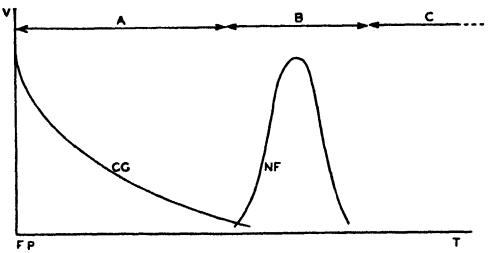


FIGURE 1. Graph representing the velocity of crystal growth (curve CG) and of the formation of crystalline nuclei (curve NF) in terms of decreasing temperatures; V, velocity; T, temperature; F. P., freezing point. (Curves drawn from Tammann's [1] data.)

(b) In the zone B, where the number of nuclei is high and growth velocity low, there will be many small crystals.

(c) In the zone C, where no nuclei are formed, and where they could not grow if they were formed, there will evidently be no crystallization. The substance will usually solidify in an amorphous state.

Action of Solutes on the Rate of Crystallization of a Solvent. Callow [2] reported that 1 per cent gelatin in solution in water reduces the rate of crystallization of the solvent to about one-half its value and that 3 per cent gelatin reduces it 350 times. Tammann and Büchner [3] made similar observations on the effect of sodium chloride, sulfuric acid, glycerin, alcohol and sucrose on the velocity of freezing of water. The presence of solutes may thus be of considerable importance in any operation involving the crystallization velocity.

Vitrification. From what has been said it follows that, if a liquid is not permitted to stay in zone B, it will not crystallize. If it can be carried from the zone A to the zone C, across B, so rapidly that the crystalline nuclei have no time to form, it will assume the vitreous state.

One can facilitate the rapid transition across zone B in two ways: (1) by reducing the bulk of the material, thus increasing the rate of cooling, and (2) by adding a solute, thus lowering the crystallization velocity.

In fact one can easily vitrify small amounts of various concentrated aqueous solutions by smearing a drop on a thin, glass or mica support and immersing this

preparation in liquid air (for the details of the procedure, see Luyet [4] and Barnes and Matthews [5]).

The vitrification of pure water requires a much higher cooling velocity, which was obtained in the following manner. Drops of water falling from a pipette were caught between a metal abutter cooled in liquid air and placed beyond the falling drops and a metal projectile also cooled in liquid air and shot against the abutter by the propelling action of the spring of a toy pistol. The drops flattened between the two pieces of metal, solidified into pellicles, a few microns thick, which were opaque when observed between crossed nicols.

Barnes and Matthews [5] made an x-ray study of the structure of gelatin gels of 50 to 70 per cent concentrations solidified by the method of vitrification described above and found no trace of crystalline structure. But it is possible that in less concentrated gelatin gels, or in various other gels or solutions, or mostly in pure water as treated in our experiments, minute crystals, either completely formed and ready to grow larger, or in some incipient stage of formation, may be detected by x-ray analysis.

Devitrification (Crystallization at Rising Temperatures). When a substance in the vitreous state is warmed from the zone C to the zone B it "devitrifies," that is, it crystallizes. Barnes and Matthews (*loc. cit.*) obtained the x-ray diffraction pattern of ice when they allowed gelatin gels of 50 to 63 per cent concentration, previously vitrified in liquid air, to warm up "until the temperature was just below the melting point" and then lowered the temperature again to its original value.

We determined the devitrification temperatures of several aqueous solutions. The procedure consisted in vitrifying by immersion in liquid air a drop of the solution to be investigated, placed between two thin glass plates, and observing at gradually rising temperatures when crystallization occurred (that is, when the preparation became opaque). It was found that, in general, the devitrification temperatures rose as the molecular weight of the solutes increased. The values observed for a series of more or less comparable compounds are as follows: formaldehyde: −80.5°, glucose: −40.6°, sucrose: −31.8°, dextrin: −9.4°C. (The reader will have a better idea of the degree of comparability of these results by consulting the original papers in which the effect of concentration is discussed and the duration of devitrification of the same amounts of different solutions is mentioned.)

As to the devitrification temperature of pure water, we have no information except that the pellicles of solid water obtained with the vitrifying gun, when allowed to warm gradually in a liquid bath, while continuously observed between crossed nicols, showed an intense reestablishment of light immediately before melting, that is, in the neighborhood of −1°C. Whether this temperature corresponds to a sudden increase in the crystallization velocity of individual crystals, to the incorporation of small crystals into larger ones, or to the formation of a new phase, we cannot yet say.

Crystallization at rising temperatures can be avoided by a rapid warming, in the same manner as crystallization at decreasing temperatures was avoided by a rapid cooling. Rapid warming is effected by immersion of the vitrified preparations in a warm liquid.

Changes of State Involving the Destruction of Life

When we use the term *life* in this section we do not mean actual vital activities such as respiration or assimilation, which are inhibited by the low temperature to which we must expose protoplasm to induce the changes of state to be studied. We mean the ability of an organism whose vital functions are inhibited, to resume them when it is brought back to normal temperatures. The problem is whether this ability, and the particular arrangement of the molecules of protoplasm that underlies it, are irreversibly destroyed when some of the molecules are subjected to the rearrangement involved in crystallization, vitrification or devitrification.

Crystallization. It seems quite well established that the *complete* crystallization of the freezable water of an organism in the vegetative state, with *full* water content, causes death. This was observed in most of the researches permitting an accurate determination of the internal temperature of the frozen organism, and of the quantity of ice formed in it. (For details see the review by Luyet and Gehenio,[6] mostly the sections concerning Amphibia and Fishes.)

It is also well known that organisms which support dehydration and which, when desiccated, have such a low water content that no ice can be formed in them (seeds, moss, nematodes, rotifers, tardigrades, brine shrimp eggs, etc.) are not injured by the action of low temperatures. The absence of injury where no ice can be formed confirms the view that ice is the injurious factor.

Vitrification. We then undertook to investigate whether the protoplasmic structure would also be destroyed by solidification in the vitreous state. We used for that purpose: (1) monocellular organisms: Amoebae, Myxamoebae, Euglenae, Paramecia, Colpoda, spermatozoa; (2) tissues of higher plants and animals: epidermis of onion, frog muscle fibers; and (3) entire organs or organisms: moss leaves, vinegar eels.

The procedure was as follows: The material was obtained in the smallest possible quantity so that the bulk effect was reduced to the minimum. Its water content was decreased by evaporation or by immersion for a short time in some hypertonic solution. The organisms or organs thus partly dehydrated were placed on thin glass or mica supports and immersed in liquid air. (With monocellular organisms we also used the method of immersing in liquid air films of the culture mounted on a loop of fine wire, or smears, on thin glass supports, of an oil emulsion of the culture, where the organisms were enclosed in droplets of the nutritive medium surrounded by the oil.) After a given length of time the material was withdrawn from liquid air and immersed rapidly in warm water. It was then examined for signs of vitality. The characters considered as criteria of life were *movement* in protozoa, spermatozoa and vinegar eels, *plasmolysis* and *deplasmolysis* in plant tissues, and *response to electric stimulation* in muscle fibers.

The amorphous character of the suddenly cooled cells was ascertained in the case of the epidermis of the onion at liquid air temperature, by observation between crossed nicols. As to the other materials we have no direct evidence of their physical state.

The vitality tests gave essentially the following results: (1) Plant epidermal cells maintained their semi-permeability if their water content had been reduced by plasmolysis in a 5 to 15 per cent solution of NaCl before immersion in liquid air. (2) The motility of frog's spermatozoa was partially restored if they had been previously treated with a 40 to 50 per cent solution of sucrose. (3) As many as 79 per cent of the vinegar eels survived the low-temperature treatment when this was preceded by immersion in a 30 per cent solution of ethylene glycol, while only 11 per cent survived without previous dehydration. All the worms were injured. (4) A large proportion of frog's muscle fibers (91 per cent) responded to electric stimulation, though they were considerably injured. (5) The semi-permeable membranes of the cells of moss leaves were in functioning condition, even if exposed to low temperatures with their full water content of 72 per cent. (6) Of the many protozoa treated, only a few myxamoebae survived.

As a rule, when the conditions for vitrification, namely, rapid cooling, rapid rewarming, small mass of material and low water content (the equivalent of the presence of a solute) are satisfied, some revival is possible. And, when none of the organisms treated survives, it is usually observed that one of the precautions for avoiding crystallization has been omitted. The presumption, therefore, is that the dangerous regions in the temperature scale are those at which crystallization occurs. As to the vitreous state, it appears safe for life; but the case of protozoa still remains to be explained.

Devitrification. Vinegar eels, which lend themselves so well to this kind of experiment because of their high degree of motility, were used in an attempt to determine, in a living organism, the range of injurious temperatures of devitrification. After being vitrified in liquid air the worms were transferred into a bath of isopentane, where they were maintained for 1, 5, 15 or 30 minutes at a constant temperature between −50° and 0°C. They were then warmed rapidly and examined. It was found that an exposure of the solidified worms to temperatures from −39° to −5° for 1 minute was always fatal. Exposures to temperatures from −43° to −50° for a period as long as 30 minutes did not prevent revival. When the temperature of exposure passed from −43° to −39° the mortality increased rapidly. These facts fit in well with the idea that in the neighborhood of −43° lies the lower limit of a zone of devitrification temperatures for some essential constituent of the body of these animals.

To *summarize*, there is a certain amount of evidence that solidification in the amorphous state does not kill protoplasm, but that crystallization, either from the liquid or from the vitreous state, is lethal.

References

1. Tammann, G., "The States of Aggregation," New York, 1925. (See p. 260, Fig. 120, for the curve of crystallization velocity and pp. 228-230, in particular Fig. 109, for curves of the velocity of nuclei formation.)
2. Callow, E. H., *Proc. Roy. Soc. London, A*, **108**, 307 (1925).
3. Tammann, G., and Buchner, A., *Z. anorg. allg. Chemie*, **222**, 12 (1935).
4. Luyet, B. J., *Biodynamica*, **1**, No. 29 (1937).
5. Barnes, W. H., and Matthews, F. W., *Biodynamica*, **2**, No. 49 (1939).
6. Luyet, B. J., and Gehenio, P. M., *Biodynamica*, **1**, No. 33, 74-83 (1938). See also Luyet, B., *Compt. Rend. Soc. Biol.*, **127**, 788 (1938).

Original Papers by the Author and his Collaborators:

Luyet, B. J., "The Vitrification of Organic Colloids and of Protoplasm," *Biodynamica*, **1**, No. 29, 1-14 (1937).

—, and Gehenio, P. M., "The Survival of Moss Vitrified in Liquid Air and Its Relation to Water Content," *Ibid.*, **2**, No. 42, 1-7 (1938).

—, and Thoennes, G., "The Survival of Plant Cells Immersed in Liquid Air," *Science*, **88**, 284-285 (1938).

—, —, "Démonstration des propriétés isotropiques de masses cellulaires vitrifiées à la température de l'air liquide," *Compt. Rend. Acad. Sci.*, **206**, 2002 (1938).

—, —, "La reviviscence de fibres musculaires vitrifiées dans l'air liquide," *Ibid.*, **207**, 1256 (1938).

—, and Hodapp, E. L., "Revival of Frog's Spermatozoa Vitrified in Liquid Air," *Proc. Soc. Exp. Biol. Med.*, **39**, 433-434 (1938).

—, "The Devitrification Temperatures of Solutions of a Carbohydrate Series," *J. Phys. Chem.*, **43**, 881-885 (1939).

—, "Vitrification of Water," *Physical Review*, **46**, 1244-1245 (1939).

—, "The Vitreous State of Matter and the Devitrification Temperatures," *in* "Temperature, its Measurement and Control in Science and Industry," pp. 420-424. Reinhold Publishing Corp., New York, 1940.

Thoennes, G., "Properties of Muscle Fibers Subjected to Vitrification by Extremely Rapid Cooling," *Biodyn.*, **3**, 145-156 (1940).

Luyet, B. J., and Hartung, M. C., "Survival of *Anguillula aceti* After Solidification in Liquid Air," *Ibid.*, **3**, 353-362 (1941).

—, "Death by Devitrification in the Nematode *Anguillula aceti*," *Ibid.*, **3**, 363-367 (1941).

Among the publications of other workers we shall mention, as concerning directly the problem investigated here, those of A. Goetz and S. Scott Goetz, on yeast: *Biodyn.*, **2**, No. 43, 1-8 (1938); *J. Appl. Physics*, **9**, 718-729 (1938); *Proc. Am. Phil. Soc.*, **79**, 361-368 (1938).

Review:

Luyet, B. J. and Gehenio, P. M., "Life and Death at Low Temperatures," pp. 1-341. "Biodynamica," Normandy, Missouri, 1940.

Some Physical Properties of Protoplasm

ROBERT CHAMBERS

Research Professor of Biology, New York University

In our present state of knowledge it is impossible to consider protoplasm otherwise than in living cells. Thus far protoplasm has never been extracted from cells without complete loss of its outstanding property of being alive. Therefore, studies on it must be done within the confines of living cells. This close relationship has caused many to regard a living cell and protoplasm as synonymous on the basis that protoplasm is an organized body of which the structural unit is the living cell. "One of the most striking characteristics of protoplasm is that mechanical crushing destroys it. This holds not only for the integrated cell but also for any viable fragment of protoplasm which may be spontaneously or experimentally separated from the cell. The implication has long been recognized by cytologists that the property of life depends upon a definite architecture of structural nature. Verworn and E. B. Wilson have defined protoplasm as a morphological concept. Hypothetical units of physiological structure have been variously termed micellae by Nägeli, plasomes by Wiesner, bioblasts by O. Hertwig, etc. An attempt to give a chemical significance to the ultimate unit of living matter was made by Pflüger (1875). Pflüger postulated a living protein (lebendiges Eiweiss) molecule with a constitution based on a cyanogen radical. Its stability and lability were supposed to depend upon interchanges of atom groups in the molecule, induced by the consumption of oxygen and the liberation of carbonic acid. "Much earlier than Pflüger is the almost forgotten account of Fletcher of the University of Edinburgh in 1835. In his 'Rudiments of Physiology' Fletcher attacked the old hypothesis of a vital spirit. As an alternative he suggested that the elements composing living matter are in a peculiar state of combination. The newer developments in the extraction of proteins, *e.g.*, by ultracentrifuging at very low temperatures, have made possible the isolation of certain viruses hitherto unobtainable by the usual chemical means. This may be a first step toward isolating the extremely unstable complex which at present can be termed only as protoplasmic protein.*

This paper deals entirely with some of the properties of protoplasm as they are found in the living cell. Hitherto, in most studies of this nature not involving the technique of micromanipulation, it has been customary to deal with the cell in its entirety or with masses of cells. The fact must be appreciated that a cell includes more than its protoplasm. Botanists have been much more aware of this than have animal cytologists largely because of the prominence in the plant cell of its extraneous coats. Hence, the use of the botanical term *protoplast* to indicate the more truly protoplasmic body lying within the cellulose and pectinaceous walls of the cell. Among animal cells the extraneous coats are far less evident and, generally, have not been taken into serious consideration. This has given rise to certain misconceptions regarding, for example, the antagonistic action of salts. Actually, the mutual antagonism of NaCl and $CaCl_2$ refers more particularly to the action of these two salts on the extraneous material which surrounds cells and which serves to bind them to-

* Chambers, R., "The Micromanipulation of Living Cells. The Cell and Protoplasm," Publ. 14, Am. Assn. Adv. Sci., The Science Press, Washington, D. C., pp. 28-29, 1940.

gether. When the extraneous coats are removed it has been found that the maintenance of the actual protoplasmic surface film or layer depends on the simultaneous presence of NaCl and KCl in definite, proportional concentrations irrespective of the presence of $CaCl_2$.

For this reason the subject matter of this paper first deals with the nature of the extraneous coats. Then it deals with the protoplasmic surface layer and finally with certain properties of the protoplasm within its surface layer or plasma membrane. The use of the term *plasma membrane* is of historic interest since it was first used to denote that part of a living cell upon which the viability of its internal protoplasm depends. In regard to the morphology of protoplasm it has been impossible, thus far, to consider the internal protoplasm without the presence of its investing plasma membrane or protoplasmic surface layer. The disintegration of the one has always accompanied the disintegration of the other. Can it be that the stability of the protein-like complex of protoplasm depends upon the presence of a plasma membrane the selective permeability of which preserves the proper environment within the cell?

Extraneous Coats of Living Cells

All living cells, without exception, at sometime have membranes or jelly-like coatings. These extraneous coats may be delicate or tough, thin or thick, soft or hard, elastic or rigid, viscous or plastic, proteinacous or pectinaceous. All coatings originate from the living cell. The important feature about extraneous coats is that each layer may be removed mechanically or by suitable chemical solvents so that only a final limiting surface remains. The remaining surface is the protoplasmic surface layer, a structure which cannot be mistreated without seriously damaging the cell, or removed without completely destroying the cell. In the absence of injury, cells which have been denuded of their extraneous coats remain viable.

Nearly all investigations dealing with surface phenomena of living cells have been made on cells which possessed some or all of their normal complement of extraneous coats. Only during recent years have serious attempts been made to investigate cells from which all extraneous coats were removed. The physical properties of the extraneous coats have been recently reviewed.[15]

The eggs of various Echinoderms, particularly those from sea-urchins, have provided useful material for the study of a variety of extraneous coats. Mature, unfertilized sea-urchin eggs possess two well-defined extraneous coats. The protoplasmic surface layer is bounded by a thin, delicate vitelline membrane. Outside the vitelline membrane is a soft, invisible layer of jelly.

On insemination, the vitelline membrane separates from the protoplasmic surface layer, elevates and becomes the fertilization membrane.* The latter, when fully developed, differs conspicuously from its precursors and possesses the greatest amount of mechanical strength and resistance to chemical action of all membranes one encounters on sea-urchin eggs. Several minutes after insemination, the hyaline layer is secreted by the zygote and this layer accumulates on the surface previously occupied by the vitelline membrane.

The jelly layer may be readily removed by repeated washing of the sea-urchin eggs in isosmotic NaCl solutions slightly diluted with sea-water. Mature sea-urchin eggs, with the jelly removed will, on insemination, develop fertilization membranes and cleave as well as untreated eggs. The jelly layer is a dispensable coat.

The vitelline membrane may be removed in several ways. The most commonly used method is to inseminate the eggs and then transfer such eggs to isosmotic KCl solutions within 1.5 minutes after insemination. The incipient fertilization membrane which originates from the vitelline membrane is dispersible in the presence of KCl. If such eggs are returned to sea-water, the development is apparently normal, except that no fertilization membrane is present.

* See article by E. E. Just in Vol. II of this series. J.A.

The tangential rigidity of the extraneous coats can be compared by employing the oil-coalescence method.[15] Under certain conditions, sea-urchin eggs will coalesce with oil drops applied to their surface. One of the conditions which determines the ease with which coalescence occurs is the mechanical resistance offered by an extraneous coat. The completely denuded egg shows the highest tendency to coalesce with oil drops.

Mature, unfertilized sea-urchin eggs when immersed in isosmotic KCl solutions show the same tendency to coalesce with oil drops as exhibited by similar eggs immersed in sea-water. One interprets this reaction to signify that KCl has no action on the vitelline membrane. On the other hand, unfertilized eggs immersed in isosmotic $CaCl_2$ solutions have a coalescence about 8 to 10 times higher than control eggs immersed either in sea-water or in isosmotic KCl solutions. Excess calcium ions tend to increase either the liquidity or brittleness of the vitelline membrane, either of which would decrease the effectiveness of the vitelline membrane as a physical barrier to coalescence. Calcium also appears to increase the adhesiveness of the vitelline membrane. An oil drop brought into contact with an unfertilized egg (jelly removed) immersed in Ca solutions and suddenly pulled away, will carry with it a small portion of the egg. This adhesiveness undoubtedly enhances the contact between the egg and oil drop, thereby increasing still further the tendency of such eggs to coalesce with oil drops.

Because of its remarkable rigidity, the fertilization membrane completely prevents coalescence. The fully formed hyaline layer likewise prevents coalescence. It can be mechanically removed with microneedles or by agitating the eggs shortly after it has formed when the membrane breaks and the eggs slip out.

A few minutes after fertilization there develops a hyaline layer several micra in thickness closely investing the surface of the egg cell. One of the outstanding properties of the hyaline layer is its instability in Ca-free media. Fertilized eggs immersed in isosmotic KCl solutions gradually lose their hyaline layers, so that after a period of 15 to 30 minutes in the Ca-free medium, the hyaline layer becomes completely dispersed. Such eggs, providing the fertilization membrane was previously removed, very readily coalesce with oil drops.

The coalescence data, as well as other observations, lead to the conclusions that sea-urchin eggs with fertilization membranes removed mechanically (by shaking) and kept in KCl solutions for at least 30 minutes have the minimum of extraneous coatings. Such cells may be considered as being essentially free of their extraneous coats. The weakness of cells treated in this way to mechanical stress and interfacial forces suggests that the external surface on these cells is actually the protoplasmic surface layer.

All extraneous coatings including jelly, vitelline and fertilization membranes, and hyaline layers are non-essential to the life of the cell. The absence of any of these coatings does not prevent a fertilized egg from cleaving. Eggs from which vitelline membranes have been removed by chemical means are capable of being inseminated and will cleave normally. The hyaline layer is a dispensable coat of the single cell. However, if it is removed from the cleaving egg, the blastomeres will not remain attached except by temporary connecting stalks. The hyaline layer is therefore essential to a multicellular organism, and must be considered as a primitive intercellular cement.

A substance released by sea-urchin blastulae at the time of hatching dissolves fertilization membranes.[16] This substance is, at present, considered as an enzyme, and it can be recovered and preserved in sea-water. The action of this substance is most rapid on the transitional membrane (intermediate stage between the vitelline membrane and the tough fertilization membrane). The enzyme, however, has no effect on the precursor of the fertilization membrane (the vitelline membrane), nor on the hyaline layer.

Experimental evidence has been recently presented [5] which indicates that the normal growth in dimensions of the fertilization membrane is independent of elastic stretching such as might be occasioned by osmotic phenomena. On activation by sperm insemination, the vitelline membrane begins to lift away from the protoplasmic surface layer, and expands until it lies an appreciable distance away from the egg's surface. The expansibility has been shown to be independent of the content of the perivitelline space which the developing fertilization membrane creates. It is entirely possible that the intrinsic expansibility of the membrane-forming substance is due to the change in configuration of the protein molecules which make up a significant fraction of the fertilization membrane.

The Protoplasmic Surface Layer

Pfeffer was the first to use the term *plasma membrane* for the protoplasmic surface layer. He postulated the existence of a semi-permeable membrane to account for the results obtained in his classic studies on the osmotic properties of plant cells. Pfeffer and Nageli, who preceded him, regarded the surface of protoplasm as a haptogen membrane, protein in nature.

Butschli was opposed to the idea of a differentiated plasma membrane, holding that the integrity of a protoplasmic body is due to the immiscibility of its material with the aqueous environment. The absorption of water by protoplasm he regarded as analogous to the way in which the water fluids diffuse into fatty oils where they may collect as aqueous drops. Lepeschkin [19] has accumulated considerable data on this point, one of his chief arguments for Butschli's view being that protoplasm can absorb water only to a limited degree. This he showed by exposing the protoplasm of marine foraminifera and of certain algae to diluted sea-water and noting that the protoplasm of the cells, after swelling to a limited degree, became filled with vacuoles, the size and number of which varied with the dilution of the medium. When the cells were returned to ordinary sea-water the vacuoles disappeared and the cells regained their normal appearance.

It is true that the penetrability into living cells of many substances, particularly those which are lipid-soluble, can be explained without postulating that a selective permeability is limited to a peripheral layer of the protoplasm. On the other hand, the membrane theory explains many more phenomena.

Experimental work employing the oil-coalescence method has shown that the protoplasmic surface layer is essential to the life of the cell. So far no method has been worked out which permits the removal of this layer without causing irreversible damage to the cell. In the absence of extraneous coats the protoplasmic surface layer is very delicate. Very slight mechanical stresses cause an irreparable rupture, as evidenced by the tearing of fertilized eggs developing in Ca-free media. Coalescence of such cells with oil drops as small as 5 microns in diameter sets off an immediate disintegration of the entire cell. Unfertilized *Arbacia* eggs with vitelline membranes intact withstand considerable mechanical strain, and coalescence with oil drops as large as 50 microns in diameter does not destroy the cell.[15]

Experiments on the electrical conductivity of cells offer another argument in favor of the idea that the external surface layer of protoplasm possesses properties decidedly different from those of the interior. These experiments were first done by Hoeber [13] on red blood-cells in serum and on frog's muscle in isotonic NaCl. Hoeber found that the impedance to the passage of a current decreases with a rise in the frequency of the alternating current and that at very high frequencies the impedance falls to a low value.

The more recent work of Cole and his collaborators [9] has shown that the cell membrane has a resistance of the order of 1000 ohms/cm^2. This is substantially a measure of the ability of the cell surface to transmit ions. The cell surface likewise possesses a capacitance of about 1 microfarad/cm^2. The latter represents the ion-

impermeable structure. Cole [9] concludes: "The mechanism of the ion permeability is not obvious from resistance measurements but the capacity and dielectric loss suggest that the impermeable structure may be highly organized and a number of molecules in thickness."

Other important measurements on the physical properties of protoplasmic surface layers have been made by Harvey and his collaborators. These deal with measurements on the tension at the surface of living cells, a subject recently reviewed by Harvey [10] and by Harvey and Danielli.[11] The tension is of a low order of magnitude, and the values suggest that the protoplasmic surface layer is polymolecular, consisting of layers of protein molecules alternating with layers of aliphatic compounds, a concept also supported by the optical data of Schmitt, Bear, and Ponder.[21]

The relative impermeability of the surface of protoplasm as contrasted with its interior has also been shown by the microinjection of various substances. Freshwater amebas, immersed in solutions of magnesium chloride at concentrations as high as 1 to 2 molar, move about freely and show no injury. On the other hand, an injection of 0.2*M* $MgCl_2$ of an amount less than the volume of the nucleus of the ameba causes, at the site of the injection, an immediate coagulation which spreads until the entire ameba is involved. The effect of $CaCl_2$ is similar to that of $MgCl_2$ except for the complication that the ameba immediately reacts by pinching off the region at the site of injection before the salt has had time to spread. The question of old and newly formed surfaces is discussed elsewhere.

The diffusibility through the interior of protoplasm is most strikingly seen when injections are made of aqueous solutions of dyes which do not penetrate the cell from the medium. This is illustrated by injecting an aqueous solution of a non-toxic acid dye, such as chlorphenol red into a variety of cells such as living muscle fibers, gastric mucosa cells, starfish eggs and fresh water amebas. If a microdroplet of the solution is injected into the cytoplasm, the color quickly diffuses until it stops at the surface boundary. Evidently, the interior of the cytoplasm is freely permeable to the dye and the inability of the dye to get in from without or to pass out from within must be ascribed to the existence of a surface film or layer which is impermeable to the dye from either side. A great number of aqueous solutions have been injected into a variety of cells and, in every instance, the solutions, provided they are non-toxic and are of the type which do not penetrate from without, will diffuse through the interior when injected.

Interesting cases are the injection of $NaHCO_3$ and of NH_4Cl into the cytoplasm of a starfish egg. It has been known that CO_2 and NH_3 readily penetrate living cells whereas the highly dissociated NaCl and NaOH do not. Jacobs [14] had shown that starfish eggs, immersed in an alkaline solution of $NaHCO_3$, develop an intracellular acidity while, on the other hand, an intracellular alkalinity occurs when the eggs are exposed to a solution of ammonium chloride. However, when these solutions are injected directly into the eggs the results are reversed. This is because the injection brings the entire solution into the interior of the cell, with the result that the internal protoplasm reacts to the alkalinity of the carbonate and to the acidity of the ammonium chloride solutions.

The very fluid state of the plasma membrane can be demonstrated in the following manner with a starfish egg. The extraneous coating of an egg is torn and a part of the protoplasm allowed to flow out as an exovate. With a delicate microneedle the naked surface of the exovate is seized and dragged out for some distance in the form of a filament. On reversing the movement of the needle the substance of the filament flows back into the exovate with no sign of wrinkling on its surface. If the filament is stretched farther it may break into a string of coherent droplets. On lessening the tension the droplets flow together and when the filament is released it flows back into the main body of the protoplasm.

Another indication of the fluidity of the protoplasmic surface is the churning movements which can be induced in an unfertilized, mature *Arbacia* egg denuded of its extraneous coatings. It is significant that this experiment can be done with eggs immersed either in sea-water, in solutions of calcium or of potassium chloride. Eggs previously stratified by centrifugation exhibit the same property. The eggs were mounted in a broad, shallow hanging drop and those adhering to the coverslip pushed gently with a blunt microneedle. As the egg was pushed along the undersurface of a coverslip, peripheral currents appeared, streaming forward along the two sides and backward along the top and bottom of the egg. Extraneous particles adhering to the surface clearly showed by their movement that the streaming involved the surface. A drop of olive oil, applied to the surface which it caps, was carried forward along the side and backward along the undersurface of the egg.

A striking result occurred when the oil-cap was large enough to be so maneuvered, during the pushing of the egg, that it lay over the region including the two opposing, peripheral currents. When the currents once got under way the oil-cap broke in two, each portion being carried along in an opposite direction.

Still another experiment was the following: A drop of oil, exuding from the micropipette but remaining continuous with the oil in the pipette, was brought into contact with the surface which it capped. By moving the pipette the oil slipped over the surface but could not be pulled off without carrying some protoplasm away with it. The cap of oil on the protoplasmic surface behaved like a drop of oil on the surface of water.[2]

Thickness of the Protoplasmic Surface Layer

Under dark-field illumination a bright line frequently is visible at the boundary of the protoplasm and some investigators have ascribed a measureable thickness to it. However, it is highly probable that such measurements include more than the protoplasmic surface layer *per se*. In the case of unfertilized sea-urchin eggs, it would be difficult indeed to distinguish between the limits of the vitelline membrane and of the protoplasmic surface layer.

Injection experiments indicate that the thickness of the protoplasmic surface layer is far too small to be microscopically visible. If, for example, an ameba or a starfish egg is immersed in a solution of a dye which cannot penetrate and if a quantity of the same solution is injected into the cell there is no evidence of a colorless zone between the color inside and outside the cell.

Measurements have been made of the thickness of the bright line which is to be seen bounding the surface of isolated cells under darkfield illumination. Such lines are very noticeable and seem to have an appreciable thickness, for example, around fibroblasts and other cells in tissue culture. The phenomenon of apparent thickness of the line is probably an optical illusion, especially since the visibility of this line is due to the effect of light scattering. When two or more such cells are in intimate contact the bounding line between them disappears completely. Indeed, a sheet of cohering cells viewed in the dark field gives the effect of a syncytium with a boundary line appearing only along the free margin of the sheet.[6]

The recent electrical measurements by Cole [9] indicate that the protoplasmic surface layer may be a membrane at least 100 Å thick (exact estimations of membrane thickness from capacitance data are impossible due to uncertain values for the dielectric constant of the membrane components). The optical data of Waugh and Schmitt [22] likewise suggest that the protoplasmic surface layer is polymolecular.

The Protoplasmic Surface Layer in Relation to Extraneous Coats

All extraneous coatings are those which normally envelop cells but are not essential for their life. On the other hand, without the protoplasmic surface layer

the cell cannot survive. This layer is extremely delicate and easily ruptured. The rupture, unless immediately repaired, is inevitably followed by cytolysis. This bounding layer must therefore be regarded as an integral and indispensable part of the protoplasm. In nature it is probably seldom, if ever, exposed, being protected by an extraneous coat which, in the fertilized Echinoderm egg, is the hyaline layer. Three separate techniques have been employed to expose the actual protoplasmic surface layer of the *Arbacia* egg: the mechanical, for coatings which are sufficiently brittle and loosely stretched so as to be removable by shaking; the chemical, effective in removing the intercellular cement and hyaline layer material which are dispersible in the presence of excess monovalent cations; and the biological, in which, for example, the act of fertilizing an egg induces the lifting of the vitelline membrane which can then be removed by shaking. A combination of the three techniques is sometimes necessary to ensure the exposure of the protoplasmic surface.[15] So far, the medium best tolerated by naked *Arbacia* eggs is a mixture of NaCl and KCl in proportions of 19:1 in concentrations isotonic with sea-water but at a pH more acid than that of sea-water, *i.e.*, 7.0. In these solutions the naked eggs continue to cleave at the normal rate for many generations, the resulting blastomeres falling apart and continuing to exist as isolated cells. This demonstrates the dispensibility of calcium in the surrounding medium as a factor not only in the maintenance, but also in the reformation of the protoplasmic surface layer, since the development of thousands of cells from the original ovum must involve an enormous increase in total surface area.

There is a significant difference in the way the protoplasmic surface is re-formed in the presence and in the absence of calcium. In the latter case, no repair is possible if the surface layer is actually torn to expose the internal cytoplasm. No interface develops; the cytoplasm flows out and a disintegration of the surface layer spreads from the margin of the tear over the rest of the cell. However, as long as no actual tear occurs, the protoplasmic body can be pinched with microneedles into several rounded bodies of viable protoplasm and strands can be drawn out which may separate from the main body with no evidence of cytolysis. The isolated pieces immediately round up into spheres as long as no extraneous material is present to restrain them. The resulting increase in area of the protoplasmic surface layer is to be accounted for not only by the stretching of a pre-existing layer but also by the addition of surface layer material from the underlying cytoplasm. This type of surface formation also seems to be the only means in media containing a disproportionately large amount of calcium in which an actual rupture, occasioned by a mere scratch of the surface, results in an irreversible coagulation of the entire protoplasm.[8]

Attempts have been made to apply a solution of calcium chloride to the inner surface of the plasma membrane of the sea-urchin egg. A micropipette is inserted through the interior and the solution expelled when the tip of the pipette is close to the inner side of the far surface. The result is always a coagulation; friable strips can then be pulled from the surface with a needle. However, it is impossible to ascertain whether the injected $CaCl_2$ coagulates the plasma membrane by attacking it from the inside, or whether the coagulation is due to the action on the cytoplasmic material underlying it.

Repair of the Protoplasmic Surface Film and its Permeability

The protoplasm of a living cell is able to repair a torn surface if the tear is not too extensive, and if the invironment is normal. This is well illustrated in microdissection experiments on starfish eggs in sea-water. No breakdown occurs if the tip of the needle is moved slowly through the protoplasm and out of its surface. But if the piercing action is performed suddenly, the surface film disrupts and the exposed cytoplasm will begin to pour out and its granules to scatter in the medium. While this is occurring, films may appear around the masses of the disorganizing material which swell and burst. Frequently films form within what appears to be normal cy-

toplasm. These films may unite and reach the intact film which surrounds the remaining portion of the egg after which no further disintegration occurs.

It is difficult to decide between the relative permeability of old and freshly formed surface films of protoplasm. When sea-urchin egg, with a freshly formed surface film where it had previously been torn with microneedles, is immersed in hypotonic sea-water the portion covered by the new film bulges outward. However, this indicates that the new film is simply weaker owing to the lack in that region of a restraining, incipient vitelline membrane and to the absence of an underlying, gelated cortex. No such bulge appears when the egg is in the normal sea-water, unless mechanical pressure is brought to bear on the egg. Numerous attempts have been made to observe the penetration of normally non-penetrating dyes across the newly formed film. None have been successful. It would appear that the new film, when once formed completely, is no more permeable than the pre-existing film surrounding a normal egg.

The Nuclear Membrane

The nucleus is surrounded by a morphologically constituted membrane which can be torn. Its repair has never been observed and the tear always results in cytolysis. It is possible to puncture the nucleus and to inject dyes into it. However, subsequent removal of the micropipette is generally unsuccessful unless done with extreme care.

Its permeability is of a different order from that of the plasma membrane. In this regard, an interesting feature is the fact that all cell non-penetrating acid dyes which have been injected into the cytoplasm or the nucleus can diffuse across the nuclear membrane. The dyes eventually accumulate either in the nucleus or the cytoplasm, according to whether the pH of the nucleus or the cytoplasm respectively is the pH of a medium for which the dye has a preference.

The Consistency of Protoplasm

The physiologically active protoplasmic surface layer is to be regarded as an exceedingly thin film which determines the selective permeability of the cell and which reacts differently to the various salts in the medium. We have already seen that this membrane generally is covered on the outside by a layer. This layer differs in its properties from those of the plasma membrane and can be removed without injury to the protoplast.

The consistency of the internal cytoplasm varies considerably. In many plant protoplasts its fluidity is evident from the continuous streaming movements which it exhibits in cyclosis. In animal cells it is viscid and semi-solid in some cases and liquid in others. In many fluid protoplasts, plant and animal, there is often a more solid cytoplasmic cortex of varying thickness lying directly beneath the plasma membrane.

These differences in consistency have been ascertained by various methods, one of which is to observe the effect of centrifugal force in dislocating the visible constituents within the cell. This method has shown that the viscosity of protoplasm even in the same cell is not constant, and also that variations in the temperature of the environment affect it. Heilbrunn [12] has found, for example, that the internal viscosity of the egg of the clam *Cumingia* is at a maximum at about 15°. On each side of this temperature the viscosity drops to a minimum at about 2° and at 30°. At 1° and at 31° there is a sudden and pronounced increase in the viscosity. He states that this increase in viscosity differs from the heat coagulation of proteins in being reversible. It is possible that the changes in viscosity are related to changes in the permeability of the cell to water at the various temperatures and to changes in the state of the fatty materials within the cell.

The existence of Brownian movement, made visible with dark ground illumina-

tion, appears to be almost a universal phenomenon in protoplasm. This movement can be observed not only in the more fluid regions of cells but also in regions which, by means of microdissecting needles, have been shown to be in a jellied state.

A remarkable feature is the fact that semi-solid regions in protoplasm can sometimes be made suddenly fluid by mechanical agitation. This phenomenon * can be exhibited in a sea-urchin egg undergoing division. At this time the fluid protoplasm is being converted into two jelly-like, semi-solid masses. If a microneedle is inserted into one of the masses and the needle is suddenly jerked, the mass becomes converted immediately into a fluid. If the needle is then removed or allowed to remain stationary for a few minutes the original jellied condition returns and the division of the egg proceeds normally.

It is noteworthy that the normal activities of a cell may continue even when the fluidity of its protoplasm is increased by exposing the cell to hypotonic conditions. This can be seen in fertilized sea-urchin eggs immersed in sea-water diluted with distilled water to as much as two-thirds of its original concentration. Under such conditions the protoplasm becomes so fluid that the visible granules within it undergo such violent oscillatory movement as to be evident even under ordinary transmitted illumination. In spite of this, nuclear division takes place and the egg segments in the usual way but at a slower rate.

The Water-miscibility of Protoplasm

Microinjection experiments present strong evidence that water forms a continuous phase in protoplasm. Aqueous solutions diffuse readily through the cytoplasm of such diverse types of cells as protoplasts of root hair cells, amebas, various ova, muscle fibers, nerve cells, and ciliated epithelial cells. To ensure success the injection must be performed gradually; otherwise a localized disintegration may occur. In the latter case the resulting injury may spread and involve complete cytolysis of the cell. Frequently, however, a membrane may form around the deintegrating region to constitute a vacuole while the rest of the protoplasm remains unaffected.

Water-immiscible fluids, when injected, always form discrete droplets, the shapes of which are conditioned by the presence or absence of structural elements in the cytoplasm. Oil drops often show a tendency to maintain a spherical configuration due to the positive nature of the tension between the oil and the protoplasmic phase.[17] In the muscle fiber of the frog, large oil drops are compressed to cylindrical shapes because of the steric hindrance offered by the shape of the muscle fiber.

The water within the cytoplasm freezes readily.[7] To demonstrate this, the interior of the sub-cooled cell must be inoculated with an ice-crystal, since the external membrane tends to obstruct the spread of freezing from outside the cell. In the ameba the freezing is in the form of feathery ice-crystals which spread in different directions from the site of the inoculation. In the case of the muscle fiber of a frog the ice forms long, slender columns between denser, longitudinally arranged constituents of the fiber. If the frozen condition is maintained for more than a few minutes the columns progressively enlarge and the dehydration of the protoplasm becomes irreversible. In the case of the ameba the feathery crystals eventually condense into blocks of ice.

Ice crystals, as they are forming, tend to grow along lines of least resistance. In the protoplasm their growing tips have never been seen to puncture relatively stiff structures or pronounced interfacial films. For example, in the ameba, when the feathery crystals reach a gastric vacuole or the contractile vacuole, the tips of the crystals spread over the vacuolar membrane and eventually surround the vacuoles. This fact opens up a field for investigating intra-protoplasmic structures.

* Termed *thixotropy* ("change by touching") by H. Freundlich, who demonstrated it with bentonite and other sols. J. A.

The Protein Phase of Protoplasm

Recent studies based on the interaction between oil drops and protoplasm indicate that proteins and protein complexes are bound together to form a continuous protein phase.[17] The surface-chemical properties of cytoplasmic proteins are considered in another paper in this volume.[18]

Evidence based on the diffusibility of water-soluble components as well as on surface-chemical properties of the cytoplasmic proteins suggests that at least two phases are continuous in protoplasm, *i.e.*, the aqueous and the protein.

The Hydrogen-ion Concentration of Protoplasm

Since protoplasm is an organized structure which is destroyed by crushing, it is obvious that a determination of the hydrogen-ion concentration of extracts of crushed tissue cannot be a reliable clue to the intensity of acidity within the dividing cell. Moreover, the unavoidable admixture of intercellular fluid from the tissue introduces an additional source of error.

Reliable results can be obtained only while the protoplasm is living. Electrometric measurements with microelectrodes inserted into living cells have not yet been successful, principally because of the difficulty of ensuring that the tip of the electrode is actually within the cell interior. The tendency of the protoplasmic film is to spread and close a gap. When a needle is inserted into a cell and left there for an appreciable time, the protoplasmic film spreads along the inserted part of the needle to its very tip. The result of this is that the needle comes to lie in a deep depression in the cell with no part of it within the cell. Colorimetric methods, which have afforded fairly consistent results, fall into three groups: the use of natural indicators already existing in cells, vital staining with dyes having a pH virage, and the micro-injection of indicators into the protoplasm. It should be noted here that the pH values obtained are of the aqueous phase of protoplasm. The only indicators to be relied on are those which are more purely water-soluble.

Natural indicators have been unreliable mainly because they are usually localized in vacuoles which do not necessarily have the same pH as the surrounding cytoplasm. This accounts for the findings by botanists of so-called "acid," "neutral" and "basic" cells. The differences are due not to cytoplasm but to the varying pH conditions of their contained vacuoles. An interesting case in this connection is the occasional finding of "blue nuclei" in the petals of purple tulip petals. This condition has been found to be due to injury in which the pigment of the cell vacuole has escaped by rupture of the vacuole followed by the coloration of the alkaline nucleus in the cytolyzed debris of the cytoplasm.[4]

Vital staining is also unreliable not only because the dyes which penetrate cells from without tend to become segregated in vacuoles but also because of the extreme difficulty of controlling the amount which penetrates. Moreover penetrating dyes are mostly basic in nature and many tend to color fatty bodies in the cell.

The above statements indicate the ease with which errors are likely to be made in determining the pH of living cells.

The micro-injection of lipid-insoluble indicators into the protoplasm has given results least open to question.[1] Even then the color tends to be more or less rapidly segregated in vacuoles. Values obtained for the inter-cytoplasmic pH by the best available methods lie between 6.8 and 7.2. These values have been obtained by injecting a series of indicators of overlapping ranges into such cells as echinoderm eggs, various somatic cells of Vertebrates, the protoplasts of plant cells and amebas. By this method it has also been found that cytoplasmic vacuoles give color indications which have no necessary relation to the observed pH of the surrounding cytoplasm. In Metazoa the cell nucleus has been found to be consistently more alkaline

than cytoplasm, the colorimetric value from injected indicators being in the neighborhood of pH 7.6 to 7.8.

Both cytoplasm and nucleus maintain their respective pH values even when the pH of the environment is shifted by means of penetrating acids and bases. The pH of the cytoplasmic vacuoles, however, changes readily. It was by the color change of neutral red in these vacuoles that the penetration of ammonia and of carbon dioxide was first determined. The buffering power of the protoplasm seems to be limited, and in some cells the pH can be upset to the extent of several tenths of a pH unit on either side of the normal pH value. Death ensues, however, when the pH is thrown too far in either direction, particularly to the acid side. Mechanical injury which initiates cytolysis, even when localized, is accompanied by the production of acidity, the intensity of which is in the neighborhood of pH 5.0 to 5.5. On the other hand the nucleus, when injured, gives no evidence of changed pH.

In short, we find within the living cell the remarkable phenomenon of regions which are circumscribed by well defined membranes and differ in the intensity of their acidity; the nucleus with a pH of 7.6 to 7.8, the cytoplasm with a pH of 6.8 to 7.2, and the cytoplasmic vacuoles with a variable pH which can be shifted readily by penetrating acids and alkalies in the medium.

The Action of Salts on the Internal Protoplasm

The relatively neutral reaction of the internal protoplasm suggests that much of the protein present is on the alkaline side of the isoelectric point. This is borne out by the ease with which coagulation occurs when salts of polyvalent cations, even in very dilute concentrations, are injected into such cells as echinoderm eggs, amebas, and the protoplasts of plant cells. For example, $CaCl_2$, in all concentrations down to $M/200$ and $MgCl_2$ down to $M/1000$ produce a coagulating effect which is either restricted or extensive, according to the amount injected. The injection of NaCl and of KCl [8] has no such effect, presumably because they form soluble salts of the protein present.

An additional argument for the existence of proteins on the alkaline side of their isoelectric point in the protoplasm is offered by the fact that picric acid, when injected into a cell, exerts no coagulating reaction. Pollack [20] found that it was possible to inject an aqueous solution of picric acid into an ameba with no other effect than to color it a vivid yellow. An explanation for this unexpected result is brought out by injecting a mixture of picric acid with a solution of bromcresol purple, an indicator which is yellow in acid solutions below pH 6.0 and purple in more alkaline solutions. When a moderate amount of the bright-yellow mixture is injected, the ameba assumes a murky green color and remains alive. When too much is injected the color remains yellow and the ameba becomes coagulated and dies. These facts are explained on the basis that with moderate injections the proteins remain on the alkaline side of their isoelectric points and the picric acid is converted into ineffective sodium picrate. However, if the amount injected is large, the excess acid carries the proteins to the acid side of their isoelectric points, thus permitting the formation of an insoluble protein picrate.

The ease with which the internal protoplasm is coagulated by $CaCl_2$ in contrast to the non-coagulating effect of NaCl and of KCl has been questioned by Heilbrunn,[12] who obtained opposite results by immersing sea-urchin eggs in solutions of NaCl and of $CaCl_2$ isotonic with sea-water. Physical changes observed within a cell from immersion experiments are difficult to interpret, since they may not always be due to penetration of substances from the medium. However, even if the effect is assumed to be due to penetration the difference in results still may be accounted for by the difference in the amounts of the salts which are brought into the cell by the two methods of experimentation. The difference in the effect of injection and of immersion is brought out strikingly if, instead of $CaCl_2$, we use the salt of a basic dye, *e.g.*, neutral

red. This dye is a vital stain and, if present in the medium in appropriate dilutions, will readily penetrate cells and color their cytoplasmic granules and vacuoles without any sign of injuring the cell. However, if an aqueous solution of the dye is injected directly into a cell, the protoplasm at the site of injection is converted immediately into a red, coagulated mass. Apparently, in the usual method of vital staining the dye is taken up by vacuoles and granules as fast as it penetrates and before it has time to combine with the cytoplasmic proteins although coagulation may even then occur if an excessive amount of the dye salt is present in the medium.

References

1. Chambers, R., *Bull. Nat. Res. Council,* No. 69, 37 (1929).
2. —, *Am. Natur.,* **72,** 141 (1938).
3. —, *Cold Spring Harbor Symposia,* **8,** 144 (1940).
4. —, *Am. J. Bot.,* **28,** 445 (1941).
5. —, *J. Cell. Comp. Physiol.,* **19,** 145 (1942).
6. —, and Fell, H. B., *Proc. Roy. Soc. London,* **B109,** 381 (1931).
7. —, and Hale, H. P., *Ibid.,* **B110,** 336 (1932).
8. —, and Reznikoff, P., *J. Gen. Physiol.,* **8,** 369 (1923).
9. Cole, K. S., *Cold Spring Harbor Symposia,* **8,** 110 (1940).
10. Harvey, E. N., *J. Applied Physics,* **9,** 68 (1938).
11. —, and Danielli, J. F., *Biol. Rev.,* **13,** 319 (1938).
12. Heilbrunn, L. V., "The Colloid Chemistry of Protoplasm," Borntraeger, Berlin, 1928.
13. Hoeber, R., "Physikalische Chemie der Zelle und Gewebe," W. Engelmann, Leipzig, 1924.
14. Jacobs, M. H., "General Cytology," Chapter III, Edmund V. Cowdry, ed., University of Chicago Press, Chicago, 1924.
15. Kopac, M. J., *Cold Spring Harbor Symposia,* **8,** 154 (1940).
16. —, *J. Cell. Comp. Physiol.,* **18,** 215 (1941).
17. —, "Micrurgical and Germ-Free Methods," Chapter II, James A. Reyniers, ed., Charles C. Thomas, Springfield, Ill., 1943.
18. —, This volume, pp. 875-883.
19. Lepeschkin, W. W., "Kolloidchemie des Protoplasmas," J. Springer, Berlin, 1924.
20. Pollack, H., *Proc. Soc. Exper. Biol. Med.,* **25,** 145 (1927).
21. Schmitt, F. O., Bear, R. S., and Ponder, E., *J. Cell. Comp. Physiol.,* **9,** 89 (1936); **11,** 309 (1938).
22. Waugh, D. F., and Schmitt, F. O., *Cold Spring Harbor Symposia,* **8,** 233 (1940).

Some Surface-chemical Properties of Protoplasmic Proteins *

M. J. KOPAC

Visiting Assistant Professor of Biology, Washington Square College of Arts and Science, New York University, New York, N. Y.

Because of the difficulties involved in the study of protoplasmic proteins it has been necessary to develop new techniques in order to evaluate the properties of such proteins either in living cells or from the residue of freshly disintegrated protoplasm. These new methods are based on the theoretical and experimental procedures of surface chemistry as adapted to microscopic dimensions.

A new interfacial micro-tensiometer has been built which permits the simultaneous measurement of interfacial tension and amounts of proteins or other substances adsorbed at oil-water interfaces.[17] Flow-pressures as well as equilibrium pressures can be measured so that interfacial tensions may be determined at any time after the interface is established.

The adsorption of proteins or other substances at oil-water interfaces is measured by a semi-automatic drop-retraction method. The apparatus permits of precise control over the size of an oil drop poised on the tip of a micropipette. The surface area of an oil drop can be diminished by decreasing its size with the drop-retraction apparatus. The glass wall of the orifice of the micropipette serves as a sweep and, if the retraction is done slowly, proteins will not enter the pipette with the oil. This is

* These investigations were made possible by a grant from the Rockefeller Foundation.

analogous to the compression of surface-adsorbed molecules by the sweep of a straight-edge barrier at the surface of a conventional Langmuir film trough.

Tension measurements at oil-protoplasm interfaces give indications of certain protein characteristics, but these data alone are insufficient to indicate any particular organization in the living cell. Differences in tension lowering, for example, as produced on various oil surfaces by the same protein species cannot be interpreted properly unless one also knows how much protein was adsorbed. Accordingly, interfacial tension measurements must be augmented with measurements on the number of protein molecules adsorbed at oil-water interfaces together with the determination of factors which can modify such adsorption.

The results obtained by applying surface-chemical methods to living cells are summarized in the following sections: (1) Surface activity of intact cytoplasm at oil-water interfaces; (2) surface activity of the residue from rapidly disintegrated Arbacia egg cytoplasm; (3) protein adsorption at oil-water interfaces; and (4) spontaneous Devaux effect at oil-protoplasm interfaces.

Table 1. Tensions at *Oil-Protoplasm* and Related Interfaces

Oil Phase	Aqueous Phase	Method	T_x in Dynes/cm	Reference
Mackerel egg oil [1]	Mackerel egg cytoplasm	Centrifuge-sessile drop	0.1-2.6	(8)
	Mackerel egg-extract [2]	Du Nouy tensiometer	0.8	(5)
Daphnia egg oil [1]	Daphnia egg cytoplasm	Centrifuge-sessile drop	0.6-1.4 [3]	(7)
Olive oil [4]	Ameba cytoplasm	Centrifuge-sessile drop	1.8-2.0 [3]	
Paraffin oil [oleic acid]	Valonia aplanospores	Oil-coalescency [5]	<2.5	(11)
	Aging Arbacia eggs		<3.0	(2)
Liquid petrolatum	Arbacia egg cytoplasm	Flow-pressure	$<6 \times 10^{0}$	(16) [6]
Oleic acid			$<5 \times 10^{-1}$	(16) [6]
Oleum percomorphum	Arbacia egg-extract [7]	Drop-retraction [10]	3.5-7.5	(16) [6]
Oleic acid		Flow-pressure	0.5-1.2	(15)
Cottonseed oil		Flow-pressure	5.7	
Oleum percomorphum [4]	Asterias egg cytoplasm [9]	Devaux effect	0 [8]	(12, 13, 14)
	Asterias egg-extract	Drop-retraction [10]	$n \times 10^{-1}$ [11]	(16) [6]
			$n \times 10^{0}$ [12]	

[1] Large oil-drop normally present in these eggs.
[2] Aqueous extract from crushed or frozen mackerel eggs.
[3] Tension increases with centrifugal force or degree of flattening.
[4] Oil-drop injected into cytoplasm.
[5] Thermodynamic maximum calculated from data on oil-coalescency experiments.
[6] Unpublished data.
[7] Arbacia eggs disintegrated in 0.53M KCl-solution.
[8] Zero interfacial tension occurs on cytolysis if oil is injected prior to cytolysis.
[9] Cells cytolyzed by puncturing the germinal vesicle with a microneedle.
[10] Calculated from $T_x = T_0(1 - \pi d_c^2/\pi d_0^2)$ but using $\pi d_c^2 = \pi d_c^2$. Accordingly, the calculated value for T_x is probably too high (cf. 17).
[11] Oil brought in contact with cytolytic residue, 1 minute after cytolysis.
[12] Oil brought in contact with cytolytic residue, 10 minutes after cytolysis.

Surface Activity of Intact Cytoplasm at Oil-water Interfaces

Harvey and his collaborators [7, 8] devised methods for measuring oil-protoplasm interfacial tensions by applying the theory of the sessile drop to those oil drops normally present in cells or to those introduced into cells by microinjection. The imperceptible flattening which occurs by gravity is accentuated by applying known centrifugal accelerations with the centrifuge-microscope. The tensions at the surfaces of oil drops can be calculated from the degree of flattening at a given centrifugal field.

The micro-tensiometer, based on the flow-pressure through the orifice of a micro-pipette, was employed by Kopac.[17] The micro-pipette which contains the oil is inserted into the cytoplasm of a cell with a micro-manipulator, so that contact between the oil and cytoplasm can be established. The micro-pipettes are calibrated by measuring the flow-pressures of the same oil with the pipette dipping in dilute buffer solutions. In the latter systems, tensions at the oil-buffer interfaces can be measured by classical methods. These data, together with those obtained by Harvey, *et al.*, are summarized in Table 1.

Oil-protoplasm interfacial tensions, when expressed in dynes/cm, are of a low order of magnitude. The tension-lowering activity of protoplasm and protoplasmic proteins can be demonstrated even more strikingly by recalculating the data given in Table 1. Oil-water interfacial tensions, except those of apolar oils, vary with the pH and salt content of the aqueous phase.[8] The T_o values which were used in recalculating the data are the tensions of the various oils in contact with an aqueous phase of about the same pH as that of cytoplasm, *i.e.*, pH = 6.8. In some cases, the T_o-values were derived from other data.

The T_e values were taken from the same sources as those given in Table 1. The ratio, θ, which equals T_e/T_o and varies inversely with surface activity, indicates the tension-lowering activity of the components in the aqueous phase. The interfacial tensions of the various oils used and their corresponding θ values are given in Table 2.

Protoplasm is capable of reducing the tension at most oil-water interfaces by very significant amounts, θ being generally less than 0.2. Protoplasmic derivatives show similar surface activities. Albumins are also surface-active but their activity at oil-water interfaces is not as pronounced as that of cellular extracts. Only when lecithin or similar substances have been added to the interface with albumin, is there a tension lowering of the same order as that found with protoplasm or protoplasmic proteins.

Surface Activity of the Residue from Rapidly Disintegrated Arbacia Egg Cytoplasm

By appropriate centrifugal fractionation, the residue from rapidly disintegrated Arbacia eggs (in 0.53*M* KCl) can be separated into the following fractions: (a) pigment vacuoles; (b) yolk granules; (c) mitochondrial granules; and (d) cytoplasmic matrix residue.[9, 17]

The proteins of the cytoplasmic matrix as obtained by centrifugal fractionation are readily adsorbed at oil-water interfaces. These proteins, however, when in solution exist in the 3-dimensional or globular state. Some of the molecules which reach the interface undergo a 3 to a 2-dimensional shift. The unfolded, 2-dimensional protein molecule becomes insoluble in the aqueous phase and it does not become more soluble in the oil phase. Such molecules are therefore 'trapped' at the interface where they may be concentrated and packed by appropriate means. However, before such molecules can be packed, some of the molecules in the mixed population (3 and 2-dimensional configurations) will be forced out of the interface. Since the 3-dimensional protein molecules are water-soluble, they will be the ones

Table 2. Tension-Reducing Activity of Protoplasm, Cellular Extracts and Proteins

Oil Phase	Aqueous Phase	T_0	T_x	$\theta = T_x/T_0$	Reference
Mackerel egg oil	Mackerel egg cytoplasm	9 [1]	0.1	0.02	(8)
			0.6 [2]	0.07	
			1.8 [3]	0.2	
	Mackerel egg-extract	7 [4]	0.8	0.1	(5)
Olive oil	Ameba cytoplasm	10 [5]	2.0	0.2	(7)
Paraffin oil	Arbacia egg cytoplasm	38	6	0.16	(16)
Oleic acid	Arbacia egg cytoplasm	7.5 [6]	0.5	0.07	(15)
	Arbacia egg-extract	7.5 [6]	0.6	0.07+	
			1.3	0.18	
Cottonseed oil	Arbacia egg-extract	12.0 [6]	5.7	0.48	(17)
			9.1	0.76	
Brombenzene	Albumin [0.5 p. c.]	—	—	0.5 [7]	(4)
Oleic acid	Albumin [0.5 p. c.]	—	—	0.5 [7]	
Oleic acid [lecithin]	Albumin [0.1 p. c.]	7.5 [5]	1 [8]	0.1+	
n-Butyl phthalate	Albumin	23.0 [9]	6 [10]	0.25	(7)
			8 [11]	0.35	

[1] Buffered solution at *pH* 6.8, drop-volume method.
[2] Average value of determinations by centrifuge-sessile drop method.
[3] More than 90 per cent of measurements were below 1.8 dynes/cm.
[4] Phosphate buffer at *pH* 6.8, du Nouy tensiometer method.
[5] Deduced value, average for *oil-water* [pH *6.8*] interfacial tensions.
[6] 0.53 M KCl at *pH* 7, Flow-pressure method.
[7] Estimated by Danielli (4), T_x, and T_0 values not given in paper.
[8] Deduced value: "interfacial tension in the neighborhood of pH 7 is as low as that met in biological systems." Danielli (4), p. 192.
[9] Distilled water, centrifuge-sessile drop method.
[10] Average at 800-1600 × gravity.
[11] Average at 1600-3200 × gravity.

forced into the aqueous phase and only those molecules in the 2-dimensional state will remain. Accordingly, the saturation capacity of an oil-water interface depends on the extent to which the 3 to 2-dimensional shift is induced (see below).

Data on the ratio of expanded/compressed areas, K, of unfolded molecules were obtained at various surface pressures. The values for calculating K were derived from the simultaneous determination of tension-lowering as produced by a given number of protein molecules and the fraction of the oil-water interface which these molecules occupied.

The calculated K-values were then plotted against corresponding tension-lowering values. The value of K corresponding to an interfacial saturation of zero (extrapolation to zero surface pressure) was found to be 2.17. Where the interface is half-saturated, K equals 1.73; while at complete saturation K equals 1. Other values are: at 0.24 saturation, K equals 2.0; at 0.93 saturation, K equals 1.10.

The above K-values show that at oil-water interfaces under the experimental

conditions used there can be more than a 2-fold difference in area between expanded and compressed cytoplasmic protein molecules in the unfolded state. These are in agreement with data published by Neurath [20] for the areas of other expanded and compressed proteins (at water-air interfaces). Calculation of his data gives the following K-values: egg albumin, 1.69; serum albumin, 1.83; insulin, 2.14; and cytochrome-C, 2.21. Here the value of K is the ratio of expanded/compressed spreading areas at water-air interfaces of 1 mg of protein.

Protein Adsorption at Oil-Water Interfaces

The oils, tricaproin, *n*-butyl phthalate, or Squibb's Liquid Petrolatum have different adsorption curves, as might be expected from their obvious chemical dissimilarity. However, if to each of these oils is added enough oleic acid to make its concentration about 1 per cent, considerably higher amounts of proteins are adsorbed. Even more striking is the fact that one adsorption curve adequately accommodates the average determinations for each basic oil phase. The similar rates and amounts of adsorbed proteins show that these oils have acquired nearly identical surface-properties. The drop-retraction method was employed for these measurements.[17]

Pure oleic acid adsorbs cytoplasmic proteins rapidly and its interface becomes completely covered in less than 1 minute. The rate of protein adsorption on those oils to which oleic acid was added is probably influenced by the rate at which the less surface-active molecules in the oil become replaced at the interface by a monolayer (or more) of oleic acid. Oils which contain higher concentrations of oleic acid become completely coated with cytoplasmic proteins in less than 8 minutes. On the other hand, oils containing lower concentrations of oleic acid require considerably longer exposures to the proteins before their interfaces become entirely coated.

It is evident that the oil-water interface has a saturation capacity for cytoplasmic proteins and that this capacity depends on the oil phase and possibly on the proteins. For example, Liquid Petrolatum after being exposed to cytoplasmic proteins for 10 minutes is completely saturated when about 0.4 of its interfacial area has become covered by adsorbed molecules. However, Liquid Petrolatum + oleic acid (1 per cent), following a 10-minute exposure, becomes saturated with cytoplasmic proteins only when its entire interfacial area is covered. Other oils show either similar or different saturation capacities. Furthermore, their saturation capacities may be altered by adding a new component to the oil-phase, *e.g.*, the capacity of Liquid Petrolatum + lecithin (0.5 per cent) becomes 1.0 in less than 30 seconds (indicating complete coverage of the interfacial area).

These data [17] show that oil-water interfaces will become completely coated with cytoplasmic protein molecules only under certain conditions. The oil must contain molecules which can form interfacial complexes with those protein molecules reaching the interfacial zone. For complete coverage of the interface, the complex must be stable enough to withstand the high surface pressures resulting from the adsorption of large numbers of surface-active molecules. In the case of oleic acid dissolved in oil, the interface becomes oil [oleic acid-protein] water (see stage E, Table 3), in which the oleic acid-protein complex has sufficient stability to endure surface pressures ranging from 12 to 15 dynes/cm.

Spontaneous Devaux Effect at Oil-Protoplasm Interfaces

Devaux was the first to describe that egg albumin forms insoluble monolayers at water-air interfaces, and later [6] he reported on the adsorption of similar proteins at benzene-water interfaces. He observed that such interfaces in the presence of proteins might become folded or crinkled.* The amounts of protein required to produce this phenomenon were estimated to be just sufficient to form a monolayer

* See also "The Influence of Colloids on Crystalline Form & Cohesion," by W. M. Ord, London, 1879.

at the interface. This remarkable crinkling or folding at an oil-water interface by adsorbed proteins has been called the "Devaux crinkling effect" by Langmuir and Waugh.[18] A similar crinkling which occurs on drops of certain oils when brought in contact with cytolyzing starfish eggs has also been called the Devaux effect.[12]

Similar phenomena were observed by F. M. Ascherson * over 100 years ago; his paper, presented to the Paris Academy of Sciences on Nov. 12th, 1838, was published in *Arch. f. Anat. Physiol.*, 1840, p. 44, and a translation forms the first paper in "The Foundations of Colloid Chemistry" edited for the Colloids Committee of the British Association by E. Hatschek (E. Benn, Ltd., London, 1925), under the title: "On the Physiological Utility of Fats and on a New Theory of Cell Formation Based on Their Co-operation and Supported by Several New Facts." Ascherson stated: ". . . coagulation in the form of a membrane occurs inevitably and instantaneously when albumin comes into contact with a liquid fat, and that consequently an oil drop cannot be surrounded even for a moment by an albuminous liquid without a vesicular membrane or a cell forming round it. For the sake of brevity I shall call the property of forming membranes by contact *Hymenogony,* and the membranes thus formed *Haptogen* membranes. . . . The formation of a few globules which do not become flattened again is perhaps the most delicate reagent for discovering in distilled water the slightest trace of an organic substance, and I must add that up to now I have not found any which passed this test completely." Ramsden studied the adsorption of albumin, which he found was thus denatured.[21a]

Kopac reported[10, 12, 14] that oil drops when injected into starfish eggs, under certain conditions, spontaneously increase their interfacial area by crinkling or folding. Experimentally introduced oil-drops (by microinjection) will not show the crinkling effect unless the cells are partially or completely cytolyzed. For example, every oil drop (Oleum percomorphum) when microinjected into *Asterias* oocytes produces the Devaux effect providing the cells are cytolyzed not more than 30 seconds later, the cytolysis being induced experimentally by puncturing the germinal vesicle with a micro-needle. The same reaction occurs on oil drops introduced within 15 seconds after cytolysis. Also 99 per cent of the oil drops microinjected into oocytes within 1 minute before cytolysis and 50 per cent, if introduced within 1 minute after cytolysis, give the Devaux effect.

The evidence, at present, suggests that the majority of proteins and protein complexes in protoplasm are bound together to form a protein-continuous phase so that they become essentially non-diffusible and hence relatively non-adsorbable. Such a conclusion follows from the fact that spontaneous Devaux effects are not obtained unless the cytoplasm is disintegrated (also discussed by Chambers[1]). Following

* Ascherson also anticipated ultramicroscopic illumination, as the following quotation shows: "It can hardly be doubted that hymenogony acts under the influence of life just as it does in the chemist's test-tube. The irregular cells which I have found in ova, and the dull surface and slight irregularities which a practised eye fairly easily discovers on the globules in milk or yolk larger than $\frac{1}{80}$ mm, prove the existence of a haptogen membrane formed in the living animal. This skin appears to be the only cause of the spherical shape and of the isolation exhibited by drops of fat in plants and animals according to my observatons; likewise in milk, where Raspail had already conjectured the existence of a membrane on the globules, and in artificial milk. I have shaken oil with distilled water and have found that all drops which can conform to hydrostatic laws assume a lenticular shape with thin, very transparent edges, while they preserve their spherical shape and black margins even in the largest volume of water, provided they have previously had an opportunity of coating themselves with a haptogen membrane by the aid of a little mucus or albumin. This very decided difference can be perceived most easily when a few drops of oil are shaken with water containing only a little albumin. It is then still possible to distinguish among the droplets, the diameter of which does not exceed $\frac{1}{700}$ mm, those surrounded by a membrane, which increase in number with the albumin content of the liquid, from those which have remained free. The latter are best seen by focusing the microscope exactly on the surface of the liquid and illuminating from the side." Zsigmondy first saw fine colloidal particles by similar illumination. ("Colloids and the Ultramicroscope," by R. Zsigmondy, Eng. trans. by J. Alexander, New York, 1909). EDITOR.

cytolysis, the protein-continuous phase is broken thereby releasing the protein molecules in a form which permits their adsorption on suitable oil-water interfaces. The Devaux effect can develop only if enough protein molecules accumulate and unfold to form at least one monolayer on the experimentally introduced oil-surface.[10, 12, 14, 16, 18]

Protein phases bound together by experimental means also inhibit the development of the Devaux effect.[16] Trichloracetic or phosphotungstic acid produce a protein-continuous phase, albeit by precipitation, but under such conditions the components are obviously non-diffusible. No Devaux effects are produced after the introduction of either of these reagents into a cytolyzing oocyte even though suitable oils were injected into the cell prior to cytolysis.

During the life of a cell, both proteins and protein complexes must at various times break away from the continuous phase and exist in solution in the aqueous phase. An equilibrium, in all probability, is maintained between the dissolved proteins and those anchored to the protein skeleton.[23] The concentration of such molecules in the aqueous phase is insufficient to produce spontaneous Devaux effects, however.

Danielli and Harvey [5] conclude that the low values of oil-protoplasm interfacial tensions are due to the adsorption of a film of protein, probably a globulin, at the interface. It is important to recognize that non-protein substances frequently enhance the adsorption of proteins [17, 22] and thereby yield low interfacial tension values. Danielli [4] has shown that the addition of lecithin (in oleic acid) increased the surface activity at an oleic acid-water [albumin] interface. Lecithins and similar substances are present in living cells and derivatives of these substances also occur in the extracted cellular residues.

Evidence from studies of the Devaux effect indicates that the proteins of intact cytoplasm are not adsorbed in quantity by experimentally introduced oil surfaces. Such a situation would make the explanation of low tensions at oil-protoplasm interfaces rather difficult. One must therefore assume that these organized complexes which come in contact with the oil must have a high degree of intrinsic surface activity.

The above assumption is supported by the fact that tensions at oil-water [virus protein] interfaces are low,[16] and it is also known that similar macromolecules (hemocyanins) do not spread at water-air interfaces unless dissociation into smaller units occurs.[21]

Our tentative interpretation is that the interface between oil and protoplasm is fundamentally an oil [PROTEIN-protein complex] water * system, instead of an oil-water [protein] system, especially in those instances where the oil is native to the cell (mackerel and Daphnia eggs). The tensions at either interface could be of the same order. The protein complex is defined here as the association of proteins with other large molecules (including proteins) by weak apolar, dipole-dipolar, ion-dipolar, or interionic bonds.

During the measurement of interfacial tensions by the flow-pressure method, it is possible that enough cytolysis is caused by the introduction of the experimental oil phase to release additional proteins and protein complexes from the protein-continuous phase. Here the interfacial system would be: oil-water [protein complex] in which the complexes may have the same status as other high molecular weight protein particles (mitochondria and viruses, for example).

General cytolysis, probably, involves not only a further breakdown of the protein-continuous phase, but also the dissociation of protein complexes out of which it is formed. If these complexes dissociate further, the interfacial conditions become similar to oil-water [protein + lecithin] systems in which the degree of tension lowering is a direct function of the oil phase, the nature of the proteins and of the

* The symbol, PROTEIN, represents the postulated protein-continuous phase of protoplasm.

complexes in association with proteins. Phospholipids of the lecithin type, because of their surface-active properties, are believed to be the most important of the complexes other than proteins to be released on cytolysis. In this review, lecithin generically represents all complex phospholipids which may be derived from protoplasm.

In summary, the change in protein organization on cytolysis in relation to surface activity at oil-water interfaces may be schematically represented as follows:

Table 3

NATIVE *OIL-PROTOPLASM* INTERFACES

A. *oil* [*PROTEIN-protein complex*] *water*

EXPERIMENTAL *OIL-PROTOPLASM* INTERFACES

Incipient cytolysis [Reversible]

B. *oil-water* [*PROTEIN + protein complex*]
General cytolysis [Irreversible]
Breakdown of protein-continuous phase and complex dissociation

C. *oil-water* [*protein + lecithin*]
Redistribution of lipid complex

D. [*lecithin*] *oil-water* [*protein*]
Interfacial complex formation

E. *oil* [*lecithin-protein*] *water*

It has been shown that various substances when introduced in the oil-phase may modify protein adsorption (p. 878), and lecithin is one of these substances. Thus at stage D, the oil phase becomes conditioned by the lecithin, which in turn promotes protein adsorption and thus increases the saturation capacity of that interface. The result is the production of interfacial complexes (stage E) in which the transfacial forces are of maximum magnitude. If these complexes completely cover the oil surface, the interfacial tension becomes zero (transfacial forces equal interfacial forces).

The ideal time for introducing the oil phase for spontaneous Devaux effects seems to be between stages B and C so that the proteins and protein complexes can be adsorbed just before their inevitable dissociation. Morphologically, such time corresponds to the first onset of cytolysis. At this time, protein complexes as well as lecithin and protein molecules are simultaneously within the field-force of the oil-water interface and, following interfacial dissociation, the solution of lecithin in the oil and adsorption of proteins can occur independently of diffusion rates. Introduction of oil surfaces at later stages is less effective because now both lecithin and proteins must diffuse toward the interface (in small aggregates or as molecules) before solution and adsorption of adequate amounts is possible. In these later stages, diffusion from the bulk phase to the interface becomes important.[19]

Cytolysis, which results in the breakdown of the protein-continuous phase, and then of the complexes, yields not only protein molecules of low molecular weight (one to several Svedberg units) but also minor complexes between these and lipophilic molecules. Most of the proteins and other protoplasmic components which have been studied *in vitro* are obtained in this condition. The study of 'native' cytoplasmic proteins *in situ* has only been broached.

References

1. Chambers, R., Publ. No. 14 of the Amer. Assoc. Advan. Sci., pp. 20-30, 1940.
2. Chambers, R., and Kopac, M. J., *J. Cell. Comp. Physiol.*, **9**, 331 (1937).
3. Danielli, J. F., *Proc. Roy. Soc. London*, **B122**, 155 (1937).
4. Danielli, J. F., *Cold Spring Harbor Symposia*, **6**, 190 (1938).
5. Danielli, J. F., and Harvey, E. N., *J. Cell. Comp. Physiol.*, **5**, 483 (1935).
6. Devaux, H., *Compt. rend. Acad. sci. (Paris)*, **201**, 109 (1935); **202**, 1957 (1936).
7. Harvey, E. N., and Schoepfle, G., *J. Cell. Comp. Physiol.*, **13**, 383 (1939).
8. Harvey, E. N., and Shapiro, H., *J. Cell. Comp. Physiol.*, **5**, 255 (1934).

9. Hutchens, J. O., Kopac, M. J., and Krahl, M. E., *J. Cell. Comp. Physiol.*, **20**, 113 (1942).
10. Kopac, M. J., *Biol. Bull.*, **71**, 398 (1936).
11. Kopac, M. J., *Carnegie Inst. Washington Year Book*, **36**, 94 (1937).
12. Kopac, M. J., *Biol. Bull.*, **75**, 351 (1938).
13. Kopac, M. J., *Anat. Rec. Suppl.*, **72**, 114 (1938).
14. Kopac, M. J., *Anat. Rec. Suppl.*, **75**, 143 (1939).
15. Kopac, M. J., *Biol. Bull.*, **79**, 341 (1940).
16. Kopac, M. J., *et al.*, Unpublished data.
17. Kopac, M. J., Chapter II, "Micrurgical and Germ-Free Methods," James A. Reyniers, ed. Charles C. Thomas, Springfield, Ill., 1943.
18. Langmuir, I., and Waugh, D. F., *J. Gen. Physiol.*, **21**, 745 (1938).
19. Langmuir, I., and Schaefer, V. J., *J. Am. Chem. Soc.*, **59**, 2400 (1937).
20. Neurath, H., *J. Phys. Chem.*, **44**, 296 (1940).
21. Philippi, G. T., "On the Nature of Proteins," Amsterdam, 1936; *Biochem. J.*, **31**, 513 (1937).
21a. Ramsden, W., *Arch. Anat. Physiol., Physiol. Abt.*, 1894, p. 517; *Proc. Roy. Soc. London*, **72**, 156 (1903); *Z. phys. Chem.*, **47**, 336 (1904).
22. Schulman, J. H., and Cockbain, E. G., *Trans. Farad. Soc.*, **36**, 651 (1940).
23. Wrinch, D. M., *Cold Spring Harbor Symposia*, **9**, 218 (1941).

Physical Changes of Muscle Related to Activity

OTTO MEYERHOF, M.D.

Research Professor of Biochemistry,
Department of Physiological Chemistry, School of Medicine,
University of Pennsylvania, Philadelphia, Pa.

Introduction

Investigators of the chemistry and physiology of muscle in the last two decades have been busy disentangling the chemical reactions concerned with the transfer of energy and with the time course and the implications of this energy transfer; but we should not overlook the fact that muscular contraction in itself is essentially a problem of colloid chemistry, or at least has many aspects of colloidal character. The contraction process can be traced to the change in shape and possibly hydration of protein particles, myosin micelles, in the anisotropic elements of the striated muscle fiber. A change in birefringence of these anisotropic discs coincides with the contraction and relaxation.

The recent work of Astbury[1a] and others on x-ray patterns of myosin and muscle fibers (reviewed by the author in this volume) supplies us with a working hypothesis: the intramolecular changes in the chain-like myosin molecule, closely resembling the α-β transformation of keratin and the supercontraction of hair, are considered responsible for the contraction of muscle.

On the other hand, the study of the energy-yielding chemical reactions revealed some unexpected physical changes in muscle, which undoubtedly must also be explained by a change in the colloidal state of the protein, *e.g.*, the increase in transparency of the muscle during the splitting of phosphocreatine into creatine and phosphate. Moreover, other physicochemical changes, like change of pH, volume constriction, etc. can be only partially explained by the formation or decomposition of intermediary metabolic substances which consist of small "crystalloid" molecules. There always remains a superimposed effect, which has to be attributed to unknown changes in the contractile protein. Therefore, these studies add emphasis to the essential role which the protein plays in the shortening and development of tension during activity.

Of course, colloid chemistry has also some bearing upon the chemical mechanism of the intermediary reactions, insofar as the enzymes concerned are proteins, which in part possess a dissociable prosthetic group, called coenzyme. The relations between the colloid enzyme protein, the coenzyme and the substrate are of too general

a nature to be dealt with comprehensively in an essay on muscle chemistry. Nevertheless, recent findings indicate that some of these enzyme proteins may form a link between the metabolic reactions and the physical change in muscle, or that the contractile protein itself may function as an enzyme in the last step of the energy-yielding reactions.

The thermodynamics of contraction leads to the conclusion that physical changes of the contractile protein are related to great changes in energy. Besides the thermoelastic heat exhibited by a resting muscle in stretching, the time analysis of the heat development of an active muscle reveals a fraction which is due to the dissipation of energy of "tension" during relaxation in isometric contraction.

Since several of these items have already been treated in other parts of this book, I do not propose to discuss them all in detail. I shall restrict myself to an outline of the connection between chemical and physical changes during muscle activity, stressing those phenomena which can be explained only partly by simultaneous chemical reactions.

Outline of the Chemical Reactions Related to Activity

The true mechanism of contraction is not known. The earlier theories of Helmholtz, Dubois-Reymond, Fick, Bernstein, von Fürth, etc. must all be discarded because they are either unfounded or based on misinterpreted experimental evidence. Various chemical reactions concerned with energy production for the mechanical work of muscle have been found since that period of classic physiology. Although these reactions are more or less indirectly connected with the contraction mechanism, knowledge of them is a necessary preliminary to the understanding of the transfer of energy and finally to all discussion regarding the possible mechanism of contraction. Only a brief review of these reactions will be given here. A more comprehensive report can be found in some recent reviews, *e.g.*, Meyerhof,[2, 3, 4] Kalckar,[5] and Lipmann.[6] A fuller discussion was made in German in a review of 1937 by the writer,[2] and the physical aspects were discussed by von Muralt.[8]

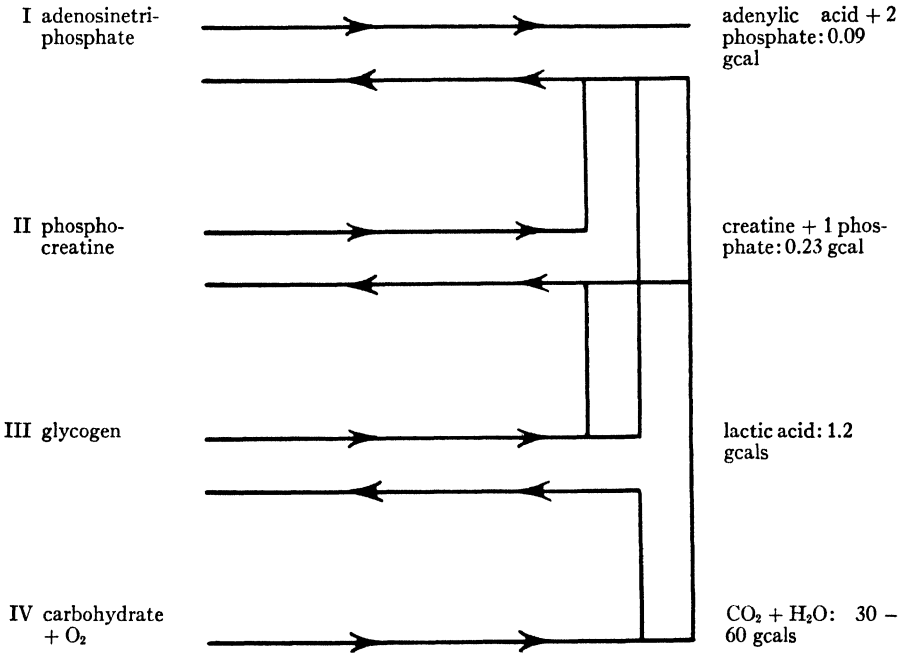
The heat measurements of Hill (1913) and of Hartree and Hill (1921-1928) and the chemical investigations of the writer (1919-1923) established the fact that the fundamental process of activity is anaerobic, and that oxidation serves for recovery. This recovery involves the endothermic resynthesis of substances split during activity, whereby the energy for this involuntary synthesis is supplied by coupling with oxidation. This fundamental fact was recently doubted by Sacks.[9] D. K. Hill [10] in 1940 repeated and extended the experiments of his father, A. V. Hill, and showed again conclusively, that the oxidative heat is really the exact expression of the simultaneous oxidation coupled with resynthesis.

Three different anaerobic reactions were found to be related to activity and also to be linked to one another: (1) splitting of carbohydrate, preformed in the muscle as glycogen and hexose-6-monophosphate, into lactic acid via phosphorylated intermediaries and pyruvic acid; (2) splitting of phosphocreatine into creatine and phosphate; (3) splitting of adenosine triphosphate to adenylic acid + 2 phosphate via adenosine-diphosphate + 1 phosphate.

The coupling of these three reactions with one another and with oxidation is shown in the accompanying diagram, the arrows pointing to the direction in which the reactions proceed. The vertical connections between the arrows correspond to the coupling. The end products of the breakdown reactions are given on the right side of the diagram, together with figures showing the calories of total energy which can be derived from the splitting of the amounts of each primary product present in one gram of an isolated frog muscle under normal conditions. These figures are obtained by multiplying the amount present in unfatigued muscles by the heat of reaction. In this way, one can see that these potential reactions represent energy stores,

which become larger from top to bottom of the diagram, that is, the farther the reaction is away from the contraction mechanism.

The interrelationship of these reactions was first shown in studies on enzyme extracts of muscle, and was confirmed in the muscle *in vivo*. The lactic acid formation restores the phosphocreatine, broken down in the earlier stages of activity, while



the breakdown of phosphocreatine is able to restore the adenosine triphosphate, and the oxidation of carbohydrate can reverse every cleavage reaction of I, II and III. Although it is not completely proved that in muscle activity the partial breakdown of adenosine triphosphate precedes the splitting of phosphocreatine, this is probable for several reasons, and we take it provisionally for granted. While the dephosphorylation of adenosine triphosphate is the first known step in the chain of reactions related to activity, it does not preclude other unknown chemical reactions preceding this step, which in such a case would belong to the recovery instead of to the immediate working phase of muscle. However, the analysis of the physical changes makes it highly probable that this reaction occurs in a very early stage of contraction and development of tension. The lack of success in finding any other reaction preceding this step has induced some authors to connect it immediately with the contractile mechanism. This subject will be discussed further (p. 890).

Time Course of Heat Production and Thermoelastic Heat

Although I am not taking into consideration the electrical changes in the active muscle, which are the expression for the excitation process rather than for the contraction, something must be said regarding the development of heat, already discovered by Helmholtz, which may be taken as an indication of the energy production for the mechanical work. The well-known studies of Hill [11] and of Hartree and Hill [12] revealed the time course of this heat production and supplied in this way

the first tool for studying the coördination of the mechanical and chemical processes in active muscle. From the point of view of colloid chemistry, those phases of heat production have special interest in that the equivalent of heat is not a simultaneous chemical reaction, but a thermoelastic process. It is true that in anaerobic heat,

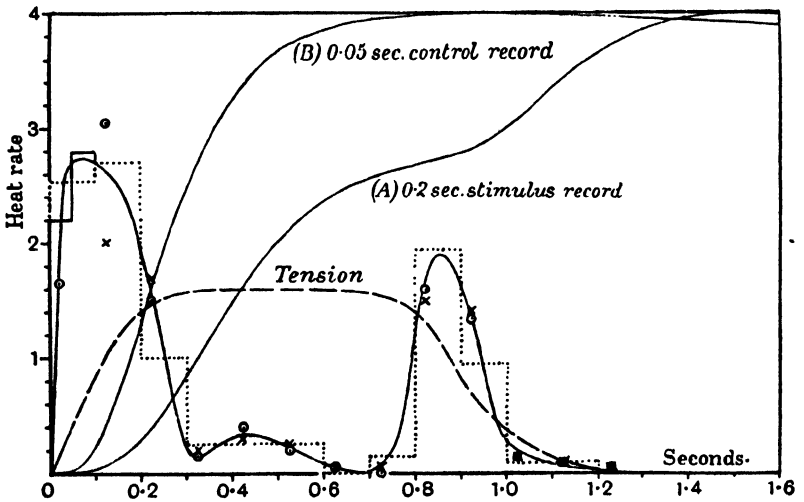


FIGURE 1. Analysis of the heat rate during tetanic contraction. Stimulus 0.2 sec 0° C. Continuous lines, heat; A, galvanometer record; B, curve of control heating. Dotted line, result of analysis in 0.1 sec. By interpolation the smooth continuous curve is obtained, showing heat outburst during tension development, small heat during maintenance of tension and large heat of relaxation. Broken line, isometric tension of muscle. (From Hartree.)

compared with oxidative heat, a fraction appears which has to be ascribed to a physicochemical change, namely, the neutralization heat of the protein "zwitterion" by the acids formed in severe fatigue. This reaction is not of a collodial nature; it occurs also with amino acids and dipeptides. Moreover, the fraction of the anaerobic heat linked to this neutralization is much smaller than was formerly assumed.[13] The muscle does not become acid in the earlier stages of anaerobic fatigue, and so long as no lactic acid is formed, the amount of heat to be attributed to a change in the ionization of protein is negligible.

The thermoelastic properties of the resting and active muscle are of much greater importance for an insight into the physical mechanism. In 1863 Heidenhain showed that a passively stretched muscle is warmed on stretching.[14] This result, doubted for a long time, was fully confirmed by Hill and Hartree in 1920;[15] they found that a resting muscle, passively stretched, behaves like rubber, in that it produces heat while being extended and cools when released. The total change is reversible except for loss of energy by viscosity. This corresponds thermodynamically to a negative coefficient of linear expansion. Woehlisch and Clamann,[16] in a more detailed study, confirmed this phenomenon and measured a negative coefficient of linear expansion for small and medium elongations (up to 30 per cent from the resting length), but for stronger extensions they found the sign of the expansion coefficient reversed. It is not quite clear what role the connective tissue plays in this reverse of sign. In 1937 the problem was taken up again by Meyer and Picken,[17] who investigated the effect of change of temperature on the elastic force exerted by a stretched, unstimulated muscle at constant length. A negative coefficient of linear expansion corresponds to a positive temperature coefficient of the elastic force.

Such a positive coefficient of force was found for moderate elongations and reversed its sign for larger ones, in agreement with Wöhlisch and Clamann. However, under a force of zero and practically resting length, the sign was also reversed—a fact not clearly shown in the results of other authors, but as Meyer and Picken point out, in complete agreement with the behavior of rubber. They consider the latent heat set free during stretching as a sort of heat of crystallization. The flexible primary-valence chain molecules become arranged in bundles by stretching and a lattice is formed, visible in x-ray diagrams of stretched muscle.

Possibly the expansion coefficient is also reversed during active contraction of muscle. Azuma[18] found in 1924 in Hill's laboratory that during the onset of contraction the muscle is cooled by stretching and behaves, not like rubber, but like a normal solid body. It is questionable whether the system is in equilibrium and whether thermodynamic arguments are immediately applicable. But assuming that a new state of equilibrium is established by stimulation, one must conclude that the mobility of the links between neighboring chain molecules has disappeared, that the liquid bonds have become solid linkages, and that contraction is accompanied by a reversible coagulation.

Although the relation to the thermoelastic heat could not be definitely established, the phenomenon, called "relaxation heat" may be mentioned in this connection. Hartree and Hill discovered in 1921 that the initial heat which coincides with the contraction is given out in three distinct phases in the isometric contracting muscle: the first outburst takes place almost simultaneously with the tension development, and is proportional to it; a second, smaller amount is liberated during the maintenance of tension, and is proportional to the duration of the tetanic contraction and dependent upon temperature; a third part, amounting to 30 per cent of the whole, in tetani of 0.1-1.0 sec; coincides with relaxation. This relaxation heat corresponds to the dissipation of the elastic energy, when the tension disappears; it is, therefore, this part of the total heat which can appear as work instead of heat, if the muscle is allowed to shorten and to lift a load, instead of remaining fixed. An example taken from a publication of Hartree and Hill[19] may illustrate this point:

Duration of stimulus 0.4 sec.		Energy in gcm.		
Form of Contraction	Total Initial Heat	Heat of Contraction	Heat of Relaxation	Work Done
Isometric	96	61.5	34.5	0
Work on ergometer	95	89.5	5.5	26

The figures given here may illustrate still an other fact, discovered by Fenn in Hill's laboratory and called Fenn-effect:[19a] the expenditure of energy is greater when the muscle is working by lifting a load then in isometric contraction, where it develops tension without shortening. As the example shows, the initial heat is in both cases 95 gcms. To this energy the work done on the ergometer has to be added: the working muscle develops an extra amount of energy just equal to the work done. The expenditure of energy in the active muscle is not fixed completely at the moment of the stimulus, but is regulated during the active contraction period according to the length and tension of the muscle fibers at every moment.

Time Course of Change of pH

The change of the pH of the muscle and also other physicochemical changes, to be discussed later, have this feature in common with the development of heat, that they reflect *in toto* the chemical changes brought about by activity. Even the time course corresponds roughly to these partly transitory (reversible), partly lasting effects. However, further change of H^+ concentration, due not only to the breakdown of metabolites but to the orientation of the contractile protein, becomes apparent by the use of methods possessing still less inertia and applicable to the muscle during onset, maintenance and disappearance of contraction.

It has long been known that the muscle becomes acid when greatly fatigued. This is due to lactic acid formation; but, contrary to former assumptions, this process has no immediate connection with contraction. In 1930 Einar Lundsgaard [20] discovered that a muscle, poisoned with iodoacetic acid, did anaerobic work without lactic acid

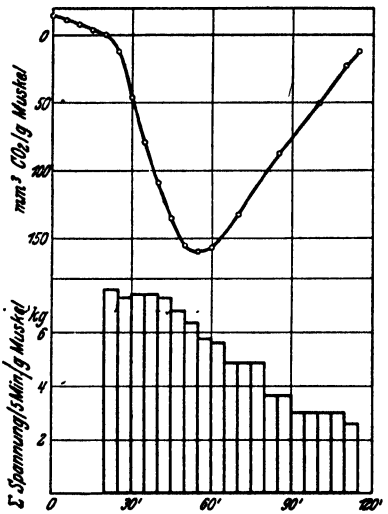


FIGURE 2. Change of pH of muscle during activity. Below, sum of tension development in kg per g muscle and 5 min, with 4 stimuli per min. Above, curve of change of CO_2 pressure. The downward slope of the curve means uptake of CO_2 and increasing alkalinity; upwards, increase in acidity. (From Meyerhof, Möhle, Schulz.)

formation ("alactacid contraction"), as the result of the breakdown of an energetically equivalent amount of phosphocreatine. After the phosphocreatine is completely split, the muscle goes into rigor, when the adenosine triphosphate is split also. In the normal muscle, under anaerobic conditions, part of the broken-down phosphocreatine is immediately resynthesized by coupling with lactic acid formation, as was shown by Meyerhof with Lohmann and Nachmansohn.[21, 22] The change of pH, which normal and poisoned muscle undergo in these phases of anaerobic fatigue, was studied by Lipmann and Meyerhof [23] and with increased accuracy by Meyerhof, Möhle, and Schulz [24] using a manometric method, measuring at the same time the isometric tension developed in activity. The more acid the muscle was at the beginning (regulated by increasing the CO_2 content of the N_2 atmosphere) the more alkalinity was developed during the first 50-100 single contractions. While the poisoned muscle went into rigor after such an amount of activity, and the alkalinity remained unchanged, the normal muscle was able to do further work, the acidity developing at an increasing rate until complete fatigue. The alkalinity during the first part of anaerobic fatigue can be explained quantitatively by the splitting of phosphocreatine, because the second dissociation constant of the phosphoric acid group in this compound is much stronger than in the free phosphoric acid. The splitting of the phosphoric acid groups of adenosine di- and triphosphate, on the other hand, develops acid, because secondary phosphoric acid valencies are liberated; but this is veiled in the rigor of the poisoned muscle by simultaneous splitting of NH_3 from adenylic acid.

While these changes of reaction are of purely chemical origin, the studies of Margaria [25] and Dubuisson [26] in recent years have demonstrated a change of pH attributable to the physical state of the contractile protein and therefore of primary importance for the contraction mechanism. Margaria found that a frog muscle stained with a suitable indicator dye (brom-cresol purple or brom-thymol blue) showed a change in color during contraction equivalent to a shift of pH of about

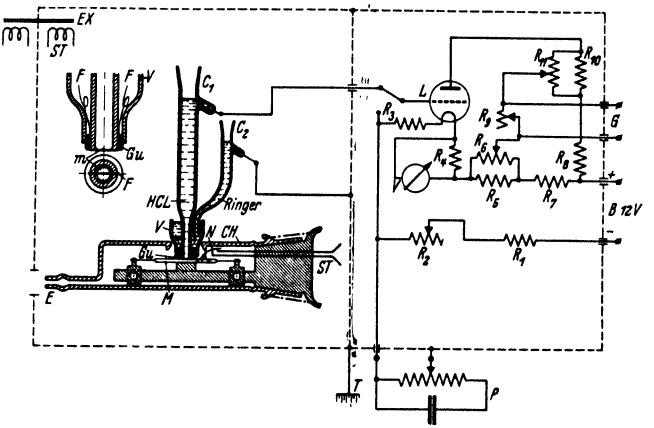


FIGURE 3. Muscle chamber with glass electrode and amplifier set for measuring changes of pH of muscle during contraction and passive stretching. C_1, glass electrode filled with $N/10$HCl and closed on the lower end by a glass membrane; C_2, indifferent electrode filled with Ringer solution and connected with a thread, F, lying on the muscle surface on the same place as the glass membrane, for avoiding disturbances by bioelectrical potentials (*cf.* the magnified drawing on the left.) The thread is fixed by a rubber cuff, (Gu) (m), muscle in isometric position. St, stimulating electrodes; CH, glass chamber with outlet E for filling with Ringer bicarbonate solution and gas mixtures. (From Dubuisson.[62])

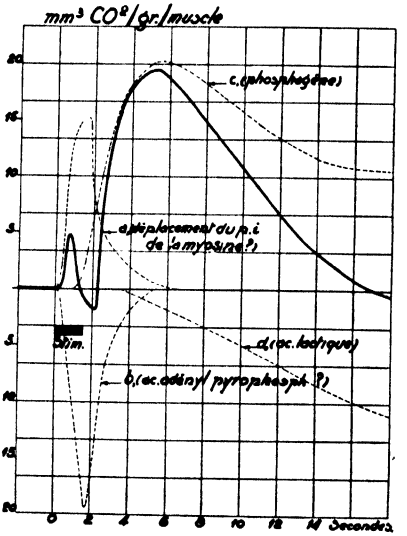


FIGURE 4. Record of galvanometer excursion in the apparatus of Figure 3, and analysis of the curve. Stim, stimulation, 1.5 sec. Continuous line, galvanometer reading. Alkalinization upwards; acidification downwards. The first peak on the alkaline side is probably due to the change of the isoelectric point of myosin during the onset of tension. See the broken line, giving the analysis of the galvanometer curve. The following peak on the acid side is to be attributed to the splitting of adenylpyrophosphate, the next big alkalinization to the splitting of phosphocreatine (= phosphagen), and the final downwards slope to lactic acid formation. (From Dubuisson.)

+0.5 to the alkaline side, and a similar, although less reproducible shift, if passively stretched during rest.

Finally Dubuisson measured with great accuracy the time course and the absolute amount of these changes of pH in different stages of activity, by the use of a glass electrode pressed on the surface of a frog muscle. While for longer time intervals a complete agreement is obtained with the manometric measurement, and while the chemical changes determined simultaneously explain quantitatively the resulting total shift of pH,[27] the measurement during the contraction itself reveals two more variations. In isotonic conditions the wave of alkalinity, attributed to the splitting of phosphocreatine, is preceded by a short deviation to the acid side, which coincides roughly with the duration of the contraction. But in isometric contraction of an unfatigued muscle this acidity change is preceded by a still sharper peak to the alkaline side, which has its maximum before the tension is fully developed. Now there are different reasons for assuming that the acidity wave, independent of tension, is the expression of the reversible breakdown of adenosine triphosphate, which is reversed by the breakdown of creatinephosphate; but the alkalinity peak during the development of tension is the same phenomenon as observed by Margaria. That this phenomenon does not originate in metabolism is evident from the experiments on stretching: if the muscle is completely unfatigued, passive stretching gives alkalinity changes amounting to a maximum of +0.2 pH (from pH 7.22 to 7.42). In this way the results of Margaria are at least qualitatively confirmed by an objective and accurate method. (The apparent shift of pH = +0.5, as given by Margaria, is probably too high.) It seems necessary to assume that the elongation of the protein chain, under stress, changes the isoelectric point. So far a similar phenomenon has not been observed on isolated myosin. Although the sign of the charge is opposite, an apparently related phenomenon has been recently reported in the stretching of nylon fibers.[28] Stretching shifts the isoelectric point approximately 1.2 pH to the acid side (from 3.9 to 2.7).

Volume Change During and after Contraction

The term "muscle contraction" refers to the diminution of length. Indeed, the question of whether the volume of a muscle changes at all during activity remained controversial during the whole period of "classical muscle physiology." As recently as 1925 the Hungarian physiologist Ernst [29] established without doubt a slight reversible diminution of volume during contraction. A more detailed study of this phenomenon in the Heidelberg Institute [30, 31] showed that, in addition to the reversible change, there is a second important effect which continues beyond the contraction. The rather complicated curve of volume change can be explained, in large part, by the change of molecular volume of those substances which undergo transformation during activity. Such volume changes can be measured separately by allowing the same chemical reactions to go on in an enzymatic extract of muscle in suitable dilatometers. Roughly, the curve of volume change can be explained by the chemical reactions already discussed: splitting of phosphocreatine into phosphate and creatine is accompanied by volume constriction of 11.5 cc per mol, splitting of adenosine triphosphate into 1 mol adenylic acid and 2 phosphate with constriction of 20 cc per mol, desamination of adenylic acid to inosinic acid with 1 mol NH_3 formed with constriction of 21 cc, splitting of glycogen into lactic acid with volume dilatation of 24 cc per mol lactic acid. The summation of all these conflicting changes during a series of anaerobic contractions until complete fatigue gives a dilatation amounting to 15.10^{-5} cc per gram of muscle. On the other hand, a muscle poisoned with iodoacetic acid shows only volume constriction, since lactic acid formation is absent, and the sum of the constrictions until onset of rigor amounts to 50.10^{-5} per gram of muscle. But a volume change of only $30\text{-}40.10^{-5}$ cc is calculated from the measured chemical reactions; the excess constriction, therefore, is to be attributed to a slow and

irreversible change of muscle protein. A similar result is found from a calculation of the effect in a normal muscle, since the sum of the chemical changes leads to a dilatation of 50.10^{-5} cc per gram of muscle, while only $15\text{-}20.10^{-5}$ mm are measured; therefore, in this case an unexplained constriction process of 30.10^{-5} cc must be assumed independent of the changes of molecular volume of the metabolic substance.

The foregoing conclusion was also reached by measurements of the volume change in different forms of rigor and during permeation of an acid in absence of rigor. There always remains an additional constrictory effect, not accounted for by the simultaneous chemical reactions, *e.g.*, in the heat rigor only 40 per cent of the dilatation is found, as would follow from the lactic acid formed, corrected for the splitting of creatine phosphate. In the alactacid rigor of muscles poisoned with iodoacetic acid, approximately 180 per cent constriction of that calculated from breakdown of phosphocreatine and adenosine triphosphate is observed. Moreover, the permeation of valeric acid into a muscle is connected with a constriction of 20.10^{-5} cc per gram of muscle without lactic acid formation or any change in the shape of the muscle.

These additional effects of constriction are always larger in muscles suspended in Ringer solution than in paraffin oil; probably they are due to shift of water. The uptake of water from outside is rendered impossible in paraffin oil; nevertheless, water can be exchanged between the fibers and the interfibrous fluid, and between fibrils and sarcoplasma, and so on.

But for any theory of contraction, the transient volume changes which vanish with relaxation seem to be of much more importance. Ernst attributed these volume changes to electroconstriction of water caused by the liberation of K ions in the excitation process. But this is improbable. The volume change begins at the earliest 7σ after the stimulus, where the action current has already passed away. The volume constriction of the isotonically contracting muscle extends over the first half of the shortening time, whereas in isometric contraction the maximum of constriction coincides with the maximum of tension development. A second argument against the excitation theory of constriction lies in the great differences of volume change for the isotonic and isometric single contraction, while the stimulus is the same in both cases.

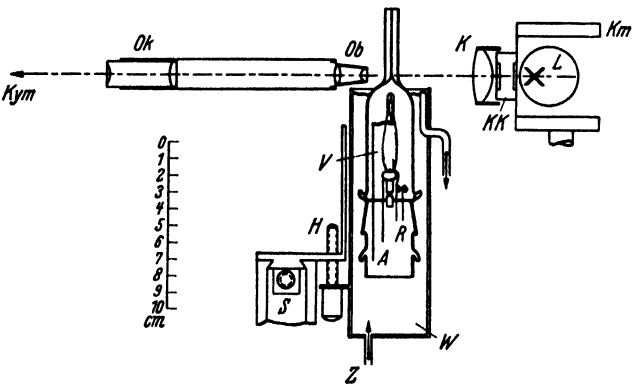


FIGURE 5. Muscle chamber for measuring volume changes during contraction. The muscle (gastrocnemius of frog) is mounted rigidly in the volume chamber (V). R, stimulating electrodes. A, electrodes for registering action currents. The muscle chamber is surrounded by a water jacket, W, maintaining constant temperature. The muscle chamber is filled with Ringer solution or paraffin oil, the capillary with hexane, and the meniscus of the hexane is photographed through a horizontally mounted microscope on a rotating photographion (kym) not shown in the figure. (From Meyerhof and Möhle.)

On the other hand, the reversible volume change is very probably linked to the fundamental physicochemical process, resulting in the tension development of the stimulated muscle. Ernst made the observation, confirmed and more closely investigated in Heidelberg, that for a smooth isotonic tetanus of 50 to 100 stimuli per sec. the corresponding curve of volume constriction is notched and exhibits as many notches as stimuli applied. Indeed, these notches correspond exactly to the summits of the constriction waves of single twitches, incompletely fused. The height of the reversible constriction in normal unfatigued muscles amounts to 10 times the constriction residue in single twitches; and the same is true for muscles poisoned with iodoacetic acid.

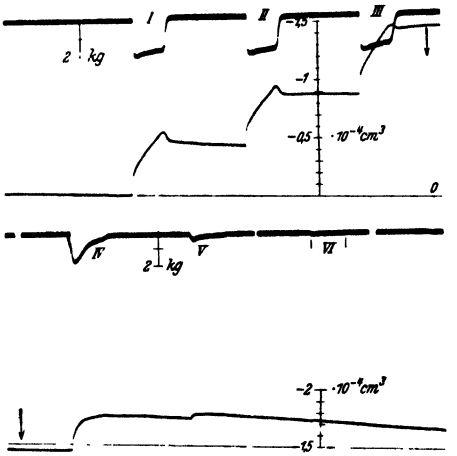


FIGURE 6. Curves of volume constriction of a muscle poisoned with iodoacetic acid during tetanic isometric contraction. For the experiments of this type, a larger volume vessel was used than that shown in Figure 5, and the tension developed by the contraction was recorded simultaneously by means of a torsion spring supplied with a mirror. Upper curve, I, II, III tension development during three 2-sec tetani. Lower curve, volume constriction. Diminution of volume is drawn upwards. Scale refers to the actual change of volume measured. Note: steep change during onset of tension, continuously increasing constriction during maintenance of tension, small decrease during relaxation and a high remainder of constriction which is cumulative. (From Meyerhof and Hartmann.)

The volume change in the gastrocnemius muscle of the frog is not influenced by the initial tension nor by passive extension of the muscle during contraction. But for muscles with parallel fibers it was reported by E. Fischer [82] that with high initial tension the sign of the volume change was reversed, and a transient volume increase occurred. If this result is confirmed, it would stress the argument that so far as the reversible part of the volume change is concerned, physical factors are involved. These factors may consist of a shift of water between the contractile micelles and the surrounding fluid, and the like.

Changes of Transparency

It has been known for a long time that some smooth muscles become opaque or less transparent during contraction. That transparency to light changes during activity in the striated muscle was first observed by von Muralt and von Baeyer in the Heidelberg Institute.[88] The effects which outlast the contraction, can be attributed to the two main processes known to occur, *i.e.*, to splitting of phosphocreatine

which causes the transparency to increase, and to lactic acid formation which causes it to decrease. The latter effect is an indirect one and happens only when the muscle is immersed in Ringer solution. It is caused by the uptake of water. Immersed in a gas or in paraffin oil, the muscle does not become less transparent by forming lactic acid. But the increase in transparency, due to splitting of phosphocreatine, is independent of the surrounding medium and can, therefore, be studied most accurately in a muscle suspended in either nitrogen or oxygen. In both cases curves are obtained which give the clearest and truest picture of the simultaneous turnover of phosphocreatine. The reason why this metabolism results in a change of scattering of light is not known, but the effect has something to do with the colloidal state of protein in the muscle.

A change of transparency can be brought about both by change in scattering and by change in absorption of light. This subject was thoroughly investigated by Buchthal and Knappeis [84] using single muscle fibers. By the use of a very ingenious method—putting the muscle fiber into a ball-photometer ("Kugelphotometer of Ulbricht"), whereby the sum of light scattered in all directions is measured—they can distinguish between increase in scattering and true absorption, and can apply this procedure also to a fiber during contraction.

The light transparency decreases during isometric contractions. This is mainly due to increased scattering, but also partly to increased absorption. About 2 to 5 per cent of the extinguished light of the primary beam is absorbed in accordance with the Lambert-Beer law. Besides the absorption increase during contraction there is a small summing up of absorption after a series of contractions, amounting

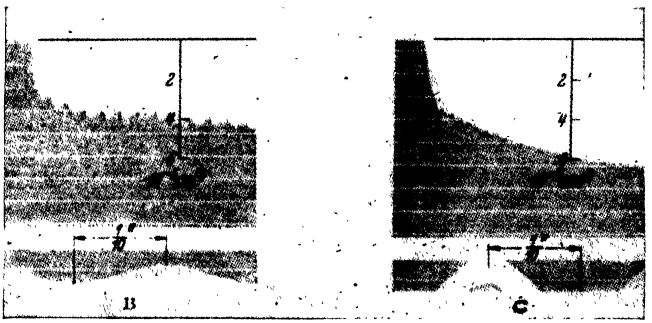


FIGURE 7. Record of volume change on a photographion, rotating with high speed. Volume chamber of Figure 5. Constriction downwards. Isotonic contraction of muscle. B, tetanic stimulation with 50 stimuli per sec; C, with 100 stimuli per sec. Notice the notches corresponding to the number of stimuli. Below, recording of time in 1/10 sec. (From Meyerhof and Möhle.)

to 0.8 per cent of the total absorption. It is not possible to give an unambiguous explanation of these effects, except that elements having the magnitude of wave lengths of visible light change their optical properties, while the length of the fiber remains unchanged. This can be made more precise by recording the diffraction spectra of single muscle fibers and the whole muscle. The cross-striated fiber gives a grating constant of 2.10μ in agreement with the height of the single compartments of 2.18μ. From the form of the spectral lines of this diffraction grating during rest, contraction, and extension of the muscle, it appears that the light becomes more scattered in the direction of longer angles during contraction, as well as during extension, which seems to show that in both cases the micellar elongated elements of the resting fiber become more oriented in parallel.

Change of Double Refraction

As is well known, voluntary muscle is striated because the muscle fibers, or their constituents the fibrils, are built up of discs with alternating optical properties, one layer (the so-called I-layer) being nearly isotropic, but according to W. T. Schmidt[35] actually possessing a weak double refraction; the other layer (so-called A-layer, formerly named Q-) possessing a high degree of birefringence. After this fact had been established by Brücke (1858), von Ebner showed (1882) that this birefringence was reversibly decreased during isotonic contraction—the so-called negative variation of double refraction. Later work proved conclusively that only the anisotropic discs are actively contractile, the I-layers following passively or being stretched in isometric contraction.

The theory which applies to this birefringence was developed by Wiener in 1909. He proved by mathematical deduction that besides a crystalline double refraction, which is shown by a condensed homogeneous substance on account of its atomic structure, another kind of double refraction must exist for a mixed body where elongated micelles or rodlets, small by comparison with the wave-length of visible light, are in regular order imbedded in a second medium of a different refractive index. Under such conditions "form birefringence" results, even when the material

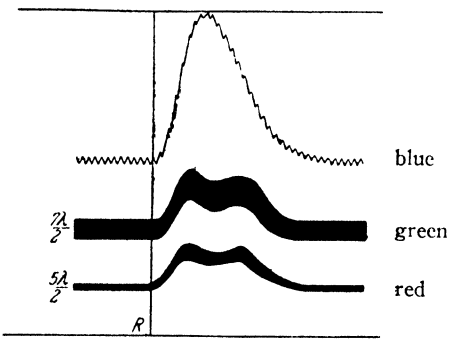


FIGURE 8. Record of change of double refraction of a frog sartorius during isometric contraction. The upper thin curve is the tension curve of the muscle notched by a tuning fork of 250 vibrations per sec for recording time. The upwards displacement of the Fizeau-Foucault fringes to shorter wave lengths corresponds to a diminution of birefringence. Note the two humps during increase and decrease of tension. Vertical line R, stimulus. (From von Muralt.)

of the rodlets and that of the surrounding medium are both singly refractive in themselves. In order to find out the amount of double refraction of form besides a possible amount of crystalline double refraction, it is necessary to substitute for the natural surrounding medium other liquids with different refractive indices. Starting with liquids of high refractive index and descending to lower indices, the total double refraction ought to decrease in the same sense until a minimum value is obtained, from which point on, with liquids of still lower values for the refractive index, the double refraction rises again. The change of birefringence from the start to the minimum corresponds to the form double refraction, while the remaining birefringence at the minimum is crystalline double refraction.

This method was applied to muscle by Schmidt in Giessen[35] and still more thoroughly by Weber and Noll.[36, 36a] It was found that the greater part of the birefringence was due to the micellar structure and only about ⅓ was crystalline. Weber showed that the same was true for threads of undenatured myosin, the contractile protein of the muscle.[38] Some years before, in 1930, Muralt and Edsall[37] had shown that solutions of undenatured myosin exhibited birefringence of flow

and abnormal viscosity, indicating that the solution consists of elongated particles of sub-microscopic dimensions, and they thought it probable that these particles were identical with the micelles of the doubly refractive discs. This was confirmed by the investigations of Weber,[38] who devised a highly valuable scheme for making undenatured myosin threads by squirting the Edsall-Muralt sol through a fine hole into water. These threads are extensible and elastic, and exhibit typical heat rigor by warming to 40° C. They show a birefringence which, in the extended threads, corresponds exactly to that in the anisotropic elements of living muscle, which take up 40 per cent of the volume of the fiber. Accordingly, the birefringence of the extended threads is 2.5 times the overall birefringence of the fiber. The results of these measurements in connection with x-ray and ultracentrifuge studies permit calculation of the dimensions of the micellar pattern in the birefringent discs. According to Weber, we must assume, per square millimeter cross section of muscle, about 15 billion rodlets, each composed of 15 to 20 myosin molecules in a more or less parallel bundle. These fill 70 per cent of the micelles, while the remaining 30 per cent is water, in which the protein molecules are swollen and can shift. The length of the rodlets, identical with the length of the "backbone" of the protein molecules, is about 500 Å, the width 50 Å, while the distance between neighboring micelles is less than 70 Å. The medium in which the micelles are suspended is a salt solution; a thread of myosin, swelling in a solution of myogen, does not take up any of the myogen, while it takes up sugar, because the distance between neighboring rodlets in the birefringent discs, as in the myosin threads, does not allow entrance of foreign protein.

As F. O. Schmitt[39] has pointed out, the rodlets are most probably interconnected by anastomosing chains. Moreover, water plays an important role in maintaining the specific structure, and may, in itself, be responsible for the long spacing equatorial diffractions in x-ray diagrams of fresh muscle. If this is true, the micelles might be considerably thicker than the 50 Å mentioned above.

In 1932, von Muralt in the Heidelberg Institute was the first to measure and to register the change of birefringence during contraction. He photographed on a highly sensitive film the Fizeau-Foucault fringes which appear by prismatic dispersion of polarized light after its passage through a birefringent medium. When the birefringence changes, then the black fringes which result from interference move to other wave-lengths. In isometric contraction when, by special arrangements, all movement of the muscle was carefully avoided, a diminution of about 30 per cent was obtained. The curve of change showed two spikes, one during onset of tension and one during relaxation, while during the interval of maintained tension a saddle was formed. Muralt interprets this curve, which is analogous to the curve of isometric heat obtained by Hartree and Hill[12] as due to the occurrence of two distinct processes which affect the molecular pattern of the micelles, one during rising and one during falling tension.

Recently several investigators have measured change in birefringence of striated and smooth muscle, using more simple and less exact methods than von Muralt. They either used the Babinet compensator mounted on a microscope, or a wedge of mica for obtaining interference fringes, which move with change of double refraction. Bozler and Cottrell[41] confirm the drop of 30 per cent in birefringence in an isometric twitch of a frog sartorius, though only with small initial tensions, and not with zero tension nor with high tension. They attribute it to the shortening of the A-bands and extension of the I-bands in the striated fibrils. However, this shortening in the length would be accompanied by an increase in thickness; therefore, pure geometrical reasons would not lead to a change and we would have to assume the same factor effective here as in isotonic contraction. Buchthal and Knappeis,[42] using single fibers, also find a decrease of birefringence in isometric contraction, and a slowly developing diminution during fatigue, which they attribute to lactic

acid formation and which is absent in muscles poisoned with iodoacetic acid. On the other hand, Fischer [43] and Bozler and Cottrell [41] found that the double refraction of smooth muscle, which concerns the muscle fibers *in toto* instead of the discs in the striated muscle, increases by extending the muscle proportionately to the length, but does not change, in general, with tension. The increase with length can be interpreted as due to an improved orientation of the chain molecules by stretching.

It may be mentioned in this connection that the increase in "impedance," the electrical resistance to alternating current, which was observed during contraction by Bozler (1935)[44] and Dubuisson (1937),[45] may have a theoretical relation to the changes in birefringence. The curves of the changes in impedance, recorded by these authors during isometric tetanic contractions of striated muscle, are very much like those of volume change and are obviously composed of two different processes, one quick and synchronous with the change of tension, the other slow and outlasting the mechanical response. While the latter part probably depends upon the permeability of the cell membranes, altered by the influence of metabolites—splitting of phosphocreatine (cf. Dubuisson)[55]—the quick reversible increase in impedance may be due to a change of dielectric constants. As Fischer [32] has pointed out, Wiener's formula determining the form birefringence of a mixed body has been derived from Maxwell's equations, and is based on the fact that the dielectric constants in the direction of the long axis of the rodlets are different from the constants in the direction perpendicular to the long axis. In consequence, any change of form birefringence of the mixed body must be accompanied by a change in impedance.

Colloidal Behavior of Muscular Tissue and of Isolated Myosin

While the actual change of the colloidal properties of muscle protein during the state of contraction can be detected only by indirect methods, the permanent change in greater degrees of fatigue can be observed by simpler means. It was reported by Saxl [46] in 1906, from Fürth's laboratory, that in the rigor of death, the solubility of muscle protein was diminished when measured by extraction of the muscle tissue with 10 per cent ammonium chloride solution. Decrease of the power to bind water, with increase of fatigue, was observed by Meyerhof and Lohmann (1924),[47] since the volume of the same amount of muscle tissue extracted with 10 per cent NaCl solution on a Buchner funnel becomes smaller with progress of fatigue and "the sheer aspect of the more or less swollen tissue enables us to state how far the muscle was previously fatigued."

A quantitative study of the solubility of muscle protein after onset of rigor and after prolonged stimulation was made by Deuticke (1930).[48] He found that the solubility of protein extracted by 0.11 mol phosphate solution of pH 7.5 decreased gradually with the time of anaerobic fatigue. The protein content of such extracts, always made in the same way, diminished after stimulation of the muscle; 3.3 per cent after ½ hour, 8.8 per cent after 1 hour, 13-17 per cent after 2½ hours of stimulation.

As Weber [38] reported in a preliminary note, this change of solubility concerns the myosin exclusively. Moreover, even slight degrees of fatigue bring about some loss of solubility, as may be seen from the accompanying chart drawn after Weber, and this change in solubility is completely reversible with recovery (cf. also Kamp [38a]). In the comprehensive studies of Weber the fact was definitely established that myosin is the contractile protein, and that it is exclusively located in the doubly refractive elements of the muscle. Myosin has the properties of a high-molecular globulin. It is insoluble in water and in dilute salt solutions, soluble in KCl solutions over 0.04*N*, but also in this case it can be dissolved up to 2 per cent only on account of its high viscosity. With higher concentrations gels are formed.

The accompanying table shows the distribution and physicochemical properties of the different proteins in muscle.

Proteins of Rabbit Muscle, According to Weber [38]

	In % of Muscle-Protein		Isoelectric Point	η	Molecular Weight
	White Muscle	Red Muscle			
Myogen	22	17	6.3	1.04	81,000
Globulin X	22	17	5.1	1.14	140,000-180,000
Myosin	39	39	5.4	10.0	$\sim 10^6$
Stroma	17	27			

Myogen, a protein of albumin nature, is located in the sarcoplasm, globulin X presumably in the isotropic discs. The content of 39 per cent of myosin corresponds exactly to the volume of 40 per cent, taken up by the anisotropic discs.

The molecular pattern of myosin was investigated by Boehm and Weber,[49] Meyer and Picken [17] and recently with much improved technique by Astbury.[1] Since the

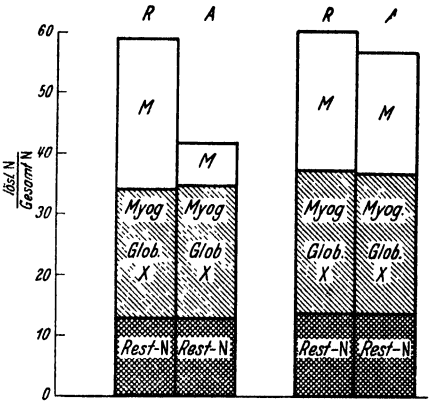


FIGURE 9. Change of solubility of muscle protein during anaerobic fatigue of the muscle. R, resting muscles. A, working muscles. Left, strong fatigue; right, slight fatigue. White area, dissolved myosin; shaded area, myogen and globulin X; dark-shaded area, residual nitrogen. Note that only myosin changes its solubility during fatigue. Scale: soluble N divided by total N. (From Weber.)

latter author has described his results in this volume, I may refer to his paper. According to him, myosin exists like keratin in the three forms, so-called α and β and supercontracted. The contraction of muscle is caused, according to Astbury, by a folding of the backbone of the myosin molecule, consisting of polypeptide linkages. Such "supercontraction" could be produced artificially in sheets of myosin, and while the α-form shows a diminution in length of 50 per cent compared with the stretched β-form, supercontraction further diminishes the length of the α-chain by 25 per cent, so that the final contracted length would be 38 per cent of that of the most extended length.

In the opinion of the writer, these quantitative relations appear as a weak point of an otherwise very attractive hypothesis. The birefringent discs can contract, as may be seen from histological photographs, to at least a quarter, that is by 75 per cent of the normal resting length, while starting with the stretched state of 200 per cent elongation of the relaxed muscle, the contraction in length, compensated for by an increase in thickness, would be to an eighth part. Some authors have reported still greater contractions. According to Studnitz,[50] muscle fibers of Astacus shorten to 30 per cent of their resting length; the actively contracting anisotropic discs, forming about 50 per cent of the fiber, therefore would shorten to 15 per cent of the resting length. Engelmann [51] in an earlier publication observed in striated muscles

of insects contractions of the anisotropic elements to 5 per cent of their initial length! It seems therefore necessary to assume either a much more complicated folding of the long-chain molecules than in supercontracted hair or, besides folding, a gliding of the chains and a redistribution of water, as was assumed by the "classical" school of physiologists.

The stability of myosin sols is strongly influenced by ions. Curiously, K salts are the best suited for keeping myosin in a stable solution in distinction to Na, NH_4, or Li, as was observed by Greenstein and Edsall.[52] L. J. Mullins [53] in Fenn's laboratory discovered recently that myosin combines much easier with K ion than with Na ion. As is well known, the interior of the muscle fiber contains exclusively K, while Na is located in the surrounding interfibrous space. This fact may be explained tentatively by the preferential affinity of myosin to K. This K ion would maintain the myosin in a state of solvation in the micelles of the anisotropic elements.

Hypotheses Connecting the Phosphorylating Mechanism with Contraction of Myosin

It was tempting to the biochemists to test whether the first known chemical process to occur after stimulation,* the partial dephosphorylation of adenylpyrophosphate (= adenosinetriphosphate) would have an influence on the physical state of myosin. Some investigations of recent years seem to show that this is possibly the case.

In 1939 Engelhardt and Ljubimova [54] found that the enzyme adenylpyrophosphatase, by which the labile phosphate groups are split off, is contained in the myosin fraction of muscle protein. Because of this, the authors themselves and several others were inspired to look for a physical change in myosin in the presence of adenylpyrophosphate. In January, 1941 [55] the Russian authors reported that myosin threads made according to Weber and attached to a sensitive torsion balance show a reversible increase in extensibility, which can be measured on the balance, when they are immersed in a solution of 0.0025 mol adenylpyrophosphate. This extensibility is dependent upon the native state of the myosin and is destroyed under the same conditions under which the enzymatic activity of myosin toward adenylpyrophosphate is destroyed.

Independently of this research, Needham, *et al.*[56] measured the double refraction of flow of myosin sols of Edsall-Muralt and the influence of different substances thereon. Adenylpyrophosphate diminishes this double refraction in very low concentration, *e.g.*, 0.004 mol, while other salts have a similar effect in a hundred times higher concentration. The effect of adenylpyrophosphate is reversible, and the return of the initial flow-birefringence is exactly proportional to the dephosphorylation of the adenylpyrophosphate caused by the enzymatic activity of myosin. Moreover, adenylpyrophosphate protects the myosin sol against the formation of an irreversible gel by other salts, allowing the sol to return to its original value of birefringence.

In recent publications from the same laboratory these findings were extended. D. M. Needham [57] and K. Bailey [58] were unable to separate by any means the enzymatic activity from myosin in its native state. The activity decreases in exact proportion to the denaturation of the myosin sol. Pure myosin causes only the first labile phosphate group to be split up, that is, the reaction: adenosine triphosphate → adenosine diphosphate; the dephosphorylation of adenosinediphosphate to adenylic acid is caused by a second enzyme. The authors assume therefore that the enzymatic activity is directly bound to the native myosin.

In their latest publication of July, 1942, J. Needham *et al.*[59] report quantitative

* I am not discussing here the excitatory process. This process may be connected with the release of acetylcholine from the nerve end-plates; but how this release starts the excitatory wave, which propagates over the muscle is not known. Only this propagation induces the fundamental process of contraction itself.

results: "When a myosin sol is treated with adenylpyrophosphate at concentrations as low as 0.004 *M*, the flow birefringence is reduced by some 40-60 per cent, the anomaly is unaffected, and the relative viscosity is reduced by some 20 per cent. The subsequent return of both to their original values, which may take from a quarter of an hour to several hours, is accompanied by the splitting off of inorganic phosphate from the adenylpyrophosphate." This is repeatable on the same sample of myosin by new additions of adenylpyrophosphate. The effect is strongly specific, and is not given by adenosinediphosphate or any other phosphorylated compound.

Kalckar,[60] in a theoretical argumentation which preceded these experiments, discussed the possibility that myosin may take up phosphate from adenylpyrophosphate in the contracted state and would relax when its phosphate is transferred to other phosphate acceptors like hexosemonophosphate. In this way the energy of the dephosphorylation of adenylpyrophosphate would be transferred to the contractile mechanism. Indeed, all considerations of the mechanism of contraction should visualize the thermodynamic aspects of this question. As was found by the thermochemical measurements of Meyerhof and Lohmann,[61] the "energy-rich phosphate bonds" of the adenylic system have accumulated the whole energy (heat and free energy) of the anaerobic breakdown of carbohydrate (24,000 gcals per mol adenylpyrophosphate or per mol lactic acid). The dephosphorylation of the adenylpolyphosphoric acid must transfer this energy to the working system, either directly or indirectly.

The experimental stage of this problem has only just started and it seems premature to draw definite conclusions regarding this mechanism from the above quoted experiments of Engelhardt and Needham. Important progress will probably be achieved in this field in the near future.

References

1. Astbury, W. T., and Dickinson, S., *Proc. Roy. Soc. London*, **B129**, 307 (1940).
1a. Astbury, W. T., "Advances in Eurymology," Vol. III, p. 63, 1943.
2. Meyerhof, O., "Biological Symposia," Vol. III, p. 239, 1940.
3. —, *Ibid.*, V, p. 141, 1941.
4. —, "Symposium on Respiratory Enzymes," p. 3, Wisconsin Press, Madison, 1942.
5. Kalckar, H. M., *Chem. Rev.*, **28**, 71 (1941).
6. Lipmann, F., "Advances in Enzymology," vol. 1, p. 99, 1941.
7. Meyerhof, O., *Asher-Spiro's Ergebnisse der Physiologie*, **39**, 10 (1937).
8. von Muralt, A., *Ibid.*, **37**, 406 (1935).
8a. Hill, A. V., and Meyerhof, O., *Ibid.*, **22**, 328 (1923).
9. Sacks, T., *Am. J. Physiol.*, **122**, 215 (1938).
10. Hill, D. K., *J. Physiol.*, **98**, 207 (1940).
11. Hill, A. V., *Asher-Spiro's Ergebnisse der Physiologie*, **15**, 340 (1916).
12. Hartree, W., and Hill, A. V., *J. Physiol.*, **56**, 367 (1922).
12a. Hartree, W., *J. Physiol.*, **61**, 255 (1926).
13. Meyerhof, O., and Schulz, W., *Biochem. Z.*, **281**, 292 (1935).
14. Heidenhain, R., *Zent. med. Wiss.*, **1**, 545 (1865).
15. Hill, A. V., and Hartree, W., *Phil. Trans.*, **B210**, 153 (1920).
16. Wöhlisch, E., and Clamann, H. J., *Z. Biol.*, **91**, 399 (1931).
17. Meyer, K. H., and Picken, L. E. R., *Proc. Roy. Soc. London*, **B124**, 29 (1937).
18. Azuma, R., *Proc. Roy. Soc. London*, **B96**, 338 (1924).
19. Hartree, W., and Hill, A. V., *Ibid.*, **B104**, 1 (1928).
19a. Fenn, W. O., *J. Physiol.*, **58**, 175, 373 (1927).
20. Lundsgaard, E., *Biochem. Z.*, **217**, 162 (1930).
21. Meyerhof, O., and Lohmann, K., *Biochem. Z.*, **196**, 22, 49 (1928).
22. Nachmansohn, D., *Biochem. Z.*, **196**, 73 (1928).
23. Lipmann, F., and Meyerhof, O., *Biochem. Z.*, **227**, 84 (1930).
24. Meyerhof, O., Möhle, W., and Schulz, W., *Biochem. Z.*, **246**, 319 (1932).
25. Margaria, R., *J. Physiol.*, **82**, 496 (1934).
26. Dubuisson, M., *Arch. Internat. Physiologie*, **50**, 203 (1940).
27. Dubuisson, M. and Schulz, W., *Pflügers' Arch. f. Physiol.*, **239**, 776 (1938).
28. Harris, M., *Nat. Bur. Standards J. Res.*, **26**, 289 (1941).
29. Ernst, E., *Pflüger's Arch. Physiol.*, **209**, 613 (1925).
30. Meyerhof, O., and Möhle, W., *Biochem. Z.*, **260**, 454, 469; **261**, 252 (1933); **236**, 533 (1935); **284**, 1 (1936).
31. Hartman, H., *Biochem. Z.*, **270**, 164 (1934).
32. Fischer, E., "Biological Symposia," Vol. III, p. 211, 1941.
33. Baeyer, E. and von Muralt, A., *Pflüger's Arch. Physiol.*, **234**, 233 (1934).
von Muralt, A., *Ibid.*, **234**, 643 (1934).
34. Buchthal, F., Knappeis, G. G., *Skand. Arch. Physiol.*, **82**, 225; **83**, 281 (1939).
35. Schmidt, W. T., "Abderhalden Handbuch d. biol. Arbeitsmethod," Abt. V, Teil 10, p. 435, 1934.
36. Weber, H. H., *Pflüger's Arch. Physiol.*, **235**, 205 (1934).
36a. Noll, D., and Weber, H. H., *Ibid.*, **235**, 234 (1934).

37. von Muralt, A., and Edsall, J. T., *J. Biol. Chem.*, **89**, 315, 351 (1930).
38. Weber, H. H., *Naturwiss.*, **27**, 33 (1939)
38a. Kamp, F., *Biochem. Z.*, **307**, 226 (1941).
39. Schmitt, F. O., "Biological Symposia," Vol. III, p. 238, 1941.
40. von Muralt, A., *Pflüger's Arch. Physiol.*, **230**, 299 (1932).
41. Bozler, E., and Cottrell, C. L., *J. Cell Comp. Physiol.*, **10**, 165 (1937).
42. Buchthal, F., and Knappeis, G. G., *Skand. Arch. Physiol.*, **78**, 97 (1938).
43. Fischer, E., *J. Cell Comp. Physiol.*, **12**, 85 (1938).
44. Bozler, E., *Ibid.*, **6**, 211 (1935).
45. Dubuisson, M., *J. Physiol.*, **89**, 132 (1937).
46. Saxl, P., *Hofmeister's Beitrage*, **9**, 1 (1906).
47. Meyerhof, O., and Lohmann, K., *Biochem. Z.*, **168**, 128 (1926).
48. Deuticke, H. T., *Pflüger's Arch. Physiol.*, **224**, 1 (1930)
49. Boehm, J., and Weber, H. H., *Kolloid Z.*, **61**, 269 (1932).
50. Studnitz, Z., *Zellforschung u. mikroskop. Anatomie*, **23**, 1 (1935).
51. Engelmann, *Proc. Roy. Soc. London*, **57**, 411 (1895).
52. Greenstein, and Edsall, J. T., *J. Biol. Chem.*, **133**, 397 (1940).
53. Mullins, L. J., *Proc. Am. Federation Biologists*, **1**, 61 (1942).
54. Engelhardt, W. A., and Ljubimova, M. N., *Nature*, **144**, 668 (1939).
55. —, and Meitina, R. A., *Compt. rend. Acad. Sci. U. R. S. S. (Doklady)*, **30**, No. 7, 644 (1941).
56. Needham, J., Shi-Chang Shen, Needham, D. M., and Lawrence, A. S. C., *Nature*, **147**, 766 (1941).
57. Needham, D. M., *Biochem. J.*, **36**, 113 (1942).
58. Bailey, K., *Biochem. J.*, **36**, 121 (1942).
59. Needham, J., Kleinzeller, A., Miall, M., Dainty, M., Needham, D. M., and Lawrence, A. S. C., *Nature*, **150**, 46 (1942).
60. Kalckar, H. M., *Chem. Rev.*, **28**, 7 (1941).
61. Meyerhof, O., and Lohmann, K., *Biochem. Z.*, **253**, 431 (1932).
62. Dubuisson, M., *Am. Physiol.*, **15**, 443 (1939).

The Capillary Circulation

EUGENE M. LANDIS

Professor of Internal Medicine, University of Virginia, Charlottesville, Virginia

The capillary network very appropriately has been called "the workshop of the body." With the heart as pump, the arteries, arterioles, venules, and veins serve as the conducting tubes which transport blood to and from this workshop. The reactions of the larger portions of the circulatory system were known in detail long before the capillaries were studied intensively. Yet, it is only in the capillary network that the circulating blood comes into intimate contact with the tissues. Through the walls of the blood capillaries must pass the water, dissolved salts, oxygen and nutrient substances required by tissue function. Water, carbon dioxide, and other products which arise from tissue activity can be carried away by the circulating blood only after they have passed by diffusion or absorption through these same walls. The brief review * which follows is concerned chiefly with the properties of the capillary network and some of the major forces which control the movement of water, dissolved substances and gases through the walls of the capillaries either from the blood to the tissue spaces, or from the tissue spaces to the circulating blood.

General Considerations

The blood capillaries differ widely in diameter, length and the nature of their connections with the smallest arterioles and venules. Blood pressure and rate of flow vary not only in adjacent capillaries but also in the same capillary from moment to moment. Hence, while the single capillary is the true unit of interchange between blood and tissue, it must be emphasized that these units are by no means identical, and that any "average" figures must take into account the wide range of normal variation. With this reservation, an average mammalian capillary can be said to be approximately 0.25 to 0.75 millimeter long and 2 to 15 microns in diameter, depending upon its state of contraction. The wall of the capillary is composed of thin, plate-like endothelial cells placed edge to edge to produce an uninterrupted

* In the interest of brevity the references given are limited to a few monographs or comprehensive reviews in which extensive bibliographies can be found.

membrane about one micron in thickness, except for the localized bulge containing the nucleus of each endothelial cell. The apposed edges of the endothelial cells are bridged, or made tight, by an intercellular substance or cement concerning which little is known except that, although usually impervious, it will under certain conditions permit particulate matter or even leucocytes to pass through. These bodies apparently pass between the endothelial elements; the openings in the intercellular substance which permit their passage; they then close again tightly without disturbing the movement of fluid to any great extent. In some diseases, however, this intercellular substance seems deficient, and minute or even larger hemorrhages can occur. It should be noted that deficiencies in this intercellular substance are not necessarily associated with abnormalities in the function of the remainder of the capillary wall.

The endothelial cells, on the other hand, normally permit water and small molecules to pass easily, but prevent molecules of colloidal dimensions from passing in either direction. When these endothelial cells are injured their permeability increases to such an extent that molecules of colloidal size, *e.g.* proteins, can pass easily, but particulate bodies such as India ink or erythrocytes are still retained as long as the membrane is not physically torn or ruptured. Therefore, even in the capillary unit itself the functions and properties of the intercellular substance differ from those of the endothelial membrane proper. It is the endothelium which is predominantly concerned with interchange between the circulating blood and the tissues.

For many years the capillary network was regarded merely as a system of thin-walled vessels whose function it was to conduct blood between the tissue cells in whatever quantity the arterioles might supply. According to this concept, flow through the capillaries was entirely dependent upon the constriction or dilatation of the arteries and arterioles. The extensive studies of Krogh,[3] and Lewis [5] have summarized an impressive mass of evidence, showing that the capillaries are independently contractile, and that they are capable of responding individually in a delicate manner to the circulatory needs of the immediately adjacent tissues.

Krogh's original observations led him to believe that the endothelium was inert and that the contraction of the capillary was produced by certain Rouget or adventitial cells, scattered along the capillary wall. Recent evidence indicates that, in certain animals at least, the endothelium is also capable of contraction. For the present discussion the exact location of this contractile power is a matter of academic interest. The important fact is that, while arterioles control blood flow to a considerable extent, the capillaries are also able to respond separately and individually to localized tissue needs.

Capillaries dilate in response to local heat, accumulation of carbon dioxide or acid metabolites, local injury, drugs such as histamine, and nervous impulses induced by the axone reflex, or by stimulation of the vasodilator fibers of the sympathetic nervous system. They constrict in response to local cold (except excessive cold which produces injury and thereby dilatation), very slightly alkaline solutions, vasoconstrictor impulses from the sympathetic nervous system, hormones such as epinephrine or pituitrin, and absence of carbon dioxide or acid metabolites.

Thus the diameter of the capillaries is determined by interaction between (a) composition of the tissue fluid surrounding the individual capillaries, (b) composition of the blood flowing through the capillaries, (c) nervous stimuli, and (d) certain naturally occurring hormones. Even with this delicately and locally adjustable system, the blood stream, which serves as a transporting medium, is still separated from the tissues it supplies by a single layer of endothelial cells forming the walls of the capillary vessels. Obviously the interchange of substances between blood and the fluid bathing the tissues must depend fundamentally upon the properties of this membrane and the forces acting upon the fluid within and without the membrane.

Factors involved in fluid movement

(a) **General nature of the capillary endothelium.** While the cross-sectional area of a single capillary is extremely small, the collective cross-sectional area of all the capillaries in the body is roughly one thousand times greater than that of the aorta. Blood moves through the aorta at an average speed of 0.5 meter per second, but the greater cross-sectional area of the capillary network slows the average movement of capillary blood to 0.5 mm per second. In individual capillaries, however, flow is extremely variable, ranging from 0 to 3 mm per second, depending upon constriction or dilatation of the capillary and of its supplying arteriole.

It follows that the area of capillary wall available for fluid interchange is relatively enormous. Krogh [8] estimated that if a man's muscles weigh 50 kg and his capillaries number 2,000 per sq mm, their total surface would be 6,300 sq meters. In the horse, dog and frog, 1 cc of blood is exposed in the capillary network to a filtering surface of 7,300, 5,600 and 2,700 sq cm respectively. Krogh estimated also that the total surface of capillary wall in the body of an average-sized man must exceed 6,300 sq meters, or 68,000 sq ft. In other words, the collective area of vascular endothelium in the adult human body can be visualized as a microscopically thin membrane three feet wide and over four miles long. Obviously, very rapid interchanges are favored by areas of this magnitude.

The capillary wall is also highly permeable, since it permits fluid to pass much more easily than do other cellular membranes so far studied quantitatively. Strictly speaking, the permeability of a membrane should be defined in terms of units of volume or mass passing through unit area and thickness of membrane in unit time under the influence of unit hydrostatic or unit osmotic pressure. Using the term "permeability" in this limited sense, the capillary endothelium in the mesenteric vessels of the frog can be compared roughly with certain other cellular membranes.

It has been shown that, under an osmotic pressure of one atmosphere, water passes through the membrane of the sea-urchin egg cell at the rate of 0.1 cubic micron per square micron of surface per minute. The membrane of the human erythrocyte is more permeable; under this same pressure water passes at the rate of 3.0 cubic microns per square micron of surface per minute. According to microinjection experiments carried out on single capillaries of the frog's mesentery, a change in hydrostatic pressure of 5 cm of water modifies the filtration rate by 0.03 cubic micron per square micron of surface per second, which corresponds to a filtration rate of 370 cubic microns per square micron of surface per minute under a pressure of one atmosphere. These calculations indicate that fluid filters through the frog's capillary wall over 100 times as easily as through the membrane of the human erythrocyte, and over 3,000 times as easily as through the membrane of the sea-urchin egg cell. This high grade of permeability enters into any consideration of the physiology of the capillary network, one function of which is to distribute water and dissolved substances throughout the body efficiently and rapidly.

The significance of an extensive filtering area combined with great permeability can be illustrated by a calculation which, on account of the assumption made, is admittedly more interesting than important. Assuming that the human capillary wall in general possesses a permeability to fluid similar to that of the frog's mesenteric capillaries, the total plasma volume of a man would be filtered through his calculated 6,300 sq meters of capillary surface within ten seconds at a capillary blood pressure of 10 mm Hg if there were no force retaining fluid within the blood capillaries. Actually, the vasomotor system normally prevents the entire peripheral vascular bed from opening simultaneously and, in addition, the colloidal constituents of blood plasma, by developing a colloidal osmotic pressure, limit the loss of fluid from the blood stream. Only in widespread injury of the vascular endothelium would it be possible for fluid to leave the vascular system at such a rapid rate.

The movement of fluid between blood and tissue spaces varies in both rate and

direction. Hydration, dehydration, change in posture and exercise affect the total volume and the water content of the blood, but the variations are kept within safe physiological limits except under the most drastic conditions. Blood volume responds quickly to physiological needs and under pathological conditions shows great alterations, but the healthy organism strives to keep the general level constant. When water is taken by mouth, its absorption from the gastro-intestinal tract rapidly produces dilution of the blood during the period when the extra fluid is being transported throughout the body. Before the excretion of excess water by the kidneys begins, some of the excess fluid passes into the tissues. If the excretion of water by the kidneys is delayed for any reason, both hydremia and temporary storage in the tissues are accentuated. Limitation of water intake is accompanied, on the contrary, by concentration of the blood, but the water content of the blood again rapidly returns to normal as soon as the fluid is administered.

Muscular movement also modifies the passage of fluid through the capillary wall in that blood loses fluid very rapidly as it passes through active muscle and the flow of lymph from the region is increased many fold. During rest, loss of fluid from the blood ceases and lymph flow diminishes. Changes in position affect the movement of fluid from the blood to the tissues; quiet standing diminishes blood volume and elevates the concentration of the plasma proteins, because high intravascular pressure in the dependent lower extremities produces greater filtration and accumulation of fluid in the extravascular spaces of the legs. In the recumbent position this extravascular fluid is reabsorbed and blood volume returns to the earlier level. Under physiological conditions, therefore, fluid is moving at varying rates, at one moment inward, at another outward, through an extensive and highly permeable system of capillary walls. Nevertheless, gross variations in blood volume and in the volume of tissue fluid are avoided.

To explain the equilibrium existing between blood and tissue fluid under these diverse conditions, the most attractive hypothesis, and the one for which there is most evidence, is that advanced by Starling.[7] The essential parts of this hypothesis can be summarized under four heads. (a) The blood flows through the capillary network under a pressure gradient, which if unopposed, filters fluid from the blood plasma through the capillary wall into the tissue spaces. (b) The capillary wall, though quite permeable to water, salts and simple organic compounds, is in most tissues relatively impermeable to the plasma proteins. (c) Those substances (such as salts, urea, creatinine, glucose, etc.) which pass easily through the capillary wall and are present in approximately equal concentration in blood plasma and tissue fluids, cannot exert a permanent osmotic pressure across the capillary wall and should not, therefore, affect the distribution of fluid, except temporarily. (d) The plasma proteins, by reason of their greater molecular dimensions, are retained by the capillary wall and thus develop a small, but physiologically important, colloid osmotic (or oncotic) pressure which, if unopposed, leads to absorption of fluid from the tissue spaces.

Starling postulated that the constancy of blood volume and of tissue fluid volume depends primarily upon the balance between capillary blood pressure and the colloid osmotic pressure of the plasma proteins, acting in association with a capillary wall which is impermeable to protein. In those areas where capillary pressure exceeds the colloid osmotic pressure of the blood, fluid should be filtered from the blood into the tissue spaces. In other areas where capillary pressure falls below the colloid osmotic pressure of the blood, fluid should be absorbed from the tissue spaces into the circulating blood.

This concept, in which physical forces are the prime agents controlling the movement of fluid through the capillary wall, can be applied only if the endothelium of the capillary network acts like an inert membrane. From the morphological point of view it is not likely that a membrane having the thinness [8] and the structural

simplicity of vascular endothelium could have a secreting function. The available evidence led Krogh[8] to state, "when I review all the facts that have come to my notice, I have no hesitation in saying that there is no trustworthy evidence of the capillary having any power of hindering or favoring the passage by diffusion of all kinds of crystalloids through the endothelium."

This view is supported by recent studies of the concentrations of electrolytes and simple organic substances in blood serum or plasma and in extravascular fluid, such as edema fluid. It has been shown also that the electrolyte concentrations observed when the capillary wall separates plasma and edema fluid are identical with those observed when the same fluids are permitted to come into equilibrium across a collodion membrane which is permeable to water and salts but impermeable to protein. It has been concluded that the normal living capillary wall does not act differently from a collodion membrane in this respect. This does not mean, however, that the concentrations of electrolytes are absolutely identical on the two sides of the capillary wall. The circulating plasma contains a considerable amount of protein, whereas the extravascular fluid contains less protein. Most of the small differences between electrolyte concentrations on the two sides of the capillary wall, or on the two sides of a collodion membrane, have been shown to depend upon differences in the protein content of the two fluids in that the observed concentrations are in general accord with the Gibbs-Donnan equilibrium. Certain additional minor discrepancies observed in the living organism, and not wholly explained by the Gibbs-Donnan equation, are probably due to as yet undefined modification of activity coefficients by the complex mixture of substances found in blood plasma or serum. In any case, the capillary wall cannot be responsible for these discrepancies because they persist when the same two fluids are separated merely by a collodion membrane. The non-electrolytes so far studied are found in approximately equal concentrations on the two sides of the living capillary membrane, or of the collodion membrane, as soon as true equilibrium has been reached.

All available evidence indicates that the capillary wall can be regarded as an ultrafilter comparable to an artificial membrane which is impermeable to protein. As Starling suggests, it follows that the osmotic pressure of the proteins must tend to absorb water from the tissue spaces, and that this absorbing force cannot be balanced by any other force than capillary blood pressure. Since this membrane has the physical characteristics of an inert (in the sense of non-secreting) membrane which is permeable to water and crystalloids, but relatively impermeable to plasma proteins, it is obligatory to regard physical forces as important factors in the movement of fluid through the capillary wall. It is also necessary to examine critically so-called "changes in capillary permeability" in which these physical forces have not been carefully controlled. The important physical factors to be considered are (1) capillary blood pressure, (2) the colloid osmotic pressure of the blood, (3) the degree to which the capillary wall is impermeable to proteins under normal conditions and after injury, and (4) the role played by tissue pressure and by the lymphatic vessels.

(b) **Capillary blood pressure.** Capillary blood pressure was originally measured by numerous modifications of the "indirect" technique in which, by means of a capsule or other device, external pressure was imposed upon cutaneous or more deeply placed capillaries to determine that pressure which modified skin color or interfered with the flow of blood through the vessels while they were being observed under a microscope. The spread of the observed values, the arbitrary nature of the criteria adopted to indicate balance between the external and internal pressure, and the lack of agreement between results obtained by different methods indicated that more exact determinations were necessary. This was finally accomplished by the introduction into individual capillaries of minute glass pipettes or needles having a diameter of four to ten microns and filled throughout with physiological salt solu-

tion.[4] These micropipettes were connected to an apparatus to measure the pressure required to balance directly the intravascular pressure. These determinations have shown, in accordance with the Starling hypothesis and with theoretical considerations, that some of the peripheral resistance to blood flow is situated in the capillaries. The blood as it flows through the vascular system sustains little drop in pressure until the minute arterioles are reached, in which the drop in blood pressure amounts to approximately 70 per cent of the total. The capillaries account for approximately 30 per cent of the peripheral resistance and the gradient of blood pressure penetrates into the capillary network proportionately.

Average pressure in the arteriolar capillaries in four species, including man, rat, guinea pig and frog, was shown to be greater than the colloid osmotic pressure of the blood, while the average pressure in the venous capillaries was less than the colloid osmotic pressure of the blood. Under average conditions the relationship between capillary blood pressure and the colloid osmotic pressure of the blood indicates that fluid is filtered from the blood during its passage through the arteriolar portion of the capillary network, while absorption takes place in the venous portion of the capillary network. Though this represents the average condition, fluid movement in individual capillaries will diverge widely from the average in accordance with changes in capillary pressure, diameter and flow.

It has been mentioned that capillary blood pressure is an extremely variable figure, changing in the same capillary from moment to moment and differing widely even in adjacent capillaries connected to the same arteriole. Arterial and arteriolar constriction increases the peripheral resistance and capillary pressure becomes lower throughout the affected area, irrespective of the diameters of the capillaries themselves. Pulse pressure decreases and a lower capillary blood pressure favors absorption of fluid. Dilatation of the arterioles, on the contrary, shifts the peripheral fall of blood pressure toward the capillary network, usually elevates capillary pressure, increases the pulse pressure in the capillary vessels and increases the filtration of fluids. Any rise of venous pressure elevates capillary blood pressure, as does also the depression of tissues below heart level owing, in the latter instance, to the hydrostatic pressure of the blood in the vertically placed veins. Since blood pressure in an entire capillary, or even in a whole network, may be at one moment far above, and at another far below, the colloid osmotic pressure of the blood, massive filtration or massive reabsorption respectively may be observed in large tissue areas. With average resting circulation, however, filtration and absorption are roughly in balance.

The micro-injection technique has made it possible also to measure the effect of capillary blood pressure on fluid movement through the walls of single capillaries. Whenever, in a given capillary, the hydrostatic pressure is greater than the colloid osmotic pressure of the blood, the volume of plasma within the observed vessel becomes less as fluid is filtered from the blood through the capillary wall into the tissue spaces. On the contrary, in those capillaries in which blood pressure is lower than the colloid osmotic pressure of the blood, fluid moves in the reverse direction and absorption is indicated by increase in the volume of the plasma in the observed capillary. When capillary blood pressure is equal to the colloid osmotic pressure of the plasma proteins, fluid moves neither inward nor outward. Above and below this point the rate of fluid movement is directly proportional to the difference between these two forces. From these relationships the permeability of the capillary wall was calculated.

In the human forearm, when capillary blood pressure is raised by congesting the veins by known pressures, the accumulation of fluid in the extravascular spaces is directly proportional to the hydrostatic pressure exerted on the blood as it passes through the capillary network. This filtration can be measured either by increase in arm volume or by concentration of the cells produced by loss of fluid from the blood

plasma. Increasing venous pressure by 1 cm of water increases the filtration of fluid by 0.0023 cc per minute per 100 cc of forearm tissue.

In summary, the movement of fluid from blood to tissue spaces varies considerably with circulatory adjustments and particularly with capillary blood pressure. Active hyperemia and high capillary blood pressure increase the filtration of fluid. Elevating venous pressure is always associated with increased filtration, the rate of filtration being proportional, within certain limits, to venous pressure and capillary blood pressure. Low capillary blood pressure leads to absorption of fluid.

(c) **The colloid osmotic pressure of the blood.**[9] Starling, in 1896, was the first to measure the magnitude of the colloid osmotic pressure produced when the plasma was enclosed within a membrane which was impermeable to protein and at the same time permeable to water and dissolved salts. In man the most recent figures indicate that the colloid osmotic pressure of the plasma proteins lies between 24 and 26 mm Hg, while average blood pressure is 32 mm Hg in the arteriolar portion of the capillary network and 12 mm Hg in the venous capillaries. In lower animals the colloid osmotic pressure of the blood is less, but capillary blood pressure is also correspondingly lower. This balance between colloid osmotic pressure and capillary blood pressure offers presumptive evidence that the plasma proteins are an important factor in the control of fluid movement.

The association of a low plasma protein percentage, and therefore a low colloid osmotic pressure of the blood, with some but not all forms of edema in man has been well established, particularly for certain stages of renal disease and in the edema of malnutrition or starvation. Edema of this particular type has, in fact, been produced in animals by reducing the colloid osmotic pressure of the blood by removing the plasma proteins and returning the red cells suspended in salt solution. Conversely, the addition of relatively large quantities of serum to salt solution with which a surviving limb is being perfused will delay, though not prevent, the appearance of edema which always follows the perfusion of simple salt solutions. To prevent edema of a perfused tissue indefinitely it is necessary also to assure adequate oxygenation of the perfused extremity for reasons that will be mentioned below. In the treatment of one type of hypoproteinemic edema a colloidal polysaccharide, acacia, has been used with some success in suitable instances. This substance acts merely by increasing temporarily the colloid osmotic pressure of the blood and thereby blood volume, so that diuresis is induced.

The effect of changing the colloid osmotic pressure of the blood can be compared quantitatively with that of changing venous pressure. Unit rise (1 cm of water) of venous pressure increases filtration through the capillary network of the human forearm by between 0.0023 and 0.0033 cc per minute per 100 cc of forearm. Conversely, a unit rise (1 cm of water) of colloid osmotic pressure diminishes filtration by 0.0027 to 0.0045 cc per minute per 100 cc of forearm.

In summary, available evidence, direct and indirect, indicates that the colloid osmotic pressure of the blood tends to retain fluid within the vascular system and acts quantitatively in opposition to capillary pressure. In accord with the view that the capillary wall is an inert (*i.e.*, non-secreting) membrane, the effects of capillary pressure and colloid osmotic pressure are opposite in sign but have the same order of magnitude. The sustained balance between these two forces will depend, of course, upon the efficiency with which the capillary wall prevents the passage of the plasma proteins.

(d) **Impermeability of the capillary wall to the plasma proteins.** The permeability of the capillary wall to protein could be defined easily and exactly if a true capillary filtrate could be obtained and analyzed for comparison with blood plasma. The collection and analysis of blood plasma offers no difficulties, but samples of normal capillary filtrate have not so far been obtainable because of the minute volume of this filtrate under normal conditions. Information concerning the permeabil-

ity of the capillary wall has necessarily been obtained chiefly by indirect methods. For this reason the three fluids upon which measurements have been made are (a) blood plasma, (b) the tissue fluid bathing the external surface of the capillary, and (c) lymph collected from small or large lymphatic ducts.

Tissue fluid consists of the original capillary filtrate mixed more or less with capillary filtrate which has been modified by partial reabsorption of fluid. Under these conditions any protein present in the original capillary filtrate would not be reabsorbed, and the resulting mixed tissue fluid may contain much more protein than was present originally in the pure filtrate.

Normally, because of active reabsorption, the volume of tissue fluid is also too small to permit collection of samples. In the edema of advanced nephrosis and malnutrition, the colloid osmotic pressure of the blood is usually much lower than capillary blood pressure. Reabsorption is then impossible and tissue fluid accumulates in the tissue spaces in large amounts. The composition of this fluid should approach that of the original capillary filtrate because reabsorption is minimal. The protein content of edema fluids, of widely differing pathogenesis, ranges from approximately 0.04 to 1.00 per cent. Very frequently edema fluids of mechanical origin contain less than 0.25 per cent protein, and when this figure is compared with the plasma protein percentage in the same patients, it appears that the capillary wall can under suitable conditions retain 95 per cent or more of the blood protein. The same conclusion is reached when the protein content of blood is compared before and after it has passed through capillaries subjected to venous congestion in order that reabsorption may be eliminated. According to this indirect method the capillary filtrate contains an average of 0.3 per cent protein, again indicating retention of approximately 95 per cent of the blood protein if anoxemia is avoided.

Analyses of lymph offer a third source of information on this point, but interpretation of results is slightly more difficult because the protein content of lymph varies from 0.2 to 4.7 per cent, depending upon the conditions under which it is collected. Lymph, like tissue fluid, contains (a) original capillary filtrate plus (b) the protein from any capillary filtrate of which the fluid and salts have been reabsorbed. Therefore, unless lymph is collected under conditions arranged to exclude absorption by the venous capillaries, the protein content will be higher than the original filtrate. The composition of lymph may furthermore be affected by its passage along the lymphatic channels or through lymph glands.

According to Drinker and Field,[1] lymph flow from a resting tissue is so slow that samples must be obtained by massage, and then contain as much as 4.7 per cent protein. It is likely that reabsorption of fluid is very active under these conditions. Contraction of the muscles of a limb increases the rate at which lymph flows along the lymphatics and diminishes the protein content to between 0.5 and 1.5 per cent. Red cells may also be found in this lymph. If the effects of exercise are avoided and at the same time reabsorption is excluded by raising venous pressure in the resting extremity, lymph flows more rapidly and simultaneously the protein content drops to as low as 0.2 per cent. Under these rigid conditions lymph probably reflects accurately the composition of the original capillary filtrate. If this interpretation is correct, indirect studies by 3 methods indicate that while the capillaries are not absolutely impermeable to plasma protein they nevertheless retain all but about 5 per cent of the protein in the circulating blood.

The fact that small amounts of protein are found in edema fluid and in lymph obtained during slight venous congestion invites inquiry into the manner in which protein leaks through the capillary wall. Protein might pass in very low concentration through the entire endothelial surface, or unmodified plasma might leak through a few capillaries which are abnormally permeable owing to trauma, or to a natural process of ageing. If the first possibility were correct, lymph would probably contain more albumin than globulin, because of the smaller molecular size of the albumin

fraction of the plasma proteins. If the second condition existed, the albumin-globulin ratio of lymph would tend to resemble that of blood. Examination of the albumin-globulin ratios of lymph and edema fluid are unfortunately not uniform enough to indicate the size of the pores through which protein passes. In most instances the albumin concentration is disproportionately high, indicating the presence of many small pores, but in other instances the ratio is practically that observed in the circulating blood.

The statements made so far apply to the capillaries of the extremities. In other tissues, particularly the intestines and liver, greater permeability is observed and lymph from these areas, even when reabsorption is avoided, contains from 3 to 4 per cent protein. Also it is only while the capillary wall is normal that it remains relatively impermeable to protein. Injury by chemical substances, heat or inflammation increases capillary permeability conspicuously. Under these conditions the capillary filtrate, edema fluid and lymph may contain almost as much protein as found in blood plasma. While the colloid osmotic pressure of the plasma, as measured *in vitro* across a protein-tight membrane, is normal, yet in the injured capillaries the effective colloid osmotic pressure is reduced in proportion to the leakage of protein. The ability of the plasma proteins to retain fluid in the affected capillaries is more or less diminished, filtration increases, and edema results.

In summary, the physical factors involved in the movement of fluid and dissolved substances through the vascular endothelium are so complex that changes in capillary permeability should be described only after careful control of physical forces known to influence the movement of substances through an inert membrane. Analyses of edema fluid, blood and lymph indicate that the capillary wall is relatively, but not absolutely, impermeable to protein. In the extremities the capillary endothelium normally retains at least 95 per cent of the total plasma protein. In some regions, however, permeability to protein is greater than this. Severe injury of any capillary network increases the permeability to protein so that all the plasma proteins pass freely, and filtration is thereby increased.

(e) **Diffusion and the permeability of the capillary wall to electrolytes.** The exchange of highly diffusible solutes to which the capillary wall is permeable need not follow the current of water during either filtration or absorption. When hypertonic solutions of electrolytes are introduced intravenously the electrolyte diffuses through the capillary wall toward the tissue spaces even while fluid is moving toward the blood. On account of the time required for the diffusion of electrolytes, fluid movement through the capillary wall is modified temporarily. This effect is transient and disappears when the concentration of electrolytes is once more equal inside and outside the capillary wall. The time required to attain equilibrium after the injection of hypotonic or hypertonic solutions depends upon the rapidity with which the substance diffuses through the capillary wall. According to Keys,[2] the approximate rates of diffusion across the capillary wall in man seem to be: H_2O > urea > K^+, Na^+, Cl^-, NO_3^- > Ca^{++}, Mg^{++}, phosphate glucose, $SO_4^=$, SCN^- > sucrose. This order is roughly similar to that observed for gelatin, Cellophane and collodion membranes. Diffusion is doubtless aided by a current of water moving in the same direction but whether diffusion can be hindered appreciably by fluid moving through the capillary wall in the opposite direction is not known, though from rough calculation this seems unlikely. The independence of fluid movement and the diffusion of electrolytes is perhaps best illustrated by the behavior of certain dyes which have been used extensively to investigate the resistance which the capillary wall offers to the passage of substances whose molecular size lies between those of plasma protein and crystalloids. The passage of dyes can be followed easily since they are usually visible in low concentration. Interpretation of results, on the other hand, is often difficult owing to complicating factors such as electrical charge, selective staining, adsorption, toxicity, decolorization or deposition in the tissues.

This subject of diffusion of dyes through the capillary wall studied sporadically for decades, was investigated in considerable detail by Rous and his co-workers,[6] who injected acid vital dyes intravenously and then studied the rate and intensity of staining of tissues about the capillaries and veins. They found that highly diffusible dyes escape so rapidly from the capillaries of mammalian muscle that the tissue surrounding the arterial end of the capillary is colored even before the venous end is reached by the dye-stained blood. Less diffusible dyes emerge all along the muscle capillaries, but in greater amount toward the venous end. Poorly diffusible dyes at first fail to appear outside the arteriolar portion of the capillary network but escape in increasing amount in the venous portion. Eventually in all cases diffusion causes the tissue to be evenly colored.

The preferential passage of poorly diffusible dyes through the walls of the venous capillaries and venules is but little affected by the changes of circulation and fluid movement which accompany muscular activity, plethora, severe hemorrhage, nerve section, shock, dehydration and injections of epinephrine and pituitrin. The independence between fluid movement and this diffusion is indicated further by the fact that capillary pressure can be reduced almost to zero and still dyes pass through the endothelium rapidly. Rous and his co-workers developed the hypothesis that the endothelium of the capillaries serving certain organs is relatively impermeable near the arterioles and becomes increasingly permeable as the venules and veins are approached. This might at first seem to invalidate the Starling hypothesis.

It is known, however, that plasma proteins pass through the vascular walls, including those of the venous capillaries and venules, only in small concentration under normal conditions. Therefore this gradient of permeability, which has so far been demonstrated only by dyes, does not affect the relation which fluid movement bears to the balance between capillary pressure and the colloid osmotic pressure of the plasma proteins, except possibly to hasten absorption and expedite the development of a state of equilibrium. These dye studies do indicate that gradients of permeability tend to equalize rapidly the distribution of nutrient materials of moderate molecular size even though they cannot appreciably influence the final equilibrium between the fluid in blood plasma and tissue fluid.

(f) **Tissue pressure and the movement of fluid through the capillary wall.** The pressure under which blood flows through the capillary network is only one part of the balance between the hydrostatic pressures inside and outside the capillary wall. If the fluid filtered by an excess of capillary pressure over colloid osmotic pressure passed into a space which could expand easily without distention, a high capillary pressure should produce filtration indefinitely at a constant rate. It is generally agreed, however, that this is not the case. In the tissues, and to a lesser extent in the serous cavities, the space available for accommodating extravascular fluid is limited. As fluid accumulates outside the capillaries, the tissue elements must be separated. Further filtration of fluid is then opposed by the pressure required to distort the tissues. If capillary pressure is elevated in an extremity by congesting the veins, the filtration rate is greatest during the first few minutes and then decreases rapidly during the first thirty minutes of congestion. At the end of thirty minutes, venous pressure and capillary blood pressure remaining the same, the rate of filtration is usually less than one quarter that observed during the first five or ten minutes. When sufficient fluid has accumulated in the tissue spaces low venous pressures will fail to produce further filtration.

Obviously the development of this tissue pressure tends to prevent edema in the dependent tissues of the human being where venous pressure for brief periods will often be much greater than the colloid osmotic pressure of the plasma proteins. It must be remembered, however, that whatever may be the importance of tissue pressure in limiting undue filtration under ordinary conditions, the mere existence of edema in disease shows in itself that the power of the tissues to resist accumulation

of tissue fluid is limited. If a tendency toward edema formation exists for a long time the tissues gradually give way. Their elasticity delays, but cannot prevent, the appearance of edema. It is a common observation that certain forms of edema appear first in the loose tissues of the orbit and face, but the mere looseness of these tissues is never the primary cause for the appearance of the edema. When filtration tends to exceed absorption over long periods of time, the less resistant looser tissues merely show pitting edema before the more compactly constructed tissues do.

Interchange of Gases

Available evidence indicates that the required interchange of gases between blood and tissue cells can be accomplished by diffusion both in rest and during exercise. This is possible, however, only because during activity of tissue the number of open capillaries increases conspicuously and mean blood flow increases in proportion. In a cross section of resting muscle only a few of the available capillaries are open, while the great majority are constricted. For example, Krogh [3] stated that in resting muscle the open capillaries, though equally spaced, were as few as 5 per sq mm, whereas after exercise the corresponding muscle of the other leg contained 195 capillaries per sq mm—a forty-fold increase. The distance between open capillaries in resting muscle is therefore great, but with exercise the larger number of open capillaries reduces the number of muscle fibers supplied by each vessel.

Oxygen diffuses through animal tissues at rates between ½ and ⅓ as fast as those observed in water, but the diffusion constant increases with rising temperature about one per cent per degree centigrade. If each capillary be supposed to supply oxygen independently of all others to the cylinder of tissue surrounding it, the difference in oxygen pressure between the inside of the capillary wall and the outermost border of this cylinder will be proportional to the oxygen consumption and inversely proportional to the diffusion rate. Knowing the radius of the capillary, the radius of the cylinder of tissue it supplies, and the oxygen tension of venous blood, it is possible to estimate roughly whether or not simple diffusion is quantitatively sufficient to cover the oxygen requirements of muscle during exercise.

In resting muscle, with very few capillaries open, these calculations [3] indicate that slight oxygen lack may exist in those parts of the muscle which are most distant from an open capillary. Usually, however, the number of open capillaries, though small, is sufficient to maintain an adequate oxygen tension everywhere. After massage, and particularly during work, opening of many previously closed capillaries reduces the radius of the cylinder of tissue to be supplied by each vessel to such an extent that the oxygen pressure in muscle becomes practically equal to that of the venous blood, and the greater utilization of oxygen is matched by a greater supply. As a matter of fact the number of capillaries opened by exercise, according to these estimates, is somewhat greater than that required for oxygen supply alone.

The diffusion constant for carbon dioxide is much higher than that for oxygen and it follows that this increased capillary bed will also provide for efficient elimination of carbon dioxide by simple diffusion into the capillary blood. Because the number of open capillaries is determined by the dilating effect of the metabolic products arising directly from tissue activity it is clear that the richness of local circulation is always suited to the metabolic needs of the immediately adjacent tissues in rest and during activity. By this means the independent contraction or relaxation of the capillaries, in association with like changes in arteriolar tone permits simple diffusion to provide both oxygen supply and carbon dioxide removal within a wide range of activity.

The Capillary Circulation In Special Conditions

(a) **Effects of heat and cold.** Though fluid movement is conspicuously modified by heat or cold there is no direct evidence yet to show that these changes are due to

any modification of capillary permeability per se, except when the change in temperature is great enough to injure the endothelium. Slight changes in capillary permeability due to variations in temperature are by no means ruled out, but even if present they would have little effect on the concurrent processes of filtration and reabsorption because, if permeability to protein remains constant, more rapid filtration in one area would be balanced by more rapid reabsorption elsewhere. Only if capillary permeability is increased to the point where protein passes with ease, will absorption cease. Thus excessive heat injures the capillary wall and produces edema, but this is the usual effect of tissue damage and not a specific effect of temperature.

It is common knowledge that the digits swell during warm weather and shrink during cold weather. This change in volume is partly due to the amount of blood in the vessels, though the rate of change is too slow to be explained satisfactorily on this basis alone. Local heat also dilates the superficial vessels and blood flow is much increased. Average capillary pressure in the skin of the finger under normal conditions amounts to 32 mm Hg in the arteriolar limb and 12 mm in the venous limb. When skin temperature at the tip of the digit is elevated to forty-two degrees centigrade capillary pressure rises to sixty and to forty five millimeters in the arteriolar and the venous limbs respectively. The excess of capillary blood pressure over the colloid osmotic pressure of the blood then leads to filtration of fluid which accumulates temporarily in the tissue spaces but eventually passes on into the lymphatic vessels. In the distal phalanges, however, there are many arteriovenous anastomoses which open widely whenever the tissues are heated or whenever vasoconstrictor tone is reduced. It is possible that the reactions of the fingertip are not quite characteristic of those elsewhere, and this difference may offer an explanation for the observation that heat is more apt to produce swelling in the distal portions of the extremities than in the proximal tissues. Local heat also increases the total filtering area available by dilating capillaries previously closed or only partially open; this factor would tend to increase the rate of filtration at any given venous pressure. It is impossible to estimate the relative importance of these two factors from the data at present available.

Plethysmographic studies indicate that as the temperature of an extremity is increased, a given venous pressure produces greater filtration in the warm forearm than in the cool forearm. At 44° C fluid accumulates twice as rapidly at a given venous pressure than is the case at a temperature of 15° C. At temperatures below 15° C, as shown by Lewis,[5] the initial constriction of the minute vessels gives way to a reactive dilatation which is stimulated by injury and the secondary release of vasodilator substances, chiefly a substance similar to histamine. In fact, excessive cooling can produce marked vasodilatation, rise in capillary pressure, edema and blistering just as burning heat does. The blister fluid obtained from such excessively cooled or heated areas contains protein in concentrations approaching that of blood, indicating gross endothelial damage. When excessive heat leads to the formation of a blister, of which the burn is characteristic, the extravascular fluid is not only rich in protein but contains additional substances produced whenever tissues are injured. The blister fluid, when injected beneath normal skin, produces there both vascular dilatation and increased capillary permeability, owing to these contained histamine-like substances.

Other investigations indicate that heat increases the rate at which dyes pass through the capillary wall but the staining of the tissues does not become maximally rapid or uniform until the capillary wall is damaged. Also, injury by excessive heat increases lymph flow and the pressure in the lymphatic vessels may reach 120 cm of water. The protein content of such lymph also approaches that of blood owing to gross endothelial injury.

The effect of temperature on filtration is particularly interesting with regard to

the appearance, or the accentuation, of certain forms of edema in response to climatic changes. Frequently patients with mild edema complain that the swelling of the lower extremities is greater during warm weather than during cold. Physicians in tropical countries describe two types of this so-called "heat edema." The milder form is extremely common all over the tropics; Europeans on their way to the East frequently develop it when their ship reaches the southern portion of the Red Sea and Aden. The feet and legs swell slightly and pitting edema occurs without any evidence of renal or cardiac abnormality. Heat edema in more severe form was observed by Castellani in Europe and America during a heat wave. The edema comes on suddenly, lasts as long as the high environmental temperature persists and then disappears.

In a warm environment the cutaneous vessels of the upper and lower extremities dilate in order to promote the loss of heat. During such vasodilatation the total area of capillary wall available for filtration in the extremities must be increased. In the standing position venous pressure in the lower extremities reaches very high levels due to the hydrostatic pressure of the column of venous blood. With such venous pressure the greater filtering area favors excessive filtration of fluid. Heat edemas may therefore be at least partially due to physical effects similar to those exerted by local heat on the filtration of fluids through the normal capillary wall.

In summary, the effects of heat on fluid movement through the capillary wall can be explained satisfactorily on the basis of (a) capillary dilatation which increases the area available for fluid movement; (b) rise in capillary blood pressure which favors filtration; and finally (c) injury by excessive heat which, like other harmful agents, increases capillary permeability to colloids and consequently lowers the effective osmotic pressure of the plasma proteins. Prolonged and excessive cold can produce injury of the same degree as illustrated by frostbite. The effect of mild changes in temperature on the permeability of the capillary wall to fluid alone has not been ruled out but, other factors remaining constant, such change cannot be expected to influence conspicuously the balance between blood and tissue fluid.

(b) **Capillary pressure and permeability in relation to tissue function.** Secretory activity of glands and contraction of muscles are accompanied by a conspicuous increase of blood flow through the active tissues. Even temporary ischemia of the tissues, for example by a tourniquet, produces, after blood flow returns, a reactive hyperemia in response to the accumulation of metabolites which with normal circulation would have been carried away as soon as they were produced. The number of open capillaries in exercising muscle is, as mentioned previously, much greater than in resting muscles and the surface available for filtration is proportionately increased. Arteriolar dilatation and the opening of previously closed capillaries are generally ascribed to the direct action of metabolic products. While there is not complete agreement concerning the specific metabolic product which produces vasodilatation, the relation between tissue activity and conspicuous hyperemia is not questioned.

Many studies have shown also that tissue activity is accompanied by a shift in the equilibrium between the circulating blood and tissue fluid. As mentioned above, lymph can be obtained only with difficulty from resting muscle but it flows freely during and after a series of contractions, indicating that the filtration of fluid from the blood circulating through the exercised muscle is greatly increased. Physiological activity also increases the flow of lymph from the salivary glands and from the kidney.

Direct measurements have shown that capillary blood pressure rises following contraction of muscle. In resting frog's muscle average arteriolar capillary pressure is 14.9 cm of water while average venous capillary pressure is 9.5 cm; during the hyperemia which follows tetanus of muscle average arteriolar capillary pressure is 20.1 cm and average venous capillary pressure 16 cm without significant change in

either arterial or venous pressure. The rate of blood flow and pulse pressure in the capillary network are increased conspicuously during this period of hyperemia indicating general arteriolar and capillary dilatation.

Indirect determinations of filtration during muscular activity indicate, however, that the amount of fluid actually moving toward the muscles exceeds the possible filtration to be expected from a simple rise of capillary blood pressure. During muscular activity, when the blood is losing fluid at the rate of two cc per minute per hundred grams of functioning tissue, the rate of filtration is at least several times more rapid than it is in resting tissue under a venous pressure of 80 cm of water during rest and at a temperature of 44° C. This suggests that some other factor is concerned in producing the high rate of lymph flow characteristic of actively functioning tissues.

Because of the close relation between tissue activity and the rate of lymph flow it has been postulated that the osmotic pressure of metabolic products must change the equilibrium between blood and lymph. These metabolic products are diffusible and, by reason of their smaller molecular size, have an osmotic pressure greater than the parent substances from which they were derived. In agreement with this concept muscle after being completely fatigued by exercise, has an osmotic pressure greater than normal by about thirty per cent. This, of course, is an extreme condition following anaerobic contraction. The change in the vapor pressure of the circulating blood during exercise shows also that significant local changes in osmotic pressure in contracting muscle are quite within the range of possibility.

Since these osmotically active substances diffuse from muscle into the extravascular fluids, and then into blood, it is logical to suppose that fluid will be withdrawn from the blood stream until diffusion restores the osmotic equilibrium which usually obtains between the diffusible constituents of blood and tissue fluid. No change in capillary permeability is required, since this movement of fluid is due solely to temporary inequality of osmotic pressure inside and outside the capillary wall. When blood flow is kept constant, while the rate of metabolism is varied, the loss of fluid from the blood changes with metabolism. However, when metabolism is kept quite constant the loss of fluid from the blood changes in general with blood flow. It can be concluded that functional activity increases filtration (a) directly, by altering the osmotic properties of the tissue fluid, and (b) indirectly by inducing vasodilatation, increasing capillary pressure, and expanding the area available for filtration. This view at least explains the observation that lymph flow during muscular activity exceeds temporarily the filtration produced during rest by simple elevation of capillary blood pressure.

That the movement of fluid toward functioning tissues might conceivably be concerned also with increase in the permeability of the capillary wall itself is suggested by the relatively high protein content of lymph collected from dogs during walking. This lymph contains from 1.0 to 1.5 per cent protein. It is known that lack of oxygen increases the permeability of the capillary wall to protein. Moreover, in perfusion experiments oxygen supply must be adequate if edema is to be avoided. In active tissues the situation is more complicated because metabolic activity might modify endothelial permeability through lowering of oxygen tension, increase of carbon dioxide tension or by local increase in acidity. These three possibilities have been explored by the micro-injection technique and it has been concluded that after complete lack of oxygen for three minutes, fluid filters through the frog's capillary wall four times more rapidly than normal. At the same time the capillary wall, which is normally relatively impermeable to protein, permits protein to pass and the effective osmotic pressure of the plasma proteins is reduced to one-half normal.

The movement of fluid through the asphyxiated capillary wall is still directly proportional to the difference between capillary pressure and the effective osmotic pressure of the plasma proteins indicating that even under these conditions the

endothelium acts merely as a passive filter, though it is more permeable than normal. When blood flow is resumed after brief oxygen lack the capillary wall rapidly recovers its impermeability to protein and the rate of fluid movement also returns practically to normal. The effects of oxygen lack, if sufficiently brief in duration, are almost completely reversible but longer periods of asphyxia produce irreversible changes in permeability.

Similar exposure of the frog's mesentery to Ringer's fluid half saturated with carbon dioxide does not modify fluid movement, and complete saturation increases the rate of fluid movement very slightly; but in both cases the wall still remains normally impermeable to protein. Likewise, increasing the hydrogen ion concentration within physiological limits produces almost no change in fluid movement until, at a pH of 4.0, the characteristic effects of injury appear.

Though complete lack of oxygen increases capillary permeability, it is still doubtful whether the oxygen tension in functioning tissues with normal blood vessels ever becomes low enough to produce this change. Anoxemia would have to occur in spite of increased blood flow and the opening of many new capillaries which, according to Krogh,[3] is so efficient that, theoretically, the oxygen tension through the entire exercising muscle is practically that of venous blood.

Lymph collected from contracting muscle ordinarily contains more red cells than are present in lymph collected from resting tissues. It has been suggested that these erythrocytes may escape through preformed stomata, probably within the cement substance, which open under increased pressure and close when pressure is reduced to normal. When muscle is strongly contracted groups of capillaries can be seen to be obviously distended, because the pressure exerted by contracted muscle fibers flattens the small venules draining the region under observation. With relaxation of muscle, blood will frequently spurt from such capillaries into the venules. Apparently any openings thus produced can appear and disappear without modifying the normal permeability of the capillary wall under ordinary conditions. Future work on the relation of capillary permeability to oxygen lack and to these temporary openings may explain some of the perplexing anomalies of fluid movement observed during tissue activity.

In summary, functional activity of the tissues produces hyperemia and with it a rise in capillary blood pressure. The increased flow of lymph which accompanies vigorous tissue activity is probably too great to be explained on the basis of a simple rise in capillary pressure or increase of filtering area. It is likely that osmotically active substances produced in the course of tissue metabolism are partially responsible for this copious and temporary filtration. Accumulation of carbon dioxide and changes in hydrogen ion concentration within physiological limits have little, if any, influence on capillary permeability, certainly not enough to modify the effective colloid osmotic pressure of the plasma proteins to any measurable extent. Lack of oxygen, if extreme, can produce a temporary increase of permeability which is great enough to permit protein to pass but it is doubtful whether anoxemia of this grade can be produced in active tissues with normal blood flow.

(c) **The effects of injury on capillary pressure and permeability.** The effectiveness with which the plasma proteins retain water within the capillaries depends upon the vitality of the endothelium. All forms of injury increase capillary permeability and reduce correspondingly the effective colloid osmotic pressure of the blood. Tissue damage therefore influences the movement of fluid through the capillary wall in a complex manner by reason of simultaneous changes in blood flow, capillary pressure and capillary permeability.

The injury that is produced when a single capillary is pierced by a micro-pipette is extremely slight, but nevertheless is often followed by clear increase in blood flow and capillary pressure. If a single capillary in the frog's mesentery be compressed

gently by means of a minute blunt glass rod the effect of local injury on the permeability of the endothelium becomes apparent at once. Colloidal dye solutions pass freely through such an injured area before they can be detected outside normal areas of endothelium in the same capillary. The colored fluid does not spurt from a single point as it should if the endothelium were mechanically ruptured, but filters out uniformly along the whole damaged section. India ink introduced intravenously flows past the normal section of capillary but the carbon particles accumulate in dense masses on the inner surface of the injured section. When a single capillary is severely injured blood corpuscles and ink are commingled in a solidly packed collection, indicating that endothelial permeability is maximally increased and that whole plasma has filtered through the damaged section of capillary wall.

Injury which, instead of being limited to a minute area, affects large portions of the peripheral vascular system, initiates a series of events which are fundamentally similar in their development irrespective of the agent used. Small amounts of harmful substances introduced intravenously, or applied to a tissue externally, produce simply vasodilatation, increased rate of blood flow and usually a rise of capillary blood pressure. More severe injury is followed by visible concentration of the red cells, slower flow, and finally complete stoppage of circulation due to complete loss of plasma through the damaged endothelium. This is true stasis and the solid cylinder of compressed red cells devoid of plasma, is the surest sign of increased capillary permeability. When injury reaches this stage proteins, colloidal dyes, and colloidal starch pass through the capillary wall rapidly, but India ink particles and red cells are retained quantitatively. On this account Krogh[3] has placed the size of the pores in damaged endothelium between 5 and 200 millimicrons, whereas the pores of normal endothelium, according to Keys,[2] have diameters between 0.7 and 2.1 millimicrons. At a given capillary pressure fluid filters from 5 to 7 times more rapidly through an injured capillary wall than through a normal capillary wall. At the same time the endothelium becomes sticky so that leucocytes, erythrocytes, and platelets adhere to the damaged capillary wall.

While the increased transudation through the walls of injured capillaries is due chiefly to greater capillary permeability it is also influenced by capillary blood pressure. The rate at which blood plasma leaves the injured capillary is greater during cardiac systole when capillary pressure is high, than during diastole, when capillary pressure is slightly lower. If systemic blood pressure is very low stasis develops much more slowly. Conversely, stasis develops more rapidly in a capillary network with rapid blood flow and higher capillary blood pressure.

If the injury is great enough to produce immediate stasis edema rarely appears for the obvious reason that blood flow ceases in the injured area before sufficient fluid can pass out of the blood vessels to distend the tissues grossly. Milder grades of damage persisting over longer periods of time permit fluid to filter in sufficient amount to produce edema.

The histamine wheal illustrates very well the complex reactions which follow injury. If 1:1000 histamine is introduced into the skin by placing a droplet of the solution on the skin and then puncturing the skin with a fine needle through the droplet the so-called "triple response" develops. (a) In 20 seconds a small red spot develops around the puncture due to stasis in the minute vessels immediately affected. (b) This is followed in 1 to 2 minutes by a pink flare from 2 to 4 cm in diameter due to dilatation produced by a local nervous reflex mechanism (the axone reflex) in which vasodilator impulses relax the minute vessels. (c) In 2 to 3 minutes a wheal, or area of local edema, forms around and over the red spot, with a diameter of 2 to 3 mm. This local edema is due to leakage of plasma from the capillaries which have been made highly permeable by the histamine. The wheal fluid contains a large amount of protein. Lewis[5] has presented detailed evidence to indicate that any form

of injury—thermal, chemical, mechanical, or physical—liberates from the tissues a histamine-like substance which is responsible for the circulatory changes in inflammation.

The inflammatory swellings seen in pathological conditions demonstrate in their early stages all of the reactions characteristic of local injury, but in later stages are complicated further by gross tissue destruction and liquefaction. There can be little doubt that the permeability of the capillary endothelium is increased in all types of inflammation because colloidal dyes pass more easily, and exudates in general contain protein in concentrations approaching that in blood plasma. Vasodilatation, increased blood flow, later sluggish blood flow and even stasis in isolated areas, with accompanying changes in capillary blood pressure are all found at some stage of inflammation. Even though the capillary wall is injured grossly, agents which reduce capillary blood pressure diminish filtration to some extent, while those which elevate capillary pressure increase the accumulation of edema fluid.

In summary, local injury initiates a complex response including vasodilatation, rise in capillary blood pressure, increased blood flow, increased endothelial permeability and, if injury is very severe, finally stasis and reduced blood flow. All forms of local or general edema produced by injury are due fundamentally to increased capillary permeability and resulting passage of proteins and fluid through the endothelium.

Bibliography

1. Drinker, C. K., and Field, M. E., "Lymphatics, Lymph and Tissue Fluid," Williams and Wilkins, Baltimore, 1933.
2. Keys, A., *Trans. Faraday Soc.*, **33**, 930 (1937).
3. Krogh, A. "Anatomy and Physiology of the Capillaries," Yale University Press, New Haven, 1922; Revised edition, 1929.
4. Landis, E. M., *Physiological Rev.*, **14**, 404 (1934).
5. Lewis, T. "The Blood Vessels of the Human Skin and their Responses," London, Shaw and Sons, 1927.
6. Rous, P. and Smith, F., *J. Exp. Med.*, **53**, 219 (1931).
7. Starling, E. H., *J. Physiol.*, **19**, 312 (1896).

[8] Structure may be a factor here, and mere morphological thinness is no positive criterion for secretion, which with some plant cells takes place through septa no thicker than endothelium, and appears also to take place through vacuolar membranes, *e.g.*, in the contractile vacuole of ameba. J. A.

[9] The attraction of plasma for water ("plasma thirst") is opposed by the corresponding attraction of the tissues ("tissue thirst") ; and the distribution of water is influenced by these interrelated factors, the intensities of which vary with the diffusion of metabolic products, *e.g.*, CO_2 which diffuses from tissues to blood (erythrocytes) while O_2 diffuses into tissues. (See J. Alexander, "Colloid Chemistry," 4th ed., D. Van Nostrand Co., 1937). A somewhat analogous situation exists in plants, where the "root pull" is opposed by the "back pull" of the soil. The percentage of water left in the soil when the plant can no longer protect itself against desiccation by taking up soil moisture, is known as the "wilting coefficient." Its value varies greatly with both plant and soil; thus, in the case of Kubanka wheat it is 2.59 for fine sand, 9.66 for fine sandy loam, and 16.3 for clay loam. (See J. Alexander, *lib. cit.*, p. 191). J. A.

EDITOR'S NOTE

In Vol. II of this series, Prof. Frank P. Underhill (pp. 723-730) pointed out that in the case of extensive burns, affecting over about one third of the body surface, death generally occurs because the viscosity of the blood becomes excessive, due to loss of water from the plasma to the tissues. "The systematic treatment of these burn cases consisted simply in the forcing of fluids, water by mouth when possible; when the patient could not cooperate because of unconsciousness, fluid was injected under the skin, directly into the blood, by rectum, etc. The quantity of fluid taken in varied from 4 to 8 liters daily. In a day or two on this treatment the blood concentration fell gradually and the patient's condition steadily improved." Prof. Underhill also points out that "burns" from "war gases" may have analogous effects.

Experimenting with mice subjected to standardized burns, Sanford M. Rosenthal (Public Health Repts., **58**, 513-522 (1943)) found that the mortality during the first 72 hours was significantly reduced by administration of sodium chloride solutions by mouth or intraperitoeally, isotonic solutions being better than hypertonic solutions. Intravenous injection was less effective. Potassium chloride caused an acceleration in the time of death, and when

administered with NaCl it antagonized the effects of the latter. Another factor to be considered was pointed out by H. Kabat and M. Levine (Science, **96**, 476 (1942)), who found that in some cases speedy death may follow burns because of protein coagulation, leading to the formation of emboli.

Biochemical and Physiological Mechanisms in Inflammation *

VALY MENKIN, M.D.

Department of Pathology, Harvard University Medical School, Boston, Mass.

Inflammation may be broadly defined as the complex vascular, lymphatic and local tissue reaction elicited in higher animals by the presence of microorganisms or non-viable irritants. It represents a basic or elemental reaction to injury whereby the deleterious agent tends to be localized and ultimately destroyed. The phenomenon of inflammation as manifested in the vertebrate kingdom may properly be regarded as the physical basis of infectious processes. It may be considered as an immunological mechanism of definite significance. The inflammatory reaction irrespective of the irritant tends to develop along a uniform pattern. Foreign substances whether viable or non-viable in contact with normal tissue will induce the reaction, the intensity of which may vary from a barely visible hyperemia to an intense suppurative process.

The purpose of this review is to summarize the available evidence on the biochemical and physiological mechanisms involved in the development of the inflammatory reaction. The immunological implications of inflammation have already been considered in detail elsewhere (Menkin, 1938a, 1940a), and therefore this important aspect of the problem will in the present chapter be treated with brevity.

The orderly sequences in the development of the inflammatory reaction can be conveniently classified somewhat as follows:

1. **The Vascular Capillaries in Inflammation.** The inflammatory reaction is initiated by an alteration in the local fluid exchange. This is primarily referable to a basic disturbance in the local filtration equilibrium between the small vascular channels, the lymphatics and the so-called intercellular fluid.

The most satisfactory hypothesis that explains filtration between blood and tissue was first offered by Starling (1896). The concept was subsequently verified by the measurements of Landis (1926, 1927). In brief, this investigator concluded that the gradient of pressure along the course of the minute vessels favors filtration at the arteriolar portion of the capillary and reabsorption at its distal part. The rate of fluid passage appears to depend on the balance between the capillary pressure and the osmotic pressure † of the plasma proteins. These original conclusions were based on the assumption that the endothelial lining resembled an inert membrane and that it was essentially impermeable to the outward passage of proteins. This concept was modified by Drinker and his collaborators who demonstrated that the endothelial wall of capillaries was not wholly impermeable to the leakage of proteins. An appreciable concentration of proteins to be reckoned with was recovered from the lymph which thus reduced markedly the effective osmotic pressure of the plasma colloids (Churchill, Hakazawa, and Drinker, 1927; Drinker and Field, 1933).

In the last ten years the importance of the capillary filtration of Starling, revived

* These studies were aided in part by grants from the International Cancer Research Foundation, the Dazian Foundation for Medical Research and the Milton Fund of Harvard University.

† Also termed onkotic pressure. See H. Schade, Vol. II, p. 635, this series. J.A.

by Schade (1927) and verified by the ingenious measurements of Landis (1926, 1927) has been critically scrutinized in view of the findings of Rous and his collaborators (1930). These investigators injected intravenously some poorly diffusible dyes into frogs and into mammals (guinea pigs, rabbits and mice). The dye was observed to escape primarily from the distal part of the capillary and from the small venules, thus giving rise to a mounting gradient of permeability.

The gradient was observed to be independent of various factors, including hydrostatic pressure (McMaster, Hudack and Rous, 1932), osmotic pressure (Smith and Dick, 1932), nervous section, hemorrhage or cessation of circulation. Rous concluded that the permeability gradient was most likely referable to structural differentiation of the capillary wall rather than to any functional alteration. To Rous the mounting permeability gradient represents an adaptive mechanism for the equalization of opportunity in the distribution of diffusible nutrients to tissue. Through such a delicate regulatory mechanism a homeostatic state is effectively maintained in tissue, in accordance with Bernard's concept of the "milieu interne."

On the basis of his results, Rous cast some doubt on the validity of the Starling hypothesis as an adequate explanation of fluid exchange. Rous and Smith (1931) failed to confirm Landis' observations on the passage of trypan red (vital red HR) through the arterial part of the capillary wall in the frog mesentery. Landis (1934) insists that observations made with a relatively poorly diffusible dye, such as have been used by Rous and his collaborators should not be stressed too far in the evaluation of data pertaining to highly diffusible constituents of blood. Peters (1933, 1935) agrees that the Rous permeability gradient may be concerned in certain tissues with the equalizing of opportunities in distribution of nutrient material of large molecular dimensions. But according to this investigator, this fact in itself alters in no way the Starling-Landis hypothesis which is still considered a reasonable explanation for the local fluid exchange of water and electrolytes through the capillary wall.

The recent work of Zweifach (1940) in regard to the structure of the capillary wall and the factors at play in determining permeability under what is interpreted as approximating physiological conditions, is doubtless of distinct interest. According to this author, the capillary barrier is referable to an intercellular cement substance and also to an adsorbed layer of protein. The cement substance seems to be a calcium proteinate. In the opinion of this worker, the cellular components of the capillary wall seem relatively unimportant in determining permeability. Danielli (1940) has recently assembled the results of numerous experiments tending to show that edema formation seems to be associated with platelets. Serum, red cells or merely solutions containing platelet material reduce edema formation more than can be accounted for by the osmotic pressure of colloids. The view is advanced that the action of the platelets may be primarily mechanical, involving thus a blockage of the protein-permeable pores. The observations of these last two investigators may yet prove to be of distinct significance in clarifying our understanding of the nature of the capillary wall and of some of the factors capable of influencing filtration.

The introduction of a chemical or bacterial irritant into normal vertebrate tissue is followed by definite disturbances in the filtration equilibrium. There is local increase in the passage of plasma as evidenced by the edema which may develop fairly early. Both the capillary pressure and the mounting permeability gradients are so profoundly altered as to favor enhanced fluid accumulation in the affected tissue. The principal forces at play in promoting these changes may be listed as follows:

1. Increased capillary pressure accompanying local vasodilatation.
2. Increase in the actual number of functioning capillary channels.
3. Increase in capillary permeability. The initial constriction of the small vessels which is frequently encountered will not be discussed here owing to the lack of exact data on the mechanism and significance of this phenomenon.

Krogh (1922) assumed that the increased filtration through the capillary wall following the use of urethane was referable to the stretched endothelium of the dilated capillary. Landis (1927) demonstrated by intravascular injection of dyes that there was no measurable increase in permeability with dilatation of the capillaries. The passage of fluid seemed to be primarily referable to the level of intracapillary pressure. This investigator regards the effect of urethane as the result of direct injury to the endothelial lining accompanied by an increased capillary pressure.

The available data indicate that capillary dilatation in hyperemia or inflammation is accompanied by increased capillary pressure (Landis, 1931). This elevated pressure constitutes a definite factor that favors the outward passage of fluid. Landis concluded that the rise in capillary pressure in hyperemic conditions was referable to arteriolar dilatation. There was no indication in his experiments of any increased venous pressure to account for the augmentation in the intracapillary level. Hyperemia was induced, for instance, by placing a crystal of silver nitrate on the skin of the web of a frog. Within 1 to 3 minutes there was a rapid rise in capillary pressure which returned, as a rule, to the resting level within 7 to 20 minutes. The average rise in pressure varied between 19.5 cm of water to 16.5 cm depending on whether measurements were made in arteriolar or venous capillaries. Landis believes that this transient rise in pressure is referable to a stimulation of the axon reflex which causes arteriolar dilatation with accompanying increase in capillary pressure. It is obvious that an increase in pressure within capillaries would be significant in promoting filtration through the endothelial wall.

Probably an even more striking factor favoring increased seepage into the extracapillary spaces of an inflamed area is an actual change in the endothelial lining which increases its permeability or capacity for permitting the passage of fluids. Cohnheim called attention to this phenomenon about fifty years ago. Landis (1927) pointed out that capillaries injured by alcohol and mercuric chloride appear to be permeable to the plasma colloids and approximately seven times more permeable to fluids than is the normal capillary wall.

The extent of injury to the capillary wall can be demonstrated in the early phase of the inflammatory reaction by a variety of means. For instance, graphite particles that normally fail to pass through the endothelial membrane readily do so in the early stage of an acute inflammatory process (Menkin, 1931a). Another convenient method of demonstrating the alteration in the permeability of the endothelial wall in injury is by the use of intravenously injected dyes. Trypan blue and other dyes introduced into the circulating blood promptly accumulate in an inflamed area. The localization of vital dyes in areas of inflammation has been demonstrated by several investigators. MacCurdy and Evans (1912) pointed out that the normal brain and cord always remain free from dye injected intravenously but that areas of damage, such as softening or inflammation, become deeply stained. Bowman, Winternitz, and Evans (1912) found that trypan blue injected intravenously stains tubercles in experimental tuberculosis. Subsequently Winternitz and Hirshfelder (1913) demonstrated that this dye when injected intravenously in experimental lobar pneumonia stains the consolidated area of lung selectively. Lewis (1916) found that if the cornea of a rabbit is inoculated with a living culture of tubercle bacilli, a progressive lesion results characterized by an intense congestion of the conjunctiva. If the animal receives an intravenous injection of trypan red 24 hours or more after such inoculation, the fluid in the anterior chamber of the inoculated eye always becomes colored. Precisely similar results were obtained when abrin was administered in the conjunctiva as an inflammatory irritant. A number of years ago McClellan and Goodpasture (1923) showed that trypan blue accumulates in lesions of herpetic encephalitis in the rabbit's brain, the injured areas presenting a striking color against the quite unstained healthy brain tissue. Siengalewicz (1925) pointed out that general damage to the nervous tissue, such as poisoning with carbon monoxide

or with salvarsan, is followed by marked staining of the damaged areas by trypan blue. Ramsdell (1928) injected trypan blue into the veins of rabbits and guinea pigs previously treated with foreign serum and found that injection of the same serum into the skin of the ear immediately caused local infiltration of the dye into the adjacent tissue. She regarded the infiltration of the dye as an indicator of edematous changes resulting from toxic injury to the capillary endothelium.

Okuneff (1924) found that a thermal irritant favors the passage of vital stains from the blood stream into the area heated. Kusnetzowsky (1925) also observed that the local application to the skin of an irritant, such as heat or mustard oil, causes an accumulation of trypan blue in the inflamed area when the dye has previously been injected into the blood stream. Menkin (1929) demonstrated that trypan blue injected into the circulating blood enters the site of inflammation rapidly and is fixed there, so that the tissues are deeply stained. Furthermore, trypan blue injected directly into the site of inflammation in the subcutaneous tissue or in the peritoneal cavity is fixed in the inflamed area and fails to reach adequately the regional lymphatic nodes. This work clearly showed that there is not merely a rapid accumulation of the dye from the blood stream into an acutely inflamed area, but also that the dye is held in such an area and is unable to drain readily away through the regional lymphatics.

The accumulation in inflamed tissue of a dye that has been introduced into the blood stream is doubtless primarily the result of an increased permeability of the capillaries which is part of the inflammatory reaction. Menkin and Menkin (1930) were able to secure quantitative data by studying directly the change in the concentration of the dye in the blood stream both in the inflamed and in the normal mesentery of the frog. They found that, with the irritants employed, the permeability of the capillaries in the injured area was almost doubled.

Sufficient evidences have been adduced indicating that increased filtration in inflammation is primarily the result of an alteration in the permeability of the capillary wall. This, in turn, is reinforced to some extent by a transient and initial increased capillary pressure.

2. **Mechanism of Increased Capillary Permeability in Inflammation.** The precise mechanism involved in the initial augmentation of capillary permeability in injury is obviously of considerable importance for an adequate understanding of the subsequent sequences in the development of an inflammatory reaction. In 1923 Ebbecke postulated that a substance is formed by irritated epithelium which on diffusing to the cutis dilates the capillaries and smallest arterioles. Subsequently Lewis (1927) postulated the development of a type reaction primarily referable to a chemical H substance liberated from injured tissue. The H substance is presumably histamine or a substance having similar properties and therefore resembling it closely. According to Lewis and Grant (1924) the type reaction elicited by the cutaneous injection of histamine manifests itself in three ways: *a*, a local vasodilatation of capillaries, venules, and arterioles by direct action; *b*, a widespread dilatation of outlying arterioles resulting from a local reflex; and *c*, an increase in the permeability of the minute vessels by direct action. This type reaction leads to local edema of the skin.

The conclusions of Lewis and Grant (1924) appear to be based largely on an analogy of the type reaction obtained by histamine with that of a variety of other injurious agents. When more direct tests were performed on the effect of the skin wheal fluid on the contraction of the guinea-pig uterus, they were unable to obtain any evidence that histamine was liberated in larger quantities from injured tissue than was found in normal plasma.

Krogh (1929) accepts unreservedly the view first foreshadowed by the work of Ebbecke (1923), and later chiefly sponsored by Lewis, that under certain stimuli the tissue cells will liberate substances having a dilator effect on capillaries; but at the same time he finds it impossible to assume that in all cases the action is referable to a

single chemical H substance. For this reason Krogh postulates the possibility of two effective substances liberated from injured cells: a diffusible factor closely related to histamine, if not histamine itself; and an H colloid substance which is probably less diffusible. Rous and Gilding (1930) raise considerable doubt concerning the validity of a hypothesis which refers all local vasodilatation to the action of a single chemical substance liberated within tissues. They point out that the vascular contraction in Bier's spots prevails over the local reddening induced by mechanical injury, whereas it is without effect upon the local vasodilatation induced by cutaneous injection of histamine. More recently Goldschmidt and McGlone (1934) have studied the failure of a reactive hyperemia occurring with arrested circulation in an oxygen atmosphere. They have compared their findings with the histamine reaction in a similar oxygenated and isochemic environment. Their observations on the human forearm seem incompatible with the view that the vasodilatation responsible for reactive hyperemia is due to an H substance identical with histamine. By comparing the reactions of the dilated minute skin vessels to adrenaline puncture and to Bier's spots, Percival and Scott (1931) have been unable to conclude that the vascular reactions in exfoliative dermatitis, ultra-violet erythema, oil of mustard erythema, psoriasis, dermatitis venenata, and tinea corporis (ring-worm) were due to a histamine-like substance or histamine itself. That histamine is not always the active substance responsible for the increased capillary permeability in injury is indicated by the observations of Grant and Jones (1929) on the frog. This substance causes no dilatation of the frog's blood vessels either when applied locally or when given intravenously. They express the belief that in the frog the active substance is a base of the histidine-arginine fraction which is responsible for the vascular reaction to injury. The contention raised by Rous and Gilding has been reinvestigated by Wayne (1931) who arrives at the conclusion that the vascular reactions resulting from injury to the skin are due to a liberated substance, either histamine or some substance closely allied to it. This worker answers the objection of Rous and Gilding by maintaining that the vasoconstrictor effect of Bier's spots can overcome the slight local vasodilatation produced by histamine, provided the latter is introduced electrophoretically into the skin in a diffuse manner and in low concentrations. The vasoconstriction will, however, not prevail over a more intense local vasodilatation induced by greater concentrations of histamine. He therefore concludes that the vessels of the skin, under the influence of either histamine or the H substance liberated by injury, behave similarly towards Bier's spots when the two are distributed in a comparable way throughout the tissue spaces.

A study was undertaken by Menkin (1936a) in order to determine whether one or more substances could be obtained from inflammatory exudates which, when introduced into normal cutaneous tissue, would induce local vasodilatation and an increase in the permeability of the capillary wall. Furthermore, the properties of the active fractions which have been obtained from inflammatory exudates have been compared with histamine in an endeavor to test Lewis' hypothesis. In brief, the writer's observations indicate that a diffusible crystalline material capable of increasing capillary permeability is present in various types of inflammatory exudates. Its liberation and presence in exudates offers a reasonable explanation for the mechanism of increased permeability of small vessels in injured tissue. By appropriate tests this active principle has been shown to lack the properties characteristic of histamine (Menkin, 1936a, 1937, 1938b).

The evidence obtained by several authors indicates that histamine is present in exudates (Loos, 1931; Bier and Rocha e Silva, 1938; Menkin, 1938b, 1939a, b); but the primary importance of histamine in explaining the mechanism of increased capillary permeability in inflammation is now open to serious doubt. The finding by Feldberg and his associates (1937) that perfusion of the lung by snake venoms, staphylococcus toxin, or peptone, induces the liberation of proteins and of histamine

is no demonstration that the latter substance is primarily involved in the mechanism of enhancement of capillary permeability during injury. Countless substances are probably liberated from a site of inflammation, but whether many of these participate as important factors in directly determining increased capillary permeability is another matter. The case of histamine has not been satisfactorily proven. The permeability factor (termed leukotaxine) recovered from an exudate, described and isolated by Menkin (1936a, 1938a, b), displays none of the properties of histamine. Yet the purified permeability factor manifests identical reactions on the capillary wall as obtained by the untreated exudate. When histamine and the permeability factor were compared with regard to their respective properties, the following observations were made: (1) Upon intravenous injection of dye, the local staining pattern differed with each of these two substances. In dilution of 1:500,000 or 1:1,000,000 histamine may induce, as recently shown by Bier and Rocha e Silva (1938a), the accumulation of dye in the center of the area; but even in such dilution the pattern of staining is much lighter and over a larger area than is the case with exudate or the purified permeability factor. Furthermore, Bier has shown that the recovered amount of histamine in exudate induced by turpentine reaches a level of concentration ranging between 1:20,000 and 1:50,000. At such concentrations the pattern of staining induced by histamine is wholly different from that obtained by leukotaxine. Bier also reports that "inflammatory exudates of low histamine content are incapable of increasing capillary permeability in the treated cutaneous areas." The earlier studies of this investigator (1938a), cited above, have definitely pointed out that it is with low concentration of histamine that trypan blue accumulates from the circulating blood into a treated skin area. These facts, in themselves, would tend to invalidate the original claim that the permeability factor and histamine are identical. (2) In contrast to the usual effect elicited by histamine, the permeability factor failed to cause a contraction of the isolated intestine. (3) Following treatment with the permeability factor the intestinal segment exhibited no refractory state; the usual contractile effect elicited by histamine occurred. (4) Evidence was obtained of the antagonistic action of histamine to the active principle when mixed *in vitro* and then tested on the capillary wall. (5) The specific Zimmermann color reaction for histamine (1930) was negative when tested with the permeability factor. (6) A rapid chemotactic effect was induced by the permeability factor (leukotaxine); histamine apparently does not cause leukocytic migration. (7) The permeability factor failed to depress blood pressure in the cat in the manner characteristic of histamine (Menkin and Kadish, 1938). Recently Abramson and his collaborators (1938) have further confirmed by an iontophoretic method the earlier conclusions of the writer that histamine is probably not responsible for the production of hyperemic wheals.

The first method used by the writer to determine the effect of the permeability factor on the endothelial wall consisted of injecting the cell-free exudate intracutaneously in rabbits (1936a). This was almost immediately followed by an intravenous injection of trypan blue. The extent of accumulation of the dye in the treated cutaneous area served as a gauge of capillary permeability.

Bier and Rocha e Silva (1938b) have recently repeated the earlier studies of the writer on whole exudates and have confirmed his findings of a permeability and of a chemotactic factor in inflammatory exudates. They have, however, neglected to purify the exudates in order to obtain the relatively pure material capable *per se* of reproducing the increased capillary permeability and migration of polymorphonuclear leukocytes. On the basis of only meager evidence, they have concluded that the permeability factor (but not the chemotactic factor) is histamine. Their conclusions have been reached after extracting histamine from exudates, by the method of Barsoum and Gaddum (1935) in order to eliminate thus the presence of any depressing substances when the end-product is tested on the isolated strip of guinea pig

intestine. They admit, as the writer has pointed out above and also previously (1936a), that with the accumulation of trypan blue there is a difference in the staining pattern of tissue as elicited by histamine and by an inflammatory exudate. Yet these investigators assume without any evidence, that this is presumably referable to the loose combination of histamine with proteins or by adsorption of colloids. In brief, their observations on the contraction of the isolated intestine of the guinea pig simply indicate that histamine probably exists in exudates, and can by appropriate extraction, be demonstrated. This, as mentioned above, has been surmised and known.

The permeability factor (*i.e.*, leukotaxine) recovered and isolated from an exudate displays none of the specific properties of histamine. Yet leukotaxine is capable *per se* of reproducing the same type of reaction as the whole exudate. The mutual, non-specific properties of leukotaxine and histamine (*e.g.*, dialyzing property, etc.), listed by Rocha e Silva and Bier seem quite insignificant for the most part, inasmuch as they are true of numerous other unrelated substances. These investigators have failed to repeat the work on the chemical extraction of leukotaxine from exudate. They have worked solely with whole exudates. They have thus been unable to establish the fact that leukotaxine is histamine; whereas the writer has definitely demonstrated that leukotaxine has a great many physiological and chemical properties differing from those of histamine (Menkin, 1936a, 1938a, b, 1939a).

A single very important concrete evidence which nullifies the contention of these investigators is indicated by the following observation: A sample of leukotaxine, capable *per se* of actively inducing increased capillary permeability and leukocytic migration, is extracted for the presence of histamine by the method of Barsoum and Gaddum (1935). According to the contention of the Brazilian workers (1938a, b), by this procedure, the presence of any depressing impurities in leukotaxine, capable *per se* of overshadowing the contractile effect on the intestine of histamine, would thus be eliminated. The final extracted material not only fails to produce a positive Zimmermann reaction for histamine, but it fails to induce an enhanced contraction of the isolated segment of guinea pig intestine. Subsequent application of histamine to the same segment is followed by a powerful contraction, indicating thus the absence of any refractory effect. (Barsoum and Gaddum, 1935). These additional observations of leukotaxine render it difficult even to consider the possibility of its identity with histamine.

In view of the accumulated evidences Bier has recently retracted his original claim that histamine and leukotaxine (*i.e.*, the permeability factor) are identical substances (personal communication, 1939). His present argument, however, that leukotaxine may perhaps liberate histamine which in turn can act on the capillary wall, is not substantiated by any observations. On the contrary, recent studies by the writer on the ova of sea urchins (*Arbacia punctulata*) indicate that leukotaxine markedly enhances cellular permeability to water (1940b). Lucké has failed to show any such effects with histamine (personal communication). This would support the view that leukotaxine affects cellular permeability without the intermediary presence of histamine.

Recently Rocha e Silva and Dragstedt have reported observations tending to correlate the histamine equivalent of extracts of various tissues with their capacity to induce increased capillary permeability (1941). These investigators have failed to take cognizance of the fact that in such extracts leukotaxine may likewise be liberated as a result of cellular injury during the mechanical process of grinding and extracting the tissue. Their views have been severely criticized by Menkin (1941a) who has advanced new observations indicating that leukotaxine added to the inactive dialyzed exudate fraction, presumably free of histamine and of leukotaxine, can reconstitute the biological effect of the whole exudate whereas the substitution of histamine fails to accomplish a similar result. These experiments in addition to

previous observations demonstrate that leukotaxine and not histamine is primarily responsible for the mechanism of increased capillary permeability in inflammation.

In connection with the method utilized to identify the presence of a permeability factor in exudates, it is interesting to note that the writer has also been able to demonstrate that cortin, an extract of the adrenal cortex, is *per se* capable of inhibiting either completely or in part, the accumulation of dye in a skin area previously inoculated with either exudate or testicular extract (1940c, 1942a). The addition of cortin to an exudate or to leukotaxine and the subsequent injection of the mixture into a cutaneous layer seems thus to counteract the usual effect elicited by the permeability factor of exudates. These observations may perhaps be of some significance in explaining the rôle of the cortical hormone in traumatic shock (Swingle *et al.*, 1933). Furthermore, the method utilized by the writer may have practical application in the assaying of extracts of the adrenal cortex. These findings regarding the inhibitory effect of extracts of the adrenal cortex on increased capillary permeability have recently been confirmed by Freed and Lindner (1941).

In conclusion, the available evidence therefore indicates the presence in inflammatory exudates of a factor, other than histamine, which induces a prompt increase in the permeability of normal capillaries. A method for its isolation and purification has already been described in detail elsewhere (Menkin, 1938b, 1940a). In its essential features, this consists in treating the exudate with pyridine followed with acetone. After separation of the protein fractions, further purification can be obtained by prolonged extraction with butyl alcohol. The butyl alcohol supernatant fraction is then concentrated to about ⅓ of its original volume and subjected to low temperature in an ordinary refrigerator. The latter procedure favors spontaneous separation of the active principle. The purified material is a crystalline, doubly refractive, nitrogenous substance which is more or less free of any gross, viscous impurities. A somewhat simplified method for the extraction of leukotaxine, which is a slight modification of the original scheme, is shown in Table 1.

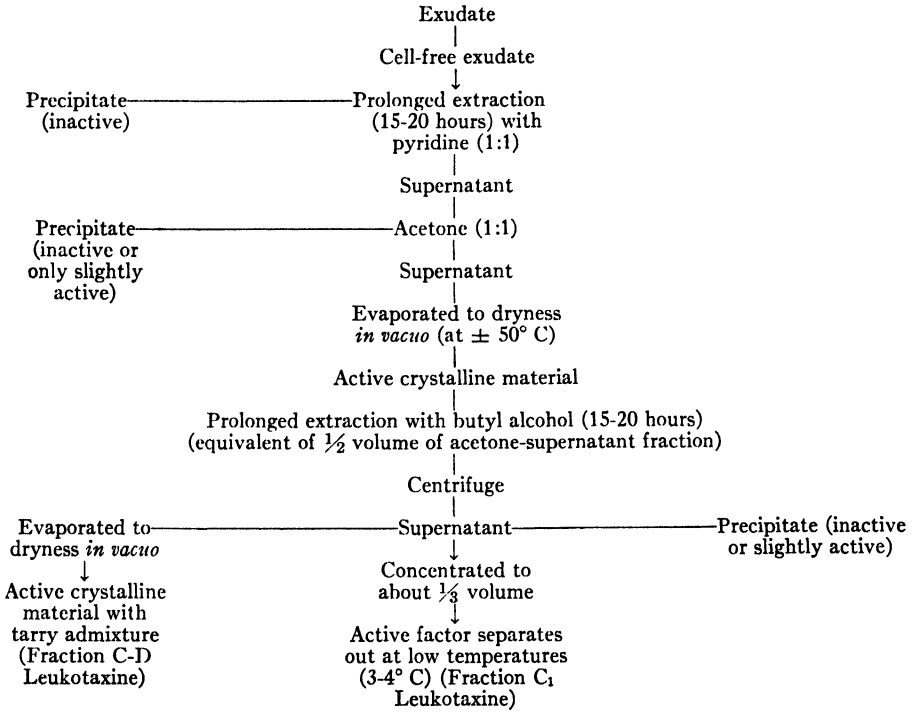
The purified crystalline substance is capable of inducing increased capillary permeability in dilution ranging from 1:100,000 to 1:10,000,000. This indicates considerable purification, for the untreated exudate displays practically no reactivity when diluted above 1 part in 10. The end-product obtained by a variety of methods (Menkin, 1938b, 1940a) is essentially of the same pattern. This is characterized by the formation of discrete aggregates of doubly refractive granules, of needle-like or small rod-like crystals, or a mixture of both types of crystalline material. The crystals usually appear as if they are bound together by the presence of an ill-defined, greyish, homogeneous matrix. From an aqueous medium the crystalline material upon gradual dehydration of the medium occasionally separates out in the form of beautiful needle-like crystals. The similarity of the end-product obtained by different analytical procedures strongly suggests that one is dealing with the same active substance. This crystalline material is extremely active in reproducing the two basic sequences encountered in the development of the inflammatory reaction, namely, increased capillary permeability and prompt leukocytic migration. In view of its additional chemotactic property, which is to be described presently, the writer has named this active substance *leukotaxine*. Minami and Inugami have recently repeated the studies of Menkin on the isolation of leukotaxine from exudates (1940). Their findings fully confirm his original observations, and besides, apparently suggest the liberation of a larger quantity of leukotaxine from an allergic than from a non-specific inflamed area.

Properties of Leukotaxine. The exact chemical identification of leukotaxine requires further studies. These are now in progress. As pointed out above, this substance is certainly not histamine. Physiological tests indicate that it is neither acetylcholine nor an adenylic compound (Menkin and Kadish, 1938). A recent quantitative analysis of a sample of leukotaxine (fraction N; see Menkin, 1940a)

has revealed the following data: C, 50.67%; H, 9.24%; N, 7.65%; Cl, 1.3%; P, 2.06%; S, none; residue, 2.81%. A trace of sodium (6.4%) has been recovered from one of the partially purified fractions studied (C-D fraction, Table 1).

When separated in large measure from inorganic substances by any of the methods described (Menkin, 1938a, b, 1940a, Table 1), the active material usually appears as a slightly brownish, hygroscopic, doubly refractive crystalline material.

Table 1. Scheme of Extraction of Leukotaxine from Inflammatory Exudates (Permeability Factor)



Its solubility in water seems to depend somewhat on the extent of what appear to be adhering impurities, possibly lipoidal in character. The C-D fraction (Table 1), for instance, as reported previously (1940a), is relatively insoluble in water, forming a cloudy suspension in this medium. On the other hand, the C_1 fraction (Table 1) which is presumably further purified is at least moderately soluble in water. Leukotaxine is dissolved by glacial acetic acid or nitric acid but not by hydrochloric acid. The purified fraction (*e.g.*, C_1) is weakly soluble in acetone. An excess of this organic solvent will precipitate out the active material from the pyridine supernatant fraction. It is insoluble in ether. It appears to be soluble to varying degrees in all the lower alcohols tested, being extremely soluble in methyl alcohol.

Hydrolysis for about 20 hours in 5N HCl or 9.5N NaOH inactivates the material. Hydrolysis, however, with normal NaOH fails either to destroy its activity or to dissociate the permeability from the chemotactic factor by inactivation of the former (cf. this fact with the statement of Bier and Rocha e Silva, 1938b).

The material is thermostable. When brought to 100° it is found still active. Heating it *in vacuo* at 85° C for eleven hours has failed to reduce its activity. In a similar fashion, when exposed for several days at −20° the material retains full potency. In its present stage of purification the active material fails to show a sharp

melting point. From about 200° leukotaxine begins to char; and at 300° it still has not melted.

The active permeability factor readily diffuses through a cellophane membrane upon dialysis of the untreated exudate (Menkin, 1936a, 1941a). Dr. Sheldon C. Sommers, in the writer's laboratory, utilizing the osmotic method of G. Barger (Pregl, 1937) determined the molecular weight of leukotaxine (fraction C_1). His results indicate values ranging from 2300 to 7400 with most values falling around 6000-7000. When dextrose was used as the comparison substance the values for the molecular weight of leukotaxine were all of a lower order being around 2500. On the other hand, when $CaCl_2$ was utilized in the method as a more hygroscopic substance and thereby as a comparative substance more akin to leukotaxine, *i.e.*, in this particular respect only, the values were of a distinctly higher order, averaging 6625. On the other hand, the diffusibility of the active material through a cellophane membrane, as revealed by the dialysis of the untreated exudative material, suggests that the molecular weight of leukotaxine is presumably somewhat lower than 5000. It is conceivable that some of the slightly higher values obtained may well be referable to impurities in the material at its present stage of purification. It is to be recalled also that results with the particular method employed in the determination of molecular weight can at best be considered merely as a first approximation. With further purification, studies with other recognized methods will doubtless be undertaken.

The purified active substance yields a negative biuret reaction indicating that it probably is not a protein. The Millon test is likewise negative. The xanthoproteic test is positive indicating thus the presence of a benzene nucleus in an intermediary product of protein catabolism. The Adamkiewicz reaction is likewise positive, pointing to the presence of tryptophane or at least of the indole nucleus. Quantitative determinations of the tryptophane content of leukotaxine by the method of Adamkiewicz and Hopkins (see Block and Bolling, 1940) indicate an average concentration of 1.37 per cent. The ninhydrin test is strongly positive, supporting thus the view that the active substance has at least one alpha-amino group that is free. The reaction with lead acetate or the nitroprusside test fails to reveal the presence of sulphur. These various tests in addition to the diffusibility and thermostability of leukotaxine point to the view that this active permeability factor is not a protein but probably an intermediary product of proteolysis. In this connection it is interesting to note that a proteolytic enzyme, incubated with an otherwise relatively inactive blood serum, favors the formation of products that increase vascular permeability and induce leukocytic migration (Menkin, 1938c). Duthie and Chain (1939) have recently confirmed in large part the studies and views of the writer on the nature of leukotaxine by obtaining a polypeptide from the peptic hydrolysate of blood fibrin that induces an identical biological effect.

Proteolysis is a cardinal feature of the inflammatory reaction (Opie, 1910; Menkin, 1938b, 1940a, 1941b). The level of amino acid nitrogen in exudates is definitely higher than that in blood serum, whereas the total protein concentration in these two fluids is reversed, that in exudate being definitely lower than in blood (Menkin, 1938b). This type of reciprocal relationship confirms the observations of a number of workers concerning the proteolytic activity of exudates, leading thus through this process to the formation of a number of split products. That one of these, or perhaps a group of related compounds may play an important rôle in the mechanism of increased capillary permeability seems a reasonable supposition. The available evidences support this view. The relative homogeneity of the active crystalline material recovered from exudates favors the idea of a single substance to explain the permeability phenomena. The possibility, however, of the existence in exudates of other active substances is not precluded by the data on hand. The present

evidence merely indicates that there exists a factor in whole exudates which alters the normal permeability of capillaries. Analysis of exudates from various species (dog, rabbit, man) reveals the presence of what appears to be a single crystalline substance which *per se* is capable of producing the same effect on the capillary wall as the whole exudate.

The various tests outlined above, in addition to the apparent zwitterionic or dipolar ionic property of the permeability factor, the capacity of saturated ammonium sulfate and of 5 per cent phosphomolybdic acid to precipitate it, and finally its diffusibility through a cellophane membrane, all suggest that the active crystalline substance may well be a relatively simple intermediary product of protein catabolism. It appears unlikely that it is either a proteose, a peptone or an amino acid. The available data are not inconsistent with the fact that this active substance can belong to the group of simple polypeptides. As will be pointed out below, however, this does not preclude the possibility of an additional prosthetic group.

Dr. Sommers has studied, by the method of Van Slyke (1929), in the laboratory of the writer, the amino nitrogen concentration of leukotaxine before and after hydrolysis. The outcome of the measurements may be briefly listed as follows: The amino N concentration of leukotaxine before hydrolysis averages 2.03 mg per 100 mg. This figure represents the results of seven determinations. Leukotaxine samples were then hydrolyzed for about 18 hours in 50 per cent HCl. The amino N content after hydrolysis was found to average 3.54 milligrams per 100 milligrams. These results indicate a ratio of 1.74 when the amino N after hydrolysis is compared with that prior to subjecting the material to such procedure. These facts at first might in themselves suggest that one is possibly dealing with a dipeptide. On the other hand, the nitrogen content of leukotaxine averages 8.65 per cent.* Since the amino nitrogen of leukotaxine was found to average 2.03 per cent, this indicates that possibly a large amount of nitrogen in the molecule is not located in the peptide linkage. This fact, therefore, may necessitate the consideration of a possible prosthetic group, such as, for instance, a nucleic or purine ring (but evidently not an adenylic compound, cf. Menkin and Kadish, 1938). These various possibilities are presented to the reader with utmost caution, for it is to be recalled, for instance, that several hours may be required for the release from the guanidine group of arginine of an amount of nitrogen equivalent to one half the total nitrogen of arginine. Since this fact is not reckoned with in the determinations of the amino nitrogen of leukotaxine by the rapid Van Slyke technique, it is entirely possible that some of the nitrogen of leukotaxine might still be bound up in one of the amino acid components. Furthermore, the magnitude of the molecular weight of leukotaxine, granted that it is acceptable as a first approximation, tends to rule out a dipeptide linkage as the sole structure involved in the leukotaxine molecule. The plausibility of a prosthetic group must, therefore, be seriously envisaged. The negative Molisch test would tend to rule out a carbohydrate linkage. A lipoidal grouping or a purine ring in the molecule, besides other possible prosthetic components, remain distinct possibilities. On the other hand, it is to be emphasized that the enzymatic work of the writer (1938c) and of Duthie and Chain (1939) in addition to the various properties of leukotaxine, as outlined above, seem to support the view that in the molecule of this substance there is a relatively simple peptide chain which *per se* appears to be responsible for at least some of its biological activity.

The mechanism of action whereby leukotaxine is capable of increasing capillary, or for that matter, cellular permeability in general, (Menkin, 1940b) is at present

* The nitrogen content of 7.65% cited previously for a sample of leukotaxine (Fraction N, Menkin, 1940a), was obtained by the method of Dumas, whereas the remaining analyses were performed by the method of Wong (1923) and of Folin and Denis (1916). This may in part account for the slight differences in results.

being investigated. The various recent studies on surface tension and particularly an application of the surface film technique of Langmuir loom as distinct possibilities in this endeavor.

Duran-Reynals (1929) described a spreading factor present in extracts of certain tissues, particularly the testis. This agent facilitates the extension of India ink from the site of its cutaneous inoculation. He referred the phenomenon to an enhancement of tissue permeability by the organ extract employed. Observations indicate that the crystalline material recovered from exudates (leukotaxine) is probably unrelated to this spreading factor. The Duran-Reynals substance is relatively thermolabile and non-dialyzable, whereas the permeability factor in exudates is both thermostable and dialyzable.* Preliminary studies have shown that in testicular extracts both factors may be present, and that the spreading factor of Duran-Reynals may be dissociated from a heat stable permeability factor which is apparently similar to, if not identical with, the one recovered from inflammatory exudates (Menkin, 1937). Furthermore, the recent studies of Rigdon (1940) essentially confirm the earlier observations of the writer. This investigator finds a permeability factor in various normal tissues by extracting them with saline. It is dubious whether his contention that such a factor is distinct from leukotaxine is fully warranted. Grinding, mashing and teasing a tissue remove all aspects of normality from it. The liberation by this severe procedure of leukotaxine, or at least of a substance closely related to it, is therefore not precluded. Thus the writer, as stated above, has been able to demonstrate the presence of a factor apparently identical with leukotaxine when subjecting normal testicular tissue to the severe treatment involved in saline extraction (1937). A few years ago in unpublished studies the writer has also succeeded in showing the presence of such permeability factor in the extracts of several other organs; but muscle tissue, in accordance with similar findings by Rigdon (1940), was found to contain a relatively low concentration of the active material.

The observations brought forth above suggest that the increased capillary filtration of inflammatory conditions is primarily referable to: 1, a transient elevation in capillary pressure; 2, an increase in capillary permeability.

Do reactions on the part of nerves play a rôle in inducing these capillary changes? This question as yet remains more or less unanswered. Cohnheim (1889) maintained that the primary *modus operandi* accounting for capillary filtration was a molecular alteration of the endothelial wall presumably effected directly by the irritating substance. The result is a prompt increase in capillary permeability. The vasomotor nerves, according to Cohnheim, play no part in the reaction. The reasons presented for arriving at this conclusion are perhaps open to doubt, particularly as Cohnheim seemed to be of the opinion that permeability changes are detectable only some hours after the application of an irritant. In this connection Menkin (1932) has demonstrated an increase in capillary filtration as early as two minutes after the introduction of certain irritants.

Bruce (1910) studied the inflammation produced in the conjunctiva of rabbits by a drop of mustard oil. He reported that the maximal capillary dilatation could be greatly diminished by local anesthesia with cocaine. Furthermore the dilatation fails to appear when the cut nerves are allowed to degenerate. He therefore concluded that the effects must be due to a local nervous mechanism. He referred the mechanism to an axon reflex located in the sensory fibers. These results have essentially been confirmed by Bardy (1915) and by Breslauer (1919).

Krogh (1920) observed that the application of iodine to the frog's tongue produced a violent contraction of the underlying muscles. The capillaries of the mucous membrane strongly dilated and the dilatation extended for a considerable distance. After the application of cocaine to paralyze the sensory nerves and the

* The recent studies of Chain and Duthie (1940) and of others indicate that the spreading factor and the enzyme hyaluronidase may be identical.

nerve endings, a drop of iodine failed to elicit the usual effect. When iodine was applied to the mucous membrane after section and degeneration of the lingual nerves, the affected capillaries became partially dilated over a localized area. Krogh (1922) concluded that, at least for certain substances, nervous reactions play some part in the initial inflammatory symptoms. The nerves which are responsible for the reaction are most likely sensory fibers that induce dilatation through local axon reflexes.

Ricker and Regendanz (1921), after studying the local action of a number of substances on the pancreas and conjunctiva of the rabbit, reject the views of Cohnheim in regard to the rôle of the vascular nervous system in inflammation. These investigators assume, with, however, only meager evidence, an intricate arrangement of vasomotor nerves consisting of constrictor and dilator fibers supplying both arteries and capillaries. They contend that tissue alterations caused by the presence of an irritant are, in the last analysis, due to local circulatory disturbances arising through stimulation of involved vasomotor nerves.

More recently the studies of Inutsuka (1928) confirmed the earlier observations of Samuel (1890) and the later ones of Meltzer and Meltzer (1903) on the effect of sectioning the cervical or sympathetic nerves in enhancing exudation. This investigator demonstrated that excision of sympathetic ganglia induces marked exudation.

The stimulation of the peripheral ends of sensory nerves induces vasodilatation of skin vessels (antidromic impulses). Lewis and Marvin (1928) and Ungar *et al.* (1936) are of the opinion that such stimulation favors the liberation of histamine in the skin. Wybauw (1936) on the other hand believes that antidromic impulses release acetylcholine. The available data fail to settle this controversial point. At any rate there seems to be no strong proof that either one of these substances, even though perhaps released, is necessarily concerned with inducing the cutaneous vascular dilatation which results from stimulation of sensory nerves.

The foregoing observations indicate that local nervous reflexes probably play a rôle in modifying the initial development of the inflammatory reaction. The experiments of Landis (1931), cited previously, suggested that the transient increase in capillary pressure following the application of an irritant was apparently referable to a local axon reflex. It is conceivable, in view of these facts, that nerves play a rôle in inflammation merely by inducing a temporary increase in capillary pressure thereby favoring outward filtration. On the other hand, there is little doubt that, as pointed out by Cohnheim, an acute inflammation may develop in the absence of all central nervous influences. The persistent and far more significant phenomenon of increased endothelial permeability in inflammatory processes appears to be primarily referable to the liberation of the permeability factor (leukotaxine) by injured tissue. Whether the release of this active crystalline substance is any way related to initial local axon reflexes; or whether, on the contrary, its presence plays some definite part in the development of such reflexes still remains to be determined.

3. The Migration of Leukocytes into Inflamed Tissue. The preceding sections have dealt with capillary changes brought about by the presence of an irritant. The increase in capillary permeability allows the outward passage of plasma proteins. The early formation of fibrin in the inflamed area plays an important rôle in immunity by circumscribing the irritant. This phase has been discussed at length previously (Menkin, 1938a, 1940a). It will be considered again but only briefly in a subsequent section.

The significance of phagocytosis in the disposal of an irritant was stressed primarily by the classical studies of Metchnikoff (1892). This phase of the field has been already discussed at length, and lack of space prevents any detailed reference to it (cf. Mudd, McCutcheon, and Lucké, 1934; Menkin, 1940a). In brief, it has been pointed out that phagocytosis may be influenced by various factors, such as, for instance: temperature, osmotic pressure, the presence of certain ions, or the pH. The most important element that seems to favor phagocytosis consists in the presence

of opsonins or tropins from either normal or immune serum. By forming a protein film around the particles or bacteria, these phagocytosis-promoting substances enhance the phagocytic capacity of leukocytes. The globulin fraction of serum seems to contain these active substances capable of favoring phagocytosis by both polymorphonuclear leukocytes and macrophages. The relation of phagocytosis to various surface properties is to be borne in mind, as well as a consideration of the problem from the standpoint of interfacial forces.

The importance of the emigration of leukocytes into injured tissue was demonstrated in the middle of the nineteenth century by Cohnheim (1867). The attraction of leukocytes to a site of injury is an old concept dating back to the work of Leber in 1879. This investigator found a chemotactic substance in the bodies of dead cultures of staphylococci. This substance, which he called *phlogosin,* was found capable of attracting leukocytes in tissues. The early work of Massart and Bordet indicated the chemotactic influence exerted by the products of leukocytic disintegration (1891). Gabritchevsky (1890) studied a number of substances, such as lactic acid, which actually have a repellent action on white cells (the so-called "negative chemotaxis"). Dead or living cultures of bacteria, on the other hand, were found to incite strong positive chemotaxis. Buchner (1890) reported the finding of a protein from the Friedlander organism, capable of exerting a strong chemotactic influence. He found that glycine and leucine were definitely chemotactic whereas tyrosine and trimethylamine failed to attract leukocytes. Wolf studied the phenomenon of chemotaxis *in vitro* (1921). She found that the calcium ion was the only inorganic ion which *per se* was positively chemotactic. She reported that, to a certain extent, all amino acids and amines are likewise positively chemotactic.

Grand and Chambers several years ago reported that mechanically injured tissues liberate thermolabile substances which are positively chemotactic for polymorphonuclear leukocytes (1937). Dixon and McCutcheon studied chemotropism of polymorphonuclear leukocytes *in vitro.* Their results led them to conclude that the chemotropic response is one of direction only. The rate of motion depends on other factors such as the osmotic pressure of the medium, temperature, or the internal condition of the cell (1936). These workers found that, whereas polymorphonuclear leukocytes were strongly attracted by staphylococci or tubercle bacilli, lymphocytes failed to exhibit any chemotropism (1935). Very careful observations on the emigration of leukocytes in the living animal (amphibian larvae and rabbits) were recorded by Clark and his collaborators (1920, 1936). In general these investigators noted that the emigration of leukocytes commences on the average about 2½ hours after the injection of the irritant (croton oil). The majority of the cells were polymorphonuclear leukocytes. Relatively fewer monocytes or lymphocytes migrated through the endothelial wall. Furthermore, the period of emigration lasted only several hours. After a while the endothelial wall seemed to undergo a reversal in consistency which prevented further migration. Strong chemotactic response by polymorphonuclear leukocytes was not necessarily accompanied by active phagocytosis on the part of these cells.

The mechanism of the migration of leukocytes to an area of injury has not yet been satisfactorily explained. Nordmann and Ruether (1930) believe that the phenomenon is dependent on a retardation of blood flow in an inflamed area. The hypothesis that migration of leukocytes is referable to changes in surface tension or that the difference in potential between normal and injured tissue favors the wandering into injured tissue of negatively charged leukocytes is still in large part the result of inference (Abramson, 1927, 1933; Weden, 1933). There is as yet little direct experimental observation to substantiate these views. Grant and Wood (1928) reported that histamine fails to induce leukocytic migration. They, therefore, concluded that the release of histamine from injured tissue does not provide a

complete explanation of the process of inflammation. Moon (1935) essentially confirmed these observations.

As described in the foregoing, the writer isolated a crystalline nitrogenous substance from inflammatory exudates which is *per se* capable of inducing a marked increase in capillary permeability. The substance is concerned not merely with an alteration in the permeability of small vessels but also with the mechanism of leukocytic migration (Menkin, 1937, 1938d). The introduction into the skin of rabbits of this active material induces an almost immediate increase in capillary permeability followed by rapid migration of polymorphonuclear leukocytes. Within 15 to 20 minutes the lumina of the small cutaneous vessels are crowded with these cells, many of which adhere to the endothelial lining. Within 30 to 40 minutes the extracapillary areas are seen to contain a considerable number of polymorphonuclear leukocytes that have migrated outward from the blood stream (Fig. 1).



FIGURE 1. Cutaneous vessels 41 minutes after the intradermal inoculation of leukotaxine. The migration and perivascular infiltration of polymorphonuclear leukocytes are striking. × 680 approximately. (From Valy Menkin, *J. Exp. Med.*, 67, 145 (1938).)

This cellular reaction does not seem to be referable to the preliminary enhanced capillary permeability. Strong irritants, such as aleuronat or turpentine, although producing a prompt increase in the permeability of capillaries fail to induce a rapid chemotactic effect. Furthermore, this active substance, which, as pointed out above, has been named *leukotaxine,* is *per se* chemotactic as indicated by *in vitro* tests, as well as by observations on the accumulation of white cells in glass capillary tubes containing the active material. For instance, placing particles of leukotaxine on a slide induces a rapid collection and characteristic clustering of leukocytes around such particles (Menkin, 1940a). The available evidence supports the view that the permeability factor and the chemotactic factor are referable to one and the same crystalline substance. Leukotaxine, as pointed out above, seems to belong to the group of relatively simple polypeptides with perhaps a prosthetic component.

It is conceivable that the formation of leukotaxine from injured cells results from an interference with normal protein catabolism induced by the presence of an irri-

tant. The basic sequences of the inflammatory reaction, namely the initial increase in capillary permeability followed by leukocytic migration, are thus perhaps referable to a common denominator: the liberation of leukotaxine. Variation in type and in intensity of cellular infiltration may be in part a function of the local concentration of leukotaxine. There are doubtless other factors involved in the development of the inflammatory reaction. For instance, fibrin formation and thrombosis of small vessels do not seem to be directly related to the presence of leukotaxine (Menkin, 1938a). Unpublished observations indicate that leukotaxine, in addition to altering the permeability of the small vessels and inducing rapid leukocytic migration, produces probably little tissue injury. The latter is apparently due to another factor termed *necrosin* which will be described subsequently. There is, however, some evidence that leukotaxine definitely affects normal cell division. Recent studies by the writer on the ova of sea urchins (Arbacia punctulata) indicate that this substance not only increases markedly the permeability of the cell to water but also influences its subsequent development. Fertilization of such leukotaxine-treated ova is followed by a retardation in the rate of cell division and by the appearance of atypical cleavage (Fig. 2). The sperm of sea urchins exposed to leukotaxine are

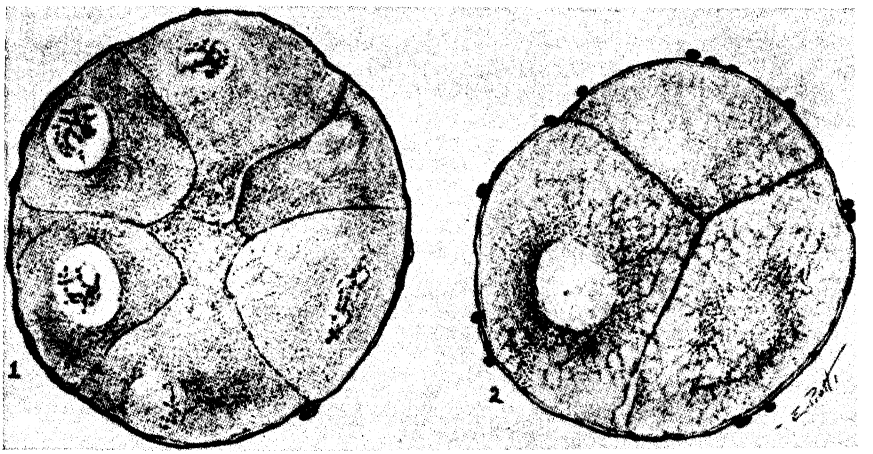


FIGURE 2. (1) Normal cleavage of a fertilized ovum of Arbacia punctulata. (2) Note the unequal cleavage and the retarded development of a fertilized ovum of Arbacia punctulata which had previously been immersed in a medium containing leukotaxine. The interval after fertilization is the same in both ova. From Valy Menkin, *Proc. Soc. Exp. Biol. Med.*, **44**, 588 (1940).

likewise inactivated (1940b). These facts, coupled with the effect on the permeability of the endothelial wall, therefore strongly suggest that leukotaxine induces some degree of relatively mild cellular injury.

Furthermore, studies to be discussed in the next section indicate that the leukocytosis which frequently accompanies severe inflammatory conditions is apparently not referable to the liberation of leukotaxine. This substance seems to be concerned with the mechanism of migration without directly affecting the number of leukocytes in the circulating blood. By such studies the writer has succeeded in dissociating the factors concerned with the mechanism of diapedesis from that which regulates the hematogenous picture of infectious processes (Menkin, 1938a, 1939c, 1940a, c).

The observations described above clearly indicate that the presence of an irritant in tissue induces the liberation of various products of protein catabolism. One of

these split products is probably leukotaxine which is *per se* capable of initiating some of the basic sequences of the inflammatory reaction. It is conceivable, however, that the irritant may also possess chemotactic properties which by superimposition might thus enhance the effect of the liberated leukotaxine. This, if true, may be a factor in distinguishing the ultimate character of various types of infectious lesions. For instance, the writer has recently succeeded in obtaining from a virulent culture of staphylococcus aureus, a substance which when injected intracutaneously in rabbits, induces rapid leukocytic migration (unpublished studies). This active material is probably of the same nature as the chemotactic substance originally described by Leber (1879). It is curious to note that living virulent cultures of this microorganism produce in rabbits relatively little cellular infiltration (Delaunay, 1938). One wonders whether the presence of leucocidin elaborated by the living microbe overcomes its chemotactic capacity. Chemical extraction may very well destroy the relatively thermolabile leucocidin without appreciably affecting the potency of its heat stable leukotaxine-like component. These questions are of definite importance and therefore are accordingly being investigated further.

4. **Mechanism of Leukocytosis with Inflammation.** Does leukotaxine *per se* induce an increase in the level of leukocytes in the blood stream? Although leukocytosis is well known not to be an invariable accompaniment of an inflammatory reaction, some authors, nevertheless, have expressed the opinion that diapedesis and the level of circulating leukocytes may be referable to one and the same factor. Moon (1938), for instance, assumes that "probably the same substances which attract the leukocytes into areas of local injury, produce a systemic mobilization of leukocytes when released in large amounts following extensive injury."

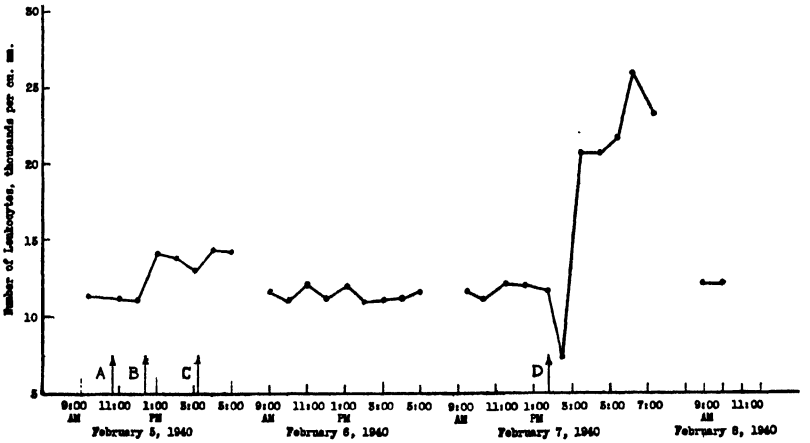
Recent observations indicate (Menkin, 1939c, 1940a, c) that the liberation of leukotaxine evidently bears no relation to the degree of leukocytosis frequently encountered in association with inflammatory processes. Furthermore, the heightened leukocyte level in the blood seems to be referable to the liberation, at the site of injury, of a leukocytosis-promoting factor. The independence of the chemotactic factor from that concerned with leukocytosis is not wholly surprising when it is recalled that certain inflammatory processes characterized by marked leukocytic infiltration can even be accompanied by distinct leukopenia.

a. **The Effect of Leukotaxine on the Number of Leukocytes in the Circulating Blood.** Experiments were performed on rabbits and dogs. When, after several preliminary counts on peripheral blood samples, the normal leukocytic rhythm had been established, leukotaxine, extracted as previously described from inflammatory exudates of dogs, was injected into rabbits or dogs, either subcutaneously or intravenously. The results of numerous experiments show that repeated injections of this substance even over a prolonged interval of time fail to increase the level of leukocytes in the circulating blood stream. The conspicuous degree of cellular infiltration and abscess formation occasionally encountered at the site of repeated leukotaxine injections in rabbits bears evidently no direct relation to the number of circulating leukocytes.

b. **Presence of a Leukocytosis-promoting Factor in Inflammatory Exudates.** Having apparently eliminated leukotaxine as the factor responsible for the state of leukocytosis frequently associated with inflammation, studies were undertaken by the writer in an endeavor to determine whether the active principle might not be liberated in the exudate as a result of tissue injury (Menkin, 1939c, 1940a, c).

Danzer (1930) showed that the injection in rabbits of various extracts of organs deflects the polymorphonuclear count. The extracts were prepared from normal muscle, liver, brain, and testis by shaking the minced tissue in isotonic sodium chloride. Nettleship (1938) recently reported experiments on the production of acute inflammation by hemolytic streptococci. The procedure induces leukocytosis in rabbits. This worker correlated the rise in white blood cell counts with the extent

of cytoplasmic damage in leukocytes at the site of inflammation. From this type of histological correlation Nettleship concluded that substances released from injured leukocytes are responsible for the developing leukocytosis. Ponder and MacLeod (1938) recently expressed the opinion that in the blood stream of rabbits with peritonitis the resultant shift to the left of polymorphonuclear counts is probably referable to the absorption of breakdown products of the cells appearing first in the exudate. These investigators demonstrated that the repeated injection of an irritant into the peritoneal cavity is correlated with a showering in the blood stream of very young polymorphonuclear leukocytes of class I or even of metamyelocytes.



A. 6 to 7 cc. Pentnucleotide intracard.
B. 9 to 10 cc. Pentnucleotide intracard.
C. 10 cc. Pentnucleotide intracard.
D. Globulin fraction derived from about 6 cc. exudate.

FIGURE 3. Repeated injections of pentnucleotide solution into the blood stream of a dog induced only a slight increase in the number of circulating leukocytes, whereas one injection of the LPF, recovered as the active globulin from a sample of exudate, induced a marked rise in the white blood cell counts.

Inflammatory exudates were obtained, as described previously, from dogs subsequent to intrapleural injections of 1.5 cc of turpentine (Menkin, 1934).* The duration of the inflammatory reaction prior to thoracentesis varied from one to several days. From 15 to 25 cc of the freshly removed exudate was injected by intracardiac puncture into the circulating blood stream of a normal dog.

The results of the experiments pointed to a conspicuous rise in the leukocyte counts several hours after the administration of an exudate. The increase in leukocytes averaged about 75 per cent. The normal variation in the leukocyte counts of a number of dogs averaged about 25 per cent. Therefore, the effect of the leukocytosis-promoting factor induced a three-fold increase in the number of circulating leukocytes. The outcome of a few of the experiments, however, suggests that an exudate from an animal with marked leukopenia contains a reduced amount of leukocytosis-promoting factor.

The identification of the leukocytosis-promoting factor in inflammatory exudates of animals with concomitant leukocytosis as well as the failure of leukotaxine to alter the level of circulating white blood cells, indicates that the migration of leukocytes to the point of injury seems scarcely referable to the same mechanism which

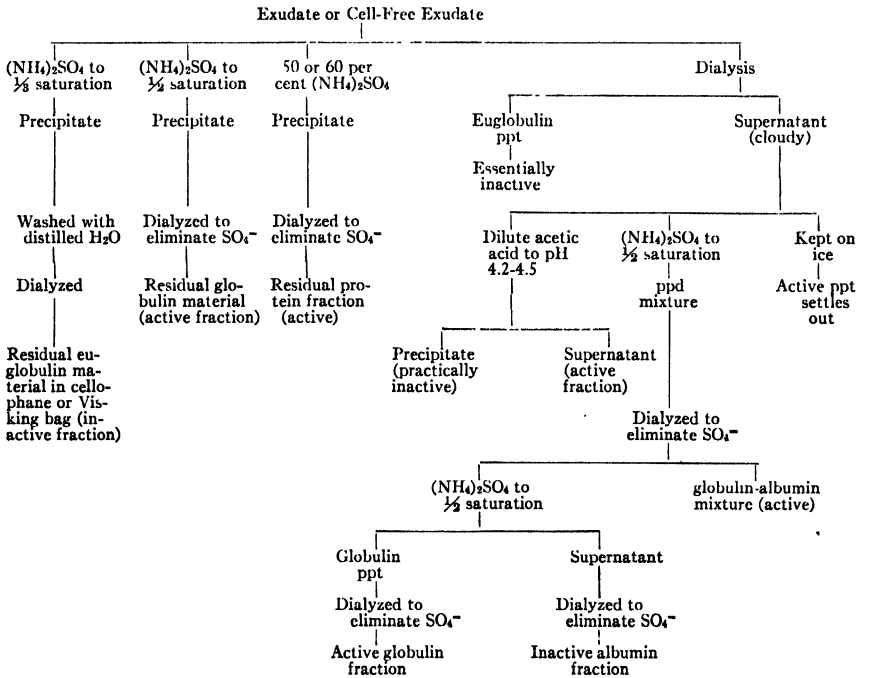
* Other types of exudates were also utilized, e.g. as obtained following a burn, injection of croton oil in olive oil, or from pleural effusions in man.

is responsible for the accompanying leukocytosis. Contrary to the properties of leukotaxine, the leukocytosis-promoting factor is thermolabile and essentially indiffusible through a cellophane membrane. The action of this factor seems to be on the bone marrow, causing a discharge of relatively immature leukocytes into the circulating blood stream (Menkin, 1940c). Studies have indicated that the leukocytosis-promoting factor fails to exert precisely the same type of effect on leukocytes as is elicited by adenosine, histamine, yeast nucleic acid or pentnucleotides. The effect of pentnucleotides on the circulating leukocytes is considerably less intense than that obtained with the leukocytosis-promoting factor (see Fig. 3).

Some of the properties of the leukocytosis-promoting factor have suggested the possibility that it may be a protein or perhaps a product of protein catabolism. Protein fractionation of exudates, presently to be described, supports the view that the leukocytosis-promoting factor is a globulin or is at least in close association with the globulin fraction of exudate.

c. **Protein Fractionation of Inflammatory Exudates.** In an endeavor to identify the leukocytosis-promoting factor in exudates, the material was analyzed by several different methods. The general scheme utilized is summarized in Tables 2 and 3

Table 2. Scheme of Extraction of Leukocytosis-Promoting Factor



(Menkin, 1940d, e). In brief, the main features indicate that the euglobulin fraction of exudates obtained by dialysis or else by precipitation with ammonium sulfate at one-third saturation is essentially inactive. On the other hand, the fractions obtained after precipitation at one-half saturation or with 50-60 per cent of $(NH_4)_2SO_4$ contain the leukocytosis-promoting factor. Precipitation at pH 4.2-4.5 with acetic acid fails to carry down the active principle. Whether this precipitate is a protein nucleate or a nucleo-protein remains to be investigated (Sevag and

Smolens, 1941). The leukocytosis-promoting factor (termed the LPF) is not present in the albumin fraction, as shown in Table 2. These facts therefore seem to localize the LPF in the pseudoglobulin fraction of inflammatory exudates. The material can be obtained in the dried state with relatively little loss in its biological

Table 3. Purification of LPF

cell-free exudate
|
$(NH_4)_2SO_4$ to saturation
|
precipitate
|
dialysis
|
residual globulin material
|
ether (⅓ volume of residual material)
|\
| ether—containing layer discarded
$(NH_4)_2SO_4$ at ½ saturation
|
precipitate
|
dialysis
|
residual content
|
evaporated by dryness in vacuo (t. 35° C)

potency. The effects of the various fractions derived from the exudative material are graphically represented in Fig. 4. The course of an experiment is also illustrated in Fig. 5.

The action of the LPF is evidently on the bone marrow, causing a discharge of immature granulocytes into the circulation (Menkin, 1940c). Recent studies indicate that this substance presumably penetrates from the site of inflammation into the circulating blood stream for it can be extracted from the serum of such animals. Under ordinary circumstances it cannot be recovered from the serum of normal

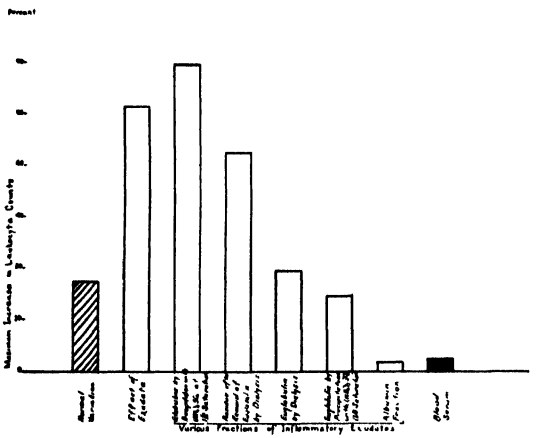


FIGURE 4. A diagrammatic representation summarizing the average effects of exudates, of blood serum and of various fractions of exudates on the level of white blood cells. The normal range of fluctuation in the leukocyte counts of several dogs is shown in the first column.

dogs (Menkin and Kadish, 1942, *Archives of Pathology*, **33**, 193). This fact coupled with the observation that neurogenic influences seem to play no part in the development of the leukocytosis accompanying acute inflammation, strongly points to the action of the LPF on the bone marrow via the blood stream (Menkin, 1940a).

The leukocytosis-promoting factor has been recovered originally from exudates of dogs and rabbits (Menkin, 1940a). Reifenstein and his collaborators have recently confirmed its finding in the exudates of rabbits (1941). Studies by the writer and his associates reveal likewise its presence in inflammatory exudates of man (*Archives of Pathology*, **33**, 188 (1942)).

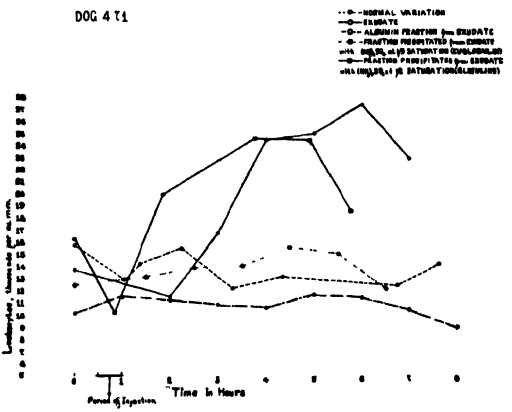


FIGURE 5. The effect of an exudate and of several fractions of exudative material, injected into the circulation, on the leukocyte counts of a dog. The respective fractions utilized and the time of their injection are indicated on the chart.

Longsworth, Shedlovsky and MacInnes (1939) have called attention to the increase in alpha globulin as shown by electrophoretic measurements and to the consequent high value of the resulting proportion of alpha globulin to albumin in the blood serum of patients afflicted with various inflammatory processes. The alpha globulin is presumably in the pseudo-globulin fraction. It is conceivable that the findings of these investigators are due to an excess of globulin containing the leukocytosis-promoting factor that has in turn penetrated from the site of inflammation into the circulating blood. The recent findings, as pointed out above, that the LPF is recovered from the serum of animals with concomitant inflammation would add support to the possibility of such an interpretation (Menkin and Kadish, 1942). It is also possible that the leukocytosis-promoting factor may prove to have clinical application in a number of infectious conditions. Its action on the leukocyte level, as shown previously (Fig. 3), seems to be considerably more effective than that obtained with a commercial preparation of pentnucleotide.*

In conclusion, the mechanism of increased capillary permeability and of leukocytic migration in an inflamed area seems primarily referable to a thermostable, diffusible, nitrogenous substance termed leukotaxine, which is presumably a split product of protein catabolism. On the other hand, the rise in the number of circulating leukocytes seems to be referable to the presence of a thermolabile, non-diffusible protein-like substance in exudates which appears to be either a pseudo-globulin or at least to be linked primarily with this protein fraction. It presumably penetrates into the circulation from the site of inflammation whence it is recovered

* The pentnucleotide preparation was obtained through the courtesy of Smith, Kline, and French laboratories of Philadelphia.

in measurable quantities from the blood serum. The leukocytosis-promoting factor induces a discharge of immature leukocytes from the bone marrow. Studies are now in progress in an effort to purify further this active globulin-like substance which *per se* offers an explanation for the mechanism of leukocytosis accompanying inflammation. The recent finding (1943) that injury in inflammation is referable to the liberation of necrosin, which is located in the euglobulin fraction of exudates, and the further observation that necrosin is responsible for the development of leucopenia, has suggested preliminary removal of necrosin in the further purification of the leukocytosis-promoting factor. This step has yielded extremely potent fractions of this LP factor, capable of increasing the leukocyte level 200 to 300 per cent. (Menkin, *Am. J. Med. Sci.*, **205**, 363 (1943).) Finally, recent observations indicate that the leukocytosis-promoting factor induces marked hyperplasia of the bone marrow (Menkin, *New England J. of Med.*, Sept. 23rd, 1943).

5. **Cellular Sequences in Inflammation.** For many years it has been known that the cytologic sequence in acute inflammation is characterized in the earliest stages by active migration of polymorphonuclear leukocytes. After a time this is followed by infiltration of mononuclear phagocytes or macrophages.

Space does not permit a discussion of the various views concerning the origin of the macrophages in inflammation. Suffice it to say that some investigators, notably Möllendorff, believe that macrophages are derived *in situ* by amitotic division of fibrocytes (1927). Weatherford has obtained further evidence indicating that in the earliest phase of inflammation a conversion of fibrocytes into histiocytes can readily take place (1933). Maximow and Bloom on the contrary deny this contention (1935). Bloom believes that most of the mononuclear cells in inflammation migrate from the blood and are transformed lymphocytes and monocytes. Relatively few of the cells originate from the macrophages already present in the affected tissue. Clark and his collaborators (1936) recently called attention to the possibility of artefacts. Injured or degenerating polymorphonuclear cells often round up and present the appearance of mononuclear cells, round cells, or lymphocytes.

In acute inflammation the polymorphonuclear cells that leave the circulating blood stream form the chief cellular constituents of the early exudate. The mononuclear phagocytes or macrophages increase in number in the later stages. This orderly cytologic sequence in the development of the inflammatory reaction was first pointed out by Borrel (1893) and then by Durham (1897) and Beattie (1903). Such a sequence seems to characterize the majority of acute inflammatory reactions caused by bacteria or by chemical irritants.

No adequate explanation had been offered for this fundamental process. The earlier studies of Müller (1888), of Opie (1905, 1910), and later of Weiss (1927) indicate the presence of proteolytic enzymes in leukocytes. Leukoprotease, found in polymorphonuclear leukocytes, is active only in a neutral or slightly alkaline medium. Lymphoprotease recovered from mononuclear phagocytes is, on the other hand, active in an acid medium. These old observations suggested to the writer that there may be a parallelism between the concentration of hydrogen ions in the exudate and the pH which is optimal for the action of the intracellular enzyme of the predominating type of phagocyte. If this assumption were correct, one might expect in acutely inflamed areas the development of a local acidosis with the shift from polymorphonuclear to mononuclear phagocytes. Lord (1919) in his studies of the pneumonic lung had concluded that in the course of the disease an increase in concentration of hydrogen ion in the exudate probably occurs. Schade and his collaborators (1921) reported that pus from acute abscesses had a pH ranging from 5.95 to 6.50. Menkin studied the relation between the pH of the medium and the cytologic picture during the course of an acute inflammatory reaction (1934). The observations indicated that the predominance of polymorphonuclear leukocytes at

the beginning of an inflammatory reaction is correlated with a pH of the exudate ranging between 7.4 to 7.25 or thereabouts. With the progress of the inflammation the pH may fall to a level of 6.9 or 6.8. At this stage the exudate is characterized by a preponderance of macrophages. The polymorphonuclear cells are found swollen, vacuolated, and appear degenerated. Below a pH of 6.8 or 6.7 all types of leukocytes are injured, and frank pus ensues. Observations indicate that there seems to exist a definite relation between the pH and the cytologic picture of an exudate. When the inflammatory reaction is mild the pH persists above 7.0, and the polymorphonuclear cell type predominates. Furthermore, Evans reported that leukocytes are sensitive to the action of acids (1931). The writer has been able to verify this observation. He has thus shown that polymorphonuclear leukocytes are incapable of survival *in vitro* when exposed to a pH of 6.6 or thereabouts, whereas the macrophages are apparently undamaged (1939d). In brief, with a fairly reasonable degree of certainty, the cytologic picture of an exudate can be predicted from its hydrogen ion concentration and vice versa.* The data suggest that the cytologic picture in inflammation is conditioned by the pH of the exudate. Subsequent studies revealed that the mechanism of local acidosis is often referable to a disturbance in the local intermediary carbohydrate metabolism and in the depletion of the alkali reserve (Menkin and Warner, 1937). When the lymphatic circulation and ultimately the vascular channels are obstructed, as is apt to occur in acutely inflamed tissue, the injured area becomes shunted off from the rest of the organism. It develops its own pH, its own fluid exchange, and its own metabolism. With such impairment in the local circulation, a state of relative anoxemia tends to develop. Perhaps an anaerobic glycolytic type of respiration may result. The plausibility of this view is supported by finding an enhanced glycolysis and a depletion in the alkali reserve. The lactic acid level of the exudate rises with a corresponding drop in sugar level (Fig. 6). The observations thus indicate that the cytologic picture seems to be conditioned by the pH which, in turn, is frequently referable to disturbance in the local carbohydrate metabolism. The possible implications of these findings in regard to a clearer understanding of the histologic manifestations of various lesions are obvious. It is conceivable that the relative absence of leukocytic migration in the late stages of inflammation may be referable to a local increased concentration of lactic acid. This acid is, as has been pointed out above, negatively chemotactic. Its formation in appreciable amounts may thus counteract the migration of white cells induced by the liberation of leukotaxine (Menkin, 1938a, d, 1940a).

6. **Chemical Basis of Injury in Inflammation.** Neither leukotaxine nor the leukocytosis-promoting factor induces any appreciable degree of tissue injury. Yet inflammation is a manifestation of severe tissue injury. Recent studies (Menkin, 1943) have indicated that the basic pattern of injury in inflammation is referable to the liberation of a factor located in or associated with the euglobulin fraction of exudates. For the sake of convenience this substance has been termed *necrosin.*

The fundamental pattern of inflammation may be influenced by the nature of the irritant or the anatomical site of the lesion. Nevertheless, the basic pattern seems referable to a common denominator, in turn released by the injured cell. This is necrosin. In brief, the stereopattern reaction of inflammation is referable to an altered cellular chemistry, which as a result liberates various by-products, such, for instance, as leukotaxine, the leukocytosis-promoting factor, necrosin, and even

* Steinberg and Dietz have recently questioned the repeated findings of a great number of investigators concerning the development of local acidosis in acutely inflamed areas (1938). As indicated in detail elsewhere (1938a) by the writer, their contention may primarily be invalidated by a questionable technique in the measurement of pH and by their failure to correlate the leukocytic picture and pH on an identical sample of exudate. Menkin's conclusions have recently received further confirmation in the hands of Lurie (1937) and in the work of Rugiero and Tanturi (1942).

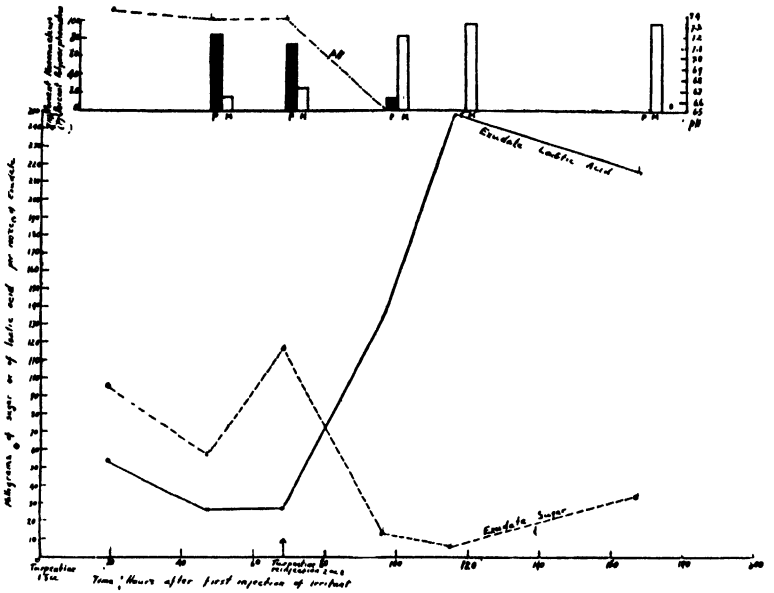


FIGURE 6. Correlation between the hydrogen ion concentration, the conversion of sugar to lactic acid, and the cytological picture of a pleural exudate (Dog 9-2). Note that with the fall in the level of the exudate sugar there is a rise in the concentration of lactic acid and the pH falls notably. This local acidosis is accompanied by a shift in cellular elements. The polymorphonuclear leukocytes are replaced by mononuclear phagocytes.

—·—·—·— pH
———·——— Lactic acid
———o——— Exudate sugar
▮ Percentage of polymorphonuclear leukocytes
▯ Percentage of mononuclear phagocytes

(From: Menkin and Warner, *Am. J. Path.*, **13**, 25 (1937).)

glucose. It is these common denominators which are in turn responsible for the stereopatterned response. Necrosin displays a considerable number of interesting properties. It is absent in normal blood serum, but it is frequently encountered in the circulation of an animal with a concomitant inflammation. This observation suggests that necrosin in absorbed from the site of inflammation into the blood stream. The importance of this fact in regard to the concept of the effect of an inflammation focus to organs at a distance, is obvious. The intramuscular injection of necrosin in dogs invariably injures the liver and very frequently the kidneys. The induced lesions suggest hepatic focal necrosis and small abscesses in the kidneys with injury to the tubules. The possible significance of necrosin absorption in a total explanation of the syndrome of shock, also deserves further study. Furthermore, necrosin induces a severe leukopenia in dogs. It is the only fraction of exudates which causes a sharp fall in the level of circulating leukocytes. This fact has been utilized in a further purification of the leukocytosis-promoting factor (1943b). The significance of this observation (*i.e.*, that necrosin injection is followed by leukopenia) in an endeavor to explain various leucopenic manifestations in certain infectious processes, *e.g.*, typhoid fever, is to be considered. Finally, necrosin is the sole fraction recovered from exudative material which is capable, *per se*, of eliciting fever in dogs. This finding offers definite possibilities in attempting to understand further the mechanism of fever production with inflammation.

Inflammation in Relation to Immunity

In the foregoing discussion an attempt has been made to trace the salient sequences involved in the development of an acute inflammatory reaction. The interdependence of the various manifestations concerned in the disturbance of local fluid exchange, increased capillary permeability, and leukocytic migration produce a complex mechanism which tends to localize and ultimately dispose of the irritant. Some of the implications of this concept of inflammation will now be briefly considered.

The inflammatory reaction plays a definite rôle in localizing the irritating factor, or at least limiting its dissemination, whether it be a bacterial or a chemical irritant. This capacity, which may appropriately be termed *fixation*, contributes a further significance to inflammation in relation to immunity. By delaying the dissemination of obnoxious substances into the circulating blood stream, the inflammatory reaction protects the organism as a whole at the expense of local injury. This view of inflammation has gained considerable ground in the last ten years. Analyses have demonstrated that fixation at the site of inflammation is primarily due to mechanical obstruction caused by a network of fibrin and to occlusion of lymphatic vessels (Menkin, 1931b, c, 1938a, 1940a). Studies on the time relationship have shown that with powerful irritants fixation is the initial phase in the development of a severe inflammatory reaction, and that it precedes even the migration of leukocytes. In view of the available evidence one can ascribe a definite function to the early formation of lymphatic blockade at the site of inflammation. *Fixation* plays a definite rôle in the relation of inflammation to immunity. It may be considered in the same light as the migration of leukocytes and the stage of fibrous repair that follow.

The earlier literature on the rôle of inflammation in immunity dating from the studies of Issayeff (1894) has been considered in detail elsewhere and will therefore be omitted from the present discussion (Menkin, 1938a, 1940a). Considerable evidence, to which the reader is referred in the foregoing publications, have indicated that various substances including bacteria, when introduced into the circulation, rapidly accumulate at the site of inflammation where the material is fixed owing primarily to the establishment of an effective lymphatic blockade (Menkin, 1931b). The conclusions of the writer on the mechanism of fixation, based on numerous experimental observations, contrary to the statement of Bloom (1940), have received support in the hands of Lurie (1939), Burrows (1932), Clark and Clark (1937), Dennis and Berberian (1934) and of Glenn, Peterson, and Drinker (1942). Minor injury favors the dissemination of injected foreign material into the tributary lymphatic channels (Hudack and McMaster, 1931, 1932, 1934). On the other hand, the "walling off" of an acutely inflamed area seems to be due to an enhanced passage of fibrinogen through the more permeable capillary wall. This fact incidentally points to the importance of a clear understanding of the mechanism of increased capillary permeability in inflammation. The mechanism of fixation, as indicated by three distinct types of evidence, is primarily referable to the formation of a fibrinous network and of thrombi occluding the lumina of draining lymphatics.

The evidence for arriving at this conclusion has been described in detail elsewhere (Menkin, 1931b, 1932, 1935, 1938a, 1940a) and therefore will merely be restated here in brief:

1. Morphological examination substantiates the view that fixation may well be referable to lymphatic blockade caused by fibrin and obstructing thrombi in lymphatic channels (Figs. 7 and 8).

2. It was reasoned that, if mechanical obstruction plays an important rôle in fixation, then substances which are unable to disseminate from an acutely inflamed area should, for the same reason, fail to enter it when injected at the periphery of

the area. This is precisely what was found to occur when trypan blue or bacteria were injected at the periphery of an inflamed focus. Burrows has recently confirmed this experiment with India ink (1932). Studying the living frog under the microscope, the writer has followed the diffusion of a dye from its peripheral inoculation to the boundary of the inflamed area where its further progress was blocked by the presence of a thrombus occluding a lymphatic channel (1931d).

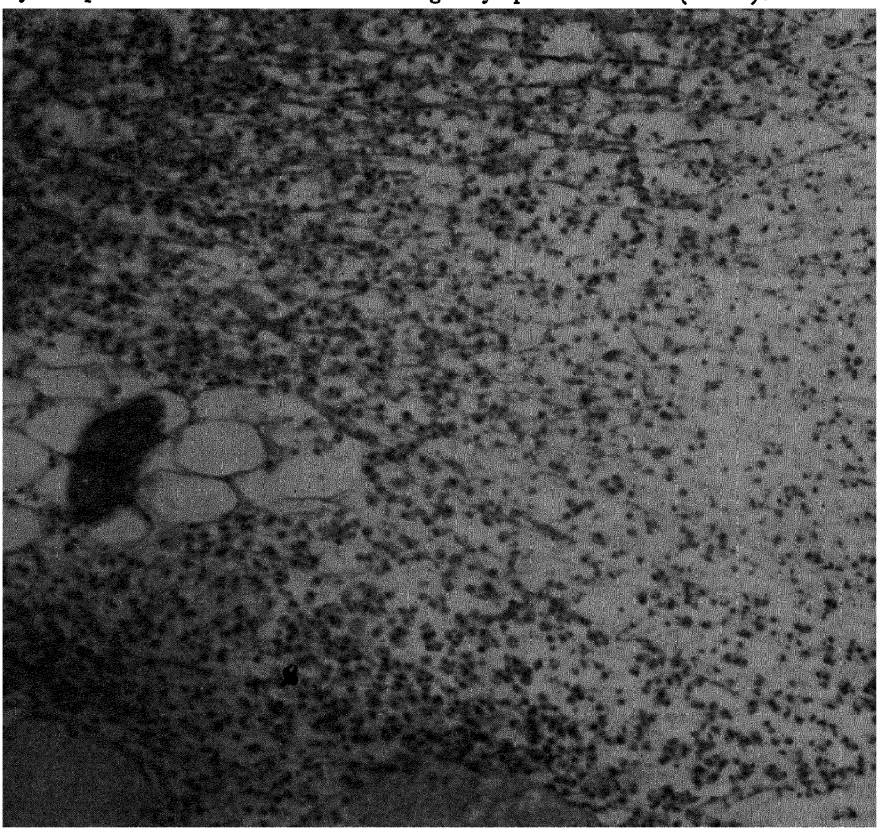


FIGURE 7. Fibrinous network in an area of dense leukocytic infiltration.

3. A third type of evidence which reinforces the hypothesis was obtained by taking advantage of the fact that concentrated urea dissolves fibrin. When this solvent was injected immediately after the irritant, obstructing thrombi failed to form in the tributary lymphatics, and fixation was inhibited. Urea had apparently no other manifest effect on the cytological character of the exudate as revealed by microscopic examination. These three distinct types of evidence render it quite likely, as stated above, that the basic mechanism of fixation in inflammation is primarily referable to an obstructing barrier in the form of thrombi in lymphatics and of coagulated plasma in edematous tissue. There is as yet no precise information on the question of the relation of increased capillary passage to the lymph flow from an inflamed area. Pending such data, the finding of increased lymph flow from a main lymphatic trunk draining a large area of tissue injured unequally by the inflammatory irritant only indicates, as would be expected, an absolute increase in lymph flow. In such an experiment one cannot preclude the possibility of a rela-

tive decrease in the lymph output when the total amount of plasma which has permeated through the capillaries is taken into consideration (Menkin, 1933a). Inflammatory edema may in part be an index of the inability of a partially impaired lymphatic circulation to take care of the excessive passage of fluid from the circulating blood into acutely inflamed tissue. In this connection the interesting observations of Drinker and Field are suggestive and instructive (1933). These investigators have succeeded in developing a conspicuous local edema or elephantiasis in the leg of dogs following obstruction of the lymphatic trunks.

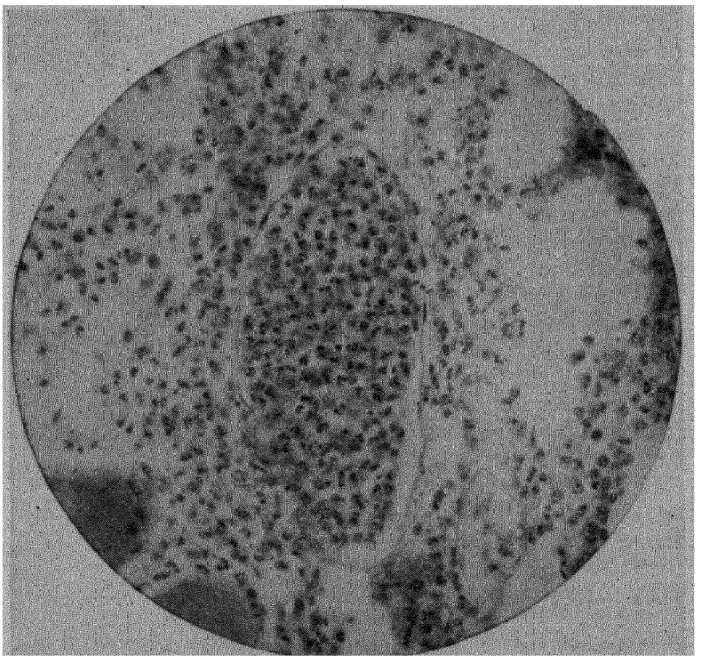


FIGURE 8. Obstructing thrombi in lymphatic channels.

Various secondary factors, such as, for instance, the presence of immune bodies in anaphylactic or allergic inflammation, may reinforce the basic mechanism (Menkin, 1940a). The early occurrence of fixation in a severely injured area plays a definite rôle in immunity, for it allows an interval in which the relatively sluggish leukocytes can assemble for phagocytosis.

a. **Inflammation and Bacterial Invasiveness.** The foregoing observations on the mechanism of fixation have offered a means of studying a factor that may account in large part for the differences in the invasive ability of various pyogenic organisms.

The intensity and rapidity with which an inflammatory irritant is circumscribed in a tissue area varies with the type of irritant. The speed with which an irritant causes a region to be walled off by thrombosed lymphatics, or a fibrinous network, or by both, is necessarily an important index in determining its ability to disseminate ultimately into the circulating blood. This holds true if the irritant can readily drain through lymphatic vessels from the site of inoculation. That bacteria disseminate from the site of inoculation through lymphatics has frequently been demonstrated (Noetzel, 1906; Wells and Johnstone, 1907). It follows from this that, if bacteria are employed as inflammatory irritants, some information may be obtained concerning the invasive capacities of different microorganisms. This would depend on the type of inflammatory reaction which a given microorganism induces

in the host. In other words, can the behavior of various pyogenic organisms in tissues be related, at least in part, to their respective abilities of inducing an early inflammatory reaction at the site of their inoculation which, by lymphatic blockade, would tend to prevent their dissemination to the tributary lymphatic nodes?

In observations described fully elsewhere (Menkin, 1933b, 1935) it was definitely shown that the respective differences in the invasive behavior of *Staphylococcus aureus*, pneumococci, and hemolytic streptococci was in large part referable to differences in the rapidity and intensity of injury induced by these respective microorganisms. A culture of *Staphylococcus aureus* or of its soluble toxin * induces prompt lymphatic blockage, as indicated by histological examination of the site of inoculation, and also by the fact that within one hour after infection an injected indicator dye fails to reach the regional lymphatic nodes. In the case of hemolytic streptococci, however, about two days are required after their cutaneous inoculation before the local reaction is of sufficient intensity to induce lymphatic obstruction. These observations suggest a paradoxical interpretation. Staphylococci produce in general little systemic effect owing to their intense local injurious reaction which fixes them *in situ*. Streptococci produce marked generalized effects on the organism as a whole because of their comparatively mild local reaction. This allows them relatively free penetration to the essential organs.

The sum total of all the observations, described in detail elsewhere (Menkin, 1936b, 1938a, 1940a), clearly indicates the rôle of the inflammatory reaction in regulating the rapidity of invasiveness of a microorganism, and therefore the significance of this factor in grading the immunity or resistance of the host is manifest. However, the rate of dissemination in no way alters the virulence *per se* of the bacteria, which still remains a most important factor in evaluating resistance. Virulence and invasiveness are therefore regarded as two separate variables in studying problems of immunity (Menkin, 1935, 1936b).

The rapidity and the intensity of local fixation in an inflamed area are factors to be reckoned with in determining the invasiveness of an infectious microorganism. Dissemination seems thus to be related directly to time and inversely to the degree of local injury. This may perhaps be conveniently expressed as follows:

$$D = \frac{Kt}{I}$$

where D refers to dissemination, t to time, I to extent of injury, and K is a constant depending on the physical properties of the irritant and perhaps on the anatomical position of the lesion. Powerfully necrotizing irritants produce, as a result of an increase in capillary permeability and lymphatic damage, an extremely prompt "walling-off" reaction (termed *fixation*). By this early process, the area of injury is mechanically circumscribed and the dissemination of the irritant is prevented. This initial response, which may occur within thirty minutes following the injection of an irritant, allows a definite interval of time for the leukocytes to assemble at the site of inflammation for phagocytosis. *Staphylococcus aureus* is an example of such a bacterial irritant. Aleuronat is a chemical irritant of similar potency (Menkin, 1929). Mild irritants, on the other hand, produce only a delayed reaction thus allowing relatively free penetration of the irritant into the circulation for a considerable interval of time. Occlusion of the draining lymphatics in such instances often takes place as late as two days following the introduction of the irritant. Hemolytic streptococci exemplify this type of irritant (Menkin, 1933b). Another

* The effectiveness of the toxin of *Staphylococcus aureus* in inducing lymphatic blockade is destroyed by heating it to about 58° C for one hour. The action of the toxin seems to be caused by the presence of a factor similar, if not identical, to leukocidin (Menkin, 1935).

instance has been recently demonstrated by McMaster and Hudack who showed that, up to 48 hours following a mere skin incision or local burn, lymph drainage is adequate (1934). Subsequently lymphatics failed to convey effectively materials contained in them. The intensity of fixation is found frequently to parallel the extent of inflammatory edema. This, as already mentioned, would suggest that in an inflamed area the local swelling is at least in part the result of blockage to normal lymphatic drainage which is thus unable to cope adequately with the excess outpouring of plasma from the capillaries.

b. **Diabetes and Inflammation.** The condition of diabetes mellitus is known to become markedly intensified when complicated by inflammation or infection. No satisfactory explanation has ever been offered to account for the basic mechanism involved. The inflammatory reaction tends to be concomitantly increased in severity and there seems to exist a lowered resistance to infections. There are some observations indicating a diminished capacity for antibody production and a reduction in the bactericidal power of the blood (Richardson, 1933, 1940). Recent observations of the writer have attempted to explain both the enhanced degree of diabetes and the associated increased severity of the inflammatory reaction in diabetic dogs having a superimposed acute pleural inflammation (Menkin, 1941b; 1941c).

All observations were made on dogs rendered diabetic by pancreatectomy. It is to be recalled in this connection that although this type of diabetes has many obvious points of similarities with the human form, nevertheless there are several points of differences (Long, 1936-1937). Three groups of animals were studied as follows: (1) dogs having a pleural inflammation induced by the introduction of 1.5 cc of turpentine; (2) depancreatized dogs with a pleural inflammation induced by the same irritant, but receiving no insulin immediately preceding and during the period of inflammation; (3) depancreatized dogs with pleural inflammation but receiving repeated injections of low doses of insulin (8 to 10 units twice daily) throughout the period of study.

The presence of an acute inflammation in the right pleural cavity of dogs fails to alter the blood sugar level. On the other hand, a similar injurious reaction in a diabetic dog induces a sharp and rapid ascendency in the blood sugar from an average in the 8 dogs studied of 253 milligrams to a level of 469.1 milligrams per 100 cc. The experimental findings wholly substantiate the well-known clinical picture of a diabetic disorder complicated with infection.

What is the basic mechanism involved to explain the enhanced diabetes? There are two considerations which must first of all be borne in mind. In the first place, as already pointed out, an inflamed area is a focus of proteolysis. This has been pointed out previously by a number of investigators including the writer (1940a). The conspicuous protein catabolic process may be regarded as a cardinal feature of the inflammatory reaction. In the second place, the earlier studies, particularly of Lusk and of his school, have demonstrated that amino products of protein degradation are in the diabetic organism readily converted into sugar (Stiles and Lusk, 1903; Lusk, 1923). Gluconeogenesis from proteins is thus a well-recognized process. It is therefore quite possible that an inflamed area in the diabetic behaves as a focus of enhanced proteolysis, favoring thus the formation of glucose from protein breakdown; the glucose, in turn would gradually diffuse into the circulation. The accentuated degree of proteolysis would offer at the same time a reasonable explanation for the increased severity of an inflammatory reaction in diabetes. Studies have therefore been carried out on both exudate and blood samples of diabetic and non-diabetic dogs to determine the degree of proteolysis at the site of an acute inflammation. Besides measuring the sugar and lactic acid, determinations have also been performed on the total proteins, non-protein nitrogen, urea and amino acid nitrogen concentrations of both exudate and blood.

The results of the determinations on exudates substantiate the view that the inflamed area is an active focus of gluconeogenesis originating from increased protein breakdown. The lactic acid and sugar levels of diabetic exudates are considerably higher than encountered in non-diabetic exudates. The average increase in lactic acid is 52 per cent, while the increase in exudate sugar averages 473.6 per cent. There is a slight drop in total proteins averaging 12.56 per cent. The rise in the products of proteolysis in diabetic exudates is striking. The non-protein nitrogen is increased 89.45 per cent; the urea, 126.3 per cent; and the amino acid nitrogen, 74.29 per cent. These facts therefore indicate an enhanced degree of proteolysis at the site of inflammation in a diabetic animal. That these products of protein breakdown are converted to glucose by deamination is further substantiated by insulin administration. Repeated injections of this substance to depancreatized dogs with inflammation are followed not only by a drop in lactic acid and sugar concentration but also by complete inhibition of the enhanced local proteolytic processes. The total proteins, NPN, urea and amino acid nitrogen are now restored to the levels found in exudates of non-diabetic animals.

The enhanced glucose formation from proteins in the acutely inflamed area allows for the gradual penetration of this diffusible substance into the circulation. Not only is the level of blood sugar, as pointed out above, markedly raised; but there is also a pronounced rise in the concentration of non-protein nitrogen, urea and to some extent of amino acid nitrogen in the systemic circulation. This state of affairs is not referable to pancreatectomy *per se*. Furthermore, the concentration of the exudate sugar is higher than that found in the blood stream, indicating thus that the induced hyperglycemia with inflammation doubtless originates by gluconeogenesis in the inflamed area. The reflection in the blood stream of the rise in a diabetic exudate of intermediary products of carbohydrate and protein metabolism does not contradict the earlier observations of the writer on the fixation or localization of material at the site of inflammation; for, as he has pointed out, particles of smaller dimensions, and of greater diffusibility are less effectively retained in such foci. Such materials therefore gradually find their way into the circulating blood (Menkin, 1938a, 1940a). Inflammation *per se* in non-diabetic dogs fails to induce in the blood any increase in either protein or carbohydrate products of metabolism. Insulin administration by inhibiting gluconeogenesis from proteins at the site of inflammation likewise induces no detectable changes in the levels of the blood.

Finally, the enhanced local proteolysis in the inflamed area of a diabetic animal manifests morphological signs of cellular injury in the form of vacuolized, degenerated leukocytes which are in many instances unidentifiable. The cells studied were from areas of inflammation of several hours' to one day's duration. Comparable studies of non-diabetic exudates revealed normal appearing leukocytes. In agreement with earlier studies by the writer the character and type of the cells in exudates were found to be correlated with the pH and the lactic acid concentration (Menkin, 1934; Menkin and Warner, 1937). Both the hydrogen ion and the lactic acid concentrations were found to be elevated in exudates of diabetic animals.

In conclusion, the mechanism of enhanced diabetes with inflammation seems to be referable to an increased proteolysis in the inflamed area. The rise in protein breakdown processes is correlated both with gluconeogenesis, and with increased severity of the inflammatory reaction. The surplus glucose formed locally gradually diffuses into the circulation thus elevating markedly the blood sugar level (Menkin, 1941c, 1942b).

In a further study (Menkin, 1943b), the writer has shown that at the very beginning of an inflammation in a diabetic dog the level of exudate sugar is considerably more elevated than it is found in the blood stream. This concentration gradient indicates further that gluconeogenesis occurs at the site of inflammation, whence the

glucose gradually diffuses into the circulation. Furthermore, studies on non-diabetic dogs definitely indicate that glucose is formed at the site of injury, but that probably owing to the presence of insulin in non-diabetic dogs, the effect is merely transitory. In diabetic dogs gluconeogenesis at the site of inflammation is sustained and quantitatively exaggerated. But nevertheless from the observations on non-diabetic animals, the principle is evident that injured cells in general are potentially gluconeogenic.

Conclusions

The term inflammation refers to the aggregation of several interdependent reactions incited by the presence in normal vertebrate tissue of a foreign body or so-called irritant. The injurious agent may be living or non-living; it may be of exogenous or even of endogenous origin. Inflammation as such, is a basic phenomenon in infectious processes. An analysis of the mechanisms involved and their implications to the organism as a whole, especially in the light of modern biological concepts, is obviously of considerable significance. Inflammation is essentially one, among others, of the various manifestations of cellular injury. It represents the complex response on the part of normal tissue to relatively severe damage. There are doubtless other forms of reaction to cell injury. Neoplasia, for instance, may yet prove to be the resultant of such a form. Cellular regeneration is perhaps merely another form of elemental reparative response to injurious stimuli. As we probe further into the intricate mechanisms of cellular physiology and pathology, it is conceivable that the term inflammation may become obsolete and may therefore eventually be discarded. The apparently distinct types of cellular injury may come to be regarded as quantitative variations of a single or perhaps of several basic physiological disturbances. What is now regarded as an inflammatory reaction may in the future be considered only as a prototype of severe cellular injury in tissues of higher animals. Reparative reactions may ultimately be analyzed and considered from a similar angle.

The inflammatory reaction is initiated by a derangement in local fluid exchange. It subsequently proceeds through a series of interdependent sequences which ultimately tend to localize and dispose of the irritant. The disturbances in the local physiology of inflamed tissue are closely associated with the immunological significance of the inflammatory reaction. The principal sequences in the development of inflammation, besides the pattern of severe cellular injury referable to *necrosin* liberation as discussed above, may be listed again as follows:

a. **Increased Fluid Passage Through the Capillary Endothelial Wall.** This seems to be primarily referable to two factors: 1. *Elevation in capillary pressure* which is perhaps the outcome of a local axon reflex affecting the caliber of arterioles. 2. *Increased capillary permeability;* this seems to be referable to the liberation by injured tissue of the permeability factor termed *leukotaxine.* Leukotaxine is a nitrogenous substance, the significant properties of which evidently do not resemble those of histamine. It appears as if it may belong to the group of relatively simply polypeptides, although the presence of a prosthetic group in the molecule remains a distinct possibility. It is postulated that its formation occurs as a result of the presence of an irritant interfering with local protein catabolism.

b. **Localization of the Irritant (fixation).** The "walling-off" of an inflamed area seems to be due to an enhanced passage of fibrinogen through the more permeable capillary wall. The mechanism of fixation is primarily referable to the formation of a fibrinous network and of thrombi occluding the lumina of draining lymphatics. This favors the development of lymphatic blockade in acute inflammation. Various secondary factors, such as the presence of immune bodies in anaphylactic or allergic inflammation, may reinforce the basic mechanism. The early occur-

rence of fixation in a severely injured area plays a definite rôle in immunity for it allows an interval in which the relatively sluggish leukocytes assemble for the purpose of phagocytosis.

c. Migration of Leukocytes. The first cells to migrate into an inflamed area are the polymorphonuclear leukocytes. The mechanism of their migration seems to be related to the liberation of *leukotaxine* by injured tissue. The properties of this crystalline nitrogenous substance recovered from exudates offer a reasonable explanation for two of the basic sequences of the inflammatory reaction: first, the initial increase in capillary permeability, and secondly, the rapid emigration of polymorphonuclear leukocytes into injured tissue. Leukotaxine appears to contain factors concerned both with permeability and chemotaxis.

The polymorphonuclear cells are gradually displaced by macrophages. Cytological changes in acute inflammation seem to be conditioned by the pH of the exudate which in turn is often referable to disturbance in the local intermediary carbohydrate metabolism. The development of a local acidosis, resulting from increased glycolysis and depletion of the alkali reserve, seems to injure the polymorphonuclear cells. Macrophages survive and predominate when the pH falls to a level of about 6.9 to 6.8. Further reduction in the pH proves lethal to all types of leukocytes and frank suppuration ensues.

The rise in the level of circulating leukocytes, frequently encountered in infectious processes, seems to be referable to the liberation in the inflamed area of a leukocytosis-promoting factor. This factor is either a globulin or at least it is associated with the pseudo-globulin fraction of exudates. The leucocytosis-promoting factor is also capable of inducing marked hyperplastic growth of granulocytic elements and megakaryocytes in the bone marrow of injected dogs (Menkin, 1943c).

The interplay and dynamic relationships of the above sequences constitute an acute inflammation. An inflamed area can be considered as shunted off from the rest of the organism. It has its own metabolism, its own hydrogen ion concentration, and its own modified circulation. The inflammatory reaction thus displays an extraordinarily complex mechanism tending to localize and dispose of a chemical or bacterial irritant. This ultimately leads to organization and repair of the affected tissue. Recent studies by the writer (1941d) on the relation of cellular injury to proliferative and neoplastic response suggest the possible presence of one or several growth-promoting substances liberated in an inflammatory exudate. Exposure for several months of normal cells to the effect of repeated injections of exudative material induces ultimately in the ears of rabbits a sustained proliferative response on the part of the cartilage and epithelium. There is likewise evidence of metaplastic bone transformation, metaplasia of the epithelium, formation of chondroma-like structures and in a few instances, characteristic papillomatous projections. It is conceivable, in view of such observations which are being studied further, that neoplasia may yet prove to be, as stated above, a manifestation of long standing, but relatively mild, cellular injury brought about either by a protracted inflammatory process or by an alteration in the normal physico-chemical equilibria of the cell. The results of such profound changes in the physiology of the cell may in turn induce the liberation of growth-promoting factors. It is quite possible that ordinary repairative processes at the end stage of the inflammatory reaction are referable merely to the presence of a minimal concentration of such growth-promoting substances.

The increase in local proteolysis at the site of inflammation in diabetes favors both local gluconeogenesis and tissue injury. The surplus glucose diffuses into the circulation and thus offers an explanation for the enhanced hyperglycemia accompanying inflammation. Furthermore, injured cells in general are potentially gluconeogenetic.

The concept of fixation and its mechanism afford a rational interpretation of the

rôle of inflammation in immunity. The capacity of pathogenic microorganisms to disseminate from their site of inoculation is apparently inversely related to the intensity of induced local injury. In this way inflammation plays a significant rôle in problems of immunity as a regulator of bacterial invasiveness. It may be truly regarded as the physical basis of infectious processes.

References

1. Abramson, H. A., *J. Exper. Med.*, **46**, 987; **46**, 1003 (1927).
2. —, *Cold Spring Harbor Symp. on Quant. Biol.*, **1**, 92 (1933).
3. —, Engel, M., Lubkin, V., and Ochs, I., *Proc. Soc. Exper. Biol. Med.*, **38**, 65 (1938).
4. Bardy, H., *Skand. Arch. Physiol.*, **32**, 198 (1914-15).
5. Barsoum, G. S., and Gaddum, J. S., *J. Physiol.*, **85**, 1 (1935).
6. Beattie, J. M., *J. Path. Bact.*, **8**, 129 (1903).
7. Bier, O. and Rocha e Silva, M., *Arqu. Inst. Biol.*, **9**, 109 (1938a).
8. —, *Arqu. Inst. Biol.*, **9**, 123 (1938b).
9. Block, R. J. and Bolling, D., "The determination of the amino acids," Burgess Publishing Co., Minneapolis, 1940.
10. Bloom, W., *American Naturalist*, **74**, 129 (1940).
11. Borrel, A., *Ann. Inst. Pasteur*, **7**, 593 (1893).
12. Bowman, F. B., Winternitz, M. C., and Evans, H. M., *Centralbl. Bakteriol. Parasit.*, **65**, 403 (1912).
13. Breslauer, F., *Deutsch. Ztschr. Chirurgie*, **150**, 50 (1919).
14. Bruce, A. N., *Arch. Exper. Path. Pharmakol.*, **63**, 424 (1910).
15. Buchner, H., *Berlin. Klin. Wchnschr.*, **27**, 1084 (1890).
16. Burrows, H., "Some factors in the localisation of disease in the body," Biallière, Tindall and Cox, London, 1932.
17. Chain, E., and Duthie, E. S., *Brit. J. Exper. Path.*, **21**, 324 (1940).
18. Churchill, E. D., Hakazawa, F., and Drinker, C. K., *J. Physiol.*, **63**, 304 (1927).
19. Clark, E. R., and Clark E. L., *Am. J. Anat.*, **27**, 221 (1920).
20. —, *Am. J. Anat.*, **60**, 253 (1937).
21. —, Clark, E. L., and Rex, R. O., *Am. J. Anat.*, **59**, 123 (1936).
22. Cohnheim, J., *Virchow's Arch.*, **40**, 1 (1867).
23. Cohnheim, J., "Lectures on general pathology," London, New Sydenham Society, 1889.
24. Danielli, J. F., *J. Physiol.*, **98**, 109 (1940).
25. Danzer, M., *Quart. J. Exper. Physiol.*, **20**, 141 (1930).
26. Delaunay, A., "De l'Immunité Antistaphylococcique," Vigot Frères, Paris, 1938.
27. Dennis, E. W., and Berberian, D., *J. Exper. Med.*, **60**, 581 (1934).
28. Dixon, H. M., and McCutcheon, M., *Arch. Path.*, **19**, 679 (1935).
29. —, *Proc. Soc. Exper. Biol. Med.*, **34**, 173 (1936).
30. Drinker, C. K., and Field, M. E., "Lymphatics, Lymph and Tissue Fluid," Baltimore, Williams and Wilkins Co., 1933.
31. Duran-Reynals, F., *J. Exper. Med.*, **50**, 327 (1929).
32. Durham, H. E., *J. Path. Bact.*, **4**, 338 (1897).
33. Duthie, E. S., and Chain, E., *Brit. J. Exper. Path.*, **20**, 417 (1939).
34. Ebbecke, U., *Pflüger's Arch.*, **199**, 197 (1923).
35. Evans, A. C., *Public Health Reports*, **46**, 2539 (1931).
36. Feldberg, W., and Kellaway, C. H., *J. Physiol.*, **90**, 257 (1937).
37. Folin, O., and Denis, W., *J. Biol. Chem.*, **26**, 473 (1916).
37a. Freed, S. C., and Lindner E., *Am. J. Physiol.*, **134**, 258 (1941).
38. Gabritchevsky, G., *Ann. Inst. Pasteur*, **4**, 346 (1890).
38a. Glenn, W. L., Peterson, D. K., and Drinker, C. K., *Surgery* **12**, 685 (1942).
39. Goldschmidt, S., and McGlone, B., *Proc. Am. Physiol. Soc., Am. J. Physiol.*, **109**, 42 (1934).
40. Grand, C. G., and Chambers, R., *J. Cell. Comp. Physiol.*, **9**, 165 (1937).
41. Grant, R. T., and Jones, T. D., *Heart*, **14**, 339 (1927-29).
42. —, and Wood, J. E., *J. Path. Bact.*, **31**, 1 (1928).
43. Hudack, S. S., and McMaster, P. D., *Proc. Soc. Exper. Biol. Med.*, **28**, 853 (1931).
44. Inutsuka, N., *Trans. Jap. Path. Soc.*, **18**, 341 (1928).
45. Issayeff., "Untersuchungen über die künstliche Immunität gegen Cholera," *Z. Hyg. Infektionskr.*, **16**, 287 (1894).
46. Krogh, A., *J. Physiol.*, **53**, 399 (1920).
47. —, "Anatomy and physiology of capillaries," New Haven, Yale Univ. Press, 1922; revised ed., 1929.
48. Kusnetzowsky, N., *Z. Ges. Exper. Med.*, **44**, 646 (1925).
49. Landis, E. M., *Am. J. Physiol.*, **75**, 548 (1926).
50. —, *Am. J. Physiol.*, **81**, 124 (1927); **82**, 217 (1927).
51. —, *Am. J. Physiol.*, **98**, 704 (1931).
52. —, *Physiol. Rev.*, **14**, 404 (1934).
53. Leber, T., *Fortschr. Med.*, **6**, 460 (1888).
54. Lewis, P. A., *J. Exper. Med.*, **23**, 669 (1916).
55. Lewis, T., "The blood vessels of the human skin and their responses," London, Shaw and Sons, Ltd., 1927.
56. —, and Grant, R. T., *Heart*, **11**, 209 (1924).
57. —, and Marvin, H. M., *Heart*, **14**, 27 (1927-29).
58. Long, C. N. H., "Harvey Lectures," p. 194. Baltimore, Williams and Wilkins Co., 1936-37.
59. Longsworth, L. G., Shedlovsky, T., and MacInnes, D. A., *J. Exper. Med.*, **70**, 399 (1939).
60. Loos, H. O., *Arch. Dermat. Syph.*, **164**, 199 (1931).
61. Lord, F. T., *J. Exper. Med.*, **30**, 379 (1919).
62. Lurie, M. B., *Am. J. Path.*, **13**, 612 (1937).
63. —, *J. Exper. Med.*, **69**, 555 (1939).
64. Lusk, G., "The science of nutrition," Philadelphia, W. B. Saunders Co., 1923.
65. MacCurdy, I. T., and Evans, H. M., *Berlin. Klin. Wchnschr.*, **49**, 1695 (1912).
66. McClellan, R. H., and Goodpasture, E. W., *J. Med. Res.*, **44**, 201 (1923).
67. McMaster, P. D., and Hudack, S. S., *J. Exper. Med.*, **56**, 239 (1932).
68. - -, *J. Exper. Med.*, **60**, 479 (1934).
69. McMaster, P. D., Hudack, S. S., and Rous, P., *J. Exper. Med.*, **55**, 203 (1932).

70. Massart, J., and Bordet, C., *Ann. Inst. Pasteur*, **5**, 417 (1891).
71. Maximow, A. A., and Bloom, W., "A textbook of histology," 2nd ed., Philadelphia, W. B. Saunders Co., 1935.
72. Meltzer, S. J., and Meltzer, C., *Am. J. Physiol.*, **9**, 57 (1903a).
73. —, *J. Med. Res.*, **10**, 135 (1903b).
74. Menkin, V., *J. Exper. Med.*, **50**, 171 (1929).
75. —, *J. Exper. Med.*, **53**, 647 (1931a).
76. —, *J. Exper. Med.*, **53**, 171 (1931b).
77. —, *Arch. Path.*, **12**, 802 (1931c).
78. —, *J. Exper. Med.*, **53**, 179 (1931d).
79. —, *J. Exper. Med.*, **56**, 157 (1932).
80. —, *Proc. Soc. Exper. Biol. Med.*, **30**, 1069 (1933a).
81. —, *J. Exper. Med.*, **57**, 977 (1933b).
82. —, *Am. J. Path.*, **10**, 193 (1934).
83. —, *Am. J. Med. Sci.*, **190**, 583 (1935).
84. —, *J. Exper. Med.*, **64**, 485 (1936a).
85. —, *J. Infect. Dis.*, **58**, 81 (1936b).
86. —, *Arch. Path.*, **24**, 65 (1937).
87. —, *Physiol. Rev.*, **18**, 366 (1938a).
88. —, *J. Exper. Med.*, **67**, 129 (1938b).
89. —, *J. Exper. Med.*, **67**, 153 (1938c).
90. —, *J. Exper. Med.*, **67**, 145 (1938d).
91. —, *Proc. Soc. Exper. Biol. Med.*, **40**, 103 (1939a).
92. —, *Rev. d'Immunol.*, **5**, 113 (1939b).
93. —, *Science*, **90**, 237 (1939c).
94. —, *Arch. Path.*, **27**, 115 (1939d).
95. —, "Dynamics of inflammation," New York, Macmillan Co., 1940a.
96. —, *Proc. Soc. Exper. Biol. Med.*, **44**, 588 (1940b).
97. —, *Am J. Physiol.*, **129**, 691 (1940c), *Proc. Soc. Exper. Biol. Med.*, **51**, 39 (1942a).
98. —, *Arch. Path.*, **30**, 363 (1940d).
99. —, *Am. J. Path.*, **16**, 685 (1940e).
100. —, *Proc. Soc. Exper. Biol. Med.*, **47**, 456 (1941a).
101. —, *Science*, **93**, 456 (1941b).
102. —, *Am. J. Physiol.*, **134**, 517 (1941c).
102a. Menkin, V., *Arch. Path.*, **34**, 182 (1942b).
102b. Menkin, V., *Am. J. Physiol.*, **138**, 396 (1943b).
103. —, *Cancer Research*, **1**, 548 (1941).
104. —, and Kadish, M. A., *Am. J. Physiol.*, **124**, 524 (1938).
104a. Menkin, V., *Arch. Path.*, **36**, 269 (1943).
105. —, *Arch. Path.*, **33**, 193 (1942).
106. —, Kadish, M. A., and Sommers, S. C., *Arch. Path.*, **33**, 188 (1942).
107. —, and Menkin, M. F., *J. Exper. Med.*, **51**, 285 (1930).
108. —, and Warner, C. R., *Am. J. Path.*, **13**, 25 (1937).
108a. Menkin, V., and Kadish, M. A., *Am. J. of Med. Sci.*, **205**, 363 (1943b).
108b. Menkin, V., *New Engl. J. Med.*, **229**, 511 (1943c).
109. Metchnikoff, E., "Leçons sur la Pathologie Comparée de l'Inflammation," Paris, G. Masson, 1892.
110. Minami, G., and Inugami, K., *Trans. Soc. Path. Japonicae*, **30**, 389 (1940).
111. v. Möllendorff, and Munchen, W., *Med. Wchnschr.*, **74**, 135 (1927).
112. Moon, V., *Arch. Path.*, **20**, 561 (1935).
113. —, "Shock and related capillary phenomena," Oxford Univ. Press, New York, 1938.
114. Mudd, S., McCutcheon, M., and Lucké, B., *Physiol. Rev.*, **14**, 210 (1934).
115. Müller, F., (Cited by H. Kossel) *Z. Klin. Med.*, **13**, 149 (1888).
116. Nettleship, A., *Am. J. Clin. Path.*, **8**, 398 (1938).
117. Noetzel, W., *Beitr. Klin. Chir.*, **51**, 740 (1906).
118. Nordmann, M., and Ruether, A., *Virchow's Arch. f. path. Anat.*, **279**, 45 (1930).
119. Okuneff, N., *Pflüger's Arch.*, **204**, 261 (1924).
120. Opie, E. L., *J. Exper. Med.*, **7**, 316 (1905).
121. —, *Arch. Int. Med.*, **5**, 541 (1910).
122. Percival, G. H., and Scott, C. M., *J. Pharmacol. Exper. Therap.*, **41**, 147 (1931).
123. Peters, J. P., *Yale J. Biol. Med.*, **5**, 431 (1933).
124. —, "Body water," Springfield, Ill., Charles C. Thomas, 1935.
125. Ponder, E., and MacLeod, J., *J. Exper. Med.*, **67**, 839 (1938).
126. Pregl, "Quantitative organic microanalysis," Philadelphia, Blakiston Co., 1937 (See G. Barger, p. 244).
127. Ramsdell, S. G., *J. Immunol.*, **15**, 305 (1928).
128. Reifenstein, G. H., Ferguson, J. H., and Weiskotten, H. G., *Am. J. Path.*, **17**, 233 (1941).
129. Richardson, R. J., *J. Clin. Invest.*, **12**, 1143 (1933).
130. —, *J. Clin. Invest.*, **19**, 239 (1940).
131. Ricker, G., and Regendanz, P., *Virchow's Arch.*, **231**, 1 (1921).
132. Rigdon, R. H., *Arch. Surg.*, **41**, 96 (1940).
133. Rocha e Silva, M., and Dragstedt, C. A., *Proc. Soc. Exper. Biol. Med.*, **46**, 303 (1941).
134. Rous, P., and Gilding, H. P., *J. Exper. Med.*, **51**, 27 (1930).
135. —, —, and Smith, F., *J. Exper. Med.*, **51**, 807 (1930).
136. —, and Smith, F., *J. Exper. Med.*, **53**, 219 (1931).
136a. Rugiero, H. R., and Tanturi, C. A., *Semana Med.*, **13**, 8 (1942).
137. Samuel, S., *Virchow's Arch.*, **121**, 396 (1890).
138. Schade, H., *Ergeb. Inn. Med. Kind.*, **32**, 425 (1927).
139. —, Neukirch, P., and Halpert, A., *Z. Ges. Exper. Med.*, **24**, 11 (1921).
140. Sevag, M. G., and Smolens, J., *J. Biol. Chem.*, in press, September, 1941.
141. Siengalewicz, S. S., *J. Pharmacol. Exper. Therap.*, **24**, 289 (1925).
142. Smith, F., and Dick, M., *J. Exper. Med.*, **56**, 371 (1932).
143. Starling, E. H., *J. Physiol.*, **19**, 312 (1896).
144. Steinberg, B., and Dietz, A., *Arch. Path.*, **25**, 777 (1938).
145. Stiles, P. G., and Lusk, G., *Am. J. Physiol.*, **9**, 380 (1903).
146. Swingle, W. W., Pfiffner, J. J., Vars, H. M., Bott, P. A., and Parkins, W. M., *Science*, **77**, 58 (1933).
147. Ungar, G., Parrot, J.-L., and Grossiord, A., *C. R. Soc. de Biol.*, **121**, 1077 (1936).
148. Van Slyke, D. D., *J. Biol. Chem.*, **83**, 425 (1929).
149. Wayne, E. J., *Heart*, **16**, 53 (1931).

150. Weatherford, H. L., *Z. Zellforsch. Mikr. Anat.*, **17**, 518 (1933).
151. Weden, H., *Arch. Exper. Path. Pharmakol.*, **172**, 161 (1933).
152. Wells, H. G., and Johnstone, O. P., *J. Infect. Dis.*, **4**, 582 (1907).
153. Weiss, C., *J. Infect. Dis.*, **41**, 467 (1927).
154. Winternitz, M. C., and Hirshfelder, A. D., *J. Exper. Med.*, **17**, 657 (1913).
155. Wolf, E. P., *J. Exper. Med.*, **34**, 375 (1921).
156. Wong, S. Y., *J. Biol. Chem.*, **55**, 431 (1923).
157. Wybauw, L., *C. R. Soc. de Biol.*, **123**, 524 (1936).
158. Zweifach, B. W., *Cold Spring Harbor Symp. Quant. Biol.*, **8**, 216 (1940).

The Modern Outlook on Blood Coagulation

JOHN H. FERGUSON

Departments of Physiology and Pharmacology, University of North Carolina
(*formerly* Department of Pharmacology, University of Michigan)

The highlights of the major viewpoints that have guided investigators in this difficult field are touched upon in a recent presentation of a new blood-clotting theory.[49] The subject has many ramifications [39, 118] involving fundamental principles of physical [2, 122] and colloid [98] chemistry, in addition to its interest for the protein,[18] lipid,[8, 64, 119] carbohydrate (polysaccharide),[88] and enzyme [96] chemist and its importance for clinical medicine.[41, 100] Briefly, the essential feature of blood coagulation is the conversion of *fibrinogen,* a plasma protein fraction, from the state of a colloidal hydrosol to that of an oriented micellar gel, *fibrin.* This process requires a specific agent termed *thrombin* which has first to be elaborated from plasma protein sources in which it exists as an inactive precursor, *prothrombin.*[102] The puzzling problem is to account for the fact that blood does not yield thrombin and clot in the course of the normal circulation, despite our ability to demonstrate, by isolation *in vitro,* even from the plasma devoid of all formed elements (cf.[26, 82]), all the known factors which lead to the coagulation of shed blood. It is true that we are still only at the threshold of dignifying these isolated factors with true chemical significance, and the important physicochemical aspects are particularly in need of further investigation. Even when the "coagulationist" is forced still to work with relatively crude materials, he can learn much by carefully controlling his tests so as to identify and evaluate the significance of suspected "impurities".[28] Recent techniques have been standardized [36, 100] to give data of definite quantitative significance. Selected references (see bibliography) are given for *assay* of (1) fibrinogen,[68, 91] (2) prothrombin,[7, 42, 99] (3) ionized calcium,[10, 82, 84, 101] (4) tryptase,[45, 47, 48] and (5) heparin.[1, 75] This "progress report" will endeavor to bring out the main points that can be deduced from the modern experimental approach to the blood-coagulation question, considered from the point of view of fundamentals and without reference to the important problems of practical applications, which would unduly extend the scope of this review. Thus, no mention will be made of purely *in vivo* factors such as the *naphthaquinones* (vitamins K[21]) and the prothrombin-lowering dicoumarins.[121]

Fibrinogen is isolated in varying degrees of purity by selective precipitation (repeated for "purification") by (a) dilution and acidification (to the isoelectric point),[102] cf.,[18] (b) "salting-out" with neutral salts,[69] or (c) cold alcohol precipitation.[19] Its preparations are likely to show some degree of (1) "spontaneous" coagulation,[102] (2) denaturation,[18, 60, 118] and (3) lytic destruction, with loss of ability to yield fibrin.[44] These are minimal in the best preparations,[19] some of which keep indefinitely in the dry state [52] and very satisfactorily in sterile saline solution. Solutions that do not yield fibrin "spontaneously" are free from thrombin,[102] and the ability to withstand the addition of the necessary "activators" (see below) of

thrombin-precursor (prothrombin) show this to be absent also.[36, 41] Failure in experimental application of these principles, incidentally, led Mills (*e.g.*[89]) to erect an erroneous concept [28] of the action of tissue extracts in blood-clotting (cf. Smith *et al.*[109]) "Denaturation" can be controlled by keeping blood cool and free from protein precipitants.[62]

We have had preparations [44] apparently free from "lytic factors" (proteolytic enzymes of blood and tissue origin); and the alcohol-precipitated human fibrinogen prepared by Dr. R. M. Ferry at the Harvard Plasma Fractionation Laboratory [10] is a refined product, all but enzyme-free. Apart from specific tests for such impurities as those mentioned as interesting to the "coagulationist," the physical chemist is checking the purification by the best modern methods [122] such as solubility tests,[17] electrophoretic analysis (Tiselius [116]) and ultracentrifugation (Svedberg [111]).

Cataphoresis,[110, 115] electrometric titration,[93] and flocculation maxima in *electrolyte-poor* buffer [44, 83] agree in establishing the isoelectric point of fibrinogen at close to pH = 5.4. There is no pH shift during blood-clotting.[23, 51] Modern applications of physical methods are excellently reviewed by Wöhlisch,[118] the techniques including (1) ultramicroscopy (dark field), (2) optical polarization, including streaming double-refraction and depolarization of diffracted light (Tyndall-beam), (3) x-ray spectroscopy, (4) electron microscopy, (5) viscosimetry, including application of the principles of "humoral rheology," [20] (6) orientation in surface films, (7) polarography, and (8) studies of hydrotropy (especially with urea). The major conclusion from these techniques is that both fibrinogen and fibrin exist in elongated (fibrillar) micelles with definite fluid and colloidal properties. In analyzing the data great caution is needed to avoid many pitfalls occasioned by the crude and impure materials. The effects, for instance, of traces of contaminating proteolytic enzymes repeatedly come up for comment.[118]

Prothrombin is the most difficult of all clotting agents to purify.[7] Some highly potent materials are currently being prepared,[86, 107] but our own experience with similar techniques affords abundant evidence that mere potency is no criterion of purity in the chemical sense.[30, 41, 49] The best thrombin preparations [4, 5, 87, 90, 105, 106] are purer, but few of them [97, 105] succeed in eliminating a proteolytic contaminant which has all the characteristics of serum tryptase.[50, 90] The cruder prothrombins keep poorly in solution [86] and are likely to pass over into thrombin "spontaneously," or under such simple physical treatment as acidification, dialysis, shaking with chloroform, and the like.[9] Long experience with preparations which do not behave thus has convinced us of the soundness of the view that the elaboration of thrombin from its inactive precursor requires certain specific "activators." [33, 35, 53, 54] The persistence of traces of these in the cruder prothrombin accounts for the phenomena cited.

Activators of Prothrombin.[31] The first of these is ionized *calcium,* almost unique in this respect.[30] But we have frequently obtained prothrombins which were not at all (or very poorly) activated by calcium salts alone.[33, 53] The second factor (or factors) needed has received a variety of "working names" such as *zymoplastic substance* (Schmidt), *thrombokinase* (Morawitz) or *thromboplastin* (Howell), but before analyzing the recent advances in the elucidation of the thromboplastic factors, we may add a few comments on the calcium story. It is the active Ca *ion* which is normally essential and it takes a considerable excess of oxalate, citrate, or other Ca-ion depressants, added a sufficient time before completion of the prothrombin conversion, to prevent demonstrable thrombin formation. Nordbö's work [94] clarifies the underlying physical chemistry in terms of the Ca-ion "activity," which is subject to special influences in the presence of the protein colloids. Fully-formed thrombin is active in the presence of considerable excess of oxalate, etc. and some purified thrombins have been prepared completely Ca-free.[88] There is, however, an interesting intermediary stage in prothrombin activation, in which the thrombin (up to a short time after the mixture has reached its maximal coagulant potency) can never-

theless be progressively inactivated by oxalate or citrate.[29, 33] Re-addition of adequate ionized Ca restores the process of thrombin formation. These facts, together with the quantitative dependence of the thrombin yield upon the amount of ionizable calcium present, below a certain "optimum," clearly indicate a Ca-prothrombin intermediary complex in thrombin formation.[33, 35] These experiments [33] were performed in the presence of abundant cephalin,[27] a phospho-lipid.[63, 119]

While the physical chemistry [80] of the *phospholipid-factor* in blood-clotting needs further study and we know of no coagulation studies as between (1) α- and β-forms [8] or (2) * serine- vs. ethanolamine-cephalins,[63] it has been known since the experiments of Howell [71] and his pupils [85] that *cephalin* has a unique rôle and is a constant chemical constituent of crude aqueous (or alcoholic) tissue extracts of "thromboplastin." [6] It is equally clear, however, that the pure phospholipid cannot account for *all* the prothrombin-activating potency of crude thromboplastin.[58] Howell speculated that cephalin became more active as the result of combination with proteins.[72, 73] Some recent studies claim to have identified such alleged compounds [14, 15, 16] and the data are deserving of consideration, but they need to be regarded very critically in the light of the following facts. It is to be expected from the physico-chemical properties of cephalin,[12, 80] and the fact is confirmed by experiments in the author's laboratory,[25] that this phospholipid is *always* in firm combination with the proteins present, and one has but to add its solution to any blood or tissue protein preparation and test the ability to recover the P-lipid by extraction procedures, to be convinced of this finding. Moreover, our prothrombins which were not activated to thrombin by calcium, alone, were shown by chemical analysis to contain significant amounts of cephalin and other P-lipids.[54] This is true, also, of all plasma-, cell-, and tissue-preparations. When comparable amounts of pure cephalin were added to Ca-prothrombin, complete activation to a very stable thrombin took but a few minutes. The clots formed with a good fibrinogen were firm gels and showed no retraction or liquefaction in many days. Hence, in the complete absence of demonstrable enzyme (lytic) factor, there was no evidence that the activation of prothrombin to thrombin required more than the two chemically identified entities, Ca-ions and "free" cephalin.[54] Further studies [85] revealed the quantitative dependence of the amount of thrombin formed upon the proportion of cephalin added, below an "optimum," and as little as one in several millions of the pure P-lipid had significant thromboplastic effects. Potent thrombins have been prepared completely free from phosphorus.[5, 72, 87] Thus, it is concluded that cephalin, like Ca, is a component of an "intermediary" *Ca-cephalin-prothrombin complex* appearing temporarily during thrombin formation.[35]

The fact still remains that crude aqueous tissue extracts are more thromboplastic than their isolated P-lipids, and this was proved by extraction and quantitative comparison in the author's laboratory.[38] Another experimental difference is the ease with which prothrombin activation can be prevented by heparin (see below) when cephalin is used as the thromboplastic agent, as compared with crude tissue extracts.[84, 88] The addition of one other pure chemical agent bridges these differences. This is the enzyme, crystalline (pancreatic) trypsin.[24, 92] An optimal mixture of Ca-salt, cephalin, and crystalline trypsin [53] is a thromboplastic agent that closely imitates the so-called natural "thromboplastin" in essentially all the phenomena of *in vivo* injections and of artificial blood-clotting systems,[38, 54] including correction (*in vitro*) of the coagulation defect in hemophilia.[37] It would appear, therefore, that tryptase-enzymes supply the "missing link" in our data with reference to the thromboplastic system.

In the cited review [49] are mentioned the chief data on the natural "tryptases" [96] of blood and tissues and the experimental work on proteolytic enzymes in relation

* Current experiments (unpublished), in collaboration with Dr. J. Folch-Pi, show that *both* these (and other) P-lipids possess thromboplastic activity (J. H. F.).

to blood clotting.[22] It may be repeated that *trypsin* digests, but does not clot, prothrombin-free fibrinogen. The enzyme *papain* does both, and the fibrin clots which it produces are indistinguishable under the microscope from those produced by enzyme-free thrombin.[57] Claims[11] for a true thrombin-like action of the simple chemical, *ninhydrin* (1, 2, 3-indantrione hydrate), are not substantiated.[57] Two "inhibitors" of trypsin actions, (1) heparin (see below)[66, 70] and (2) the crystalline polypeptide first isolated by Northrop and Kunitz,[92] retard and may prevent blood-clotting.[46]

Lytic Phenomena. Not immediately germane to the processes of blood clotting, but of great significance in linking the commonly encountered lytic phenomena (*e.g. progressive* destruction of prothrombin,[24, 42] thrombin,[65] fibrinogen,[47] and fibrin[45, 95, 112]) with the ordinary protein-digesting actions of "tryptases,"[96] are recent correlations of these phenomena in natural (isolated) and purely chemical systems.[65, 81] A practical result of such studies has been the development of sensitive and selective "assay" methods with which to detect and measure the natural tryptase enzymes. Timing the lysis of fibrin[45, 48] or of fibrinogen,[47] in which "activated" tryptase unknowns are compared with a reference series of standard trypsin dilutions, are two such methods of wide applicability.

There is clear evidence that "active" tryptase is not normally present in the blood, which is indeed capable of *inhibiting* added trypsin.[50, 78] It is necessary to free the enzyme from its naturally inactive state by such procedures as acidification, dialysis, or shaking with chloroform, etc.[78, 95, 104, 112] (cf. "spontaneous" activation of prothrombin[9]). It is not yet certain whether these procedures (especially the last) can be relied upon quantitatively, but they do, at least, have possibilities.[45, 47, 114]

It should be of particular interest to the colloid chemist to attempt an explanation of the exact way in which tryptase "activation" occurs in shed blood and with special reference to the cited procedures. At present we can only theorize that "wetting" (contact with foreign surfaces) and the whole set of conditions encountered *in vitro* may introduce new conditions favorable to activation of the enzyme, perhaps by releasing the system from some "stabilizer" (in the colloid sense).[98] A plausible analogy is the autolysis of damaged tissue cells by their cathepsins (proteolytic enzymes) when they are "damaged" by withdrawal from their natural environment and subjection to various traumatizing conditions. If Schmitz[104] is correct in postulating not only an inhibitor-combination, but also an inactive precursor form ("tryptogen"[49]) for the plasma tryptase, analogies may be drawn between the blood enzyme system and that so clearly worked out by Northrop and his colleagues for the pancreatic trypsins.[92] The blood tryptase, however, is not dependent upon the pancreatic secretion[118] nor is it related to the digestive cycle (author's unpublished experiments). That *tryptogen* activation proceeds "autocatalytically" is suggested by some recent experiments in the author's laboratory, the data indicating a close parallelism between enzyme liberation and actual clotting.[50]

Clot Inhibition. The field of clot inhibitors is too extensive to cover in this review, but the major facts relating to the heparins are included in recent articles on this topic.[56, 79] The Ferguson clotting theory[49] attributes the normal fluidity of the blood to the absence of enzyme catalysts of the thromboplastic system. In part, at least, such absence may be due to the demonstrable tryptase-inhibitors of plasma (and serum). *Heparins,* which are now believed to be mucoitin-polysulfuric esters[79] derived from the metachromatic (toluidin-blue-staining) granules of the Ehrlich mast-cells (basophils),[117] may very well appear in the blood as the result of dissociation of protein combinations of varying potency (Jaques).[76, 77] It is possible that these heparins or their protein combinations are the anti-enzymes[66] and anti-coagulants[67] we are looking for. It should be mentioned that in artificial clotting-systems heparin requires a *co-factor,*[74] which has so far been but partly identified as a component of certain plasma albumin fractions.[18, 120] As to the mode

whereby heparin (or heparin *plus* co-factor) inhibits clotting,[3] mention will merely be made of three experimentally demonstrated reactions: (1) "anti-thromboplastic," which may be due partly to displacement of cephalin [12, 34] and partly to inhibition of tryptase enzyme; [40, 55] (2) "anti-prothrombic," by which we indicate such results as show a lessening of the thrombin yield; [55] and (3) "anti-thrombic," *i.e.*, interfering with the final thrombin-fibrinogen reaction.[67, 108] These actions are imitated, in whole or in part, by numerous other chemical agents [40, 67] and they undoubtedly have a common physicochemical basis [58, 61] which can be summed up in terms of the colloidal properties of the proteins involved, including the enzyme-proteins as well as their substrates.

In summary, the author's view of the coagulation phenomena is as follows: Blood plasma contains all the necessary factors for clotting, but some of them are not normally in "available" form because the whole unit has the characteristics of a "stabilized colloidal system." [59, 98] We can definitely point to inhibition of tryptase enzymes in this natural system and can demonstrate the enzyme by appropriate activation procedures. It is, therefore, believed that, with the removal of colloidal "protectors" during the physicochemical disturbances attending blood shedding, or damaging of the blood vessels *in vivo*, new conditions result in the appearance of active enzyme. This catalyzes the formation of thrombin from prothrombin via an "intermediary" complex containing prothrombin, calcium, and cephalin. The final thrombin is enzyme-like in potency and in the kinetics of its interaction with fibrinogen to form fibrin. It would be interesting to attempt a more modern comparison (cf.[103]) of this reaction with other crystalline gels, forming from sol precursors under the influence of small amounts of initiating agents.

Thrombin itself is not proteolytic in the ordinary sense, but continuing action of the serum protease may result in the variable and incidental accompaniments of blood coagulation, (1) thrombinolysis (progressive "antithrombic" action), (2) fibrinolysis, and its preliminary, "clot-retraction," which may readily be explained as the digesting away (by tryptase) of the attachments of the fibrin threads, thus permitting elastic retraction from the stretched state in which they are deposited. This clotting theory incorporates all the major features of earlier viewpoints and permits due interpretation of the newer facts, while still leaving room for supporting and supplementary data to be expected from the work to come. As this article constitutes a "progress report" rather than a strictly scientific review, the exigencies of space preclude full reference to the extensive experimental literature upon which the facts assembled are substantially based, but the appended bibliography may be consulted for a wealth of reference data in the specifically indicated fields.

References

1. Astrup, T., *Enzymol.*, **5**, 12 (1938).
2. —, *ibid.*, **5**, 119 (1938).
3. —, *ibid.*, **6**, 64 (1939).
4. —, and Darling, S., *J. Biol. Chem.*, **133**, 761 (1940).
5. —, —, *Acta Physiol. Scand.*, **2**, 22 (1941).
6. —, —, *ibid.*, **3**, 168 (1942).
7. Brinkhous, K. M., *Medicine*, **19**, 329 (1940).
8. Burmaster, C. F., *Federn. Proc.*, **2**, No. 1, 59 (1943).
9. Cekada, E. B., *Am. J. Physiol.*, **78**, 512 (1926).
10. Chanutin, A., Ludewig, S., and Masket, A. V., *J. Biol. Chem.*, **143**, 737, *et seq.* (1942).
11. Chargaff, E., and Ziff, M., *J. Biol. Chem.*, **138**, 787 (1941).
12. —, —, and Cohen, S. S., *ibid.*, **136**, 257 (1940).
13. —, —, and Moore, D. H., *ibid.*, **139**, 383 (1941).
14. —, Moore, D. H., and Bendich, A., *ibid.*, **145**, 593 (1942).
15. Cohen, S. S., and Chargaff, E., *ibid.*, **136**, 243 (1940).
16. —, —, *ibid.*, **140**, 689 (1941).
17. Cohn, E. J., *Physiol. Rev.*, **5**, 349 (1925).
18. —, *Chem. Rev.*, **28**, 395 (1941).
19. —, *et al.*, *J. Am. Chem. Soc.*, **62**, 3386; 3396 (1940).
20. Copley, A. L., Krehma, L. C., and Whitney, M. E., *J. Gen. Physiol.*, **26**, 49 (1942).
21. Doisy, E. A., Binkley, S. B., and Thayer, S. A., *Chem. Rev.*, **28**, 477 (1941).
22. Eagle, H., *Medicine*, **16**, 95 (1937).
23. —, and Baumberger, J. P., *J. Gen. Physiol.*, **18**, 809 (1935).
24. —, and Harris, T., *ibid.*, **20**, 543 (1937).
25. Erickson, Nims, B., and Ferguson, J. H., *Proc. Soc. Exp. Biol. Med.*, **45**, 579 (1940).

26. Ferguson, J. H., *Am. J. Physiol.*, **108**, 670 (1934).
27. —, *Proc. Soc. Exp. Biol. Med.*, **31**, 929 (1934).
28. —, Chap. XV. *Blood;* Harrow, B. and C. P. Sherwin's "Textbook of Biochemistry," W. B. Saunders, Phila., 1935.
29. —, *Proc. Soc. Exp. Biol. Med.*, **34**, 797 (1936).
30. —, *Physiol. Rev.*, **16**, 640 (1936).
31. —, *Am. J. Physiol.*, **117**, 587 (1936).
32. —, *J. Lab. Clin. Med.*, **22**, 462 (1937).
33. —, *Am. J. Physiol.*, **119**, 755 (1937).
34. —, *Proc. Soc. Exp. Biol. Med.*, **37**, 23 (1937).
35. —, *Am. J. Physiol.*, **123**, 341 (1938).
36. —, *J. Lab. Clin. Med.*, **24**, 273 (1938).
37. —, *Am. J. Physiol.*, **126**, 669 (1939).
38. —, *Proc. Soc. Exp. Biol. Med.*, **42**, 33 (1939).
39. —, *Ann. Rev. Physiol.*, **2**, 71 (1940).
40. —, *Am. J. Physiol.*, **130**, 759 (1940).
41. —, *J. Lab. Clin. Med.*, **26**, 52 (1940).
42. —, *Proc. Soc. Exp. Biol. Med.*, **46**, 80 (1941).
43. —, *ibid.*, **47**, 538 (1941).
44. —, *J. Gen. Physiol.*, **25**, 607 (1942).
45. —, *Proc. Soc. Exp. Biol. Med.*, **51**, 73 (1942).
46. —, *ibid.*, **51**, 373 (1942).
47. —, *ibid.*, **52**, 243 (1943).
48. —, *Federn. Proc.*, **2**, No. 1, 79 (1943).
49. —, *Science*, **97**, 319 (1943).
50. —, *J. Lab. Clin. Med.*, **28**, 1156 (1943).
51. —, and DuBois, D., *J. Lab. Clin. Med.*, **21**, 663 (1936).
52. —, and Erickson, B. N., *Proc. Soc. Exp. Biol. Med.*, **40**, 425 (1939).
53. —, —, *ibid.*, **40**, 625 (1939).
54. —, —, *Am. J. Physiol.*, **126**, 661 (1939).
55. —, and Glazko, A. J., *ibid.*, **134**, 47 (1941).
56. —, —, *J. Lab. Clin. Med.*, **26**, 1559 (1941).
57. —, and Ralph, P. H., *Am. J. Physiol.*, **138**, 648 (1943).
58. Fischer, A., *Jap. J. Exp. Med.*, **13**, 223 (1935).
59. —, *Enzymol.*, **1**, 81 (1936).
60. —, *Z. physik. Chem.*, **176**, 260 (1936).
61. —, and Schmitz, A., *Z. physiol. chem.*, **210**, 129 (1932).
62. Florkin, M., *J. Biol. Chem.*, **87**, 629 (1930).
63. Folch, J., *ibid.*, **139**, 973 (1941).
64. —, *ibid.*, **146**, 31; 35 (1942).
65. Glazko, A. J., and Ferguson, J. H., *J. Gen. Physiol.*, **24**, 169 (1940).
66. —, —, *Proc. Soc. Exp. Biol. Med.*, **45**, 43 (1940).
67. —, —, *Am. J. Physiol.*, **134**, 54 (1941).
68. Ham, T. H., and Curtis, F., *Medicine*, **17**, 413 (1938).
69. Hammarsten, O., *Nova Acta Reg. Soc. scient., Upsala, Ser. III*, **10**, 1 (1875).
70. Horwitt, M. K., *Science*, **92**, 89 (1940).
71. Howell, W. H., *Am. J. Physiol.*, **31**, 1 (1912).
72. —, *Harvey Lect.*, Ser. XII, 273 (1916-17).
73. —, *Physiol. Rev.*, **15**, 435 (1935).
74. —, and Holt, E., *Am. J. Physiol.*, **47**, 328 (1918).
75. Jaques, L. B., and Charles, A. F., *Quart. J. Pharmacy Pharmacol.*, **14**, 1 (1941).
76. —, and Mustard, R. A., *Biochem. J.*, **34**, 153 (1940).
77. —, Waters, E. T., and Charles, A. F., *J. Biol. Chem.*, **144**, 229 (1942).
78. Jobling, J. W., and Peterson, A. W., *J. Exp. Med.*, **19**, 459 (1914).
79. Jorpes, J. E., *"Heparin, Its Chemistry, Physiology and Application in Medicine,"* Oxford University Press, London, 1939.
80. Jukes, T. H., *J. Biol. Chem.*, **107**, 783 (1934).
81. Kaplan, M. H., Tagnon, H. J., Davidson, C. S., and Taylor, F. H. L., *J. Clin. Investig.*, **21**, 533 (1942).
82. Lebel, H., Schønheyder, F., and Muus, J., *Skand. Arch. Physiol.*, **78**, 179 (1938).
83. Mainzer, F., *Biochem. Z.*, **246**, 164; 182 (1932).
84. McLean, F. C., and Hastings, A. B., *J. Biol. Chem.*, **108**, 285 (1935).
85. McLean, J., *Am. J. Physiol.*, **41**, 250 (1916).
86. Mellanby, J., *Proc. Roy. Soc. London, B*, **107**, 271 (1930).
87. —, *ibid.*, **113**, 93 (1933).
88. Meyer, K., *Cold Spring Harbor Sympos. Quant. Biol.*, **6**, 91 (1938).
89. Mills, C. A., *Chinese J. Physiol.*, **1**, 235 (1927).
90. Milstone, H., *J. Gen. Physiol.*, **25**, 679 (1942).
91. Mylon, E., Winternitz, M. C., and de Sütö-Nagy, G. J., *J. Biol. Chem.*, **143**, 21 (1942).
92. Northrop, J. H., "Crystalline Enzymes, The Chemistry of Pepsin, Trypsin, and Bacteriophage," Columbia University Press, New York, 1939.
93. Nordbö, R., *Biochem. Z.*, **190**, 150 (1927).
94. —, *Skand. Arch. Physiol.*, **75**, Suppl. 11 (1936).
95. Nolf, P., *Medicine*, **17**, 381 (1938).
96. Oppenheimer, C., "Die Fermente and ihre Wirkungen, Bd. 2 (1926) Lief. 7, (1937). G. Thieme, Leipzig.
97. Parfentjev, I. A., *Am. J. Med. Sci.*, **202**, 578 (1941).
98. Pickering, J. W., "The Blood Plasma in Health and Disease," Longmans, London, 1928.
99. Quick, A. J., *J. Am. Med. Assoc.*, **110**, 1658 (1938); correction, **111**, 1775 (1938).
100. —, "The Hemorrhagic Diseases and the Physiology of Hemostasis," Chas. C. Thomas, Springfield, Ill., 1942.
101. Ransmeier, J. C., and McLean, F. C., *Am. J. Physiol.*, **121**, 488 (1938).
102. Schmidt, A., "Zur Blutlehre," Leipzig, 1892.
103. —, "Weitere Beiträge zur Blutlehre," Wiesbaden, 1895.
104. Schmitz, A., *Z. physiol. chem.*, **250**, 37 (1937).
105. Seegers, W. H., *J. Biol. Chem.*, **136**, 103 (1940).
106. —, Brinkhous, K. M., Smith, H. P., and Warner, E. D., *ibid.*, **126**, 91 (1938).
107. —, Smith, H. P., Warner, E. D., and Brinkhous, K. M., *ibid.*, **123**, 751 (1938).
108. —, and Smith, H. P., *Proc. Soc. Exp. Biol. Med.*, **52**, 159 (1943).

109. Smith, H. P., Warner, E. D., and Brinkhous, K. M., *Am. J. Physiol.*, **107**, 63 (1934).
110. Stenhagen, E., *Biochem. J.*, **32**, 714 (1938).
111. Svedberg, T., *Ind. Eng. Chem., Anal. Ed.*, **10**, 113 (1938).
112. Tagnon, H. J., *J. Lab. Clin. Med.*, **27**, 1119 (1942).
113. —, *Arch. internat. physiol.*, **51**, 472 (1942)
114. —, Davidson, C. S., and Taylor, F. H. L., *J. Clin. Invest.*, **21**, 525 (1942).
115. Tiselius, A., *Biochem. J.*, **31**, 313; 1464 (1937).
116. —, *Kolloid-Z.*, **85**, 129 (1938).
117. Wilander, O., *Skand. Arch. Physiol.*, **81**, Suppl. 15, 89 (1938).
118. Wöhlisch, E., *Ergebn. d. Physiol.*, **43**, 174 (1940).
119. Working, E. B., and Andrews, A. C., *Chem. Rev.*, **29**, 245 (1941).
120. Ziff, M., and Chargaff, E., *J. Biol. Chem.*, **136**, 689 (1940).
121. Various authors. 4 articles and editorial on Dicoumarin, in *J. Am. Med. Assoc.*, **120**, 1009, *et seq.* (1942).
122. —, Symposium on physicochemical methods in protein chemistry, *Chem. Rev.*, 30, 321, *et seq.* (1942).

Immunology *

WILLIAM C. BOYD

Evans Memorial and Mass. Memorial Hospitals, School of Medicine, Boston University

Progress in the science of immunology has been so rapid in recent years that a large part of the subject can now be treated purely from the point of view of the chemist. This is mainly due to the preponderant role in immunity of the defensive substances—*antibodies*—produced by an animal in response to the entrance of certain foreign chemical substances, called *antigens*. Writing primarily as a chemist, the present author will therefore limit this article arbitrarily to a consideration of antibodies, antigens, and the nature of the reactions which take place between them.

Antibodies

Since antibodies were originally purely hypothetical substances, they were named and characterized according to their effects. Just as there was a period in science when a group of chemists and physicists decided to dispense with the (then unprovable) atomic theory, and to deal solely with observable phenomena—a point of view which has turned up again in different guise in quantum mechanics—so in the development of immunology there was a period when many workers wondered if it would not be better to speak, not of "antibodies," but of "a modification of the physical state of the serum," so as to give it antibody properties, being careful to avoid making assumptions as to the cause or nature of this modification. We now know that antibodies are actual chemical substances (modified globulins), and we can give up this excessive caution and, on the basis of modern knowledge, begin our discussion of antibodies immediately.

Specificity of Antibodies. An antibody produced in response to one pathogenic organism, A, will generally not protect against, or otherwise react with, organism B, and *vice versa,* unless the two are closely related. This fact, among others, is one of those leading us to speak of the specificity of antibodies.

The specificity of serological reactions, extraordinarily sharp, but still limited, has been well defined by Landsteiner [78] as "the disproportional action of a number of similar agents on a variety of related substrata." This definition seems the best that has been proposed, although it would also include many chemical reactions, as might be expected if serological specificity were fundamentally chemical in its nature.

The specificity of antibodies has been explored in great detail by numerous workers in the present century, notably by Landsteiner.[73] The work has been done in part by the study of reactions with related natural proteins, but largely by the

* Thanks are due to the Interscience Publishers for permission to make use of material which formed part of the author's "Fundamentals of Immunology," New York, 1943.

use of "conjugated antigens." By the term "conjugated antigen" we mean an antigen which has had a new specificity grafted onto it by chemical treatment. It has thus been possible to create new categories of specificity, so that a chemically conjugated protein will engender antibodies which will react with other unrelated proteins similarly treated. Methods of doing this are discussed in the various papers devoted to the subject.

The great significance of Landsteiner's work on artificial antigens was to show that antibodies could be directed toward relatively small, well-defined chemical groups on an antigen, instead of toward the whole molecule, and *could react with these groupings separately.* From the chemist's point of view, this is probably the greatest single step forward yet taken in the study of immunology, since it advanced the science at once from the stage where an unknown, special constitution, peculiar to proteins, was thought to be necessary to incite the production of antibodies, to the stage where the chemical composition of the groups determining antibody production and antibody reactivity could be studied.

Study of synthetic, artificial antigens has shown that antibodies produced by them may be of several sorts. In the first place, especially if the chemical treatment has not been too drastic, the serum of an animal immunized with such a treated protein may contain antibodies directed towards the protein substrate, having no relation to the new specificity. It may contain antibodies directed chiefly towards the introduced groups, or towards these groups and the parts of the protein molecule with which they combine. Also, antibodies may occasionally be found which are directed simultaneously towards the original protein and the introduced group.[53, 51, 86]

If we study only the antibodies directed towards the introduced chemical group (hapten), we may still find that although in some cases the specificity seems complete, *i.e.,* the immune serum reacts only with the homologous antigen, more frequently cross reactions are found, even though the reactions with the homologous antigen are most intense. In other words, the immunological mechanism is not able to distinguish absolutely between very closely related compounds. We may suppose that the antibodies, although they reflect in some way the characteristic electronic pattern of the antigen or hapten, do not reflect every detail perfectly. There will be some blurring of outlines. Or we might change the metaphor, and say that antibody is a lock fitted by the hapten; the lock, however, is of somewhat simpler construction than the key, so that keys for other locks may also open it, though perhaps not so readily.

Experiments further indicate that a large number of serologically different compounds could be made up from amino acids, in conformity with the large numbers of specificities found in natural proteins. They indicate that the antibody-forming mechanism can recognize as a unit a fairly large chemical compound. Just how large a group can still be responded to as a whole is not known; the maximum must be smaller than the surface of the smallest protein molecules, but probably larger than any synthetic compound yet studied.

Nevertheless, the immunological mechanism can on occasion single out parts even of chemical compounds of moderate size, so that the specificity in some cases is chiefly or solely directed towards one part of the molecule.

The differences observed in antibodies to aromatic compounds containing the same substituents, but in different positions, indicate that the spatial arrangement of groups, as well as their chemical nature, is of importance in immunology. This conforms to our earlier knowledge of other aspects of biological chemistry, where spatial arrangement is known to be important. For example, enzymes, as is well known, have frequently been found to be specific for one of the stereo-isomers of a substance, different isomers of optically active dyes have been found to have different staining effects, optical isomers of some of the amino acids have different tastes, and optical isomers of drugs usually exhibit different potencies.

Chemical nature of antibodies. Antibodies are serum proteins, with physical properties indicating that they belong in the "globulin" class. An antibody clearly must differ in some respect from the other globulins in the serum, since it bears the imprint of its characteristic specificity, not possessed by the others; this must reside in some chemical difference in at least part of the molecule. Nevertheless, no marked chemical or physical differences have been observed between antibodies and so-called "normal" globulins.

It has, however, been found that the antipneumococcous antibody in horse serum has an isoelectric point of about 4.8, while that of the normal horse globulin is 5.7.[59] Green, McKhann, Kapnick and Fahey [48] found that the water insoluble globulin from horse antipneumococcous sera has a viscosity different from any of the normal proteins, and its titration curve differed slightly from that of normal globulin. See also Goodner and Horsfall.[47] The antipneumococcal antibodies in horse serum have a much higher molecular weight than the majority of the horse serum globulins, but a small amount of a globulin with this molecular weight is also found in normal horse serum.

In contrast to the horse, the rabbit produces antibodies to the pneumococcus which have the same molecular weight as normal rabbit serum globulins.

There was probably at first a tendency on the part of certain immunologists to think and speak of antibodies as if their nature would be a function solely of the antigen towards which they were directed. This may have been easier because it was not yet fully established that the antibodies were themselves proteins. Since the serum proteins of different species exhibit chemical differences, it is to be expected *a priori* that their antibodies will also have different properties. This has been found to be the case. Aside from the difference just mentioned between the horse and rabbit antipneumococcus antibody in regard to molecular weight, it has been found [70] that all antipneumococcus antibodies so far examined fall into two groups as regards molecular weight. In the cow, horse, and pig a heavy molecule of molecular weight 990,000 is formed; in man, rabbit, and monkey, the molecular size is that of the normal serum globulin, about 160,000. On the other hand, horse anti-toxin (diphtheria) has only the normal molecular weight of serum globulin.[92]

Goodner and Horsfall [46] observed a distribution of antibodies into two groups with respect to complement fixing ability and liquid composition. They found that complement (probably from the guinea pig) was fixed by antibodies from rabbit, rat, guinea pig, and sheep, whereas complement was not fixed by antibodies from horse, man, dog, mouse, cat, and goat. Dingle, Fothergill and Chandler [27] also observed similar differences. It will be observed that the above list does not quite correspond to the classification based on molecular weight. Boyd [14] found indications that antibodies fall into two classes in regard to flocculative behavior.

The solubility behavior of antibodies is generally the same as that of the class of serum proteins to which they belong, but here too species differences are found. Rabbit antibodies to proteins seem generally to appear first in the fraction of the globulins precipitated in 13.5% sodium sulphate solution ("euglobulin"); then as immunization progresses, they are found also in the fraction precipitated between 13.5 and 17.4% ("pseudoglobulin I"); and only in powerful sera, as a rule, are they found in the fraction precipitated between 17.4 and 21.5% sodium sulphate concentration ("pseudoglobulin II").[15] During immunization these globulin fractions in the rabbit tend to increase roughly in this order. In the horse, antibodies to toxins and other proteins tend to fall chiefly in the pseudoglobulin fraction, and the results of Reymann [100] suggest that in this animal it is this fraction which increases most on immunization.

Antitoxin in horse serum is usually found to some extent in all three globulin fractions, and variations in its distribution among the fractions are found, even in the sera of the same animal examined by the same method at different times.[42, 79]

The findings depend partly on the method of protein precipitation used. In the goat it appears that diphtheria antitoxin has chiefly the properties of a euglobulin.[6] Felton [81, 82, 83] found the antipneumococcal antibodies in horse serum to be of euglobulin character, and on this fact based his widely used method of concentrating these antibodies.

It is a noteworthy observation that immune antibodies, at any rate, are never found in the albumin fraction of the serum in any species.

In their susceptibility to the action of various denaturing and destructive agents, antibodies closely resemble proteins in general. A good summary of their reactions will be found in Marrack's book.[85] They are destroyed by heat very much as proteins are denatured. In many instances the course is that of a monomolecular reaction, with a very high heat of activation, as in denaturation; probably the process is essentially just that. Antibodies are destroyed by alcohol, but in the cold, or if the alcohol concentration is brought rapidly to over 90%, this is much less marked. Treatment with protein-altering reagents, such as diazonium compounds, iodine, formaldehyde, and ketene, usually reduces or destroys the activity of antibodies.

Antibodies are destroyed by pepsin, less rapidly by trypsin. Note, however, that Parfentjev [98] found that treatment of diphtheria antitoxic serum with pepsin until 70-80% of the protein was rendered non-coagulable by heat, resulted in considerable purification, giving a product a high proportion of which was specifically antitoxic.[cf. 110]

The species differences between rabbit and horse antibody have recently acquired a practical aspect as the result of studies by Goodner and Horsfall,[47] and rabbit antipneumococcus serum has to some extent replaced horse serum for clinical use.

Neurath [89] has calculated that in a molecule of rabbit antipneumococcus antibody, the longer and shorter axes are in the ratio of 9.2, the shorter axis being 37 Å in length, and the longer axis 338 Å. For rabbit antibody the axis ratio is 7.5, and the lengths of the axes 37 and 274 Å. For horse antipneumococcus antibody these figures become 20.1, 47 Å and 950 Å.

The majority of antibodies which have been studied in the Tiselius apparatus have proved to be α globulins. Certain horse antisera, however, may show a component which migrates as a slow β (β_2) globulin. In some horse anti-diphtheria sera two sorts of antitoxin may be found, a β_2, and a γ. They show definite differences in serological properties.

The electrophoretic mobilities of various purified antibodies and normal serum proteins have been determined by Tiselius and Kabat [106] and others. The results are shown in Table 1. It will be seen that the mobilities of the various normal components in the sera from the several animal species are approximately the same. In

Table 1. Electrophoretic Mobilities of the Components of Sera from Various Species (Modified from Tiselius and Kabat.[106])

Species	Antibody for	ab $N/\Sigma N$ (%)	$\mu \times 10^5$ cm^{2}sec^{-1}volt^{-1} Al	α	β	Ab	γ
horse [1]	pneumococcus type 1	20.7	−5.5	−3.7	−3.0	−2.1	−0.9
horse [1]	pneumococcus type 1	21.8	−5.7	−4.0	−3.2	−2.2	−1.3
horse [1]	——— [4]	0	−5.7	−3.8	−3.1	..	−1.0
pig [1]	pneumococcus type 1	1.6	−5.7	−3.5	−2.7	..	−1.1
rabbit [1]	ovalbumin	36.4	−6.0	−3.6	−2.9	..	−1.1
monkey	pneumococcus type 3	6.6	−5.2	−4.3	−3.0	..	−0.7
horse [2]	diphtheria toxin	18.9		−4.9	..	−2.6	..
horse [3]	diphtheria toxin	...	..	−4.0	..	..	..

[1] Determined at pH 7.72 ± 0.02 in buffer containing 0.15M NaCl and 0.02M PO_4^{----}. Temperature +0.5°.
[2] Determined at pH 7.35, ionic strength 0.1.[92]
[3] At pH 7.3.[90]
[4] Normal horse serum.
α, β, γ = globulin fractions. Al = albumin. Ab = (special) antibody component.

the pig, rabbit, and monkey (in which it has been found that the antibodies have the same molecular weight as the γ globulin fraction) no new electrophoretic component was found, but in the horse a new component, between the β and γ components, appeared. This new component seems to be all antibody, since it is absent from the same serum after the antibody has been removed by "absorption" with antigen. This serum showed in the ultracentrifuge a heavy component which was the bearer of the antibody activity. Pig serum, which also possessed such a heavy component, did not show the component with the new electrophoretic mobility at pH 7.72, but did at another pH.

There have been a number of analyses for nitrogen, amid nitrogen, mono-amino, and di-amino nitrogen and amino acids in antibodies, and also in specific precipitates containing a non-nitrogenous antigen. On the whole, the differences between samples, and between antibody and normal globulin, do not seem to be greater than could be accounted for by the use of different methods by different workers.

Valence of antibodies. It is common to speak of a polyvalent serum, meaning one which contains antibodies for more than one antigen, *e.g.*, against several types of pneumococcus. The antibodies in such sera are separate and distinct. There is another question of valence which is the one we wish to discuss here, namely, the number of specific combining groups on a single molecule of antibody.

It is hardly likely that the specific reactivity of an antibody molecule could be distributed uniformly over the whole surface. Also, either of the theories of antibody formation now considered tenable [1, 96] could lead us to predict that a molecule of antibody could not have many more than two localized reactive groups. What we know of specificity shows that orientation in space is important, and the combining group of an antibody must therefore possess spatial differentiation.

It has been found [90, 86] that when slides are coated with a film of antibody *ca.* 45 Å thick, antigen can be specifically deposited, on top of this another film of antibody, and finally another film of antigen. If the antibody molecules are all oriented alike in such films, this would seem to demonstrate that the antibody is at least divalent. However a colleague has pointed out that it is perhaps not excluded that each film might consist of molecules arranged alternately up and down, in a head-tails-head-tails . . . arrangement, so that even antibody possessing only one reactive site per molecule might still give the "sandwich effect."

It has been observed in a number of cases [55, 91, 60] that there may be present in a serum, antibody which has the power of combining with antigen, but does not form a precipitate. Such antibody has been variously referred to as low grade, incomplete, imperfect. Heidelberger [52] calls it "univalent," but states elsewhere [60] that the word "univalent" is not used in the literal sense. The special properties of this antibody may perhaps be due to imperfect specificity, or physical properties different from those of "good" or "polyvalent" antibody, or both.

There is one theory of antigen-antibody reactions which necessarily supposes the antibodies to be polyvalent or at least divalent, but since we cannot regard this theory as completely established, the polyvalency of antibodies cannot be considered proved. The proper way to settle the point seems to be a study of the molecular composition of antigen-antibody compounds made with excess antigen. Pauling [96] has proposed the determination of the valence of antibodies by analysis of precipitates of antibody and a synthetic conjugated antigen. From certain data of Haurowitz, Kraus and Marx,[51] which led these authors to conclude that the antibody was monovalent, Pauling concludes that it was probably divalent. Such studies apparently have not so far revealed any compound containing more than two molecules of antigen in combination with one of antibody, so that the maximal valence so far found for antibody is two. Antigen, on the other hand, is almost always multivalent, and compounds have been observed in which over 100 molecules of antibody were combined with one of antigen.[66]

The only experimental work directed precisely at the investigation of the valence of antibodies is that of Haurowitz and collaborators.[51a, 51b] These results suggest that antibodies do not have a valence greater than one.

Antigens

Conditions of antigenicity. A very large number of substances have been investigated for their antigenic power, and it is natural to consider them as a group and try to see what characteristics they have in common. Unfortunately the knowledge which has thus been acquired is very meager; we are not yet able to say, from the chemical properties of a substance, that it will *certainly* be antigenic or non-antigenic. However, certain general principles do emerge.

(1). Antigenic substances are always large molecules, of molecular weight of the order of 10,000 or higher (excluding simple compounds which have been found to *sensitize*) and consequently are usually colloidal in solution. Correlated with the large molecular size is the inability of antigenic substances to go through collodion or cellophane membranes which pass smaller molecules such as sugar. There is some indication that in general the larger the molecules the better the antigenic power.

Another indication of the importance of particle size in antigenicity is the fact that non-antigenic or weakly antigenic substances can sometimes be made antigenic by adsorbing them on particulate matter such as collodion particles, kaolin, or charcoal.[45, 75, 98, 108]

(2). The majority of antigenic substances are proteins, and it was formerly believed that only proteins could act as antigens. It has since been found that some "lipoids," when mixed with antigenic protein and injected, produce antibodies. Further than this, it has been found that some carbohydrates (usually of bacterial origin) and lipoid-carbohydrate compounds, even when carefully purified to remove protein, will produce antibodies when injected.[34, 5, 38, 35, 90] In some cases these latter classes of substances are inferior to proteins as antibody producers.

Aside from these three classes of substances, it has not been certainly established that any other class of compound can be antigenic.

Not all proteins are equally good antigens. Here, and in much that follows, we mean "antigens for rabbits," for this is the animal most often used; some species seem definitely inferior to the rabbit as precipitin producers. Even within this species, individual differences may be found in antibody producing power for a given antigen. The best antigens are probably complex mixtures, such as whole blood serum, or bacteria or blood cells. Then probably come certain proteins, such as hemocyanins, ovalbumins. Hemoglobin is a poor antigen, and has in fact several times been erroneously reported to be non-antigenic. Finally gelatin (a derived protein) is non-antigenic, except as a conjugate. The simplest proteins, the protamines, are also non-antigenic.[111]

(3). As a rule the proteins of an animal are not antigenic for itself, or for another individual of the same species. In immunology this principle has received the name of *horror autotoxicus,* which is stated thus: an animal will never produce antibodies to any substance normally found in its own circulation.

Some workers apparently believe that antibodies to an individual's own antigens may be produced, but never observed, since they would be continually removed by combination with the big excess of the antigen constantly present. Others feel that this violates some sort of "principle of laziness" in nature and that such antibodies are probably not produced at all.

The qualification which restricts the principle to substances in the circulation is needed, for it has been found that protein from the lens of the eye [cf. 111] will produce antibodies in the same species,[cf. 72, 84] and it has been claimed [cf. 81] that a guinea pig can be sensitized with the lens of one eye and later shocked anaphylactically with the

lens of the other. Lewis [81] found that lactating goats could be caused to form antibodies to their own casein; these antibodies reacted to an equal degree with cow casein. Hektoen (cited by Lewis [81]) reported obtaining other *iso-antibodies* in rabbits, against thyroglobulin (the iodine-containing protein of the thyroid gland) and fibrinogen from the blood. Stokinger and Heidelberger [103] however, found definite differences in their antisera to thyroglobulin from man, hog, cow, and sheep. Thus it seems established that certain proteins of the body, which never normally come into the circulation, can act as iso-antigens.

In general we may venture to surmise: (1) the antigenic power of the proteins is greater the more removed is their source from the experimental animal in the zoological scale; (2) very simple proteins are not antigenic. These guesses may not be wholly correct, of course. Plant proteins seem in general good antigens for animals, although perhaps not as much better than animal proteins as their greater taxonomic remoteness would suggest.

Gelatin contains no tyrosine or tryptophane, and but little phenylalanine. The fact that it is not antigenic led Obermeyer and Pick [cf. 73] and Wells [111] to infer that the aromatic radicals of the protein molecule are of importance in determining antigenic activity. In support of this idea Wells mentioned several facts, among them the observation of Obermeyer and Pick that chemical substitution in the tyrosine nucleus markedly alters the immunological specificity of proteins. It is difficult to test this idea, but there are some observations rather contrary to it.

In the first place, it has since been found that substitution in other parts of the protein molecule also alters the specificity. Aromatic groups can be introduced into gelatin without making it antigenic.[67] Landsteiner points out that not all proteins possessing aromatic groups are antigenic; examples of this are proteins which have been altered sufficiently by acid or alkali. In other experiments [62, 22] introduced aromatic groups did make gelatin feebly antigenic. Hooker and Boyd concluded that their experiment did not show that the non-antigenicity of gelatin was due solely to its deficiency in aromatic amino acids, since the possibility that non-aromatic prosthetic groups, especially if they contained polar groups, might also make gelatin antigenic had not been tested.

Some carbohydrates, not possessing aromatic nuclei, are nevertheless somewhat antigenic. The power to immunize seems to depend on general properties of the molecule; or perhaps, though this is much less likely, on some specific characteristic as yet unknown to us which antigens have in common; the possession of aromatic groups does not appear to be one of these characteristics.

Finally, Landsteiner's observations that simple substances not causing antibody formation, and not necessarily containing aromatic groups, can still react specifically with antibodies, were rather against the idea, which was partly based on the notion that antigenic capacity and ability to react with antibodies were inseparable.

When proteins are racemized by treatment with alkali, they are no longer antigenic. Since it was found by Dakin and Dudley [24] that such proteins were no longer susceptible to hydrolysis by enzymes, it was suggested that these facts were causally related.[105] This suggestion recalls the "digestion" hypothesis of antibody formation. However other workers have more recently found [82] that these racemized proteins are not completely resistant to enzyme action. It has also been found that the antigenicity is lost more rapidly than the optical activity.[114] Landsteiner [73] has pointed out that the alkali might destroy structures significant for the antigenic function at the same time that it produced racemization. Another explanation for the lower antigenicity of alkali-treated proteins may be smaller molecular size. Boyd [12] found that on treatment of crystalline ovalbumin with alkali there was at first a marked rise in optical activity to several times the initial value, followed by a slower fall to a minimum, indicating possibly that the molecule was first split, then racemized.

That racemization does not in itself mean non-antigenicity is indicated by the

observation of Landsteiner and Barron [74] that nitration of racemized proteins restored some antigenicity, as, to a lesser degree, did iodination.[69]

There seems to be no proof, therefore, that optical activity is essential to antigenicity, although it may play an important role in specificity.

The specificity of antigens. The specificity of antigens resides in the structural peculiarities of their molecules. While it has not been possible to correlate immunological differences in substances, particularly proteins, with their chemical structure in every case, there is no doubt that this failure is due simply to our ignorance of the detailed structure of these compounds.

The dependence of specificity on chemical structure is proved by several lines of evidence. (a) Purified proteins that exhibit chemical differences can nearly always be differentiated serologically.[111, 26, 49, 25, 63, 70, 57] (b) Carbohydrates related structurally give serological cross reactions.[4, 44, 43] (c) Simple chemical substances (haptens) give cross reactions when chemically similar.[73] (d) Chemical alteration of antigens generally alters their specificity. (e) Corresponding proteins of different species which are functionally, and thus probably structurally, related, generally cross react.

Antibody-Antigen Reactions

General characteristics. The majority of those actively engaged in studying immunological reactions are convinced that the reaction between antibody and antigen is a chemical reaction. Certain difficulties have arisen in attempting to interpret all the observations in a consistent way, probably because of the size and complexity of the molecules involved. The reaction between antibody and antigen, and the reaction between proteins and proteolytic enzymes, are the only instances which have been well studied of reactions occurring between molecules of very large size; it need not be surprising if some new features emerge. We omit discussion of the reaction with haptens, which are usually much smaller molecules, and reactions with intact cells such as bacteria or erythrocytes.

Several salient characteristics of the reaction may be mentioned:

(1). The reaction is specific. This point has already been discussed, and it will not be necessary to give examples here. We should keep in mind the specificity of the reaction when we speculate on the mechanism of union, that is, the nature of the forces holding antibody and antigen together.

(2). The *entire* molecule reacts, not some fragment split off from it. This is clearly shown by the ultracentrifugal experiments such as those of Heidelberger and Pederson,[58] and of Pappenheimer, Lundgren and Williams.[92]

(3). The antibody molecule, and molecules of protein antigens, at any rate, behave in these reactions as fairly rigid *ellipsoids* of greater or lesser eccentricity. This is obvious if we accept the views of Neurath [89] on the shape of protein molecules. Boyd and Hooker [16, 17] found that the composition of antibody precipitates is compatible with the idea that antigen molecules are roughly spherical and antibody molecules behave as elongated units, such as ellipsoids, or chains of connected spheres. It is also found in antibody-antigen reactions between films of the reagents [99] that the thickness (30 — 50 Å) of the films is such as to suggest that the molecules are behaving like ellipsoids. Rothen and Landsteiner [101] found that thinner (9 Å) films of antibody would no longer react specifically, suggesting that antibody no longer exhibits its specific behavior if it loses its globular shape.

Evidence that protein molecules act as fairly rigid spheres or ellipsoids comes from their behavior in intense gravitational fields [104] and their possession of a definite electric moment.[115]

Certain earlier workers apparently supposed that the antibody spread as a film over the surface of the antigen. Such a film would resemble a condensed protein

film, and having the polar groups turned away from the water, would account for the hydrophobic character of the compound.

Modern evidence however indicates that antibody does not behave as such a film. Also, the amount of antibody observed to combine with a molecule of antigen does not correspond to a film approximately 10 Å, as this would lead us to suppose, but to a thickness of about 30 Å. The electrokinetic properties of the antibody-antigen complex are still those of protein, and do not suggest that all the polar groups are turned away from the surface.

(4). *No splitting, digestion, or other profound chemical alteration* takes place in the reacting molecules after they have combined.

There are numerous observations showing that no profound alteration takes place in the antigen and antibody molecules after they combine. It is a well known fact that toxin is not permanently detoxified, but simply neutralized, by combination with antitoxin, and illness has resulted from the injection of originally neutral toxin-antitoxin mixtures in which the anti-toxin has been damaged or destroyed, as by freezing. Similarly, various workers have been able to recover unaltered antibody from antibody-antigen precipitates or "agglutinates."

It is evident that if any alteration takes place, it is reversible.[21] A number of workers have thought that the antibody is denatured after combination with antigen. Probably without fully sharing this idea, Eagle [28] and Mudd [cf 87] have emphasized the similarity in properties of combined antibody and denatured globulin. This similarity certainly exists, and is probably significant for theories of serological reactions, but it need not mean that the antibody is actually denatured. Eagle suggests that it is due to the (assumed) turning of the polar groups of the antibody towards the polar groups of the antigen, leaving a surface composed predominantly of non-polar groups, with a consequent reduction of solubility in polar solvents.

Completely denatured proteins are found to have reverted to the fibrous form, which may be detected by x-ray analysis; therefore Marrack [85] points out that if the antibody in specific precipitates were completely denatured, this should be detectable by x-ray study, but actually experiments have not produced any evidence that the antibody in such precipitates has changed from the globular form. We may therefore conclude that the antibody in such precipitates is not wholly denatured.

Not only is the antibody not immediately denatured, but it has been found [18] that the solubility of ovalbumin-antiovalbumin precipitates in excess of antigen was not altered by storage. Since such resolution is a specific reaction, the result indicated that no progressive denaturation of either of the reagents had taken place during the time of storage.

There is other evidence that the serological specificity of the antigen and antibody is not altered by their combination. Precipitates will specifically sensitize animals to the antigen, and to the serum proteins of the animal from which the antibody was derived; non-precipitating (inhibition zone) compounds of antibody and antigen can be precipitated, after the lapse of an indefinite time, by the addition of more of the deficient reagent, or by the addition of a different antibody directed against either the antigen or the first antibody. This again indicates that no significant denaturation takes place.

(5). The combination of antibody and antigen takes place at the *surface* of the molecules. This follows from the third point discussed above. Also Marrack [85] has pointed out that no molecule larger than the lower fatty acids can pass between the constituent atoms of a protein molecule, in view of their known distances from each other. Evidence for the role of surface in serological reactions has already been mentioned; it has been stressed by Hooker.[61] The experiments of Marrack and of Rothen and Landsteiner [101] suggest that when the antibody molecule is unfolded, thus producing a new "surface," it loses its power of specific combination, and the experi-

ments of Porter and Pappenheimer [99] indicate that the reactive capacity of the antitoxin molecule is restricted to a portion of its surface.

(6). The union between antibody and antigen is a *firm* one, but is at least partially reversible. It is well known that it is quite difficult to dissociate diphtheria antitoxin and toxin after their combination.[cf. 92] It is possible to dissolve the precipitate formed by the conjugate protein arsanilic acid-azo-casein and its antibody by the addition of dilute alkali; addition of dilute acid to this solution, which would not precipitate the antibody alone, throws down the antibody and antigen together, with hardly a trace of antibody left in the supernatant fluid.[12] Studies on the heat of serological reactions [18] indicate a considerable energy change, implying a rather firm combination.

Nevertheless, the reversibility of serological reactions is shown by several lines of evidence, such as dissociation of antibody from precipitates and agglutinated cells, solution of precipitates in excess of antigen, etc.

(7). *Both antibody and antigen* enter into the specific precipitate which is formed. There is now abundant evidence for this statement, but before the introduction of quantitative methods into immunology the question was sometimes debated.

(8). Antibody and antigen can combine in *varying proportions*. This is an obvious consequence of the multivalency of one or the other of its reagents, and has been shown by numerous analyses. In the light of what we know now it may seem surprising that it was ever doubted, yet it is precisely the failure to recognize the possibility of combination in different proportions which caused the early writers on antigen-antibody reactions so much difficulty. However, once it was realized that antigen or antibody, or possibly both, was multivalent, the observation that antibody-antigen compounds may vary in composition becomes quite understandable.

(9). Serological reactions display certain analogies with *colloidal* reactions. These similarities are so striking that they will be discussed later at greater length.

The two stages of serological reactions. It is customary to consider serological reactions as occurring in two stages. Such a subdivision should not be taken as necessarily implying that the two stages do not overlap, for it is quite likely that the second stage sets in before the first is entirely completed.

These two stages are: (A) the specific combination between specific groups of the antibody and corresponding groups of the antigen or hapten; (B) the secondary, visible reactions which may follow this, such as precipitation, agglutination, complement fixation, etc.

There are several features of these reactions which make it convenient for us to retain the concept of division into stages.

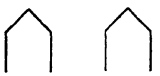
(1). The first stage proceeds without visible alteration, and can be detected only indirectly, while the second may be quite conspicuous and easily detected; (2) in the case of haptens no second stage follows at all; (3) under some conditions, as in the absence of salts, the first stage can take place but the second cannot; (4) their speeds are very unequal, the first stage being extremely rapid, and the second sometimes very slow; (5) the energy change seems to take place during the first stage, the second being accompanied by little energy change; (6) there is some evidence (see below) that the specificity of the second stage may be of a lower order than that of the first stage.

First Stage: Combination

(1). **Mode of combination between antibody and antigen.** Because of the chemical nature of serological specificity, it is logical to suppose that the union between antibody and antigen is a chemical one, due to combination of the specific reactive groups of the two reagents. This is supported by such facts as the power of simple chemical substances (haptens) to react with antibodies, and by the analogy with enzyme reactions, known to be chemical.

The principal question in this connection would appear to be: is the union solely due to Coulomb forces (attraction between positive and negative charges, *e.g.*, $-NH_3^+$ and $-COO^-$) or are other forces involved? This is a difficult problem and has not been satisfactorily solved. This fact should not be too surprising when it is realized that until rather recently chemists were quite unable to give any rational account of how coordinate compounds (such as compounds containing water of crystallization) were formed.

If the forces acting are Coulomb forces, we should suppose that the combination between antibody and antigen is chiefly or entirely due to attraction between amino and carboxyl groups. There are points in support of this idea, such as the rapidity of the primary reaction, and the strong influence of polar groups observed in studies of specificity. In particular the immunological equivalence of such groups as Cl and CH_3 and the acidic character of serologically reactive non-protein



antigens, such as the pneumococcal polysaccharides, are suggestive. Goebel and Hotchkiss [44] found that anti-pneumococcus horse sera of types 1, 3, and 8 gave vigorous precipitation with artificial antigens containing benzenecarboxylic and sulfonic acid radicals, which are quite unrelated in chemical constitution to the pneumomoccus polysaccharides. Chow and Goebel [20] suggested that combination between the pneumococcus polysaccharides and the corresponding antibodies might be largely due to attraction between the ionized uronic acid groups, containing $-COO^-$, and the ionized amino groups, $-NH_3^+$, of the protein. In support of this idea, it was found that treatment of the antibody with formalin, which reacts with the amino groups, destroys the precipitating power of the antibody, and acetylation, also affecting the amino groups, is almost equally effective.

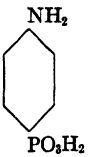
Landsteiner [73] lists in favor of the idea that such a mechanism, which he refers to as the formation of salt-like compounds, may operate, the following: the prominent influence of acid groups (see above), the change in specificity following esterification of aromatic acids or proteins, and the similarity of hemagglutination and hemolysis produced by serological agents and by colloidal inorganic acids.

It has been objected that such a simple mechanism could not account for the specificity of serological reactions, but this is perhaps not a necessary difficulty. If a number of amino and carboxyl groups in the antibody are concerned when one specific determinant group in the antigen reacts, the specificity might be due to the spatial arrangement of these groups. The weaker union observed with the related heterologous antigens could be explained by a failure of the oppositely charged groups to correspond perfectly in position.

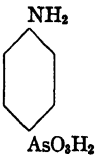
It is however possible that other forces, such as those classified as secondary valence, might play a role in serological reactions. It does not seem likely that the covalent bond is concerned, for as Landsteiner says, ". . . in velocity and easy reversibility antigen-antibody reactions differ from those due to primary valences and resemble the formation of ionic and molecular compounds; and a strong argument against the assumption of covalent bonds is the fact that quite different substances, regardless of their chemical nature, are capable of reacting with antibodies all in like manner." It is likely that similar objections apply against the co-ordinate link, or semi-polar double bond.[cf. 95] Other types of bonds which may possibly play some role in serological reactions are dipole-dipole bonds, dipole-ion bonds,[cf. 50] and the hydrogen bond. The latter consists essentially of a hydrogen atom which is attracted simultaneously to two different atoms, as in salicylaldehyde. The hydrogen bond, which is thoroughly discussed by Pauling [95] appears to be one of the forces keeping proteins in their characteristic configurations. Pauling believes that the hydrogen bond will be found to have greater importance for physiology than

any other single structural feature. It may well figure importantly in serological reactions for, as Pauling says, ". . . the hydrogen bond is the only strong and directed intermolecular interaction which can come into operation quickly."

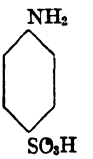
Other writers have mentioned cases of specificity which compare with serological reactions. Thus Landsteiner [78] calls attention to the very interesting examples of the reactions of Cu with the $HO-\overset{|}{C}-\overset{|}{C}=NOH$ and Tl with $-CO-CH_2-CO-$, and the observations of Bergmann on the precipitation of proline with some peptides by rhodanilic acid, and the distinctly specific precipitation of glycine, the simplest amino acid, with potassium trioxalatochromiate [$Cr(C_2O_4)_3K_3 \cdot 3H_2O$]. Erlenmeyer and Berger [30] have stressed the analogy with the forces in crystals, and have drawn attention to the immunological similarity of CH_2, NH, and O, of CH_3 and halogen, (groups which are mutually replacable in crystals of certain substances), and of the correspondence of antigens made from



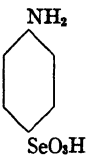
and



and from

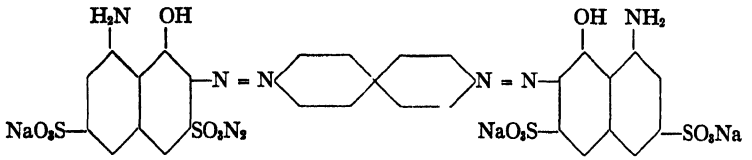


and

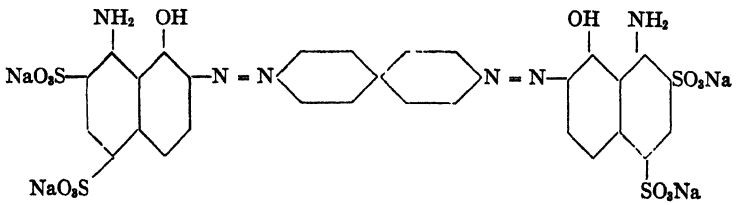


(corresponding with the isomorphism of the salts of H_3PO_4 and H_3AsO_4, etc.).

Erlenmeyer and Berger have stressed the importance of local similarities in electric field, resulting from equality in number of valence electrons of an atom or part of a molecule. Marrack [85] has drawn attention to the analogy with certain cases of remarkably specific *adsorption* to crystal surfaces.* Thus methylene blue is adsorbed by diamond and not by graphite, succinic acid by graphite and not by diamond. The only difference between these two crystals is the spacing of the carbon atoms; it will be recalled that we have already found spatial arrangement to be highly important for specificity. The dye I (below) is adsorbed on the crystal faces of potash alum, while its isomer, II, is not.



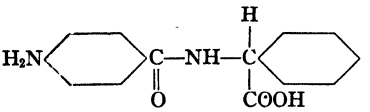
(I)



(II)

* See paper by W. G. France in this volume. J. A.

Differences have even been found in the degree to which optical isomers are adsorbed. Thus the dextro and levo forms of the dye made from phenyl-(*p*-aminobenzoylamino)-acetic acid



and dimethyl aniline, have been found to be adsorbed unequally by wool. It is certainly by no means inconceivable that the reaction between antibody and antigen may be at least in part caused by the same forces which act in these cases. Immunologists, however, no longer consider these forces any less chemical than the forces holding sodium and chlorine ions together in a crystal of sodium chloride. As Langmuir [78] puts it, there is ". . . no fundamental distinction between chemical and physical forces." *

(2). **Analogies with colloidal reactions.** Serological reactions have a number of points of similarity with many colloidal reactions, as was well brought out by Bordet.[10, 11] A consideration of these analogies will be helpful in understanding antibody-antigen reactions.

One important similarity is the ability of the reagents to enter into composition in almost any number of varying proportions, depending upon the ratio in which they are brought together. This is in sharp contrast to the behavior of simple chemical agents, and is evidently in both cases a consequence of the large number of reactive groups possessed by one or both of the reagents. Thus in serological reactions we find, as in colloidal reactions, that if an excess of one reagent is added, much more of it will be combined with than would be necessary to exert its typical effect; but *not more in proportion* to the excess added. That is, as more reagent is added, the additional amount taken up begins to taper off, giving a curve which flattens out, like an adsorption isotherm.

The two sorts of reactions also resemble each other in the importance of electrolytes, which in each case have considerable effect on the rate of reaction and the stability of the reagents. In both cases it may be found that without electrolytes the reaction does not go, at least to the extent of producing a visible change. The electrolyte concentration is also sometimes observed to have an important effect on the composition of the product which is formed. We may suppose that the effect of electrolytes depends in both cases on the important role played by electrical charges in the stability and reactivity of the substances.

The slow or imperfect reversibility of the reaction is also characteristic. It is well known that time is required for attainment of many colloidal equilibria, and that if an excess of one agent is added, the supersaturated compound which is formed is often only slowly converted to a compound of more "normal" composition when a corresponding amount of the second reagent is added. This is extremely reminiscent of the "Danysz phenomenon" in toxin-antitoxin reactions.

The effect of temperature, which may in some cases markedly diminish the amount of product or prevent its appearing, is also analogous.

Just as colloids often precipitate each other, antibody and antigen when mixed often produce a precipitate. The colloidal particles usually carry charges of opposite sign, which is not the case with antibody and antigen, though colloidal particles of the same sign may also combine.[89] Most striking of all is the fact that in both cases there is usually an "optimal" proportion in which the two reagents precipitate each other most rapidly and completely; if an excess of a certain one, or sometimes of either, reagent is used, precipitation is diminished, or absent. A good illustration

* See also Langmuir's paper in Vol. I, p. 525 *et seq.* J. A.

of this is seen in the reaction between ferric oxide sol and arsenious sulphide sol (Table 2).

Table 2. Flocculation of an As_2S_3 sol by a Fe_2O_3 sol[7]

Total vol. 10 cc		Degree of precipitation	Charge on particles in suspension
mg Fe_2O_3	mg As_2O_3		
0.61	20.3	cloudy	−
6.08	16.6	immediate, partial	−
9.12	14.5	nearly complete	
15.2	10.4	immediate, partial	+
24.3	4.14	delayed clouding	+
27.4	2.07	none	

The Fe_2O_3 particles are positively charged, the As_2S_3 particles negatively.

It will be noted that a well defined optimum, quite similar to those observed in some serological reactions, is obtained. In this case it evidently occurs at the point where the charge on the particles resulting from union of the two oppositely charged colloids is at a minimum; the reduction in the electrical repulsive forces allows the particles to approach each other closely enough to stick together. There are ways in which serological reactions may differ from this.

Second Stage of Serological Reactions (Neutralization, Complement Fixation, Precipitation, etc.)

When antibody reacts with simple haptens usually no second stage is observable, and it is likely that the primary stage of combination is the whole extent of the reaction. When toxins react with antitoxins, the toxic property is found to be wholly or partially neutralized. The combination of toxin and antitoxin in vitro is sometimes followed by another, visible phenomenon, and the combination of antibody with proteins, carbohydrates, and bacterial and other cells usually is. In spite of rather extensive study, the exact mechanism of these secondary reactions is still a matter of spirited controversy. The attempt here will be to present the facts, leaving controversy for other publications.

(1). **Toxin-antitoxin reactions.** These were among the first examples of immunological reactions to be studied, because of their practical importance, and the ease of determining the end point by test for toxicity on animals. We may consider them here as examples of serological reactions in which the first stage and second stage are probably one and the same, and consequently the reaction is particularly simple.

In the process of neutralizing toxins, antibody combines with them, in varying proportions, dependent on the ratio in which the reagents are mixed. Erlich early demonstrated that an antibody to the vegetable toxin, ricin, would neutralize *in vitro* its hemolytic power. Similarly it has been found that neutralizing antibodies for vaccinia virus can be adsorbed by the elementary bodies,[102] which points to union between the virus and the antibody.

It is not known how the combination of antibody with toxins and viruses neutralizes them. Toxins are evidently not permanently changed by combination with antitoxin, for it has been found that diphtheria toxin, for instance, can be recovered from the combination by destroying the antitoxin.[19, 109] Similarly it has been shown by Todd,[107] Andrewes,[2] and others that neutral mixtures of antibody and virus can in some cases be made active again by simple dilution or centrifugation. Perhaps the action of the antibody is simply to cover the toxin or virus and thus prevent it from coming in contact with the susceptible tissue. Salaman[102] suggests that antiviral antibodies act by preventing the entry of the virus into the cell.

The quantitative course of such reactions was studied by Arrhenius. He assumed that the mass law was applicable. In its simplest form, however,

$$\frac{\text{(free toxin) (free antitoxin)}}{\text{(toxin-antitoxin compound)}} = K$$

(the parentheses indicate concentrations of the respective substances) this formula does not apply, so Arrhenius [3] assumed that the reaction took place according to the following equation:

$$\text{toxin} + \text{antitoxin} = 2\ \text{(toxin-antitoxin)}$$

that is, that the reaction of one molecule of antitoxin with one of toxin gave two molecules of product. This led to the following relation:

$$\frac{\text{(free toxin) (free antitoxin)}}{\text{(toxin-antitoxin)}^2} = K$$

The values of q shown in Table 3 were calculated from this, assigning K the

Table 3. Observed Toxicity (q) of Tetanolysin after Addition of n cc of Antilysin, Showing also q Calculated in Various Ways

n	q (obs.)	q (calc. by mass law) ‡	q (calc. by "adsorption") *	q (calc. by "adsorption") †
0	100	100	100	100
0.05	82	82	82	83
0.1	70	66	65	67
0.15	52	52	50	52
0.2	36	38	37	39
0.3	22	23	21	22
0.4	14.2	13.9	13.4	14.1
0.5	10.1	10.4	9.5	10
0.7	6.1	6.3	5.8	6.1
1.0	4.0	4.0	3.7	3.8
1.3	2.7	2.9	2.7	2.8
1.6	2.0	2.5	2.1	2.1
2.0	1.8	1.9	1.6	1.7

* Equation of Langmuir.
† Equation of Ghosh.
‡ Equation of Arrhenius.

value of 0.115, and assuming that 0.276 cc of the antitoxin was equivalent to the quantity of toxin used. Arrhenius considered that the numerical agreement proved that the assumptions were correct, but his formulation is really in error in two respects. Toxin and antitoxin combine in more than one proportion, instead of giving only a single type of compound as Arrhenius assumed, and a molecule of toxin and one of antitoxin give only one, not two molecules of product. The agreement between the observed and calculated values of q is therefore fortuitous.

Since toxins and antitoxins combine in multiple proportions, the above simple mass law expression is not applicable, and we must assume a set of equilibria of the type

$$\frac{(A)\,(A_{n-1}G)}{(A_nG)} = k \qquad \frac{(A)\,(A_{n-2}G)}{(A_{n-1}G)} = k' \qquad \frac{(A)\,(A_{n-3}G)}{(A_{n-2}G)} = k'' \text{ etc.}$$

where A = antibody, G = antigen, and A_nG, $A_{n-1}G$ etc., represent compounds of antibody and antigen in various proportions. It can be shown that if $k = k' = k'' = \cdots\cdots$ and a molecule of antibody occupies one combining site on the antigen, the quantitative relation between combined antitoxin and free antitoxin, or between combined and free toxin, will reduce to the Langmuir "adsorption" isotherm

$$R = nkx \quad (1 + kx)$$

where R = the amount of antibody combining with a fixed amount of antigen (or vice versa), n and k are constants, and x is the final concentration.

Kinetically this amounts to assuming that antibody combines reversibly with a limited number of points on the surface of the antigen, which points are all alike for our purposes.[cf. 88]

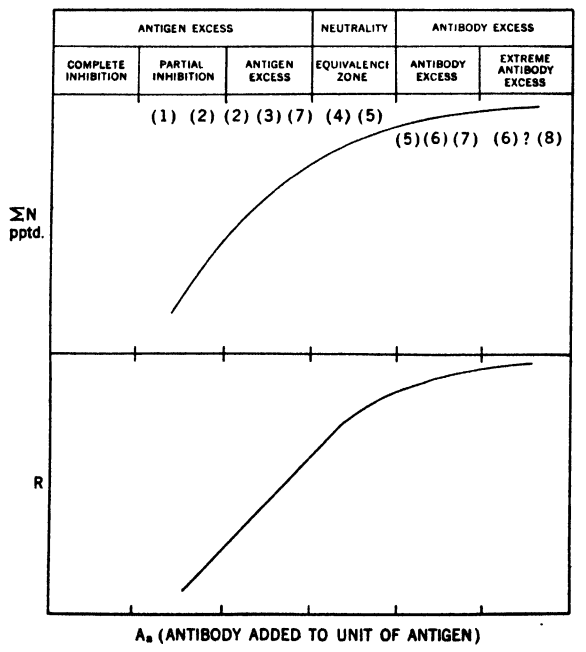
The results of many experimenters [20, 23, 8, 40, 41, 80] can be fitted by equations of the adsorption type. One of the most used has been the "Freundlich" adsorption isotherm

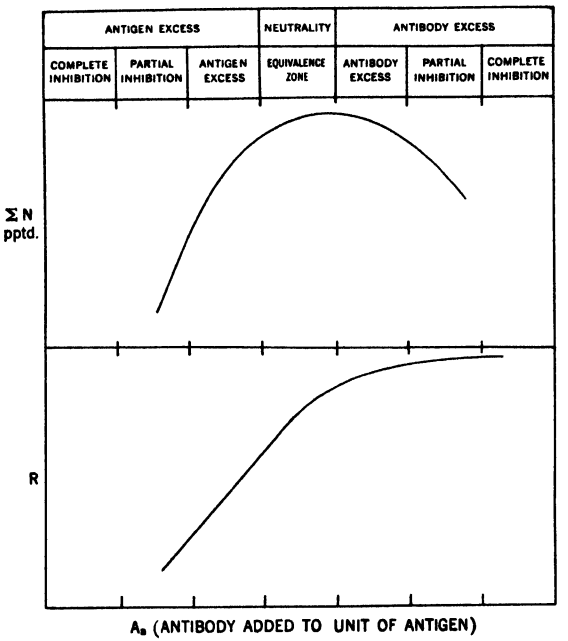
$$R = kx^n$$

where R is the amount of substance adsorbed per unit of adsorbent, x is the final concentration of the adsorbed substance in the supernatant, and k and n are constants. The Langmuir [77] equation is perhaps somewhat to be preferred since it can be derived from statistical mechanical considerations, without any reference to the mechanism of the reaction,[87] and would thus be expected to apply, allowing for its somewhat oversimplified character, to any action taking place reversibly at a surface where many combining sites are present.

(2). **The precipitin reaction.** The formation of a precipitate when antibody and soluble antigen are mixed is one of the most striking phenomena of serological reactions; the recent studies of it have yielded results of great interest and importance.

It has been found convenient to divide the whole range of the reaction of rabbit precipitin with antigen into the following five zones in order of increasing antibody and antigen ratio: complete inhibition, partial inhibition, antigen excess, equivalence, antibody excess. In the case of horse anti-protein precipitin it would be necessary to add two more zones, partial antibody inhibition, and complete antibody inhibition. The meaning of these terms will become clear in the following pages. A schematic





presentation of this arrangement, which also serves to indicate approximately the meaning to be attributed to each term, is given in Figures 1a and 1b.

Amount of precipitate: zones. Either the amount of antiserum or the amount of antigen can be varied; it is usual to keep the concentration of one reagent constant, and add different amounts of the other. Precipitates are now usually analyzed by some modification of the micro-Kjeldahl technic employed by Parnas and Wagner.[94] From the results so obtained we can calculate back to the basis of constant amounts of serum or of antigen. A series of such determinations made with different proportions of reagents gives a system of data which if graphed will form a surface which shows how the composition of the precipitate depends on the concentrations of the two reagents; two dimensional graphs can be made by showing appropriate sections of this.[14]

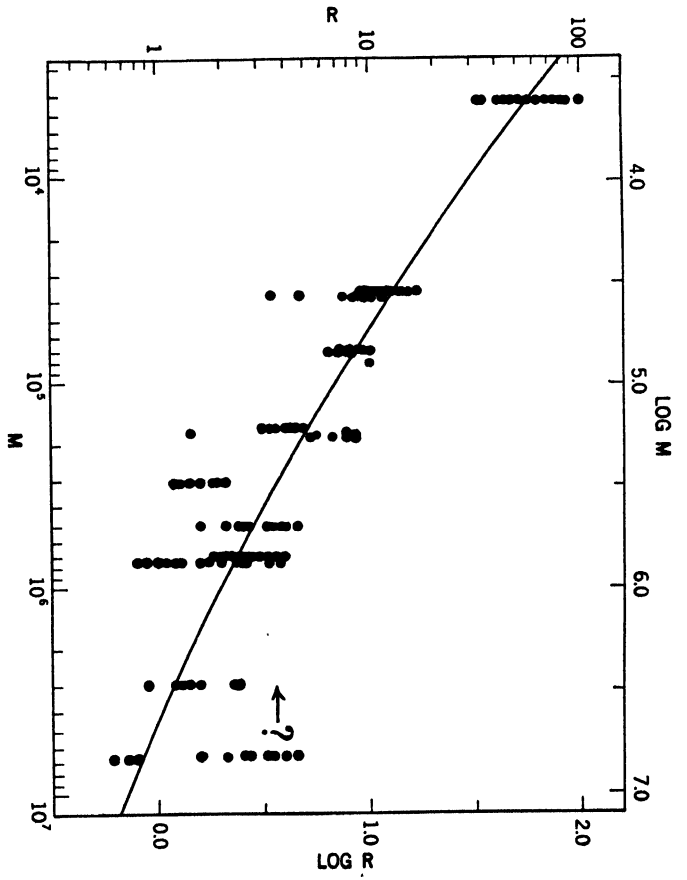
Most of the precipitating antisera that have been studied quantitatively thus far were obtained from the horse or the rabbit. In the horse there is a striking difference in the behavior of anti-protein antibodies and antibodies against bacterial polysaccharides; in the rabbit this difference is not found, but both kinds of antibodies resemble more the anti-carbohydrate antibody of the horse.

Sera from the two species differ in this way: in the horse it is found that if more than a certain relative amount of antibody is added no precipitate is produced, soluble compoonds being formed (*i.e.*, horse anti-protein sera exhibit a prezone); whereas carbohydrates are precipitated even with large excess of horse antibody. Rabbit antibodies, whether against protein or carbohydrate, do not show a prezone with excess antibody. It may be suspected that the difference is at least partly due to a difference in solubility of the various antibodies,[14] for it will be recalled that the anti-carbohydrate antibodies in the horse, and most antibodies of the rabbit, are predominantly found in the euglobulin (least soluble) fraction of the serum proteins, while the anti-protein antibody in the horse is mostly of pseudo-globulin (more soluble) character.

All precipitating systems exhibit the phenomenon of the postzone which is unfortunately usually called the prezone (since in precipitin reactions it is usually the antigen which is diluted); for this reason and for others, it seems best to use Heidelberger's term "inhibition zone" for the zone in which there is insufficient antibody to precipitate the antigen present. The antibody in this zone forms soluble compounds with the antigen.

The actual amount of precipitate obtained for a given amount of antiserum depends in the first place on the amount of precipitating antibody present, and second on the ratio of antibody to antigen in the precipitate, which in turn depends on the amount of antigen added. As increasing amounts of antigen are added, the amount of precipitate increases up to a maximum, then declines as the inhibition zone is reached. In the horse anti-protein systems a prezone with antibody excess is observed.

If the supernatants from such precipitates are tested for antibody and antigen, it is found that antibody is present in the first tubes (where insufficient antigen was added), and antigen present in the last tubes (where excess antigen was added). In the central part of the range one or more tubes will be found where there is either no antibody and no antigen, or small traces of both, in the supernatant. This is called the equivalence zone (Heidelberger). In some systems it is found that this zone



may coincide with, or lie near, the point of maximal precipitation. In other systems, however,[56, 88] the point at which the maximal precipitate is obtained may lie in the region of antigen excess.

Influence of molecular weight of the antigen. Zinsser[116] showed that the relative surface involved would explain the well known observation that less antibody is required to agglutinate bacteria than to precipitate, for example, a protein antigen. Boyd and Hooker[16, 17] called attention to the fact that the ratio of antibody to antigen in precipitates made at the equivalence point (mid-point of equivalence zone) depends on the molecular weight of the antigen. The dependence is not absolute, as will be seen in Figure 2, which includes data up to 1940, but it was found that there was at least a highly significant degree of correlation between the two variables. Boyd and Hooker interpreted this tentatively as due to the fact that at the equivalence point the surface of the antigen molecule is just covered with molecules of antibody, which was supposed to act as a chain of three (or better, four) roughly spherical units of molecular weight about 35,000. It would do equally well to suppose that the antibody behaved as a long flexible ellipsoid. On the basis of such a picture it was possible to calculate by spherical geometry that the theoretical relation between molecular weight of the antigen and ratio by weight of antibody to antigen in the precipitates ought to be:

$$R = 37{,}800M^{-0.8} + 179M^{-0.35}$$

In Figure 2 this theoretical relation, shown by the solid line, is seen to fit the data rather well, considering the roughness of many of the measurements. It is doubtful, however, if we are justified in considering the polysaccharide molecules, here plotted at the value $M = 4{,}000$, as spherical, so it may be that part of the agreement is fortuitous. At any rate, it can be seen that the ratio does tend to diminish with increasing molecular weight of the antigen, falling from values of around 10 for antigens of molecular weight 35,000-40,000 to values of less than 1.00 for the large hemocyanin molecules of M W ca. 7,000,000.

How[68] has more recently derived an improved formula giving somewhat better agreement with the data.

When less than the equivalent amount of antibody is used, the precipitate contains less antibody, and we may suppose the surface to be only partly coated. When more antibody is used, the ratio of antibody to antigen in the precipitate goes up. If antibodies are really to be visualized as just described, it is necessary to suppose that some of the molecules in this case are not in full contact with the antigen surface, and when there is considerable excess of antibody we should have to visualize them as "standing on their heads."[64]

Influence of proportion in which reagents are mixed. The composition of the precipitate depends on the final concentration of antibody or antigen in the supernatant, as would be expected if an equilibrium is reached.

But few direct measurements of such equilibrium concentrations have been made, however, because of analytical difficulties, whereas we can easily make precipitates by mixing antibody and antigen in various proportions. Therefore it is desirable to investigate also the way in which the composition of the precipitate depends on the ratio in which the reagents are mixed.

If to a constant amount of antibody there is added more than the equivalent amount of antigen, the compounds formed will have a lower antibody/antigen ratio Above a certain point these compounds are soluble, in the case of all systems, and no precipitate is formed. We may refer to this region of no precipitation as the "zone of complete inhibition" or prezone. The ratio of antibody to antigen in the precipitate therefore never falls below a certain value which, if it could be measured, would give the approximate composition of the compound having the smallest proportion of antibody which will still precipitate. For small antigen molecules, this is

estimated at about two molecules of antibody to one of antigen, but for larger molecules as many as 30-40 molecules of antibody may be required to precipitate one molecule of antigen. This is shown in Table 5.

When the proportion of antigen remaining unprecipitated is small, serological methods, [56, 64] can be used to estimate it, but the only analytical methods of general applicability to this problem depend on the use of antigens which are colored, or contain an inorganic atom which can be tested for chemically, or which contains some introduced prosthetic group which enables them to be detected.[112, 113, 65, 55, 83] Only the last two groups of investigators made any extensive studies of the region of antigen excess. Malkiel and Boyd [83] found in their two systems that there was a linear relation between the ratio of antibody to antigen and the amount of antibody added to a given amount of antigen. Investigation showed that this relation probably held for other systems, for if the data of Heidelberger and Kendall [55] were plotted similarly, a straight line was also obtained.

In the region of antibody excess, Heidelberger and Kendall [55, 56] have also reported another linear relation, between R and the amount of antigen, N, precipitated (from a constant amount of serum). Since in this region all (or nearly all) of the antigen is precipitated, this is in most cases equivalent to a linear relation between R and the amount of antigen added, thus: $R = a - b\ G_a$

Farther in the region of antibody excess, where the antibody content of the precipitate approaches an upper limit asymptotically, a linear relation cannot be expected to hold, and it has been found that it does not. Heidelberger and Kendall [55, 56] found that better agreement was obtained by plotting, not the linear equation referred to above, $R = a - b\ G_a$, (where R is the ratio of antibody to antigen in the precipitate, and G_a is the antigen added to a given quantity of antibody), but the curvilinear relation $R = a - b\ G_a^{1/2}$. The relation expected between R and G_a on the basis of the Freundlich equation gives for this region, when graphed, a line having a slight curvature, but nearly straight.

Table 4. Quantitative Relations between Amount of Antigen (Brucella) Added, Amount of Antigen and Antibody Precipitated, and Composition of Precipitate (data of Pennell and Huddleson [97]); together with Predicted Results from Various Theoretical and Empirical Equations

				Amounts in milligrams				
Antigen added	Antigen pptd.	Antibody pptd. (obs.)	"Free antibody" (p)	Antibody pptd. (eq. 1)	Antibody pptd (eq 2)	Antibody pptd (eq 3)	Antibody pptd. (eq 4)	Antibody/antigen in ppt. (R)
0.05	all	0.066	0.557	0.064	0.068	0.066	0.066	1.32
0.10	all	0.128	0.495	0.125	0.127	0.128	0.127	1.28
0.15	all	0.175	0.448	0.182	0.180	0.189	0.184	1.16
0.20	all	0.229	0.394	0.236	0.229	0.244	0.236	1.14
0.30	all	0.324	0.299	0.333	0.318	0.324	0.324	1.08
0.50	0.338	0.362	0.261	0.367	0.349	0.375	0.352	1.06
0.70	0.616	0.506	0.117	0.545	0.505	0.505	0.498	0.82
1.00	0.946	0.575	0.048	0.614	0.605	0.464	0.576	0.61
1.20	0.998	0.580	0.043	0.611	0.612	0.449	0.579	0.58
Σx^2	...	....	....	....	4.428	2.156	29.991	0.297	...

The equations used for the calculations of the predicted values of antibody precipitated were:
(1) $R = 1.325 - 0.714\ G_a$ (Heidelberger and Kendall)
(2) $R = 1.58 - 0.968\ G_a^{\frac{1}{2}}$ (Heidelberger and Kendall)
(3) $R = 14.84\ p/(1 + 957p)$ (Langmuir, Ghosh)
(4) $R = 1.57\ p^{0.31}$ ("Freundlich")

In these equations R = the ratio by weight of antibody to antigen in the precipitate, G_a = antigen added, p = the concentration of unprecipitated antibody (mg in each 2 cc), shown in column 4 (obtained by subtracting the values in column 3 from 0.623), x = (true value of antigen precipitated − predicted value) × (10^2).

The expected values of "antibody precipitated" are calculated from the predicted values of R by multiplying by the amount of antigen precipitated.

In Table 4 will be found data on the composition of precipitates in the region of antibody excess and an example of the fitting of experimental data by the two equations of Heidelberger and Kendall and by simple equations of the "adsorption" type. All predict the composition of the precipitate moderately well. The least squares technic indicates that in the present case definitely the best fit is obtained with the "Freundlich" type of equation.

The suitability of the others decreases in the order: Heidelberger and Kendall's empirical equation, their theoretical equation, the Langmuir-Ghosh equation. The fit obtained with the Freundlich equation is about as good as could be expected from the data. Other sets of data might conform better in some cases to one or the other equations.

Molecular composition of the precipitates. The application of quantitative methods to the study of the precipitin reaction, plus the recent determinations of molecular weights of various proteins, including antibodies, by the ultracentrifugal technic, have enabled the calculation of the actual molecular composition of precipitates made under various conditions. In Table 5 are given the results of the studies of Heidelberger and collaborators [52] to which have been added results calculated from the data of Malkiel and Boyd,[83] Pappenheimer,[91] Heidelberger, Treffers and Mayer,[60] and Pappenheimer, Lundgren and Williams.[92]

Table 5. Molecular Composition of Specific Precipitates

Antigen	Antibody	Empirical composition of precipitate at extreme antibody excess	antibody excess end of equivalence zone	antigen excess end of equivalence zone	zone of partial inhibition	Composition of soluble compounds in zone of partial inhibition
pvalbumin	rabbit	A_6G	A_3G	A_6G_2	A_2G	(AG)
dye-ovalbumin	rabbit	(A_6G)	(A_3G)	A_6G_2	A_3G_4	(AG_2)
serum albumin	rabbit	A_6G	A_4G	A_3G	A_2G	(AG)
thyro-globulin	rabbit	$A_{40}G$	$A_{14}G$	$A_{10}G$	A_2G	(AG)
viviparus hemocyanin	rabbit	....	$A_{120}G$	$A_{83}G$	$A_{36}G$	....
diphtheria antitoxin	horse	A_6G	A_4G	A_3G_2	AG	AG_2
ovalbumin	horse	(A_4G)	A_2G		AG	(AG_2)

A = antibody, G = antigen
Formulas in parentheses are somewhat uncertain.

From Table 5 it may be concluded, since the ratio of antibody to antigen in precipitates is observed to vary continuously, that if the above formulas really represent the composition of the various antibody-antigen possible, the precipitate obtained at any particular point is a mixture of two or more of these compounds.

References

1. Alexander, J., *Protoplasma,* **14**, 296 (1932). See also paper in this volume, "Catalysis as a Biological Factor."
2. Andrewes, C. H., *J. Path. Bact.,* **31**, 671 (1928).
3. Arrhenius, S., "Immunochemistry," 1907.
4. Avery, O. T., Goebel, W. F., and Babers, F. H., *J. Exp. Med.,* **55**, 769 (1932).
5. —, ——, *J. Exp. Med.,* **58**, 731 (1933).
6. Banzhaf, E. J., and Famulener, W., *Studies, N. Y. City Dep't Health,* **8**, 208 (1914).
7. Billitzer, J. *Z. physik. chem.,* **51**, 129 (1905).
8. Biltz, W., *Biochem. Z.,* **23**, 27 (1909).
9. Boivin, A., and Mesrobeanu, I., *Rev. Immunol.,* 1, **553** (1935).
10. Bordet, J., "Traité de l'Immunité," Paris, 1920.
11. —, "Traité de l'Immunité" Paris, 1939.
12. Boyd, W. C., unpublished.
13. —, *J. Immunol.,* **38**, 143 (1940).
14. —, *J. Exp. Med.,* **74**, 369 (1941).
15. —, and Bernard, H., *J. Immunol.,* **33**, 111 (1937).
16. —, and Hooker, S. B., *J. Gen. Physiol.,* **17**, 341 (1934).
17. —, —, *J. Gen. Physiol.,* **22**, 281 (1939).
18. —, Conn, J. B., Gregg, D. C., Kistiaskowsky, G. B., and Roberts, R. M., *Science,* **92**, 604 (1940).
19. Calmette, A., *Ann. Inst. Pasteur,* **9**, 225 (1895).
20. Chow, B. F., and Goebel, W. F., *J. Exp. Med.,* **62**, 179 (1935).

21. —, and Wu, H., *Chinese J. Physiol.*, **11**, 139 (1937).
22. Clutton R. F., Harington, C. R., and Yuill, M. E., *Biochem. J.*, **32**, 1111 (1938).
23. Cromwell, H. W., *J. Immunol.*, **7**, 461 (1922).
24. Dakin, H. D., and Dudley, H. W., *J. Biol. Chem.*, **15**, 271 (1913).
25. —, and Dale, H. H., *Biochem. J.*, **13**, 248 (1919).
26. Dale, H. H., and Hartley, P., *Biochem. J.*, **10**, 408 (1916).
27. Dingle, J. H., Fothergill, L. D., and Chandler, C. A., *J. Immunol.*, **34**, 357 (1938).
28. Eagle, H., *J. Immunol.*, **18**, 169 (1930); **18**, 393 (1930).
29. Eisenberg, P., and Volk, R., *Z. Hyg. Infekt. kr.*, **40**, 155 (1902).
30. Erlenmeyer, H., and Berger, E., *Biochem. Z.*, **252**, 22 (1932); **255**, 429 (1932).
31. Felton L. D., *J. Infectious Diseases*, **37**, 199, 309 (1925).
32. —, *J. Infectious Diseases*, **42**, 256 (1928); **43**, 543 (1928).
33. —, *J. Immunol.*, **21**, 357 (1931).
34. —, *J. Immunol.*, **27**, 379 (1934).
35. —, and Kauffmann, G., *Bull. Johns Hopkins Hospital*, **62**, 430 (1938); *U. S. Pub. Health Report*, **53**, 1855 (1938).
36. Fourt, L., Harkins, W. D., and Wright, G. G., Abstracts 98th Meeting Am. Chem. Soc., 1939.
37. Fowler, R. H., "Statistical Mechanics," Cambridge, 1936.
38. Francis, T., *Proc. Soc. Exp. Biol. Med.*, **31**, 493 (1933).
39. Freundlich, H., "Kapillarchemie," Leipzig, 1923
40. Ghosh, B. N., *Indian J. Med. Res.*, **23**, 285 (1935).
41. —, *Indian J. Med. Res.*, **23**, 837 (1936).
42. Gibson, R. B., and Collins, K. R., *J. Biol. Chem.*, **3**, 233 (1907).
43. Goebel. W. F., *J. Exp. Med.*, **68**, 469 (1938).
44. —, and Hotchkiss, R. D., *J. Exp. Med.*, **66**, 191 (1937).
45. Gonzales, P., and Armangué, M., *Compt. rend. soc. biol.*, **106**, 1006 (1931).
46. Goodner, K., and Horsfall, F. L., *J. Exp. Med.*, **64**, 201 (1936).
47. —, —, *J. Exp. Med.*, **66**, 413, 425, 437 (1937).
48. Green, A. A., McKhann, C. F., Kapnick, I. R., and Fahey, K. R., *J. Immunol.*, **36**, 245 (1939).
49. Harris, T., and Eagle, H., *J. Gen. Physiol.*, **19**, 383 (1935).
50. Haurowitz, F., "Chemie der Antigene und der Antikörper," in "Fortschritte der Allergielehre," 1939.
51. —, Kraus, F., and Marx, F., *Z. physiol. chem.*, **245**, 23 (1936).
51a. Haurowitz, F., and Schwerin, P., *Brit. J. Exp. Path.*, **23**, 146 (1942).
51b. —, —, *J. Immunol.*, **47**, 111 (1943).
52. Heidelberger, M., *Bact. Rev.*, **3**, 49 (1939).
53. —, and Kendall, F. E., *J. Exp. Med.*, **59**, 519 (1934).
54. —, —, *J. Exp. Med.*, **61**, 559, 563 (1935).
55. —, —, *J. Exp. Med.*, **62**, 467 (1935).
56. —, —, *J. Exp. Med.*, **62**, 697 (1935).
57. —, and Landsteiner, K., *J. Exp. Med.*, **38**, 561 (1923).
58. —, and Pederson, K. O., *J. Exp. Med.*, **65**, 393 (1937).
59. —, —, and Tiselius, A., *Nature*, **138**, 165 (1936).
60. —, Treffers, H. P., and Mayer, M., *J. Exp. Med.*, **71**, 271 (1940).
61. Hooker, S. B., *J. Immunol.*, **33**, 57 (1937).
62. —, and Boyd, W. C., *J. Immunol.*, **23**, 465 (1932).
63. —, —, *J. Immunol.*, **26**, 469 (1934).
64. —, —, *J. Gen. Physiol.*, **19**, 373 (1935).
65. —, —, *J. Immunol.*, **30**, 33 (1936).
66. —, —, unpublished.
67. Hopkins, S. J., and Wormall, A., *Biochem. J.*, **27**, 740 (1933); **27**, 1706 (1933).
68. How, A. E., *J. Immunol.*, **37**, 77 (1939).
69. Johnson, L. R., and Wormall, A., *Biochem. J.*, **26**, 1202 (1932).
70. Kabat, E. A., *J. Exp. Med.*, **69**, 103 (1939).
71. —, and Heidelberger, M., *J. Exp. Med.*, **66**, 229 (1937).
72. Kunifusa, J., *Keijo J. Med.*, **1**, 409 (1930).
73. Landsteiner, K., "The Specificity of Serological Reactions," 1936.
74. —, and Barron, C., *Z. Immunitäts.*, **26**, 142 (1917).
75. —, and Jacobs, J., *Proc. Soc. Exp. Biol. Med.*, **30**, 1055 (1933).
76. —, and van der Scheer, J., *J. Exp. Med.*, **71**, 445 (1940).
77. Langmuir, I., *J. Am. Chem. Soc.*, **40**, 1361 (1918).
78. —, Nobel Lecture, 1932.
79. Ledingham, J. C. G., *J. Hyg. Camb.*, **7**, 65 (1907).
80. Levine, B. S., *J. Immunol.*, **35**, 131 (1938).
81. Lewis, J. H., *J. Infectious Diseases*, **55**, 168 (1934).
82. Lin, K., Wu, H., and Chen, T., *Chinese J. Physiol.*, **2**, 131 (1928).
83. Malkiel, S., and Boyd, W. C., *J. Exp. Med.*, **66**, 383 (1937).
84. Markin, L., and Kyes, P., *J. Infectious Diseases*, **65**, 156 (1939).
85. Marrack, J. R., "The Chemistry of Antigens and Antibodies," in Med. Res. Council Special Report Series, London, 1938.
86. Meyer, K., *Compt. rend. soc. biol.*, **124**, 430 (1937).
87. Mudd, S., Nugent, R. L., and Bullock, L. T., *J. Phys. Chem.*, **36**, 229 (1932).
88. von Muralt, A. I., *J. Am. Chem. Soc.*, **52**, 3518 (1930).
89. Neurath, H., *J. Am. Chem. Soc.*, **61**, 1841 (1939).
90. Northrop, J. H., *Science*, **93**, 92 (1941).
91. Pappenheimer, A. M., *J. Exp. Med.*, **71**, 263 (1940).
92. —, Lundgren, H. P., and Williams, J. W., *J. Exp. Med.*, **71**, 247 (1940).
93. Parfentjev, I. A., U. S. Patent 2,065,196, 1936.
94. Parnas, J. K., and Wagner, R., *Biochem. Z.*, **125**, 253 (1921).
95. Pauling, L., "The Nature of the Chemical Bond," 1939.
96. —, *J. Am. Chem. Soc.*, **62**, 2643 (1940).
97. Pennell, R. B., and Huddleson, I. F., *J. Exp. Med.*, **68**, 73 (1938).
98. Plaut, F., and Rudy, H., *Z. Immunitäts.*, **81**, 87 (1933).
99. Porter, E. F., and Pappenheimer, A. M., *J. Exp. Med.*, **69**, 755 (1939).
100. Reymann, G. C., *Z. Immunitäts.*, **39**, 15 (1924).
101. Rothen, A., and Landsteiner, K., *Science*, **90**, 65 (1939).
102. Salaman, M. H., *Proc. III Int. Cong. Microbiol.*, (New York), 356 (1940).
103. Stokinger, A. E., and Heidelberger, M., *J. Exp. Med.*, **66**, 251 (1937).
104. Svedberg, T., and Pederson, K. O., "The Ultracentrifuge," 1940.

105. Ten Broeck, C., *J. Biol. Chem.*, **17**, 369 (1914).
106. Tiselius, A., and Kabat, E. A., *J. Exp. Med.*, **69**, 119 (1939).
107. Todd, C., *Brit. J. Exp. Path.*, **9**, 244 (1928).
108. Uhlenhuth, P., and Remy, E., *Z. Immunitats.*, **92**, 171 (1938).
109. Wassermann, A., *Z. Hyg.*, **22**, 263 (1896).
110. Weil, A. J., Parfentjev, I. A., and Bowman, K. L., *J. Immunol.*, **35**, 399 (1938).
111. Wells, H. G., "The Chemical Aspects of Immunity," 1925.
112. Wu, H., Cheng, L. H., and Li, C. P., *Proc. Soc. Exp. Biol. Med.*, **25**, 853 (1928).
113. Wu, H., Sah, P. P. T., and Li, C. P., *Proc. Soc. Exp. Biol. Med.*, **26**, 737 (1929).
114. Wu, H., Ten Broeck, C., and Li, C. P., *Chinese J. Physiol.*, **1**, 277 (1927).
115. Wyman, J., *J. Biol. Chem.*, **90**, 443 (1931).
116. Zinsser, H., *J. Immunol.*, **18**, 483 (1930); **18**, 227 (1930).

Allergy and Anaphylaxis

CARL A. DRAGSTEDT

Professor, Northwestern University Medical School, Chicago, Ill.

The terms *allergy* and *anaphylaxis* are used to include a considerable variety of biological reactions in which some foreign substance (antigen, allergen, anaphylactogen) provokes a reaction not by virtue of its inherent or specific pharmacological characteristics, but rather by virtue of some changed reactivity, hypersensitiveness, or hypersusceptibility of the animal or organ to which it is administered. Such reactions include the "true anaphylaxis" induced in experimental animals, as well as serum reactions, pollen, food and drug idiosyncrasies occurring in man. Richet introduced the term *anaphylaxis* to signify the antithesis of prophylaxis, with special reference to the hypersusceptible condition induced in experimental animals by a suitable preceding inoculation with various toxins. Von Pirquet introduced the term *allergy,* meaning altered reactivity, with special reference to the reactions of serum sickness in man and which he felt were analogous to those of the experimental animal. The experimental condition lent itself readily to analysis; and it soon became evident that any substance, whether toxic or not, which can induce antibody formation by injection into an animal (and therefore called an antigen), can initiate the change in the animal called sensitization, and, when reinjected after an appropriate interval or incubation period, can provoke a chain of symptoms of a highly characteristic nature, referred to as anaphylactic shock.

A convincing demonstration that this acquired sensitive state is dependent upon antibody formation was made by conveying a similar sort of sensitivity to a normal animal by injecting the blood serum of either a "sensitive" or an "immune" animal into a normal animal. Such a procedure, which is called passive sensitization, seems absolute proof of the association of sensitization with some sort of humoral antibody, since the recipient animal becomes sensitive after a latent period of a few hours, which seems too short a time for it to have generated its own reactivity. Laboratory methods can demonstrate the presence in the blood from sensitized animals of precipitins in a concentration which appears correlated with the degree of sensitivity of the animal and with the potency of the blood in conveying passive sensitization. Inasmuch as the apparently spontaneous conditions occurring in man were not so readily proved to be associated with an antigen-antibody reaction, but had nevertheless a certain similarity in basic pattern, Doerr extended the meaning of the term *allergy* to include all forms of changed reaction capacity, whether the substance which provoked the altered reaction was demonstrable as an antigen or not. Thus in Doerr's nomenclature, anaphylaxis is one form of allergy.

Cocoa advocated *hypersensitiveness* as a general designation, classifying this as anaphylaxis where the reactions clearly depended upon an antigen-antibody reaction,

and as allergy those reactions in which no such relationship exists.' There are two objections to this use of terms. In the first place it seems to pass final judgment that the allergic reactions are not antigen-antibody reactions, while the facts seem to indicate that it is only rather difficult to demonstrate specific antibodies in this case. In the second place, the term *hypersensitiveness* seems more appropriately descriptive of a quantitative rather than qualitative change in sensitivity. For example, we can artificially induce an enhanced sensitivity to epinephrine in an animal by the prior administration of cocaine, but the character of the reaction remains unchanged and is characteristic of epinephrine. Such examples of bona fide hypersensitiveness are not common, but they do exist, and they are different both in pattern and pathogenesis from allergic or anaphylactic reactions, in which the character of the response is not characteristic of the provocative agent.

The space allotted to this resume does not permit a comprehensive discussion of these phenomena, but a digest of the important characteristics of true anaphylaxis provides a good foundation for the consideration of all related reactions. Almost any protein, when injected parenterally, even in a dose which has no observable effect, can render almost any animal sensitive, so that the reinjection of this protein, after a suitable interval, will provoke an anaphylactic response. This general statement requires considerable qualification. All proteins are not equally effective (*i.e.*, equally antigenic or anaphylactogenic). Gelatin, hemoglobin, etc., are relatively ineffective. There is some evidence to indicate that the character of the amino acid content of a protein may determine its antigenic properties, but there is also the possibility that molecular size, etc., may be determining factors. All animals are not equally susceptible to sensitization. Guinea pigs are perhaps the most readily sensitized. Dogs, rabbits, monkeys, and rats are respectively less so. Other species demonstrate varying degrees of susceptibility. Such evidence as is available from inadvertent experiences in medical practice seems to indicate that most people are not readily sensitized. While some explanations for the species variations can be offered, no ultimately determining factors have been established. Within a given species there may be wide variations in susceptibility to sensitization. Whether this is due to genetically determined constitutional differences, to environmental factors, or both, is not fully known. There is considerable evidence to indicate a hereditary predisposition to allergic disease in man, and it may be that there is a corresponding hereditary factor in the experimental animal.

A suitable interval (incubation period) must elapse between the sensitizing and the assaulting injection for maximal reactions to occur. Seven days is an approximately minimum time, the optimal times varying in different animals up to three or four weeks. The sensitive state may persist in some animals (*e.g.*, guinea pigs) for many months or decline fairly rapidly in others (*e.g.*, rabbits). The assaulting injection is most effective in producing shock if made intravenously; a certain adequate dose is necessary, but further than this there is no absolute correlation between the size of the dose and the degree of the shock reaction produced. We may properly infer from this and the preceding characteristics that the antigen is not toxic *per se* but acts by some trigger mechanism to initiate a reaction whose severity rests upon a more complicated basis. As mentioned, the reaction may vary in severity from one of almost negligible symptoms to that of a violent and rapidly fatal form. For a certain period after recovery from a non-fatal reaction, the animal fails to react if the antigen is readministered. This abrupt loss of reaction capacity is called desensitization. This desensitization lasts only a few days in most cases, and the preceding sensitive state recurs. Experiments have shown that certain isolated smooth-muscle organs (such as uterus, intestine, etc.), when washed as free as possible from the blood of the sensitive animal from which they have been obtained, will react (by contraction) when exposed to the antigen to which the animal was sensitized. For this (and other reasons) the conception prevails that in an

intact animal an anaphylactic reaction occurs as a result of the interaction between the antigen and antibodies which are somehow bound to the fixed tissue cells. Since such isolated tissue preparations also show the phenomenon of desensitization, and since the blood of an animal which has been desensitized may not lose its ability to convey passive sensitization, it is generally assumed that desensitization is related to a temporary binding of fixed or sessile antibodies with antigen in some loose manner.

A comprehensive consideration of anaphylaxis, therefore, involves a variety of problems. For example, what are the ultimate factors which determine whether a given substance can function as an antigen, or more precisely as an anaphylactogen? What are the factors which determine the sensitizability of a given animal? If this is directly related to the animal's antibody production capacity, then what are the antibodies concerned, what is their origin and fate, and what conditions promote or inhibit their production? What is the physiological reaction which takes place when the antigen is injected, or, in other words, how can all the phenomena of anaphylaxis be accounted for? What is the explanation for the occurrence and the loss of the desensitization which follows a reaction? And finally, what appears to be Nature's purpose or design that such a reaction should occur? Is it some sort of defense reaction which has greatly overextended itself, or one which has taken an anomalous turn? Since it is not possible to answer all these questions, attention will be devoted primarily to the question of the proximal pathogenesis of the symptoms and phenomena of the anaphylactic reaction.

That all anaphylactic reactions in all animals have a similar fundamental mechanism is indicated by two general considerations. First, the basic pattern as to sensitization, incubation period, desensitization, etc., is strikingly similar in all cases. Secondly, the same foreign protein may serve as the anaphylactogen in several species of animals on the one hand, while on the other the character of the anaphylactic reaction in any given species of animal is always the same, even though widely different anaphylactogens are used. This is true even when the latter has its own inherent toxic properties, in which case the end effect is a composite of the anaphylactic and specifically toxic reactions. Since, however, the symptomatology of anaphylaxis is different in the various species of animals in which it has been studied, the necessity arises of scrutinizing and harmonizing this apparent discrepancy.

First known and often referred to as the Theobald Smith phenomenon, the anaphylactic reaction in the guinea pig was extensively studied by Otto, Rosenau and Anderson, and many subsequent investigators. After the intravenous injection of antigen into a sensitized animal, there is a prodromal period of ½ to 3 minutes during which the animal usually sneezes, scratches its nose, becomes restless, discharges urine and feces, becomes weaker and lies down; acute respiratory difficulty then develops, and death from asphyxia may follow in 2 to 10 minutes. Intraperitoneal or subcutaneous injection of the antigen is much less certain, but may produce death with similar symptoms. The first analysis of the physiological mechanism involved was that of Auer and Lewis. They demonstrated that the respiratory difficulty and asphyxia were due to a swiftly developing stenosis of the bronchioles, and that in spite of violent respiratory efforts the animal was unable to effect any air exchange. They demonstrated that this bronchial reaction occurred in animals which had been curarized, whose vagi had been cut, or whose spinal cord, medulla and basal brain had been destroyed, indicating that the reaction occurred peripherally. The observations of Auer and Lewis have been confirmed by numerous workers and the demonstration by Dale that the bronchospasm can be reproduced with the isolated, perfused lungs corroborated their conclusion that it was due to a direct action on the muscular walls of the bronchi and not due to a central or reflex nervous reaction. Auer and Lewis, Anderson and Schultz, and Loewit have recorded the systemic blood pressure in anaphylactic guinea pigs and report that there is an initial moderate rise followed by a gradual fall, such as occurs in asphyxia from any cause. It is

thus generally accepted that the predominant reaction is pulmonary and that the circulatory effects are secondary in importance.

Less prominent in the symptomatology are a pronounced fall in body temperature and a slight to moderate deficiency in the coagulability of the blood, which may be preceded by a brief period of noticeable hypercoagulability, a temporary leucopenia, etc. The most prominent post-mortem finding is the characteristic emphysematous distension of the lungs, while less noticeable are passive congestion of the liver and petechial hemorrhage in the intestines.

The first experimental study of anaphylaxis in the rabbit was made by Arthus. In this animal the reaction reveals itself either as a local or a general manifestation, depending upon the method of administration of the antigen. When the shocking injection is made subcutaneously (or intracutaneously, etc.), there may result infiltration, edema, sterile abscess or gangrenous slough at the site of the injection. This local reaction was first described by Arthus and is generally referred to as the Arthus phenomenon. When the reinjection of antigen is made intravenously the general reaction is seen. The respiration quickens, the animal sinks upon its abdomen, feces and urine are usually expelled, a fleeting hyperemia followed by pallor of the ears may be seen, and the animal may die in a few minutes. The respiration does not become dyspneic as it does in the guinea pig, and although some bronchospasm occurs it is seldom marked enough to cause asphyxia. The carotid blood pressure falls greatly, this fall sometimes being preceded by a temporary rise. There is an accompanying marked rise in the pulmonary artery pressure, and a number of workers have shown that the circulatory failure is mainly due to a pronounced obstruction to the flow of blood through the pulmonary circuit. Gilbert has shown that a marked contraction of the pulmonary arterioles occurs which may partly account for this, while an additional factor is probably the aggregation of leukocytes into emboli which obstruct the smaller pulmonary vessels, a phenomenon witnessed by Abell and Schenk in the ear vessels. There may be a minor cardiac element to the circulatory failure, as Rocha e Silva has shown that occasionally section of the vagi during the fall of the systemic blood pressure may produce a prompt recovery. This vagal inhibition of the heart seems to be a reflex secondary to the pulmonary phenomena and is probably comparable to that which is often seen associated with pulmonary embolism. Diminished coagulability of the blood, leucopenia, and fall in body temperature are usually found.

The analysis of the anaphylactic reaction in dogs began with the work of Richet, was soon amplified by Biedl and Kraus, Arthus, and many subsequent workers. In the unanesthetized animal the prominent symptoms are salivation, vomiting, diarrhea, dyspnea, and profound weakness. There is a marked fall in the systemic blood pressure. This is not due to cardiac weakness, but is readily shown to be due to peripheral vascular reactions. There is a marked swelling and engorgement of the liver, associated with a marked rise in the portal vein pressure, indicating a marked impairment to the flow of blood through the liver. Weil ascribed this obstruction to the swelling of the hepatic parenchymal cells; and Simonds has presented evidence that it may be accounted for by a contraction of the smooth muscle in the hepatic veins, which is particularly abundant in this animal. A number of workers have shown that anaphylactic shock is prevented entirely or greatly reduced in severity if the liver is either removed or short-circuited out of the circulation. The hepatic reaction, therefore, plays a dominant role in the genesis of the symptoms in the dog. In this animal, defective coagulability of the blood is a prominent finding, occurring in a large percentage of shocked animals, and not infrequently achieving a state of complete incoagulability.

Although there is thus a difference in the gross pictures of anaphylaxis as expressed in the guinea pig, rabbit and dog, there is a striking uniformity in the underlying physiological reactions. A contraction of smooth muscle, a dilatation of

capillaries, an increased secretion of various glands, a deficiency in the coagulability of the blood, a temporary leucopenia, and a drop in body temperature are common to all species. Simonds was one of the first to point out that contraction of strategically located smooth muscle (*i.e.*, in the bronchi of guinea pigs, in the pulmonary arterioles of rabbits, in the hepatic veins of dogs) could explain the divergence in symptomatology which is met with in different animals on the basis of an identical physiological reaction. Since the phenomena which can be interpreted as consequences of such smooth-muscle contraction include the most prominent elements in the category of symptoms (*e.g.*, bronchospasm in guinea pigs, circulatory failure in rabbits and dogs) there was for a time a tendency to view smooth-muscle contraction as the essential anaphylactic response, and to consider all the other phenomena as subjacent thereto. For many years, therefore, the conviction prevailed that if there were an anaphylactic poison (anaphylatoxin) into which the antigen was metamorphosed upon its injection, or which developed from any other source, that it must possess smooth muscle stimulating properties.

During the past ten years evidence has accumulated which permits the statement that histamine has been proved to be the smooth-muscle stimulating factor which is probably largely responsible for many of the symptoms of anaphylactic shock. The evidence seems unequivocal for the dog and the guinea pig, and virtually so for the rabbit. The evidence indicates that the histamine does not originate from the injected antigen, but is liberated from the tissues of the sensitized animal. In the dog the liver appears to be the major source, in the guinea pig the lungs, and in the rabbit the cellular elements of the blood. With the recognition that an agent mediating at least some of the symptoms of anaphylaxis could be isolated and identified in the blood and lymph as an initial consequence of the anaphylactic response, efforts to detect other such agents were intensified. Heparin has been isolated from the blood of anaphylactic dogs, shown to be adequate to account for such defective coagulability as occurs, and shown to be derived almost exclusively from the liver. That heparin is responsible for the defective coagulability of the blood in rabbits and guinea pigs is highly probable. Kellaway and Trethewie have presented evidence for the liberation of an unidentified principle which they call Slow Reacting Substance (S-R-S), during *in vitro* anaphylaxis in guinea pigs. There is a certain amount of suggestive but unconvincing evidence that choline or acetylcholine and potassium may be liberated from tissue sources and mediate some of the symptoms. If the role which such agents play is similar to, but relatively inconspicuous in comparison with that of histamine, etc., it is obvious that it is difficult to clearly establish or definitely deny their causal significance.

The cardinal symptoms of anaphylaxis can thus be explained as being due, in the immediate instance, to an autointoxication produced by physiologically active substances liberated from the tissues of the sensitized animal. If we attempt an inventory of all of the anaphylactic phenomena, in a much more detailed fashion than that presented previously, we find that there are a number of reactions which are not accounted for by the major anaphylatoxins, histamine and heparin, nor by the minor anaphylatoxins, choline, acetylcholine, S-R-S, or potassium. Some of these are the transitory hypercoagulability of the blood, which precedes the hypocoagulability, the decreased sedimentation rate of the erythrocytes, etc. Though histamine can account for the fatal bronchospasm in the guinea pig, the violent circulatory effects in rabbits and dogs, and heparin for the characteristic incoagulability of the blood during anaphylaxis, neither of these agents, alone or combined, can completely account for the local necrosis, etc., of the local anaphylaxis, or Arthus reaction. That the latter is a counterpart of the general anaphylactic reaction, dependent upon an antigen-antibody reaction, etc., is clearly established. Careful morphologic studies of the processes underlying the Arthus reaction demonstrate such a marked similarity to those of an acute inflammatory reaction, that one seems warranted to state

that an anaphylactic reaction is merely an exaggerated inflammatory reaction. The imprint of histamine in the capillary dilatation, permeability and local edema is clearly evident, but histamine alone cannot account for either the leukocytic response or the necrosis of tissue. It is thus clear that, though we can explain most of the phenomena of anaphylaxis, we cannot at present explain them all. There are some hints from both early and late investigations which suggest that some of the unexplained phenomena may be related to enzyme reactions, but the present status of this evidence does not warrant discussion here.

As yet little can be said regarding the manner in which the injected antigen provokes the liberation from the tissues of these varied cellular reactants. From studies on reactions which closely simulate anaphylaxis in symptomatology (peptone shock, venom reactions, etc.), it is known that these same anaphylatoxins can be liberated without any known antigen-antibody reaction being concerned. Feldberg, viewing the composition of cell boundaries as that of a mixed lipoprotein film, has pointed out how either a protease or lysolecithin could act as the intermediary histamine-releasing factor. That either or both are concerned seems highly probable, but how they in turn are provided the opportunity to act is not known.

As to the allergic reactions of man, our information is not so extensive or so definitive. Since reactions which are virtually identical with anaphylaxis, both in symptoms and pathogenesis, can be induced in the experimental animal without the participation of an antigen-antibody reaction, the inability to incriminate an antigen-antibody reaction in human allergies does not necessarily set them apart as fundamentally different reactions. It has been shown, however, in a great many cases of varied types of allergy that some sort of humoral agent was present in the blood of the allergic patient which by suitable injection could convey a temporary passive sensitization to the skin of a normal individual. For various reasons of apparent dissimilarity to the precipitins of anaphylactic animals, these humoral agents have been called reagins. Recently, however, using improved methods, Cannon has been able to show precipitins in the blood of patients who were allergic to various substances. It seems highly probable that many reports indicating the absence of antibodies in the human case represent failures to find them rather than conclusively negative evidence. That the symptoms of allergic reactions, which display such manifold patterns as those of hay fever, asthma, angioneurotic edema, etc., are in turn produced by cellular by-products such as histamine, heparin, etc., can neither be categorically affirmed nor denied. It can be stated that these diverse reactions could largely be accounted for by these agents and that there is a considerable amount of circumstantial evidence to substantiate this. There is in addition a certain amount of direct evidence that merits mention. Randolph and Rackemann have observed a significant increase in the blood histamine of patients during paroxysms of asthma. Katz has shown that the application of the specific allergen to the skin of ragweed-sensitive patients leads to a liberation of histamine which parallels the severity of the local reaction. That the demonstration of a relationship between the various symptoms and the agents which may mediate them is not so simple in the human case is readily understandable. In the first place, in the animal experiment we can deal with a reaction which is of explosive violence and perhaps fatal in its termination. In the second place, it must be considered that many, if not most, of the allergic reactions of man are confined to restricted tissues in which the reaction may occur and direct access to the detection of mediating agents is difficult and inconclusive. In summary, it seems warranted to accept Zinsser's view that "all forms of hypersensitiveness represent essentially one and the same basic phenomenon when analyzed with understanding of superficial variables" and to deny that man "possesses a sort of immaculately conceived hypersensitiveness that, like his soul, distinguishes him from the animal kingdom."

References

Abell and Schenk, *J. Immunol.*, **34**, 195 (1938).
Anderson and Schultz, *Proc. Soc. Exp. Biol. Med.*, **7**, 32 (1909, 1910).
Arthus, *Arch. internat. de Physiol.*, **7**, 471 (1908, 1909).
Auer and Lewis, *J. Am. Med. Assn.*, **53**, 458 (1909).
Biedl and Kraus, *Wien. Klin. Wchnschr.*, **22**, 363 (1909).
Cannon and Marshall, *J. Immunol.*, **38**, 365 (1940).
Coca, "Hypersensitiveness," in Tice's "Practice of Medicine," 1920.
Dale, *J. Pharmacol. Exp. Therap.*, **4**, 167 (1912, 1913).
Doerr, Kolle and Wassermann, "Handbuch," 2nd Ed., Vol. 2.
Feldberg, *Ann. Rev. Physiol.*, **3**, 671 (1941).
Gilbert, *J. Pharmacol. Exp. Therap.*, **62**, 228 (1938).
Katz, *Proc. Soc. Exp. Biol. Med.*, **49**, 272 (1942).
Kellaway and Trethewie, *Quart. J. Exp. Physiol. Cognate Med. Sci.*, **30**, 121 (1940).
Loewit, *Arch. exper. Path. Pharmakol.*, **68**, 83 (1912).
Otto, "Leuthold Gedenkschrift," Vol. **1**, 1905.
Randolph and Rackemann, *J. Allergy*, **12**, 450 (1941).
Richet, *Compt. Rend. Soc. Biol.*, p. 170 (1902).
Rocha e Silva, *J. Immunol.*, **38**, 333 (1940).
Rosenau and Anderson. U. S. Pub. Health and Hyg. Lab. Bull. **29**, 1906; **30**, 1906; **36**, 1907. *J. Med. Res.*, **15** (1906); **16** (1907).
Simonds, *J. Am. Med. Assn.*, **73**, 1437 (1919).
Von Pirquet, Schick, "Die Serumkrankheit," 1906.
Weil, *J. Immunol.*, **2**, 525 (1917).

Homeostasis: The Maintenance of Steady States in the Organism

WALTER B. CANNON

Professor, Harvard Medical School

The extreme instability of parts of the body structure in the higher vertebrate, their readiness for disturbance when only slight external forces are applied, and their rapid deterioration in the absence of favorable conditions, render the persistence of a normal state and of normal activity through many decades an extraordinarily remarkable phenomenon.* Its extraordinary character is emphasized by the fact that the organism is an open system engaging in free exchange of material with the outer world, and by the further fact that bodily structure itself is not permanent but is being continuously broken down by activity and continuously restored by the process of repair.

The ability of organisms to maintain a fairly constant state has long impressed biologists. The ancient phrase *"vis medicatrix naturae"* recognized factors which repair the normal state when that has been damaged. An important contribution toward understanding the phenomenon of organic stability was offered by Claude Bernard about ninety years ago. He pointed out that metazoa have two environments: an external environment as have other objects, and also an internal environment. This internal environment is made of the fluids which bathe the living parts of the body: tissue fluids, lymph, and especially blood. Bernard called this the *milieu interne*. It may be regarded as the "fluid matrix" of the organism (cf. Cannon, 1929, 1939a). In so far as the fluid matrix is kept constant, organisms are independent of limiting changes which might be imposed by the external surroundings or by the disturbances produced by bodily activity. The freedom thus assured is a privilege of the higher vertebrates, birds and mammals. It is a freedom which has been gradually evolved. For example, the frog, in which control of body water and temperature is lacking, is restricted in the summer to the vicinity of a moist region,

* The relatively slow changes due to ageing are considered in a paper by A. J. Carlson in this volume. J. A.

and in the winter must sink to the bottom of a frigid pool. Only when spring returns is he freed from the paralyzing effects of icy water. Birds, on the contrary, with their feathery covering, may spend the cold winter months in vigorous activity because of special devices for keeping warm.

The constancy of the fluid matrix is controlled by the organism itself. In that control many complex physiological processes may be involved. The term "equilibrium," which is technically applied to simple physical and chemical systems, does not imply the complicated arrangements in the body for preserving uniformity. Consequently, the relatively stable state maintained in the fluid matrix has been given the special name "homeostasis" (Cannon, 1929).

Essential Role of the Fluid Matrix

When there is danger that the blood may be lost from the body by injury to blood vessels, the blood coagulates at the point of injury. A clot is attached to the edges of the broken surface and thus, sealing the opening, it stops the blood loss. By physiological reactions within the organism, excessive bleeding is accompanied by an increased speed of the clotting process. In an experimental hemorrhage, loss of 13 per cent of the estimated blood volume reduced the clotting time from 7 to 3 minutes. Removal of 10 per cent more reduced the time from 3 minutes to 1 minute (Gray and Lunt, 1914). For later interest it may be stated that more rapid coagulation of blood can be induced by injection of adrenalin and that a large hemorrhage is associated with extra activity of the sympathetic nervous control of the adrenal medulla and consequent discharge of adrenalin into the blood stream.

When muscles become active many capillaries, closed during muscular rest, are opened to such a degree that they may appear to be from 40 to 100 times as numerous in the contracting as in the relaxed muscle; the blood flow may be 5 times as great through the active muscle as through the inactive, thereby supplying extra oxygen and energy-yielding material for the vigorous processes which are taking place. Also the greater volume flow of blood carries away extra waste resulting from these processes. In time of need, therefore, adjustments provide for local homeostasis, *i.e.*, for local preservation of uniform conditions (Cannon, 1936b).

Furthermore, there are special provisions for maintenance of steady states in essential and peculiarly sensitive structures, notably the delicate neurons of the cerebral cortex and the muscle of the continuously active heart. Unless a sufficiently high arterial pressure is present, the delivery of oxygen to these organs will be inadequate. If there is a loss of 15 to 20 per cent of the estimated blood volume, a sharp fall of blood pressure ensues. Only a few minutes elapse, however, before the pressure is restored to approximately the former height, thus assuring a proper blood supply to the brain and to the heart. The restored blood pressure results from reflexes from the vascular system itself, which so operate that the blood vessels generally throughout the body, except in the heart and the brain, are contracted to a degree which fits the capacity of the vascular system to the reduced volume of blood and thereby maintains in these essential organs an adequate oxygen delivery. There is in this preferential treatment a lessened delivery to other organs in which the blood vessels are constricted. These organs, however, are more hardy and can endure the condition for a relatively long period (Cannon, 1923).

Associated with a loss of the fluid matrix by hemorrhage is another compensatory device, the contraction of the spleen, an organ in which corpuscles are stored in concentration. By splenic contraction, the stored corpuscles are discharged into the blood stream and become useful at once as oxygen carriers (Barcroft, 1925).

Types of Homeostatic Regulation

Two general types of homeostatic regulation can be distinguished, dependent upon whether the steady state involves supplies or processes.

Homeostasis by the Regulation of Supplies

Homeostasis of supplies is provided for by *storage,* as a means of adjustment between occasional abundance and later privation and need, and also by *overflow* or discharge from the body when there is excess. Two types of storage can be distinguished: a temporary flooding of the interstices of loose tissue by material in solution, *i.e.,* storage by inundation; and an inclusion of material in cells or in other relatively fixed structures, storage by segregation.

Storage by Inundation. Somewhat as a stream passes through a bog or swamp, the circulating blood in the capillaries sweeps by the more stagnant extravascular fluids. Just as in time of flood swamp water rises, so in time of abundance there is an outflow from the capillaries into the loose connective tissue under the skin and around and between muscles and muscle bundles. In this meshwork and bound by it in some manner, water and dissolved substances appear to be held.

Water. Water storage in the extravascular regions passes out from them as it is needed, *e.g.,* a dog deprived of water for three days showed no change in the freezing point of the blood (Mayer, 1900). Clearly, constancy was preserved in spite of continuous loss of water through the lungs and the kidneys, a constancy due to seepage into the blood from the inundated reservoirs. Evidence indicates that most of the water which leaves the reservoirs, *e.g.,* after a hemorrhage, comes mainly from the muscles and the skin (Skelton, 1927).

Probably storage by inundation, and release from storage, are managed through a balance between filtration pressure through the capillary walls and osmotic pressure of the proteins of the blood plasma, working in opposition. Doubtless there are other factors. With the taking of sodium chloride, there is a markedly increased retention of water in the body. Probably other electrolytes play a role. That the thyroid gland is influential is indicated by the great increase of protein in the plasma and of albumin in the tissues in conditions of low thyroid activity, and the disappearance of these conditions, together with a large release of water and sodium chloride when thyroxine is administered. How these various factors coöperate when water and sodium chloride are needed in the circulation, as after hemorrhage, is not clear.

The importance of retention of water in the body to preserve homeostasis is indicated by physiological provision for saving it when there is dearth. It is known, for example, that an extract of the posterior lobe of the pituitary gland will check the water loss through the kidneys. If the water load in the body is reduced, the amount of such an extract administered intravenously may be considerably less as a check on urine production than if the water load is abundant. It would appear, therefore, that a large water load inhibits the release of the antidiuretic substance from the pituitary gland; consequently, the amount of extract required to check the flow of urine is greater than when the load is small. With a small load the requirement of the extract would be less because a protective antidiuretic substance is secreted from the pituitary source (cf. Pickford, 1936).

Sodium Chloride. Concentration of sodium chloride in the blood plasma and lymph is remarkably steady. Among the electrolytes it is the main agency for maintenance of a uniform osmotic pressure in the circulating fluids. It is stored in times of excess, as shown by its retention in the body. If concentrated sodium chloride is taken by mouth, only a small part of the amount ingested may appear in the urine; most of it may be stored in the body; and even thereafter if enough water is drunk to produce a diuresis the urine has a low salt content, *i.e.,* the salt is not readily released from its storage place. With a fairly constant chloride intake, a loss of chloride through the skin by sweating is accompanied by a great reduction of chloride output in the urine—a condition which continues, although thereafter a diet rich in salt is taken; in such circumstances a compensatory retention of as much as 14 grams of sodium chloride has been observed.

When a search is made for the location of the sodium chloride reserve, the highest percentage is found in the skin; indeed, on a chloride-rich diet 33 per cent of the chloride of the body may be in the skin. If animals are fed on a chloride-poor diet, between ⅓ and ⅔ of the chloride content of the body is lost, and of this amount between 60 and 90 per cent comes from the skin (the skin constitutes only about 16 per cent of the total body weight (Padtberg, 1910)). Whether sodium chloride is held in the skin by adsorption on surfaces in areolar tissue or by solution in interstitial fluid of the areolar spaces has not yet been definitely determined. Probably it is osmotically inactive. The evidence appears to be clear, however, that sodium chloride and water are closely related in their storage in the organism.

Glucose. Like excessive sodium chloride, excessive blood sugar finds its first temporary depository in the skin. If sugar or other readily digested carbohydrate is taken in abundance, the concentration of glucose in the blood commonly rises from about 100 to 170 mg per cent. After the digestive process has subsided the glycemic percentage likewise subsides. During the period of high concentration of glucose in the blood there is also a high concentration in the skin (Folin, Trimble and Newman, 1927). The drop in the glycemic level occurs as the circulating glucose is utilized or placed in more permanent storage in the liver and in muscle cells. At the same time with the drop in the glycemic level, the more concentrated glucose which has overflowed into the subcutaneous net of areolar tissue and possibly into other regions where this tissue is abundant, gradually seeps back into the blood again and then follows the usual courses of blood glucose into use or into fixed reserves.

Storage by Segregation. This mode of storage, commonly within cells, is stable and lasting. It differs from storage by inundation in being subject to more complicated control and in involving, as a rule, changes of physical state or of molecular configuration. In some instances it is known to be subject to nervous or neuro-endocrine government. There are, however, large gaps in the knowledge of ways in which substances are segregated in times of abundance and released from segregation in times of bodily need.

Carbohydrate. The best example of homeostasis by means of segregation is offered by arrangements for storage and release of carbohydrate. Plentiful intake of carbohydrate food results in large reserves of glycogen in the liver; prolonged muscular work draws upon these reserves until they may be largely depleted; and yet the rate of release is so nicely regulated that the blood sugar is maintained at a degree of concentration which results neither in sugar loss through the kidneys nor in disturbances due to hypoglycemia.

The control of the glucose concentration of the blood is complex. Various glands of internal secretion are involved: the anterior lobe of the pituitary, the thyroid, the adrenal cortex, the adrenal medulla and the islet cells of the pancreas. It appears that the endocrine agencies most directly concerned with maintenance of the normal glycemic level are the pancreatic islets and the adrenal medulla.

When insulin is injected in sufficient proportions, it lowers the glycemic concentration of the blood. As the concentration falls to about two-thirds of the usual degree, a "hypoglycemic reaction" occurs, characterized by pallor, rapid pulse, dilated pupils and profuse sweating—all signs of sympathetic activity. This hypoglycemic state calls into action the nervous organization which discharges adrenaline from the adrenal medulla into the blood stream. Adrenaline, by acting on the liver cells, mobilizes the glycogen stored within them and thus increases the sugar in the blood. It is clear that reduction of the glycemic percentage below a critical level calls forth an agency which acts to correct the condition (Cannon, McIver and Bliss, 1924).

If the sympathico-adrenal system is ineffective in checking the fall of blood sugar, convulsions are likely to occur when the concentration is reduced to about

half the normal concentration. Each convulsion is associated with a maximal display of sympathico-adrenal activity. If the liver is well supplied with glycogen, such extreme activity can restore the blood sugar to the normal level and thus abolish the condition which occasioned the convulsive attacks. That it is secreted adrenaline which protects the organism against coma and death (if the glucose concentration falls to a very low level), has been shown by evidence that adrenaline and not nervous control of the liver cells is the effective factor (see Cannon, 1940).

Agencies opposed to the sympathico-adrenal operations are those which operate with excess of glucose is ingested. The blood sugar then rises to a level close to that at which it escapes through the kidneys, but ordinarily it does not surpass that level. The excess sugar, apart from that set aside by inundation, is either stored in the liver or in the muscles, is converted into fat, or is promptly utilized. Evidence indicates that this process of storage is dependent on the insulin secreted by the pancreatic islets. The items of this evidence are as follows: (1) If the islet cells are removed hyperglycemia promptly occurs, and with it a great reduction of the hepatic glycogen reserves; (2) the administration of insulin to sugar-fed depancreatized dogs reduces the glycemic concentration to the normal range and causes glycogen to accumulate again in the liver; (3) in the absence of the liver, injected insulin augments the glycogen deposit in muscles, especially when extra glucose is provided; (4) when only a remnant of the pancreas is present and carbohydrate is fed, the islet cells degenerate and show signs of overwork; it would appear, therefore, that hyperglycemia stimulates the cells to secrete.

Although the internal secretion of the pancreas can operate effectively when wholly separated from the nervous system, there is testimony that the vagus nerves can in certain special circumstances cause an augmented discharge of insulin. Much as the sympathico-adrenal system serves to bring forth sugar from the liver rapidly in emergencies, so does the vago-insular apparatus seem to serve for storing reserves when there is excess glucose in the blood. Neither the sympathetic nerves nor the vagi are required, however, to maintain the usual glycemic concentration, for in the absence of these nerves that concentration is preserved. By determining simultaneously the rate of blood flow through the liver and the glucose content of the inflowing and outflowing blood, it has been shown that during control periods the liver secretes glucose into the blood stream, but that when glucose is abundantly supplied secretion ceases and the sugar is retained (Soskin, Essex, Herrick and Mann, 1938). It is very possible that reversible enzyme action provides for routine adjustment between plenty and dearth, and nervous control serves for conditioning or emergencies.

The manner in which the anterior lobe of the pituitary gland is involved in the homeostasis of blood sugar is not well demonstrated. If a diabetic hyperglycemia has been produced by removal of the pancreas, a subsequent removal of the pituitary lobe reduces the hyperglycemia (Houssay and Biasotti, 1931). Similar lowering of the glucose concentration of the blood occurs after removal of the adrenal cortex. Injection of an extract of the anterior lobe of the pituitary produces hyperglycemia, and if continued can cause diabetes. In this complex of endocrine influences the role played by each factor is still obscure. The possibility of disorder due to a variety of agencies which can affect the glycemic concentration, renders limitation of the range of that concentration in normal conditions all the more admirable.

Protein. The constancy of body structure is perhaps the most widely manifested evidence of the homeostasis of protein. That constancy includes also the blood, and since the steady states of the blood are of prime concern particular attention is now paid to that phase of the subject. Because plasma proteins exert osmotic pressure and ordinarily do not escape through capillary walls, they prevent the salt solution of the plasma from passing freely into perivascular spaces or escaping from the

body through the renal glomeruli. When, after an experimental hemorrhage, the corpuscles are returned to the body suspended in salt solution (the process of plasmapheresis), the immediate effect is a reduction of plasma proteins from about 6 to 2 per cent. Within fifteen minutes, however, there is a prompt rise in the concentration, followed by a gradual restoration thereafter until at the end of 24 hours there has been a 40 per cent recovery. Full recovery occurs in 2 to 7 days. Possibly the prompt rise is relative, due to an escape of salt solution from the blood vessels into perivascular spaces or through the kidneys. The slower recovery seems certainly to be dependent on the liver, for if the liver has been poisoned, restoration of the plasma protein is delayed; or if the return of blood from the intestines is shunted directly into the vena cava, thus sidetracking the liver, there may be no restoration for the first three days; or if the liver is unable to act, fibrinogen, which is usually completely restored within twenty-four hours, is not thus restored (cf. Whipple, 1938).

That the liver is important for homeostasis of plasma proteins raises the question whether protein is stored in that organ. Histological studies have shown that when animals are fed abundant protein food there appear in the hepatic cells fine droplets or masses reacting as protein material does, disappearing during a fast and reappearing when proteins or amino acids are fed. This histological observation is confirmed by biochemical analyses, which have shown that in well-fed animals the total nitrogen of the liver in relation to that of the rest of the body is two to three times as great as in fasting animals. It has also been found that in animals maintained on a high protein diet, various globulin and albumin fractions of the liver protein are increased 50 to 60 per cent over controls (Luco, 1936).

Unfortunately the modes of storage and release of protein are almost wholly unknown. There is evidence that protein is stored in other places than the liver and also that the thyroid gland exerts an important control. Thus the edema of the myxedema resulting from hypothyroidism may be an exaggeration of the deposit of protein in and beneath the skin. The effect of thyroxine in reducing the increased proteins of the tissues in cases of myxedema supports the view that the thyroid gland is somehow associated with protein regulation and metabolism. The increase of fibrinogen in the blood when adrenaline is injected, indicates that the sympathicoadrenal mechanism may affect storage of that protein, for it is known to come from the liver (Riecker and Winters, 1931).

Fat. Although the concentration of fat, cholesterol and lecithin in the blood is fairly constant in the same species of animals, it may differ greatly in different species. To be sure, the ingestion of fat produces an alimentary lipemia, in which the fat content of the blood may rise as high as 2 per cent. A relatively large increase in the concentration of fat in the blood appears to be without serious consequences. In pathological states, as for example in diabetes, the lipemic percentage may rise to 10, 15 or even to 20 per cent without producing obvious symptoms. On the other hand, normal blood fat is remarkably persistent. Whether carbohydrate and protein alone are fed for considerable periods or there is an experience of fasting, the lipemic level may undergo no downward change. Indeed, in a short fast the level may actually rise before it undergoes a gradual descent.

The relative constancy of the lipemic level for many days in spite of complete starvation implies a governing agency which brings fat from storage into the blood stream. Fat is stored in the cells of the liver if carbohydrate is not fed; it is also stored in adipose tissue under the skin, beneath serous coats (*e.g.*, around the kidneys), in the omentum and between and in the muscle fibers. Excessive intake of any of the foodstuffs may lead to excessive fat storage. In hypothyroidism and also after lesions in the neighborhood of the pituitary gland there is an abnormal adiposity. The manner in which these two glands may affect control of the obese state is unknown.

If the regulation of fat storage is obscure, the regulation of its release is even more so. Recent studies have shown that the stores are not fixed but are being continuously broken down and renewed. Certainly when fat is needed to maintain the energies of the body it is removed from adipose tissue until the fat cells are practically empty. It is possible that a reversible reaction mediated by tissue lipase is an important factor in maintaining homeostasis of the lipemic level—the enzyme favoring storage when the level is raised and favoring release when the level falls.

Calcium. Calcium has many diverse functions. It is used for the growth of the skeleton and the teeth, for the repair of broken bone, for maintaining proper conditions of irritability of nervous and muscular tissue, for the coagulation of blood, and for the production of milk. Like sugar and protein and fat, calcium may be in great demand on exceptional occasions, some of which have just been indicated. In such circumstances, however, the amount in the blood must not be reduced, for serious consequences ensue. Normally the calcemic concentration is held at approximately 10 mgm per cent; a variation to half that percentage may be accompanied by convulsions, and a variation to twice that percentage may be accompanied by greatly increased viscosity of the blood. Obviously, homeostasis of blood calcium is of capital importance. As with other substances useful in the organism, the homeostasis of calcium is made possible by storage in times of abundance and release from storage in times of need. Calcium is stored in the trabeculae of long bones; these trabeculae may be made to disappear by a persistent diet deficient in calcium and by growth. They are readily restored, however, by feeding a calcium-rich diet (Bauer, Aub and Albright, 1929).

How the homeostasis of calcium is regulated has not been determined, though the parathyroid glands are surely involved. Thus partial or complete removal of these glands results in a reduced calcium content of the blood, in defective deposit of dentine in growing teeth, and in a failure of proper development of a callus around a bone fracture. Again, a diet poor in calcium induces parathyroid hyperplasia. The stress of pregnancy and lactation does likewise, without, however, reducing the calcium percentage in the blood. Furthermore, diseases which are characterized by defects in the calcification of bone (for example, rickets and osteomalacia) are attended by hypertrophy of the parathyroid glands. Finally, experiment has shown that the implantation of parathyroids in a parathyroidectomized rat restores the power to deposit dentine having a normal calcium content. Unfortunately the manner in which the parathyroid glands control the homeostasis of calcium is still unknown.

The pharmacodynamic action of thyroxine seems to implicate the thyroid as well as the parathyroid glands in calcium metabolism, for administration of that agent greatly increases the loss of calcium from the body by the excretions. There is also evidence that in hyperthyroidism, bones develop a porous condition and that calcium excretion is much augmented.

Homeostasis by Overflow

The use of overflow as a means of checking an upward variation of constituents of the blood has been mentioned. Not only excess glucose but also excess water and excess sodium and potassium and chloride ions are discharged through the kidneys. These are all "threshold substances"—they are resorbed by the kidney tubules to a greater or less degree.

It is interesting to note that these substances are primarily stored by flooding or inundation. When the supplies are adequate, however, the ability of the overflow factor to maintain homeostasis in spite of excessive intake is little short of marvelous. The feat of drinking 5.5 quarts of water in 6 hours—an amount more than 33 per cent greater than the estimated blood volume, passed through the kidneys with such nicety that at no time was the hemoglobin percentage observably reduced—is a

revelation not only of the efficacy of the kidneys as a spillway but also of the provisions in the body for the maintenance of constancy of its fluid matrix (Haldane and Priestley, 1915).

The lungs as well as the kidneys serve for overflow. A slight excess of carbonic acid in the arterial blood is followed by greatly increased pulmonary ventilation. The extra carbon dioxide is thus so promptly and effectively eliminated that the alveolar air is kept normally constant. By this means extra carbon dioxide resulting from bodily activity flows out from the blood over a dam which is set at a fixed level. In consequence, the hydrogen-ion concentration of the blood is, in normal conditions, fairly evenly maintained and the harmful effects of an excessive shift are avoided in either the alkaline or the acid direction.

Homeostasis by Regulating Processes

There are homeostatic conditions which, though involving the use of materials, are so much more dependent on altering the rate of continuous processes that they can reasonably be placed in a separate category. The maintenance of neutrality and the maintenance of uniform temperature (in homeothermic animals) will illustrate the methods.

Maintenance of Neutrality. The hydrogen-ion concentration in the blood is determined by the ratio of H_2CO_3 to $NaHCO_3$. On going to a high altitude the tension of carbonic acid is lessened; thus the ratio is lowered and the pH rises. In these circumstances the blood alkali also is lessened until the normal pH is restored. On returning to sea level the opposite process occurs and continues until the usual relations are restored—due probably to the passage of acid substances from the blood into the tissues or out through the kidneys.

As is well known, accumulation of the acid metabolites, which are continuously being produced in living cells, would interfere with or actually prevent continued functioning. That contingency is avoided by elaborate arrangements. The lactic acid, for example, which is developed in muscular contraction, is in part promptly neutralized; another part is soon oxidized; and the rest may be rebuilt in the liver into neutral glycogen. For continued effectiveness of these methods of disposal an adequate supply of oxygen must be provided.

Muscles and probably other tissues are able to go into "oxygen debt" by action in spite of accumulating lactic acid. That state, however, is characterized by a diminished capacity to do work, by a limitation on the amount of debt allowable, and by a definite necessity of ultimate payment. An oxygen debt can be avoided if the non-volatile lactic acid is burned to volatile carbonic acid; the volatile acid can be carried away and disposed of to an almost unlimited amount with only a slight change in the reaction of the blood. Obviously, therefore, during vigorous muscular work, when the production of lactic acid is great, as much oxygen as possible should be delivered to the active tissues. Probably because air-breathing animals are surrounded by an unlimited supply of oxygen there is practically no store of it in their bodies. The problem, therefore, is that of conveying as much as possible of the boundless external supply to the needy tissues. For that purpose the continuous process of respiration and the circulation of the blood must be greatly accelerated. The adjustments required to rid the body of the volatile acid waste are, by good fortune, precisely those required to bring to the tissues the oxygen which serves to make the acid volatile and thereby readily discharged.

In vigorous exercise the pulmonary ventilation may be increased from 10- to 13-fold, due to the stimulatory effects of acid on the respiratory mechanisms. Also the return of blood to the heart per minute is greatly augmented by contraction of the abdominal vessels, by pressures rhythmically exerted by the laboring muscles on the capillaries and valves in the veins within them, and by pumping action of the diaphragm. The heart, more amply charged with blood, puts forth a greater volume

per beat, and because of adjustments through the nervous system it may beat more than twice as fast as it does when at rest.

The larger cardiac output of blood per minute meets a constricted splanchnic area, and the result is a marked rise in arterial pressure. In the laboring muscles, arterioles are dilated and the capillaries, many of which have been closed, are abundantly opened; through these dilated and more numerous channels the high head of arterial pressure drives a rushing stream of blood. The total circulation rate may be augmented as much as 4-fold. Not only are the respiratory gases conveyed more effectively by rapid movement of the red blood corpuscles, but the conveyance of these gases is also increased by a release of extra corpuscles from storage in the spleen (Barcroft, 1926). Contraction of the spleen, like constriction of the splanchnic area and acceleration of the heart, results from operation of the sympathetic system. Furthermore, in the laboring muscles where acid is being produced and where there is special need for oxygen, the excess carbon dioxide, which results from the burning of lactic acid, itself serves to facilitate the unloading of oxygen from the red corpuscles and its carriage away to the lungs. In these ways the local blood flow in a vigorously active region may be increased as much as 9 times and the oxygen delivery may be increased as much as 18 times its amount during rest. By these admirable adjustments, lactic acid, which can be produced in a short time by muscular work to a degree which could neutralize the buffer substances in the blood, and which could induce the total cessation of muscular work, is not allowed to accumulate—indeed, the reaction of the arterial blood is altered to only a minor degree.

Maintenance of Uniform Temperature. Uniform temperature is obviously important for providing conditions favorable for a fairly constant rate of the chemical changes in the body. In normal conditions the diurnal range of temperature variations in human beings is scarcely more than 1°C. If the temperature tends to rise, relaxation of peripheral vessels exposes warm blood to cooler surroundings; and if that is ineffective, sweating occurs and the blood flowing through the skin, cooled by evaporation, loses extra heat. Panting performs a service similar to sweating, especially in animals such as the dog that are not well provided with sweat glands.

If the temperature tends to fall there is a constriction of peripheral vessels, and in animals provided with fur or feathers, an erection of these appendages of the skin which enmesh around the body a layer of poorly conducting air. When these means of conserving heat do not check the fall of temperature, adrenalin is set free in the blood stream and exerts a thermogenic influence. And if the heat thus produced does not suffice, shivering is automatically induced as a final protection against a temperature drop. These arrangements for maintaining a homeothermic state involve only acceleration or retardation of processes which are constantly going on.

Recent evidence has shown that in the subthalamus there are two thermostatic regulators; one of them is brought into operation when the temperature tends to rise, the other when the temperature tends to fall (Ranson, 1936). The regulation can be influenced directly by the temperature of the blood and also reflexly. The noteworthy features of the complex of arrangements are the varieties of bodily devices for the maintenance of thermo-homeostasis, their appearance in a sequence of defenses against change, and the intimate involvement of the sympathetic system in the conservation, production and dissipation of heat.

Role of the Autonomic Nervous System in Homeostasis

For the most part, the regulation of steady states in the blood is not under voluntary government. Commonly the autonomic system, or that system in coöperation with endocrine organs, is called into action. Illustrations of these facts are found in the vago-insular and the sympathico-adrenal influences on the glycemic level, the vagal and sympathetic effects on the heart rate and the sympathetic effects on blood

vessels during vigorous muscular effort, and the sympathico-adrenal function in accelerating heat production when the body temperature tends to fall (cf. Cannon, 1939a).

The nervous system produces its effects in two relations. It operates on the muscles which move the bony levers and thereby changes the external environment or the position of the organism in that environment, as in laboring, running or fighting. These may appropriately be regarded as exteroeffective activities.

The other relation of the nervous system is toward the internal environment, *i.e.*, toward inducing changes in the heart, in smooth muscles and glands such as to preserve a fitness of the internal environment for continued exteroeffective action. Inactivity of the exteroeffective system establishes a basal state for the organism because minimal functioning of that system is accompanied by minimal functioning also of the interoeffective system. When exteroeffective activities produce changes in the internal environment, *e.g.*, by utilizing blood sugar or by discharging into the blood acid waste and extra heat, the interoeffective system is set in action to correct the alterations; when they are corrected its activities subside. Thus the two parts of the nervous system—the exteroeffective and the interoeffective—are engaged in close coöperation.

Although the cranial division of the autonomic system may be regarded as concerned with the upbuilding of reserves, *e.g.*, in promoting gastrointestinal movements, in secretion of digestive juices and thereby in storage of provisions and materials for bodily use, it is the sympathetic division which is the main agency for mobilizing bodily forces for action or correcting the alterations produced by varying internal or external conditions. By mobilizing reserves and by altering the rate of continuous processes, this division works to keep constant the fluid matrix of the body, and may properly be regarded as the special and immediate agent for control of homeostasis.

References

Barcroft, J., *Lancet*, **1**, 319 (1925).
—, *Ergebn. d. Physiol.*, **25**, 818 (1926).
Bauer, W., Aub, J. C., and Albright, F., *J. Exper. Med.*, **49**, 145 (1929).
Bernard, C., "Les Leçons sur les Phénomènes de la Vie," Vol. **1**, p. 113, Paris, J. B. Bailliere et fils, 1878.
Cannon, W. B., "Traumatic Shock," chaps. 7-10, New York, D. Appleton and Co., 1923.
—, *Physiol. Rev.*, **9**, 399 (1929).
—, "The Wisdom of the Body," 2nd ed. chap. 1, W. W. Norton and Co., New York, 1939a.
—, *J. Mt. Sinai Hosp.*, **5**, 587 (1939b).
—, *Am. J. Med. Sci.*, **199**, 98 (1940).
—, McIver, M. A., and Bliss, S. W., *Am. J. Physiol.*, **59**, 46 (1924).
Folin, O., Trimble, H. C., and Newman, L. H., *J. Biol. Chem.*, **75**, 263 (1910).
Gray, H., and Lunt, L., *Am. J. Physiol.*, **34**, 332 (1914).
Haldane, J. S., and Priestley, J. G., *J. Physiol.*, **50**, 296 (1915).
Houssay, B. A., and Biasotti, A., *Endocrinol.*, **15**, 511 (1931).
Luco, J. M., *J. Biol. Chem.*, **115**, 491 (1936).
Mayer, A., *Compt. rend. Soc. de Biol.*, **52**, 154, 389, 522 (1900).
Padtberg, J. H., *Arch. exper. Pathol. Pharmakol.*, **63**, 78 (1910).
Pickford, M. J., *J. Physiol.*, **87**, 291 (1936).
Ranson, S. W., and Teague, R. S., *Am. J. Physiol.*, **117**, 562 (1936).
Riecker, H. H., and Winters, M., *Proc. Soc. Exper. Biol. and Med.*, **28**, 671 (1931).
Skelton, H. P., *Arch. Int. Med.*, **40**, 140 (1927).
Soskin, S., Essex, H. E., Herrick, J. F., and Mann, F. C., *Am. J. Physiol.*, **124**, 558 (1938).
Whipple, G. H., *Am. J. Med. Sci.*, **196**, 609 (1938).

The Causes and Nature of Cancer *

LEO LOEB
Washington University, St. Louis, Mo.

Introduction

Tumors in general, and malignant tumors in particular, were first studied by pathologists and clinicians as the cause of disease and death in man. A tumor, in the specific sense in which this term is now used, means an abnormal, localized condition in the body, characterized by an increase in the number of cells or in intercellular substances, or both; in some cases, the number of cells, and in others, the quantity of intercellular substances preponderating. Tumors may be classified as benign or malignant. Benign tumors have slow growth, which may come to a standstill after some time, and they remain localized; if they produce injurious effects, these are largely due to the pressure they exert on neighboring tissues or organs. Malignant tumors, which are called cancers, grow more rapidly than benign tumors. Their growth consists in a multiplication of cells which are essentially of the same type as those of the organ or tissue in which they originate, such as the epidermal cells of the skin, the cells of the intestinal surface epithelium or glands, the specific cells of the liver or thyroid acini, and so on; or they may originate in the connective tissue cells of the stroma of these and other organs, or in the cells of bone and cartilage. Also, the endothelia of the blood vessels and components of the nervous system may give origin to both benign and malignant tumors. The growth of the latter is usually so rapid that the conditions are unfavorable for the production of specific intercellular structures.

Cancers differ from benign tumors not only in their more rapid growth, but also in the manifestation of structural abnormalities. The relation of the proliferating cells to stroma and blood vessels may be altered as the result of the more rapid growth; thus, instead of a single layer of glandular tubular or acinar epithelium, the proliferating epithelial cells may produce strands of cells, which show an irregular arrangement. There is usually associated with the increase in cell multiplication an increase in ameboid motility of these cells. As the result of both the increased proliferation and movements, cancerous cells infiltrate the neighboring tissues and destroy them. They may penetrate into blood vessels as well, especially into veins and into lymph vessels, and be carried with the blood or lymph stream to distant organs. If in these they are held back by mechanical factors, they may settle, multiply, and give origin to secondary cancers, which closely resemble the primary tumors and are called metastases. Occasionally there may be found transitional stages between benign and malignant tumors, as, for instance, tumors which have the main characteristics of benign growths but which give rise to metastases.

All kinds of cells can give origin to tumors, including cancers, provided they have the power to propagate. The large majority of the cells composing the various tissues and organs possess this power, but certain cells which represent endstages of differentiation have lost it, such as the polymorphonuclear leucocytes or the ganglia cells of the adult organism, although the precursor cells of the leucocytes, as well as the embryonal ganglia cells, still have the ability to proliferate and to produce

* Written with the aid of a grant from The International Cancer Research Foundation.

tumors. If tumors found in postembryonal life consist of cells which proliferate only during the embryonal period, it may be assumed that such tumors have developed from embryonal tissues.

Both benign tumors and cancers are classified according to the organs and tissues from which they originate. The principal types of cancers are carcinomas, derived from epithelial tissues, and sarcomas, the offspring of connective tissue cells. Carcinomas and sarcomas, as well as benign tumors, are further subdivided in accordance with the special tissues and organs from which they are derived.

The large majority of cancers develop during later adult life. As a rule, they increase in frequency with increasing age, but malignant tumors of certain organs have a maximum incidence at an age which is influenced by the activity curve of the given organ in its interaction with other organs. In certain organs the incidence may, therefore, decrease again in old age. Other cancers, whose origin can be traced to abnormalities developing during embryonal life, may start during the intrauterine growth phase, although they may become manifest only in postuterine life. Likewise, benign tumors may begin to develop before birth.

Clinicians and pathologists have studied the origin and growth of various types of tumors, their structural behavior, and their effect on the health and metabolism of the patients, and pathologists in particular have described the characteristics of the cells and intercellular substances of a great many of such tumors in order to trace their origin from certain tissues.

This constitutes the clinical and descriptive morphological phase in the study of tumors. While this phase continues today, there have been added to it during the last fifty years experimental investigations into the biology and biochemistry of tumors. With the introduction of the experimental method, new problems arose, in which cancerous growth was no longer considered of clinical concern only, but also of general biological significance and interest; compared with other types of growth, such as embryonal, regenerative and hormonal-correlative growth, this study suggested the analysis of the relations of various chemical and physical agents which are growth stimulators to the substratum on which they act.

Transplantation of Tumors

The earliest experimental procedure which greatly extended our knowledge of the cancerous process, was the transplantation of tumors in animals, and the first successful transplantations were carried out by Hanau, Morau, Velich, Eiselsberg and Firket. These investigators used for this purpose carcinoma in the rat and mouse, as well as sarcoma in the rat. These experiments established the fact that certain tumors can be transplanted to other animals of the same species, at least for a limited number of generations. In the beginning of this century, consecutive transplantations of sarcoma of the thyroid gland of rats by the writer (1901) and of a mammary gland adenocarcinoma of the mouse by Jensen (1902), which extended through many more generations, were first used as a method for the analysis of the characteristics of the tumor cells and of the interaction between tumors and hosts; there was thus initiated the experimental study of the biology and causes of cancer, which has continued with increasing intensity until the present day.

It is possible to excise and transplant normal adult tissues to other positions in the same organism or to other organisms belonging to the same species. However, the transplantability of normal tissues is very restricted, being limited by the genetic relationship of the donor of the transplant and the new host and of their individuality differentials. The results of transplantation of cancers depend likewise upon the relationship between the individuality differentials—and organismal differentials in general—of host and transplant, but transplantation of cancer differs from that of normal tissues in several important respects. While many normal tissues as well as almost all cancers can be transplanted into individuals in which the individuality

differentials have become very similar as the result of long-continued close inbreeding or of certain other conditions, the growth of grafted normal tissues is very limited in contrast to that of grafted cancers, which latter, as a rule, continue to grow after transplantation at least as actively as the original cancers; they can therefore be subdivided into small pieces and transferred from generation to generation into many animals in an unending series. Furthermore, there are certain cancers which are much more independent of the nature of the individuality differentials than others; they can be transplanted even into individuals of the same species in which the individuality differentials between host and transplant differ very markedly. Great differences exist in this respect between different cancers; some can be transplanted only into a limited group of different strains, others into all or almost all strains of the same species. This remarkable degree of transplantability is partly due to the strong growth momentum of the tumor transplant, which is able to overcome obstacles to which normal tissues would succumb. This growth momentum in the case of many cancers increases in the course of the first serial transplantations, or sometimes during the later transplantations, and with this increase in growth momentum there is associated a widening in the range of transplantability. This increase, observed so frequently in the course of the first transplantations, occurs also, as a rule, if serial transplantations are carried out, for instance, with a mammary gland carcinoma originating spontaneously in a closely inbred strain, into mice belonging to this strain. Such an effect is not due to a chance selection of hosts among non-homozygous animals. In a few tumors it has been possible to carry out even transplantations serially to different though relatively nearly related species, but this succeeds usually only if the tumor is retransplanted to a new host, before immunity against the graft has fully developed in the first host. Putnoky found that a certain line of adenocarcinoma of the mammary gland of mice could be readily transplanted into successive generations of a Hungarian strain of rats. This mouse tumor possesses a very great growth momentum and this is probably one of the reasons why it succeeded in overcoming the obstacle which a heterogenous, although nearly related, species differential presented. It is probable that, in addition, processes of adaptation may have taken place in this tumor in the course of serial transplantation in the rat, but this latter process did not alter the species differential of the tumor cells, the mouse differential not being changed into that of the rat. Whether these adaptative changes, if present, consist in an increased resistance of the transplanted cells to injurious factors emanating from the host, or in a diminution in strange substances given off by the transplant and consequently to a diminished reaction of the host against the transplant, is not known. On the other hand, transplanted cancers as a rule are less invasive than the original cancers and they can be more readily removed from the host without leaving behind cancerous cells which would give rise to recurrent growths, which latter often develop in the animal in which the tumor originated even after an apparently complete extirpation.

It is thus possible to increase the growth energy of cancerous tissues by means of experimental transplantations.[36] Even in metastases, which represent spontaneous, autogenous transplantations, the growth energy is increased in many cases over that of the original tumor. It is furthermore possible, through mechanical means, to increase the growth energy of a tumor which has lost a part of its growth momentum, as for instance, as a result of bacterial contamination of the grafted tumor. If such a tumor is growing with diminished proliferative activity, it can, in some instances, be made to grow again more actively by excising a piece or by pulling a thread through it. Conversely, the growth energy of a tumor can be diminished experimentally in a graded manner by subjecting the excised cells in vitro to various degrees of heat, or by exposing them to the action of injurious chemical substances previous to transplantation.[36] It is then possible to raise again, by experimental

means, the growth energy which had previously been diminished; this often occurs in the course of further transplantations. In addition, it has been shown by MacDowell in the case of acute leukemia, which is a cancer of certain leucocytic cells, that the immunizing properties of cancer cells may change in the course of transplantation. While these various characteristics have so far been established only in the case of cancer cells, it is not certain that they are peculiar to these and that they are lacking altogether in normal cells; it may well be that they are potentially present also in the latter but cannot become active because of the low growth momentum of these tissues.

There are some locations in the body where the transplanted tumors in some way are more or less protected against the injurious action of heterogenous hosts, which means individuals belonging to different species. Such locations are the brain (Shirai, Murphy), and the anterior chamber of the eye (Smirnova, Greene, Saxton, and others). Tumors from distant species grow also in the allantois of the chick (Murphy). Again, this is not a characteristic of cancers, but it applies also to normal heterogenous tissues, which likewise remain alive in the chick allantois and may be somewhat better preserved in the anterior chamber of the eye than in other locations.

The fact that different types of cancer composed of diverse tissues may be transplanted indefinitely from generation to generation has led to the important conclusion that isolated tissues of mammalian organisms may have the potentiality to immortal life, although the composite organisms undergo ageing processes and invariably die. This conclusion follows, since cancers are merely transformed normal tissues and it is possible experimentally to change a normal into a cancerous tissue.[33, 34, 35, 36, 45]

Benign tumors differ from cancerous tumors in a lesser growth energy. They stand, in this respect, midway between normal tissues and cancers. Furthermore, they lack invasive powers. Correspondingly, their transplantability is intermediate between that of cancers and normal tissues; but there are benign tumors which also can be transplanted indefinitely from generation to generation; however, under these conditions they grow relatively more slowly than cancers, and in contrast to the latter, they may still be accessible to the action of hormones produced by their hosts or introduced into the hosts from the outside. Cancerous tissues as a rule can no longer be influenced by hormones, but some variations may perhaps exist in this respect in the case of certain malignant tumors.

The ability to undergo permanent growth, with increase in mass, after transplantation into other individuals has been made use of in order to recognize whether a normal tissue, subjected to certain cancerigenic stimulations, had become converted into fully cancerous tissue, in which case it would continue to grow in a cancer-like manner in the new host after cessation of the original growth stimuli which had induced the cancerous state. On the other hand, tissues which merely simulated cancerous growth, or which had not yet fully reached the cancerous state, would cease to proliferate in the typical cancerous manner after transplantation, since it is one of the characteristic features of cancer that if this stage has once been attained tissues may die under unfavorable circumstances, but, as far as is known, they do not return to the state of normal tissues. However, as a rule it is not necessary to have recourse to such a method for the diagnosis of malignant tumors, because this can be made by the observation of an irregular structure and mode of growth and of invasive activities.

Cancerous tissues during the process of serial transplantation can be seen to grow as the result of the continued multiplication of their own cells; they do not commonly transmit their cancerous characteristics to the neighboring tissues with which they come in contact. But such a transmission of cancerous properties has been observed in some cases, in which transplanted carcinomatous tissue converted the stroma with which it came in contact into a sarcomatous tumor. In addition,

there is some evidence that a carcinomatous tissue under exceptional conditions may change also adjoining normal epithelium into carcinoma. We must assume that the cancerous tissue exerts this effect on the neighboring cells through some substance which it gives off; but as to the nature of this substance, nothing is known at the present time. In the center of the transplanted, as well as of spontaneous, cancers, the tissue often degenerates and dies, because of the insufficient circulation at a distance from the periphery of the tumor.

We have already mentioned that cancer cells may elicit in the host immune reactions which are injurious to the cancerous transplant, and either prevent its full development after transfer of the cancer cells into a new host or cause their retrogression after they have established themselves and have grown vigorously for some time. Such immune processes may occur also as a result of the active growth of a first tumor, which prevents a second implanted tumor from growing, or they may develop after extirpation of a first, strongly growing tumor which had the ability to neutralize the immune mechanisms developing in the host as a result of this growth (Uhlenhuth's phenomenon). Another type of active immunity may develop after implantation of various normal tissues which belong to the same species as the tumor cells and the host. Identity of the species differential of the tumor and of the immunizing tissue is necessary for the establishment of this type of immunity, but implantation of tissue possessing the same individuality differential as the host is unable to cause immunization. There has been much discussion as to the mechanism underlying these types of immunity against cancer. Do they depend upon the development of antibodies which can be transferred to other hosts, which are thus made passively immune? It has not been possible to demonstrate such a passive immunity in a clear cut way. Nevertheless, there are experimental indications that substances are present in the bodyfluids of the immunized host which injure the transplant (Woglom).[68] In addition, the transplant is injured as a result of cellular reactions, in which lymphocytes, blood vessels and connective tissue may participate. While immune processes are most striking in the case of cancerous tissues, they are not limited to or characteristic of the latter; they may be demonstrated also in the case of implanted embryonal tissues, and even transplanted normal adult tissues may induce certain immune reactions against normal tissues transplanted subsequently (Blumenthal). While this type of immunity is directed against the individuality or species differentials of the transplants, there is evidence that an immunity may develop also against substances which are distinct from these differentials but are characteristic of certain types of cancer.

It is an interesting fact, observed by Loeb and Fleisher, and more recently by Blumenthal, that spontaneous tumors of the mammary gland of mice, which as a rule cannot be transplanted into mice belonging to strains other than those in which they originated, can be successfully transplanted in a considerable number of cases into such animals if they are bearers of spontaneous tumors. There must, here, be active some factor which counteracts the effects of the strange strain differential, which is unfavorable to the long-continued life and growth of the grafted tissues. Whether this is a growth-promoting substance or a merhanism counteracting immunization against the transplant has not yet been determined.

Transplantable tumors present a material favorable for attempts to separate, if possible, from tumor cells microrganisms or viruses able to induce tumor formation if injected into susceptible animals. Early experiments of this kind by the writer and Maxmilian Herzog showed that with rat sarcoma, only injection of macerated tumor tissue, passed through filters which permitted also cells to pass through the pores, gave origin to tumors, whereas the filtrate from Berkefeld filters, which was free of cells, was ineffective. Similar results were obtained with all other mammalian cancers which have been tested so far, and also with the leukemic tissues or cells from mice affected with this tumor-like condition. However, when Ellermann

(1908) in the case of avian leukemia, and especially Peyton Rous (1911) in the case of various types of fowl sarcomas, carried out similar experiments, they succeeded in separating from the tumor cells a filtrate which, on injection into other chickens, stimulated leucocytes to become leukemic or mesenchymatous cells, especially those which were actively proliferating, to give origin to sarcomata composed of the same type of connective tissue which characterized the original tumor. These experiments of Rous, which have proved very fertile, have since been confirmed by Fujinami in Japan and by many others. It is furthermore possible to produce neutralizing antibodies and precipitins against this agent. Murphy and his collaborators, Claude and Sturm, purified the sarcoma filtrates or desiccates and studied the physical and chemical characteristics of these substances. They found that after purification, the active constituents of these filtrates no longer were able to produce antibodies. As to the nature of the agent, Claude believes that it is a phospholipid-ribonucleo-protein complex, similar to that which constitutes the mitochondria of cells. Further investigations must show whether the particles separated by the ultracentrifuge and resembling mitochondria in size and constitution are the real agents of chicken sarcoma, or whether they are the substratum to which the real agent adheres. Jobling, Schemin and Sproul obtained an active substance showing qualitatively a similar constitution to that found by Claude. These investigators, as well as Beard, believe, therefore, that the agent causing the formation of chicken sarcoma is a very complex, non-living substance.

As to the heterogenous transfer of Rous chicken tumor I, either by injection of the virus-containing tumor extract or by transplantation of minced tumor tissue containing cells, Murphy and Rous first found that this chicken tumor I could develop in the allantois of chick embryos, but transplantation into adult ducks did not succeed. Purdy was not able to infect ducks or ducklings with such tumors by injection of large amounts of virus-containing extracts from a Rous or an endothelial chicken tumor, but by transmitting great quantities of minced tumor tissue, he could transfer Rous sarcoma through several generations of ducklings; however, this could be accomplished only if very young ducklings were used—the tumor could not be transplanted into adult ducks. Des Ligneris likewise had negative results when he used adult ducks and geese, but he succeeded in transferring the tumor to turkeys and guinea fowls.

If the Fujinami fowl sarcoma was used instead of Rous sarcoma I, the results were more favorable. Fujinami transferred this tumor through 40 generations of ducks. Gye, by means of filtrates or of cell suspensions, could transfer the same tumor serially to ducklings, but in half-grown or adult ducks the tumor could grow only for some time and it later regressed; however, Purdy succeeded in transmitting it serially to ducklings, and even to adult ducks, by injecting very large amounts of minced tumor tissue. A chick in which such a tumor happened to regress, had thereby acquired an immunity against a Fujinami tumor but not against a Rous tumor.

In general, it may then be concluded that fowl sarcoma can be transferred much more readily to newly hatched, heterogenous birds than to somewhat older ones; with increasing age, the transmission becomes more difficult. However, different types of fowl tumors differ in their ability to grow in heterogenous hosts. In this connection it may be stated that also normal tissues are tolerated better in very young hosts than in older ones. Moreover, different races of a certain avian species differ in their suitability as hosts. This transfer of avian tumors can be accomplished by means of tumor extract (virus) as well as by means of tumor cells; but the latter on the whole seems to be more successful, especially if large quantities are used. A transfer of these tumors into more distant species, such as mammals, does not succeed.

Quite recently, Duran-Reynals has shown that if we transfer Rous fowl sarcoma

I to ducklings, at first the unchanged chicken tumor virus or cells cause the tumor development in these ducklings, as indicated by the fact that the tumors thus produced have not only the same morphological characteristics as the chicken tumor, but they have also the same tissue affinities and the same tendency to be localized. However, after the tumors have lived for some time in ducklings, they may change their characteristics, the chicken-adapted virus becoming a duck-adapted virus; it then tends to cause sarcomas in different organs, such as bones or lymph glands, calling forth a lymphosarcoma in the latter organ. This changed tumor also tends to be generalized. A similar adaptation seems to occur if cells are inoculated; these also cause the same kind of tumors as the duck-adapted virus. Such a duck-adapted virus or cell suspension not only causes tumor development in ducklings, but also in adult ducks. Cells and virus are now no longer heterogenous but homologenous for the duck. If such duck-adapted virus or cell suspension is transferred into chicken, it seems at first to behave like material heterogenous for the chicken, but after some time the duck-adapted virus or cell suspension can again become chicken-adapted, being converted into a homologenous virus. However, it cannot very well be concluded from these experiments that a chicken cell was actually changed into a duck cell, but merely that the changed virus altered at the same time the tumor-producing characteristics of the chicken cells in which the virus lived. It has also been found by Duran-Reynals that the Rous chicken sarcoma I virus may, occasionally, instead of giving rise to tumors, produce a generalized hemorrhagic disease. We have discussed these conditions somewhat more fully, because they are of importance also as far as the problems relating to the causes of cancer in general are concerned.

From this brief resume, it may be seen that by means of the method of transplantation of tumors many interesting facts have been established, which are not only of great significance from a biological point of view, but which have also contributed much to the understanding of the nature of tumors.

The Biochemical Constitution of Cancer Tissues

Cancer tissues have acquired certain new characteristics which differentiate them from normal tissues and also from various other types of growing tissues, as, for instance, those undergoing correlative, hormone-stimulated growth. The increased growth of cancer cells is a permanent characteristic and not a transitory condition, as is the growth of non-cancerous growing cells. We may now inquire whether in addition to these growth and structural differences between cancer and normal tissues there exist also differences in their biochemical constitution and metabolism.

Several interesting differences of this kind have been reported; we shall consider here only the most important ones. In 1926, Warburg observed an altered carbohydrate metabolism in malignant tumors. Oxidative processes were diminished and aerobic as well as anaerobic glycolysis was increased; lactic acid was produced mainly by the splitting of sugar instead of CO_2 by oxidation. The behavior of benign tumors was midway between that of normal and of cancerous tissue. It was found, however, that glycolysis was also increased in the retina, in embryonal tissues, placenta (Murphy), regenerative tissue (Pentimalli), and leucocytes. Moreover, Murphy and Hawkins, as well as others, observed that in malignant mammary gland tumors of the mouse, glycolysis tends to be low or zero. According to Berenblum, even normal skin may behave like malignant tumor tissue in this respect. On the other hand, Dickens and Burk support the view of Warburg that the carbohydrate metabolism of malignant tumors is specific. They consider a relatively low respiratory quotient, together with very high anaerobic glycolysis, as characteristic of tumor metabolism. The increase in aerobic glycolysis is held by Burk to be a secondary condition, due to the fact that notwithstanding an active Pasteur effect, anaerobic glycolysis is so strong in malignant tumors that the respiration in many

cases cannot prevent glycolysis even under aerobic conditions. In mouse leukemia, Hall and Furth find that the oxygen consumption and the Pasteur effect are normal, and that the rate of aerobic glycolysis is not increased in contrast to the anaerobic glycolysis, which is higher in leukemic than in normal lymph nodes. Regenerating or embryonal tissue, like tumor tissue, shows a high anaerobic glycolysis, but this is to some extent compensated by a strong respiration. While in normal tissue the respiration is below that of noncancerous but growing tissue, the anaerobic glycolysis is here also relatively low. The chicken sarcomas caused by a filtrable virus possess the same carbohydrate metabolism which is characteristic of other malignant tumors; the agents transmit therefore, to the normal cells in which they live and multiply, the metabolic characteristics of cancer cells. On the other hand, it seems that the cells of fowl leukemia, which is likewise caused by a filtrable virus, do not possess this type of carbohydrate metabolism.

We may then conclude that while the carbohydrate metabolism of cancer cells shows certain peculiar properties, which, in a general way, distinguish them from other types of cells, the degree to which these distinguishing features are developed differs apparently in different types of tumors and in different species. There occur also variations in the intensity of the oxidation processes of different types of tumors. Aerobic glycolysis is less pronounced, or may be lacking entirely, in the breast carcinoma of mice, while it may be marked in rat and human tumors, and also in chicken sarcomas. Moreover, in certain normal cells, such as monocytes, the carbohydrate metabolism may be similar to that of cancerous tissues. Notwithstanding the utilization of oxygen in the Pasteur reaction by cancer tissue, it seems that the oxidation-producing cytochrome C system of this tissue is diminished. These alterations in the tumor metabolism may occur either at the time of the conversion of non-cancerous into cancerous tissue or shortly preceding this change.

These data suggest that this change in the metabolism of cancer cells is not the cause of the cancerous process and, perhaps, not even directly connected with its essential cause. It seems more probable that it is one of the results of the cancerous transformation and possibly due to secondary conditions. Moreover, this type of carbohydrate metabolism is present in various kinds of cancer to an unequal degree, and a condition similar to it, although perhaps not identical with it, may be found also in certain noncancerous tissues.

An alteration in the chemical constitution of tumors which aroused a great deal of interest was announced a short time ago by Kögl and Erxelben (1939). They found in protein hydrolysates of malignant tumors a considerable amount of glutamic acid and smaller percentages of certain other amino acids in dextrorotatory form, while in normal tissues only the levorotatory stereoisomers occur. However, other investigators were unable to confirm these observations. Still more recently, West and Woglom found that the amount of biotin, a vitamin which is required for the growth of many microorganisms and also of higher animals, is different in cancers and in the corresponding normal tissues, and this variation seems to be in the same direction in which embryonal tissues differ from normal tissues in this respect. In some cases the development of cancer in a certain tissue leads to an increase in the biotin content, but in the normal adult liver, heart and kidney the biotin content is greater than in the embryonal tissues, and greater likewise than in cancerous liver tissue. It is conceivable that liver, heart and kidney, in addition to functioning active biotin, contain a certain amount of inactive biotin serving as reserve material, and that this may explain the different conditions found in these various organs. This biochemical difference from normal tissues is not, therefore, characteristic of the malignant state as such, but it is shared by growing tissues of an embryonal nature. However, still more recent data published by West and Woglom seem to indicate that the biotin content of tumors may be intermediate between that of embryonal and normal tissues, or that it may be lower than either

adult or embryonal tissue, although in one instance it was equal to that of adult tissue. These investigators moreover found the interesting fact that two transplanted mouse sarcomas are able to grow actively in mice depleted of their biotin content by avidin administration; these sarcomas themselves are very poor in biotin. Sarcoma cells thus seem to differ markedly from normal cells, which need biotin for their functioning. Shifts in the biotin content, when the normal tissue passes into the cancerous state, may be the consequence rather than the cause of the cancerous transformation. Furthermore, the biotin content of the Shope papilloma does not increase at the time when it undergoes a cancerous change. Young transplanted carcinoma of the liver in mice has a biotin level similar to that of the normal liver in this species. These data indicate that biotin does not play a significant role in the origin of tumors. However, it seems that biotin may, perhaps indirectly, favor the development of cancer at least in one particular instance; this would follow from the observation that administration of casein, with the addition of cystine and choline, protects rats against the formation of cancer in the liver, which otherwise would be induced by butter-yellow feeding. However, after the addition of biotin to the diet of such rats, the protection was lost or diminished and hepatomas developed in a number of animals.

In tumors, a number of different enzymes occur, and among them, catalase, arginase and amylase. It is of some interest that as a rule in tumor-bearing rats and mice the liver catalase activity is lower than in normal animals, and this change in the amount of liver catalase is proportional to the rate of the growth intensity of the tumor (Greenstein and Andervont).

Vitamin C (ascorbic acid) seems to be increased in tumor tissue as it is in some other actively growing tissues, and according to Lasnitzki and Brewer, the ratio between the isotopes of potassium K_{39}:K_{41} is somewhat different in tumors from that in normal tissues. As to the significance of these differences between the biochemical constitution of tumor and of normal tissues, it seems doubtful whether they are connected with the factors leading to cancer formation; it seems more probable that they are the consequence of the cancerous growth processes directly preceding the initiation of the cancerous growth. To decide this question, it would be necessary to compare the changes found in the chemical constitution of cancer cells with the alterations in the chemical constitution which are noted in the various noncancerous growth processes.

There is one particular type of cancer, namely, the experimentally produced hepatoma of rats, in which biochemical processes which influence the initiation of this kind of growth have been studied more recently. This carcinoma in the liver of rats can be produced by oral administration of *o*-aminoazotoluene (Sasaki and Yoshida), or of dimethylaminoazobenzene (butter yellow) (Kinosita); the latter substance is somewhat more effective than the former. The liver tumor develops only if the rats are kept on an inadequate diet of non-polished rice. When butter yellow is used, cirrhosis and cholangioma formation usually precede the development of hepatoma, while *o*-aminoazotoluene may induce the formation of hepatoma directly, without a preceding cirrhosis and with or without formation of cholangioma. Rhoads and his collaborators have found a deficiency of coenzyme I and of riboflavin in this kind of tumor. Several investigators have observed that various supplements to the rice diet may diminish or prevent the appearance of this type of cancer. Such supplements are yeast, liver, kidney, ricebran oil, and various extracts of yeast, ricebran or liver. If a combination of riboflavin and casein, or of the vitamin B complex and cystine and choline was added to the rice diet of rats to which butter yellow was administered, carcinoma of the liver did not develop. A substitution of wheat, rye or millet for the rice diet likewise decreased the cancerigenic action of butter yellow and of *o*-aminoazotoluene. Kinosita and Rhoads, Stevenson and Dobriner have studied the split products found in the urine of rats fed with butter

yellow or *o*-aminoazotoluene; there was an indication that it is some of these secondary substances which have the cancerigenic effect, and as stated, Rhoads and his associates suggest that this effect consists in an interference with coenzyme I, and perhaps also with the riboflavin present in the normal liver cells. However, the outcome of further investigations must be awaited regarding these problems.

There is still another method suitable for the study of specific chemical constituents in cancer tissue, if such should exist, namely, the production of antibodies against cancer. Specific substances present in the cancer cells might function as antigens and induce the production of immune substances in animals belonging to different species (rabbits). The interaction of the antigens with the immune substances (antibodies) in the blood serum of the animal injected repeatedly with cancer cells or their constituents might be recognized by precipitation, agglutination, complement fixation, anaphylaxis, or otherwise by neutralization of the specific antigens. In other investigations, the blood serum of a person or animal affected with cancer has been mixed directly with a tumor-specific antigen, in order to demonstrate the presence of an antibody preformed in the blood serum of such an organism. Experiments of the former or latter kind have been carried out especially by Hirszfeld and his collaborators, by Lehmann-Facius, Lumsden, Kidd, Dmochowski, Welker, and a number of other investigators. Their results agree in that there is noted in cancer tissue the presence of certain antigens not present in normal resting adult tissues; but there are differences between the conclusions arrived at by these investigators as to the degree of specificity and as to the nature of these antigenic substances.

Hirszfeld and Halber found that a complement fixation reaction takes place when alcoholic extracts of human cancer, to which cholesterol has been added, are tested with sera of persons affected with cancer. However, this reaction did not occur with every serum and it was observed occasionally also in sera from patients suffering from other diseases and those from pregnant women. Hoyle, on mixing lipoid antigens extracted from transplanted mouse tumors and the sera of mice bearing these tumors, obtained complement fixation reactions. But according to Hirszfeld and Halber, also immune sera against rat embryo extracts may react with alcoholic extracts of human cancers, and immune sera against extracts of normal rat organs reacted in some instances even more strongly with human tumor extracts than with the antigen used for the preparation of the immune sera. Conversely, immune sera against human tumors reacted also with human embryo extracts. Furthermore, antigens obtained from necrotic human tissue and from degenerating leucocytes behaved in a similar manner to antigens present in human tumors. Likewise, Dmochowski noted cross reactions when he mixed extracts of experimental (benzpyrene) rat tumors, rat muscle, or rat embryos with immune sera. On the other hand, Hirszfeld, Halber and Laskowski found similar reactions of rabbit immune sera against extracts of human cancer and against extracts of the liver of patients affected with carcinoma of the stomach.

In contrast to the results of these experiments are those of Lumsden, who observed that antisera, obtained by immunization of rabbit or sheep with human, rat or mouse cancer material, exert in vitro a specific injurious effect on cancer cells of various kinds, but not on normal tissues. This would indicate that there exist antigens which are common to all kinds of cancer, irrespective of species and tissues in which they originated. However, it seems that this specificity is not absolute, because normal spleen cells, and perhaps also spermatozoa, possess similar antigens.

Different again are the conclusions of Welker and his associates. They subjected the autolysates of cancer tissue to long-continued treatment with alumina cream and in this manner eliminated admixtures of serum proteins from this material, with which they then immunized rabbits. They thus obtained immune sera which reacted in a specific manner with extracts of cancers from the same species but not with those from a different species; in addition, these immune sera showed a

preference for extracts of cancers which originated in the same kind of tissues or organs that had served as antigens. These specific cancer antigens possess, therefore, not only species specificity, but also organ or tissue specificity, and they may be considered as modified organ or tissue differentials. However, such antisera should be tested with a still greater variety of noncancerous tissues before the conclusion as to a complete specificity of these antigens can be fully accepted.

The fact that tumors may contain antigens which are specific, differing from those found in normal tissues, and that these antigens can give rise to specific immune substances in the serum of the bearer of the tumor is indicated also by the experiments of Kidd and of Dmochowski, who found that rabbits which bear transplants of Brown-Pearce carcinoma contain in their blood serum antibodies which, when mixed with tumor extract, fix complement, while extracts of normal tissues do not have this effect. But whereas Dmochowski assumes that the Brown-Pearce tumor, as well as Murphy's liposarcoma of the guinea pig have acquired this specific antigen as a result of long-continued transplantation, Kidd believes that the antigen is a virus, which he considers as the cause of the Brown-Pearce tumor. He arrives at this conclusion because of the similarity in the behavior of the Brown-Pearce tumor antigen and the antigen from the Shope rabbit papilloma, which is known to be a virus disease. Both antigens can be filtered through Berkefeld V filters, and the serum of the rabbits affected with papilloma not only fixes complement when combined with the papilloma antigen, but in addition is able to neutralize the papilloma virus, and both these antibody effects seem to be a function of the same substance. This immune substance which thus can be used as a test for the virus may develop even under conditions when the presence of the papilloma virus in the tumor cells can no longer be demonstrated by the usual methods. In this instance, then, a specific virus would act as the specific antigen present in tumor cells and distinguishing it from normal tissues. However, while the findings in regard to the Shope papilloma antigen and antibody seem to be definite, there is apparently a certain variability as to the results obtained with the Brown-Pearce tumor, and further experiments in these directions may perhaps clear up this point. Moreover, there is the possibility that virus and antigen are not identical.

We may then conclude that specific antigens exist in tumors, differing from those obtained from the corresponding normal tissues. On the other hand, tumors possess the same individuality and species differentials as normal tissues of the individual and the species in which the tumors originated.

The Causes of Cancer

After this discussion of the most characteristic features of cancer, we may now proceed to the analysis of the factors which cause the transformation of normal tissues into cancerous growths. This transformation can be induced in all tissues which have the potentiality to proliferate, but it takes place with unequal readiness in different types of tissues in the same species and in the analogous tissues of different species. The principal factors which have to be considered as potential causes of cancer are the following: (a) The constitutional factors of the tissues of individuals and species, which are largely genetic in nature. There is, in addition, a substance which in certain strains of mice is transmitted with the milk from the mother to the offspring and which helps to determine the readiness with which mammary cancer and perhaps one other type of cancer develop in these mice; this substance is likewise hereditarily transmitted in various strains of mice. (b) Stimulating factors originating either within the organism itself or in the outer environment, which act on the cells and make them cancerous. The distinction between these two sets of factors, however, while very useful as an approximation, is not as sharply defined as it might appear, inasmuch as some hereditarily transmitted factors may function as growth stimuli, and conversely, stimulating factors may also be, at least partly,

genetically determined. As to the character of the various stimulating factors, these may differ greatly in specificity. Among the most specific ones are (1) certain viruses, (2) hormones, and (3) certain metazoic parasites; next in order come (4) tar and the so-called carcinogenic (cancerigenic) hydrocarbons, which seem to be able to act on many tissues endowed with growth potentialities, and are specific mainly because of the great intensity with which they act. The least specific factors are those which not only may affect many tissues indiscriminately, but which act also with a lesser intensity, and are well known for other effects they may exert. To this group belong various radiations (Roentgen rays, radium, untraviolet rays), as well as ordinary chemical and mechanical irritations, usually acting over long periods of time. The hereditary genetic factors function largely by determining the degree of responsiveness of a certain tissue to the various stimuli affecting it; but in some instances they may act indirectly by inducing certain diseases, which themselves predispose to the development of cancer. In regard to the milk factor, which is especially potent in the case of breast cancer, it is not yet certain whether it acts by intensifying the response of the tissue to stimuli, thus supporting the genetic factor, or whether it functions as a growth stimulus.

We shall now discuss these various factors, although not necessarily in the above order.

(a) **Heredity in Cancer and Its Interaction with Hormones and the Milk Factor.** All activities of living organisms are determined essentially by the inherited characteristic structure and metabolism of these organisms, and these constitutional factors are the most important ones, although they are effective only in coöperation with the environment in which the organisms live. All observations likewise show that these characteristics are hereditarily transmitted from generation to generation, in a way specific not only for each kind of organism but also for each organ, tissue and function. Cancerous growth is one of these reactions of the living substratum. In this very general sense, then, all cancers are predominantly determined by heredity. But in a more restricted sense, only conditions differential for smaller groups, such as species, strains (races) and individuals are included in the genetic analysis. We are interested in the hereditary, constitutional factors, mainly in their contrast to environmental factors, and especially we wish to compare the relative importance of these two sets of conditions. The environmental factors consist not so much in the basic conditions which affect all vital phenomena in a non-specific sense, such as oxygen, food, temperature, but above all in the specific factors, which, within a certain range, make possible the experimental control of these phenomena. Furthermore, the mode or mechanism of inheritance, expressed in terms of Mendelian genetics, has become a leading subject of scientific analysis.

The study of the significance of heredity in cancer started in a preliminary way towards the end of the last century and at the beginning of this century, but systematic investigations were begun not quite forty years ago. The first observations concerned some colonies of mice and rats, and cattle on some ranches in which certain cancers seemed to be endemic, owing to hereditary factors. Findings of this kind were made in mice and we described such observations in rats and among cattle. A number of years later, Tyzzer and the writer, and subsequently Murray in London, noted the hereditary transmission of cancer of the lung and mammary gland, respectively, among certain strains of mice. From 1910 on, in association with Abbie Lathrop, we made a systematic study of the inheritance of mammary gland carcinoma in a considerable number of strains of mice which had been inbred for a variable number of generations. It could be shown that there existed extreme differences in the cancer rates of different strains of mice; while in some of them the incidence was about 90 per cent or more, in others it approached zero, and these incidences remained approximately constant in subsequent generations, although it was possible

in some instances to develop substrains which differed in their cancer rate from that of the main strains. Not only was the cancer rate an inherited characteristic, but also the cancer age. In the majority of strains a high cancer incidence was associated with an early cancer age and a low cancer incidence with a later cancer age, indicating that a common intensity factor determined both incidence and age. But in some cases a low tumor rate was associated with an early development of the tumor, as an indication that if in a mouse belonging to such a strain the conditions for cancerous growth of the mammary gland existed, the necessary stimuli reached the threshold of effectiveness in a relatively short period of time, whereas, in some other strains the hereditary conditions for the development of cancerous growth were present in many individuals, but became manifest only if the stimuli acted over a long period of time.

If strains with a low and with a high incidence of cancer were hybridized, the parent strain with the higher rate dominated in the majority of cases; there was in successive generations of the hybrids no indication of a hereditary transmission of mammary gland cancer which would correspond to a one factor ratio, but intermediate conditions, which suggested the presence of multiple factors, were usually observed. Likewise, the great number of different cancer incidences among the different strains was favorable to this interpretation.

In addition, it was found that if mice belonging to strains with a high and with a low cancer incidence were hybridized, the cancer rate of the mother strain dominated over the cancer rate of the father strain in a considerable number of instances. This fact, taken together with several observations in which in reciprocal crosses, the hybrids, followed the tumor rate of the mother strain, suggested to us the conclusion that the mother strain may be of greater significance in determining the tumor rate of the mammary gland than the father strain.

However, it was also found in the course of these investigations that the hereditary factor was not the only one that determined the development of mammary gland carcinoma in mice, but that it needed the coöperation of ovarian hormones, and in particular, of the estrogenic hormone. By extirpating the ovaries of the mice belonging to high cancer rate strains at various periods of life, it was shown that the development of mammary gland cancer appeared the more frequently and the earlier in life the longer these hormones had had a chance to act. If these hormones had been active only during a very short period, tumors did not appear at all in the majority of strains tested, and subsequently, Cori found that in eliminating the ovarian influence still earlier in life, cancer of the mammary gland did not appear in a single case. We concluded, furthermore, that the fact that male mice were free of mammary gland carcinoma was due to the lack of the ovarian hormones in these animals. This was confirmed later by Lacassagne, who, by injecting estrogen into male mice, was able to induce cancer of the mammary gland provided the animals did not belong to low tumor rate strains. It is possible, also, that the difference in the rate of mammary gland cancer between breeding and virgin mice, which we had previously noted and which exists in a large majority of strains of mice, although it varies greatly in intensity in these strains, is due to the action of hormones given off during the period of pregnancy and lactation; however, other factors may perhaps coöperate in this condition. This was taken to indicate that the hormone, in addition to causing the transformation of normal mammary gland tissue into carcinoma, exerted still another function. Previously we had found that the corpus luteum sensitizes the uterine mucous membrane in such a way that an ordinary mechanical stimulus calls forth growth processes of a very great intensity, leading to the production of temporary tumors, the placentomata. This concept of a gradually increasing tissue sensitization by means of hormones was later transferred by us to the mammary gland of mice, where the sensitization took place under the influence of ovarian hormones.

We also suggested, in the course of our earlier investigations, that the influence

of hormones in the origin of cancer may not be limited to ovarian hormones, but that different hormones may be effective also in various other types of cancer. A further conclusion which could be drawn from these experiments concerned the sensitization of the mammary gland tissue of the mice belonging to high cancer rate strains, a sensitization which took place after the hormones had been active for a certain length of time. It was found that if the hormone action was interrupted before cancer had developed, the ordinary stimuli active during the normal metabolism of the organism were able to transform, step by step, the state of sensitization which had been induced, and which could not be differentiated histologically from a non-sensitized condition, into real cancer. In conformity with this interpretation, more recently Burns and Schenken observed that if estrogen had been injected for a period of five months into male mice of a high tumor rate strain, in many animals cancer of the mammary gland developed in the course of time subsequent to the cessation of the stimulation by the injected estrogen.

Returning to the hereditary factors, Maud Slye, whose investigations followed some of those already mentioned, observed that the relative incidence of various types of cancer differs greatly in different strains of mice. There are types of tumors that are frequent in a certain strain and others that are rare, and heredity largely determines the incidence. This investigator assumed that all forms of cancer are transmitted as a single recessive factor. This conclusion is contrary to the evidence already mentioned, which points to the presence of multiple factors and of dominance rather than recessiveness; which showed furthermore that heredity is effective only in combination with factors of a different nature, a condition which complicates and renders difficult the genetic analysis of the origin of cancer of the mammary gland in mice. Conditions seem to be somewhat simpler in the case of pulmonary carcinoma of the mouse, which again is much more frequent in certain strains than in others. Clara Lynch could show that this type of cancer is transmitted as a dominant factor, and Bittner has confirmed this conclusion; but the incidence can be increased under the influence of carcinogenic hydrocarbons. In this instance, too, it has been at first assumed that a single dominant genetic factor was responsible for the hereditary transmission of this tumor, and this view is still held by Bittner at the present time. However, it seems to us that the evidence points more to the conclusion that, as regards heredity, the origin of these tumors is determined by multiple factors. Differences in interpretation which have developed are due largely to differences in the criteria used in the genetic analysis of this process. There are quantitative elements involved in the process of cancerization, the study of which in many instances has been neglected; instead, it has been considered an "all or nothing" process. If these quantitative conditions are also taken into account, a multiplicity of factors becomes the more probable assumption. Confirmation is furthermore found in the work of Andervont, regarding the grading of the cancer incidence in hybrids, in accordance with variations in the choice of the paternal strain, which transmits genetic elements but not the milk factor.

An advance in the study of these and related problems as to the origin of cancer has been made possible in more recent years by Little and his collaborators, Strong, Murray, Bittner and Cloudman, who by long-continued brother-and-sister matings increased the homozygous condition of the strains used for the genetic analysis of cancer, and in particular of mammary gland cancer. Each of these strains differs from the others in the incidence and type of cancer to which it gives origin. Because these strains are approaching—although not fully attaining—a homozygous condition, in which all individuals composing a strain have about the same genetic constitution and receive about the same amount of milk factor, it would be expected that, given the same amount of stimulation by hormones, all the mice of a certain strain should be affected in the same way with cancers; they should all acquire cancer of the mammary gland at about the same age, or should all remain free of cancer. This

is, however, not the case. These closely inbred strains behave in this respect as did the less closely inbred strains used formerly; each strain is characterized by a definite average percentage of individuals which are affected by cancer, inclusive of leukemia, and by the average age at which the cancer appears, and these figures vary greatly in different strains. However, Strong, in the case of mammary gland carcinoma, and MacDowell, in the case of leukemia, showed that if in a certain strain the cancer incidence of the offspring of mice affected by cancer is compared with that of the offspring of mice not affected by cancer, both behave in the same way; they have the cancer incidence typical of the strain. This would suggest that all mice belonging to the same inbred strain have the same genetic constitution and that the difference in the cancer rate of different individuals within the same strain must be due to non-genetic factors. But if a variety of chance conditions cause these differences in the appearance of cancer in different individuals of the same strain, it is difficult to explain why there should be such a constancy in the cancer incidence characteristic of each strain as a whole; such an explanation would postulate a corresponding constancy of the averages of these environmental factors affecting the individuals belonging to a given strain. However, it is necessary to consider also the fact that in reality none of the strains tested so far by the criteria developed for the analysis of individuality differentials has been found completely homozygous; and this fact may be responsible, at least to some extent, for these variations among the individuals of an inbred strain, although it is probably not sufficient to explain the principal data in this respect.

Reciprocal crosses between closely inbred strains with a high and a low incidence of mammary gland carcinoma made the preponderance of the mother strain, in determining the appearance or absence of cancer, more evident than it had been in the earlier experiments carried out with less closely inbred strains. In order to determine which of the non-genetic factors might be responsible for this difference in the cancer incidence found in reciprocal crosses, Little and his collaborators tested the effect of the mother's milk and of factors transmitted by uterus and placenta to the developing embryo and fetus. This was done by varying the females nursing the newborn mice and by transferring fertilized ova to the uterus of other females. In carrying out the first kind of experiments, Bittner was able to show that this difference between the effectiveness of father and mother in determining the cancer rate of the mammary gland of the hybrid offspring was due to a substance which is transmitted with the milk of the mother to the child, and which is very influential in determining this cancer rate in the offspring. The mice belonging to high cancer rate strains possess this substance in large quantities, and the mice belonging to low tumor rate strains lack the milk factor or possess it only in small amounts. This milk factor occurs, however, not only in the milk of the mothers belonging to a high tumor rate strain, but also in various organs of these mice as well as in the blood and in the cancer developing from the mammary gland. Little interprets the difference in the cancer incidence of hybrids between high and low tumor rate strains, in cases in which father or mother belong to the high cancer rate strain, as indicating that the milk factor is about ten times more effective in determining the development of cancer than are the genetic factors. However, there are indications that the genetic factors are of much greater significance than these hybridization experiments indicate. The milk factor plays such a prominent role in determining the cancer rate in these hybrids because they possess a genetic constitution which is sufficient for the development of mammary gland cancer. Under these conditions the action of the milk factor, together with that of the hormones, is the determining variable condition. But in strains which do not possess the appropriate genetic constitution, the milk factor has very little power to induce cancerous growth in the mammary gland. This holds good, for instance, in the case of strain C57, in which the genetic constitution is not favorable for the development of mammary gland carcinoma. Trans-

mitting the milk factor to these mice has only a relatively slight effect; the milk factor seems to function only if the genetic constitution is present, which permits it to act. However, it seems that the incidence of mammary gland cancer of C57 mice which have been nursed by female mice belonging to high tumor rate strains, can be increased considerably if these mice have been allowed to breed; the low incidence applies only to non-breeding mice. It is probable that this effect of breeding depends upon the increased action of ovarian hormones on the mammary gland during pregnancy and lactation, although other possible explanations exist. The combined action of hormones and milk factor is therefore able to raise very much the cancer rate of a strain in which normally the cancer rate is as low as it is in strain C57.

Furthermore, the genetic constitution appears to vary in a graded quantitative manner in different strains, in accordance with the observations on which we have already reported. In this way we would be inclined also to interpret certain experiments of Andervont, in which he found that the cancer rate of the father strain helped to determine the cancer rate of the hybrids. In accordance with these differences in the genetic constitution, the effect of the milk factors varies in different strains. This is evidently the reason why in non-breeding mice it is so much easier to lower the cancer rate in mice belonging to high cancer rate strains by withholding the milk factor, than to increase it in mice belonging to low cancer rate strains by supplying this factor through foster-nursing by mothers from high cancer rate strains. Without the appropriate genetic constitution, the milk factor is evidently ineffective, or only very slightly effective; but if the genetic constitution is favorable for the development of cancer, the milk factor has a chance to produce breast cancer, provided the third factor, the stimulation by hormones, is also active. Without the coöperation of milk factor and hormones, the otherwise suitable genetic constitution is unable to produce mammary gland carcinoma. However, this requirement, as far as the milk factor is concerned, seems to obtain only in the case of mammary gland carcinoma of the mouse, and not in carcinoma of the lung; nor has it as yet been determined in other types of cancer. However, in the case of leukemia MacDowell has shown that in hybrids between strains with a high and with a low incidence of leukemia, the influence of the mother strain preponderates, and according to Furth and Cole, an influence favoring the development of leukemia is transmitted through the milk of high leukemia strains; but this influence of the milk factor seems to be much weaker in leukemia than in the mammary gland carcinoma and it seems to be transmitted only to the F_1 generation of hybrids and not to successive generations of these hybrids.

The genetic factors, therefore, are of great importance in the development of mammary gland carcinoma in mice, and great differences exist in the cancer-determining constitution in different strains. Moreover, there seems to be a definite connection between these differences in the genetic constitution and the amount of milk factor a certain strain possesses. Strong indications exist that strains with a marked genetic tendency to the growth of mammary gland carcinoma possess a large amount of milk factor, and that strains with a low genetic breast cancer tendency possess only a slight amount. Thus it seems, as already stated, that strain C57, in which the genetic constitution is unfavorable to the development of mammary gland carcinoma, possesses only a very slight amount of the milk factor, or none; in strain CBA, likewise, the genetic constitution is relatively unfavorable to the formation of breast cancer, but it is less unfavorable than in strain C57; correspondingly, the amount of milk factor present seems to be somewhat greater. Data published by Andervont seem to indicate that in strain C the genetic constitution is somewhat more suitable for the development of breast cancer than in strain C57, and that, correspondingly, the milk factor is here more effective; but whether it is present in this strain in a greater amount than in strain C57 is apparently not certain, although this is possible. In the high tumor rate strains A, C3H and D, the milk factor seems to be present in larger

quantities. While this parallelism between a suitable genetic constitution and the presence of the milk factor has not yet been established in every case, the data already on hand strongly suggest such a connection.

This parallelism is a very interesting condition and it raises the question as to the origin and nature of the milk factor, and as to whether it may not be an agent produced in the mice themselves. Such an interpretation is perhaps suggested also by the statement of Bittner, if the writer understands him correctly, that if mice belonging to a high tumor rate strain have been nursed by mice belonging to a low tumor rate strain, the offspring of these mice may regain the milk factor in the course of continued propagation by inbreeding. That the milk factor probably originates in the mouse itself is further indicated by the parallelism which exists, at least in a certain degree, between the genetic constitution of an inbred strain and the amount of milk factor possessed by the female mice of this strain; otherwise we would have to assume that the affinity of an extraneous virus to various strains of mice is regulated by the degree of their genetic tendency to develop mammary gland carcinoma, an assumption which does not appear likely. However, it would be necessary to assume that the milk factor multiplies in the strains which have a high tumor rate, but that it does not multiply in strains with a very low tumor rate. If this were not so, it should be easy to transform a strain with a very low tumor rate and possessing only a very small amount of milk factor, or none, into a strain which retains its low genetic tendency to mammary gland carcinoma but possessing a large amount of milk factor. All findings published so far make such a change very improbable.

There is, however, an observation made by Bittner which it seems difficult to reconcile with this conclusion; he found that if high tumor rate strain A mice are foster-nursed by very low tumor rate C57 mice, the tumor incidence of these A mice, in accordance with expectations, is very low; however, some of these mice may develop tumors. If this latter group of mice, possessing a greater incidence of cancer, are bred among themselves, the cancer rate in the offspring diminishes from generation to generation, until in the end it becomes very low. This is contrary to the result which should have been expected if these A mice would themselves have produced the milk factor and transmitted it to their offspring; but this did not take place. A possible explanation of this condition may be that strain A mice have a high cancer rate only if they are allowed to breed and that virgin A mice have a low breast carcinoma rate. Their genetic constitution is therefore such that it needs to be supplemented by strong hormonal stimuli, which are provided by pregnancy and lactation, and that the amounts of ovarian hormone active in virgin strain A mice is unable to cause the cancerous transformation of the mammary gland in a large number of individuals. In addition there is the possibility that in strain A, also, the milk factor as a rule is produced in large amounts only in breeding mice. If in some exceptional virgin A mice enough milk factor was present to allow the development of mammary gland tumors, it is quite likely that this condition might not be transmitted to their offspring. It might perhaps be of interest to repeat this experiment with mice belonging to strain C3H, in which also virgin mice have a relatively high tumor rate.

There is also an observation of Murray and Little which seems to be opposed to the assumption that the mice produce the milk factor themselves. They found that if hybrids (C57 ♀ × D ♂) F_1 or (D ♀ × C57 ♂) F_1 are backcrossed to their parent strains, C57 ♂, D ♂, or D ♀, in the first backcross generation the development of breast cancer is determined by the milk factor rather than by the genetic constitution. If the genetic constitution of strain D is designated by the symbol CCCC and the genetic constitution of C57 by cccc, and the presence of the milk factor by E and its absence by e, then they found in the first backcross generation the following carcinoma rates: CCCce = 2.40%; CCCcE = 35.7%; Cccce = 0.41%; CcccE = 34%. The

mammary gland carcinoma incidence of strain D is 80%, while that of strain C57 is very low. From these experiments they concluded that it is almost exclusively the milk factor which in these cases determines the breast cancer incidence. However, it seems more probable that the cancer incidence here also was determined by both the genetic constitution and the milk factor. We have to assume that the presence of C in CcccE is sufficient to allow for a cancer rate of 34%, although it does not make possible a cancer rate of 80%, which requires the genetic constitution CCCC.

If two of these backcrosses were continued for eight consecutive generations, approximately the following constitutions were obtained: (1) CCCCe and (2) ccccE; both of these have a low tumor incidence. Murray and Little concluded therefore that after eight backcrossings, the effect of the genetic constitution, as well as of the milk factor, was lost. But these results might be interpreted also as indicating that both milk factor and a favorable genetic constitution are prerequisites for a high tumor rate. In mice with the constitution CCCCe, which is genetically very favorable, the tumor rate is low because the milk factor is lacking; in mice with the constitution ccccE, in which the milk factor is present, the incidence is low because the genetic constitution is very deficient. It would be of interest to test thhe cancer rate of the eighth backcross generation, which possesses the constitution CCCCE; here the tumor rate incidence should be relatively high if the interpretation just given is correct. While there is thus some evidence in favor of the view that the milk factor is produced by the mice themselves, and in the largest quantity by those belonging to high tumor rate strains, the possibility that it is an extraneous, virus-like material cannot as yet be excluded.

The milk factor has been designated by Little, Murray, and Bittner as the extrachromosomal factor in contradistinction to the genetic factors localized in the chromosomes. However, the hormone actions which are necessary for the production of mammary gland cancer may in a similar way be considered as extrachromosomal; both milk factor and hormones can be transmitted to the organism from the outside. Yet, the genetic constitution of an animal, among other factors, is responsible for the development of endocrine organs and therefore also for the production of hormones that lead to the so-called spontaneous development of mammary gland carcinoma; likewise, observations already mentioned suggest very strongly the possibility that the production of the milk-producing factor may also, directly or indirectly, depend upon chromosomal factors, as indicated by the mode of distribution of the milk factor among different types of strains. However, there is a significant difference between hormones and milk factor, insofar as the former are secreted in about the same quantities in the various strains of mice, while the amounts of the milk factor present differ very much in these strains; the genetic constitution of the latter is therefore more important in the case of the milk factor than in the case of the hormones.

Bittner first suggested that the milk factor was a hormone-like substance; however, the long-continued effect, extending over many months, of relatively small amounts of this substance made this interpretation improbable and suggested instead a complex, presumably colloidal, constitution of this factor. We expressed therefore, a number of years ago, the opinion that it might be a virus-like substance, without implying thereby that it needed to be a living organism. In subsequent investigations Bittner also adopted this view, and the recent results of several investigations, by investigators who used the ultra-centrifuge in determining the molecular size of the milk factor, led to the conclusion that it is a large molecular substance, not present in the cream partition of the milk. It passes through a Seitz filter and can be obtained from milk as well as from extracts of glycerin-treated tissues, and it is not inactivated by desiccation. The size of the active particles is less than that of some of the larger known viruses, as, for instance, vaccinia virus (Visscher, Green and Bittner; Bryan, Kahler, Shimkin and Andervont). The recent work of Fekete and Little indicate that not only through the milk can a cancerigenic effect be transmitted

to mice, but that such an effect may perhaps also be transmitted to the embryo or fetus by way of the uterus. However, these experiments do not appear to be quite decisive and further investigations should determine definitely whether such an effect actually exists.

While these observations concerning the milk factor have been made in recent years, our knowledge concerning the significance of hormones in the origin of cancer has also been extended during this time. Thus we have found that the activity of the ovarian hormones may be much increased by transplantation of anterior hypophysis from mice belonging to the same inbred strain; these transplants, as a rule, remain alive during the lifetime of the host. This effect has been especially noted in virgin mice of strain A, which normally have a low rate of mammary gland carcinoma. Transplantation of the pituitary is ineffective in strain C57, in which the obstacles to the extensive growth of the mammary gland are very great; it is also ineffective in males, where ovaries are lacking.

Cramer and Horning have observed that contrary to the action of pituitary transplants, certain extracts of the anterior hypophysis may inhibit the development of carcinoma of the mammary gland which takes place under the influence of ovarian hormones; they attribute this effect to the thyrotropic hormone which is produced in the anterior pituitary. However, it seems to us possible that it is due rather to a constituent of the pituitary, which has an injurious influence on the ovarian follicles and promotes their atresia. Also, testosterone may counteract the effect of estrogen, at least in the case of benign tumors of the mammary gland in the rat (Heiman). There exists, then, a double kind of hormone action on the development of mammary gland carcinoma, namely (1) a direct action, and (2) an indirect action which affects the primary endocrine organs and the hormones given off by the latter.

Following the observation of Spiegel that adrenal cortical tumors may appear in old male guinea pigs which had been castrated at an early age, and that there developed in these animals also masculinizing effects, Woolley, Little and Fekete noted that if in mice belonging to high tumor rate strains the ovaries are extirpated soon after birth, nodular hyperplasia of the adrenal cortex develops in a number of animals, and vagina, uterus and mammary gland are stimulated; even mammary gland tumors develop, evidently as the result of the adrenal hypertrophy, which presumably produces estrogenic hormones. In low breast tumor rate strains, these effects of early ovariectomy were not observed. Of interest are also the experiments of Geschickter, who produced mammary gland carcinoma in rats under the influence of estrogenic hormones, probably in strains hereditarily predisposed to this transformation, as well as the experiments of Nelson, Moricard and Cauchoix, and of Lipschütz, who produced fibromyomata in the uterus of the guinea pig. Lipschütz found such tumors even in organs adjoining the uterus. It is known that estrogen may exert certain effects also in the male organism; correspondingly, Hooker, Gardner and Pfeiffer observed the development of interstitial gland tumors in the testicle of mice injected with estrogen. These tumors have, however, not yet reached the stage of fully developed cancer, as indicated by the fact that as in the case of some benign tumors of the mammary gland, their transplantation succeeds only if the hormone action, to which they owe their origin, continues in the new host. There are some observations which seem to indicate that in cancer of the male breast in man, administration of testosterone propionate favors the formation of osseous metastases and that orchidectomy may inhibit the growth of the mammary gland tumor as well as of the bony metastases (Farrar and Adair). Previous to these investigations, Huggins and his associates had observed a favorable effect of orchidectomy and of administration of the estrogenic stilbestrol on carcinoma of the prostate and on bony metastases of this tumor; concomitantly, the acid serum phosphatase decreased in amount. But no complete regression of these tumors occurred. On the other hand, testosterone seems to aggravate these conditions. Also, Schenken and

Burns noted regressive changes in prostatic carcinoma after administration of stilboestrol. There are some indications that in the exprimental production of teratomas in the testis of birds, which, according to Michalowsky, can be accomplished by injection of zinc salts into this organ, gonadotropic hormones may play a role (Falin); a similar effect was attributed to these hormones by Champy and Lavedan in the development of seminomas, which may follow partial extirpation of the testicle in fowl. It is very probable that future investigations will add other examples in which hormones are or may be significant in the origin of cancer, and, in general, wherever hormones stimulate or participate in the stimulation of growth processes of certain tissues or organs, they may play a role also in the development of cancer of such tissues and organs. But as we have seen, it is essentially the stimulating factors in interaction with genetic factors which determine the transformation of normal into cancerous tissues, with the additional effect of the milk factor in certain instances.

So far, we have referred mainly to the differences in the genetic constitution which exist between different inbred strains of mice. However, genetic differences exist also between different species. Thus in mice, carcinoma of the mammary gland is the preponderating tumor, whereas in the rat there is apparently a much greater tendency to the development of sarcoma. The tumors which develop in the mammary gland in this latter species are largely benign tumors, in which both glandular and connective tissue elements participate. In the guinea pig, spontaneous tumors seem to be much less frequent than in the mouse or rat, this difference being essentially due to genetic conditions.

The genetic constitution determines the kind of tissues which respond readily to certain growth-producing stimuli and it determines also the mode of growth-progression in response to long-continued stimulation. To some extent, genetic factors play, here, a rôle similar to that seen in the embryonic development of individual tissues and organs. The genetic factors which act on the different organs and tissues are the same in a given individual; but notwithstanding this fact, these genes make possible the specific and distinct structure and function of these various organs and tissues constituting the organism. In the same way, the hereditary tendency to the development of cancer is specific and distinct for all organs and tissues in an individual, strain and species, notwithstanding the identity of the gene sets acting on these various organs and tissues. But whereas in the building up of the organism these inner factors—genetic factors in combination with cytoplasm and formative contact substances—essentially determine the development of the organism and its parts, in the development of cancer there is required in addition the coöperation of special stimulating factors of an endogenous or exogenous nature.

(b) **The Stimulating Factors.** Physicians and pathologists have known for a long time that chronic irritations of various kinds, chemical as well as mechanical, and also those caused by radiation and metazoic parasites, may in the end lead to the development of cancer in man. The common factor underlying these irritations is long-continued, growth stimulation of the tissues which are exposed to these injurious conditions. In the preceding account we have already described one type of stimulating factors in their interaction with hereditary factors, namely, hormones. We shall now continue the analysis of hormone action and then discuss in succession the various other types of stimulating factors active in the production of cancer and the mode of their action.

(1) *The action of hormones.* In regard to the mode of action of hormones in the production of cancer, a comparison between the rôle they play in the mammary gland, and in vagina and cervix, is of interest. Hofbauer attempted to produce abnormal proliferation in the cervix of the guinea pig by administration of anterior pituitary substance. Overholser and Edgar Allen elicited precancerous prolifera-

tions in the cervix of a monkey by injections of estrogen. In mice, precancerous lesions in the cervix have been produced by us, by Lacassagne, and by Gardner, Allen, Smith and Strong, and we as well as the last named investigators induced also real cancer growth in these organs. Allen and Overholser, in order to effect an abnormal growth in the cervix, combined administration of estrogen with incisions made into this organ. However, such incisions merely produce regenerative processes of an acute character, which in all probability are not able to intensify noticeably the action of hormones. On the other hand, we observed, in vagina as well as in cervix, indications that mechanical irritation of a long-continued, chronic nature, may make the action of hormones more effective and may determine the location in which precancerous proliferations shall take place first. In both the development of mammary gland carcinoma and of vaginal and cervical carcinoma, the cancerous growth is preceded by a preparatory growth period, the duration of which varies in accordance with the hereditary tendency to the development of cancer in these strains; and gradually, step by step, this preparatory phase passes into the precancerous and cancerous phase as the result of the hormone action. Under normal conditions, in the course of the sexual cycle the hormone affects the epithelium as a whole of a certain organ, as it does also during the preparatory growth period. Yet, one place may be in advance of another one in these proliferative processes and one area may enter first the cancerous phase; others then follow in the course of time. Both the strength and the concentration of the hormone and the strength of the genetic tendency to cancer determine the size and number of the areas that enter the precancerous and cancerous phase; in general, the stronger the hormone action and the greater the genetically determined tendency towards cancer, the larger the individual areas and the larger the number of areas that undergo the cancerous transformation and the shorter the time required, and vice versa; if both these factors are weak, then cancer may develop in a single area, while only relatively slight growth processes take place elsewhere. Gardner found, grossly, nodules consisting of proliferating mammary gland tissue more frequently in some high tumor rate strains than in a low tumor rate strain, and he believes that tumors originate from these nodules; the latter develop therefore during the preparatory growth period. Besides the hormone and hereditary factors, there are evidently other accessory factors which may exert an influence on the intensity of the preparatory growth period and the transitions of the latter into the cancerous phase. It is of interest, also, that the genetic constitutional factors differentiating the strains of mice affect differently the development of mammary gland carcinoma and of vaginal and cervical carcinoma; this is in accordance with the fact, stated above, that the genetic constitution of the whole organism in its interaction with cytoplasmic factors acts on each organ and tissue in a specific way. In some instances at least, it can be shown that this specificity in the tissue and organ constitution consists in a different degree of responsiveness to certain stimuli.

In addition to the direct stimulation by hormones and the genetic constitution, there is still another factor which plays a role in the transformation of the normal mammary gland tissue into cancerous tissue. As mentioned above, we found that the so-called spontaneous development of mammary gland carcinoma in mice, as a rule, takes place more frequently in breeding than in nonbreeding mice; but that the intensity with which this condition acts varies greatly in different strains. In some strains, the difference in the breast cancer incidence between breeding and nonbreeding mice is very great, in others less great, and we even found one strain in which nonbreeding mice seemed to develop this cancer more readily than breeding mice. Similar differences in the effects of breeding were noted in the very closely inbred strains.

As to the manner in which breeding affects the development of mammary gland carcinoma, several possibilities exist. In the first place there is, during pregnancy, a

very strong action of hormones, which induces a marked though temporary growth of the mammary gland tissue. The cancerigenic effect of pregnancy may therefore really represent merely an intensified hormone action. Secondly, it is possible that during pregnancy an intensified production of the milk factor takes place, and lastly, it is possible that a combination of both of these effects occurs. The experimental evidence which exists does not yet enable us to decide which of these possible interpretations is the correct one.

Experiments by ourselves and Suntzeff have shown that a stimulation of the mammary gland by large amounts of estrogen during the later stages of pregnancy does not tend to lead to a direct transformation of this organ into carcinoma, although the mammary gland proliferates very strongly during pregnancy, nor does the great increase in the number of cells on which estrogen now is able to act make more easy a rapid transformation of these cells into cancer.

However, in accordance with what has been established as to the influence of long-continued stimulation of mammary gland tissue by hormones, it may perhaps be assumed that during pregnancy, as the result of repeated proliferation of the mammary gland tissue due to the strong hormone action, these cells, to a certain extent, become sensitized to the subsequent action of hormones, so that they respond to the latter more readily with the production of cancer. Bagg has made use of this effect of breeding on the development of mammary gland cancer, and in mice and rats has increased the incidence of cancer by forced breeding. Such an effect was especially evident if he used hybrids between high and low breast cancer rate strains, in which hybrids the cancer rate was low but in which it may be assumed that there was still preserved the genetic potentiality to the development of cancer, which the activating effect of hormones then set into motion.

As to the mode of interaction between the various factors, genetic constitution, hormones and milk factor, in association with secondary factors such as the effect of breeding, all of these are of importance in the production of mammary gland cancer. Without the presence of the genetic factors or the milk factor, hormones are ineffective; without the presence of the milk factor, the genetic factors and hormones are ineffective. But likewise, without the action of hormones, a favorable genetic constitution and the presence of a strong milk factor are unable to produce cancer. All of these factors are necessary for the cancerous transformation of the mammary gland tissue and all are causes of mammary gland carcinoma, although each may function in its own specific way in producing this effect.

But in addition, there are indications that there exists a quantitative relation in the interaction of the three factors mentioned. As already stated, the stronger the genetic constitution in coöperation with the milk factor, the more readily can the cancerous transformation be accomplished under the influence of hormones acting as stimulating factors. But there is, likewise, some evidence that an increase in the amount of hormones acting in association with the milk factor may overcome to a certain extent an unfavorable genetic constitution. By these means, Twombley was able to produce mammary gland carcinoma in a considerable number of C57 male mice, in which the degree of genetic tendency to breast cancer is very low. This relationship between the various factors, in a general way, although presumably not in an exact quantitative manner, has been expressed by us in the equation H (hereditary or, more generally, constitutional factors) X S (stimulating factors) = C (cancer). This equation seems to apply to various kinds of cancers in various species. To cite an example from human tumors: it is possible to produce cancer through the application of light or ultraviolet rays. But under normal conditions and in normal persons this result is achieved only in a small minority of all individuals exposed to these rays and it requires a relatively long period of time in order to accomplish it. There is, however, a condition, known as xeroderma pigmentosum, in which the skin is genetically sensitized to the production of cancer under the in-

fluence of light or ultraviolet rays; in persons affected by this abnormality, the cancerous transformation occurs quite regularly.

The causes of human cancer are similar to those established experimentally in animal cancer. Human cancer also results from the quantitative interactions of genetic and stimulating factors. We must assume that in certain cases, especially in cancers which develop in early life—often on the basis of embryonal malformations—genetic factors cause a sensitization of certain tissues, which are transformed into cancer either by environmental stimulating factors or even by ordinary metabolic conditions. The large majority of cancers develop, however, during later periods of life and in these types of cancers specific stimulating factors play a more important role; they first sensitize the tissues to growth stimuli and in the end accomplish the actual cancerous transformation.

(2) *The cancerigenic action of polycyclic compounds and of some other substances.* In 1914, two Japanese investigators, Yamagiwa and Ishikawa, discovered that if the ears of rabbits are painted with tar over long periods of time, at first warts, papillomas or cutaneous horns develop, and in the end true cancer of the epidermis may be produced. There seem to be, between the papillomatous stage and the stage of irreversible cancerous growth, intermediate stages in which, in the beginning, the stimulated epidermal cells invade in an abnormal manner the underlying tissues, but in which a spontaneous regression of these invasive processes sets in after cessation of the tarring. They interpreted these observations as an experimental proof that long-continued irritation may eventuate in the production of cancer. Subsequently, Tsutsui showed that the skin of the mouse responds much more readily with the production of epidermal carcinoma to the painting with tar than that of rabbits.

It was found later that the active substances present in tar which were capable of inducing cancer under these conditions were contained in the fraction which passed into the distillate at a high temperature. These investigations suggested the search for pure substances contained in the tar which might be responsible for such a cancerigenic effect. Kennaway, with his collaborators Cook, Hieger, Burrows, Hewett and Badger, Mayneord, discovered a group of cancer-producing (cancerigenic) hydrocarbons which were derivatives of anthracene and phenanthrene, and as a result of these discoveries they have greatly aided cancer research, and afterwards, Fieser, Shear, Shields Warren and others extended these investigations. The most effective ones among these substances are, in decreasing order of their effectiveness: 9:10 dimethyl-, as well as 5, 9, 10 trimethyl- 1.2-benzanthracene, 20-methylcholanthrene, 3:4-benzpyrene, and 1:2:5:6-dibenzanthracene. These substances induce cancer when they are painted on the skin of mice or injected subcutaneously at regular intervals over long periods of time, or when they are introduced in various places into the organism in the form of pellets. They cause a cancerous transformation in the organs and in tissues on which they act.

There are, in addition, a great many substances which exert a weaker effect of the same kind, and the intensity of their effectiveness is measured by the number of individuals which, when exposed to their action, become cancerous and by the length of the latent period which precedes the appearance of cancer. However, the degree of potency varies greatly with different preparations; slight alterations, such as the introduction of radicals CH_3, OH, or of the cyano groups, may cause great changes in the intensity of the action of these compounds; also, the position of these radicals is of importance. Usually, the introduction of oxy-radicals diminished the effectiveness of such substances and the 5, 6, 9, 10 position in the anthracene ring seems to be especially favorable for the production of active preparations. Thus the Kennaways, and subsequently Warren, found that among five dimethylanthracenes, the 9, 10- compound alone produced epithelioma of the skin, and adenocarcinomas of the lung were most frequent in mice treated with this substance, but none of these preparations produced sarcoma after subcutaneous injection. However, there is no gen-

eral rule concerning the factors determining the potency of these compounds which would apply to all, and which would make it possible to foretell whether a given compound will or will not be active. At the present stage of our very rudimentary knowledge of the biochemistry of the various tissues and organs, it is to be expected that we are not yet able to interpret the mode of interaction between certain substances and specific cell constituents, which will lead first to ordinary growth processes and in the end to cancer.

There are certain other substances which probably have specific cancerigenic effects, similar to those of the typical carcinogenic hydrocarbons, although they are not anthracene or phenanthrene preparations. The most important ones among them are, perhaps, 4-amino-2:3 azotoluene and 3:dimethyl-4 amino azobenzene (butter yellow), and some related preparations, which when fed to rats or injected into mice may induce carcinoma of the liver, and these as well as certain napthalene preparations may give rise not only to liver tumors, but may also induce papilloma formation and cancer in the urinary bladder. Here, likewise, the introduction of certain radicals into the molecule modifies their effectiveness very greatly.

There is a third miscellaneous group of many, quite dissimilar substances, which act in a much less specific way and which, after subcutaneous injection, may occasionally produce sarcoma, which is a cancer of connective tissue cells. In this group may be included estrogens, luteal hormones, preparations of liver substance used in the treatment of pernicious anemia, concentrated sugar solutions, and even buffered solutions of diluted hydrochloric acid. The action of these substances, as, for instance, estrogen, in the production of sarcoma is independent of their other more specific effects. They stimulate connective tissue cells with which they come into contact and in which they call forth regenerative growth processes. In this respect it is of interest that different types of cells or tissues seem to show a different degree of responsiveness to the action of such substances and it seems, moreover, that in general, connective tissue cells are less selective as to the kind of stimuli to which they respond than are various epithelial tissues. This is true at least as far as mouse tissues are concerned, and it suggests that an analysis of the differential physiology of tissues as distinct from those of organs may be a prerequisite to the more complete understanding of the origin of cancer.

It will also be necessary to analyze the difference in response of analogous tissues in different species. We have already referred to differences in distribution of various types of cancer in different strains of mice and in different species. There are also indications that in different human families and races the tendency to the development of cancers in the various organs and tissues differs. In evaluating the data concerning this problem, however, we have not infrequently to face the difficulty that environmental factors affecting individuals, families and races may differ in some essential respects and it is therefore not always a simple task to decide how far these differences should be attributed to genetically determined differential tissue and organ characteristics and how far to environmental factors. We have already mentioned some facts which demonstrate that different cancerigenic substances may affect organs and tissues in a specific manner. It has also been noted that a certain type of carcinogenic hydrocarbons affects the subcutaneous connective tissue more readily than the epidermis, while another type affects more readily the epidermis, and that the differential actions of these substances may differ in different species. Such differences may be due either to conditions which determine the retention, modifications or elimination of these various substances in different parts of the organism, or to specific chemical interactions between certain substances and tissues or cells. The analysis of these conditions is complicated by the fact that the results obtained may vary also with the concentrations of the substances applied. Moreover, the latent period in the action of different substances varied greatly, and apparently negative results may be due to the fact that the life span of the animals used was too

short to obtain visible results, although more subtle changes may have taken place; and lastly, the medium in which the various substances are dissolved or suspended may be of some importance (Peacock and Beck, Morton and Mider, Stowell and Cramer, Dickens and Weil-Malherbe). It has been maintained that sesame oil, which has been mentioned as a solvent for polycyclic cancerigenic substances, may, in rare cases, itself induce cancer. The term, "co-carcinogenic substances," has been used to signify the action of media which intensify the carcinogenic action of these polycyclic compounds.

(3) *The cancerigenic effects of certain metazoan parasites.* It has been observed that cancers in some organs may be caused by metazoan parasites. Thus Bilharzia (Schistosoma) is known to cause carcinoma of the urinary bladder; a preparatory growth period, in which papillomas may form, precedes this cancerous change. In rats and mice, Fibiger (1913, 1920) found that the nematode Spiroptera neoplastica (Gongylonema neoplasticum), which uses the cockroach as an intermediate host, may induce first papilloma and subsequently also carcinoma of the forestomach; this condition again is preceded by a preparatory growth period, in which hypertrophic changes occur in the mucosa. However, it has recently been contended by Cramer and others that in these experiments a true cancerous state has not yet been reached. On the other hand, there is no doubt that the cysticercus of Taenia crassicollis, which infects the liver of rats, may give rise to the development of sarcomas in the wall of the parasite-containing cysts (Bullock, Curtis and Dunning). The factor responsible for such cancerigenic effects is not known; but it seems probable that chemical substances are given off by these parasites, which exert a growth-stimulating action on the neighboring tissues.

(4) *The cancerigenic action of physical agents.* There are also physical agents which may induce cancer but which act in a less specific manner; these usually need a long preparatory period and various tissues on which they happen to act may be thus affected. Among these agents are radiations, such as Roentgen, gamma and beta rays, the latter two given off by radium; also ultraviolet rays may exert this effect. In the large majority of cases it is those parts of the epidermis, which have been exposed to these radiations, which become cancerous, but in some instances, sarcoma has developed. It is interesting that almost all human cancers due to sunlight are epidermal, because very little radiation penetrates the human skin to the dermis. On the other hand, in the albino mouse a large proportion of radiation is not absorbed by the epidermis but it penetrates to the cutis, where it causes sarcoma formation; furthermore, only wave lengths shorter than 3200 Angstroem have cancerigenic action. In guinea pigs, the injection of radium emanation has caused the appearance of periosteal sarcomas. However, the cancerigenic effect does not depend only upon the character of the radiations and their absorption by different tissues, but in addition the responsiveness of these tissues seems to vary greatly. Thus, in rats the skin responds much more readily with cancer formation to the action of ultraviolet rays than that of rabbits, although the structure of the epidermis is similar in these two species (Hueper). Furth and Butterworth, through application of large doses of Roentgen rays, produced in the ovaries of mice tumor-like new formations of the granulosa and of the germinal epithelium. Even often repeated mechanical injury of the epithelium may in the end lead to cancer formation. This has been observed to occur in the oral cavity of man, as well as in other parts of the body, and this effect is probably, in the majority of cases, due to the long-continued regenerative growth processes which follow chronic ulcerations. As to the mode of action of the radiations, it has been assumed that, primarily, this consists in setting free, or in modifying certain substances, which then cause the cancerous transformation of the tissues with which they are in contact. In the case of ultraviolet rays Roffo has attributed great significance to the action of cholesterol; however, no definite experimental proof for this hypothesis has been given.

(5) *The mode of action of cancerigenic substances.* In contrast to the estrogenic substances, the carcinogenic hydrocarbons induce cancer apparently in all, or in almost all, the cells with which they are in contact for a sufficient length of time, but, as stated, the effectiveness of different substances in this respect differs greatly. Their action may take place either locally in directly adjoining tissue, or at a distance. Some of these substances when injected subcutaneously may cause, in mice, leukemia and allied conditions (Furth and others), or tumors in the lung (Lynch), and it seems that certain strains are more prone to react in this way than others. The substances causing, primarily, carcinoma of the liver likewise are not entirely selective as to the kind of epithelium which they transform into cancerous tissue; they act on cells as unlike as liver cells and bile duct epithelium on the one hand, and surface epithelium of the bladder on the other; the effect which they exert depends probably on the length of time during which, in the process of excretion, they are in contact with different tissues. In the lung, for instance, cancer is produced by particles of carcinogenic hydrocarbons, which, after intravenous injection, are retained in the capillaries of this organ (Shimkin). The action of estrogenic substances is different; it takes place in those distant tissues to which they are physiologically and biochemically adapted. Owing to this relationship between the chemical constitution of estrogenic compounds and the receptive cells, these compounds exert both estrogenic and cancerigenic effects. Their cancerigenic function evidently depends upon their estrogenic effectiveness; the natural as well as the experimentally produced estrogens—as, for instance, stilbestrol (Lacassagne)—may differ in their chemical constitution from the typical carcinogenic hydrocarbons; still, they produce strong cancerigenic effects on secondary sex organs owing to the physiological functions which they exert.

As to the mechanism by means of which these substances bring about the transformation of normal into cancerous tissues, the following data are of importance. In the case of hormones we have seen that, step by step, they stimulate the mammary gland tissue to increased proliferation, until in the end the limit of normal growth intensity has been reached; at that point, further stimulation produces changes exceeding those seen under normal conditions and transitions into the cancerous equilibrium take place. These intensified growth processes gradually gain in momentum; they may at first be limited to a restricted area, but more and more they affect wider areas, until in the end large parts of the gland are on their way towards the cancerous transformation. The advance towards this goal takes place the more rapidly, the more the organism on which the stimuli act is genetically predisposed to this change; in other words, in proportion to the degree of genetic tendencies towards effective cancerization, the mammary gland responds more readily with growth processes to the hormonal stimuli.

This applies not only in the case of mammary gland, but it applies equally in the case of vagina and cervix; here, also, wider areas of the epithelium become stimulated, undergo hyperplasia and hypertrophy, and grow downward invasively into the underlying connective tissue and into vascular channels. However, the genetic tendencies differ in these two types of tissue and stronger stimuli are probably required in order to produce the cancerous transformation of the vaginal and cervical epithelium than that of the mammary gland tissue.

If we compare with the cancerigenic actions of these hormones those of the carcinogenic hydrocarbons, the principles seem to be similar. Differences which may exist apparently depend on the fact that while hormones are normal products of the organism and the tissues on which they act are adapted to these substances, the carcinogenic hydrocarbons are strange to the organism into which they are introduced and to its tissues. While some of the latter substances may cause the cancerous transformation of tissues with which they are kept in contact with greater intensity and rapidity than hormones do in the case of the sex organs,

they may also have secondary toxic effects on the tissues and on the organism as a whole. In the skin, these carcinogenic hydrocarbons may call forth toxic changes in the underlying stroma, and even in the epithelium. To these toxic conditions may be due certain modes of reactions in the epidermis, such as a temporary inhibition of mitotic cell division, or the greater effectiveness of fewer rather than more frequent applications of these compounds to the epidermis (Cramer). Furthermore, after administration of these carcinogenic hydrocarbons, either by means of subcutaneous injection or in the form of pellets, the sarcomatous transformation of the connective tissue cells usually is found at some distance from their strongest concentration, an occurrence which suggests that toxic effects predominate in the immediate vicinity of these substances, while at some distance a stimulation of proliferative processes takes place. The loss in weight of animals in these experiments is also in all probability due to secondary, nonspecific processes. The organism into which the hydrocarbons have been introduced uses sulfur-containing amino acids for the purpose of detoxification and the body is thus deprived of some of its necessary building stones; but it is possible that, in addition, still other injurious mechanisms, which result in harmful metabolic changes, come into play. There is no indication that the injurious action of such substances parallel their growth-stimulating effects and, as stated, it is not probable that the injurious effects are causally connected with their growth-promoting and cancerigenic action. If these compounds in solutions are painted on the skin at regular intervals there occurs, preceding the invasion of the underlying connective tissue, a period of hypertrophy and hyperplasia, with increase in the number of mitotic cell divisions in the epidermis. This increase in the number of mitoses sets in soon after the beginning of the applications of the carcinogenic hydrocarbons and it probably passes into the cancerous period. However, as Champy and Vasilin found in tarred mice, the mitotic curve is not an evenly rising one, but there are periods when the number of mitoses is lowered and, more recently, Cooper and Reller noted in mice to whose skin methylcholanthrene was applied, a marked increase of mitoses during the first 37 days, which was followed by a relative decrease in the number of dividing cells between the 37th and 65th days, and still later by another period of rapid increase in mitotic cell proliferation. The exact significance of these variations is not yet known; however, we find also in the case of some other growth stimulants, acting on normal tissues, that a positive phase of stimulation may be followed by a negative one; moreover, the level of mitoses under the influence of methylcholanthrene does not seem to return to the low level of mitotic activity, as it is characteristic of normal tissue, but it remains always elevated. As in all growing plant and animal tissues, the turgor of the cells and nuclei increases and the viscosity of the nuclei decreases after application of methylcholanthrene in tissues which are still in a stage preceding cancer; this has been shown by centrifuging these cells according to the method of Heilbrunn (Cowdry and Paletta). The same effects can be observed in the epidermal cells during the process of wound healing, but when Guyer and Claus subjected to centrifugation cells which had already become cancerous, they found an increased viscosity of cytoplasm as well as of the nuclear contents. Other effects of a marked growth stimulation, some of which may have passed already the threshold of normality, such as unequal increase in size of nucleus and cytoplasm in different cells, abnormalities in nuclear and cell division, may be hastened and intensified if strongly-acting hydrocarbons are used. In the course of the hypertrophy and hyperplasia of the epidermis which take place during wound healing, amitotic division of nuclei frequently occurs in cells which are disadvantageously situated as regards the supply of oxygen or other needed substances. Such amitotic divisions are observed likewise after application of these cancerigenic hydrocarbons. We may then conclude that, as in the case of hormone-cancer of the mammary gland and of other tissues, the production of epidermal carcinoma, caused by hydrocarbons, is preceded by a period of preparatory

growth, in which intense stimulation leads to certain structural abnormalities. A similar sequence of events is found in the small intestine (Stewart). Here, also, the growth processes of the preparatory period become abnormal at a relatively early stage. There develops an increased number of gland cells, the nuclei increase in size and become hyperchromatic. Atypical polarization of some cells is noted. These changes may accompany the invasion of the underlying tissue under the influence of active carcinogenic hydrocarbons. The probable occurrence of a preparatory growth period was likewise foreseen by the writer in the development of sarcoma following subcutaneous application of these hydrocarbons, and it was actually observed by des Ligneris. In this instance it consists in the formation of a granulation tissue in which the number of mitoses is increased. There is an increase in the number and size of cells, and abnormalities in the nucleus and cytoplasm are found. However, while under the influence of non-carcinogenic substances this granulation tissue after some time regresses, it persists and gradually changes into cancer if strong and long-continued growth stimuli are applied.

Also, in the liver the feeding of butter yellow or related substances seems to cause increased mitotic activity in the liver cells, leading at first to a nodular hyperplasia and an adenomatous proliferation, which at last changes into carcinoma. Degenerative processes, followed by cirrhosis in the liver, which may be observed in this condition, are due probably to the secondary toxic action of these substances; they do not seem to be essential in the process of cancerization and are, as a rule, not found in spontaneous hepatoma in certain strains of mice, although in spontaneous carcinoma of the liver in man the regenerative processes preceding or associated with cirrhosis may aid in or lead to the production of cancer in this organ. As we have mentioned already, cirrhosis of the liver is common if butter yellow is used as a carcinogenic substance, but occurs more readily after application of *o*-aminoazotoluene.

There is, then, very strong evidence that in all these cases a preparatory growth period precedes the development of cancer, and that this period is of significance in the cancerous transformation. We see also that if the growth-promoting stimuli affect a tissue as a whole, instead of being localized at a certain point, large areas may become changed into cancer. Such a process, however, does not need to take place everywhere at exactly the same time; more or less accidental factors may accelerate it in some places and retard it in others. We have found this tendency towards a generalized transformation into cancer in the mammary gland, as well as in vagina and cervix of mice, and it occurs also in the liver of mice. Similarly, in several types of human cancer such a condition has been noted.

These observations make it very probable that the development of cancer under the influence of various substances is due to the long stimulation of growth processes in the cells and tissues on which these substances act. A further argument in favor of this interpretation is furnished by the evidence that hormones which aid in the production of cancer, either stimulate certain tissues directly or bring about an increase in the amount of such hormones, or otherwise promote their action, while hormones which interfere with the cancerigenic effect of a hormone, interfere with the growth-stimulating effect of the latter.

This interpretation is also supported by experiments in which it was shown that appropriate concentrations of carcinogenic hydrocarbons induce in various freeliving unicellular organisms and in cultures of tissues of higher organisms, an accelerated reproduction of cells (Goldstein, Wolman, Hearne-Creech, and others). Similarly, Hammett and Reimann found that these hydrocarbons caused a stimulation of cell proliferation and development in the hydroid Obelia geniculata, and Owen, Weiss and Price observed that dibenzanthracene stimulated regeneration of cut segments of Euplanaria dorotocephala. Various authors (Pourbaix, Cook, Hare and Joly, Wright and Anderson), in corresponding experiments, induced changes in the carbohydrate

metabolism similar to those found in many types of cancer. However, while Spencer and Melroy observed similar stimulating effects of carcinogenic hydrocarbons on on Paramecium, they do not believe that these effects are specific for carcinogenic compounds; also, other substances may exert such effects. It is known, of course, that many substances which are toxic if used in a certain concentration, may exert a stimulating action in weaker concentrations. But these investigators found that individual paramecia, when exposed continuously to methylcholanthrene for more than a year, acquired a greater resistance to the action of an unfavorable medium, to which non-adapted organisms succumbed very readily. So far, the long-continued addition of non-cancerigenic substances to the culture medium has not produced such an adaptation. These latter types of experiments would be more in accord with the fact that also in experimental cancerization the stimuli have to be applied over extended periods of time, and it is only as a result of a prolonged change in conditions that permanent alterations of the cells exposed to these stimuli take place. However, in prolonging these experiments Spencer and Melroy observed that still longer-continued, uninterrupted application of the various agents which at first may stimulate, or at least have no injurious effect, at last begin to exert an injurious effect on these organisms. This took place with bacteria and paramecia, as well as with worms; moreover, it was not limited to carcinogenic substances, such as methylcholanthrene, but applied also to non-carcinogenic hydrocarbons, such as phenanthrene and other agents, *e.g.*, radium. These investigators suggest that such an injurious effect might perhaps be avoided by using these substances rhythmically instead of continuously, a suggestion which is in harmony with certain conceptions concerning the origin of cancer developed some years ago by the writer, to which we shall refer later.

As we have seen, the cancerigenic response of the mammary gland to the endogenous ovarian hormones differs greatly in different strains of mice (Loeb and Lathrop), and the readiness with which the strains react to the injection of estrogenic hormones with the formation of mammary gland carcinoma varies in a corresponding way (Lacassagne; Loeb and Suntzeff). These differences are due to genetic differences in the constitution of these animals. For similar reasons, the cancerigenic response to tar and carcinogenic hydrocarbons varies in different species. But the genetic factors which determine the responsiveness of an organ to stimulation by estrogen, on the one hand, and by tar or by a carcinogenic hydrocarbon, on the other, do not seem to be the same; likewise, the genetic basis for the response of different tissues to the same stimulant (*e.g.*, methylcholanthrene) differs in different inbred strains and species. However, as a rule the animals in which cancer develops relatively quickly under the influence of these carcinogenic compounds pass more rapidly through the preparatory growth period, and if these substances are applied to the skin, warts appear more readily in these animals than in animals in which the cancerigenic reaction takes place more slowly. In general, it has been found that if a strain has a genetic tendency to the spontaneous development of a certain type of cancer, it also responds more readily to experimental stimulation with the formation of cancer of the same type. Thus, according to Lynch, those strains of mice in which there is a greater tendency to the spontaneous development of pulmonary carcinoma, react also more readily to the application of tar with the production of this tumor. Indications that the same condition holds good also in other tumors have been obtained more recently by Strong in various substrains of his mixed NH strain. It appears that treatment with a carcinogenic compound called forth, in each substrain, the development of a type of tumor to which it is genetically predisposed, and which, under favorable conditions, would therefore have a tendency to develop spontaneously. Similarly, Engelbreth-Holm noted that by application of 9, 10, dimethyl-1, 2 benzanthracene, leukemia could be more readily induced in those strains of mice in which spontaneous leukemia was a frequent occurrence. These findings agree

with what has been established in the case of mammary gland carcinoma. Here, also, hormones are most effective in producing this type of tumor in strains of mice in which it most readily develops spontaneously, that is, under the influence of hormones given off in the course of the normal function of the sex glands. This takes place in strains which have a genetic constitution favoring this mode of reaction of the mammary gland. In all these respects there is, therefore, a complete parallelism between the action of hormones and that of tar and the carcinogenic compounds; both of these types of cancerigenic substances coöperate in a similar manner with specific inherited factors in the production of cancer. This similarity is also noted in the effect of the time factor on the production of cancer, if ovarian hormones and hydrocarbons are compared. In the case of ovarian hormones, it follows from our earlier experiments that mammary gland carcinoma appears the more frequently and the more rapidly, the longer the time was during which the ovarian hormones have acted. In the same way, Burns and Schenken found a corresponding relation between the length of time during which estrogen was injected into mice and the frequency with which mammary gland carcinoma developed. In the case of carcinogenic hydrocarbons, Andervont and Shear noted that the percentage of sarcomas which formed after subcutaneous introduction of pellets of methylcholanthrene was the greater the longer the pellets had had a chance to act on the connective tissue.

In various strains of mice it can be shown that very young animals respond much more readily with mammary gland carcinoma to the administration of estrogenic hormones than do somewhat older animals, according to some as yet unpublished experiments of Loeb and Suntzeff, as well as of Burns. A similar observation has been made in the induction of leukosis by means of carcinogenic compounds (Morton and Mider; Furth and Barnes). The greater responsiveness of young organisms to cancerigenic stimuli may perhaps be merely an example of the greater intensity of growth processes in general in the various organs and tissues of higher organisms in early life, and of the gradual decrease of this activity with advancing age.

We have referred already to the fact that the development of mammary gland carcinoma is initiated by a substance which is produced in the normal organism in the course of the normal function of the female sex organs with the coöperation of the proper genetic constitution and the so-called milk factor. Subsequently, when it was found that, similarly to tar, various polycyclic hydrocarbons could induce cancer in different organs, numerous attempts were made to prove that these substances may cause also spontaneous cancers and that they may therefore appear spontaneously in the body and may here exert a function leading to abnormal growth processes. If it could be proved that within the organism carcinogenic hydrocarbons may develop, under certain conditions, from normal metabolic products, the conclusion might be justified that they could in some instances be the hidden cause of spontaneous cancer. Such a possibility suggested itself in view of the close chemical relationship between bile acids and cholesterol on the one hand, and these carcinogenic compounds on the other. Moreover, it has been possible in vitro to prepare methylcholanthrene from desoxycholic acid, a constituent of bile. However, there is no indication that a similar process occurs in the organism. Desoxycholic acid itself does not seem to be cancerigenic, although some investigators have reported the production of sarcoma in some rats and mice after subcutaneous injection of this substance; but in this connection the fact must be recalled that the sarcomatous transformation of the fibroblasts, following subcutaneous injection, may be caused by substances with very diverse chemical constitutions, and the experiments with desoxycholic acid cannot, therefore, be used as a proof that these hydrocarbons may induce spontaneous carcinoma formation in various organs.

Another indirect method to demonstrate the significance of these substances in the origin of spontaneous tumors has been used by Shabad and various other investigators. Shabad and his collaborators, and subsequently Hieger, des Ligneris,

and Steiner, could show that extracts of human liver and lung, made with benzene or with similar lipoid dissolving media, if painted on the skin of mice or injected subcutaneously, induced in some cases the formation of squamous cell carcinoma or sarcoma in the place of application or of tumors in various distant organs. But while Shabad and his associates used benzene extracts of liver, des Ligneris, Steiner and others found only the non-saponifiable lipoid fraction effective. There is also some difference in the results obtained by different investigators as far as the location of the cancers which develop under these conditions is concerned, some observing merely a local effect, while others noted cancer formation also in distant organs. Furthermore, it was observed that the liver of persons suffering from cancer in other organs was more potent in this respect than the liver from noncancerous persons. Des Ligneris, and subsequently Hieger, found that petroleum ether extracts of liver from cancerous, and also from noncancerous Bantus, especially those of Portuguese East African descent, gave rise to the development of cancers in mice whose skin had been painted or which had received subcutaneous injections; however, only certain liver extracts were actively carcinogenic. It is known that in these, as well as in some other African races, carcinoma of the liver is a frequent type of cancer. On the whole, the number of experimental cancers thus produced in mice was relatively small and the number of substances which may produce a small number of cancers under such conditions is very great. These experiments do not, therefore, furnish a definite proof that carcinogenic substances may develop in the organism spontaneously and induce the formation of cancer.

Somewhere on the curve representing the change from the normal to the cancerous equilibrium there is a point where the mammary gland tissue becomes sensitized to growth stimuli, so that from then on the specific hormonal stimuli no longer are needed for the progression of growth, but nonspecific extraneous or the ordinary intrinsic stimuli are able to accomplish the cancerous transformation. During the process of cancerization of the mammary gland under the influence of ovarian hormones, we noticed (1919) that if the hormone action was removed at an early period of life, the appearance of mammary gland carcinoma was prevented, but if the ovarian action had extended over a somewhat longer period, its appearance was merely retarded; it could develop a long time after the factors which were responsible for the cancer formation had ceased to act. We interpreted this relationship as indicating a double action of the removal of the ovary, and, by implication, also a double action of the ovarian hormones, namely (1) one of prevention of cancer as well as the prevention of sensitization to cancer if hormones had acted only for a short time, or, on the contrary, the production of cancer at the usual rate as the direct result of the action of hormones extending over a sufficient length of time; and (2) an action of hormones of intermediate duration, which made it possible for the mammary gland to develop cancer a long time after removal of the source of the hormones. We concluded furthermore (1920) that the first tissue changes, which eventuate in the development of cancer, occur at a much earlier period of life than might have been expected, and that castration affects these primary tissue changes rather than the secondary transformation of these changed tissues into fully developed cancer. We distinguished, therefore, between two conditions: namely, (1) the preparatory tissue changes, and (2) the change of the prepared tissues into cancer. Subsequently, we interpreted the preparatory changes as constituting a sensitization comparable to that caused by the hormone of the corpus luteum, which produces a state of sensitization in the uterus. After the uterine mucosa has been sensitized by the specific action of a hormone, much less specific actions, such as incisions or other kinds of physical changes setting to work certain growth processes, lead to the production of placentomata, which may be regarded as temporary tumors. We considered, therefore, the mammary gland as being sensitized primarily by genetic factors, but a secondary, similar preparatory or sensitizing action is accom-

plished by the ovarian hormones, in particular, by estrogen, in such a way that subsequently stimuli of a nonspecific nature, namely, some stimuli acting on the tissues during ordinary life processes, were able to induce cancer formation. It may then be stated that castration at an intermediate or relatively late period prevents a direct production of cancer but does not prevent sensitization, which leads to a later formation of cancer under the influence of nonspecific stimuli.

Through the work of a number of investigators it was then shown that the same condition exists also in cancer induced experimentally by tar and by carcinogenic hydrocarbons. Leitch observed (1922) that after tar had acted for a certain length of time, the tissues had changed in such a fashion that cancer could develop a long time after this action had ceased. He concluded that a considerable time before cancer had begun to develop, a change had occurred in the tissue exposed to the tar action, which was not visible under the microscope. Similar observations were made at about the same time by Bang. In tar cancer, induced in mice, Bang (1922) divided the time which elapsed between the first tarring and the occurrence of invasive growth into a preparatory period, which lasted for several months of the treatment, until the latter had accomplished the potentiality to the cancerous change, and a latent period which began at the end of the preparatory period and extended to the beginning of the invasive cancerous growth, and in which the application of tar was no longer needed. In this paper, the term "preparatory period" includes the entire time from the beginning of the action of the cancerigenic agents to the transformation into cancer. Subsequently it was shown that incisions in previously tarred tissue could induce cancer formation (Döderlein and Lipschütz, Deelman, des Ligneris), or that a second tarring brought out papilloma formation sooner than the first (des Ligneris), and that a third and a fourth tarring was still more effective and induced papilloma formation in places in which the first tarring had been inefficient (Rous and Kidd). Twort and Twort, des Ligneris, Rous, Kidd and MacKenzie, as well as Berenblum, noted that after a first period of tarring or application or carcinogenic hydrocarbons, the subsequent application of less specific irritants or stimulating substances, such as liquid paraffin, oleic acid, turpentine, croton resin, was able to induce the formation of papilloma, or even of cancer. Another interesting related observation was made by Burrows, Mayneord and Roberts, who found that granulomatous proliferation caused by kaolin could be converted into sarcoma by the application of Roentgen rays. We may assume that the granulomatous cells were in a sensitized state tending towards tumor formation, which was activated by other agents. It may then be concluded that cancerigenic substances and also physical factors may act in a double way: 1) they may induce cancer, and 2) there is associated with and preceding this process one of sensitization, which causes the process leading towards the production of cancer to become more and more intensified if the same cancerigenic agent continues to act, or which makes it possible for factors, which ordinarily are not cancerigenic, to exert such an effect. However, there does not need to be a marked difference in principle between the action of the specific cancerigenic factors and the others which are less so or nonspecific. The difference is probably merely a quantitative one, the latter substance requiring an action extending over a much longer time in order to accomplish what a more specific substance can accomplish in a shorter time.

It follows from the experiments of Rous, Kidd and Friedewald that warts and carcinomatoid downgrowths caused by tar or methylcholanthrene in the rabbit's skin behave like sensitized non-papillomatous skin, or like mammary gland tissue sensitized by ovarian hormones in the mouse. The sensitization has occurred probably at some time previous to the development of these newformations. Of great interest is the marked difference which they find in the behavior of these papillomas in different rabbits. In some, the sensitization is quite evident; after cessation of tarring, the warts continue to grow, and, in rare cases, to become cancerous. In other in-

stances, continued tarring is required in order to maintain them, and without it they would regress; in still others, they regress even if tarring is continued. But although the warts or precancerous downgrowths (carcinomatoids) have regressed and the epidermis has apparently returned to about the normal state, these cells still remain sensitized and can be re-awakened to abnormal growth by the use of various activators; even cells that formerly had not been a part of the warts will now show sensitization and develop into warty outgrowths, presumably because they had previously been affected to a certain degree by the tar or carcinogenic hydrocarbons, although perhaps not with the same intensity as some of the neighboring cells.

The great variation which different strains of individuals manifest in their reactions to these stimulating cancerigenic factors corresponds to the great difference in the reactivity of the mammary gland to endogenous or extraneous estrogenic hormones which has been established in different strains of mice. Both are due largely to genetic differences in the constitution of these strains or individuals. This constitution may represent the primary and predetermined type of sensitization of the tissues, a sensitization of the first order; as we have seen, it determines the intensity of stimulation which is needed to induce a response of a certain magnitude, in accordance with the relation expressed in the equation $H \times S = C$. In curves representing the response of the tissue to various kinds or various intensities of stimulation, this genetic sensitization could be expressed as the degree of elevation of the starting point of the curve above the level of the baseline. The curve of growth response of the tissue could then be represented either as a straight line which turns sharply upwards at the point of sensitization, in case the latter process consisted in a single abrupt change. On the other hand, the possibility exists that the process of sensitization is not a single change, but that it consists of a succession of sensitizations in which the individual steps may be very small. In this case the response of the tissues to new stimulation would become, step by step, more intense and the curve would rise gradually upwards. There are certain indications which favor the second interpretation. If we follow the development of mammary gland carcinoma under the influence of estrogen, it seems that the progress of tissue proliferation is at first very slow and gradually becomes more and more intense. Other observations, however, merely make possible the conclusion that during the preparatory period the growth momentum of the affected tissues gradually increases, until at last the cancerous state is reached. Thus, with the continued administration of estrogen the incidence of mammary gland carcinoma increases step by step with the increasing length of time during which ovarian hormones have exerted their growth stimulation. To a similar conclusion point the observations of Andervont and Shear concerning the effect of methylcholanthrene; here also, the incidence of cancer increased with the increasing length of time during which this substance acted. In the same way, the benign formation, such as papilloma and adenoma which may develop preceding the formation of cancer, may be interpreted as intermediate steps towards cancerization. We may then conceive of the action of the stimulating factors as increasing, step by step, the growth momentum of the tissue on which they act, until in the end it has reached such an intensity that from this point on even weaker stimulations are able to carry the growth to the cancerous stage. It seems probable that different quantities of an autocatalytic growth substance correspond to the different intensities of the growth momentum, and it seems, moreover, probable that this process of gain in growth momentum continues after the point has been reached when nonspecific growth factors are able to complete the cancerous transformation of the sensitized tissue. In accordance with this interpretation, we found in our earlier experiments indications that with the increasing intensity of the hormone action, the time necessary for the appearance of cancerous growth is shortened; and other experiments, to which we shall refer, agree with this conclusion. We must further assume that in different individuals belonging to the same inbred

strain, either different intensities of growth momentum are necessary to carry the tissue into the cancerous phase, or that the growth momentum is the same in all individuals, but that the amount of stimulation needed to achieve this critical growth momentum and to maintain it varies in different individuals. This question, as well as the shape of the growth curve of the stimulated tissues during the preparatory period are problems which future investigations must solve. In accordance with this assumption of a step-by-step process in the degree of sensitization is also the observation of Rous and Kidd that a third and a fourth tarring of the skin accelerates the sequence of tissue responses, ultimately leading to cancer, more than a second tarring.

There are, furthermore, indications that even the process of cancerization consists of a series of steady changes, rather than of a single sudden change which is not followed by further progress. Accordingly, we observed that in many cases in the course of the first transplantations of a tumor the growth intensity increases. These gradations may be taken to represent different intensities of cancerization. Similarly, metastases of human cancer may grow more actively than the original tumor. Mechanical stimulation may likewise exert a stimulation on tumor cells. In our early experiments we noted, as already stated, that a cancer which, as a result of unfavorable conditions had begun to grow slowly, may be made to grow more actively under the influence of mechanical stimulation, and in the course of time, following repeated transplantations, which also act as stimulating factors, may assume a more rapid growth. In the same sense may be interpreted the findings of Rous and his collaborators that certain stimuli, in particular those exerted by papilloma virus, cause a more intense and varied cancerous response in the skin of the rabbit than does tar. On former occasions we have interpreted in this way, also, the fact that the stage of irreversible cancerous growth is preceded by a series of conditions in which at first there exists still the chance of regression, but apparently to a different degree in different instances. Of interest in this connection are also the experiments of Emmart, who recognized different stages in the process of cancerization of the liver induced in rats by feeding *o*-aminoazotoluene. In the tumor phase designated as hepatoma, metastases may occur; the structure of the tissue differs from that of normal liver tissue and there is increased cell proliferation; but this may be followed by a more advanced process of cancerization, in which the structural variations between the different tumor cells become pronounced, the growth intensity greater, and deviations from the normal growth more marked; also, neighboring areas may be invaded. The increase in growth energy in the latter phase is also indicated by a greater ability to survive in tissue culture, as compared to that shown by the hepatoma phase. We would then have to conceive of the process of sensitization and cancerization as a continuous series of short, progressive, step-by-step changes, in which, however, a discontinuous alteration takes place when a certain threshold of stimulation has been reached; at this point the change becomes irreversible; this readiness with which the threshold point is reached would vary in proportion to the primary genetic sensitization of the tissues, and, perhaps, also in proportion to certain other non-genetic factors. That such sensitizations of tissues exist also in the human species is indicated by the experience that certain industrial cancers may occur many years after a person has been removed from the original cancer-inducing environment. Furthermore, in conditions such as xeroderma pigmentosum there exists a sensitized state of certain parts of the skin; if these are exposed to the rays of the sun, cancer may develop.

Under normal circumstances, a stimulation of a tissue is followed by a reaction which can be represented by a curve which rises in a first period and then descends again to the baseline. It differs therefore from the curve representing the process of sensitization and cancerization. Similar also is the curve seen in certain rabbits in which tarring causes only a temporary papilloma formation and is followed by an apparent return to the normal state, according to the observations of Yamagiwa and

Ishikawa and of subsequent investigators. As an intermediate condition, we found in strains of mice in which the mammary gland is resistant to the action of estrogenic hormones, that the stimulation as a rule progresses so slowly that the point of cancerization is reached not at all, or only in exceptional individuals. It is possible to interpret these differences to the readiness of analogous tissues to react to various long-continued stimuli, which is noted in various species or strains of animals, as due at least in part to differences in the degree of the sensitization of the first order, which depends upon the genetic constitution of an organ or tissue in a given strain or species. Thus mice, rats, guinea pigs, rabbits and dogs differ in the readiness with which their mammary gland responds to hormonal stimulation. Similarly, rabbits and mice differ very greatly in the response of their skin to stimulation by tar, and again, some strains of mice likewise show corresponding differences.

We may here discuss briefly some further data concerning the relations of the process of sensitization to the preparatory growth period which precedes the process of cancer formation, and some of the characteristic features of the preparatory growth period. The process of sensitization takes places during this period and it is the result of the growth processes occurring in the earlier phases of the preparatory period. The more intense the growth during this preparatory phase, the more quickly it passes into the cancerous stage; this has been found in hormonal cancer and it seems equally true of cancer of the rabbit skin caused by the Shope papilloma virus, where Rous, Beard and Kidd observed that the cancer appears early in cases in which the papillomas proliferate very vigorously. Rapidity of transformation into cancer depends upon genetic constitutional factors in the mammary glands of mice as well as on the strength and efficiency of the stimulating factors, and there are indications that genetic factors may play a role also in the tar cancer of mice and rabbits. In cancer induced by hormone stimulation the preparatory period represents a continuous series of progressive growth changes. However, while in cancer of the mammary gland caused by the action of endogenous estrogenic hormone there is, in general, a relation between the frequency of mammary gland carcinoma and the rapidity of preparatory growth, in addition, peculiarities in the character of the different phases of the growth curve leading to cancer formation may distinguish various strains of mice. Likewise, the carcinogenic hydrocarbons, tar and papilloma virus may differ in the manner in which they affect different phases of the growth curve of the epidermal cells. Thus, application of an efficient tar preparation seems to cause a rapid development of papillomas, but this process is followed only farely by cancer formation, and the period of latency intervening between the papilloma and cancer development, if the latter should occur, is long. In the case of methylcholanthrene, warts do not seem to develop as rapidly as after application of tar, but the transformation of warts into cancer occurs more readily. Papilloma virus induces a rapid formation of warts, but the subsequent formation of carcinoma, if it takes place at all, is a rather slow process (Rous and collaborators).

In the case of hormone cancer, the preparatory growth period seems to be essential for the induction of cancer. In not a single instance have we observed that carcinoma of the mammary gland develops directly from resting ducts or acini, nor is it sufficient that merely a moderate mitotic proliferation occurs in these gland structures. A period must be initiated in which step-by-step, intensified growth phases follow one upon the other in succession, and apparently only in this way is the stage of cancerization ultimately reached. The function of the preparatory growth period, and in particular of the process of sensitization, is to prevent the affected tissue from returning to its initial normal resting stage at which the growth started, and to push it on to continuously intensified growth. This requirement seems to obtain, irrespective of the kind of stimulating factors which are used, whether they are hormones, carcinogenic hydrocarbons or papilloma virus. One exception to this rule is found in the production of the cancerous stage by injecting

the agent of fowl sarcoma into normal chickens. In this instance, the normal cells are changed into cancer cells in an acute way; they seem to acquire the characteristics of cancer without first passing through a series of preparatory growth processes and of corresponding tissue changes; however, it is also possible that in this case the preparatory growth process takes a very rapid course. If the first alternative should hold good, then the difference between this condition and the usual mode of initiation of the cancerous process in certain respects is similar to the difference between passive immunization through injection of antitoxin, and the process of active immunization of an organism, which enables it to produce antitoxin. However, the transfer of the agent of fowl sarcoma to other chickens does not leave the mesenchymatous cells in the new host, which become sarcomatous, otherwise unchanged; as we have mentioned already, they acquire thereby the alterations in their carbohydrate metabolism which are characteristic of the majority of cancers.

There are indications that a relation exists between the length of the preparatory growth period and the length of life of a species. Thus the preparatory growth period seems to be much longer in man than in rodents and rabbits, and such a relation between length of the preparatory growth period and duration of life of a certain species may hold good in general. However, accurate comparisons between different species, made under strictly comparable conditions, regarding this point have not yet been recorded, and conclusions are furthermore made more difficult because of the existence of differences in the cancerigenic tendency of various tissues.

It seems, then, that the mechanism of cancer formation under the influence of various stimuli is in some way connected with the occurrence of a preceding preparatory growth period, during which processes of tissue sensitization occur. These stimulating factors interact with genetically determined conditions in the tissue substrata and certain quantitative relations seem to exist between these two sets of conditions; moreover, this rule seems to apply whatever the nature of the stimuli which are active.

(c) **Virus as a Cause of Cancer.** We have seen that certain metazoic parasites may, in special instances, induce cancer formation. There was a time when the possibility that bacteria or some parasitic protozoa might be the real cause of cancer was seriously considered. In some plants, bacterium tumefaciens may produce a tumor-like formation, which in certain respects resembles cancer in higher organisms. It has also been stated by Jensen that tubercle bacilli may give origin to sarcoma. However, if this microorganism was actually involved in the development of cancer, it can only have been in an indirect manner by inducing the formation of granulation tissue, which, under the influence of unknown stimuli, changed into sarcoma. Attention was directed towards protozoa mainly because of the occurrence in certain cancers of products of cell differentiation, which resemble encysted protozoa. These interpretations as to the origin of cancerous growths in higher animals have quite generally been abandoned.

On the other hand, in recent years another concept has gained very serious consideration, namely, that all cancers are caused by viruses. This concept owes its origin primarily to the experiments of Peyton Rous. That certain virus-like substances, as well as also a typical virus, may induce cancer formation has been definitely shown by this investigator. But Rous has, in addition, suggested that all animal cancers are essentially caused by viruses, and that all the other factors that play a role in the origin of cancer are merely subsidiary factors, which act presumably by influencing the virus or by giving it a chance to function as the true cancer agent. This view is based, in the main, on certain data which we shall now briefly discuss.

As stated previously, it has been shown by Rous and his collaborators that certain fowl sarcoma are caused by agents which can be separated from the tumor cells

by various processes, the best known of which and the one most commonly employed is filtration through filters which do not permit the tumor cells to pass through its pores. There are a number of such chicken tumors caused by distinct agents, which latter differ from one another in the type of host cells to which they are adapted, as, for instance, ordinary connective tissue cells, endothelial cells, and perhaps also monocytes, and which, furthermore, exert specific morphogenic effects on certain tissues; thus, one of these agents induces cartilage and bone formation in the mesenchymatous cells, which they invade and make their habitat in which they multiply. While they are thus to a certain extent specific, by means of immune reactions it can also be shown that they are—at least some of them—more or less related to one another. Moreover, they resemble the cells in which they live in regard to the degree of species specificity which they manifest. Primarily, these agents are adapted to the life in fowl cells, but as already stated, they may become adapted also to certain tissues of pheasants and ducks; and they resemble their host cells also insofar as they adapt themselves better to different though related species, if the host organisms to which they are transferred are very young.

The typical effect of the injections of these agents into chickens consists in the rapid change of the connective tissue, or, in general, of mesenchymatous cells, into sarcoma. A preparatory growth period does not seem to be necessary for this transformation, but these agents are particularly effective if they meet mesenchymatous cells which are in the process of proliferation (Rous). It has furthermore been observed that in chickens in which, under these conditions, a sarcoma forms, there develop at an early period, immune substances in the blood, which have the power to neutralize these agents.

However, the reaction of such agents is not limited to the transformation of connective tissue cells into sarcoma, as found by Rous, and soon afterwards, by Fujinami; but as we have likewise discussed already, several years previously Ellerman had discovered that leukemia in fowl could be transferred by a cell-free filtrate to other chickens. Leukemia also may be considered as a special type of cancer, namely, one originating in the leucocyte-producing organs, the cancerous cells in this instance being peculiar insofar as they readily invade and circulate in the vascular system.

It seems, then, that cancers from which, in mammals, a filtrable agent cannot be separated, may, in birds, contain such an agent. It was therefore of interest to inquire whether cancers induced in fowl by tar also would differ from mammalian tumors of this kind by the presence of a filtrable agent in the tumor cells. Murphy and Landsteiner, who first carried out such an experiment, found this not to be the case. But more recently, Andrewes stated that after all, the existence of such an agent has been made very likely by the demonstration in the blood-serum of fowls, in which such tar cancers had been produced experimentally, of the presence of antibodies which neutralize the filtrable agent of the Rous tumor. In a similar manner, Kidd has recently observed that in papilloma of the skin of domestic rabbits, produced by the Shope papilloma virus originally obtained from warts of Western cottontail rabbits, the presence of a virus can be shown in an indirect way by the demonstration of neutralizing antibodies in the blood-serum, although the virus cannot be demonstrated directly in the extract of the papilloma itself, in contrast to the extract of the papilloma of cottontail rabbits, in which the existence of the virus can be proven by its ability to produce papilloma in other rabbits. Analogous experiments suggest the presence of a virus also in the Brown-Pearce tumor of the rabbit (Kidd), but there seems as yet to be some doubt as to the possible interpretation of these latter results. These conclusions agree with the findings of Shope that in the papilloma of domestic rabbits a specific antigen is present, which is able to immunize normal rabbits against the papilloma virus.

As to the nature of the agents in these fowl tumors, Rous believes that they are

extraneous viruses which are adapted to certain avian cells. Murphy, on the other hand, holds that they are substances produced by the host organism itself; at first, he considered them as enzyme-like products, but more recently he compared them to organizer-like substances (mutagens) not unlike those which can be extracted from certain type-specific pneumococci. Murphy's collaborator, Claude, attempted to separate the cell constituents which are bearers of the agent or which represent the agent. As stated previously, he assumes that the agent constitutes modified mitochondria, because the chemical constitution of mitochondria and of the agent appears to be similar, and chemical alterations of the mitochondria-like tissues tend to destroy also the activity of the agent. The findings of Jobling, Schemin and Sproul seem to agree to some extent with those of Claude. The unsuccessful attempts of Furth and Kabat to separate virus-bearing cell constituents from normal ones suggest the possibility that adsorption processes may introduce a complication into experiments of this kind.

The avian sarcomas and leukosis furnish, then, one line of evidence in favor of the conclusion that virus-like substances may play a role in the development of cancer. Further evidence is presented by the findings of Rous and his collaborators that the papilloma on the skin of Western cottontail rabbits, caused by a virus discovered by Shope, can be converted into cancer. Such a transformation may take place spontaneously, but it can be hastened very much by the application of tar and carcinogenic hydrocarbons, such as methylcholanthrene, to these virus papillomas. Conversely, the application of the rabbit virus to tar papillomas may very much stimulate their growth and transform them into carcinomas. The intensity of the action of tar, methylcholanthrene and the virus differs very much and the character of the carcinomatous changes brought about by these substances is not exactly the same. Likewise, as mentioned above, the time relation as to the appearance of these warts and as to their conversion into cancer differs in the case of these three substances; furthermore, the areas in the skin affected by the Shope virus in the same animal are different from those affected by tar, and in addition, individual rabbits differ in their reactions to tar and to virus.

From these facts, Rous draws the conclusion that the increase in the effect which is observed if virus, tar, or methylcholanthrene are combined, is not due to a summation of the action of these three substances, but that the virus is the essential wart-producing and cancerigenic agent, while tar or methylcholanthrene merely intensify the effect of the virus, perhaps by modifying its chemical constitution in such a way that the virulence is increased.

However, we believe it more probable that all these substances act, in principle, in the same manner; they cause first a sensitization of the tissue by intensifying its growth momentum, then activate its further growth, and finally convert it into cancer; under some conditions, a combination of these substances may be more effective than a single substance, because the time and intensity curves of these substances differ. Of interest in this connection is the observation that continued action of a stimulating substance on a tissue or organ may gradually lead to an exhaustion or reversal of the growth effects, suggesting the inhibiting action of local, or perhaps also of generalized immune processes. We and our collaborators noted such curves in the action of iodine and of anterior pituitary extracts on the thyroid gland, and also in the action of the corpus luteum on the uterus of the guinea pig. Such a gradual diminution in the effectiveness of these substances might perhaps be prevented by the use of new growth-stimulating substances, to which the tissues are not yet accommodated.

We have already mentioned the fact that if the rabbit papilloma virus is applied to the skin of domestic rabbits and here calls forth papillomas, the virus cannot as a rule be demonstrated in the cancerous tissue which may develop subsequently, but that by means of immune reactions its presence in these tissues can be made prob-

able. As to the nature of the papilloma virus, Beard and his associates believe that it is a macromolecular nucleo-protein. It has been maintained that also the ordinary wart which develops in the human skin can be transmitted to other places of the skin by cell-free material from the warts; this would indicate that these small papillomatous tumors, too, may be caused by a virus. There are still other observations which point to the more general significance of viruses in the origin of cancer. In some amphibia, tumors have been observed, which seem to be produced by a virus. Thus the Champys found an epithelioma in Triton alpestris, which apparently could be transferred to other animals belonging to this species, but not to other species of Triton, by the water in which the bearers of this cancer had lived. They concluded that the successful transmission of this tumor was due not to the growth of the transplanted cells in the host, but to the action of a virus transferred with the cells. Lucké noticed in a Massachusetts race of Rana pipiens the frequent occurrence of kidney tumors, which varied in structure and activity between adenoma and metastasizing adenocarcinoma. These tumors likewise seemed to be caused by a virus, as indicated by the presence of intranuclear inclusion bodies in the cells of these new-formations and by the fact that transplantation of this tumor succeeds only in the anterior chamber of the eye, and that inoculation with this material of other frogs of the same species and race, as well as of frogs belonging to a different species or race, to other parts of the body, induces tumor formation solely in the kidney, irrespective of the place where the tissue has been implanted or injected, and that such a tumor formation occurs also if the inoculated material has previously been frozen to −80° C and powdered, a procedure which presumably kills all the tumor cells. This virus would therefore be organ-specific and at least approach species or race specificity. According to Briggs, in frog larvae transplants of this tumor may grow in various locations; but in these larvae also, the tumor inoculated elsewhere may induce multiple tumors in the kidney, which seemed to have originated through cancerous transformation of the kidney tubules and therefore were not metastases developing from inoculated cells. For this reason, Briggs considers these renal tumors as caused by a virus.

We have already discussed the significance of the socalled milk factor in the origin of mammary gland carcinoma of mice, and we shall now consider the bearing of this factor on the virus problem. The milk factor is a substance which in certain respects resembles a virus in its action and may be designated as a virus-like substance. In accordance with our previous discussion, there is at least the possibility that it originates in the mice which are bearers of the mammary gland carcinoma and that its occurrence corresponds, to a certain extent, to the degree of genetic tendency to tumor formation in this organ in various strains of mice. Its mode of action is not an acute one, comparable in this respect to the avian sarcoma agent, but on the whole it resembles rather that of the rabbit papilloma virus. However, the milk factor differs from the latter not only because of its possible origin from the tissues of the bearer, but also because it exerts, probably, a physiological function, inasmuch as there are indications that it participates in the noncancerous growth of the mammary gland ducts and presumably promotes the preparatory growth period of the mammary gland. In favor of the view that all kinds of cancer are essentially caused by viruses, two additional observations might be cited. In the course of serial transplantations of tumors in mice, sometimes as early as after the first or second transplantation, the growth of an epithelial tumor, a carcinoma, seemed to induce sarcoma formation in the adjoining connective tissue stroma. In another series of transplantations, the writer noted that at a point where a transplanted adenocarcinoma came in contact with the overlying epidermis, the latter began to grow downward into the underlying stroma and thus to assume a carcinomatous invasive growth. These observations suggest the possibility that in these cases virus-like substances were transferred from epithelial cells to neighboring connective tissue cells, or from glandular

epithelium to squamous cell epithelium, and here induced the newformation of cancerous growth. However, an objection might be raised against this interpretation. As a rule, when a virus has been shown to be active in the production of cancer, it was tissue- or organ-specific. Such a specificity would not very well be compatible with the assumption that a virus can be transferred to a different type of tissue, and would then be able to transform the latter into cancer. Instead, it may be suggested that in the instances mentioned we have to deal with the action of substances present in the cells of the primary tumor, which exert a stimulating effect on neighboring tissues. Similar contact-growth substances may be active even under normal conditions. Thus, we were able to demonstrate that epithelial tissues, as a result of processes taking place during their growth, may stimulate the proliferation of the adjoining stroma. Cancer cells may perhaps possess substances which initiate intense growth processes of an autocatalytic nature primarily in the tumor cells themselves, but which, under certain conditions, may also be transferable to adjoining tissues. On the other hand, the observations of Duran-Reynals, to which we have referred already, indicate that in the course of time a virus may undergo certain changes, which enable it to adapt itself to a strange host and, within certain limits, to induce tumor formation in tissues different from those in which it had grown originally. However, the new tissues which were invaded by virus were, on the whole, related to the original ones; a change from connective tissue to leucocytes or their progenitors being observed; but a transfer to epithelial tissues did not take place.

A second observation which likewise was made in the early period of experimental cancer investigation, which we have already noted, is the increase in growth energy which cancer cells often acquire during the first or, sometimes, the later transplantations. Such an increase in growth energy or in the intensity of certain metabolic processes, as a result of the serial transfer to other hosts, has been found also in bacteria and in viruses, but not, so far as we know, in normal cells. But then we must consider the fact that cancer cells in certain respects differ, in their mode of growth and reaction to hosts, from the normal cells from which they originated, and it is possible that they may acquire some new growth characteristics during the process of cancerization.

From all these data we may conclude that certain viruses may stimulate normal tissues in such a way that they become converted into cancerous tissue, especially under conditions in which the action of the viruses is associated with various other stimulations of a chemical or mechanical kind; and furthermore, that in birds an agent of an exogenous, or possibly of an endogenous, nature, which is responsible for the cancerous growth of certain cells, can be separated from the cancer cells and transmitted to analogous tissues of the same species or of a few other nearly related species, and here induce in the affected cells the development of cancer of the same or of a related type. The experiments which we have discussed also suggest that viruses or virus-like substances of an exogenous or endogenous nature may be present in some instances, in which their presence cannot be demonstrated in a direct manner, and lastly, they show that certain cancers may be due to a more complex system of conditions than had previously been believed. These conditions include: (1) Inherited genetic factors transmitted through the germ cells; (2) continued stimulation of certain tissues, not only by a variety of well defined physical and chemical factors but also by certain viruses; (3) a virus-like factor, which may be active in the production of cancer only in coöperation with the other factors mentioned; this virus-like substance may possibly be of an endogenous nature and may perhaps exert even some physiological function.

However, Peyton Rous and Andrewes go farther and conclude that all cancers are essentially caused by viruses and that all other factors are merely subsidiary, either modifying the virus, increasing its cancerigenic efficiency, or making the recipient tissues more responsive to the cancerigenic action of the virus. Andrewes,

in particular, has suggested that exogenous viruses may enter tissues at an early age and here lead a passive existence until certain injuries affect the host cells or some changes act directly on the viruses. In this connection it may again be stated that, as far as is known at present, the cancerigenic viruses or virus-like substances seem to be, to a large extent, organ- or tissue-specific, at least as far as their ability to induce cancerous growth is concerned, and that adaptation of the viruses to different tissues has been observed only within a relatively narrow range.

Inasmuch as all tissues which still possess the potentiality to grow may become cancerous, it would then need to be assumed, according to Andrewes, that the number of exogenous viruses which have invaded the animal organism in early life and from then on make certain organs and tissues of the latter their permanent habitat, must be considerable; indeed, a large number of different tissues and organs would have to possess their own specific viruses. Furthermore, the additional possibility must be considered that the cancers caused by different stimuli in the same organ or tissue may differ as far as the genetic factors involved are concerned. Thus it is known that estrogen produces mammary gland carcinoma readily only in mice belonging to certain strains, while in mice belonging to other strains estrogen is ineffective. However, one of the latter strains, owing to its genetic constitution, may be especially prone to respond with the formation of mammary gland carcinoma to the application of certain carcinogenic hydrocarbons. If, in each of these instances, a virus were the principal factor in the initiation of this type of cancer, the possibility might have to be taken into account that more than one virus resides in certain organs, as for instance, the mammary gland, one responding to the cancerigenic action of estrogen and another responding to the action of carcinogenic hydrocarbons. The number of viruses residing in normal tissues would then have to be still further increased.

But another interpretation also seems possible, which would limit the significance of viruses in the etiology of cancer in the following way: (a) Certain viruses may act essentially as growth stimulators for those cells or tissues to which they are adapted, resembling in this respect growth-promoting hormones, carcinogenic hydrocarbons, and various other less specific growth stimuli, which all are capable of causing the transformation of normal tissues into cancers. According to this view, each of these factors exerts its cancerigenic action in its own way, which may differ in intensity and also in certain other particulars from the mode of action of the others. Some combinations of two or even more of these stimulating factors may lead to a considerable acceleration and intensification of the effect, as compared to the effect exerted by a single one of them. (b) In one type of cancer, however, the avian sarcoma, and in avian leukemia, it has been possible to separate from the cells a growth-promoting substance, which when injected into other individuals of the same or a related species, very rapidly transfers to certain cells, as it were, passively, a very effective agent, which propagates within the cells and acts like an autocatalytic growth substance, inducing cancerigenic growth and at the same time changing some important metabolic functions of the invaded cells. (c) There exists a virus-like substance, the socalled milk factor, which acts as a partial causative agent in the cancerization of the mammary gland and which is active only in association with specific stimulating and genetic factors, all of these three conditions being of importance. It is possible that in the future there will be found still other instances in which such a partial cancerigenic virus-like substance supports other factors in producing a certain type of cancer. According to this view, in all these cases, with the exception of avian sarcoma, viruses would not be supra-ordinated but coördinated factors in this process, and some of them acting in principle not unlike many other stimulating substances. Furthermore, in this connection it is necessary to recall the fact that the nature of the viruses is not fully understood at the present time. There still remains the possibility that they represent complex macromolecular chemical compounds

rather than active microorganisms, and that they thus may be related to the autocatalytic growth substances which we have assumed to be the direct cause of cancer.

In favor of this second interpretation as to the significance of viruses, there may also be cited the fact that, as a rule, cancerization represents a step-by-step, intensified growth process, such as may be produced by hormones, by cancerigenic hydrocarbons, or by various other substances and physical conditions conferring a continuously increasing growth momentum on the tissue's substratum on which they act. In this process, the preparatory growth period is an essential component, and the stimulating agent may fail to be effective because, on account of the hereditarily determined inertia of certain tissues, the more advanced stages in the preparatory growth cannot be reached. There are very marked differences in this respect between different strains or individuals, depending upon their genetic constitution and upon the character of the stimulating factor. This can be readily understood if we accept the second interpretation of the character of viruses, but it would be more difficult to understand if we assumed that all cancers are induced essentially by viruses and that all the other factors mentioned are merely of subsidiary significance, which help the virus, inhabiting all the cells of a certain organ or tissue, to set the cancerous process into motion.

Moreover, the latter assumption would bring some other difficulties. If all cancers are caused essentially by specific viruses inhabiting all the cells of certain organs and tissues, what would then be the mode of action of an extrinsic virus, such as the rabbit papilloma virus, which when rubbed into the skin gives origin first to warts, and, ultimately, to cancer? There would then be two cancerigenic viruses superimposed upon each other: first, the indigenous virus of the rabbit epidermis, which is able to give origin to squamous cell carcinoma if it is spurred on to do so or in some way is transformed by tar or other carcinogenic substances into an active cancerigenic agent, and secondly, the extraneous cancerigenic virus. Does the latter act on the intrinsic virus in the same way as tar or a hormone is assumed to act? Or do these two viruses coöperate? As to the mode of action of a cancerigenic virus on the tissue substratum, is it supposed, step by step, to intensify the growth processes in the cells in which it lives? If this should be so, its action would be about the same as that which hormones as well as some other growth-promoting factors are known to exert on these cells.

In view of all these considerations, we think it more probable that, with the exception of the agents active in avian sarcomas, viruses are coördinated as cancer-producing factors with various other specific or less specific factors; in principle, they act like hormones and the other cancerigenic agents which we have discussed, although the milk factor may perhaps affect the responsiveness of the tissues to the stimulating condition. Such a view leaves unimpaired the great interest and importance of the facts established by the study of the action of cancerigenic viruses.

(d) **Age and Cancer.** We have discussed so far the genetic factors and the stimulating factors, including viruses, in the origin of cancer. There are some additional data concerning the origin of cancer which are based primarily on clinical observation and on microscopical study of human cancer, but which can be correlated also with experiments and observations made in animals. These concern, for instance, the effect of age on the development of cancer and the origin of cancer during embryonal life which has been noted in certain types of tumors. In general, the frequency of cancer increases with increasing age; but in some organs cancer reaches a maximum frequently at an intermediate age, after which it declines again, a condition presumably due to peculiarities of the life curves of various organs and of the stimuli reaching them. There is reason for believing that the increase in the incidence of cancer with increasing age is due primarily to the fact that the sum of stimuli affecting all the tissues and organs increases with increasing age. Furthermore, certain changes which the tissues undergo with advancing age, such as hyalinization

of the stroma, may modify the response of the tissues to stimuli, impede their return to the original normal state, and cause injuries to the tissue, resulting in long-continued recurrences of regenerative growth processes. In addition, diseases which predispose to the development of certain cancers by initiating growth processes, exert their effects with increased frequency in adult life and in old age. Similar relations of age to the incidence of cancer may be observed also in animals. Thus, in mice, cancers of the mammary gland usually appear only in middle-aged or older mice; however, they develop the earlier the more favorable the genetic constitution of an individual or inbred strain to the cancerous transformation of the organ, and the greater will be, therefore, the number of individuals affected by cancer. If the genetic constitution is favorable to the production of this type of cancer and if the stimuli have been very effective, cancer may appear even in young individuals. However, as has been mentioned already, there exists, besides, certain strain or family peculiarities as to the time of appearance of mammary gland cancers.

Of special interest are certain tumors which begin to develop during embryonal life and are associated with abnormalities of embryonal tissues, and which may be of two kinds. In the first place, disturbances may occur in some organs, such as the kidney, apparently at definite stages of embryonal development, which predispose these organs to tumor formation. Kidney tumors of this type as a rule are malignant. Secondly, there is a type of embryonal cancer which resembles a whole embryo, inasmuch as a number of varieties of tissues, and even rudimentary organs of various kinds irregularly arranged, constitute this tumor; one or more of these constituents may assume cancerous growth. Such tumors occur most frequently in the ovaries and here they are due presumably to the parthenogenetic development of eggs. This interpretation is supported by the relatively fargoing, apparently spontaneous, parthenogenetic development of eggs which we have observed relatively frequently in the ovary of the guinea pig. Similar formations may originate also in some other places in the body, where, probably, aberrant germ cells have settled and developed. Chorio-epitheliomatous structures which may be found in ovary and testicle, have very likely a similar origin. However, these abnormal embryonal structures as such would not yet constitute cancers, but not infrequently it happens that secondarily one or several of the tissue constituents of these complex tumors assume cancerous growth. Regarding the factors which cause the transformation of these various embryonal tissues into cancers, no definite knowledge exists, but there can be little doubt that some growth stimuli, perhaps even of a nonspecific nature, are able to induce the cancerous transformation. Such an effect may readily be understood if we assume that these embryonal tissues behave like sensitized adult tissues, which, as we have pointed out, may readily be transformed into cancer by various nonspecific stimuli, such as occur even under ordinary conditions of life.

Thus, the variety of types of cancers which have been observed is very great and the variety of conditions under which they develop is equally great; but the principles underlying their formation seem to be similar in all instances known so far.

(e) **The Effect of Nourishment on the Development and Growth of Tumors.** The foodstuffs which an animal consumes also influence tumor growth and may therefore be included among the etiological factors already discussed. In appraising the significance of diet on tumor growth, we have to differentiate between its effects on the growth of already established tumors and on the transformation of normal into cancerous tissue. As to the former, it has been observed by a number of earlier investigators that undernourishment may interfere with the growth of tumors after transplantation into a new host. More recently, Voegtlin and his associates have found that certain amino acids, such as lysine, cystine and methionine, which are necessary for normal growth, are necessary also for the growth of mammary gland carcinoma. Bischoff and Long showed that general undernourishment, without particular deficiency in one of the specific basic food requirements, inhibited the growth

of mouse sarcoma 180. Likewise, a deficiency in panthothenic acid resulted in growth inhibition (Bischoff and Long, Morris). However, inasmuch as this substance interferes with the state of nutrition, as a whole, of the animal, it is possible that this latter factor may also be involved in the retardation of tumor growth under these conditions and that we may have to deal with complex effects.

Distinct from these investigations are experiments in which the effect of dietary factors on the origin of cancer was studied. Of special importance in this line of investigation are the experiments of Tannenbaum, who showed that the development of spontaneous mammary gland carcinoma and of pulmonary adenocarcinoma in mice, as well as the experimental production of epidermal carcinoma or subcutaneous sarcoma by the administration of carcinogenic hydrocarbons, was markedly diminished or long delayed as the result of undernutrition, without regard for the lack of any particular dietary component. These results were confirmed by Visscher. Tannenbaum, moreover, noted that the dietary deficiencies which sufficed to prevent the newformation of tumors did not need to interfere with the growth energy of tumors after they had once developed. If once the normal cells have changed in such a way that a self-perpetuating growth stimulus has developed in the cells, this stimulus may be able to overcome the impediment to further growth caused by a deficiency in foodstuffs. In accordance with the experiments mentioned, it has been shown in our laboratory that a certain parallelism exists between the development of mammary gland carcinoma in mice and the relative state of nourishment of these animals, as expressed in their weight curves. In general, it is probable that dietary deficiencies act on tissues and tumor growth by interfering with the supply of building stones which organs and tissues need during the process of growing, and also with the maintenance of the grown organism; but it is probable also that, to a certain extent, a tumor once established can draw to itself foodstuffs which under ordinary conditions would be used by the normal tissues. In addition to these direct actions, undernourishment may indirectly affect the development of mammary gland carcinoma by causing atrophic conditions in the fat tissue and an increased formation of hyaline connective tissue, in which the normal mammary gland tissue is embedded. These changes secondarily induce an atrophy and destruction of the mammary gland tissue and thus diminish the quantity of tissue substratum available for carcinoma formation.

While, therefore, undernourishment as a rule tends to inhibit or prevent cancerous growth, there are some conditions where it has the contrary effect, and when it may help to initiate cancerous growth in certain organisms, or may lead to the production of precancerous papillomas in the stomach or to leucoplakia in the mouth (Movus and Lippincott, Rhoads and associates). The changes found in these organs may be due to general undernourishment, or, more probably, to vitamin B deficiencies; the latter seem to be responsible for leucoplakia in the tongue. It appears that lack of foodstuffs in general, or of special foodstuffs which usually tend to restrict growth processes, may under certain conditions, on the contrary, cause an increase in mitotic proliferation; this has been observed by Blumenthal and the writer in the adrenal cortex of the guinea pig, where undernourishment may cause inhibition as well as intensification of growth processes, and here the conditions for these two opposite effects can be separated. It is probable that undernourishment in this organ may stimulate mitotic proliferation, by causing the removal from the cells of some substances which tend to inhibit growth.

An inhibiting effect on the growth of transplanted tumors has likewise been observed by Ball, Samuels and Schott in hypophysectomized rats. This effect is due presumably to metabolic deficiencies caused by the lack of the hypophysis; but it is possible that some other factors are also involved in this process. This result of hypophysectomy has been confirmed by several investigators, and more recently by Gardner, who found that also the spontaneous development of mammary gland carcinoma in mice may be inhibited by lack of the hypophysis. In general, these experi-

ments confirm the conclusion that all processes which tend to interfere with normal growth, may also interfere with the development of cancer, and that they exert this effect mainly by inhibiting in the tissues the preparatory growth processes which lead to the cancerous transformation.

(f) **Somatic Mutations as the Cause of Cancer.** Cancerous growth represents a change in somatic cells, which, as far as is known, is irreversible; there is no evidence that a cancerous cell can ever regain its normal state of proliferation and function. This origin from adult somatic cells applies to the large majority of all cancers, although in a limited number of cases cancerization in all probability affects the germ cells and embryonic tissues. The permanency of the changes induced in somatic cells by cancerization is similar to that caused by mutations primarily in the germ cells, but occasionally, also in somatic cells. This fact has suggested the conclusion to a number of investigators, and especially also to Tyzzer, and subsequently to Little and Strong, that somatic mutations might be the essential cause of cancer and all other participating factors would merely be subsidiary. Accordingly, it has been assumed and is still held by many investigators that long-continued stimuli acting on a somatic cell may, in the end, bring about a permanent change in chromosomes and genes, representing a mutation which is responsible for the development of cancer.

Several observations seem to support this view—thus, through application of Roentgen rays it has been possible to produce mutations in germ cells experimentally, and, correspondingly, it is known that long-continued use of these rays has induced epidermal cancer in a number of instances. Some other changes observed in the experimental study of cancer have likewise been explained on the same ground. Thus the fact, already mentioned, that very often in the course of the first transplantations of a tumor, the growth energy increases, and that accordingly the number of successful transplantations also may increase—a change which in some instances may occur in the course of later transplantations—has been attributed to somatic mutations in transplanted cancer cells. This would presuppose that in very many, perhaps in the majority, of cases, the process of transplantation induces a somatic mutation, which invariably increases the growth energy of a tumor cell in a definite way; and furthermore, that the change induced in a single cell then determines the character of the whole tumor resulting from the transplantation and that in the succeeding transplantations this process may be repeated; but that if once these mutations have occurred, no further somatic mutation of this type as a rule takes place. As pointed out in a preceding chapter, it appears much more probable that the process of transplantation acts as a stimulus, setting into motion changes corresponding to regenerative growth and thus increasing the growth energy of the tumor cells to a maximum and bringing about an intensification of the cancerous growth. In a similar manner, adherents of the mutation theory have explained the fact that if two or more mammary gland tumors originate in the same mouse, or in two mice of the same inbred strain, they may differ from each other in the readiness with which they can be transplanted into other mice, these differences being attributed to differences in the somatic mutations which give origin to the two tumors. Instead of this interpretation, it appears more probable that we have to deal in these cases merely with different intensities in the process of cancerization which these two tumors have attained, and that the farther a tumor has progressed in this direction, the greater its growth energy will be. In principle, the same process has occurred in both of these tumors, but in one of them it happened to progress more rapidly than in the other.

Somewhat related observations have been made more recently by Reinhard and his associates; a tumor originating in a mouse belonging to the cancer-rich D strain could, as is so commonly the case, at first be transplanted solely into other mice belonging to the same strain, but not into individuals of the C57 strain. However, after application of Roentgen rays to this tumor, it grew, also, in a certain proportion of C57 mice, the proportion increasing in the beginning with increasing doses of

Roentgen rays, but then, after having reached a maximum, decreasing again. This change in the transplantability of these tumors in the C57 mice was maintained in serial transplantations and it was therefore attributed to a somatic mutation which had occurred in a tumor cell exposed to the Roentgen rays. However, another explanation which would explain these observations may be suggested. As a result of the Roentgen ray treatment, the metabolism of the tumor cells was affected in such a way that the individuality differential of these cells was modified in certain respects. On account of the change brought about in the cells thus treated, they were no longer able to produce this differential substance in the same quantity as previously, or the individuality differential they produced and gave off had lost some of its specific features. As a consequence of these alterations, the strength of the reaction of a genetically different host, such as a C57 mouse, against the transplant was diminished and the tumor grew and was able to develop in a certain proportion of these otherwise unfavorable hosts; but if still larger doses of x-rays were used, also the power of resistance to injurious conditions on the part of the transplant was diminished and consequently the ability of these tumors to grow in unfavorable hosts decreased. The occurrence of a somatic mutation which would render the genetic constitution of the tumor cells more similar to that of the C57 strain, would presuppose a concurrent alteration in their genetic constitution of such a kind that their behavior after transplantation into their own strain should be altered. While apparently an experimental test of this effect has not been made, it seems probable that it would not show a diminution in the degree of transplantability of the tumor into its own strain as the result of the Roentgen ray treatment.

Similar difficulties exist also in other instances in which the transplantability of tumors has been attributed to somatic mutations. Bittner, for instance, assumes that the occurrence of mammary gland carcinoma in certain low tumor stock mice, and in some descendants of high tumor stock mice which had been foster-nursed by low tumor rate females, cannot be accounted for by the assumption that it is due to the same factors that induce the development of mammary gland carcinoma in normal high tumor rate mice. He attributes the cases of breast cancer which develop under these circumstances, to somatic mutations taking place in the mammary glands of the mice developing the tumors. However, it might also be assumed that these low tumor rate strain mice are not entirely devoid of the required genetic constitution, and that the high tumor rate strains foster-nursed by low tumor rate strain females still possess a small amount of the milk factor necessary for the occurrence of mammary gland cancer. Certain secondary factors would then make possible the occasional occurrence of a tumor in such mice. The nature of these secondary factors is unknown, but their existence must be assumed also in order to explain why the incidence of mammary gland tumors, which differs so much in different strains of mice, is associated with great differences in the behavior of the individual mice belonging to these strains, and with a constancy of the percentage of tumor incidence and of the average tumor age in the strains as such.

However, aside from these difficulties, there are others of a more general kind, which make the interpretation of somatic mutations as the essential cause of cancer improbable. A somatic mutation is a relatively rare chance phenomenon, which, if it occurs at all, should occur in an isolated cell and here induce a sudden abrupt change in the conditions and reactions of this cell and its descendants. But if we study cancer development in the mammary gland, as it takes place under the influence of endogenous or exogenous estrogenic hormones, we find that this process affects whole complexes of gland cells simultaneously; it is a change which affects the structural gland units, the tubules or acini, and may affect several units at the same time or in succession, first in one area of the gland and then in another area, until in the end great parts of the gland may become cancerous, a condition which does not harmonize with the theory of somatic mutation as the cause of cancer, which presumes

a one-cell, discontinuous origin of cancer. Furthermore, the mammary gland tissue shows not an abrupt but a continuous structural change from the normal to the cancerous state, and before a definite precancerous stage has been reached, a continuous series of graded cell-changes has occurred, which leads, as we maintain, to a sensitization of these cells. All these successive alterations require a considerable time; they take place gradually, step by step, and represent a graded series of processes which depend primarily upon the sensitizing action of hereditary factors. Such a curve of changes, eventuating in the cancerous transformation, is not very well compatible with the concept of a somatic mutation as the cause of cancerous growth. It is hardly conceivable that somatic mutations should be produced in numerous mammary gland cells and in various places almost simultaneously, exactly at the time when the estrogenic stimulation, progressing steadily, has at last reached the point where the transformation into a cancerous state of growth may occur.

In principle, we find a similar condition if in vagina and cervix the epithelium undergoes precancerous or early cancerous changes at various points and at about the same time, under the influence of stimuli which act on these cells also under normal conditions; besides, the processes leading to cancerous transformation seem to be the same also in the epidermis when it is stimulated by tar or carcinogenic hydrocarbons. However, in the latter case we have to deal with abnormal extraneous substances, which may cause accessory injurious effects. Here there may occur, in association with the increased proliferative activity, certain pathological deviations from the usual modes of growths, perhaps as an indication of the great intensity of the stimulation or in consequence of toxic side-effects induced by these substances. But in this case, as well, a series of preparatory growth processes ultimately leads to the appearance of cancer. Different from these processes is the rapid production of various types of fowl sarcoma following the injection of specific agents, where it is hardly conceivable that a somatic mutation of a single cell would be induced at once by the transmission of the agent, a point which has been emphasized by Rous. None of these processes can be explained in a satisfactory manner if we assume that a somatic mutation is the cause of cancer in general.

Lastly, it may be stated that in embryonic development we have an example of a continuous series of irreversible cell-changes which are not due to mutations—germinal or somatic. The genetic constitution of the segmenting eggs, as far as known, is the same as the genetic constitution of the adult epidermal cell, gland, muscle or nerve cell. During embryonic development, a step-by-step diminution in the proliferative activity and in the movements of the various cells and tissues takes place. Associated with these changes are alterations in the chemical constitution of the developing cells and tissues, which are different in kind from those characteristic of somatic mutations and yet are permanent. During cancerigenesis, the opposite process occurs; instead of a diminution, there is a graded increase in the proliferative and motor activity of cells and again there are associated with these changes, chemical and structural modifications which, in certain respects, make the reactions of these cells similar to those of embryonal cells. Such alterations likewise are permanent. It seems, therefore, conceivable, and even probable, that also the cancerigenic cell-changes do not depend upon somatic mutations. In addition, there is no case known of an experimentally produced somatic mutation which has resulted in cancerous growth, and on the other hand, in no case has it been possible to prove that a somatic mutation was the cause of a cancer.

These data are, then, not favorable to the conclusion that somatic mutations are the essential cause of the cancerigenic processes. On the other hand, mutations which alter the constitution of germ cells may be an essential factor in the origin of cancer, insofar as such mutations determine the degree of hereditary tendency towards the development of cancer which these organs and tissues possess and their responsiveness to cancerigenic stimulations. However, in the discussions concern-

ing the cause of cancer the term "somatic mutation" has been used by some investigators, including the writer, not in the strict genetic sense, but in a less well defined way, in order to signify that irreversible changes of some kind have taken place in cells during their transformation into cancer, irrespective of the mechanism through which they have been accomplished. However, this use of the term somatic mutations may lead to misunderstandings and it would be advisable to abandon it.

(g) **Theories as to the Causes and Nature of Cancer.** Cancer is a growth phenomenon; to understand the cancerous process completely, a full knowledge of the mechanism, and particularly also of the biochemistry, of normal growth processes would be needed, and moreover, it would be necessary to show wherein cancerous growth differs from normal growth. Our knowledge in these respects is still very fragmentary. However, it can be stated that the cancerous process, like many other vital phenomena, represents a chain phenomenon, in which the first link of the chain is designated as the cause; in many cases, the stimulating factor is considered as the first link and regarded therefore as the cause. But, there is hidden in the cancerous process, another factor, representing the readiness of the tissue substratum to respond with growth processes to stimuli. The degree of responsiveness varies in different cases and also in different types of cells and tissues; primarily, it is determined genetically. In cases in which the state of the tissue substratum is a very important factor in the ultimate cancerous transformation, emphasis has been placed on this condition as the cause of cancer by some investigators. Furthermore, in the course of the cancerous chain process the character of the substratum undergoes certain changes; it becomes sensitized and thus gains in importance at the expense of the significance of the stimulating factors. But the process cancerization is not a simple chain process; it is represented by a very long complex chain in which the stimulating factor continues to act over extended periods of time, either perpetually in small quantities, or intermittently at intervals of different length, and each time it functions it sets into motion a reaction chain. Under normal conditions, life phenomena are cyclic or rhythmic; stimuli cause a change, or a series of changes, which normally end by a return of the system to the original state. But in many cases the end-stage reached in such a cycle is not exactly the same as the state in which the organism was at the beginning of the cycle; the return is not complete, and ultimately this incompleteness leads to certain irreversible changes as those observed in the process of ageing. The chain process leading to the cancerous tissue transformation consists also of a series of cyclic processes, in which in each case the endstage differs from the stage present in the beginning. The substratum is altered continuously, step by step, in a certain direction, until in the end the threshold of cancer is reached, which represents a new and irreversible tissue equilibrium, and, as we have seen, this process may continue in the same direction, even after this point has been attained.

Several phases in this complex chain process are understood, namely, the nature of the initial stimulating factors and, in many cases, also the degree of responsiveness of the tissue substratum; furthermore, the gradual structural and growth changes occurring in the tissues and the process of sensitization are known to some extent, as well as the mode of reaction of cancerous tissues and some of the metabolic accompaniments of the cancerous condition. There are, besides, at least certain indications of the chemical and metabolic changes which are produced by the initial stimulus during the first chain. This, in a general way, is the present status of our knowledge of the cancerous process. If all the elements of this composite chain were known, no theory of cancer, in the sense in which this term has been used in the past and is still used at the present time, would be necessary, the various theories being very incomplete substitutes for a full understanding of the causes and nature of cancer. Especially in the earlier periods, the discussion of theories concerning the causes and nature of cancer played a very prominent role. Yet, on the whole, each one of the older theories merely expressed a single feature more or less characteristic of cancerous

types of growth, at the same time ignoring a number of other equally important, or perhaps even more important, characteristics of this process. Thus, Virchow considered chronic irritation as the cause of cancer; this is designated as the irritation theory of cancer. Cohnheim stressed the importance of abnormal or retarded embryonal development in a certain organ in the origin of cancer, and he thus stated the embryonal theory of cancer. Ribbert compared cancer tissue to regenerating tissue, such as the epidermis during the process of wound healing, but differing from the latter in that the growth-restraining influence of the union of the originally separated tissues, which represents the endstage in the process of healing, is lacking. Cancer would then be tissue which is forced to regenerate indefinitely, and the cancerous process would be a state of regeneration that continues without end. Regeneration of this kind may be induced by chronic inflammatory conditions, or, in other instances, by the separation of embryonal parts during embryonal development. This explanation assumes, then, that cancer cells are essentially normal, regenerating cells, and that they would be able to return to their old normal state if environmental conditions permitted them to do so. However, cancer cells are not normal cells; they behave like cells spurred on to grow indefinitely and to move without rest because of some inner change which they have undergone, or, in some instances, because of the presence within their cells of a virus-like agent.

At the present time, three interpretations of the cancerous process have to be considered. The first theory regards somatic mutation as the essential cause of cancer; the second theory holds viruses to be responsible for cancerous growth. The third theory considers as the principal factor in this process, a series of inner consecutive changes in the cell structure and metabolism, perhaps consisting in an increase or new creation of substances acting as autocatalytic growth stimulators. These changes are connected with and caused by the long-continued growth processes which precede cancer formation, and lead, step by step, to an increased growth momentum and to a sensitization of the tissues, which makes them more and more independent of the specific stimuli in their continuation of these growth processes.

We have discussed already the significance of somatic mutations and of viruses as the possible causes of cancer. The first of these theories subordinates all other causative factors to cell mutations. It is assumed that such mutations occur in the course of the chain events set into motion by the action of stimulating factors. The second theory subordinates all factors concerned with the causation of cancer to the presence and activity of a virus. All other factors are supposed either to activate the viruses, or to prepare the substratum on which the virus acts, in such a way that the virus can fix itself to the cells. While this theory does not state definitely wherein the action of a virus, as for instance, that of the rabbit papilloma, would differ from that of other stimulating factors, such as hormones or carcinogenic hydrocarbons, it seems that in contrast to the latter, which bring about gradual changes in the cells, eventuating in cancer, and which are no longer needed once the cancerous state has been reached, the virus is supposed to be present continuously in the cells and to exert a perpetual growth stimulus on the tissues, which would not need, therefore, to undergo fargoing changes in order to maintain the continuous growth. There is some evidence that the agents of fowl sarcoma may function in this way, but there are, on the other hand, also indications that other viruses may perhaps act as cell stimulators, similar to enzymes and carcinogenic hydrocarbons. The third interpretation was stated in principle by the writer in 1916 and has since then been still further elaborated.

3. Graded Cell Changes as the Cause of Cancer. In stating the third theory, we shall recapitulate some of the essential data and conclusions at which we have arrived. This theory considers cancer as the result of step-by-step intensified growth processes, due to the interaction between growth stimuli, graded in strength as well as in time, and a graded responsiveness of tissues and cells which is primarily deter-

mined by hereditary, genetic factors. As a result of this interaction, the threshold of conditions, which makes possible the return of the stimulated tissue to its original equilibrium, is exceeded. In the course of the preparatory growth period, the tissue passes by way of sensitization to precancerous and finally to irreversible cancerous states, of which again there are various degrees of intensity. The ultimate state of the cancerous equilibrium is irreversible. Tentatively, it may be conceived as being due either to the increased, constantly renewed formation of a new autocatalytic growth-promoting substance, not yet present in the tissue in its ordinary equilibrium, or to the intensified production of autocatalytic substances, which are present already in normal cells. While it is not difficult to designate the first two theories which we have discussed by a single term, such as the virus or the somatic mutation theory, because each of these theories considers one single factor as paramount and all other factors as subsidiary, the third theory just mentioned is based on a series of graded cell changes and attempts to combine hitherto, more or less isolated data concerning cancerous growth into one connected system, in which various factors which are considered equally important coöperate. Hereditary, constitutional and stimulating factors, through their interaction, bring about these graded cell changes. However, as stated already, the distinction between these two sets of factors is not absolute, but only one of degree. The stimulating factors produce sensitizations of the tissues, and likewise, hereditary factors may produce in the tissues a primary sensitization to growth stimuli. In regard to the virus-like agent, the socalled milk factor, which coöperates with the stimulating hormones and with the genetic constitution in inducing mammary gland carcinoma in mice, it is not certain to which category it belongs. In general, stimuli of the variable kind, specific and nonspecific, strong and weak, may possess a cancerigenic function. Exogenous viruses like the Shope papilloma virus also seem essentially to exert a stimulating function not unlike that of other cancerigenic agents, such as hormones and the specific carcinogenic hydrocarbons.

The formation of growth substances, which are constantly renewed autocatalitically and which may also stimulate the invasive activity of the cells, might explain the irreversible condition of the cancerous cells, but the gradual increase in this substance, in addition, might be assumed to parallel the gradual increase in growth momentum of the cells during the preparatory growth period under the influence of the stimulating substances. At a certain stage in this process the growth momentum has reached an intensity which makes it possible for the cells to undergo the cancerous changes under the influence of the normal metabolic or mechanical conditions. Moreover, it would have to be assumed that the activity of these self-propagating growth substances would become irreversible as soon as a certain degree of intensity of their action has been reached. This interpretation of the cancerous state which, was suggested by us in 1916 and subsequently, as far as the existence of these autocatalytic growth substances is concerned, is tentative, but the essential underlying principle of the theory, namely, that cancer is the endstage of preparatory growth processes which depend upon the quantitative interaction between stimulating and hereditary factors, and that the hereditary factors determine mainly the intensity of response of certain tissues to stimuli, and furthermore, that this response is a specific characteristic of the various tissues and organs, seems to be in harmony with the essential facts more recently published.

There is then one effect which, as far as we know, all the cancerigenic chemical substances as well as physical agents have in common, namely, that of initiating a preparatory growth period, during which sensitization takes place, and there is only one condition known in which it appears that cancer is initiated without such a preparatory growth period preceding the cancer formation. The agent causing fowl sarcoma seems to represent not a stimulating factor which gradually induces the receptive tissue to undergo step-by-step changes which eventuate in cancer, but in this case the end-factor which directly induces intensified cancerous proliferation, to-

gether with certain metabolic changes characteristic of cancer, is passively transmitted to the cells and begins very soon afterwards to be effective.

In all other instances, as far as is known, the growth processes as such, or processes intimately connected with these growth processes, lead in the course of time to those cell changes which make them the active factors in the development of cancer. That it is the growth processes initiated by the various stimulating agents, or at least phenomena closely associated with these growth processes, which are essential for the production of cancer, is indicated also by the fact that all conditions which accelerate these preparatory growth processes seem, as a rule, to accelerate also the process of cancerization, as shown in the parallelism which exists between the intensity and rapidity of the preparatory growth processes and the readiness with which cancer is initiated, and further, by the fact that hormones which enforce or diminish the normal growth stimulating effects of estrogen, likewise enforce or diminish, respectively, its cancerigenic effects. However, not every growth process causes, as a rule, this transformation; ordinary regenerative changes or certain hormone actions which are quite efficient in producing the formation of new tissue, such as the development of the mammary gland during pregnancy, do not, as a rule, lead to cancer production. The cancerigenic growth processes must be of such a kind that more and more intensified growth reactions take place in the course of time; merely extensive growth which, after a certain period, allows the return of the growing tissue to its original starting point and which represents, therefore, an approximately cyclic process, does not, as far as is known, induce cancer formation. It follows also from these considerations, that in a strict sense there exist no specific cancerigenic substances, but merely growth substances differing in the intensity of their action, and this is true also of the socalled carcinogenic hydrocarbons. The only probable exception to this conclusion known so far are the agents which induce avian sarcoma, and in this case it must be remembered that they affect connective tissue structures, which are, in general, more readily converted into cancer than are epithelial and certain other tissues. However, there still remains the problem as to why different growth-promoting substances differ so greatly in their ability to cause the cancerous transformation of a given tissue. In order to answer this question it would, above all, be necessary to compare the processes taking place during the preparatory growth period initiated by these various compounds.

In general, there is a tendency inherent in tissues and organs of the organism to undergo cyclic changes in response to stimuli; these reactions may consist in functional, metabolic or growth processes. The cyclic nature of these reactions is, however, not perfect, and because of this deficiency the tissues and organs, and the organism as a whole, do not retain their youthful efficient nature, but gradually they approach the conditions characteristic of old age, or, as a result of a similar process, they become cancerous if stimuli which induce marked growth processes interfere with the return of the tissue to the original equilibrium. In certain respects, cancer may then be included among the ageing processes, even in cases in which it develops in young individuals. There is another important point of resemblance between cancer and the ageing of the organism as a whole. One of the characteristic features of old age is the partial dissociation which occurs in this condition in the harmony of the well organized functional and, in particular, metabolic interactions between the various constituents of the normal body in its youthful state. In the course of life, injuries and deficiencies develop in various organs and tissues, which proceed with unequal rapidity among these different components. These disturb the normal interaction and the harmony between adjacent tissues, as well as the interactions between distant organs, and as one of the results of this disturbance, cancer may develop. The cancerous condition quite commonly is designated as one of autonomy of a component part of the body, which was formerly dependent on regulations emanating from tissues directly surrounding it and also controlled by regulations

originating in distant places. However, in reality cancer does not represent a cell community that is autonomous in the strict sense. There is a diminished cohesion, or even a lack of it, between the units composing the cancerous tissue to which a state of autonomy has been attributed, and there is no general or local control over the individual cancer cells within the cancerous formation. The actual condition is, rather, one of more or less advanced dissociation between the constituents of the cancer tissue. The fact that in cancerous growth a certain part of a tissue becomes independent of the central organizing factors in the organism and independent also of the restricting effects of contact with neighboring tissues, is primarily the result of the predominance of the stimulating over the restricting factors in the life of the affected cells and tissue. Even long-continued, rhythmic stimulations may accomplish this effect, if there is a strong reactivity of the tissues to these stimuli, or if there is a summation of slight deviations from the complete return to the resting state of the tissues.

These conditions obtain notwithstanding the fact that cancer may appear in some instances in young individuals. This occurs if the dissociative processes set in in early life, under conditions in which the stimulating factors are very active and if, in addition, the genetic constitution of the individual is favorable to the dissociative effects of the stimulating processes. In certain cases, the dissociative processes affect embryonal cells, or even germ cells, and here likewise the cancerous state may be reached in a young organism.

While sensitization of tissues during the preparatory growth period is an indication of the continuously increasing effectiveness of the tissue stimulation, on the other hand, immunity, developing either in the tissues themselves or in distant organs, specifically endowed with the ability to produce immune substances of various kinds, tends to cause the return of the reactions to their starting point and to render them approximately cyclic. These immune processes must therefore be overcome if cancer is to develop as a result of the long-continued growth reaction. The ability of the stimuli to induce non-cyclic processes, at last leading to irreversible cancerization, depends upon three conditions: (1) the nature and intensity of the stimulations, (2) the kind of substratum on which they act, and (3) the character of the stimulus-reaction circuit. We have already discussed the various types of stimuli which may be cancerigenic. But we shall here once more state the differences between hormones and carcinogenic hydrocarbons in their function as cancerigenic stimuli. Hormones may initiate cancer by exerting their normal growth-stimulating functions on organs to which they are specifically adapted; secondarily, and much less effectively, they may in some instances exert a cancerigenic action on connective tissue or leucocytes; this effect is perhaps due to a chemical relationship of these hormones to carcinogenic hydrocarbons, and it is different from their primary cancerigenic action, which is based on their physiologic function. Carcinogenic hydrocarbons act on tissues with which they come into contact rather indiscriminately; they lack to a very large degree the tissue or organ specificity and they do not normally exert physiological functions as growth stimulators; but some of them possess what might be designated as an intensity specificity, acting on the tissues with which they are in contact with great intensity. There is therefore no reason to assume that a hidden carcinogenic hydrocarbon is the cause of all types of spontaneous cancer; in fact, the only spontaneous cancer in which an endogenous chemical factor is known as the cause is the hormone cancer; it is due to substances which occur normally in the healthy organism, and the discovery of hormones as cancerigenic agents and of their mode of action preceded that of the carcinogenic hydrocarbons, which differ in their cancerigenic action in some important respects from the hormones. The search for carcinogenic hydrocarbons as the cause of spontaneous cancers is based on the supposition that they are the only potent cancerigenic factors. However, as stated, these compounds are not in the strict sense carcinogenic substances; like all other

cancer-producing agents, they are in all probability primarily growth substances. Many stimuli may initiate growth processes and thereby cause the production of cancer. The more specific factor in this process is evidently not the nature of the stimulating factor and its mode of action in calling forth growth processes; the most specific factors are those which are common to all cancers, and the essential problem concerns the question as to how and under what conditions the transition from the preparatory growth processes to cancer takes place and whether there are processes associated with these growth phenomena which are required for this transition. The mechanism through which these stimuli set into motion the preparatory growth processes may vary according to the kind of stimulus which is applied, and it is presumably of secondary significance as far as the production of cancer is concerned. Conditions are different in this respect, if we consider the mode of action of infectious diseases; in these, the initiating factor, the microorganism, is of primary importance; it is the most specific factor and largely determines the nature of the disease which follows its entrance into a host, in contrast to cancer where the various types of stimuli all lead to the same kind of pathological change, in which the symptoms and the outcome are the same, irrespective of the factor which set this sequence of events into motion.

As to the inherited individual, strain and species differences in the substratum, these are very potent in modifying the results of stimuli. Thus, in the mouse and guinea pig the effects of estrogen on the mammary gland differ greatly. In the former, long-continued application of estrogen initiates the active growth processes only very slowly, and then, step by step, these reactions are intensified, until in the end cancer is produced. But even in the mouse the effectiveness of these stimuli differs greatly in different strains. On the other hand, in the guinea pig the early results of the stimulation of the mammary gland by estrogen are relatively much more marked than in the mouse, but after some time these growth effects come to a standstill and no further proliferation takes place; the tendency to the formation of mammary gland carcinoma under the influence of estrogen is lacking, or at least is very weak in this animal. The growth curves of the mammary gland caused by ovarian hormones differ in these two species, and, correspondingly, the readiness with which carcinoma develops also differs. Moreover, the differences in the physiology of various tissues are important factors in determining the readiness of these tissues to become cancerous. Thus connective tissue cells seem to respond as a general rule, more readily to stimuli of different kinds with cancer formation than do epithelial structures. Stimuli of a nonspecific nature may be effective in the former to a much greater extent than in the latter. It will therefore be necessary to know fully the physiology of various tissues in order to understand the difference in the readiness with which they may become cancerous. In regard to the third point, the reaction curves are different in different growth processes, as for instance, in the regenerative growth process and in embryonal development. In some instances the initial element in two growth chains may differ greatly, while the end-results may be the same. Thus injuries, leading to defects in the tissue, may initiate growth processes, and, likewise, hormones may be able to do so, but without depending upon primary injurious conditions for their effects. These differences in the character of the stimulus-reaction circuits in different types of growth processes, affect as well the manner in which the various types of growth finally change into cancerous growth.

This interpretation of the cancerous process has to meet one difficulty. According to all the data known at present, it is characteristic of cancerous cells that they retain their specific reaction modes and transmit them perpetually to the following cell generations. Thus, cancerous cells may die, but they do not seem to revert to the state of normal cells. This point has been proven especially by means of the continuous serial transplantation of tumors; but it agrees also with the observation and experimental study of individual tumors. It is furthermore certain that the cancer

cells develop from normal cells, which thus transmit their newly acquired characteristics to their descendents from generation to generation. This seems to contradict the findings in unicellular organisms where acquired characteristics survive, at best, for a considerable number of generations but ultimately lose again these acquired properties ("Dauermodifikationen" of Jollos). In protozoa, according to the interpretation of Jennings, the return in these cases to the original condition of the cells is presumably due to the controlling effect of the nucleus over the cytoplasm. It should then be possible to perpetuate a newly acquired characteristic of a whole cell or of the cytoplasm, provided the constitution of the nucleus is not opposed to this alteration. Such a condition is realized during embryonic development, when changes take place in cells which are irreversible and the endstage of which may be transmitted from generation to generation. The egg ultimately becomes converted into various types of cells, such as those of thyroid or epidermis, and there is every indication that this change—an acquired condition—can be transmitted indefinitely to successive generations of cells. Even intermediate developmental stages, as for instance, the characteristics of an embryonal tissue cell, which is also derived from mother cells possessing entirely different characteristics, can be propagated endlessly in tissue culture. In all these instances it seems that the nucleus and the rest of the cell coöperate. However, it must be noted that these changes take place only when they are in accordance with the genetic constitution of the organism of which the cells involved form a part. Correspondingly, we would have to assume that in order to be permanent, also the transition from the state of a normal cell to that of a cancerous cell would have to be in accordance with the genetic constitution of the individual to which it belongs; and that not only ageing in general, but likewise the cancerous growth would represent a continuation of embryonal development in the same sense as other ageing processes. The formation of cancer may therefore be considered as an endstage in the ontogenetic evolution of cells or tissues which occurs under certain conditions.

As stated previously, it would then be necessary to assume as preformed in the genetic constitution of the organism, the existence of two principal cell equilibria: (1) the equilibrium of the adult, normally functioning cell, and (2) the cancerous equilibrium. These two equilibria are joined to each other by intermediate stages, represented by growth processes of varying intensity and by processes of sensitization. The beginning stage, the various intermediate stages, and the endstage may be considered as a connected series comparable to the stages leading from young adult tissues to those of the senescent organism; these again are joined together by intermediate stages and they all are potentially determined by the structure of the fertilized germ cells. Furthermore, in accordance with our earlier interpretations of the cancerous process, it would be possible to conceive of the cancerous equilibrium as being due to the increase in, or to the new creation of, autocatalyzers capable of intensifying cell growth. Associated with the stimulation of growth processes there may take place also a stimulation of ameboid movement, which may help in the transfer of the cancerous tissues to various parts of the body, and which at the same time may intensify the disorganization of these tissues.

The interpretation of the phenomena of cancerigenesis has, then, narrowed down to the consideration of the virus theory of cancer and of the theory of cancer as representing the endstage of a series of graded cell changes, due to stimuli exceeding the threshold permitting normal growth and acting on a tissue substratum endowed with the potentiality to sensitization, which varies in degree in different individuals in accordance with the genetic constitution. Such stimuli may be supplied, also, by certain viruses. There are in addition the agents of fowl sarcoma, which seem to induce the cancerous transformation directly. It is possible that these viruses instead of being living microorganisms, represent very complex macromolecular proteins associated with other groups, such as nucleins. The possibility exists,

furthermore, that they originate in the higher organisms, in which they act and function here directly as cancerigenic growth stimulators; but in contrast to other growth substances which originate and multiply within the cells, they can be transferred to other individuals of the same, or nearly related species, whose tissues they are able to invade. These viruses would thus act in a similar manner to the autocatalytic growth substances, which ultimately also would transform a normal into a cancerous cell. If finally it could be shown that the agents of fowl sarcoma and the autocatalytic growth substances act in a comparable manner, the difference between these two theories would be greatly diminished. On the other hand, the ability of viruses to undergo adaptive changes after transfer to different types of hosts, in accordance with the observation of Duran-Reynals, seems to be more compatible with the concept that viruses are real organisms, able to pass through ultra-filters, than that they are protein substances. However, the experiments of Pauling and his associates, which indicate that serum globulins, through contact with various cell constituents, can be changed into immune substances which act specifically on the constituents with which they have been in contact, suggest that also complex protein molecules may be self-propagating and readily undergo changes of an adaptive character. However, these considerations are at present of a purely hypothetical nature. Future investigations must determine whether all or the large majority of tissues represent symbionts between cells and viruses, or whether various types of cell equilibria exist, of which the cancerous equilibrium represents a possible endstage potentially present in all tissues endowed with the ability to proliferate, an endstage which is activated by stimuli exceeding certain thresholds which lead to the perpetual acceleration and intensification of the processes underlying cell multiplication and ameboid movement. The nature of this threshold and the mode of reaction of the various tissues to such stimuli are essentially determined by genetic factors.

References

1. Alexander, Jerome, and Bridges, C. B., "Some physico-chemical aspects of life, mutation and evolution," "Colloid Chemistry," Vol. II (this series) p. 9, edited by Jerome Alexander, New York, 1928.
2. Bang, F., "Processus histologique au cours de l'évolution du cancer du goudron chez les Souris blanches," *Compt. rend. de la Soc. de biol.*, **87**, 757-759 (1922).
3. Beard, J. W., and Rous, P., "Virus-induced mammalian growth with characters of tumors (Shope rabbit papilloma); experimental alterations of growth on skin; morphological considerations; phenomena of retrogression," *J. Exper. Med.*, **60**, 723-740 (1934).
4. Bittner, J. J., "Some possible effects of nursing on the mammary gland tumor incidence in mice—preliminary report," *Science*, **84**, 162 (1936).
5. Bittner, J. J., "Relation of nursing to the extrachromosomal theory of breast cancer in mice," *Am. J. Cancer*, **35**, 90-97 (1939).
6. Blumenthal, H. T., "Effects of organismal differentials on the distribution of leucocytes in the circulating blood," *Arch. Path.*, **27**, 510-545 (1939).
7. Blumenthal, H. T., "The effects of spontaneous and transplanted rat and mouse tumors on the red and white cells in circulating blood and bone marrow," *Cancer Research*, **1**, 196-204 (1941).
8. Burk, D., Sprince, H., Spangler, J. M., Kabat, E. A., Furth, J., and Claude, A., "The metabolism of chicken tumors. J. Nat., Cancer Institute," **2**, 201-240 (1941).
9. Cannan, R. K., "The metabolism of tumors," *Cancer Review*, **2**, 289-302 (1927).
10. Claude, A., and Murphy, J. B., "Transmissible tumors of fowl," *Physiol. Rev.*, **13**, 246-275 (1933).
11. Claude, A., "Concentration and purification of chicken tumor I agent," *Science*, **87**, 467-468 (1938).
12. Cook, J. W., Hieger, I., Kennaway, E. L., and Mayneord, W. V., "The production of cancer by pure hydrocarbons—Part 1, *Proc. Roy. Soc. London*, **111**, 455-484 (1932).
13. Cook, J. W., "Cancer-producing chemical compounds," *Nature*, **145**, 335-338 (1940).
14. Cook, J. W., and Kennaway, E. L., "Chemical compounds as carcinogenic agents. Second supplementary report; literature 1938 and 1939," *Am. J. Cancer*, **39**, 381-428 (1940). "II. Action of carcinogenic compounds in different species and tissues," *Am. J. Cancer*, **39**, 521-582 (1940).
15. des Ligneris, M. J. A., "Precancer and carcinogenesis," *Am. J. Cancer*, **40**, 1-46 (1940).
16. Ewing, J., "Neoplastic diseases," 4th Edition, Philadelphia, W. B. Saunders and Co., 1940.
17. Furth, J., and Butterworth, J. S., "Neoplastic disease occurring among mice subjected to general irradiation with x-rays; ovarian tumors and associated lesions," *Am. J. Cancer*, **28**, 66-95 (1936).
18. Furth, J., and Breedis, C., "Attempts at cultivation of viruses producing leukosis in fowls," *Arch. Path.*, **24**, 281-302 (1937).
19. Gardner, W. U., Estrogens in carcinogenesis," *Arch. Path.*, **27**, 138-170 (1939).
20. Hartwell, J. L., "Survey of compounds which have been tested for carcinogenic activity," Nat. Cancer Institute, U. S. Public Health Service, 1941.
21. Jensen, C. O., "Experimentelle Undersogelser over Kraeft hos Mus. Hosp." *Tid. Kobenh.*, **10**, 489 (1902); **11**, 549-581 (1903).
22. Jensen, C. O., "Experimentelle Untersuchungen über Krebs bei Mäusen," *Centralbl. f. Bakt.*, **34**, 28-122 (1903).
23. Kennaway, E. L., and Hieger, I., "Carcinogenic substances and their fluorescence spectre." *Brit. Med. J.*, **1**, 1044-1046 (1930).
24. Kennaway, E. L., "On the cancer producing factor in tar," *Brit. Med. J.*, **1**, 564-567 (1934).

25. Kensler, C. J., Dexter, S. O., and Rhoads, C. P., "The inhibition of a diphosphopyridine nucleotide system by split products of dimethylaminoazobenzene," *Cancer Research*, **2**, 1-10 (1941).
26. Kidd, J. G., "Immunological reactions with a virus causing papillomas in rabbits. 1. Demonstration of a complement fixation reaction; Relation of virus-neutralizing and complement-binding antibodies," *J. Exper. Med.*, **68**, 703-724 (1938).
27. Kidd, J. G., "Immunological reactions with a virus causing papillomas in rabbits. II. Properties of the complement-binding antigen present in extracts of the growths; its relation to the virus," *J. Exper. Med.*, **68**, 725-736 (1938).
28. Kidd, J. G., "Immunological reactions with a virus causing papillomas in rabbits. III. Antigenicity and pathogenicity of extracts of the growths of wild and domestic species: General discussion," *J. Exper. Med.*, **68**, 737-759 (1938).
29. Kidd, J. G., "The masking effect of extravasated antibody on the rabbit papilloma virus (Shope)," *J. Exper. Med.*, **70**, 583-604 (1939).
30. Lacassagne, A., "Apparition de cancer de la mammelle chez la souris mâle, soumise à des injections de folliculine." *Compt. rend. Acad. sci.*, **195**, 630-632 (1932).
31. Lathrop, A. E. C., and Loeb, L., "Further investigations on the origin of tumors in mice. III. On the part played by internal secretions in spontaneous development of tumors," *J. Cancer Research*, **1**, 1-20 (1916).
32. Leitch, A., "The effect of cessation of the irritant on the development of experimental tar cancer," *Brit. Med. J.*, **2**, 1101-1103 (1922).
33. Loeb, L., "On transplantation of tumors," *J. Med. Research*, **1**, 28-38 (1901).
34. Loeb, L., "Further investigations in transplantation of tumors," *J. Med. Research*, **3**, 44-73 (1902).
35. Loeb, L., "Ueber Transplantationen eines Sarcome der Thyroidea bei einer weissen Ratte," *Arch. f. path. anat. u. physiol. (Virchow)*, **167**, 175-191 (1902).
36. Loeb, L., "General problems and tendencies in cancer research," *Science*, **43**, 293-302 (1916).
37. Loeb, L., "Further investigations on the origin of tumors in mice. VI. Internal secretion as a factor in the origin of tumors," *J. Med. Research*, **40**, 477-496 (1919).
38. Loeb, L., "Inheritance of cancer in mice," *Am. Nat.*, **55**, 510-528 (1921).
39. Loeb, L., "Quantitative relations between the factors causing cancer and the rapidity and frequency of the resulting cancerous transformation," *J. Cancer Research*, **8**, 274-284 (1924).
40. Loeb, L., "Heredity and internal secretion in the etiology of cancer," pp. 48-60, Rep. Internat. Conf. on Cancer, London, 1928.
41. Loeb, L., Transplantation and individuality, *Physiol. Rev.*, **10**, 547-616 (1930).
42. Loeb, L., "Estrogenic hormones and carcinogenesis," in "Glandular Physiology and Therapy," Chapt. 13, p. 177, American Med. Assoc., Chicago, 1935.
43. Loeb, L., "Interaction between hereditary and stimulating factors in the origin of cancer," *Acta Union Internat. contre Cancer*, **2**, 148-191 (1937).
44. Loeb, L., "The significance of hormones in the origin of cancer," *J. Nat. Cancer Institute*, **1**, 169-195 (1940).
45. Loeb, L., "Hormones and the process of ageing," The Harvey Lectures, Series 36, p. 228, 1940/41.
46. MacDowell, E. E., "What and why is leukemia. A medico-biological study on mice," No. 501, p. 353, Carnegie Institute, Washington, 1938.
47. Moore, R. A., and Melchionna, R. H., "Production of tumors of the prostate of the white rat with 1:2-benzpyrene," *Am. J. Cancer*, **30**, 731-741 (1937).
48. Murphy, J. B., and Rous, P., "The behavior of chicken sarcoma implanted in the developing embryo," *J. Exper. Med.*, **15**, 119-132 (1912).
49. Murphy, J. B., "The lymphocyte in resistance to tumor grafting, malignant disease and tuberculous infection," Monograph of the Rockefeller Institute for Med. Research, No. 21, 1926.
50. Richter, M. N., and MacDowell, E. E., "Experiments with mammalian leukemia," *Physiol. Rev.*, **15**, 509-524 (1935).
51. Rous, P., "A transmissible avian neoplasm (Sarcoma of the common fowl)," *J. Exper. Med.*, **12**, 696-705 (1910).
52. Rous, P., "Transmission of a malignant new growth by means of a cell-free filtrate," *J. Am. Med. Assn.*, **56**, 198 (1911).
53. Rous, P., "The virus tumors and the tumor problem," *Am. J. Cancer*, **28**, 233-272 (1936).
54. Rous, P., and Beard, J. W., "Virus-induced mammalian growth with characters of tumor (Shope rabbit papilloma); growth on implantation within favorable hosts," *J. Exper. Med.*, **60**, 701-722 (1934).
55. Rous, P., and Beard, J. W., "Virus-induced mammalian growth with character of tumor (Shope rabbit papilloma); further characters of growth: General discussion," *J. Exper. Med.*, **60**, 741-766 (1934).
56. Rous, P., and Kidd, J. G., "Conditional neoplasms and subthreshold neoplastic states; study of tar tumors of rabbits," *J. Exper. Med.*, **73**, 365-390 (1941).
57. Rous, P., and Murphy, J. B., "Tumor implantations in the developing embryo. Experiments with a transmissible sarcoma of the fowl," *J. Am. Med. Assn.*, **56**, 741-742 (1911).
58. Rous, P., and Murphy, J. B., "The histological signs of resistance to a transmissible sarcoma of the fowl," *J. Exper. Med.*, **15**, 270-286 (1912).
59. Shear, M. J., and Leiter, J., "Studies in carcinogenesis. XVI. Production of subcutaneous tumors in mice by miscellaneous polycyclic compounds," *J. Nat. Cancer Institute*, **2**, 241-258 (1941).
60. Shear, M. J., and Leiter, J., "Studies in carcinogenesis. XV. Compounds related to 20-methylcholanthrene," *J. Nat. Cancer Institute*, **2**, 99-112 (1942).
61. Slye, M., "The incidence and inheritability of spontaneous tumors in mice," *J. Med. Research*, **30**, 281-298 (1914).
62. Staff of the Roscoe B. Jackson Memorial Laboratory, "The biology of the laboratory mouse," Philadelphia, Blakiston, 1941.
63. Tannenbaum, A., "The initiation and growth of tumors. Introduction. I. Effects of underfeeding," *Am. J. Cancer*, **38**, 335-350 (1940).
64. Warburg, O., "Über d. Stoffwechsel. d. Tumoren," Berlin, 1936.
65. West, P. M., and Woglom, W. H., "The biotin content of tumors and other tissues," *Science*, **93**, 525-527 (1941).
66. Woglom, W. H., "Studies in cancer and allied subjects. Vol. 1. The study of experimental cancer," New York, Columbia University Press, 1913.
67. Woglom, W. H., "Critique of tumor resistance." *J. Cancer Research*, **7**, 283-311 (1922).
68. Woglom, W. H., "Immunity to transplantable tumors," *Cancer Review*, **4**, 129-214 (1929).

* The list of references given is incomplete; only papers on which the views expressed are mainly based and publications which contain further references are included. Additional literature will be found in the *Journal of Cancer Research*, in the *American Journal of Cancer*, in *Cancer Research*, and in the *Journal of the National Cancer Institute*.

Gerontology: The How and Why of Ageing *

A. J. CARLSON †

Department of Physiology, University of Chicago

The science of biology and medicine has not yet reached the stage when a factual chapter on the physiology of ageing can be written. At the end of an excellent fifty page review of the chemical aspects of ageing, McCay concluded (1939) that "almost nothing is known of the biochemistry of ageing." The physiologist need not be quite so pessimistic, but instead of writing on what we do not know, biochemists and physiologists might be better employed in research on these important unknowns.‡ Many age changes in the animal body have been clearly established. The cause or causes of these age changes in structure and function are still obscure. We are not in a position to determine to what degree the changes in the machinery of the body, which parallel advancing age, are due to the genetic constitution of the individual and the species; that is, to the ageing process *per se,* and how much are due to the accidents of living, such as faulty diets, infections, overwork, laziness, gluttony, and what have you. As regards the life span of the species, plant or animal, undoubtedly the genetic constitution plays the primary role, for on no other basis can we account for some species of plants living only one year or less and other species living several thousand years. Disease and other accidents of living cannot account for such great differences in life span in plants, nor can it account in my judgment, for similar differences in the life span in different species of animals.

There is growing evidence that the hereditary constitution is also a factor in the life span of individuals within the species, irrespective of ignorance, faulty diets, infections, and other accidents in the life of the individual. This appears also to be true for the life span of individual organs in the individual. For while it is true that faulty diets, infections, etc., may hasten the atrophy of the ovaries in primates and thus bring on a premature menopause, that similar accidents and disease may hasten the loss of elasticity of the lens of the eye, it is nevertheless true that the life span of these two organs, the primate ovaries and the lens of the eye, run on the whole such a fairly regular course in the individuals of the species that the early atrophy of the ovaries in the female and the relatively early decrease in and loss of elasticity of the lens can hardly be accounted for by the accidents and stresses in

* As this Chapter is written primarily for Physicians, familiarity with the normal physiology of the human adult is assumed, or can be renewed by the aid of any up-to-date medical textbook of physiology, biochemistry, and physiologic pathology. The lay reader may find this volume a useful aid: *The Machinery of the Body,* by A. J. Carlson and V. Johnson, University of Chicago Press, revised edition, 1941.

† Parts of this essay will appear in a volume on geriatric medicine, Dr. E. J. Stieglitz, Editor, Saunders Co., 1943.

‡ Some colloid chemical aspects of ageing are referred to by H. Schade and by A. Lumiere in Vol. II of this series. Dr. Carlson's authoritative summation of the clinical aspects and their interrelations poses many problems which must be considered in the light of a number of the theoretical and biological papers in this series. The statue by Gian Bernini (Borghese Gallery, Rome), representing the stalwart Aeneas carrying the shrunken form of his aged father Anchises and leading his chubby son Ascanius, beautifully illustrates the results of the complicated colloid chemical changes involved in ageing. J.A.

the life of the individual. The life span of these two organ systems must be primarily determined by the physiological time clock, perhaps particular genes of the hereditary constitution. But when we consider the other organ systems of the body, the slowing up of function, or the impairment of function is on the whole so gradual (at least in all cases where the life of the individual is not specifically terminated by *overwhelming* accidents in individual organs such as malignant growths, gastric ulcer, pneumonia, tuberculosis, nephritis, or arteriosclerosis) that it becomes practically impossible to separate the factors of ageing from the factors of impairment due to the accidents of disease. Of course, the germ cells themselves may be acted on unfavorably by the accidents and strains of living, and the poisons of metabolism, even though the germ cells are not the link in the life chain of the individual; at least this may be so in the case of the species. If this were true in the case of the species, we would have degeneration of species and shortening of the life span of the species and this would be a rule instead of the exception. The relative importance in the life span and in the process of senescence of the genetic constitution as opposed to the environment is illustrated by the fruit fly, where, with the identical environment but with experimentally modified genetic constitution, the life span may vary as much as 400 per cent, the most extensively modified genetic constitution leading to the earliest senescence. To be sure, it is a long distance from the fruit fly to man, and by and large, science has yet been able only to degrade rather than to strengthen the genetic constitution by experimental means. Nevertheless, similar factors probably play a role in the phenomena dimly foreshadowed in the life of man, that is, in the relatively long lived and relatively short lived human families.

If further justification of our primary thesis, namely, the practical impossibility of separating the factors of ageing *per se* from the effects of the accidents of living is needed, we might cite the case of many plants in which the tissue degenerations once thought to be due to ageing (that is, to the genetic constitutions) seem now to be due to virus disease, an injurious factor in living, of which we have at present only the scantiest of information. And in plant seeds, age degeneration seems to be due to progressive chemical changes in the cell nuclei, changes which in turn interfere with cell division and growth.

Unicellular animals are commonly said to be immortal. Nevertheless, these lowly cousins do seem to degenerate with age. Again, this degeneration appears to be in the physiologically active nucleus, and unless this aged nuclear material is replaced by the reserve micronucleus, as occurs in conjugation and endomyxis, decline and death of these unicellular organisms seem inevitable.* This is, in principle, not so different from sexual reproduction and death of the soma or body in the higher plants and animals.

Despite this discursive and disappointing introduction, the following principal age changes in the physiology of man and other animals can be listed, appearing at different age levels in the individual and at different age levels in the organ systems of the individual. The universality of these age changes, even in individual persons whose hereditary constitutions have been able to meet the accidents of living for 100 or more years, renders it highly probable that we are here dealing with age changes primarily inherent in the constitution of living matter, no matter how greatly they may be speeded up by the accidents of living.

Progressive age changes, not as yet shown to be due to specific diseases:

1. Gradual tissue dessication.
2. Gradual retardation of cell division, capacity of cell growth and tissue repair.
3. Gradual retardation in the rate of tissue oxidation (lowering of the B. M. R.).†

* See paper by H. S. Jennings in this volume. J. A.

† B. M. R. is basal metabolic rate, that is the rate of oxidation when the body is at rest and the individual has rested reclining at least 30 minutes prior to test. No food should be taken over the preceding 12 hours.

4. Cellular atrophy, degeneration, increased cell pigmentation, and fatty infiltration.
5. Gradual decrease in tissue elasticity, and degenerative changes in the elastic connective tissue.
6. Decreased speed, strength, and endurance of skeletal neuromuscular reactions.
7. Decreased strength of skeletal muscle.
8. Progressive degeneration and atrophy of the nervous system, impaired vision, hearing, attention, memory, and mental endurance.*

This picture of ageing is, in fact, not as dark and depressing as it seems, for the following reasons: (1) An efficient mental life is possible, even with reduced factors of safety. (2) Moderation in all things, plus competent medical guidance in regard to the accidents of disease will tend to prevent undue corrosion of specific life links, so that all of them pass down the hill gradually and in step. (3) The man or woman who has passed the first 60 or 70 years in honest toil and persistent efforts at understanding has accumulated a *great reserve of wisdom,* now at the disposal of the younger generation; a reserve of wisdom which antidotes the stresses and strains induced by the fears, the vanities, the greeds, and the ignorance of the earlier years. This is true at least up to the point where the impairment of cerebral function leads to the so-called "second childhood." Intensive research on the many unknowns in the ageing processes will some day enable the science of medicine to delay greatly this involution. This is not "rejuvenation" either in the popular or in the medical quack sense. *It is merely better care and better driving of the living machine, provided by our growing understanding. This is the goal of geriatric medicine.*

The Digestive System

The salivary glands usually show some evidence of atrophy with advancing years, and competent investigators have reported definite decrease in salivary volume and in percentage of ptyalin with advance in age. This is not a vital link in the life span since most animals have no ptyalin and we can adequately digest all our food without the aid of ptyalin, the most important constituent of saliva being the water and mucin. To what extent the decrease in saliva or the age changes in the composition of the saliva permit or favor growth of the aciduric flora of the mouth, which in turn predispose to caries and enhance degenerations in teeth and of gums, is still an open question.

The work of Blumfield and Poland indicates that from the age of 20 years on, the volume and the acidity of human gastric juice decreases definitely but gradually, and it is well established that the incidence of achlorhydria increases with age. Thus achlorhydria appears in only about 5 per cent of people under the age of 20 years, while it appears in 35 per cent in population above the age of 60. Again, we must note that the higher incidence of achlorhydria with age may be due to the accidents of living rather than to the process of ageing. Nevertheless, the existence of achlorhydria in a child of 5 and the presence of hydrochloric acid normally in gastric juice of a person at the age of 100, points to an hereditary time clock in this function of the gastric mucosa, as the primary factor, no matter how greatly this time clock may be accelerated by the accidents, the ignorance, and the follies of living.

It is also well established that the incidence of pernicious anemia increases with age, indicating in all probability an ageing factor in the gastric component or intrinsic factor of red cell formation. The gastric pepsin factor seems less subject to ageing and to the accidents of living than does the hydrochloric acid factor.

* These changes are not listed in the order of their primary significance, for this is not yet known. Nor does the list imply that these are all the changes of senescence. It is already evident that ageing is a continuous and complex process, some changes starting early, others later in the life of the individual.

On the motor side of the stomach, in the absence of such diseases as cancer or ulcers, there is little clear evidence except that dealing with the strength and rate of the contractions of the empty stomach, that is, the so-called hunger contractions. These do definitely decrease with age, as shown particularly in man and dog. The periods of relative motor quiescence of the empty stomach are shorter in the young, become much longer in advancing years, and the periods of active contractions shorten, and the contractions themselves become feebler with advancing years. But this evidence of impairment of gastric motor activity with ageing appears to be sufficient to cause significant delay of the emptying time of the stomach, but not significant enough to produce impairment of digestion.

Atrophic gastritis might readily be a chronic effect of dietary indiscretions if not of chronic infections, such processes wearing out, as it were, the regenerative or recuperative capacities of the gastrointestinal mucosa. However, there is no good evidence of any gradual increase of atrophic gastritis with age.

In the case of the pancreas, the incidence of diabetes increases with advancing years, at least up to 50 or 60. The pancreas shows many scars from accidents and injury of living. Goodpasture has described degenerative changes in the pancreas of old dogs and similar changes have been noted by Allen and others in post mortem examinations of elderly people dying from a variety of causes. But we have no reliable data on deficiency of pancreatic juice or pancreatic juice enzymes with advancing age uncomplicated by pancreatitis, liver, or biliary tract disease. The fact that in some individuals, diabetes may develop in early childhood, while in other individuals the hormone mechanism of the pancreas is able to meet the metabolic need of an individual 100 years old, might induce one to conceive that the failure of this phase of pancreas function with increasing age, is due to hereditary weakness rather than to the accidents of living or duration of life.

While the gross weight of the liver decreases with age, at least from the fourth to the seventh decade, the present available liver function tests do not reveal any liver impairment with ageing in the absence of specific liver disease. In fact, the liver of an old dog appears to be more resistant to the injurious effects of chloroform than is the liver of a pup, as is also clearly the case in the human species. But it is well known that the incidence of liver pathology increases with age and there appears to be some decrease in concentration of vitamin C in the liver with the advancing years. Whether or not this is secondary to decreased absorption of this vitamin from the intestine is not known. The incidence of gall stones increases with age, but the gall bladder, free from disease, apparently empties normally even in very old people and very old animals.

There appear to be no reliable data on atrophic changes in the mucosa and in the musculature of the small intestine with advancing age. To be sure death from intestinal obstruction increases with advanced age, but this may be due to adhesions rather than to weakness in the intestinal musculature. Atrophy of the colon musculature with advanced age is uncertain. The motility and absorption in the colon appears normal in the aged in the absence of definite colon disease, but diverticulosis of the colon seems definitely to increase with age from the fourth decade on. This may indicate atrophy or decreased tone of the colon musculature or decrease in the elasticity of the connective tissue layers of the colon. According to Dr. Ivy, "Death in the aged is apparently only rarely due to the wearing out of the organs of the digestive system. In the absence of gastro-intestinal cancer or local toxic or infectious processes, the digestive system seems capable of functioning beyond the ordinary life span."

The Cardio-Vascular System

Failure of health to the point of death, primarily due to failure of the cardiovascular system, is known to increase with age, at least beyond the fourth or fifth

decade of life. At the same time both cardiac and blood-vascular efficiency may persist in man to a degree rendering life possible, if not efficient, at the age of 100 years. But in the way of comparison, either the hereditary time clock is shorter or the accidents of living are more serious in the case of the cardio-vascular system than in the case of the digestive system. Also, at present, it is somewhat more difficult to separate purely age factors from accidental disease in the cardio-vascular system.

Instead of atrophy with age, there is, with advancing years, an actual increase both in size and weight of the heart—due in part perhaps to cardiac incompetency, even when the increased fat in the heart of old people is deducted, but this does not mean increased efficiency; in fact, it may be secondary to cardiac incompetency. There is decrease in the elasticity of the heart valves, and increased calcium or cholesterol deposits in the valves. There is increased thickness of the endocardium, there is an increase in brown pigment in the muscle cells, and a diminution of the cross striation of the muscle cells, especially close to the nuclei.

On the functional sides, cardiac irregularities such as premature contractions and auricular fibrillations show a tendency to increase with advancing years. Cardiac irregularities from hyperthyroidism are more frequent after the age of 45, and, while this may be an indication of decreased cardiac reserve, it may also mean an increased incidence of hyperthyroidism beyond that age. There is a decrease in oxygen consumption in the heart with advancing years, the heart in this respect showing a parallel with the gradual lowering of the general basal metabolic rate of the body as a whole. The electrocardiogram shows lessened voltage and some slowing of conduction in old people. This probably indicates some impairment of the conduction system. On the other hand, there appears to be an increased sensitivity of the carotid sinus and depressor cardio-inhibitory reflex with the advancing years.

Where spasm or atheromatosis of the coronary arteries intervene, we obviously have the same impairment of the heart as occurs in all other tissues where the efficiency of the circulation is impaired, by chronic functional spasms of the blood vessels, or by mechanical narrowing of the lumen and by decreased elasticity and calcification of the blood vessel walls.

The old saying that a person is as old as his arteries is still partly true. There is decreased elasticity of the aorta, decreased elasticity and increased calcification in all arteries in all people with advancing years. But these ageing degenerations or disease factors do not appear to the same degree in all arteries at the same age in the same individual, or to the same degree at the same age in different individuals. So here again we are faced with the problem, how much are these arterial impairments due to ageing *per se*, and how much to the accidents of living. If it should turn out, as the work of L. R. Dragstedt indicates, that one main factor in the genesis of arterial sclerosis is faulty fat metabolism induced by the deficiency in a hormone or hormones necessary for normal fat metabolism, we are well on the road to a partial understanding and possibly prevention of this well nigh universal corrosion of a vital link in the life chain of the individual. Snapper suggests that the practical absence of meats and milk in dietary of the people in North China may be a factor in the low incidence of arterial sclerosis in that population. In view of the work of Dragstedt, the overall low caloric diet in that part of China may be a more significant element.

Owing to the gradually decreasing arterial elasticity with increasing age, there is a gradual increase in the speed of propagation of the arterial pulse rate from an average of little over 6 meters per second at the age of 20, to around 8 meters per second at the age of 60. Parallel with these changes in the structure and function of the blood vessels, there is, at least beyond the age of 40, a gradual increase in the systolic blood pressure even in the absence of so-called essential hypertension and in the absence of an excess rate of the heart beat. Some investigators have estimated

this rise in systolic blood pressure beyond 40 as around 1 mm of mercury per year. This is, of course, subject to a great deal of individual variations and to many exceptions, since the systolic blood pressure depends not solely on the condition of the arteries, but also on the efficiency of the heart and on the cardio-regulatory reflexes. Judging from the appearance of the capillaries in the human skin, there appears to be decreased cellular tone leading to capillary dilatation with advancing age. In the kidneys, actual atrophy of the capillary tufts of the glomeruli with advancing age has been described.

The Kidneys

The kidneys, like the digestive system and the cardio-vascular system, serve the entire physiological economy of the individual, hence serious age impairment of either necessarily impairs the function of all. The kidneys show progressive reduction in weight after the fourth decade. The rat's kidneys show progressive involution of the glomeruli after the fourth month of life. In some fish the glomeruli disappear entirely in later life, and the renal organ becomes an aglomerular kidney. The one year old rat exhibits about 50 per cent reduction in hypertrophy of the remaining kidney after extirpation of the other kidney, as compared to the renal hypertrophy of the remaining kidney following this operation in rats only 30 days old. McNider has shown that the kidneys of old animals display increased susceptibility to injury by poisons (ether, chloroform, uranium), and decreased power of regeneration after such injuries. But whether this impairment of growth and repair is primarily a matter of the hereditary renal time clock or aggravated by the accidents of living is an open question.

In the ageing individual, renal involution or atrophy (tubular and glomerular) appear to be secondary to impairment of renal blood flow due to structural involution of the renal blood vessels. But despite such renal histo-pathology of the ageing man, the kidney usually has sufficient reserve, in the absence of specific renal disease, to meet the requirements of the individual at the age of 100 years and beyond. It is now well established by Goldblatt and others that decreased renal blood flow initiates functional hypertension of renal origin. Is this renal anemia (ischemia) initiated or merely aggravated by the accidents of living, that is, is it an unavoidable turn of the renal time clock of the individual? The fact that some individuals can reach the age of 90 to 100 years without sufficient renal ischemia to induce hypertension would seem to indicate that the accidents of living are by themselves not a sufficient cause for this type of renal failure. But it is clear that the serious affliction of the renal life link, by inducing chronic hypertension and further impairment of the circulation, will seriously accelerate the depletion of reserves in all other organs and not only shorten the life span but seriously weaken the life of the aged, years before the death of the individual.

The Neuro-Skeletal System

The gradual slowing and weakening of reflexes and general body activity in the ageing mammal is so obvious as to be well known both to physicians and laymen. Decreased functional capacity both in the nervous tissues and in the skeletal muscular tissue seems to be at the base of this gradual decline. Actual atrophy of the Purkinje cells of the cerebellum has been described in the aged and, since this part of the nervous system is seriously concerned with skeletal muscle tone and coördination of skeletal muscle contractions, it may be a factor in the growing muscular weakness of old people irrespective of the cause or causes of this atrophy in the cerebellum.

As regards the cerebrum of aged people, general atrophy has been described especially in the frontal and occipital lobes, and actual disappearance of cells in

some of the layers of the cerebral cortex, as well as pigmentation and fat infiltration of the nerve cells and actual hyperplasia of the neuroglia cells. Similar degenerative changes with age occur in the spinal cord, that is, atrophic pigmentation, actual loss of cells, and degeneration of the axones of many ventral horn cells. In the case of the brain, thickening of the meninges occurs with advancing age, but it is difficult to see how this in any way should interfere with nervous action or nervous function.

Recent investigations appear to demonstrate a very gradual but significant decrease in the myelinated fibers of the dorsal nerve roots with advancing age. This must be secondary to an atrophy and death of spinal ganglion cells and is probably the basis of the reduction in cutaneous and protopathic sensibility of aged people. The sense of pain seems to be the least affected by ageing. In view of such evidence of atrophic and degenerative changes in the central and peripheral nervous system, irrespective of the primary cause or causes of these changes, it is not surprising that neuromuscular weakness, slowing of the reaction time, decreased capacity to learn, etc., etc. is part and parcel of the physiology of ageing. The speed of learning seems indeed to decrease gradually in man from the fourth decade on. But this handicap of the aged is on the whole more than made up for in some individuals in their greater speed of correlation and evaluation of the new experience.

There is very little evidence of ageing changes in smooth muscle. Smooth muscle seems on the whole to retain its normal histologic character into advanced old age. The diminished tone in smooth muscle, as may be seen in the blood vessels in the gut, and in the smooth muscles of the skin and other structures in old people, may be secondary to the impairment in the nervous system, indicated above. But not all the impairment of body motility with age can be ascribed to the degenerate changes in the nervous system itself because the striated skeletal muscle system shows fatty infiltration and brown atrophy with advancing age. The strength of the biceps at the sixth decade of life is only about 50 per cent of that at the age of 25 to 30. The trunk muscles decline in power somewhat slower. However, the recent investigation by Kubo (1938) reports little evidence in decrease in muscle strength and endurance in people that would ordinarily be called old, that is, people 70 to 90 years of age. There is some increase in connective tissue and elastic fibers in the skeletal muscle of old people and there is clear evidence of dessication, that is, decrease in intracellular fluid. But in this respect, the skeletal muscle of the aged falls in line with all the other tissues of the body.

Vision and Hearing

Because of the accessibility of the organs themselves and the availability of quantitative tests of functions, we have more accurate information regarding the ageing changes in the physiology of the eye and the physiology of the ear than is the case with most of the other systems in the human body.

(a) Vision: In the case of vision, there is a gradual decrease in visual acuity (central vision), and a gradual narrowing of the visual field, as well as a slowing of the dark adaptation (peripheral vision), and a gradually higher threshold for light stimulation for man past the fourth decade. The narrowing of the visual field is probably due to the actual degeneration of the nerve cells (cones), starting in the periphery of the retina. We are not yet in a position to say whether these visual impairments occur independent of, or are secondary to, impaired retinal circulation.

It is equally well known that the incidence of cataract increases with age irrespective of whether or not the tendency to cataract formation is hereditary. Arterial sclerosis would undoubtedly accelerate any such hereditary weakness, and so would certain faulty diets and certain endocrine and other metabolic disorders.

The gradual decrease of the elasticity of the lens is another well known and

accurately measured phenomena of ageing man, with this exception that diminution in lens elasticity actually starts in childhood and practically all lens elasticity is lost before 60 years of age. The lens continues to grow at the periphery (vertex), and thus approaches closer and closer to the cornea with advancing years. At the same time the material at the center of the lens becomes more dense. Both of these factors, and the lens swelling from increased water content, are responsible for the well known phenomena of so-called "second sight" of people 60 years of age and beyond. This lens change tends to counteract the presbyopia, or impairment of accommodation, due to the loss of lens elasticity.

Other age changes that may contribute to the gradual impairment of vision with age are diminished translucency of the cornea and of the vitreous humor. It need not be pointed out that the retina, being actually a lobe of the brain, is necessarily as seriously impaired by local vascular pathology as is any other part of the brain. However, because of the accessibility of the retinal vessels to direct inspection, we have probably earlier factual information regarding such pathology in the retina than we have in most of the other deep organs of the body.

(b) Hearing: From the age of about 20 on, there is a gradual loss of acuity to all tones, but the loss of sensitivity is greater to the high tones. This deterioration of hearing is somewhat greater in the male, but the degree of retrogression is not predictable in chronological age, as some people at 80 have no greater auditory impairment than other normal people at 50. This impairment of auditory acuity is present even when tested by bone conduction. The cause for this decline in auditory acuity appears to be a gradual but distinct atrophy of the nerve cells in the basal coil of the cochlea. But anemia, due to incipient arterial sclerosis, may also be a factor, since in experimental anoxia, the perception of the high tones goes out first.

We have at present no information regarding impairment of vestibular reflexes with advancing years. One would expect such impairment on the basis of the evidence of age atrophy in the central nervous system.

The Endocrine System

The Gonads: Owing to the relative ease of observation as well as experimentation and measurements of such phenomena as menstruation, size and function of secondary sex characters, libido, sex behavior, etc., we have the clearest information, at least regarding the factor of ageing, in the case of the ovaries and the testes. In the case of the gonads, we have the additional special phenomena, that, irrespective of their probable role as hormone producers during intrauterine life and the preadolescent years, they are organs that attain their special function rather late and lose their function rather early in the life of the individual. It is now certain that the gonads are not a link in the life chain of the individual, their absence does not shorten the life span. Prolongation of youth, or of life itself, by gonad hormone therapy is a mirage, a product of wishful thinking.

The rate of decline of the human ovaries at the normal menopause may be as speedy as a few months or may be as prolonged as a few years, this varying with the individual woman. When ovarian atrophy appears much earlier than the 35th to 45th year, one can usually find specific disease contributing to such decline, although we cannot with certainty eliminate specific hereditary weakness in the constitution of the individual, except where destructive tumors or serious impairment of thyroid or pituitary function are clearly involved. The life span of the testes is longer and the decline is more gradual. The hormone production mechanism in the testes, the factor involved in libido and copulation, appears to fail earlier than the spermatogonia and the sperm producing factors. At least in individual instances in the human male, sperm production (not necessarily normal sperm) has been described up to, if not beyond, the ninth decade of life. In the ageing human testes, the basement membrane of the tubules is reported to grow thicker, and the

spermatogonia change or retrogress into the small round cells seen in the immature testes.

In the case of the interstitial cells, increased pigmentation has been described in the old testes. But there is at present no evidence of decrease in the number of interstitial cells. It is difficult to interpret the usual decrease in libido and sexual potency in the aged human male as related to the gradual atrophy of the testes, in view of the fact that in some males between 20 and 35, libido and sexual potency appear to persist to some degree even after total castration. In other words, the mental and vasomotor conditioning associated with the above physiological processes apparently can persist in the total absence of the primary conditioning factor, the male hormone.

Increased pigmentation and gradual atrophy of the acini of the prostate gland have been described in old people and old males of other species, but there appears to be no decrease in gross size of the prostate, despite the fact that castration in the young male leads to atrophy of the prostate and its adnexa. We are here facing the well known dilemma in this physiological system, namely, the dependence of the prostate on the male hormone during adult sexual years, the independence of the prostate of the male hormone during the later decades of life, and the frequent appearance of actual prostate hypertrophy parallel with the decline of the testicular function in old men.

Thyroid. As regards this endocrine gland, we have evidence indicating some shrinkage in size and some evidence of cellular atrophy in the thyroid in old people and old animals. Some people have inferred that this is a factor in the decreased oxidation of tissues with advancing years, but since the thyroid of youth and adult life have such large factors of safety or reserves, we must have more direct evidence than this for such a far reaching conclusion. So far as we know, the thyroid gland even of very old people is capable of meeting the hormone needs of the body, in the absence of specific thyroid disease or iodine deficiency in the diet.

Pancreas. In the case of the pancreas and insulin, the evidence of increased histopathology in the pancreas with advancing years, the fact of increased incidence of actual diabetes, at least up to the age of 50 or 60, as well as the gradual decrease in sugar tolerance with advancing age, as evidenced by the glucose tolerance tests, all these facts point to some ageing in the insulin producing mechanism parallel with the advancing years. The perennial puzzle in this field is the evidence in such animals as rabbits, guinea pigs, and dogs, of the great regenerative power of the pancreas, regeneration both of island tissue and the pancreatic juice producing cells and the apparent absence of such regenerative power in the pancreas of the human species, at least in those people whose pancreatic reserves have, through heredity or the accidents of living, been reduced to the point of actual diabetes. But whether we are here dealing primarily with ageing or with accidents of living, it is nevertheless a fact that, in some individuals at least, the hormone link of the pancreas may stand the wear and tear of living up to and beyond the age of 100 years in man, provided the pancreas reserves have survived or met the needs up to the age of 50 or 60.

The work of L. R. Dragstedt and associates on the second pancreas hormone or hormones (Lipacaic) as a fundamental factor in normal fat metabolism may be a more important element in the ageing processes than is insulin, for fatty degeneration or infiltration of cells and tissues (such as the blood vessels in sclerosis) seems to be an almost universal phenomenon in the ageing and the very old animal. We have no information on the requirement of lipocaic with age, or the ability of the ageing pancreas to supply this necessary factor.

Pituitary. As regards the pituitary gland, structural changes have been described in the hypophysis of old animals as well as old people, but in the absence of specific pituitary disease, this endocrine gland appears to be remarkably stable at least up to the age of 80 or 100 years. There is no evidence at present for ascribing

the early failure of the ovaries, the later failure of the testes, and the alleged atrophy and hyposecretion of the thyroid in advancing years to decreased production of the gonadotropic and the thyrotropic hormones respectively.

Adrenals. Armchair biologists have found little difficulty in pointing out the almost perfect parallel between the symptomatology of ageing and that of impaired adrenal cortical function, such as general asthenia, impaired digestion, increased pigmentation of the skin and low basal metabolic rate. Thus one might make out a plausible case for gradual adrenal-cortical failure as a causative factor in ageing. But the usual arterial hypertension with advancing age does not fit this picture, and significant adrenal cortical failure is always associated with the disturbance in the balance of the inorganic salts of the blood plasma, a phenomena which is absent in old people. There is at present no evidence that deficiency or excess in adrenal hormone production is a significant factor in ageing.

Parathyroid. The parathyroid hormone is important in the physiology of bone and in the regulation of the calcium and phosphorus of the blood. The latter is usually practically normal in the old and the very old people. The bones of the aged are not normal, but obviously we cannot ascribe the bone impairment of the aged to either hypo- or hyper- function of the parathyroid glands.

Thymus. The thymus was once classed with the endocrine glands and it may indeed have some endocrine function in the fetal and preadolescent life of man, as well as in animals below man (possibly the birds). But after adolescence or body maturity, the thymus becomes essentially a lymph gland, the findings of Rowntree and associates that injection of thymus extract induces accelerated growth and precocious sexual maturity in rats have not been substantiated, and the still earlier work indicating that the thymus may be necessary for the life of bone, if not for the life of the individual animal, was long since wiped off the board by the conclusive work of Park and McClure. Indeed the so-called "persistent" thymus was once thought to be the cause of sudden death.

Bone, Teeth, and Connective Tissue

A gradual solution and absorption of bone in old mice has been described. In man, the bones become more brittle with advancing years. This may be due to a change in the chemical combination of the bone calcium, a change towards a more amorphous state. The demineralization of bone that proceeds with advancing age does not go on with equal rate in all the bones of the body. Todd is inclined to charge the age changes in bones and connective tissue to arterial sclerosis, that is, to impaired circulation and nutrition. There is a gradual reduction in the red bone marrow in the vertebra, and decrease of the lecithin in the bone marrow with advancing years. Parallel with these structural and chemical changes in bone there is, at least in the very old, a decreased capacity of bone healing in the case of fractures.

In cartilage, atrophic and regenerative changes in advancing age tend to proceed parallel. In true tendons and ligaments, age-changes, apart from specific disease and malnutrition, have not been made out.

In the case of teeth, decreased permeability of the enamel, thickening of the secondary dentine, and increased calcium content with age have been reported. In some herbivora, an actual mechanical wearing out of the teeth from mastication occurs. Such mechanical wearing out is surely a minor factor in modern man. The decay and failure of teeth with age in modern man is probably more a matter of nutrition and other accidents of living than the rapid ticking of the hereditary time clock.

Lymphatic System

The changes in the number and kind of lymphocytes and leukocytes with age does not seem significant. These cells are individually short lived, and the tissues

producing them seem to be able to do this job adequately, at least up to the age of 90 to 100 years. Nevertheless, there is some evidence of atrophy and fatty infiltration of the lymph nodes in old animals and old people. The lymphoid tissue in the vermiform appendix decreases with age. There is some decrease in size of the tonsils and adenoids in people 50 to 80 years old, and a similar but slight decrease in the lymphoid tissue of the spleen. On the whole, both the ageing factor *per se,* as well as the accidents of living, appear to deal lightly with the lymphatic system.

The Blood

The blood of the newborn and the very old is almost as accessible to quantitative study as are the more superficial tissues of the human body, and such studies have grown apace in the last 50 years, driven not so much by the curiosity concerning ageing as the curiosity and necessity for information in relation to specific diseases. Increased fragility of the erythrocytes with age has been reported. But in the absence of specific disease, this is not sufficient to cause anemia in old age. Indeed in the entire blood system, both cell count and cell character, plasma volume and plasma composition, there is a greater uniformity in this tissue throughout the entire span of life than in any other organ of the body. It is, however, probable that all the subtle age changes in the blood of man have not been discovered, because of lack of adequate methods. It seems to be a fact that the serum of old chickens retards growth and also induces fatty degeneration in connective tissue cells in tissue cultures, in comparison with plasma from chick embryos. Nevertheless, in the absence of serious impairments in diets, in the renal, respiratory, circulatory, and endocrine mechanisms, the fundamental homeostatic mechanisms in which the blood plays such an important part (body temperature, acid base balance, blood sugar level, regulation of blood pressure and respiration) are able to meet the ordinary needs of the individual man or woman at the age of 100 and beyond. But the blood is not able to meet the extraordinary needs of excessive strain, at least not to the degree that similar physical or dietary strains are met in people of 15 to 40 years of age. This loss or diminution in the factors of safety with age, on the basis of the present knowledge, is probably to be ascribed to impairment of kidneys, lung, and circulation, and capillary permeability rather than to changes in the blood itself.

The Skin

The common changes with age in the hair, skin, and nails are familiar even to laymen. In the case of the hair on the human scalp, it is also known that acute infections, disorders of metabolism, and endocrine disorders, intensify these involutional changes. The skin changes include increased pigmentation of the exposed parts, decrease of water, decrease of fat and decrease of elasticity in the skin, as well as decrease in its growth and regenerative capacity. According to Todd, a skin wound of 40 sq cm in a person 20 years old heals on the whole in 40 days. The same size skin wound in a person 40 years old requires for healing about 80 days. At the age of 60, the same skin would require 5 times longer to heal than in a child of 10. In this regard the skin follows the descending curve of regenerative capacity with age seen practically in all the other tissues of the body. The decrease in skin elasticity appears to run parallel with the actual degeneration of the elastic connective tissue fibers. The decrease in the subcutaneous fat may play a minor role. The dryness of the aged skin is probably secondary to decreased secretion of the sebaceous glands, although these glands show, on the whole, hyperplasia with the advancing years.

On account of its cosmetic significance for modern man, the failure of color and the gradual falling out of the hair of the scalp has received a great deal of unscientific attention. There are indications of an actual hereditary factor. Diet, disease, and vascularity, unquestionably play a role, but in the absence or with the minimum

of these factors at play, there seems no doubt that the hair on the scalp at least would sooner or later run its course of graying and death, considerably before the zero hour of the time clock for the individual person. The complexity and the individuality of some of these mechanisms, hereditary and accidental, are indicated by the fact that parallel with the graying and falling out of the hair on the scalp, some of the body hairs, such as eyebrows, the hair in the nose, and on the external ear, may show an increased growth if not an increased pigmentation with the lengthening shadows of the accumulated years.

Except for immunities to specific infection and the pathological hyperplasias and tumors, all the changes in the machinery of the body which parallel, and, in fact, constitute the ageing process point to a gradual depletion of the hereditary tissue reserves or "factors of safety," *so that the ageing individual becomes gradually less able to meet the usual accidents and the unusual stresses of living.* As long as these can be avoided life goes on until the corroded life links break, even in the absence of specific storm and stress. But this way out appears to be very exceptional even in our own species.

Diet and Life Span

Since the scientific information on the kinds and the quantity of foods required for growth and health is very recent, considering the history of the human race, and that this information is still fragmentary, it is evident that in the past man's diet, both as to kind and quantity, was determined by appetite, hunger, and the availability of the foods. We may therefore assume that underconsumption and excess consumption of foods, starvation and gluttony, emaciation and obesity, are phenomena as ancient as man himself.

Life insurance statistics show clearly that underweight as well as obesity tend to shorten the life span of man. McCay finds that on a diet adequate in quality, but inadequate in amount fed so that growth of the animal is practically at a standstill, the life span of rats is significantly prolonged. In this experiment the control group of rats was fed the same kind of food ad libitum, so that this group may have had its life span shortened by excess eating, that is, by obesity. It is well established that many, if not all, persons showing marked and persistent underweight or emaciation have functional disorders interfering with the digestion, absorption, or the metabolism of food stuffs, excess metabolic rate, or chronic disturbance of appetite and hunger (anorexia nervosa). That such persons die prematurely presents no particular problem, since we have here clearly an excess corrosion of essential links in the life chain. But why does the overweight person, the obese person, who seems otherwise healthy, die prematurely? This is a fundamental problem in geriatric medicine, for in the absence of specific disease, overweight or obesity tend to increase in men and women past 40 years of life. Is obesity itself a disorder, a disease? (Also, consideration must be given to direct mechanical interference with respiration and circulation by crowding of intra-abdominal contents against diaphragm.) Or is the shortening of the life span merely a consequence of obesity as implied in the popular saying that "fat people dig their graves with their teeth?" Admitted that there may be hereditary factors in some species, such as the whale and the hog, favoring deposition, or excess deposition of body fat, it is nevertheless true that all animals, otherwise normal, can be fattened by eating in excess of the expenditure of energy. In man obesity is the result of the same fundamental cause, that is, eating in excess of the energy expended. The work of Newburgh goes to show that intake of food below the energy requirements reduces body weight in all types of obese people.

During the last twenty-five years considerable, though superficial, attention has been given to the possible relation of thyroid, gonad, and pituitary disorders to human and animal obesity. It now seems established, or at least highly probable,

that there is no such thing as a primary endocrine obesity, apart from endocrine influences on appetite, hunger, and physical activity. The marked obesity that may follow destructive injuries at the base of the hypothalamus seems to be largely due to increased food consumption because of increased appetite. Endocrine therapy of obesity has been disappointing, except in the case of thyroid extract, and the latter has to be administered to the point of inducing the initial symptoms of the disease, thyroid toxicosis, before weight reduction is secured, in absence of reduced food intake. On the basis of present information it seems likely that the development of obesity in some people past 40 years of age is primarily due to these people persisting in excessive eating and the satisfying of specific appetite for foods, when the rate of tissue oxidation and general physical activity are both on the decline.

It is not difficult to understand why ingestion of food to the point of obesity is injurious to people with reduced factors of safety in the matters of insulin, pancreas, sugar, and fat metabolism. Such dietary excesses damage by overwork an already impaired mechanism. But in the absence of diabetes, actual or incipient, why does obesity, maintained for years, initiate or aggravate cardiovascular, renal, and other disorders that shorten the life span? While the answer to these questions is being sought by experiments and accurate observation on mice and men, the prevention of obesity in all people past 30 appears to be a prophylactic imperative, a must, in gerontology. And as a general proposition, it is safer for the patient to do this by reducing the diet than by taking toxic drugs.

An extensive bibliography, including the most significant original investigations, is to be found in the author's book "The Problems of Ageing," 2nd revised edition, Williams & Wilkins Co., Baltimore, Md., 1942.

The Formation of Concretions

L. LICHTWITZ †

Concretions found in man and animals are inanimate structures, the majority spherical or egg-shaped, with a consistency varying from stony-hard to soft-elastic. They also vary in many other qualities and particularly in constituent material, but they have in common two dominant characteristics: an albuminous substance and a concentric structure.

Biliary and urinary calculi are best known to physician and laymen. They are considered to be pathologic in character; but harboring such concretions is not at all identical with disease. They are often harmless, more so in the biliary system than in the urinary tract, and if they do not cause any signs and symptoms they become known as incidental findings on x-ray examination and on necropsy. They are, however, potential trouble makers and certainly do not serve any useful purpose.

This latter statement, however, cannot be generalized. There are concrements, physiologic in character, universally present in the animal kingdom and of great functional importance. I here refer to the otoliths which in most instances are produced within the internal ear, while some species of fish (torpedo *Akanthias*) pick up sand or gravel to serve as equilibrium-maintaining apparatus. Otoliths, as far as chemical composition and architecture are concerned, do not differ from urinary calcium carbonate calculi. It can be assumed with a great degree of safety that both these concretions are produced by the same process. Following the formation

† The Editor regrets to state that Dr. Leopold Lichtwitz died on March 18, 1943. Following an eminent career in Germany he had served Columbia University for eight years as Professor of Clinical Medicine. His passing is a loss to this country and to science. J. A.

are essential for the production of calculi: First, an organic compound of mucus character, in which insoluble or poorly soluble material can be deposited, and second, a fluid super-saturated with petrifying or incrusting substances. Meckel continues by saying that the formation of stones depends exclusively upon the incrustation of an organic substrate by precipitable material.

Considering that at this time, almost eighty years ago, there was no colloid chemistry, Meckel's conception represents an amazing case of prescience. Of course, his ideas were not heeded and were almost forgotten, and it took a long time before we reached his point of view. Since we are equipped with more information in the basic sciences, we are able to clarify the mystery of stone formation.

Three physico-chemical principles participate in this process: the laws ruling the behavior of colloids, the laws of surface activity and the laws of solubility and supersaturation.

The body fluids (particularly bile and urine) represent complicated systems, in that substances, electrolytes as well as non-electrolytes, are kept dissolved in higher concentrations than corresponds with their solubility in water. There are a number of factors by which a well stabilized supersaturation is maintained, the most important one being the presence of colloid material. These colloids give protection against precipitation of substances in supersaturated solution, as long as their degree of dispersity is sufficient.

The stability of the colloids, here as elsewhere, is limited. This is easily observed in the urine where, shortly after evacuation, a mucus-like coagulate appears, the nubecula (Figure 2). This phenomenon is specific for acid urine, while in alkaline urine a different, ether-soluble colloid is present, which coagulates on the surface. Here it forms a membrane glittering with interference colors, like fuel oil on the surface of water (Figure 3).

When the urine is deprived of the protecting principle by these or other means of colloid coagulation, sediments appear, *e.g.*, urates, calcium oxalate, phosphate and carbonate of calcium and magnesium. In the bile where the same principle is active, bilirubin calcium and cholesterol are the most frequent sediments.

Sediments, crystalline as well as amorphous, contain colloid material which represents an impurity but has a definite influence on the crystal form.[4]

The nature of sediments and the mechanism of their formation may be left out of consideration, because their relationship to the genesis of concretions is very limited. It is extremely rare that crystals agglomerate to form a small body which, since it has neither concentric nor radial structure, resembles rubble or a heap of paving stones. Crystals may grow within the urinary tract to the size of sand. This "gravel," commonly consisting of uric acid, may bring on slight discomfort, such as burning on urination.

In most cases, sediments are evacuated readily and rapidly. When they are not, they may form nuclei around which concretions are formed.

The presence of a nucleus or of nuclei is essential for stone production. The nature of the nucleus is of no significance. Besides sediments, coagulates of blood, fibrin, cellular detritus, casts, agglutinated bacteria, foreign bodies of any description may serve as nuclei. Obviously, this variety of particles has nothing in common chemically. It can be safely assumed that their ability to act as nuclei is a matter of surface activity. Coagulable colloids become absorbed on the surface, forming layers, which in the majority of instances become incrusted by such substances which are present in super-saturated solution.

Activity, and inactivity as well, play an important rôle in biology. The stability of the fibrinogen in blood plasma depends upon the integrity of the vascular endothelium and its surface inactivity. Injury of these linings is followed by agglutination of thrombocytes and thrombus formation. The same principle seems to prevail in the biliary and urinary tracts. Very little attention has been paid to the physical

are essential for the production of calculi: First, an organic compound of mucus character, in which insoluble or poorly soluble material can be deposited, and second, a fluid super-saturated with petrifying or incrusting substances. Meckel continues by saying that the formation of stones depends exclusively upon the incrustation of an organic substrate by precipitable material.

Considering that at this time, almost eighty years ago, there was no colloid chemistry, Meckel's conception represents an amazing case of prescience. Of course, his ideas were not heeded and were almost forgotten, and it took a long time before we reached his point of view. Since we are equipped with more information in the basic sciences, we are able to clarify the mystery of stone formation.

Three physico-chemical principles participate in this process: the laws ruling the behavior of colloids, the laws of surface activity and the laws of solubility and supersaturation.

The body fluids (particularly bile and urine) represent complicated systems, in that substances, electrolytes as well as non-electrolytes, are kept dissolved in higher concentrations than corresponds with their solubility in water. There are a number of factors by which a well stabilized supersaturation is maintained, the most important one being the presence of colloid material. These colloids give protection against precipitation of substances in supersaturated solution, as long as their degree of dispersity is sufficient.

The stability of the colloids, here as elsewhere, is limited. This is easily observed in the urine where, shortly after evacuation, a mucus-like coagulate appears, the nubecula (Figure 2). This phenomenon is specific for acid urine, while in alkaline urine a different, ether-soluble colloid is present, which coagulates on the surface. Here it forms a membrane glittering with interference colors, like fuel oil on the surface of water (Figure 3).

When the urine is deprived of the protecting principle by these or other means of colloid coagulation, sediments appear, *e.g.*, urates, calcium oxalate, phosphate and carbonate of calcium and magnesium. In the bile where the same principle is active, bilirubin calcium and cholesterol are the most frequent sediments.

Sediments, crystalline as well as amorphous, contain colloid material which represents an impurity but has a definite influence on the crystal form.[4]

The nature of sediments and the mechanism of their formation may be left out of consideration, because their relationship to the genesis of concretions is very limited. It is extremely rare that crystals agglomerate to form a small body which, since it has neither concentric nor radial structure, resembles rubble or a heap of paving stones. Crystals may grow within the urinary tract to the size of sand. This "gravel," commonly consisting of uric acid, may bring on slight discomfort, such as burning on urination.

In most cases, sediments are evacuated readily and rapidly. When they are not, they may form nuclei around which concretions are formed.

The presence of a nucleus or of nuclei is essential for stone production. The nature of the nucleus is of no significance. Besides sediments, coagulates of blood, fibrin, cellular detritus, casts, agglutinated bacteria, foreign bodies of any description may serve as nuclei. Obviously, this variety of particles has nothing in common chemically. It can be safely assumed that their ability to act as nuclei is a matter of surface activity. Coagulable colloids become absorbed on the surface, forming layers, which in the majority of instances become incrusted by such substances which are present in super-saturated solution.

Activity, and inactivity as well, play an important rôle in biology. The stability of the fibrinogen in blood plasma depends upon the integrity of the vascular endothelium and its surface inactivity. Injury of these linings is followed by agglutination of thrombocytes and thrombus formation. The same principle seems to prevail in the biliary and urinary tracts. Very little attention has been paid to the physical

relationship between body fluids and the epithelial linings of channels and cavities. As far as urine is concerned it seems that there is no adhesion between urine and the epithelial layer of the urinary tract. Adhesion within the renal tubules would greatly interfere with the flow through these channels. It well may be that there is a physical contact, similar to the relationship between mercury and glass.

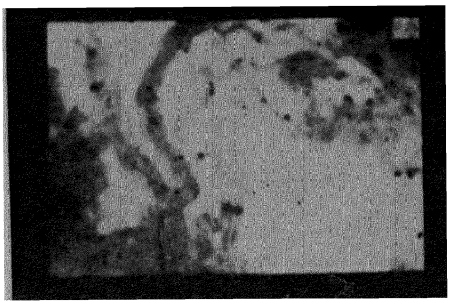


FIGURE 2. Nubecula with deposits of salts. (C. Posner)

This indifference between fluid and linings may change by alteration of the surface, as well as by variation of the fluid. Both these processes are most common. The appearance of an increased quantity of colloid material in the urine, as in the case of albuminuria and colloiduria, enhances the chance of precipitation, while even a slight alteration of the epithelial layer leads to actual colloid coagulation.

This is best evidenced by the formation of casts, which are due both to the presence of coagulable material in the kidney and to the loss of smoothness or surface indifference on the part of the renal epithelium.

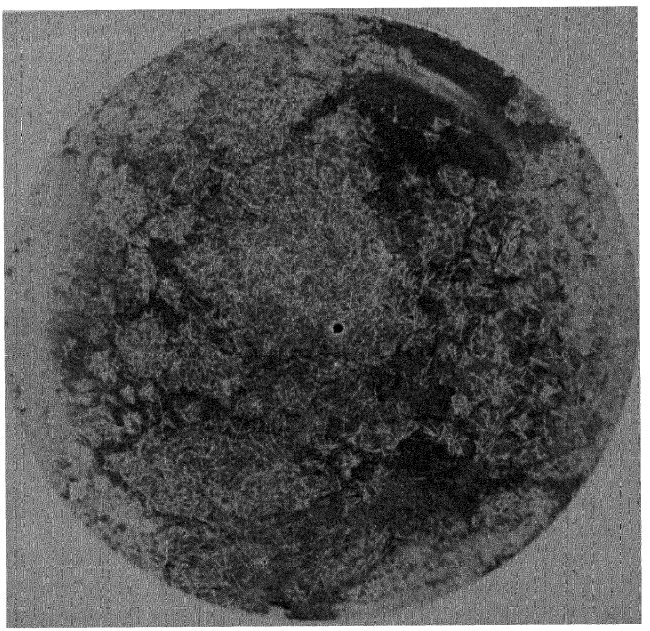


FIGURE 3. Calcified colloid precipitate on the surface of an alkaline urine.

Two important discoveries point to the effectiveness of this very mechanism in the process of stone formation.

The one is the appearance of concretions in the pelvis of the kidney, urinary bladder, and biliary tract, as sequela of vitamin A deficiency. Lack of A, the epithelium-protecting vitamin, leads to a deterioration of epithelial cells and eventually to keratinization. In this condition nothing is known of changes in the composition of the fluids, while the pathology of the wall undoubtedly involves a profound alteration of the surface.

The second fact is the appearance of kidney stones in hyperparathyroidism, that is, in hypersecretion of parathyroid hormone or in super-response to physiological production of this compound.

Parathyroid hormone represents the tool by which the metabolism of the skeleton is controlled and its dissimilatory phase is brought about. Hyperparathyroidism may lead to decalcification of bones, hypercalcemia and increased output of calcium. This has led, rather invitingly, to the widely accepted doctrine that kidney stones, which in this condition consist of phosphate and carbonate of alkaline earths, are due to augmented urinary calcium. We will see later that the concentration of compounds which petrify concretions, have almost no bearing on their formation. In addition, hyperparathyroidism brings about stone formation without any involvement of bones and without any calcium mobilization. It is due to the fact that this hormone exerts a very powerful influence upon the kidneys. It is a strong diuretic and when it acts for a considerable time produces chronic nephritis. A group of individuals respond more readily with the kidneys than with the skeleton, and develop kidney stones without the bones being affected, or before they are affected.

The mechanism of its action is elucidated by the influence of the parathyroid hormone on protoplasmic structures. This hormone is a potential poison for cells and fibers; it produces necroses in various tissues, such as the heart muscle, the medial coats of the smaller arteries, the gastrointestinal tract, the walls of the bronchi, the alveoli of the lungs and subcutaneous tissue. These lesions readily become calcified, according to the rule that calcification generally follows coagulation and the loss of colloid protection. The kidneys participate in hyperparathyroidism not only by appearance of stones in the kidney pelvis, but also by lime deposits in the lumina of the proximal tubules and in the kidney substance proper. There is scarcely any doubt that here also degeneration precedes calcification, and that alteration of the epithelial surface is part of the process.

Thus, the mechanism of parathyroid hormone in stone formation is of the same nature as the effect of A avitaminosis, with the difference that in hyperparathyroidism the biliary tract is spared, the hormone having no influence upon its structures.

The alteration of the epithelial surfaces from loss of smoothness and surface inactivity to outright necrosis, with, as well as without calcification, brings about colloid-precipitation and makes sediment stick to the lining, thus preventing its evacuation. In this way a potential stone nucleus is produced. Quite often necrotic plaques, particularly when calcified, act as nuclei. In both cases the growing concretion will be primarily adherent, and may become disengaged at a time when it is too big to pass the ureter without difficulty, if at all.

Processes of this kind occur also in the tissue below the epithelium of the renal pyramids. In this area deposition of uric acid takes place during early infancy (the uric acid infarct) and also later in life in gout. The "calcium infarct" of the pyramids, first described by Jacob Henle (1809-1885) is a quite common phenomenon in all periods of life, its incidence increasing with ageing. These calcium deposits converge on the top of the papilla, where they protrude, rendering the surface rough and grinder-like (O. Lubarsch). In recent years it has been shown by A. Randall that these calcium infarcts exert pressure on the covering epithelium, which eventually disappears, leaving the calcified plaque as carrier of an active surface and as

potential nucleus in free contact with the urine. In a series of excellent investigations Randall has demonstrated the development from the subepithelial calcium infarct to a small calculus attached to the pyramidal wall.

This formation of nuclei within the wall and their migration into the cavity, in all probability plays a great rôle in the development of urinary calculi. It also happens in the biliary system though here it is of minor practical consequence. In the case of the so-called strawberry-gallbladder (that is, infiltration of the wall by cholesterolesters) we have found minute cholesterol concretions attached to the epithelium (Figure 4). Commonly gallbladder stones form around nuclei, which origi-

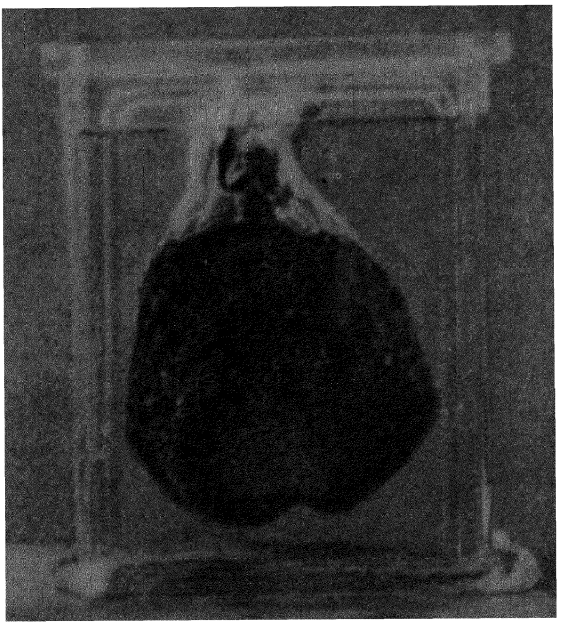


FIGURE 4. Strawberry gallbladder. Small cholesterol concretions are attached to the epithelial lining.

nate from the bile proper. Decomposition of the colloid system of the bile is brought on by a number of factors, such as albuminocholia, pH, diminution of bile acid concentration by destruction or absorption, possibly also by heavy metals, and leads to the precipitation of bilirubin calcium and albumen. By this process the bile becomes the consistency of a coarse gruel and forms a black-greenish mass of small particles (Figure 5). These are the common nuclei; they are found in the vast majority of gallstones when their transformation is not too far advanced. In a small number of cases the nuclei originate in the intrahepatic bile ducts and migrate into the gallbladder. These intrahepatic stones are amorphous conglomerates of bile pigment and calcium, completely or almost free of cholesterol with a remarkable great content of heavy metals, particularly of copper. In another small minority, foreign bodies, such as parasites and eggs of parasites, serve as nuclei.

To grow to a concretion the nucleus must remain in the cavity. A floating nucleus has a very poor chance in the urinary tract when the flow of urine is not obstructed. It is different in the gallbladder, which represents a receptacle, which in a number of cases, increasing with ageing, is not capable of complete emptying.

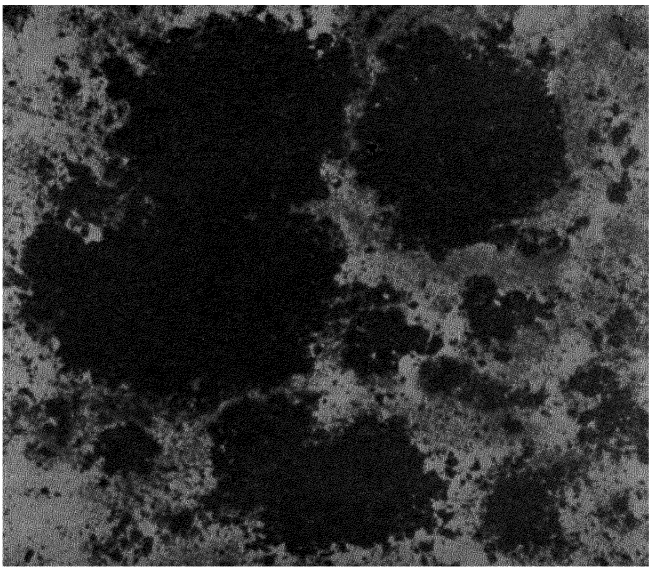


FIGURE 5. Coagulated bile, gruel-like in consistency. The black particles consisting of bile-pigment, calcium and protein, represent potential nuclei.

In these instances there is a dead space, in which residual bile may produce nuclei.

It is a remarkable fact that as a rule in more than 90 percent of the cases, formation of gallstones takes place but once. This is borne out by the fact that the common gallstones (be there three, four or any number up to a few thousand in the gallbladder) represent one generation. The stones are equal in every detail with the exception of the size, which depends upon the size of the nucleus (Figure 6) or is

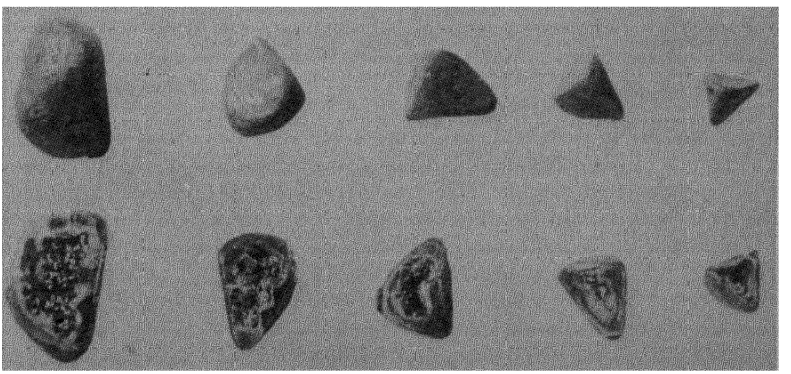


FIGURE 6. Family of common gallstones. The size depends upon the size of the nuclei.

caused by conglomeration of a number of original stones. There are a few exceptions, in which two and even three generations of gallstones are present. A second generation is bound to appear, when a gallstone is broken. The fragments act as nuclei; they possess the same specific gravity as the whole stones and herd with them in this particular part of the bladder cavity which we designated as the non-evacuable

or dead space. In this space the concretions are fairly safe. Compared with the tremendous number of stone-carrying gallbladders (in European countries stones are found on necropsy in about twelve percent of the cases; this, if valid generally, would account for about 18,000,000 cases in the United States), stones leave the bladder to enter the cystic duct or to pass the common duct in but a very small number of instances.

The recognition of the fact that practically all gallstones found in a gallbladder are produced simultaneously and that this happens only once, leads to the conclusion, that the presence of gallstones prevents the formation of new gallstones.

Resting securely in the gallbladder, the young concretions seem to have plenty of time to develop and grow. However, the time required for final development is surprisingly short; x-ray and other types of examination indicate that two weeks may suffice. It may be a matter of days and even of hours, until the final size and shape are determined. As a matter of fact, a biliary calculus in the stage of formation cannot be found, either at operation or on necropsy. Precipitation of colloids is probably as rapidly accomplished as the clotting of blood. As is known from studies on calcification, and particularly on mercury poisoning, calcium incrustation is discernible after a few hours.

Very soon after its inception the concretion stops growing. Its smooth and polished surface is inactive and apposition of new layers does not take place any longer. Stones evacuated after an interval of years are absolutely unchanged in size.

The kidney pelvis is no receptacle and stone formation takes place regardless of the presence of stones. Here also the growth of stones, is often limited, particularly in the absence of infection. In a number of instances the stone grows to and beyond the capacity of the cavity, filling the calices and destroying the renal parenchyma (Figures 7 and 8). The time of final development of smaller stones is just as short as with the common variety of gallstones.

Two factors have been emphasized as of importance and even as essential for formation of concretions, stasis and infection. The non-evacuable space of the gallbladder can be considered as equivalent to stasis. Generally, however, stasis is of significance only in that it facilitates infection. Infection leads to damage of the epithelial linings, to the appearance of a larger quantity of colloid material in dissolved and coagulated form, to agglutination of bacteria; in the urine to fermentation of urea leading to ammonia production and alkaline reaction, in the bile to destruction of bile acids and this facilitates stone formation. In the majority of cases, however, in both bile as well as urine, concretions originate in the absence of any infection.

When a nucleus is present, colloid layers are rapidly formed and thus the concentric structure of the concretion is established. When, as is usually the case, this soft primary concretion permits diffusion, electrolytes migrate and are precipitated when their concentration in the fluid is greater than aqueous saturation, and when the colloid layers lack in colloid protective capacity. There are almost no pure urinary stones. The petrifying material may vary in different layers. The urinary concentration of electrolytes above the point of saturation is of negligible influence. The electrolytes invariably crystallize but never in those forms which they show in the urinary sediment. They crystallize in radial direction, thus producing the radial architecture of the concretion. By dissolving the crystals the radiating design disappears, while the colloid concentric layers remain.

It has been assumed (H. Schade*) that the concentric architecture is brought about by alternating deposition of colloid and crystalloid. There is no evidence and very little probability that concretions are made like a layer cake. It is certain that urinary sediments do not participate in the development and hardening of stones.

The concentric pattern which characterizes the cross section is only to a small

* See paper on Concretions by H. Schade in Vol. II of this series. J. A.

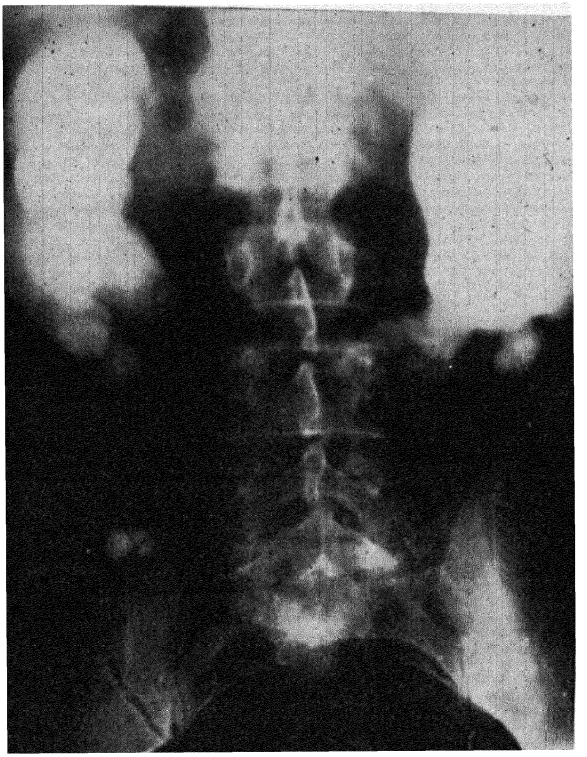


FIGURE 7. Bilateral giant renal calculi. Each stone is larger than a normal kidney.

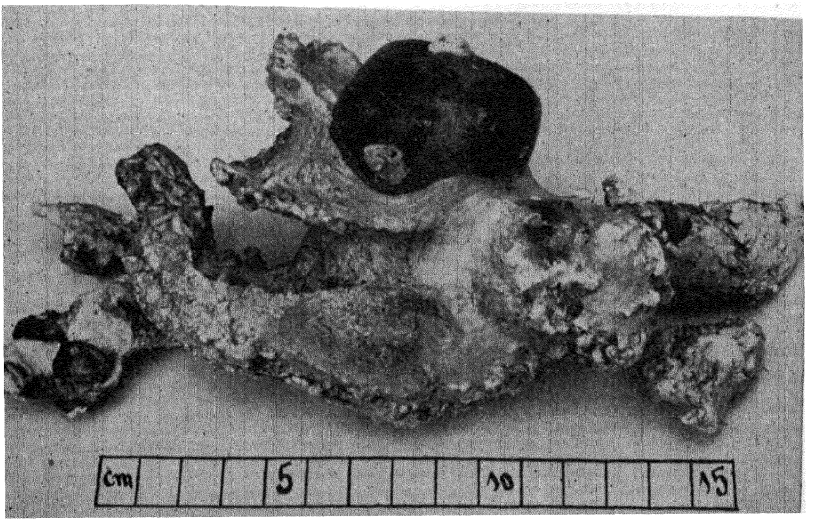


FIGURE 8. Giant renal calculus. The stone fills the pelvis of the kidney completely and intrudes into the calices.

part due to the apposition of colloid layers, the majority representing *Liesegang rings.*

This interesting phenomenon, which can easily be reproduced in a test tube (Figure 9) or on a gelatine film (Figure 10), is evidence that substances which form the

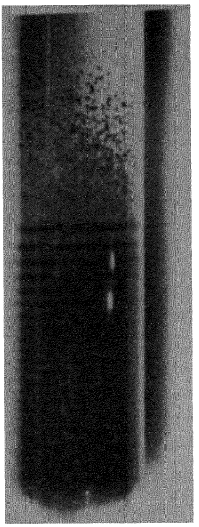


FIGURE 9. Liesegang rings produced in a test-tube.

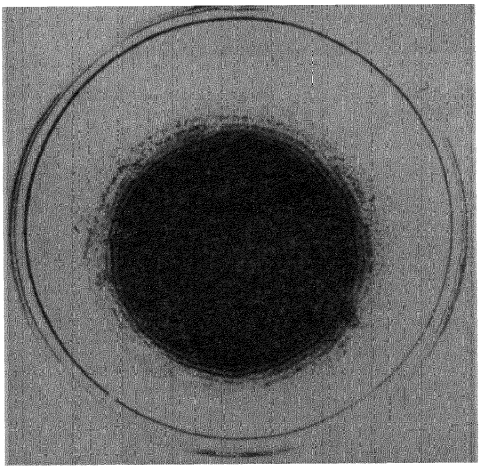


FIGURE 10. Liesegang rings produced on a gelatine film.

nucleus of the stone, are continuously dissolved, diffuse through the concretion, and are repeatedly precipitated in its colloid medium (Figure 11). This will be shown in gallstones where the development can be traced through all its stages.

The stone, even after having reached its final shape and size, undergoes a metamorphosis.

The vast majority of gallstones are not spheroidal. Multiple stones start in globular shape but are rapidly transformed to tetrahedral and hexahedral forms, that is, those forms which have the largest surface in relation to their mass. B. Naunyn and V. Goldschmidt have analyzed this process. Two young concretions, still soft but having their final size, become attached to one another. The tangential

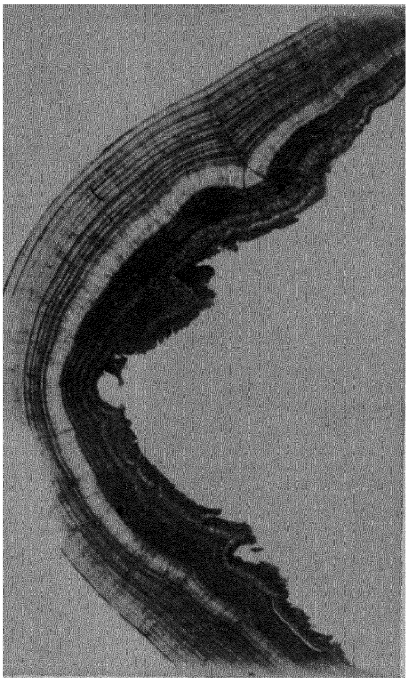


FIGURE 11. Kidney Stone. Calcium Oxalate grown around a blood coagulum which by desiccation and contraction has transformed the concretion to a hollow body. The picture shows a part of the shell with a few layers and many Liesegang rings. Residues of the blood coagulum are visible on the interior surface.



FIGURE 12. Family of four young biliary calculi. The nuclei are still present. The body shows fan-like cholesterol crystallization. The cortex consists of three layers with very little evidence of pigment migration.

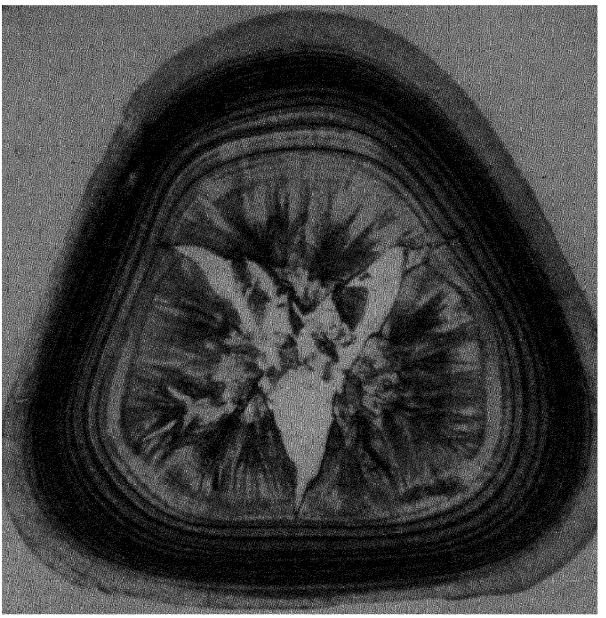


FIGURE 13. Common gallstone. The nucleus has completely emigrated. There is the typical triangular space in the center. The body shows fan-like cholesterol crystallization between the crystalline rays. Liesegang rings are in the cortical layers.



FIGURE 14. Common gallstone. The triangular space in the center is filled with amorphous cholesterol.

point of contact widens to a circle of adhesion, from the periphery of which the profile rises in angles of 60 degrees or 90 degrees. Simultaneously, by loss of water, contraction takes place, to which the surface responds with transition to polyedral forms. The nucleus, initially semi-fluid (Figure 12), disappears, rendering the center of the concrement a hollow space of a characteristic triangular shape on cross section (Figure 13). In this stage of development the cholesterol in the peripheral part of the nucleus has started crystallizing in a fan-like radial pattern with dark-brown pigment squeezed between the crystalline rays. The ground-substance of the concretion and the surface, despite its inactivity as far as absorption is concerned, still are gel-like and permit diffusion. Cholesterol fills the central hollow space first in amorphous form (Figure 14) but soon crystallizes (Figure 15). Simul-



FIGURE 15. Common gallstone. The cholesterol immigrated into the center has crystallized. (Pseudo-nucleus—B. Naunyn)

taneously bilirubin calcium, which first has formed the nucleus, later surrounded the empty space migrates and on its way out is precipitated in the colloid layers in series of Liesegang rings, thus contributing largely to the colorful pattern of the cross section. Migration of cholesterol inwardly and bile pigment outwardly depends upon the gel character of the ground substance. When by ageing or by circumstances of unknown nature the substance becomes rigid and impermeable, the metamorphosis ceases and the composition and structure of the concretion becomes fixed.

In a smaller number of cases only one nucleus is present and only one concretion

develops. A solitary stone maintains its original globular or ellipsoid shape. Thus, throughout its substance intraconcremental pressure, diffusion and migration are evenly distributed. There are no vectors and sectors, no corners and edges, no formation of a central cavity. The surface, uninfluenced by a change in shape, maintains a wax-like appearance.

In any other respect the solitaire undergoes the same metamorphoses as the common stone but often proceeding further, probably due to the quality of the surface. The pigment in many instances disappears completely; cholesterol may fill the stone to capacity and, on crystallizing, reduce the ground substance to such a degree, that no concentric pattern remains. When the metamorphosis of a solitaire has advanced thus far, it is not possible to recognize the course of the development. There are, however, many specimens in which rigid crystallization arrests the migration of pigment in various stages. Then nucleus substance and Liesegang rings remain permanently (Figures 16 and 17). This is particularly evident in Figure 18, show-

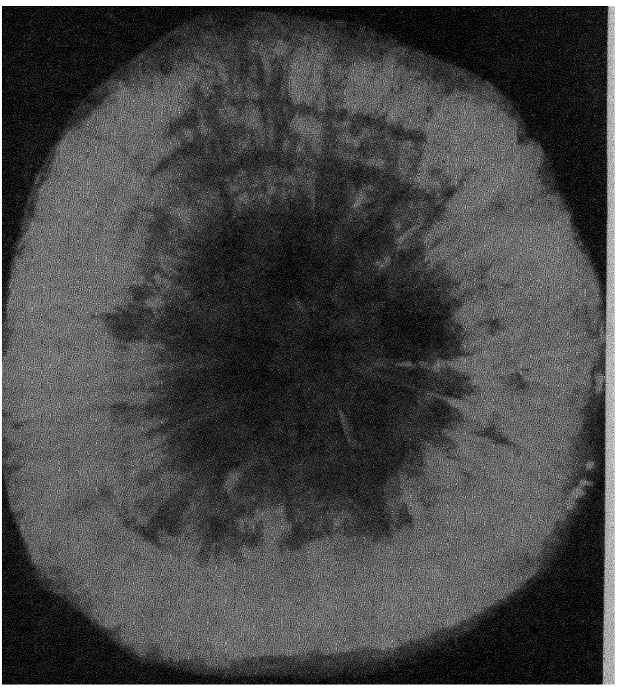


FIGURE 16. Cholesterol solitaire with large nucleus and Liesegang rings.

ing a stone with exhausted nucleus and two rings trapped in the periphery by crystallization. In the stone shown in Figure 19 the nucleus has disappeared and the radial cholesterol crystals have pierced the rings and transformed them into dotted circles.

The course of development is clearly visible in solitaires of ellipsoid shape. In these the crystallization of cholesterol, which has migrated into the apposition layers, starts (we do not know how), in the centre, thus developing a globular cholesterol solitaire within the primary concretion. The layers remain recognizable in the periphery and particularly at the poles, thus revealing the original concentric structure (Figure 20). In instances, rare but definitely demonstrating the inward mi-

gration of cholesterol, the primary stone of concentric structure is divided into two parts which crown the cholesterol solitaire in perfect symmetry (Figure 21). In other instances intraconcremental growth of a cholesterol solitaire leads to the stretching of the original stone, as may be seen from the course of the layers and Liesegang rings in Figure 22.

It has been claimed, particularly by L. Aschoff, that common stones and solitaires differ in etiology, in that pure cholesterol stones originate in a non-inflammatory

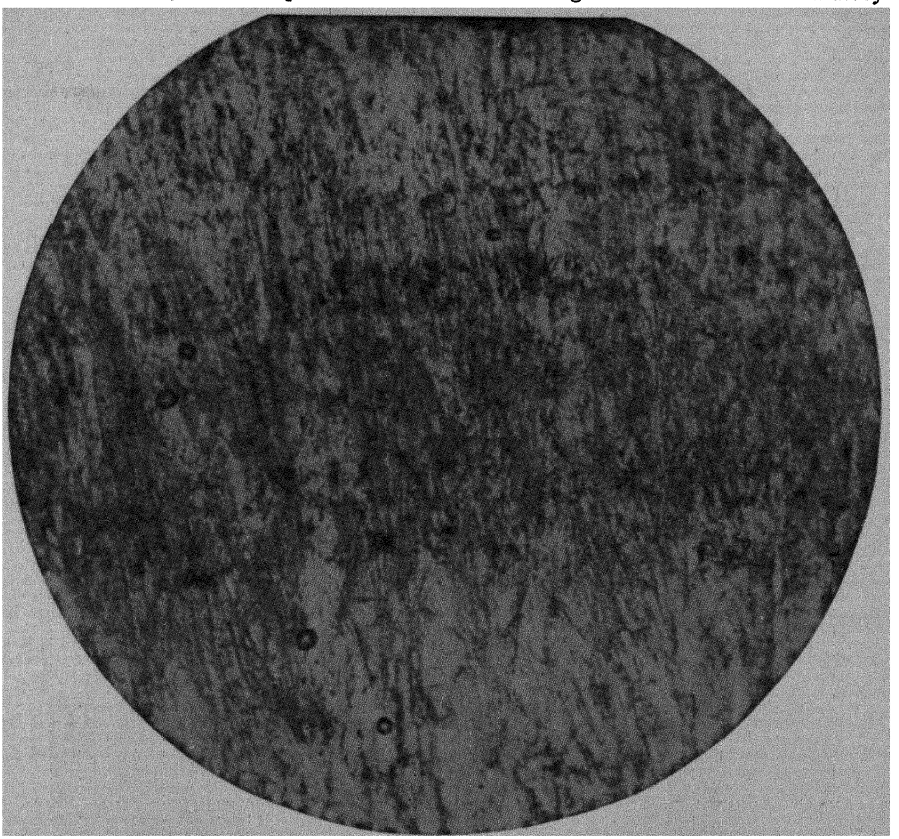


FIGURE 17. Cholesterol solitaire with Liesegang rings.



FIGURE 18. Cholesterol solitaire showing a small residue of nuclear pigment, complete radial crystallization and in the periphery two Liesgang rings.

manner from a faulty cholesterol metabolism, while the common stones are products of an inflammatory or infectious biliary process.

This statement, pathetically erroneous, has entered textbooks, where it will persist for a long time. Whatever "faulty cholesterol metabolism" may mean, it is certain that the cholesterol content of the bile does not depend upon cholesterolemia

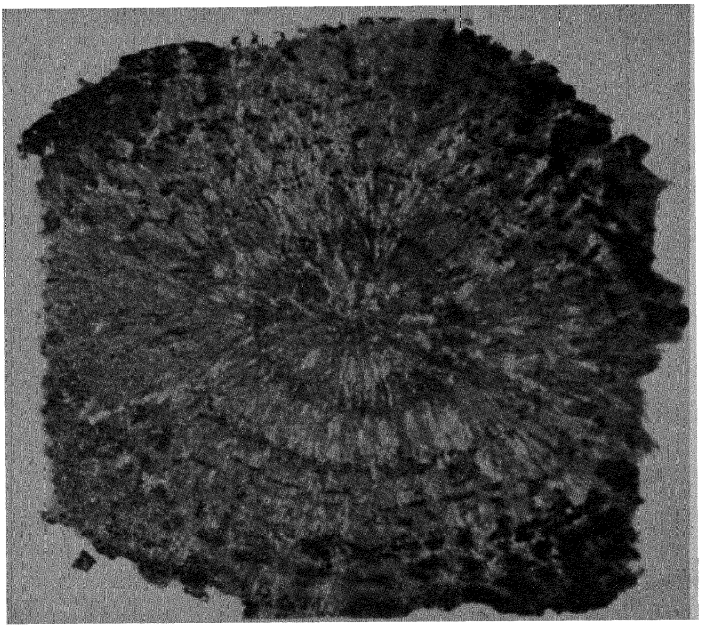


FIGURE 19. Cholesterol solitaire. The nucleus has disappeared. Liesegang rings are pierced by cholesterol crystals and thus transformed into dotted circles.



FIGURE 20. Ellipsoid solitaire of concentric pattern. A nucleus is not visible and certainly not present in the original form. A globe-shaped cholesterol concretion has developed in the center. (R. Brühl)

and cholesterol intake, and that the formation of gallstones is not favored by hypercholesterolemia. Liver bile is poor in cholesterol; in man it averages between 60 and 160 mgm percent. This variation in concentration is insignificant compared with the rise occurring in the gallbladder as a result of its concentrating capacity. This concentrated bile normally represents a well stabilized system. It does occasionally develop a deposit of cholesterol in the form of a few characteristic crystalline plates or of double-refractory droplets. Precipitates of this kind, as a rule, cannot be found in any sort of gallstones and do not act as nuclei. Cholesterol does not play a leading rôle in the formation of biliary concretions; it enters the stones secondarily, its mass increasing with advancing metamorphosis. Then, by its quantity and by its obtrusive physical and chemical qualities, it represents an important and impressive part of the concretion.

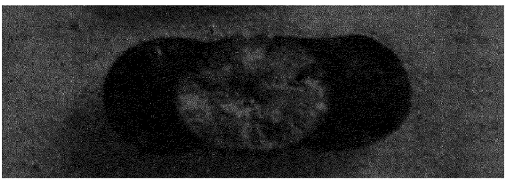


FIGURE 21. Stretched stone. A radial cholesterol concretion has developed within a solitaire of concentric pattern, consisting chiefly of ground substance and bile pigment calcium. By the intrusion of the cholesterol concretion the original stone is stretched, the two separated parts crowning the central stone in perfect symmetry.

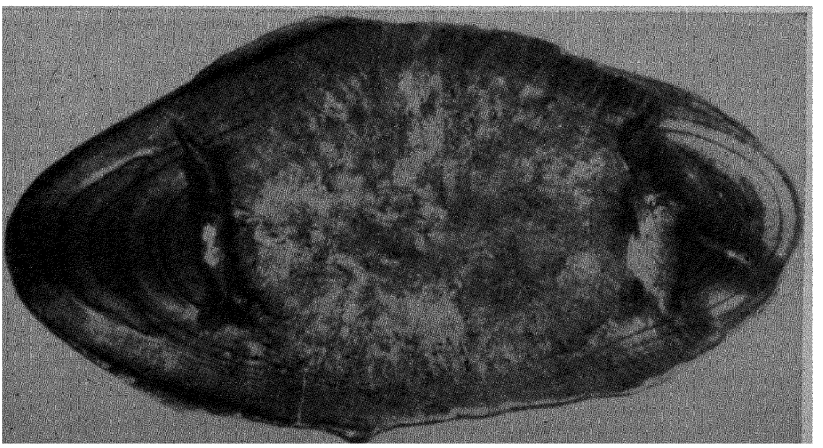


FIGURE 22. Stretched Stone. A radial cholesterol concretion grown within a gallstone of concentric pattern. (B. Naunyn)

It has been emphasized that a ball-shaped biliary concrement can develop only when there is but one nucleus. In this case, exclusively, the conditions for complete cholesterization and radial crystallization are established. The truth of this statement is borne out by a specimen, in which a radial cholesterol stone developed around the fragment of a common stone as nucleus (Figure 23). In this gallbladder a number of unsevered common stones were present. These, however, due to their solid and inactive surface, could not and did not interfere with the formation of the

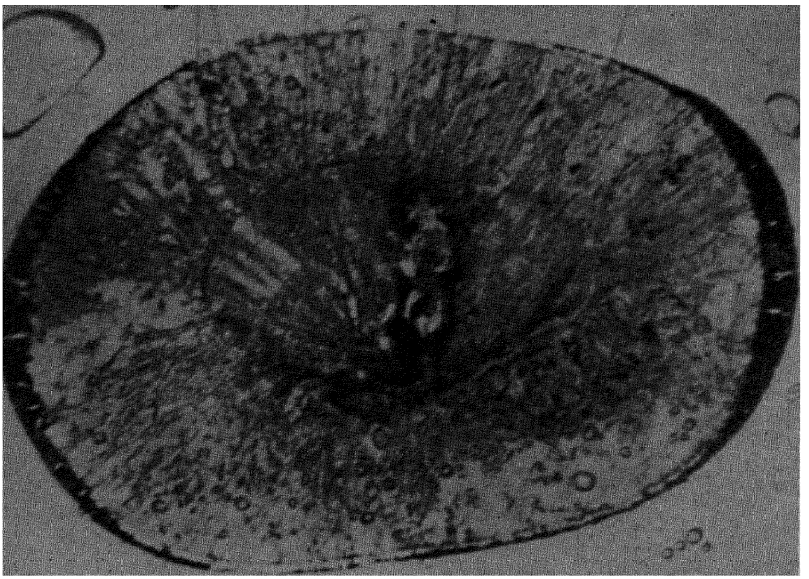


FIGURE 23. Combination stone. Around a fragment of a common gallstone as nucleus, a radial pure cholesterol concretion has formed. Within the fragment the disintegrating process, which has led to fragmentation, continues.

new specimen, which developed according to the processes governing the genesis of a solitaire.

This concretion belongs to a group of gallstones, which have been designated as *combination stones.*

It has been mentioned that the growth of the solitaire is limited. Growth being due to apposition of layers on the surface, it can be assumed that the termination of growth follows the loss of surface activity. The surface of solitaires is inactive to the constituents of normal bile. When, however, by a pathological process, such as infection, cholecystitis or albuminocholia, the bile is rich in precipitable colloids, new apposition of layers may take place around the old solitaire. Then a combination stone arises. In the beginning the new shell shows distinctly concentric layers (Figure 24), pigment and Liesegang rings. Later cholesterolization and crystallization

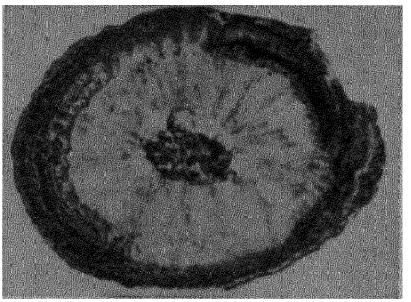


FIGURE 24. Combination stone. Around a cholesterol solitaire with well-preserved (trapped) nucleus, a shell has formed. The shell shows a concentric pattern, in parts eliminated by cholesterol crystallization.

proceed, starting from the crystal rays of the central stone and advancing throughout the shell, eventually to the final pattern characteristic of the mature cholesterol solitaire (Figure 25).

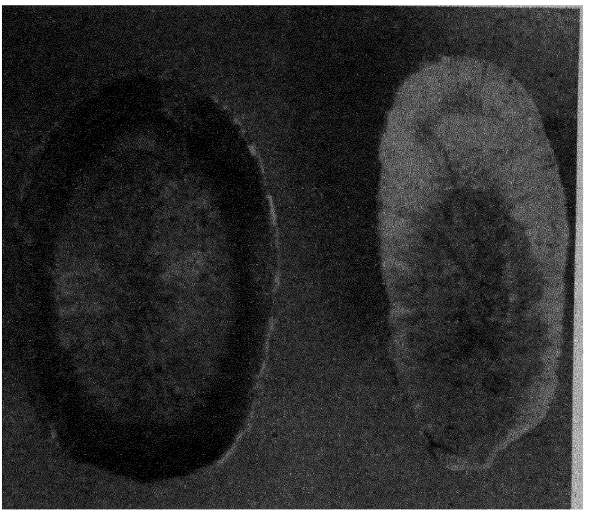


FIGURE 25. Combination stones. LEFT: The shell, incompletely cholesterized, still shows the concentric pattern. RIGHT: The cholesterolization of the shell is complete.

Another type of gallstone is represented by the barrel-shaped stones (Figure 26).



FIGURE 26. Three large stones filling the gallbladder to capacity. The largest stone is barrel-shaped. Between the stones are articulation-like surfaces.

They are the largest concretions, often filling the gallbladder to capacity. Little is known about the early phases of their development. It may be that in a cholecystitis the bile filling the bladder coagulates en masse and becomes a giant stone or a system of two or three stones, which are in contact by articulation-like surfaces. They undergo in principle the same metamorphosis as the other types of concretions. Their architecture is influenced by pressure from the outside (the gallbladder wall) and by the fact that the interconcremental surfaces are exposed to greater resistance and less open to cholesterol migration. It can often be seen that the surface of the stone

in contact with the gallbladder wall is covered with cholesterol, which seems to originate from epithelial cells and leucocytes.

In a small number of cases the gallbladder contains calcium carbonate stones or bilirubin calcium stones. The latter, very hard in consistency and cholesterol-free, develop more often in the extrahepatic bile ducts, particularly in the common duct.

A great problem is the dissolution of biliary and urinary calculi. There is disintegration leading to fragmentation and formation of cavities and holes. Nothing is known of the character of this process. There is more probability that it is peptisation than chemical dissolution. Generally concretions are more stable than patients. Chemical solution of petrifying material would leave the colloid shadow behind. When a stone is broken up the fragments remain. Shadow and fragments represent nuclei around which new stones will readily be formed.

References

1. Belonischkin, B., *Arch. f. Ohren-, Nasen- und Kehlkopfheilkunde*, **128**, 208 (1931).
2. Meckel von Hemsbach, H., "Mikrogeologie," Berlin, 1856.
3. Lichtwitz, L., Verhandl. des Dtsch. Kongresses fur Innere Med., **29**, 516 (1912).
4. Ord, W. M., "On the Influence of Colloids upon Crystalline Form and Cohension," London, 1879. See also paper by W. G. France in this volume.
5. Lichtwitz, L., "Phosphaturie, Oxalurie, Uraturie, Cystinurie," In Kraus-Brugsch, "Spec. Pathologie und Therapie," Berlin-Wien, 1919.
   "Klinische Chemie," 2nd ed., Berlin, 1930.
6. Lichtwitz, L., "Nephritis," New York, 1942.
7. Lubarsch, H., "Ueber die patholog. Ablagerungen, Speicherungen und Ausscheidungen in den Nieren," In "Handbuch der Spec. Pathol. Anatomie," Vol. **6**, 1, p. 525, Berlin, 1925.
8. Randall, A., "The Origin and Growth of Renal Calculi," *Ann. of Surgery*, **105**, 1009 (1937).
9. Lichtwitz, L., "Principien der Konkrementbildungen," In "Handbuch der normalen u. patholog. Physiologie," Vol. IV, p. 591.
10. Lichtwitz, L. and Stern, K. G., "Grundlagen der Konkrementbildung," In Lichtwitz-Liesegang-Spiro, "Medizinische Kolloidlehre," Dresden-Leipzig, 1935.
11. Schade, H., "Beiträge zur Konkrementbildung," *Münch. Med. Wochschr.*, **77**, (1909); 723 (1911).
12. Naunyn, B., "Klinik der Cholelithiasis," Leipzig, 1899.
    "Kolloidmorphologie der Gallensteine des Menschen," *Arch. exp. Path. Pharm.*, **99**, 38 (1923).
13. Goldschmidt, V., "Ueber tetraedrische und würfelformige Gallensteine," *Arch. exp. Path. Pharm.*, **99**, 23 (1923).
14. Aschoff, L., "Ueber die Entshehung der Gallenblasensteine," *Klin. Wochschr.*, **1345** (1922).
15. Aschoff, L., and Bacmeister, A., "Die Cholelithiasis," Jena, 1909.

Aero-emphysema and Caisson Disease, a Problem of Colloid Chemistry *

JEAN PICCARD
Professor of Aeronautical Engineering at the University of Minnesota, Minneapolis

Caisson Disease and Supersaturation

It is known that under certain conditions an excess of nitrogen dissolved in the blood or in other parts of the body will coalesce to form bubbles. These nitrogen bubbles are harmful or even fatal to man. To prevent the phenomenon we must understand it. The problem itself is old and well-known to deep-sea divers, but it is equally important in aeronautics. Since aerial battles are fought today at ever-increasing altitudes, the problem is becoming more and more acute.†

* We wish to express our appreciation to the publishers of the *Proceedings of the Staff Meetings of the Mayo Clinic* for permission to use in this article tables and excerpts from "Aero-emphysema and the Birth of Gas Bubbles" by Jean Piccard, *Proceedings*, **16**, 700-704 (October 29, 1941).

† In the case of too rapid decompression, divers and caisson workers are afflicted by what is called *caisson disease* or *diver's palsy*. An analogous affliction suffered by the aviator after rapid ascents is called *aero-emphysema*. The word *bends* is used in both cases, especially for less severe cases of the affliction. The word *emphysema* by itself means

When a diver ascends rapidly from great depth without special precautions, nitrogen dissolved in his body is liberated. The bubbles formed by this "cavitation" or effervescence, if carried along in the blood stream, will enter the capillaries, clog them and thus obstruct normal circulation. The damage done if cavitation occurs in parts of the body other than the blood is less well understood. It is certain, however, that most cases of "bends" are due to cavitation outside the blood vessels. The result in any case is skin itching, pain, paralysis or death, depending on the magnitude of the effect and on the location of the obstructions. Since a certain *supersaturation* is essential to initiate cavitation, and since any excess of nitrogen is rapidly liberated by the action of the lungs, the disease does not appear unless the decompression ratio has reached a certain value. This minimum value is about two to one, *i.e.*, it is reached if the diver, starting from any given depth, ascends to any level where the absolute pressure (atmosphere plus sea) has dropped to one-half its original value. The question of supersaturation seems to be most important in the description and explanation of aero-emphysema. Theory shows indeed that any bubble forming in water or in any other liquid must first pass through a stage in which it is too small to be stable. The whole phenomenon of supersaturation hinges on this fact: A very small air bubble in water is under such a pressure (produced by surface tension) that its content will go into solution even though the solution is already supersaturated.

The same phenomenon holds true, *mutatis mutandis*, for the formation of a water droplet in air. A very small water droplet is under such a high capillary pressure that its molecules will be forced into the atmosphere and the droplet will collapse, although the surrounding air is fully saturated or even supersaturated with water. This explains why atmospheric air is quite frequently supersaturated. Supersaturated solutions of air in water and supersaturated dispersions of water in air have a certain analogy with unstable colloidal dispersions of solids or liquids in water. With ordinary colloidal solutions the dispersed phase, if precipitated, comes out in "amorphous" form. The same holds true of water-in-air as well as of air-in-water dispersions. In both cases an unstable dispersion may be maintained for a certain length of time and may then suddenly "break" for no apparent reason with the onset of precipitation. There seems to be something mysterious about this unstable equilibrium and its sudden breaking. Perhaps this is as far as the analogy with ordinary and better known colloidal solutions goes. Perhaps there is a deeper analogy which will some day become apparent.

We shall, in the following pages, discuss more specifically the breaking of the equilibrium existing in supersaturated solutions of gas in water. First, however, we shall draw some comparisons between the "birth of a fog droplet" in free air and the "birth of an air bubble" in aqueous solution.

There is one very great difference between the two phenomena. If air contains ultramicroscopic droplets of dilute sulfuric acid originating, *e.g.*, from the sulfur contained in coal, or minute crystals of sodium chloride or droplets of sodium chloride solution from sea water spray, these particles or droplets will act as *nuclei* for larger fog droplets. Nuclei are, of course, likewise under a great capillary pressure, but due to their osmotic pressure and correspondingly small vapor pressure, they do not disappear under conditions which would cause water droplets of the same diameter to vanish. This persistence is especially true for sulfuric acid nuclei because sulfuric acid, if concentrated or slightly diluted, has no appreciable vapor pressure. Nuclei in water, analogous to water droplets in air would be formed by minute bubbles of any insoluble gas. The *difference* between the two phenomena lies in the fact

"puffing or inflation of the tissue by air or gas." *Embolism* means occlusion (usually of a blood vessel) and *aero-embolism* means occlusion by an air bubble. See Paul Bert: "La Pression Barométrique" (1878). J. S. Haldane and J. G. Priestley: "Respiration" (1935) and H. G. Armstrong, U. S. A.: "Principles and Practice of Aviation Medicine" (1939).

that while there are many *liquids* of exceedingly low vapor pressure there are no insoluble gases. A glance at a table of solubility of gases shows indeed that there are no gases whose insolubility in water could be compared with the non-volatility of sulfuric acid. *Persistent gaseous nuclei are, therefore, entirely out of the question.*

An *analogy* between formation of bubbles and formation of droplets is that in both cases solid insoluble objects, like minute sand grains, may act as nuclei.

In supersaturated air a small amount of water may collect on a surface with relatively large radius of curvature, and this water plaque will then be stable in an atmosphere where an equal amount of water, if in the form of spherical droplets with smaller radius of curvature, would not be stable. Solids which are easily wetted (hydrophilic solids) will, therefore, produce better nuclei than hydrophobic solids. There is little question that particles of rubber or paraffin will give less efficient nuclei for fog formation than equally small particles of quartz sand, although both are insoluble in water and connot produce hygroscopic solutions. Modern meteorology has, however, largely discounted the importance of solids as nuclei; salts or acid nuclei are much more potent than ordinary (insoluble) dust particles. Ionized molecules act also as nuclei. They need, however, in order to grow a much greater supersaturation than salts or acids.

What then will be the effect of solids in an aqueous solution? Will they be able to produce cavitation where otherwise it would not occur? There are no "aerophilic" solids in the same sense that there are hydrophilic solids. It is well known, of course, that a grain of porous clay will start cavitation and ebullition, if thrown into superheated water. This, however, does not mean much, because porous clay contains air pockets which act as nuclei, instead of the silicate itself. Wet tissue as found in the human body will never act like dry porous clay. Even porous clay, if thoroughly wetted by being boiled out and allowed to remain in water below the boiling point, loses all activity. Objects better comparable to hydrophilic units are grains of platinum or palladium. Superheated water starts boiling when a minute particle of one of these metals is dropped into it. These metals, however, absorb gases and they do not act differently, in principle, from porous clay. Base metals, on the other hand, may start cavitation in superheated water, because they produce minute particles of hydrogen under a local pressure sufficient to form hydrogen bubbles, and these, once formed, will act as nuclei.

All these cases of solid nuclei in water can be demonstrated *in vitro,* but always under conditions which are unimaginable *in vivo.*

It is quite possible that a bubble produced in human blood will start on a blood corpuscle. It may be that a white corpuscle is more active for this purpose than a red one, or *vice versa.* However, the cavitation will not start as a flat surface on the solid, but as a more or less rounded bubble. We may, therefore, assume that as a whole, cavitation in human tissue is not produced mainly by solids. Only if supersaturation is such that cavitation impends, may solids present in the human body be of importance for nuclei formation. A piece of raw meat, for instance, has very little effect, if any, on superheating of water.

If solids play a restricted role in cavitation, we can assume *a fortiori* that colloidal particles in the physiological fluids play a very small role in bubble formation in the human body.

In What Body Fluids Does Cavitation Occur?

It was originally assumed that caisson disease was produced by bubble formation in the blood and that the damage done was due to clogging of the capillaries (aeroembolism). There are specific cases reported where air bubbles have been seen in human blood *in vivo.* In other cases bubbles have been seen in other fluids and tissues, in the spinal fluid or in the eyes of living animals subjected to rapid decompression. A few years of the medical study of modern aviation have done more to

cast light on the problems of aero-emphysema than a hundred years of deep-sea diving.

The complete (double) circulation of the blood in the human body takes about 20 seconds.* Every time the blood passes through the lungs, which have a surface of about 85 square meters, it is thoroughly aerated and certainly brought exceedingly near to equilibrium. If, therefore, a man breathes pure oxygen for only a few minutes he should be entirely protected from hemato-aero-emphysema.

Aero-emphysema in the blood. Human blood contains only about one twenty-sixth † of the total free nitrogen dissolved in the human body, and, therefore, nitrogen will flow back into the blood even if pure oxygen is breathed for a considerable length of time. Under conditions most unfavorable to health, under which blood traversing the circulation would become saturated with body nitrogen in the course of each cycle before it reached the lungs, the body would be completely denitrogenized in $26 \times \frac{20}{60} = 9$ minutes. Thus, after nine minutes no more bends could occur in the blood. If the blood were not saturated once during each cycle, denitrogenation would take longer, but the danger of hemato-aero-emphysema would be decreased, not increased. This calculation is subject to several corrections, which would, however, not alter the result. Aviators subjected to low pressure in the "high-altitude chamber," do not get bends within a few minutes; but since bends may strike them as long as 45 minutes or more after they are in the chamber, it becomes evident that this kind of aero-emphysema cannot correctly be considered as hemato-aero-emphysema. Just where this cavitation occurs we do not know. The pain produced by it, usually in the articulations, may be very severe, but this kind of bends is usually not considered as dangerous to life. The older and well substantiated cases of fatal emphysema suffered by deep-sea divers, on the other hand, must probably still be attributed to hemato-emphysema, at least in some cases.

The fact that emphysema occurs only if a human being or test animal has been subjected to a great reduction of pressure shows that the body can stand a considerable amount of supersaturation without ill effects. Only if this supersaturation reaches the surprisingly high value of two to one will dissolved nitrogen coalesce and form bubbles. In the following pages we shall discuss not the whole problem of aero-emphysema, but mainly the conditions under which this cavitation is most likely to take place.

The fundamental reason why supersaturation of gas solutions in water occurs easily and regularly is very simple. As we have already noted, gas bubbles in the course of formation must pass through a stage where they are exceedingly small. In this pre-natal stage, as physicians call it, the gas in a bubble is not only under the hydrostatic pressure of the liquid but also under the capillary pressure of the spherical surface itself. The surface tension of water is about 73 mg/cm (milligrams per linear centimeter). The pressure, p, inside a sphere according to a well known formula ‡ is:

* Older literature assumed a longer cycle, about one minute; but the observation of radioactive salts introduced into the human blood stream has indicated the lesser value.

† The remaining nitrogen is dissolved in the extra-vascular aqueous fluids and, mainly, in the fat present in the body. For the adult man the total is not far from one liter.

‡ If a sphere contains a fluid under a pressure of p kg per sq cm, the two halves of the sphere, cut in two along any great circle, will be pushed apart by a force which is equal to the product of the pressure per sq cm and the area, in sq cm, included within the great circle. This is $\pi r^2 p$. This force is balanced by the stresses in the fabric of the sphere. It is divided among as many centimeters as the circumference measures. If we call T the stress per linear centimeter of the fabric, the total force is equal to $2\pi rT$. Hence we have $2\pi rT = \pi r^2 p$, and $p = \frac{2T}{r}$ mg/cm^2.

$$p = \frac{2T}{r} \text{ mg/cm}^2, \text{ and } r = \frac{2T}{p} \frac{\text{mg/cm}}{\text{mg/cm}} = \frac{2T}{p} \text{ cm} \qquad (1)$$

where T is the stress in mg/linear centimeter of the fabric and r is the radius of curvature of the sphere in centimeters.

From this formula it follows that the smaller the radius of a sphere, the greater must be the pressure inside the sphere, so long as the stress, T, in the surface remains constant. This means that inside an infinitely small bubble the pressure is infinitely great, so that the air in the bubble must pass into solution and the bubble "collapses." *

Now we can calculate from the formula given above how large a bubble must be, for any given degree of supersaturation, in order not to collapse but to grow. A bubble which is exactly between the state of collapse and the state of growth is said to be of *critical size*. The internal pressure can, obviously, not be greater than the pressure under which the gas has originally been forced into solution. The internal pressure is equal to this original pressure, *if* the bubble is of critical size and *if* no appreciable ventilation (loss of dissolved air or denitrogenation) has taken place; and p, which is the difference between internal and external pressure once dynamic and osmotic equilibrium having been established,† is equal to the difference between the atmospheric pressure to which the subject was exposed before and after ascent. If, for instance, a diver has ascended to sea level from a depth where the total hydrostatic pressure (that is, the sum of the pressures produced by the weight of the water and the weight of the air above the sea level) was five atmospheres, the difference p will be 4 kg/cm.^{2} From this we calculate, according to equation (1), the critical radius r_1 for this specific condition.

$$r_1 = \frac{2 \times 73}{4 \times 10^6}\text{cm} = 0.365\mu \tag{2}$$

This shows that when the diver has ascended rapidly from a depth of about 40.4 meters to sea level, his blood will be supersaturated to such an extent that any nitrogen bubbles larger than 0.365μ radius will grow, while any bubble smaller than this critical size will collapse.

Later, we shall discuss the apparent paradox contained in the fact that every bubble of 0.365μ radius has once passed through a phase during which it was smaller than this. The importance of the notion of the *Critical Radius* will then be demonstrated more specifically.

In actual practice the blood of the diver is not saturated with pure nitrogen but with a mixture of nitrogen, oxygen, and carbon dioxide. It is assumed that since oxygen is readily eliminated by normal metabolism, it takes a small part in the bubble formation. On the other hand human blood always contains a considerable amount of carbon dioxide, part of which is loosely bound in the alkaline serum, and part merely in solution. Since the pH of the blood changes considerably (approximately from 7.3 to 7.4) ‡ in the course of one circulatory cycle, and since the carbon

* We must differentiate between dynamic equilibrium and osmotic equilibrium in a bubble. If the external hydrostatic pressure is greater than the internal hydrostatic pressure the bubble is not in the dynamic equilibrium and it will immediately grow smaller. If the internal pressure is greater, the bubble will grow larger. For osmotic pressures just the opposite is true. If the bubble is in dynamic equilibrium and if the sum of the external vapor pressures (vapor pressure of water plus vapor pressure of nitrogen plus vapor pressure of oxygen) is greater than the internal air pressure, the bubble will grow larger. If, however, the external vapor pressure is smaller than the internal pressure, air will go into solution and the bubble will collapse. Since in our deductions we always consider any bubble as being in dynamic equilibrium, we must consider collapse as due to greater internal air pressure.

† It is necessary that the internal partial air pressure, that is, the gas pressure minus the internal partial pressure of water, be equal to the vapor pressure of air in water. For the temperature covering our experiments we can neglect the vapor pressure of water.

‡ The absorption of the oxygen while the blood flows through the lungs lowers its pH and in this way raises the vapor pressure of carbon dioxide. This phenomenon of changing alkalinity actively assists the absorption of carbon dioxide from the tissues into the blood, and its expulsion from the blood when the blood passes through the lungs.

dioxide concentration itself changes also during this cycle, it is hard to say what the exact composition of the gas inside a bubble in the human blood will be at any given time.

The calculation giving us a critical radius of 0.365μ is, therefore, accurate only for experiments *in vitro,* where pure nitrogen is dissolved in pure water. Aviation medicine assumes, however, that the bubbles in the blood or in other physiological fluids are mostly nitrogen. The following discussion of the critical radius will not be affected materially by the assumption that the bubbles consist of nitrogen and that the solubility of nitrogen in blood is the same as the solubility of nitrogen in pure water. Since the solubility of nitrogen is not very different from that of air, actual numerical results of experiments made with air will correspond closely to the results which we would have obtained if we had used a gas of the exact composition of the gas mixture dissolved in the living human body.

Difference Between Deep-sea Diver's Emphysema and Aviator's Aero-emphysema

So far we have considered the emphysema of deep sea divers and the aero-emphysema of aviators to be identical. There is, however, one great difference, the understanding of which may help considerably in explaining emphysema as such, and in explaining the general phenomenon of gaseous supersaturation. We shall now consider more closely the importance of the notion of the *critical radius. It is incomparably more dangerous for the deep-sea diver to ascend rapidly from a depth of 40.4 meters to sea level than it is for the aviator to rise from sea level to an altitude of 11,706 meters.* Yet in both cases the decompression ratio to which the human body is subjected is the same, *i.e.,* five to one. In the first case, the pressure decreases from 5 atmospheres to 1 atmosphere; in the second case it decreases from 1 atmosphere to one-fifth of an atmosphere. As shown in Table 1, the volume of liberated gas is the same in both cases.*

In order to see whether or not the difference in effect is caused by special physiological reaction, the experiment was performed *in vitro,* pure water, saturated with air, being decompressed from five to one, and from one to one-fifth of an atmosphere respectively. The result of these experiments shows that we are confronted with a purely inorganic problem. In case 1, the cavitation was rapid and strong effervescence was observed; in case 2, the cavitation was much slower. In the first case the bulk of the dissolved air had escaped after five seconds, while in the second case, several minutes elapsed before the gas production even approximately ceased. The experiment was repeated frequently, and though the time differences were not always so great, they were in all cases quite striking. This is a rather surprising result, since, as stated above, the volumes of gas produced are exactly the same in both cases and what we observe *in vitro* is volume, not weight. What injures the diver also is volume, not weight.

Since the experiment *in vitro* gives the same result as the well known observation *in vivo,* we shall restrict ourselves to seeking an explanation of the former.

Table 1 shows that in case 1, five times more nitrogen (measured by weight) escapes than in case 2, but that, since the density of the escaped gas is five times greater in case 1 than in case 2, the volume is the same in each case, *i.e.,* 80 cc per liter. One thing, however, is different in the two cases—the critical radius.

Table 2 shows that, as was found from equation (2), the critical radius r_1 of case 1 is only 0.36μ. In case 2 the pressure p is 0.8 kg/cm^2 and the critical radius r_2 is

$$r_2 = \frac{2 \times 73}{0.8 \times 10^6}\text{cm} = 1.8\mu \qquad (3)$$

* To be quite accurate one should add (to the hydrostatic pressure of the water surrounding the diver) the normal pressure produced by the body tissues. Since this is constant it would explain part of the phenomenon by somewhat cutting down the ratio of decompression. As the experiments *in vitro* show, this cannot explain the whole phenomenon.

Table 1. Air Dissolved in, and Escaping from, One Liter of Water

Pressure (atm)		Air dissolved		Air escaping	
		Weight* (mg)	Volume (cc)	Weight *(mg)	Volume (cc)
Case 2	0.2	4	20	16	80
	↑				
	1.0	20	20		
Case 1	1.0	20	20	80	80
	↑				
	5.0	100	20		

* The air volumes given are correct for a temperature of 17° C. The same table is correct for nitrogen at a temperature of 7° C. The weights are given in round numbers.

Table 2. Critical Dimensions of Air Bubbles

Case 1. Water is saturated under a pressure of 5 atm abs. and the pressure is released to 1 atm abs.
Case 2. Water is saturated under a pressure of 1 atm abs. and the pressure is released to 0.2 atm abs.

	Critical radius	Critical volume	Number of molecules	Water volume from which air is taken
Case 1	0.36μ	$0.20\mu^3$	28×10^6	$12.8\mu^3$
Case 2	1.8μ	$26\ \mu^3$	690×10^6	$1620\ \mu^3$

Columns 3 and 4 of Table 2 will be discussed later. The mathematical meaning of the critical radius as derived in Table 2 is very simple. It means: (1) if the water is saturated at five atmospheres' absolute pressure and the pressure released to one atmosphere absolute, no gas bubble can be formed unless its radius is first equal to, or greater than 0.36μ, and (2) if water is saturated at one atmosphere absolute pressure and the pressure then released to one-fifth of an atmosphere, no bubble can be formed unless its radius is first equal to or greater than 1.82μ. This is obviously absurd, since as a matter of fact, we see bubbles coming up where there were no bubbles before, and no bubble can possibly be formed unless it is first a very small bubble.* It would be like the case of an animal the embryos of which are never capable of living unless they are first, say a quarter of an inch long.

Since the result of the calculation given above does not agree with the observed facts, it is certain that at least one of the assumptions made was incorrect. As far as we can see, the assumption which does not strictly correspond to the facts is that

* Vibrations, sound waves and ultrasonic waves, and especially nodes, because of the temporary reduction of pressure which they engender, can diminish locally and temporarily the number of molecules needed to produce cavitation. In this case two possibilities exist: either the embryonic bubble has time, during the short period of favorable conditions, to grow into a bubble of more than critical size, or it has not the time needed for this growth. The bubbles may then be large enough and numerous enough to produce visible opalescence. This opalescence will then disappear at the end of each low-pressure period. To the human eye such an intermittent cloudiness would seem to be steady, and the clear phase of the phenomenon could be seen only by means of a stroboscope. The effect of acoustic and ultrasonic waves has been discussed by several authors. (See: Reynolds, O., "Papers on Mechanical and Physical Subjects," vol. **2**, p. 578, New York, The Macmillan Company, 1901, Bondy, C., and Söllner, K. "Mechanism of emulsification by ultrasonic waves," *Trans. Faraday Soc.*, **31**, 835-843 (1935); Boyle, R. W., "Ultrasonics," *Science Progress*, **23**, 75-105 (1928); Kundt, A., and Lehmann, O., "Ueber longitudinale Schwingungen and Klangfiguren in cylindrischen Flüssigkettsäulen," *Ann. Phys. Chem.*, **153**, 1-12 (1874). The latter have especially observed the formation and disappearance of bubbles in non-saturated liquids. It is quite possible, as has been pointed out by Maurice B. Visscher (private communications), that under complete absence of vibrations cavitation as observed in supersaturated solutions would not occur. See also paper by A. Einstein in Vol. I of this series.

air bubbles grow continuously and not discontinuously, *i.e.*, that air is a homogeneous fluid, whereas we know that it is composed of individual molecules. We shall now show that this incorrect assumption must be held responsible for the disagreement between calculation and observed facts.

If a single molecule were large enough to produce a nucleus or seed, it is obvious that a bubble could be started by any one molecule of nitrogen or oxygen. But it takes millions of molecules to produce a bubble equal to or greater than the critical size. Let us, therefore, calculate exactly how many molecules are contained in a bubble of critical radius in case 1 and in case 2. From the radii given we calculate the volume as $V_1 = 0.205\mu^3$ and $V_2 = 25.7\mu^3$. It is known that under standard conditions (760 mm of mercury and 0°C) a volume of 22.4 liters of any gas contains 6.02×10^{23} molecules. In case 1, the inside pressure is 5 atmospheres absolute, and in case 2 it is 1 atmosphere. The respective numbers of molecules in case 1 and in case 2 are, therefore

$$n_1 = \frac{6.02 \times 10^{23} \times 5 \times 0.205}{22.4 \times 10^{15}} = 27.6 \times 10^6$$

and

$$n_2 = \frac{6.02 \times 10^{23} \times 1 \times 25.7}{22.4 \times 10^{15}} = 690 \times 10^6$$

This proves that no single molecule of nitrogen or oxygen could possibly be considered as a seed for the birth of a bubble. How then is it possible for any bubble to pass through its embryonic stage before reaching critical dimensions and starting to grow normally? Now this question can be answered only by the theory of probabilities. It may seem exceedingly improbable that at any given point of the supersaturated solution 27.6 or 690 million molecules could meet and start a bubble growing; but the field is very large. A liter of water does not contain millions but billions times billions of air molecules.* What is exceedingly unlikely to happen at any one point of the field will probably occur somewhere in the whole field, not only once but many times.† Whenever molecules of nitrogen or oxygen, or both, happen to meet at one point of the solution and whenever their number is smaller than 27.6 or 690 million respectively, nothing happens; or, to express the same idea in other words, a bubble is formed which dies *in statu nascendi*. (See column 3 of Table 2.)

This forms the great majority of cases; but it occasionally happens that at one place or another more than 27.6 or 690 million molecules do meet, or at least come close enough to break the film of water separating them. Just how thin this separating film must be in order to break, we do not know. Then, and only then, nuclei or seeds are produced of sufficient vitality to grow and after they have grown, to make the whole solution appear milky. Bubbles of the critical size are in the colloidal zone, which may be reached by the aggregation of sub-colloidal particles, as well as by the dispersion of super-colloidal particles. In the case of gas dispersions in water, the latter phenomenon cannot take place because a bubble, once it is small enough to go into solution, will always disappear entirely, since the smaller it is the greater will be the interior pressure.

* A simple calculation shows that a liter of water, saturated at 1 atmosphere and therefore holding about 20 cc of dissolved air (standard conditions) contains 5.4×10^{20} air molecules and, if saturated at a pressure of 5 atmospheres, contains 27×10^{20} air molecules.

† In physics as well as in medicine or in any other science many instances are known where it is probable that a certain "accident," very unlikely to happen in any one case, will, in a larger field, occur with a probability which is so great that it is a certainty. Any particle of iron hydroxide, for instance, resting on the bottom of a container filled with water, will probably get as many molecular blows from one side as from the opposite side, and yet if we observe long enough and under proper conditions we are sure to see it making a sudden side motion as a result of having received during a certain fraction of a second many more blows from one side than from the other. This is the basis of Brownian motion.

The turbidity of gases and liquids in the vicinity of the critical state is due to the formation of colloidal particles of condensing liquid in the gaseous phase because of pressure and/or cooling. A. Einstein demonstrated this mathematically in his paper in Vol. 1. of this series.

Now we are able to give an answer to the question posed above: Why do we get more bubbles in case 1 than in case 2? The obvious answer is that it is much more probable that 27.6 million air molecules will meet at one place in the water solution than that 690 million will meet, especially if the volume from which the 690 million have to assemble is less densely populated. A simple calculation shows, indeed, that in case 1 the necessary air must diffuse from a volume of 12.8 cubic microns, whereas in the second case the required number of molecules must come from a volume which is 5^3 times larger, that is, 1606 cubic microns. The larger the volume from which the scattered molecules must congregate, the less the probability that a congregation of sufficient size will be formed.

Table 2 demonstrates the difference between case 1 and case 2 (see especially column 4). This is strikingly illustrated by Fig. 1, which we owe to the courtesy of Dr. W. Boothby.

AIR BUBBLE AND WATER VOLUME

	CASE 1 Pressure Reduced from 5 to 1 Atm.	CASE 2 Pressure Reduced from 1 to 1/5 Atm.
Radius of Air Bubble	0.365μ	1.82μ
Volume	$0.205\mu^3$	$25.7\mu^3$
No. Molecules Air	27.6×10^6	690×10^6
Minimum amount of water containing these air molecules		
Radius	1.43μ	7.15μ
Volume	$12.28\mu^3$	$1538\mu^3$

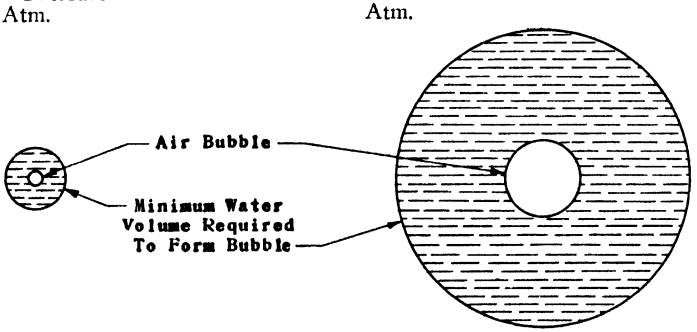


FIGURE 1. Relative size of the air bubbles and the water volume from which the molecules must come.

As far as we know today the theory of probabilities enables us to understand why it is more dangerous for a deep-sea diver to come up from a depth of 40.4 meters than it is for an aviator to go up to an altitude of 11,706 meters. In both cases the volume of nitrogen eventually escaping from tissues and blood is the same, but in the second case the cavitation is slower and for this reason a larger proportion of the nitrogen escapes through the lungs and the probability of harmful cavitation is very much less.

Methods for the Prevention of Emphysema

We may add here a résumé of the methods used to prevent bends in deep-sea diving and in aviation. Since bends are produced by nitrogen bubbles, it follows directly that if we exclude nitrogen we prevent bends altogether.

In deep-sea diving an obvious solution would be the administration of pure oxygen to the diver from the moment he begins his descent into the sea. This solution, though simple to the physicist, is prohibited by the physician. If a guinea-pig or a goat is kept for any length of time in an atmosphere of pure oxygen the epithelium of its lungs is affected by this five-fold excess of oxygen somewhat as it is affected by chlorine. It causes damage to the enzyme systems and the usual obvious result is severe lung injury which may lead to death. If such results follow when the partial pressure of oxygen is raised in the ratio of 21 to 100, how much more marked would they be with a ratio of 21 to 500, which would, for example, develop if the diver had to heach a depth of 50 meters!

For many years, therefore, deep-sea divers were given ordinary compressed air. Depths greater than 50 meters were considered somewhat dangerous and no useful work was ever done at greater depths than 90 meters.* (U. S. Navy, Honolulu, 1915)

Three elementary precautions against bends are prescribed for deep-sea diving: (1) The diver must under no conditions go deeper than 100 meters, preferably not deeper than 80 meters. (2) He must not stay too long at great depths—not more than ten minutes beyond 64 meters. (3) From any depth beyond ten meters he must come up slowly; and the longer he has remained down, the slower he must ascend.

J. S. Haldane was the first to work out a convenient table indicating at what levels intermediary stops must be taken during the ascent of the deep sea diver. The number of these stops and their length depend obviously upon the time the diver had remained at his greatest depth. With permission of the Yale University Press, we reproduce here part of this table as given in "Respiration," by J. S. Haldane and J. G. Priestley. (See p. 1092.)

From a purely biological point of view it would certainly be more efficient not to prescribe complete stops and rapid ascents between these stops, but to prescribe for every level reached during ascent, the permissible speed of ascent as a function of the maximum depth reached, and of the time spent there. From a master diver's point of view, however, it would be far more difficult to establish specific and continuously changing speeds than to arrange for a certain number of complete stops of prescribed lengths and at prescribed levels.

The safety of deep-sea divers has furthermore been greatly increased by the invention of the decompression chamber, also described by J. S. Haldane. If for any reason a diver had been raised too rapidly or if he shows any signs of bends, he is immediately introduced into this chamber. There the pressure is raised to about one-half of the greatest pressure to which the diver had been submitted, and after he has remained for some time at this pressure, it is gradually reduced, according to Haldane's tables, or more slowly if any new signs of bends appear. This is far more comfortable than the older method of re-submerging a diver if he suffers from bends, and it is also much safer.

The isolation of large amounts of helium during World War I has, however, permitted another solution of the problem—inhalation of a mixture of oxygen and helium. Helium is less soluble in water than nitrogen, and what is even more important, due to its lower molecular weight it diffuses out of the body more rapidly than nitrogen. Another important advantage of helium over nitrogen is that the former does not possess the anesthetic properties which the latter shows at high pressures. For these reasons the mixture of oxygen and helium has allowed modern divers to break the older depth records of about 100 meters without danger.

A similar solution had been suggested for aviation when a very fortunate physiological discovery was made: Man, in contradistinction to the guinea pig and the

* This does not, of course, apply to rigid diving suits or bathyspheres in which the diver is not subjected to the exterior water pressure.

Table 3. Stoppages During the Ascent of a Diver after Ordinary Limits of Time from Surface

Depth (feet)	Time from surface to beginning of ascent	Approximate time to first stop (min)	Stoppage in minutes at different depths [1] 60 ft	50 ft	40 ft	30 ft	20 ft	10 ft	Total time for ascent (mins)
0-36	No limit	1	—	—	—	—	—	—	0-1
36-42	Over 3 hours	1	—	—	—	—	—	5	6
42-48	Up to 1 hour	—	—	—	—	—	—	—	1½
	1-3 hours	1½	—	—	—	—	—	5	6½
	Over 3 hours	1½	—	—	—	—	—	10	11½
48-54	Up to ½ hour	—	—	—	—	—	—	—	2
	½-1½ hours	2	—	—	—	—	—	5	7
	1½-3 hours	2	—	—	—	—	—	10	12
	Over 3 hours	2	—	—	—	—	—	20	22
54-60	Up to 20 mins	—	—	—	—	—	—	—	2
	20-45 mins	2	—	—	—	—	—	5	7
	¾-1½ hours	2	—	—	—	—	—	10	12
	1½-3 hours	2	—	—	—	—	5	15	22
	Over 3 hours	2	—	—	—	—	10	20	32
60-66	Up to ¼ hour	2	—	—	—	—	—	—	2
	¼-½ hour	2	—	—	—	—	—	5	7
	½-1 hour	2	—	—	—	—	3	10	15
	1-2 hours	2	—	—	—	—	5	15	22
	2-3 hours	2	—	—	—	—	10	20	32
66-72	Up to ¼ hour	2	—	—	—	—	—	2	4
	¼-½ hour	2	—	—	—	—	3	5	10
	½-1 hour	2	—	—	—	—	5	12	19
	1-2 hours	2	—	—	—	—	10	20	32
72-78	Up to 20 mins	2	—	—	—	—	—	5	7
	20-45 mins	2	—	—	—	—	5	10	17
	¾-1½ hours	2	—	—	—	—	10	20	32
78-84	Up to 20 mins	2	—	—	—	—	—	5	7
	20-45 mins	2	—	—	—	—	5	15	22
	¾-1¼ hours	2	—	—	—	—	10	20	32
84-90	Up to 10 mins	2	—	—	—	—	—	3	5
	10-20 mins	2	—	—	—	—	3	5	10
	20-40 mins	2	—	—	—	—	5	15	22
	40-60 mins	2	—	—	—	3	10	15	30
90-96	Up to 10 mins	3	—	—	—	—	—	3	6
	10-20 mins	2	—	—	—	—	3	5	10
	20-35 mins	2	—	—	—	—	5	15	22
	35-55 mins	2	—	—	—	3	10	15	30
96-108	Up to 15 mins	3	—	—	—	—	3	5	11
	15-30 mins	3	—	—	—	3	7	10	23
	30-40 mins	3	—	—	—	5	10	15	33
108-120	Up to 15 mins	3	—	—	—	2	3	7	15
	15-25 mins	3	—	—	—	5	5	10	23
	25-35 mins	3	—	—	—	5	10	15	33
120-132	Up to 15 mins	3	—	—	—	2	5	7	17
	15-30 mins	3	—	—	—	5	10	15	33
132-144	Up to 12 mins	3	—	—	—	3	5	5	16
	12-25 mins	3	—	—	2	5	10	12	32
144-156	Up to 10 mins	3	—	—	—	3	5	5	16
	10-20 mins	3	—	—	2	5	10	12	32
156-168	Up to 10 mins	3	—	—	2	3	5	5	18
	10-16 mins	3	—	2	3	5	7	10	30
168-180	Up to 9 mins	3	—	—	2	3	5	5	18
	9-14 mins	3	—	2	3	5	7	10	30
180-192	Up to 13 mins	3	—	2	3	5	7	10	30
192-204	Up to 12 mins	3	2	2	3	5	7	10	32

During each stoppage the diver should continue to move his arms and legs. Haldane gives also a table, not reproduced here, to be used in case the diver had been subjected (accidentally) to a longer time of submersion.

goat, can tolerate pure oxygen at partial pressures up to one atmosphere for long periods of time. Instead of breathing ordinary air until the altitude is reached (about 10,000 to 12,000 feet) at which oxygen is needed, the aviator should breathe pure oxygen from the very start, or, better yet, for some time before the start of any flight which is likely to reach an altitude of 34,000 feet or more. If the pilot intends to reach the lower part of the stratosphere, say 36,000 feet, if the ascent is to be made with a fast-rising plane and if the pilot intends to stay in the stratosphere more than a very short time, it becomes imperative that pure oxygen be breathed a considerable time before the actual start of the flight.

Greater altitudes than 40,000 feet must not be sought without pressure suit or pressure cabins because, although bends could be avoided, the partial pressure of oxygen, even if pure, is insufficient to sustain life beyond this altitude.*

We have discussed the fact that it is much less harmful for an aviator to ascend rapidly to an altitude where the barometric pressure is one-fifth of normal than it is for the diver to ascend to an absolute pressure of one-fifth of the pressure to which he was subjected. This now well known observation was anticipated to a certain degree by results from experiments conducted by the British Admiralty Committee as described by J. S. Haldane. Haldane states that it is permissible for deep-sea divers to decompress rapidly from relatively low pressures in the ratio of 2.25 to 1. For somewhat higher pressures, the safe ratio of rapid decompression is 2 to 1 and for pressures above six atmospheres it is only 1.75 to 1. The fact that it is safe for the aviator to decompress rapidly in the ratio to 4 to 1 (ascending from sea level to 33,000 feet) is nothing more than an extrapolation of the phenomenon described by Haldane.

This suggests that, since it is safe for the aviator to ascend rapidly from sea level to 33,000 feet, it would be equally safe to ascend rapidly from 6000 feet to about 39,000 which (for lack of oxygen) is about the upper level to which any aviator should ascend without a pressure suit. This indicates a great superiority of lighter-than-air aircraft as carriers of airplanes. If such a carrier, which is always of the Zeppelin type, cruises at a level of 6000 feet, its passengers are ready at any moment to ascend to the greatest heights without any previous denitrogenation and without danger from aero-emphysema.

* Although the total pressure of pure oxygen (141 mm at 40,000 feet) would be sufficient to sustain normal life, too large a part of this oxygen is replaced inside the lungs by water vapor (the partial pressure of which is 47 mm at blood temperature) and by carbon dioxide (the partial pressure of which is also about 47 mm). At 40,000 feet the remaining partial pressure of oxygen in the lungs even if pure oxygen is given, is, therefore, only 141 minus 94 = 47 mm.

The Action of War Gases

CHAUNCEY D. LEAKE and DAVID F. MARSH
Medical Schools of the University of Texas and the University of Georgia

Before taking up the mechanism of the action of war gases, it may be well to consider the psychological and political aspects of this military weapon, which, apart from its biological effects, imposes the threat of panic among civilians conditioned by fear. Once the popular ignorance relative to "gas" is replaced by knowledge as to how relatively weak a weapon it is against civilians and how effective are simple protective measures against it, the more serious danger of panic based on ignorance will be overcome.

Since high explosives are vastly more effective against cities than is an equal tonnage of gas, the limited carrying capacity of airplanes will preferentially be devoted to demolition and incendiary bombs, whose effects do not evaporate or dissipate. Furthermore, while fatal military casualties due to gas in the last world war amounted to only about 2 per cent, with all other weapons the average was about 28 per cent; and with gas civilians are in a much safer position than the military. Soldiers have to seize, occupy and hold territory in defiance of chemical attack, whereas civilians may take shelter or may move to unexposed positions. Thus, apart from exceptional situations, civilians do not need the elaborate protection necessary for soldiers.

However, the mere threat to use gas, made by the Germans, was sufficient to make the British at the outbreak of the present war divert to the manufacture of gas masks and the like considerable amounts of steel and rubber, which might otherwise have been utilized in making offensive weapons. Gas has not yet been used by the military in the present war of movement, though it probably will be used in static battles or in desperate situations.*

It is very important to give civilians as much information about war gases as possible—as to how they act biologically, and how their effects may be prevented or reduced. If this is carefully done it will become readily apparent that one may reasonably take care of oneself no matter what kind of war gas may be suspected.

It is extremely important to point out that all explosions release "war gases." In addition to carbon monoxide, which is odorless, colorless, and non-irritating, but which fortunately dissipates very rapidly, there are also "nitrous fumes" and the effects of "blast." The action of any of these accompaniments of explosions may be confused with the action of war gas. "Nitrous fumes" are brownish, heavy vapors which will irritate the eyes and nose and lungs, and cause effects very similar to the war gases. The methods of protection devised for war gases are equally effective against "nitrous fumes," or even against hot oil smoke or the smoke and fumes from other incendiaries. Reasonable appreciation of the biological action of war gases will tend to remove the untoward psychological effects of fear of exposure to them.

Biological Action of War Gases

The intensity of biological action of a chemical agent (I) is dependent upon (1) *dosage* (D), in terms of mass of chemical per mass of living material; (2) *ratio* of the *rate of absorption* (rA) and distribution of the drug through the living tissue to its *rate of excretion* (rE) or destruction; (3) *physico-chemical properties* (Ch) of

* See paper on Chemical Warfare in Vol. III of this series, by J. Alexander.

the drug, such as its differential solubility in different solvents, its polarity, its molecular configuration and energy organization, its dissociation characteristics and its optical properties, and (4) *peculiarities (P) of the particular type of living tissue involved,* such as its age, its metabolic and allergic states and its enzyme balance. These factors may be succinctly summarized in the following non-mathematical formula:

$$I = \left[D\frac{rA}{rE}\right], Ch, P.$$

The concentration (C) of the chemical in the tissues at any given moment after administration is given by the product of D and the ratio of rA to rE.

Appreciation of these factors may help to understand the difference in action of various war gases and the variation in intensity of effect of the same war gas in the same concentration on different individuals. A suitable analogy to the latter situation is the difference in response of different people to the same intensity of sunlight or poison ivy.

For our present purposes we may consider as ordinary war gases the lung irritants, like phosgene or chloropicrin, and the vesicants, like mustard gas and lewisite. We may thus disregard such unusual possibilities as catalyzed cyanides or metallic carbonyls, and such gaseous associates of demolition bombs and incendiaries as carbon monoxide, "nitrous fumes," "blast," hot-oil smoke or phosphorus. However, the tissue aggressiveness of "nitrous fumes" suggests that these deserve attention in the same way as ordinary war gases.[1]

As shown in Table 1, the ordinary war gases may be indicated to be chemical relatives of such types of aliphatic hypnotic and inhalation anaesthetic agents as alcohol, chloroform and ether. People generally recognize the locally irritating powers of these common compounds. Their gas relatives may owe an increased irritative action to aggressive factors associated with altered halogenation and polarity.

Table 1. Chemical Relations Between Common Irritant Drugs and War Gases

Aliphatic irritant	Corresponding war gas
Alcohol $H—CH_2CH_2—OH$	Ethyl dichloroarsine $H—CH_2CH_2—AsCl_2$
Chloroform $Cl_3C—H$	Chloropicrin $Cl_3C—NO_2$
Ether $(H—CH_2CH)_2O$	Mustard gas $(Cl—CH_2CH_2)_2S$

These war gases usually contain a rather labile halogen, like chlorine or bromine, which, with the hydrocarbon portion, may be considered to be relatively more attracted to fat and protein than the rest of the molecule. On the other hand, the war gases also contain more potent polarizing elements or radicals, like oxygen, sulfur, arsenic, a nitro group or oxime, which may be relatively more attracted to water or which may reduce the strength of the halogen bond. Differences in relative water-fat solubilities and in ease of hydrolysis may be important factors in the site of action or in the onset or duration of action, as exemplified in the contrast between lacrimators and vesicants.

A common theory explains the action of war gases on the basis of splitting off

[1] Proceedings of a Board of the Chemical Warfare Service appointed for the purpose of investigating conditions incident to the disaster at the Cleveland Hospital Clinic, Cleveland, Ohio, on May 15, 1929. Edgewood Arsenal, Maryland, Lieutenant-Colonel Walter C. Baker, C. W. S., commanding. U. S. Government Printing Office, Washington, 1929, 104 pp.

halogen, with immediate irritant effect from the resulting halo-acid. This may occur promptly on the wet surfaces of eyes, and of mucous membranes of the nose, mouth, throat and lungs, with such agents as the lacrimators, phosgene, and lewisite. On the other hand, as with mustard gas, the partition coefficient may favor absorption into the cells, after which the halogen may split off. The resulting halo-acid within the cell may alter enzyme systems, permeability of the surface membrane or protein equilibria, in such a way as to kill the cell. While such formation of acid may occur, it would have to exceed the buffering capacity of cells and tissues, and this might require relatively large amounts in order to pass the threshold. Neutralization by cellular buffers would be expected to produce the corresponding halide ion which would not markedly affect cellular function. At any rate, exhaustion of the buffer mechanism should reduce further hydrolysis. Direct experiment has shown that molecularly intact mustard gas may be isolated from deep skin layers many hours after absorption. Again, acid injury usually involves protein denaturation and precipitation, whereas war gas injury is characterized more by disturbances of cellular permeability, with swelling, protein hydrolysis and cellular disintegration.[2] Another theory of mechanism of action may be considered. This relates to the relatively rigid molecular configuration of the war gas molecule as compared to the cell membrane. The latter is interpreted as a water-lipoprotein interface.[3] Portions of the war gas molecule seem to be relatively lipo-proteophilic, while other portions seem to be more hydrophilic. If enough war gas molecules are present at the cell surface, distortion of the interface may occur. This would result from orientation of the war gas molecule in accordance with the selective affinity of different parts of the molecule for water and lipoprotein, respectively. If this affinity and the interatomic angle forces in the war gas molecules are greater than the surface-tension forces which maintain the normal cell surface, torsion may follow, with changes in permeability of the surface film, with resulting swelling and further distortion and strain of the surface membrane. This may comprise the initial inflammatory response to war gases, which may go on to cellular rupture, vascular breakdown, autolysis and necrosis, as so well described by Livingston and Walker.[2] Tight packing of cells, as may be accomplished by high ascorbic acid intake,[4] would tend to reduce the intensity of this reaction, as Livingston and Walker noted.[2] Another factor of importance may be combinations of parts of war gas molecules with tissue proteins or amino-acids, with resulting allergic reactions to such combinations, and followed by cellular necrosis.

Whichever mechanism occurs, the prolonged tissue response to war gases would subsequently include the slow removal of necrotic debris, to be followed by gradual repair. In the case of lung irritants, this sluggish process indicates the need for protracted oxygen administration as well as for prophylaxis against psychiatric pneumophobia.

In the biological effects of war gases, therefore, it seems that one or more of the following factors is concerned: (1) relative water, fat and protein solubility, both in transport and in relation to cell surface; (2) relative ease of hydrolysis, with relation to possible formation of halogen acid and the effects of the rest of the molecule; (3) distortion of cellular surfaces due to the molecular configuration of war gas molecules or to their secondary valence forces, and (4) effects of war gas molecules on pH, redox potential and colloid, interface and enzyme equilibria.

As in the case of sunburn or exposure to poison ivy, once the process of war gas injury is under way, one may hope for benefit only on the basis of symptomatic relief, of aiding the removal of necrotic tissue and of promoting repair. It would seem wise,

[2] P. C. Livingston and H. M. Walker, *British J. Ophthal.*, **24**, 76 (1940). See also paper on Inflammation by V. Menkin, in this volume. J. A.

[3] J. F. Danielli and H. Davison, *J. Cell Comp. Physiol.*, **5**, 495 (1934); J. F. Danielli, *Proc. Roy. Soc.*, **B121**, 605 (1937); A. J. Clark, "General Pharmacology, Hndb. Exper. Pharmakol," Erganzungswerk, **4**, 14 (1937).

[4] S. B. Wohlbach, *Am. J. Path.*, **9**, 689 (1933); J. F. Rinehart, L. D. Greenberg, M. B. Olney and F. Choy, *Arch. Intern. Med.*, **61**, 552 (1938).

therefore, to train civilians in "self-aid" against war gases in suspected contact with war gas, since first-aid or professional care is apt to be too late.

In order to reduce confusion of thought to a minimum and thus to help prevent panic in suspected attack with war gas, "self-aid" should be devised in as simple a manner as possible. Recommendations should be based on the least common denominator of effectiveness for whatever is likely to be used by a smart enemy. Since mixtures of war gases are certain to be employed, it seems unwise to worry about specific identification and specific management of potential injury, if such identification is based on such an indefinite procedure as smell.

Absorption of the ordinary war gases and their many obvious chemical relatives may be inhibited by neutralizing hydrolysis, oxidation or adsorption. For civilian use, these methods may be improvised from materials readily available in homes. Since the war gases in general are decomposed or poorly soluble in water, a wet cloth tied over the nose and mouth is a relatively effective barrier to the passage of such vapors, including oil smoke and "nitrous fumes," to the nose, throat and lungs.

The most readily available effective oxidants are the common kitchen bleach solutions, such as "Clorox." These are buffered 3 to 5 per cent sodium hypochlorite solutions and are non-irritating for blotting on the skin, but should be diluted for application to mucous membranes, for washing the skin or for wetting cloths to breathe through. As is well known, such a solution reacts promptly with mustard gas, 2,2′-dichlorodiethyl sulfide (B.P. 217°), converting it to the non-toxic crystalline 2,2′-dichlorodiethyl sulfoxide (M.P. 110°) and probably to other non-toxic sulfones. The use of such sodium hypochlorite solutions for the prevention of mustard gas injury has been widely advertised in England.[5] Confirmation of their effectiveness against both mustard gas and lewisite has been obtained by Professor T. D. Stewart, of the University of California, on scores of human subjects, and by ourselves on humans and experimental animals. It is immaterial whether oxidation of mustard gas produces the sulfoxide or sulfone, or further decomposition, or what is produced on treating lewisite with hypochlorite. Direct experiment shows that such treatment of these compounds or their obvious chemical relatives results in non-toxic residues.

For alkaline hydrolysis, sodium bicarbonate solutions around 2 per cent may be readily prepared in a black-out room by dissolving a teaspoonful of baking soda in a glass of water. Such a solution is helpful in washing out the eyes, nose and throat in suspected war gas irritation, or for wetting cloths to breathe through.

The most suitable and readily available detergent adsorbent is lather from ordinary soap and water or soap flakes or tincture of green soap. This is particularly useful, as are hypochlorite solutions, in preventing skin injury from suspected contact with blister gases. The data in Table 2 show the value of soap and hypochlorite in reducing skin injury (in a rather sensitive test subject) from mustard gas application, in comparison with such a mustard gas solvent as kerosene.

The common blister gases are soluble in kerosene, gasoline, acetone, carbon tetrachloride and similar fat solvents. During World War I, it was naturally assumed that such solvents would be useful in removing liquid blister gas splashes from the skin. We have found no data to support this idea. However, current advice to civilians retains this recommendation. It is to be remembered that kerosene, gasoline and acetone may be absorbed through the skin, and that, like carbon tetrachloride, they are themselves skin irritants. They are also solvents of low viscosity and tend to spread easily. It is unlikely that they would be used carefully under the conditions of excitement existing in the crisis of suspected war gas contact. Our experiments show (Table 2) that even under controlled conditions they are much less satisfactory than lather or hypochlorite.[6]

[5] Half-page advertisement *British Med. J.*, opp. page 445, April 4, 1942.

[6] D. F. Marsh and C. D. Leake, *Calif. West. Med.*, 57, 8 (1942). Acknowledged in spite of printer's many typographical errors.

Pharmacologists have the obligation of establishing and explaining the facts regarding the action of chemicals on living things. They have the privilege of applying such information to whatever practical problem may be appropriate. With respect to war gases, present pharmacological information suggests that the simplest and most effective advice for civilian protection against such gases might be: (1) obey air-raid rules, taking refuge during an alarm in an air-raid shelter or black-out room, with doors and windows shut and the windows screened or heavily curtained on the inside to prevent injury from flying glass, if bombing occurs; (2) if the shelter is broken open by bombing, and if war gases are suspected by fogs, peculiar odors, smarting or stinging in the eyes, nose or throat, or by coughing, sneezing or gasping, or by any other symptoms, tie a cloth soaked in baking-soda solution, or diluted

Table 2. Average Character of Skin Response in Rabbits to 0.05 CC 10 Per Cent Mustard Gas (HS) in Ether, with Treatment Consisting of Thrice Blotting Area of Application (Rough Circle 10 mm in Diameter) One Minute After Applying HS with Gauze Soaked in Kerosene, Soap and Water, or 3 Per Cent NaOCl (Clorox), Respectively *

Day	Untreated	Kerosene	Soap	3% NaOCl
1	Intense erythema and edema	Moderate erythema, slight edema	Diffuse erythema, edema	Diffuse erythema
2	Diffuse erythema and edema, central blanched area	Diffuse erythema and edema, central blanched area	Blanched area, 10 × 15 mm	Blanched area, 10 × 15 mm
5	Deep hemorrhagic necrotic area, 10 × 12 mm	Hemorrhagic necrotic area, 12 × 15 mm, with diffuse necrosis at edges	Thin scaley necrosis, 8 × 10 mm	Thin scaley necrosis, 8 × 10 mm
15	Heavy adherent scab, 10 × 12 mm	Broad adherent scab, 15 × 20 mm	Thin flakey scab, 8 × 10 mm	Thin flakey scab, 8 × 10 mm
22	Heavy adherent scab, 10 × 12 mm	Broad adherent scab, 15 × 20 mm	Light scar	Light scar

* No significant difference from untreated areas observed after application (as above) of either 3 per cent H_2O_2, acetone or "bleach paste." Treatment with 5 per cent NaOH in 30 per cent glycerine seems to increase inflammatory reaction during the first week, producing a deeper and slower healing necrotic area. Ten per cent benzoyl peroxide in nona-ethylene glycol seems to have little effect on HS reaction during first day or so, but seems to reduce necrotic reaction and time required for healing. However, 10 per cent benzoyl peroxide in talc affords no protection when dusted on skin previous to exposure. Observations similar to the above have been obtained with lewisite; healing, however, is more rapid.

kitchen bleach solution, over the nose and mouth to breathe through, keep it wet, shut one eye and squint through the other, lie down with head in arms; (3) if eyes, nose or throat are irritated, wash them with a solution of a teaspoonful of baking soda in a glass of water; (4) if splashes of liquid are suspected on the skin or clothes, throw the outer clothing out the window, blot the skin, splash promptly and repeatedly with a cloth wet with kitchen bleach solution, lather thoroughly and frequently with soap and rinse copiously with water. If subsequent injury results, the management is symptomatic at a casualty station or hospital.

These considerations were fully reviewed early in 1942 by the San Francisco and Alameda Committees on the Medical Aspects of War Gases. Special discussions along these lines have been widely published on the West Coast for civilian informa-

tion.[7] Experience has shown that these suggestions for "self-aid" in handling suspected war gas exposure are appreciatively received by the public because they are simple and sensible. Recently these suggestions in substance have been included in "official" recommendations.[8]

Summary

War gases act on people economically, politically and psychologically, as well as in direct biological ways. It is wise to offer all available information on war gases to the public in order to dispel ignorance, reduce fear, and prevent panic from their possible use against civilians.

Infective Aerosols

PROF. M. W. JENNISON

Department of Biology and Biological Engineering, Massachusetts Institute of Technology, Cambridge, Mass.

Introduction

Broad consideration of the aerial dissemination of disease involves, first of all, setting up certain categories of the disease-producing agents, of the mode of their transmission in air, and of the organisms or substances which they may infect. In this review we are concerned particularly with those *air-borne, living* entities which, as a result of their *growth* in the *living* body, are potentially able to produce an *infection* or disease. These atmospheric colloids we shall term *infective aerosols.* Various bacteria, viruses, and molds (as spores) are the types of microbes most commonly involved as infective aerosols. The transmission of disease agents by insects will not be considered as belonging in this category. Also excluded by this definition, though they will be touched upon briefly, are "diseases" of inanimate objects, and pollen allergies. (Pollens, although they may be air-borne and may produce reactions in the animal body are not, strictly speaking, *infective* agents.) Although many plant diseases are air-borne, primary attention will be given to infective aerosols in relation to human diseases. Fundamental points as regards aerial infection are the viability of the organisms in the air, that is, how long they will remain alive and *infective,* and their dilution in the air.

The infective agents may be borne *by* the air, or the air may merely be the medium *through* which infection is carried—for example, by insects. In the former case (strictly air-borne infection), both air movement and particle size are factors which determine whether or not the infective agents will remain suspended in the air, that is, whether they are *aerosols.* Aggregates of the single units or cells of infective agents may occur, so that in the air the average size of infective aerosols probably is greater than that of the single components. A typical bacterial cell is about 1-2μ in size; virus particles are all less than 1μ, and some are as small as 0.01μ. The size range of various kinds of mold spores is from 3μ to 50μ or more. Most wind-borne pollens are between 20μ and 60μ.

Both living plants and living animals are susceptible to various diseases that may be air-borne. Infection of non-living organic susbstances by microbes also may

[7] Articles on war gases by J. F. Hildebrand (*The Commonwealth,* 1942; *San Francisco Chronicle,* Feb. 15, 1942), M. Silverman, (*San Francisco Chronicle,* March 15, 22, 29, Apr. 5, 1942) and W. F. Mould, leading West Coast newspapers through June and July, 1942

[8] *J. Am. Med. Assn.,* 119, 889 (July 11, 1942).

take place, resulting in spoilage, as, for example, in the "diseases" of beer, and in the mildewing of textiles. The principal types of air-borne microorganisms causing diseases of animals are bacteria and viruses; of plants, mold spores and to a lesser extent bacteria. Non-living organic substances may become infected by molds, bacteria, or yeasts.

It may be well here to stress the point that the demonstrable presence, in the air, of living pathogenic microbes cannot in itself be taken as proof that infection is necessarily caused thereby, but only makes evident the possibility of such infection. As will be brought out later, there is abundant evidence for the presence of potentially infective aerosols of both plant and human origin. However, since some degree of contact between persons is hard to rule out, and for other reasons, it is difficult to obtain sound experimental proof of the *relative* importance of the transmission of *human* diseases by strictly air-borne means compared to other methods of spread.

Historical Perspective

The idea that disease might be transmitted from person to person through the air dates back to antiquity. Some of the outstanding historical developments in connection with the modern concepts of air-borne human infection have been discussed by Chope and Smillie,[44] and some of the following historical background is summarized from their paper.

Four hundred years before Christ, Hippocrates stated that it appeared that the cause of disease should be found in the air, particularly when it was tainted by *morbid miasmas*. The concept of "miasmas" flourished during the middle ages. The germ theory of disease was anticipated by more than 300 years by Fracastorius when he suggested (1546) that contagious diseases were caused by "seeds of contagion" which probably were carried in the air. The disease malaria ("bad air"), the germs of which are now known to be carried by certain kinds of mosquitoes, was long thought to be caused by atmospheric miasmas, although its connection with the mosquito was hinted at over 400 years ago.[2]

The germ theory of fermentation and of disease, and the first scientific demonstration of air-borne infection (1861), were due to Pasteur.[158] As a consequence of these discoveries many persons concluded that all or nearly all diseases were spread through the medium of the air. However, the rise of bacteriology during the subsequent forty or fifty years did much to break down this hypothesis, particularly as regards the intestinal diseases. Many diseases were shown to be transmitted by other routes, *e.g.*, by contaminated water and food, by contact, and by insects, and the number of diseases thought to be air-borne decreased correspondingly. This decrease was the more real since investigators were unable to demonstrate recognized pathogenic organisms in the air, even in those diseases which are spread through discharges from the respiratory tract. Flügge [82] in 1899 was unable to obtain any appreciable numbers of organisms on culture plates exposed farther than a few feet from persons coughing and sneezing, and concluded that the expelled droplets settled out of the air quickly. Thus *droplet* infection (which resembles contact infection more than it does strictly air-borne infection) was shown to be possible within a short radius, but at the same time strictly air-borne microbes were not demonstrated, and therefore air-borne contagion seemed eliminated, even in respiratory infections. The negative evidence obtained by the experiments of Flügge and contemporary investigators (see [252]) carried great weight, but as has been remarked [286] the case against air-borne infection was dismissed with a Scotch verdict of "not proven." In 1912 Chapin [87] reviewed the literature on the transmission of contagious diseases. He concluded that close personal contact was the most important factor in the spread of respiratory diseases, and that there was no good clinical evidence that the common diseases are airborne.[88] These conclusions regarding contact (including droplet)

infection were accepted generally, the theory of air-borne infection discredited, and for the next two decades little attention was paid to the latter mode of disease transmission.

The whole question was reopened following an epidemic of psittacosis (parrot fever) in the National Institute of Health at Washington in 1930.[150] Epidemiological studies indicated that the infection in this instance was air-borne. Consequently a new stimulus was given to investigations of the role of air in the transmission of infection. Development of more sensitive bacteriological air-sampling techniques [49] has led to abundant proof of the presence of pathogenic aerosols. Recent bacteriological and epidemiological evidence indicates that respiratory disease of air-borne origin is, in certain situations at least, an actuality, and that this method of spread is more important than has hitherto been supposed.[18, 25, 240] There is no question but that close personal contact is a very important factor in the transmission of respiratory diseases, although it is not the only factor involved. It has been stated [140] that the spread of certain diseases such as measles and smallpox, and the 1918 influenza epidemic, stretched the theory of contact (or droplet) infection to the breaking point.

For a complete discussion of the many phases of the subject of the aerial dissemination of infection, including air-borne diseases of plants, and pollen allergies, reference may be made to the recent symposium on Aerobiology.[18]

Introduction of Microorganisms into the Air

Particulate matter is dispersed in the air from its source in two ways: by dynamic projection and by air carriage.[106] The sources of atmospheric contamination with pathogenic aerosols are infected organisms or their products. In man, the most important place of origin of such aerosols is the respiratory tract. The infective agents of the respiratory diseases are ordinarily contained in the secretions and discharges from the respiratory tract of the person having the disease, and are "atomized" into the air in the process of sneezing, coughing or talking. The agents of certain animal diseases may be given off in a similar manner. Desquamation, or scaling off of the skin in certain diseases, also may give rise to infective aerosols. "Dust" may contain large numbers of microorganisms of human origin which can remain viable for long periods of time. Dust also may be the source of infection of various organic, non-living materials. Many plant diseases are caused by molds, which produce large numbers of reproductive spores that are readily air-borne. Pollen grains, which are involved in reproduction in higher plants, are given off when the male reproductive structures of the flower are mature.

In relation to *human* diseases, infective aerosols must be studied chiefly in the semi-enclosed spaces of human habitation; aerosols infective for *plants* are found in the open air. These are the fields of intramural and extramural aerobiology respectively.

Infective Aerosols from Respiratory Activities. "In sneezing, coughing, and talking, droplets of saliva and other secretions, containing microorganisms, are expelled into the air. These droplets, particularly in diseases of the respiratory tract, constitute a potential source of infection. The larger particles are deposited onto nearby objects or fall to the ground before they can dry, although subsequently they may evaporate and their residue be lifted into the air as dust. Smaller droplets evaporate before reaching the ground, thus leaving their content of microbes suspended in the air. Such air-borne 'droplet nuclei' [217] may remain in suspension for long periods of time and be carried long distances. The aerial transmission of infection, therefore, is possible in two ways, depending upon the size of the infected particle, namely, by droplets proper and by air-borne droplet nuclei. Although droplet infection has long been recognized, experimental evidence in support of strictly air-borne infection has, until lately, been lacking." [118]

Previous work on the production and physical characteristics of respiratory drop-

lets, together with the writer's investigations of the atomizing of mouth and nose secretions into the air as revealed by high-speed photography, have recently been reviewed in detail.[118] Therefore, only a few of the outstanding points will be summarized here.

In connection with the interpretation of studies on respiratory droplets, the question of the technique employed is significant, since many techniques *disproportionately underestimate* the numbers of smaller particles. This was an important reason for the inability of earlier investigators to demonstrate appreciable numbers of bacteria in the air. Particles usually have been captured by settling, by impingement, or by filtration. The particles may be caught on glass slides and the actual droplets counted, or in a liquid medium or on agar and the resulting colonies of bacteria enumerated. Animals,[226] and the chick-embryo technique [29] have been employed to demonstrate virus particles. These methods and some of their limitations have been discussed by several authors.[22, 44, 49, 73, 112, 118, 148a, 149, 160, 216, 230, 234, 245] Recently, photographic methods have been used to study respiratory droplets during and immediately after their expulsion.[21, 117, 118, 119, 121, 205, 244]

The number of actual droplets obtained from a cough, or from speaking for a few minutes or enunciating letters, particularly consonants, varies from a few to a few hundred.[80, 81, 82, 89, 92, 118, 125, 131, 155, 195, 196, 244, 252] In speaking, the greatest amounts of spray are given off in enunciating consonants like *p*, *t*, and *f*.[89, 118, 125, 252] In sneezing, much larger numbers of droplets are produced than in coughing or talking.[21, 117, 118, 119, 236, 244] Photographically, up to 40,000 particles per sneeze have been demonstrated (see Fig. 1), and 20,000 are common;[118] up to 100,000 bacteria per sneeze have been recovered by bacteriological methods.[21, 21a, 218] It appears that much greater numbers of droplets are produced than has previously been supposed, particularly in sneezing. In both stifled and unstifled sneezes, the *number* of par-



FIGURE 1. Introduction of micro-organisms into the air—a violent sneeze. High-speed photograph (exposure 1/30,000 of a second) showing the droplets during their evaporation to air-borne droplet nuclei (aerosols). Over 40,000 particles are shown. Many of the droplets were out of sharp focus, thus producing images much larger than actual droplet size. The actual diameter of many of the droplets depicted is about 10μ. A few droplets may be seen coming from the nostrils.

ticles issuing from the nose usually is meager compared with the number expelled from the mouth, while the *size* of the former is greater than that of the latter.[118] This is of interest in relation to *air-borne* infection, since it is the *small* droplets—those originating chiefly in the mouth—which evaporate while in suspension in the air. For an interesting note on the appearance in three human generations of an atypical pattern (double sneezes) of the sneezing reflex, see [76].

The number of bacteria expelled in respiratory droplets seemingly is at least equal to the number of even finely-atomized droplets.[109] Specific pathogenic bacteria usually are present in respiratory droplets in relatively small numbers compared with the total number of droplets and with the numbers of other organisms.[18, 47, 91, 92, 108, 162, 191, 198, 258] Unpublished experiments by the author give the average number of bacteria in human saliva, which are cultivatable on nutrient agar, as about 6,000,000 per cc; up to 50,000,000 per cc may be found.

The horizontal distance to which the great majority of particles from the respiratory tract are actually expelled before they hit the ground or evaporate is not over 2-3 feet, as shown by the early work cited (see [252]), and by later studies,[41, 92, 109, 117, 118, 130, 155, 195, 196, 244] although a few large droplets or masses may be projected up to 10-12 feet.[117] Certain mathematical calculations of the distance of projection of sneeze droplets have recently been made,[106] based on an initial velocity of 152 feet per second.[117]

The manner in which droplets are formed, that is, the mechanics of the "atomization" of liquids, is intimately related to the size, velocity and subsequent behavior of the particles. In coughing, sneezing and talking, droplets are formed by air-stream atomization, in which air at high speed is passed over the surface of the secretions, and the liquid is violently torn by the air-stream. Most sneeze droplets appear to originate from secretions present at the front of the mouth, more and smaller ones being produced when this orifice is restricted,[117, 118] A diameter of about 10μ is the minimum droplet size producible from *water* atomized in a high-speed air-stream (100 meters per second and above).[86] Saliva and nasopharyngeal secretions are more viscous than water [148, p. 88, 197, p. 486]; and consequently can be atomized less finely under the same conditions.

The air pressures and velocities produced in respiratory activities are related to the number, size, velocity, distance of expulsion, and evaporation of droplets. In "loud" speaking, air velocities up to 16 meters per second have been found.[195] According to one report,[48] the velocity of air past the glottis in coughing may reach 100 meters per second. Sneeze droplets with an initial velocity of over 152 feet per second have been photographed.[117, 118] Such a velocity, particularly for the smaller droplets, will produce almost instantaneous evaporation to their ultimate size.[118, 217] Because of their viscosity, and in spite of the high velocities recorded, it is improbable that in respiratory activities droplets as small as 10μ in diameter are *originally* formed.[117]

Droplet size largely determines the physical behavior of droplets after expulsion, and involves several interrelated factors such as velocity, surface tension, viscosity, chemical composition, particulate inclusions, and evaporation of the droplets. Droplet size has usually been determined from measurements on glass plates upon which the particles settle or impinge. For physical reasons the smallest particles must be poorly represented, if at all, in samples so obtained. Such measurements of particle size *in the air at the time of collection* of droplets produced in *talking and coughing* show the range of diameter to be from about 20μ to 2,000μ, with a majority between perhaps 100μ and 500μ, and appreciable numbers below 100μ.[11, 109, 155, 195, 196, 218]

Measurements of *sneeze* droplets on high-speed photographs show that many evaporate at least to 7-10μ in diameter within a fraction of a second after expulsion,[118] which particles would be among the smallest of air-borne droplet nuclei. The same technique shows that droplets produced in coughing and in talking are in

general somewhat larger than those resulting from a sneeze. These studies demonstrate that a higher proportion of *small* droplets are produced in respiratory activities than has hitherto been supposed. Particle size is of importance not only in relation to the state of suspension of particles, but also as regards their inhalation and other processes concerned in their removal from the air.

Experimental studies on droplet dynamics are meager; some general aspects have been considered.[20, 86, 106, 249] The final size of atomized droplets depends upon the original size and upon subsequent evaporation; the latter is determined by several interdependent factors, some of which have been discussed in relation to respiratory droplets.[213] Both the kind and amount of dissolved substance have been shown to alter markedly the rate and degree of evaporation of minute droplets.[207] It is probable that the final size of droplets is about that of the included bacteria or other particles.[213] Wells [217] has discussed the basic principles of the formation of "droplet nuclei" from droplets proper, particularly in respect to the time of, and distance to, evaporation of droplets. The life of small droplets in unsaturated air is much shorter than their time of fall if they settled without evaporation. Droplet size [217] and velocity in respect to the air [118] are of predominating importance as regards evaporation, while air temperature and humidity have a relatively smaller effect.[217] Wells [217] concludes: "Somewhere between 0.1 and 0.2 mm lies the droplet size which identifies droplets of mouth spray that reach the ground within the life of the droplet as against droplets that evaporate and remain in the air with attached infection."

It is hardly necessary to offer proof that individuals with respiratory disease harbor the causative organisms in their respiratory tract. It is perhaps more significant to mention that many otherwise "normal" individuals, *i.e.*, without disease symptoms, carry pathogenic microbes in their nose and throat and on the skin.[34, 59, 92, 93, 96, 104, 116, 144, 147, 152, 200, 257] These organisms may be expelled even in ordinary respiratory activities. The presence of such normal "carriers" in hospital wards and other situations appears to be an important source of infective aerosols, [4, 5, 47, 136, 137, 161, 255] since weakened hospitalized persons tend to be more susceptible to infection. That infective aerosols may be given off from the respiratory tract of certain animals is evidenced from investigations on psittacosis (parrot fever), [150, 194] canine distemper,[67] and tuberculosis in rabbits.[229]

Infective Aerosols from Other Sources. The production of aerosols from the scaling off of skin has long been recognized. An example quoted by Alexander [8, p. 408] is the observation in 1868 by Dr. James Ross of epidermal scales from a case of confluent smallpox. In a beam of sunlight ("Tyndall beam" effect [208]) many scales were seen to remain in suspension in the air. Desquamation is also produced in other diseases, and even from normal skin, although strict proof of air-borne disease thus caused is lacking. Certain bacteria have been found in large numbers in the air about patients with burns infected with the same organisms,[51] presumably having been given off from the infected area.

"Dust," which is a heterogeneous mixture of organic and inorganic, dead and living particles, becomes suspended in the air through movement, such as dry sweeping or beating. Many studies,[4, 24, 48, 93, 113, 114, 162, 198a, 209, 212, 246, 248, 255] to mention a few recent ones, particularly of dust in hospitals, show infective aerosols to be present in dust from floors, and from bedclothes and the skin and clothing of infected persons. Some of such dusts consists, of course, of dried, settled particles from respiratory activities. The length of survival of microorganisms in dust varies greatly with the type of organism and with environmental conditions such as light, temperature and humidity; [27, 28, 48, 114, 191, 248, 255] some bacteria may remain alive for years.

Obviously, dust containing various microorganisms may also be the source of infection of non-living materials such as food products,[9] textiles, etc., which contain some moisture and food for the microbes. One study has been made of the contami-

nation of the air of textile mills by evaporation of droplets of contaminated water used for humidification, and from manipulation of the fibrous material being processed.[234]

The dissemination of many plant diseases by air-borne spores of the invading mold is well recognized.[193] The parasite may be superficial or otherwise; in either case the reproductive spores are usually produced on the surface. Spores may be forcibly discharged into the air in various ways, as well as being passively released.[123] In some plant diseases spores are produced in astronomical numbers; it has been shown that various kinds of fungus spores may be carried hundreds of miles by air currents.[45] The inhalation of certain types of fungus spores may also produce allergic conditions in man.[70] Further discussion of this subject may be found [45, 70, 123, 166, 193] in the symposium on Aerobiology.[1a]

Pollen grains may be insect- or wind-borne. They are exposed in the flowers of higher plants when the male reproductive structure is mature. The pollens most commonly causing "hay-fever" come from various grasses and ragweeds, although other kinds of plants produce pollens to which some persons are allergic.[253, 254]

Conditions of Suspension of Infective Aerosols

Particle size and air movement are the chief factors determining whether aerosols will remain in suspension or will sediment (physical stability). In addition, from the standpoint of infection, the type of organism and various environmental conditions determine how long the aerosols will remain alive and infective (biological stability or viability).

Presence of Living Aerosols in the Air. Under practical conditions it is not easily determinable what proportion of atomized respiratory droplets remain in the air as droplet nuclei or become resuspended as dust, but the aggregate may be large under certain conditions. Many investigators, particularly during the past twenty years, have isolated from air not artificially contaminated, considerable numbers of air-borne bacteria characteristic of the respiratory tract [24, 30, 31, 52, 60, 71, 97, 102, 104, 115, 137, 138, 139, 161, 162, 186, 191, 198a, 236, 240, 248, 251]; also references in [231]. These nasopharyngeal organisms have been found to be most prevalent in the semi-enclosed spaces of human occupancy, such as schools, hospitals, institutions, and the like, their numbers being roughly proportional to the degree of occupancy and activity and to the amount of respiratory infection. Usually the numbers of pathogenic bacteria are small in comparison with the total air flora.

A few thousand bacteria per 10 cubic feet of air were reported in a room in which sneezes were artificially induced in the occupants.[236] In other occupied spaces—schools, theaters, subways, hospitals—under a variety of conditions the total numbers of air-borne bacteria recoverable by the Wells centrifuge [216] have been found to be up to a few hundred per 10 cubic feet of air, with nasopharyngeal organisms present in larger numbers than has previously been thought.[31, 137, 138, 139, 161, 186, 231, 236, 240] It is obvious that the dilution of bacteria in air, particularly with artificial ventilation, is important in relation to quantitative air bacteriology. In outdoor air the total numbers of bacteria are in general much less than in indoor atmospheres,[240, 251, 259] although some have been found at altitudes of 4-5 miles.[163-166]

Fungus spores also are widely distributed throughout the atmosphere. For aerial mold spore surveys, see [16, 45, 70, 148]. Numerous studies of the prevalence of pollens in the air have been made.[49a, 55, 68, 253]

Physical Stability of Aerosols. The settling of aerosols is of practical importance, even in still air, only when the diameter of the particle exceeds about 10μ.[86] That particles less than 10μ in diameter result from respiratory activities, and that many mold spores and some pollen grains are smaller than this has been mentioned above. Air movement will, of course, keep particles much larger than this in sus-

pension for varying periods of time. (For methods of sampling air bacteriologically, see [22, 29, 44, 49, 78, 112, 148a, 165, 178, 216, 226, 245]; for sampling of air-borne pollens, fungus spores and insects, see [49, 69, 70, 87, 106a, 165, 253, 254].)

It is demonstrable that a proportion of atomized infected droplets may be carried for long distances by air currents, and the bacteria remain alive. This was recognized by the earlier investigators,[252] and later studies have confirmed it.[41, 48, 135, 155, 284, 286] Living aerosols may be disseminated through mechanical ventilating systems [57, 110, 170, 284]; (see also references in [188]). The length of time that aerosols will remain in suspension in *still* air may be estimated from Stokes' Law.[62, 159, 217, 280]

It is technically difficult to determine experimentally the true settling rates of living aersols, since the cultural methods used give an apparent rate which may be influenced by the death of the organisms. Apparent settling rates of infective aerosols have been measured in closed chambers, using finely atomized suspensions of bacteria representing droplet sizes that are produced in respiratory activities. Settling appears to be at a rate which decreases geometrically, and much of the material remains in suspension for many hours. Using a resistant organism, and with a mean droplet diameter of about 120μ, it was found that after the immediate settling of the heavier material and evaporation of the remainder, the material disappeared from the chamber at a rate, including the death rate, of 7 per cent per hour over a 24-hour period.[235] As measured by the development of colonies upon agar plates, short-chain alpha hemolytic streptococci in droplets of a mean diameter of 10μ deposited about 22 per cent of its ultimately settleable particles during the first hour of suspension in the dark in a room 8 feet high. About 16 per cent of the initial number remained suspended for over 8 hours. As calculated from the settling rate, two-thirds of this evaporated material was less than 1.34μ in diameter, and 95 per cent of it less than 2.68μ.[159, 160] "It represents, therefore, exactly that state of suspension in which sanitarians are most interested; that is, particles of the sort produced by sneezing large enough to carry bacteria and so small that the aerial life is measured in hours rather than minutes." [159] The apparent rate of settling (including the death rate) of vaccinia virus in the dark is about twice that of streptococci.[29] Probably this greater rate of disappearance is largely accounted for by a higher death rate.

A bacteriological study of the settling rates of natural sneeze droplets in an experimental room showed that of the total recovered, 20 per cent of the particles remained suspended for 15 minutes, and 4 per cent for 30 minutes.[21a]

One investigator,[145] exposing culture plates in an operating room during operations, estimates that 30,000-60,000 viable bacteria may fall on the "sterile" field of operation in an hour. Other studies [7, 98, 102, 104, 140, 156, 215] show varying numbers of bacteria present in and sedimenting from the air of operating rooms. The numbers are much higher in winter (due presumably to the greater incidence of respiratory infections) than in summer; when the room is occupied than when unoccupied; and in rooms not air-conditioned.[98, 104] The relation between the number of bacteria settling out of the air, and those remaining suspended, has been studied in a textile mill.[234] Other data on the size and settling of infective aerosols may be found in [238]. For settling rates of fungal spores, see [45].

As indices of the bacterial pollution of air by human beings, streptococci,[89, 218, 281, 282, 283, 288] *Neisseria catarrhalis* [120] and *Staphylococcus aureus* [104] have been suggested. Streptococci have been studied most thoroughly in this regard.

Viability of aerosols in normal air. The viability of aerosols is of outstanding significance as regards air-borne infection. Controlled experimental studies on this point are rather meager. It is apparent that any type of bacteriological air-sampler partially simulates conditions of possible human infection, in that only those aerosols which are viable will be enumerated by the methods.

There is a general agreement that many bacteria and viruses—including such

respiratory forms as influenza virus, streptococci, pneumococci and diphtheria bacilli —may remain alive in "normal" air for hours.[26, 27, 28, 29, 83, 159, 192, 226, 227, 285] The death rate increases at higher temperatures.[27] The effect of humidity on the death rate is difficult to separate from its possible effect on the settling rate, and has not been clearly shown. Some investigators [184, 204] believe that microbes survive longer in humid than in dry air; others [284] claim the reverse to be true. Survival of streptococci and pneumococci at least, is much longer in the dark than in the presence of sunlight, diffuse daylight, or illumination from fluorescent lamps,[26, 28] survival in the dark being measured in days, and in light in minutes or hours. Even ordinary daylight appears to be a more potent lethal agent for microorganisms than has been generally recognized. The relation between light intensity and lethality appears to be linear. Influenza virus suspended in air has been found to remain viable for at least an hour in the dark,[226] and vaccinia virus for eight hours.[29] Influenza virus is rather resistant to drying, and can be demonstrated in dust.[72] Not only may pathogenic bacteria remain *alive* in air for days, but their *virulence* (infectivity or disease-producing power) likewise appears not to decrease.[32, 33, 248] The viability of pathogenic aerosols under the influence of specific lethal agents will be discussed in the next section.

Spores of various fungi causing plant diseases, despite the many unfavorable conditions during transit, do survive the vicissitudes of wind dissemination.[45]

From the foregoing it is apparent that air-borne living entities, on the order of a few microns in size, do exist. Aerosols of less than 10μ diameter are physically very stable even in still air,[86] and under conditions of normal air movement much larger particles will tend to remain in suspension. Aspects of particle size in relation to inhalation by man, particularly as regards the removal of particles in the respiratory tract, have been considered by [106, 180, 185, 218, 230].

Control of Infective Aerosols

A number of methods are available for the control of pathogenic aerosols and thereby of human air-borne infection. Among the techniques are the use of masks; spatial separation, including mechanical barriers, of infector and infectee; ventilation and air-conditioning, including dilution and filtration; keeping down dust; irradiation of air with ultraviolet light, including light barriers; and the use of germicidal sprays. Generally speaking, the practical application of these methods has lagged, perhaps largely because the necessity for the control of air-borne microorganisms has only lately become apparent. Recently attention has been given to the control of air-borne infection in air-raid shelters and similar situations.[6]

As regards the control of aerosols infective for plants, destroying infected plants and their sources of infection have been practiced. Attempts have been made in some communities to control the pollen of ragweed, in particular, by destroying the plants. As a practical matter, however, this method of trying to prevent "hay-fever" is of rather questionable merit, because of the long-distance dissemination of pollen from other sources.

Masks. Masks *tend* to prevent the introduction into the air of particles from the respiratory tract. Protection of patients against direct expulsion of nasopharyngeal organisms by the use of masks during operations is now common surgical practice. Less common is their general use in hospitals and elsewhere to protect debilitated patients, particularly children, from the nasopharyngeal flora of the attendants.[151]

Masks are most commonly employed to prevent the *expulsion* of droplets rather than the *inhalation* of dried droplet residues. The various types of fabric masks in common use differ greatly in their permeability to air and to particulate matter, as shown by tests using the air volumes and pressures involved in respiratory activities.[118] We have found this range of variation for air passage to be about 10-fold in terms of *volume* of air passed through the fabric per unit area per unit time at

constant pressure,[118] which corresponds to an average of a 60-fold variation in *resistance* to air flow (pressure-drop across the fabric) at constant volume. Experimental tests of the efficiency of such face masks against droplet expulsion reveal that varying numbers of respiratory particles pass through the fabric or around the edges of the mask.[10, 17, 85, 118, 124, 134, 151, 168, 187, 202, 214] Deflector masks prevent only direct droplet expulsion, not air-borne particles.[118, 187] Clinical observations also suggest that many types of masks are not entirely adequate to prevent the transmission of infection.[58, 59, 96, 100, 104, 137, 145, 151] Hence, as has been stated,[151] the adequacy of the mask must be viewed with scepticism, although some types are over 90 per cent efficient under stringent test conditions.[118, 187] It is apparent that great comfort in the wearing of a mask, together with high filtering efficiency, are very difficult to attain. Extra heavy masking of operating room personnel who are "carriers" of pathogenic bacteria in the nose and throat will reduce the number of aerial microbes and the incidence of post-operative wound infections.[47, 59, 96, 144, 145, 214]

Ordinary face masks sometimes are used as a protection against the *inhalation* of infective aerosols. The efficiency of such masks *as usually worn* probably is low, due in part to the fact that they do not fit tightly around the face. (For numerous other references on masks and masking, see [151].)

Separation of infector and infectee. This type of control may be essentially a diluting of the aerosol in the air, or may be based on preventing or directing the dispersal of the infective material or even upon destroying it. The so-called barrier method of nursing, or bed isolation, has been used in contagious disease hospitals for years. The barrier may be air (spatial separation), some type of solid screen, or, more recently, screens of ultraviolet light. Barriers have been employed in infants' hospitals and similar situations, each patient being placed in a separate cubicle. The early success of some of these methods in contagious disease hospitals was taken as evidence for the contact theory of infection; undoubtedly the methods do help in preventing contact. A few examples will serve to indicate the methods used and the results obtained.

By allowing more space between beds in dormitories where a number of persons sleep in the same room, the incidence of certain respiratory diseases can be decreased.[52, 64, 66, 71, 88] Likewise, in hospitals where contagious disease is present, cross-infections are fewer when the beds are farther apart.[94, 189, 255] This presumably is due in part at least to a diluting of the air-borne microbes. However, some diseases of respiratory origin cannot be controlled by simple spatial separation.

Preventing or directing the dispersal of infective aerosols may be accomplished in some measure by placing patients in separate cubicles.[5, 23, 58, 75, 151, 176, 188, 255] These cubicles may or may not have a common air supply; ultraviolet light recently has been used in conjunction with such cubicles for infants. Bacteriological air-sampling in cubicles shows that mechanical barriers alone prevent to varying degree the dissemination of living aerosols. The principles of the control of cross-contamination by the use of mechanical barriers, and the application to a positive-pressure cubicle system for infants,[188] and to the rearing of germ-free animals have been discussed by Reyniers.[173, 174, 174a]

Separate incubators, each with its own supply of circulated outside air, have been employed with good results to protect premature infants against infection.[39, 40] Various methods of control of aerial respiratory contagion have been studied under controlled conditions at the Cradle, an adoption nursery at Evanston, Illinois.[50, 188] Here, as elsewhere,[23, 58, 176] it appears that mechanical barriers (cubicles) alone are a less efficient means of preventing the aerial dissemination of microorganisms than when used in conjunction with air-conditioning or ultraviolet light. Ultraviolet light screens alone have given encouraging results. (For further discussion of control by ultraviolet radiation, see below.)

Ventilation and air-conditioning. While it is generally assumed that the bacterial pollution of occupied rooms may be greatly reduced by dilution with clean outside air, certain studies do not entirely confirm this view.[110, 256] The number of living aerosols may actually be increased by stirring up infected dust particles off the floor, or from inside the ventilating system. Much depends upon the air-distribution system used, the number and activity of the occupants, and the general cleanliness of the room. The direction of air flow into the room should be transverse or upward, not downward.[256] Improper air flow into a room may more than offset any advantage gained by filtering the air. Under proper conditions, however, ventilation, particularly with air-conditioning, may reduce the numbers of suspended bacteria in occupied rooms.[52, 97, 98, 188, 256]

Commercial air filters, particularly if treated with a viscous oil, and air-washers, are fairly efficient (40-80 per cent) in removing bacteria,[56, 256] although the effectiveness of filters in actually reducing the bacterial population in occupied rooms is small or nil.[256] Filters are also effective for pollen. For a method of testing filters against pollens, bacteria and molds, see [63].

The measurement of sanitary ventilation has been discussed by [231, 232, 233, 237, 240].

Control of dust. As mentioned previously, dust contains large numbers of microorganisms which may remain viable for long periods of time. Dust becomes resuspended in the air through movement, such as sweeping or the activity of the occupants of a room. Many studies have shown the relation between activity and the number of living aerosols, hence keeping down dust particles is hygienically important. The treatment of floors and bed-clothes with mineral oils, to reduce dust-borne bacteria, has recently been advocated,[198a, 199, 209, 210, 211, 212, 246] as has the treatment of floors in breweries to prevent infection of beer.[9]

Radiant disinfection of air. In recent years irradiation of air with ultraviolet light, as a method of controlling living aerosols, has been rather extensively studied, both experimentally and under practical conditions in hospital wards and operating rooms, nurseries, and schools.

Fundamental aspects of the action of ultraviolet light on microorganisms, including those suspended in air, have been discussed.[65, 74, 84, 111, 126, 132, 133, 137a, 171, 172, 190, 221, 222, 227, 236] For many kinds of bacteria and fungus spores the most lethal wavelength is close to 2650 Å; exact data for viruses infective for human beings are lacking. Bactericidal action falls rather sharply, although not uniformly, on either side of this peak to about 2300 Å on one side and 3000 Å on the other. The activity rises again below 2300 Å, although this lower region has been investigated but little in this regard. Bacteria, fungi and some viruses can be inactivated by radiation between 3400 Å and 4400 Å. The mechanism of inactivation in this range is apparently of a different nature from that of killing produced by radiation below 3000 Å.[111] Wavelength 2537 Å, which is emitted most intensively by the low-pressure mercury vapor lamps commonly used for many practical purposes, is close to the maximum of sensitivity of bacteria.

The amount of radiant energy, time, distance, velocity of air movement, etc., are factors which must be considered in bactericidal irradiation of air.[137a, 221, 236] Different types of bacteria vary, although often relatively little, in their response to ultraviolet radiation;[111] with proper energy output a few moments exposure may be lethal within a short distance. It is possible to show, however, that a given species of bacteria is not equally resistant to ultraviolet radiation throughout its life span.[172] Bacterial sensitivity is influenced but slightly by temperature within ordinary ranges,[84] although surviving irradiated organisms may show increased sensitivity to heat [54] and to disinfecting agents. The intensity of radiation multiplied by the time of exposure is a constant for a certain percentage of killing.[126, 171, 172] With bacteria suspended in air, some investigators have found ultraviolet to be more

lethal in dry air,[126, 187a, 221, 222, 236, 247] while others[171] claim that this effect is more apparent than real, the seeming difference in resistivity being due to the selectivity of the bacteriological sampling device.

In the use of ultraviolet radiation in occupied rooms, its physiological effects on the persons exposed must be considered. These effects include photosensitization, action on the skin and eye, on wound healing, and several others.[10, 128, 182, 158] In the operating room the skin and eyes of the personnel can be adequately protected, while still supplying radiant energy sufficient to kill infective aerosols.[95, 97, 99, 100, 102, 128] It appears established that ultraviolet light of low intensity stimulates wound healing,[153] although higher intensities are injurious in various ways.[182] Hence, in radiant sterilization of air in the operating room, it is essential that exposed wounds are not subjected to destructive intensities of light. Studies in this regard,[100, 102, 103, 105, 127, 128] show that air sterilization can be accomplished without producing any deleterious effects on the patient, such as adhesions, abnormal temperature, or slow wound healing.

The production and control of bactericidal radiation, the characteristics and standardization of lamps, and considerations in the design of ultraviolet sanitation equipment for air sterilization, have been discussed by several authors.[85, 46, 100, 128, 129, 187a, 169, 183, 224, 225]

The ultraviolet irradiation of air in operating rooms has been shown to bring about a great decrease in the number of living aerosols, and in the incidence of post-operative infections [74, 95, 97, 98, 99, 100, 101, 102, 105, 127, 128, 175] without evident harm to the personnel or patients. In one hospital,[100] for example, post-operative infections were reduced to from 1/20 to 1/100 of the previous number, largely by this means. Some opinion holds, however,[7, 8, 59, 215] that transmission of bacteria by other than air-borne means is chiefly responsible for post-operative infections, and that the benefits of bactericidal irradiation of air in this regard are over-rated.

Studies of the efficacy of ultraviolet light screens between cubicles, and of irradiation of the general air of wards and corridors in children's hospitals and nurseries, show that pathogenic aerosols can be greatly decreased by such methods. Concomitantly, there is a drop in the number of cross-infections and in the incidence of respiratory diseases.[14, 15, 23, 58, 90, 175, 176, 188, 246] Similarly, in a school,[239, 242, 248] the spread of several epidemic diseases was checked among highly susceptible groups of children by radiant disinfection of the air.

Recent quantitative studies [26, 28] indicate that sunlight and ordinary daylight probably play an important role in the suppression of streptococcal and other infections of the respiratory tract, through their potent action on infective aerosols. The region 3400-4400 Å is very intense in sunlight,[111] and is an important factor in the low survival of infectious agents in the outside air.

Within the last few years, ultraviolet light has been used in connection with the storage, processing, and packaging of various foods, in order, among other things, to prevent their infection by air-borne microorganisms.

For general articles on air-borne infection and its control by irradiation of air, see [182, 219, 223, 224, 225].

Germicidal aerosols. Although the idea of using germicidal mists and vapors for air sterilization dates back to Lister, until recently they have had little use, since most of the agents which have been previously employed are toxic or irritating to human beings. The literature has been reviewed.[10a] Experimental evidence on the disinfecting power of non-irritating aerosols is now available,[179, 207, 250] but little practical application has as yet been made of the data.

Among recently investigated compounds which appear to fulfill the desiderata of a germicidal aerosol, namely, good killing power for microorganisms, cheapness and availability, and lack of toxicity for human beings, are sodium hypochlorite, hexylresorcinol, propylene glycol, and triethylene glycol. Sodium hypochlorite has been

investigated by several workers.[11, 12, 21a, 61, 141, 142, 167, 203, 207] One gram of 1 per cent NaOCl atomized in as much as 40,000,000 cc of air will effect sterilization;[141, 142] 1 gram of 2 per cent NaOCl in glycerol-water to 6,000,000 cc of air was found highly effective by other investigators.[167] One very effective aerosol developed in England contains 10 per cent hexylresorcinol and 0.05 per cent alkyl sulfate, "Lorol," * in alkaline propylene glycol;[207] one gram of this in 4,000,000,000 cc of air is bactericidal.

Propylene glycol alone has been investigated as a germicidal aerosol rather thoroughly, both as to its bactericidal activity and its toxicity for human beings.[178, 179, 180, 181] In addition to its other desirable characteristics, this material is odorless and tasteless. Its toxicity for man is very low, but for microorganisms very high. One gram of glycol vapor in three to four million cc of air effects immediate killing of staphylococci, pneumococci and streptococci.[179] Rapid but not immediate sterilization occurs in considerably higher dilutions. The above concentration also protects mice from influenza virus.[107, 181] Other members of the aliphatic glycol series have recently been investigated as regards their bactericidal and viricidal action.[181a] Triethylene glycol, in particular, compares favorably with propylene glycol as a germicidal aerosol. The mode of action of gycol on bacteria has been discussed by [179, 180.]; glycol *aerosols* (mists) can be replaced by glycol *vapor* with equally effective results.

Hydrogen peroxide has been suggested as a germicidal aerosol,[1] but its use in this regard has not yet been thoroughly investigated. Ozone, in concentrations which can be breathed without irritation by persons for long periods of time, cannot protect against air-borne infection.[72a]

Methods for the production of aerosol mists, and detailed studies of particle size and evaporation may be found in [207].

Certain kinds of smokes have been proposed for air sterilization.[79, 206] Incense smoke and smoke from ignited cardboard soaked with potassium nitrate were found to be highly effective. It is reported that 1 gram of the chemical substance dispersed in smoke form in 500,000,000 cc of air causes destruction of 95 per cent of air suspended bacteria within 15 minutes.* According to some investigators [13] germicidal aerosols, and particularly smokes, show an increase in their lethal effects with increased humidity; other workers find the opposite.[243a] Apparently the effects of humidity have not yet been properly evaluated.

The practical applications of germicidal aerosols to air sterilization have as yet been few. It is stated [21a] that the British Ministry of Health proposes to issue sodium hypochlorite sprays for use in London air-raid shelters. Sodium hypochlorite has been used in factories and offices, and it is reported [141, 142] that there is a marked diminution in the air-borne bacteria and a decrease in the incidence of acute respiratory infections in the population exposed to the mist. Other investigators [58] report the successful use of an aerosol in controlling the spread of streptococcus infections in a hospital ward, after the usual measures had failed. In an investigation recently reported,[98a] a great reduction in the incidence of respiratory infections was found in a convalescent ward during a period of vaporization of propylene glycol in the ward.

From the preceding discussion it is evident, first, that the bacterial content of the air of certain occupied spaces can be greatly diminished by suitable control measures, and second, that this control results in a lower incidence of respiratory infections and in fewer post-operative wound infections. Since the bacteriological control of

* "Lorol" is the sulfuric acid ester of mixed fatty alcohols derived from coconut oil.
J. A.

† This raises the question as to the bactericidal effect in churches of incense and other "burnt offerings," and brings to mind the lines of William Cowper:

"God moves in a mysterious way
His wonders to perform."
J. A.

air decreases disease incidence, one must conclude that air is a vehicle of infection, although not necessarily the only means of spread.

Infective Aerosols and Disease

As yet it is not possible to evaluate fully the *relative* importance of strictly air-borne transmission and of direct contact (including droplets proper) as modes of spread of infection, particularly human respiratory infection. The hypothesis that respiratory infections are largely air-borne is newer, and is considered by some investigators, particularly Wells,[236, 239, 240] to be more satisfactory than the older theory that these diseases are spread by direct contact. Recent work offers much support to the former hypothesis, and this evidence for the actuality of air-borne infection in certain situations at least, cannot be ignored. On the other hand, much of this evidence, while supporting the air-borne hypothesis, does not necessarily invalidate transmission by direct contact.

The evidence has been adequately reviewed by several investigators,[25, 180, 236, 240] and the various aspects of the subject formed the basis for a recent symposium on aerobiology.[1a] As has been pointed out,[25] the evidence is based on three rather distinct lines of research: first, laboratory studies and surveys of the presence, distribution and survival of respiratory organisms in the air; second, the control of infections by measures designed to rid the environment of pathogenic aerosols; and third, investigations of the bacteriology and epidemiology of experimental air-borne infection, and of respiratory infections in hospitals and other institutions. The first two of these lines of evidence have for the most part been discussed above; the third, and certain aspects of the others will be considered below. As a background we shall first outline some of the interactions of the host and its invading microbe as they relate to infective aerosols.

Some interactions of microbe and host. Many factors are involved in the production of disease by infective aerosols, even after assuming that this is the only method of spread. The type of infective agent and its virulence, its concentration in the air and the amount inhaled, the fate of the organisms in the respiratory tract, and the resistance of the host are some of the chief considerations. Only a few outstanding points in this regard can here be touched upon.

It is well known that different types or strains of the same microbe, as well as different kinds of microbes, vary in their disease-producing ability or virulence. It is possible experimentally to change the virulence of microorganisms. As a practical matter virulence tends to increase with the passage of the organisms through the animal body; in epidemics increased virulence is noticed as the contagion spreads through the population. Certain aspects of the virulence of bacteria in air and dust have previously been referred to.

Although in practice it may be necessary to use the concentration of bacteria in the air as a measure of sanitary ventilation, this number obviously bears little relation to the number inhaled, unless one considers the duration of exposure, the fate of the microbes after inhalation, and the type of organism in relation to that part of the respiratory tract which it must reach in order to produce an infection. Very little evidence is available concerning the number of organisms which must be inhaled to produce infection. Undoubtedly this will vary greatly with the type of organism and its virulence, and with the resistance of the host. In certain diseases at least, the survival period, the type, and progress of the ensuing disease are functions of the dose of naturally inhaled microorganisms.[229] In the experimental production of air-borne tuberculosis [229, 241] it was found that as few as 20 tubercle bacilli would produce the disease; tuberculosis was produced sooner, and was of a different type when greater numbers of the organism were inhaled. Experimental air-borne streptococcus infection of mice required several hundred organisms per mouse to

produce 18 per cent mortality in a group, while over 10,000 organisms per mouse resulted in about 90 per cent mortality.[241]

The fate of microorganisms after inhalation is of evident significance as regards disease production. Suspended particles tend to be removed in the upper respiratory tract, whereas some diseases are produced only in the lower part of this system. The air passages are an important first line of defense [200, 201] in this regard; not only may a large proportion of suspended particles be filtered out in the nose,[185] but the nasal secretions are also definitely inhibitory to some microorganisms.[83, 200] Turbulence tends to throw inhaled particles out of suspension in the upper parts of the respiratory tract [218] where they may be removed by ciliary action.[201] Particles which reach the lungs tend to be removed or killed by phagocytes.[177]

The subject of host resistance or immunity to infection is broad and complex, but only one or two points will be mentioned here. Both natural immunity and acquired immunity are part of the picture. There is great variation in susceptibility to the same and to different infections among different persons. The same person also may vary greatly at different times in his resistance to a disease, depending upon physical condition and other factors. Hospitalized, weakened and debilitated persons are especially susceptible to infectious diseases, and children are in general more susceptible than adults. In this latter connection, it is worth noting that these are the groups on which most of the investigations of air-borne infection have been carried out.

Experimental air-borne infections. Air-borne infections have been produced in experimental animals under strictly controlled conditions which exclude contact infection, and which are quantitative as to dosage.[220, 241] These infections include influenza and pneumonia,[220, 228] tuberculosis,[229] poliomyelitis,[77] and streptococcus infections.[220, 241] In some of these cases it was demonstrated that the inhaled organisms were carried deeply into the lungs. Other significant observations as regards air-borne infection and its development were made in these investigations, but space does not permit their consideration here. The basic demonstration was that strictly air-borne infection is readily possible.

Infection of wounds. Another type of infection—that of the post-operative infection of wounds—apparently may be air-borne, although the introduction of the responsible organisms certainly is not limited to the aerial route. As mentioned previously, many persons are "carriers" of organisms pathogenic for other individuals, and the organisms may be expelled from the nose and throat even in ordinary respiratory activities. The number of "carriers" is particularly high in the winter, when the incidence of respiratory infections is greatest. The observed relation between such "carriers" among operating room personnel, and the incidence of post-operative wound infections,[59, 96, 104, 145, 146, 214, 215] seems not to be fortuitous. This belief is strengthened by the fact, discussed under control of infective aerosols, that such measures as masking of "carriers" or their exclusion from the operating room, and irradiation of the air, can be shown to result in a decrease both in the aerial contamination and in the number of wound infections.

Hospital infections. The spread of diptheritic, pneumococcic and streptococcic infections in isolation, children's and maternity hospitals are well-recognized problems.[25] The consensus is that while direct contact is undoubtedly a frequent mode of transmission in these cases, there is also evidence for air-borne spread. Concerning pneumonia, although there is evidence that pneumococci can survive in air and dust there is very little direct and positive proof of pneumococcal disease in man acquired through droplet or air-borne infection.[78]

Streptococcic infections offer perhaps the most evidence for the hypothesis of transmission by the air-borne route. Streptococci are of different types which can be identified serologically, and it is often possible to trace infections by identifying

the person from whom a particular type is derived. The most significant studies have been carried out in the scarlet fever wards of isolation hospitals, where so-called complications or cross-infections are a major problem. Various studies in this regard [5, 24, 25, 60] show that the majority of the complications are actually new infections caused by a strain of streptococcus serologically distinct from the patients' original one, and whose source can usually be traced to some other individual in the hospital. Up to 75 per cent of patients show the presence of new types in the throat after some time in the hospital,[25] although these new types do not always reproduce the disease. The bacteriological evidence for the air-borne transmission of these cross-infections is chiefly that ". . . streptococci are widespread in the ward air and the serological types identified were not confined to the neighborhood of patients infected with a particular type." [24] The possibility of direct or indirect contact infection was not ruled out in these cases; it was, in fact, often known to occur. The air-borne spread of streptococcic infections is strongly suspected in other hospital situations.[25, 122, 154] Although there have been no extensive studies on the transfer of infection from patient to patient by nurses, it is observed that patients nursed in isolated cubicles show an almost negligible incidence of reinfection as compared with those treated in multiple-bed wards.[25] In infants' and children's hospitals where contact among patients is ruled out, and transmission from attendants to children is believed to be small, the decrease in the incidence of various infections by cubicle nursing and by ultraviolet irradiation of the air is evidence for the air-borne spread of respiratory contagion.[15, 23, 58, 176, 188]

That the air-borne spread of diphtheria is bacteriologically possible is shown, for example, by the fact that the organism can be demonstrated in dust from the floor of wards.[255] Further evidence for aerial dissemination was found in the fact that a number of patients nursed in cubicles showed no signs of reinfection, which is not what would be expected if attendants play an important part in the transmission of this disease.

The spread of epidemics. The epidemiological evidence for the aerial spread of certain diseases has been reviewed by [240]. Studies of the epidemic spread of respiratory diseases, particularly virus infections such as influenza, measles and chickenpox, indicate that their widespread occurrence almost requires the hypothesis of air-borne transmission.[236] The velocity of spread, the rise and fall of the epidemics, and the epidemiological patterns manifested, are considered by some investigators [240] to be more accurately explained by the air-borne than the contact theory. In a recent four-year epidemiological study of radiant disinfection of air in day schools in relation to the control of epidemic contagion,[242] epidemic spreads of measles were checked in three irradiated primary schools during the severest epidemic ever experienced by the city of Philadelphia.[239] Evidence also was obtained for a lower incidence of chickenpox and mumps in irradiated rooms as compared with the unirradiated. One conclusion from these studies was that "Since the dynamic spread of epidemic contagion can be controlled by the radiant disinfection of air, it follows that air must be a vehicle of infection." [239]

Evidence for the production of plant diseases from air-borne fungus spores is given by [45, 123]. The aerial transmission of many diseases of plants is unquestioned, since direct contact generally is not a factor.

It is apparent from the brief discussion above that the air-borne transmission of infection, both respiratory and otherwise, is an actuality, and is more important than has hitherto been recognized. As regards human respiratory diseases, the importance of this mode of spread relative to other means is yet to be delineated, but the significant fact is that it does occur.

References

1. Abramson, H. A., *Science*, **96**, 238 (1942).
1a. "Aerobiology," Edited by F. R. Moulton, Washington, D. C., American Association for the Advancement of Science, Publication No. 17, 1942.
2. Alexander, J., *Science*, **83**, 230 (1936).
3. Alexander, J., "Colloid Chemistry. Principles and Applications," 4th ed., New York, D. van Nostrand, 1937.
4. Allison, V. D., Third International Congress for Microbiology, Report of Proceedings, p. 266, New York, 1940.
5. Allison, V. D., and Brown, W. A., *J. Hyg.*, **37**, 153 (1937).
6. Andrewes, C. H., *et al.*, *Lancet*, **2**, 770 (1940).
7. Anonymous, *Brit. Med. J.*, **2**, 733 (1939).
8. Anonymous, *Brit. Med. J.*, **2**, 707 (1940).
9. Anonymous, *Brewing Trade Review*, **54**, 405 (1941).
10. Arnold, L., *Arch. Surg.*, **37**, 1008 (1938).
10a. Baker, A. H., *Chem. Products*, **4**, 25 (1941).
11. Baker, A. H., and Finn, S. R., *Nature*, **146**, 747 (1940).
12. Baker, A. H., Finn, S. R., and Twort, C. C., *J. Hyg.*, **40**, 560 (1940).
13. Baker, A. H., and Twort, C. C., *J. Hyg.*, **41**, 117 (1941).
14. Barenberg, L. H., Greene, D., and Greenspan, L., *Am. J. Dis. Children*, **59**, 1219 (1940).
15. Barenberg, L. H., Greene, D., Greenspan, L., and Greenberg, B., in "Aerobiology," p. 233, Washington, D. C., American Association for the Advancement of Science, 1942.
16. Bernstein, T. B., and Feinberg, S. M., *J. Allergy*, **13**, 231 (1942).
17. Blatt, M. L., and Dale, M. L., *Surg., Gyncol. Obstet.*, **57**, 363 (1933).
18. Bloomfield, A. L., and Felty, A. R., *Bull. Johns Hopkins Hosp.*, **35**, 115 (1924).
19. Blum, H. F., "Photodynamic Action and Diseases Caused by Light," New York, Reinhold Publishing Corp., 1941.
20. Boelter, L. M. K., *Heating, Piping and Air Conditioning*, **11**, 639 (1939).
21. Bourdillon, R. B., and Lidwell, O. M., *Lancet*, **2**, 365 (1941).
21a. Bourdillon, R. B., Lidwell, O. M., and Lovelock, J. E., *Brit. Med. J.*, **1**, 42 (1942).
22. Bourdillon, R. B., Lidwell, O. M., and Thomas, J. C., *J. Hyg.*, **41**, 197 (1941).
23. Brooks, G. L., Wilson, V., and Blackfan, K. D., in "Aerobiology," p. 228, Washington, D. C., American Association for the Advancement of Science, 1942.
24. Brown, W. A., and Allison, V. D., *J. Hyg.*, **37**, 1 (1937).
25. Buchbinder, L., *J. Am. Med. Assoc.*, **118**, 718 (1942).
26. Buchbinder, L., in "Aerobiology," p. 267, Washington, D. C., American Association for the Advancement of Science, 1942.
27. Buchbinder, L., and Phelps, E. B., *J. Bact.*, **42**, 345 (1941).
28. Buchbinder, L., Solowey, M., and Phelps, E. B., *J. Bact.*, **42**, 353 (1941).
29. Buchbinder, L., and Solotorovsky, M., *J. Bact.*, **42**, 635 (1941).
30. Buchbinder, L., Solotorovsky, M., and Solowey, M., *Proc. Soc. Expl. Biol. Med.*, **38**, 570 (1938).
31. Buchbinder, L., Solowey, M., and Solotorvsky, M., *Am. J. Pub. Health*, **28**, 61 (1938).
32. Buchbinder, L., Solowey, M., and Solotorovsky, M., *J. Bact.*, **42**, 615 (1941).
33. Buchbinder, L., Solotorovsky, M., Solowey, M., and Ruhl-Koupal, J., *J. Bact.*, **42**, 631 (1941).
34. Burky, E. L., and Smillie, W. G., *J. Expl. Med.*, **50**, 643 (1929).
35. Buttolph, L. J., in "Aerobiology," p. 171, Washington, D. C., American Association for the Advancement of Science, 1942.
36. Castleman, R. A., Jr., *Bur. Standards J. Research*, **6**, 369 (1931).
37. Chapin, C. V., "The Sources and Modes of Infection," New York, John Wiley & Sons, 1912.
38. Chapin, C. V., *J. Am. Med. Assoc.*, **62**, 423 (1914).
39. Chapple, C. C., in "Aerobiology," p. 251, Washington, D. C., American Association for the Advancement of Science, 1942.
40. Chapple, C. C., and Kenny, A., *Am. J. Dis. Children*, **57**, 1058 (1939).
41. Chaussé, P., *Comp. rend.*, **156**, 638 (1913).
42. Chaussé, P., *Ann. inst. Pasteur*, **30**, 613 (1916).
43. Chaussé, P., and Magne, H., *Arch. de médecine expérimentale et d'anatomie pathologique*, **27**, 213 (1916).
44. Chope, H. D., and Smillie, W. G., in "The Environment and its Effect upon Man," p. 32, Boston, Harvard School of Public Health, 1937.
45. Christensen, J. J., in "Aerobiology," p. 78, Washington, D. C., American Association for the Advancement of Science, 1942.
46. Coblentz, W. W., in "Aerobiology," p. 138, Washington, D. C., American Association for the Advancement of Science, 1942.
47. Colebrook, L., *Brit. Med. J.*, **2**, 723 (1933).
48. Colvin, M. G., *Am. J. Hyg.*, **15**, 247 (1932).
49. Committee on Apparatus in Aerobiology, National Research Council, *Phytopathology*, **31**, 201 (1941).
49a. Committee on Pollen Survey of the U. S. *J. Allergy*, **13**, 516 (1942).
50. Cradle Society, "Aseptic Nursery Technique as used at the Cradle," 2nd ed., Evanston, Ill., Cradle Society, 1941.
51. Cruickshank, R., *J. Path. & Bact.*, **41**, 367 (1935).
52. Cruickshank, R., and Godber, G. E., *Lancet*, **1**, 741 (1939).
53. Cruikshank, R., and Muir, C., *Lancet*, **1**, 1155 (1940).
54. Curran, H. R., and Evans, F. R., *J. Bact.*, **36**, 455 (1938).
55. Dahl, A. O., and Ellis, R. V., *Pub. Health Rpts.*, **57**, 369 (1942).
56. DallaValle, J. M., and Hollaender, A., Reprint No. 2063, *Pub. Health Repts.*, **54**, 695 (1939).
57. DallaValle, J. M., and Holaender, A., Reprint No. 2055, *Pub. Health Repts.*, **55**, 1268 (1940).
58. del Mundo, Fé, and McKhann, C. F., *Am. J. Dis. Children*, **61**, 213 (1941).
59. Devenish, E. A., and Miles, A. A., *Lancet*, **1**, 1088 (1939).
60. de Waal, H. L., *J. Hyg.*, **40**, 172 (1940).
61. Douglas, S. R., Hill, L., and Smith, W., *J. Ind. Hyg.*, **10**, 219 (1928).
62. Drinker, P., and Hatch, T., "Industrial Dust," New York, McGraw-Hill Book Co., 1936.
63. Drinker, P., and Wells, W. F., *Heating, Piping and Air Conditioning*, **6**, 408 (1934).
64. Dudley, S. F., *Lancet*, **2**, 849 (1928).
65. Duggar, B. M., in "Biological Effects of Radiation," **Vol. II**, Ch. 36, New York, McGraw-Hill Book Co., 1936.
66. Dunham, G. C., "Military Preventive Medicine," 3rd ed., Carlisle Barracks, Pa., Medical Field Service School, 1938.
67. Dunkin, G. W., and Laidlaw, P. P., *J. Comp. Path. & Therap.*, **39**, 201, 213 (1926).

68. Durham, O. C., *J. Allergy*, **6**, 128 (1935).
69. Durham, O. C., *J. Aviation Med.*, **12**, 153 (1941).
70. Durham, O. C., in "Aerobiology," p. 32, Washington, D. C., American Association for the Advancement of Science, 1942.
71. Eagleton, A. J., *Am. J. Hyg.*, **18**, 264 (1919).
72. Edward, D. G., *Lancet*, **2**, 664 (1941).
72a. Elford, W. J. and van den Ende, J., *J. Hyg.*, **42**, 240 (1942).
73. Elliott, S. D., *Lancet*, **2**, 514 (1941).
74. Ellis, C., Wells, A. A., and Heyroth, F. F., "The Chemical Action of Ultraviolet Rays," New York, Reinhold Publishing Corp., 1941.
75. Erikson, C. A., and Sauer, L. W., *Modern Hospital*, **55**, 54 (1940).
76. Erikson, M. H., *J. Genetic Psychol.*, **56**, 455 (1940).
77. Faber, H. K., and Silverberg, R. J., *Science*, **94**, 566 (1941).
78. Finland, M., in "Aerobiology," p. 212, Washington, D. C., American Association for the Advancement of Science, 1942.
79. Finn, S. R., and Powell, E. O., *J. Hyg.*, **41**, 473 (1941).
80. Flügge, C., *Z. Hyg. Infectionskrank.*, **25**, 179 (1897).
81. Flugge, C., *Deut. med. Wchnschr.*, **23**, 665 (1897).
82. Flügge, C., *Z. Hyg. Infectionskrank.*, **30**, 107 (1899).
83. Francis, T., Jr., in "Problems and Trends in Virus Research," p. 41, University of Pennsylvania Bicentennial Conference, Philadelphia, U. of P. Press, 1941.
84. Gates, F. L., *J. Gen. Physiol.*, **13**, 231, 249 (1929); **14**, 31 (1930).
85. Gauthier, P. P., *Canad. Med. Assn. J.*, **37**, 250 (1937).
86. Gibbs, W. E., in "Colloid Chemistry. Theoretical and Applied," Edited by Jerome Alexander, **Vol. 1**, p. 293, New York, Chemical Catalog Co., Inc. (Reinhold Publishing Corp.), 1926.
87. Glick, P. A., in "Aerobiology," p. 88, Washington, D. C., American Association for the Advancement of Science, 1942.
88. Glover, J. A., *Proc. Roy. Soc. Med. (Epid. Sec.)*, **21**, 95 (1928).
89. Gordon, M. H., Local Government Board (Great Britain), Suppl. Cont. Rept. Med. Officer Health for 1902-03, 32nd Annual Rept., p. 421, 1904.
90. Greene, D., Barenberg, L. H., and Greenberg, B., *Am. J. Dis. Children*, **61**, 273 (1941).
91. Hamilton, A., *J. Am. Med. Assoc.*, **44**, 1108 (1905).
92. Hare, R., *Canad. Pub. Health J.*, **31**, 539 (1940).
93. Hare, R., *Lancet*, **1**, 85 (1941).
93a. Harris, T. N. and Stokes, J., Jr., *Am. J. Med. Sci.*, **204**, 430 (1942).
94. Harries, E. H. R., *Lancet*, **1**, 491 (1924).
95. Hart, D., *J. Thoracic Surg.*, **6**, 45 (1936).
96. Hart, D., *Arch. Surg.*, **34**, 874 (1937).
97. Hart, D., *J. Thoracic Surg.*, **7**, 525 (1938).
98. Hart, D., *Arch. Surg.*, **37**, 521 (1938).
99. Hart, D., *Arch. Surg.*, **41**, 334 (1940).
100. Hart, D., in "Aerobiology," p. 186, Washington, D. C., American Association for the Advancement of Science, 1942.
101. Hart, D., Devine, J. W., and Martin, D. W., *Arch. Surg.*, **38**, 806 (1939).
102. Hart, D., and Gardner, C. E., *Trans. Southern Surg. Assoc.*, **49**, 377 (1937).
103. Hart, D., and Sanger, P. W., *Arch. Surg.*, **38**, 797 (1939).
104. Hart, D., and Schiebel, H. M., *Arch. Surg.*, **38**, 788 (1939).
105. Hart, D., and Upchurch, S. E., *Ann. Surg.*, **110**, 291 (1939).
106. Hatch, T. F., in "Aerobiology," p. 102, Washington, D. C., American Association for the Advancement of Science, 1942.
106a. Hawes, R. C., Small, W. S., and Miller, H., *J. Allergy*, **13**, 474 (1942).
107. Henle, W., and Zellat, J., *Proc. Soc. Expl. Biol. Med.*, **48**, 544 (1941).
108. Heymann, B., *Z. Hyg. Infektionskrank.*, **30**, 139 (1899).
109. Hippke, E., *Z. Hyg. Infektionskrank.*, **93**, 122 (1921).
110. Hollaender, A., *Ind. Med.*, **2**, 60 (1941).
111. Hollaender, A., in "Aerobiology," p. 156, Washington, D. C., American Association for the Advancement of Science, 1942.
112. Hollaender, A., and DallaValle, J. M., *Pub. Health Repts.*, **54**, 574 (1939).
113. Horwood, M. P., *J. Bact.*, **21**, 14 (1931).
114. Horwood, M. P., Gould, B. S., and Shwachman, H., *J. Bact.*, **27**, 66 (1934).
115. Huddleson, F., and Hull, T. G., *Am. J. Pub. Health*, **10**, 583 (1920).
116. Jacobson, L. O., and Dick, G. F., *J. Am. Med. Assoc.*, **117**, 2222 (1941).
117. Jennison, M. W., *Scient. Monthly*, **52**, 24 (1941).
118. Jennison, M. W., in "Aerobiology," p. 106, Washington, D. C., American Association for the Advancement of Science, 1942.
119. Jennison, M. W., and Edgerton, H. E., *Proc. Soc. Expl. Biol. Med.*, **43**, 455 (1940).
120. Jennison, M. W., Tapley, G. O., and Sullivan, J. C., *J. Bact.*, **43**, 100 (1942).
121. Jennison, M. W., and Turner, C. E., *Trained Nurse and Hospital Review*, **106**, 186 (1941).
122. Keevil, N. L., and Camps, F. E., *Lancet*, **2**, 207 (1937).
123. Keitt, G. W., in "Aerobiology," p. 69, Washington, D. C., American Association for the Advancement of Science, 1942.
124. Kellogg, W. H., and McMillen, G., *Am. J. Pub. Health*, **10**, 34 (1920).
125. Koeniger, H., *Z. Hyg. Infektionskrank.*, **34**, 119 (1900).
126. Koller, L. R., *J. Appl. Physics*, **10**, 624 (1939).
127. Kraissl, C. J., Cimiotti, J. G., and Meleney, F. L., *Ann. Surg.*, **111**, 161 (1940).
128. Kraissl, C. J., Wilson, D. A. W., and Wilson, M. H., in "Aerobiology," p. 197, Washington, D. C., American Association for the Advancement of Science, 1942.
129. Kruetzkamp, J., in "Aerobiology," p. 184, Washington, D. C., American Association for the Advancement of Science, 1942.
130. Lange, B., and Keschischian, K. H., *Z. Hyg. Infektionskrank.*, **104**, 256 (1925).
131. Laschtschenko, *Z. Hyg. Infektionskrank.*, **30**, 125 (1899).
132. Laurens, H., "The Physiological Effects of Radiant Energy," New York, Chemical Catalog Co., Inc. (Reinhold Publishing Corp.), 1933.
133. Laurens, H., in "Aerobiology," p. 142, Washington, D. C., American Association for the Advancement of Science, 1942.
134. Leete, H. M., *Lancet*, **1**, 392 (1919).
135. Leitner, J., *Z. Hyg. Infektionskrank.*, **109**, 102 (1928).
136. Long, A. P., McKhann, C. F., and Cheney, L. L., *Am. J. Dis. Children*, **57**, 1363 (1939).
137. Long, A. P., McKhann, C. F., and Cheney, L. L., *Am. J. Dis. Children*, **60**, 222 (1940).

137a. Luckiesh, M., and Hollaender, L. L., *Gen. Elec. Rev.*, **45**, 223 (1942).
138. MacDonald, K., *Am. J. Hyg.*, **31 (B)**, 74 (1940).
139. MacDonald, K., *Am. J. Hyg.*, **31 (B)**, 88 (1940).
140. Major, C. B., and Wilder, T. S., *Hospitals*, **11**, 87 (1937).
141. Mastermann, A. T., *J. Ind. Hyg. & Toxicol.*, **20**, 278 (1938).
142. Masterman, A. T., *J. Hyg.*, **41**, 44 (1941).
143. Matthews, A. P., "Physiological Chemistry," 4th ed., New York, William Wood, 1925.
144. Meleney, F. L., *J. Am. Med. Assoc.*, **88**, 1392 (1927).
145. Meleney, F. L., *Surg., Gyncol. Obstet.*, **60**, 264 (1935).
146. Meleney, F. L., and Stevens, F. A., *Surg., Gyncol. Obstet.*, **43**, 338 (1926).
147. Milam, D. F., and Smillie, W. G., *J. Expl. Med.*, **53**, 733 (1931).
148. Morrow, M. B., Lowe, E. P., and Prince, H. E., *J. Allergy*, **13**, 215 (1942).
148a. Moulton, S., Puck, T. T., and Lemon, H. M., *Science*, **97**, 51 (1943).
149. McConnell, W. J., and Thomas, B. G. H., Reprint No. 1045, *Pub. Health Repts.*, **40**, 2167 (1925).
150. McCoy, G. W., *J. Infectious Diseases*, **55**, 156 (1934).
151. McKhann, C. F., Steeger, A., and Long, A. P., *Am. J. Dis. Children*, **55**, 579 (1938).
152. Noble, W. C., Jr., Fisher, E. A., and Brainard, D. H., *J. Preven. Med.*, **2**, 105 (1928).
153. Nutini, L. G., in "Aerobiology," p. 193, Washington, D. C., American Association for the Advancement of Science, 1942.
154. Okell, C. C., and Elliott, S. D., *Lancet*, **2**, 836 (1936).
155. Olsen, O., and Strauss, W., *Z. Hyg. Infektionskrank.*, **105**, 552 (1926).
156. Overholt, R. H., and Betts, R. H., *J. Thoracic Surg.*, **9**, 520 (1940).
157. Paine, C. G., *Brit. Med. J.*, **1**, 243 (1935).
158. Pasteur, L., *Ann. Sc. Naturelles, Partie Zoologique*, **16**, 5 (1861).
159. Phelps, E. B., in "Aerobiology," p. 133, Washington, D. C., American Association for the Advancement of Science, 1942.
160. Phelps, E. B., and Buchbinder, L., *J. Bact.*, **42**, 321 (1941).
161. Pincus, S., and Stern, A. C., *Am. J. Pub. Health*, **27**, 321 (1937).
162. Pressman, R., *Am. Review Tuberculosis*, **35**, 815 (1937).
163. Proctor, B. E., *Proc. Am. Acad. Arts Sci.*, **69**, 315 (1934).
164. Proctor, B. E., *J. Bact.*, **30**, 363 (1935).
165. Proctor, B. E., and Parker, B. W., *J. Bact.*, **36**, 175 (1938).
166. Proctor, B. E., and Parker, B. W., in "Aerobiology," p. 48, Washington, D. C., American Association for the Advancement of Science, 1942.
167. Pulvertaft, R. J. V., and Walker, J. W., *J. Hyg.*, **39**, 696 (1939).
168. Reece, G. M., Massachusetts Institute of Technology, Master's Thesis, 1939.
169. Rentschler, H. C., in "Aerobiology," p. 166, Washington, D. C., American Association for the Advancement of Science, 1942.
170. Rentschler, H. C., and Nagy, R., *Heating, Piping, and Air Conditioning*, **12**, 127 (1940).
171. Rentschler, H. C., and Nagy, R., *J. Bact.*, **44**, 85 (1942).
172. Rentschler, H. C., Nagy, R., and Mouromseff, G., *J. Bact.*, **41**, 745 (1941)
173. Reyniers, J. A., in "Aerobiology," p. 254, Washington, D. C., American Association for the Advancement of Science, 1942.
174. Reyniers, J. A., and Trexler, P. C., in "Aerobiology," p. 260, Washington, D. C., American Association for the Advancement of Science, 1942.
174a. Reyniers, J. A., in "Micrurgical and Germ-Free Techniques," (J. A. Reyniers, Editor), pp. 95, 205, Springfield, Ill., C. C. Thomas, 1943.
175. Robertson, E. C., and Doyle, M. E., *Ann. Surg.*, **111**, 491 (1940).
176. Robertson, E. C., and Doyle, M. E., in "Aerobiology," p. 223, Washington, D. C., American Association for the Advancement of Science, 1942.
177. Robertson, O. H., *Physiol. Rev.*, **21**, 112 (1941).
178. Robertson, O. H., Bigg, E., Miller, B. F., and Baker, Z., *Science*, **93**, 213 (1941).
179. Robertson, O. H., Bigg, E., Puck, T. T., Miller, B. F., and Baker, Z., in "Aerobiology," p. 271, Washington, D. C., American Association for the Advancement of Science, 1942.
180. Robertson, O. H., Bigg, E., Puck, T. T., Miller, B. F., and Appell, E. A., *J. Expl. Med.*, **75**, 593 (1942).
181. Robertson, O. H., Loosli, C. G., Puck, T. T., Bigg, E., and Miller, B. F., *Science*, **94**, 612 (1941).
181a. Robertson, O. H., Puck, T. T., Lemon, H. F., and Loosli, C. G., *Science*, **97**, 142 (1943).
182. Robinson, F. W., *Ind. Eng. Chem., (Ind. Ed.)*, **31**, 23 (1939).
183. Robinson, F. W., in "Aerobiology," p. 181, Washington, D. C., American Association for the Advancement of Science, 1942.
184. Rochaix, A., *Le Movement Sanitaire*, **7**, 524 (1931).
185. Rooks, R., *Am. J. Hyg.*, **30**, 7 (1939).
186. Rooks, R., *J. Iowa State Med. Soc.*, **31**, 100 (1941).
187. Rooks, R., Cralley, L. J., and Barnes, M. E., *Pub. Health Repts.*, **56**, 1411 (1941).
188. Rosenstern, I., in "Aerobiology," p. 242, Washington, D. C., American Association for the Advancement of Science, 1942.
189. Rundle, C., *Lancet*, **2**, 376 (1922).
190. Sharp, D. G., *J. Bact.*, **39**, 535 (1940).
191. Sim, R. P., and Flinn, F. B., *Am. J. Hyg.*, **30 (B)**, 135 (1939).
192. Solowey, M., Solotorovsky, M., and Buchbinder, L., *J. Bact.*, **43**, 545 (1942).
193. Stakman, E. C., in "Aerobiology," p. 1, Washington, D. C., American Association for the Advancement of Science, 1942.
194. Stimson, A. M., *Pub. Health Repts., Suppl. No. 141*, p. 48, 1938.
195. Strauss, W., *Z. Hyg. Infektionskrank.*, **96**, 27 (1922).
196. Strauss, W., *Z. Hyg. Infektionskrank.*, **105**, 416 (1925).
197. "Tabulae Biologicae," (W. Junk, Editor), **Bd. 2.**, W. Junk, Berlin, 1925.
198. Teague, O., *J. Path. and Bact.*, **12**, 398 (1913).
198a. Thomas, J. C., *Lancet*, **2**, 123 (1941).
199. Thomas, J. C., and van den Ende, *Brit. Med. J.*, **1**, 953 (1941).
200. Thomson, D. & R., "The Common Cold. With special reference to the part played by streptococci, pneumococci, and other organisms," **vol. 8**, Annals of the Pickett-Thomson Research Laboratory, Baltimore, Williams & Wilkins, 1932.
201. Thomson, St. Clair, *J. Laryngology and Otolaryngology*, **51**, 1 (1936).
202. Tolman, R. C., Guernsey, E. W., Charleston, V. D., and Dougherty, R. H., *J. Infectious Diseases*, **24**, 637 (1919).
203. Trillat, A., *Ann. Hyg. Publ. Ind. Sociale*, **119**, 64 (1938).
204. Trillat, A., and Mallein, *Comp. rend.*, **170**, 1291 (1920).
205. Turner, C. E., Jennison, M. W., and Edgerton, H. E., *Am. J. Pub. Health*, **31**, 319 (1941).

206. Twort, C. C., and Baker, A. H., *Lancet*, **2**, 587 (1940).
207. Twort, C. C., Baker, A. H., Finn, S. R., and Powell, E. O., *J. Hyg.*, **40**, 253 (1940).
208. Tyndall, J., "Essays on the Floating-matter of the Air," 2nd ed., London, Longmans, Green & Co., 1883.
209. van den Ende, M., and Andrewes, C. H., in "Aerobiology," p. 281, Washington, D. C., American Association for the Advancement of Science, 1942.
210. van den Ende, M., Edward, D. G. F., and Lush, D., *Lancet*, **1**, 716 (1941).
211. van den Ende, M., Lush, D., and Edward, D. G. F., *Lancet*, **2**, 133 (1940).
212. van den Ende, M., and Spooner, E. T. C., *Lancet*, **1**, 751 (1941).
213. von Angerer, C., *Arch. Hyg.*, **89**, 262 (1920).
214. Walker, I. J., *Surg., Gyncol. Obstet.*, **50**, 266 (1930).
215. Walters, W., and McGath, T. B., *Ann. Surg.*, **112**, 271 (1940).
216. Wells, W. F., *Am. J. Pub. Health*, **23**, 58 (1933).
217. Wells, W. F., *Am. J. Hyg.*, **20**, 611 (1934).
218. Wells, W. F., *J. Ind. Hyg & Toxicol.*, **17**, 253 (1935).
219. Wells, W. F., *Heating and Ventilating*, **36**, 26 (1939).
220. Wells, W. F., *Science*, **91**, 172 (1940).
221. Wells, W. F., *J. Franklin Inst.*, **229**, 347 (1940).
222. Wells, W. F., *J. Bact.*, **39**, 18 (1940).
223. Wells, W. F., *Ind. Med.*, **10**, 281 (1941).
224. Wells, W. F., *Heating, Piping and Air Conditioning*, **2**, 143 (1942).
225. Wells, W. F., *Archiv. Phys. Therapy*, **23**, 143 (1942).
226. Wells, W. F., and Brown, H. W., *Am. J. Hyg.*, **24**, 407 (1936).
227. Wells, W. F., and Fair, G. M., *Science*, **82**, 280 (1935).
228. Wells, W. F., and Henle, W., *Proc. Soc. Expl. Biol. Med.*, **48**, 298 (1941).
229. Wells, W. F., and Lurie, M. B., *Am. J. Hyg.*, **34 (B)**, 21 (1941).
230. Wells, W. F., Phelps, E. B., Robertson, E. C., and Winslow, C.- E. A., American Public Health Association Year Book, p. 99, 1939-40, *Suppl. to Am. J. Pub. Health*, **30(2)**, (1940).
231. Wells, W. F., Phelps, E. B., Robertson, E. C., and Winslow, C.- E. A., American Public Health Association Year Book, p. 129, 1940-41, *Suppl. to Am. J. Pub. Health*, **31(3)**, (1941).
232. Wells, W. F., Phelps, E. B., and Winslow, C.- E. A., American Public Health Association, Ninth Annual Year Book, p. 69, 1938-39, *Suppl. to Am. J. Pub. Health*, **29(2)**, (1939).
233. Wells, W. F., Phelps, E. B., Robertson, E. C., and Winslow, C.- E. A., American Public Health Association Year Book, p. 137, 1941-42, *Suppl. to Am. J. Pub. Health*, **32(3)**, (1942).
234. Wells, W. F., and Riley, E. C., *J. Ind. Hyg. & Toxicol.*, **19**, 513 (1937).
235. Wells, W. F., and Stone, W. R., *Am. J. Hyg.*, **20**, 619 (1934).
236. Wells, W. F., and Wells, M. W., *J. Am. Med. Assoc.*, **107**, 1698, 1805 (1936).
237. Wells, W. F., and Wells, M. W., *Am. J. Pub. Health*, **28**, 343 (1938).
238. Wells, W. F., and Wells, D., *J. Bact.*, **39**, 755 (1940).
239. Wells, W. F., and Wells, M. W., in "Aerobiology," p. 99, Washington, D. C., American Association for the Advancement of Science, 1942.
240. Wells, W. F., Wells, M. W., and Mudd, S., *Am. J. Pub. Health*, **29**, 863 (1939).
241. Wells, W. F., Wells, M. W., Mudd, S., Lurie, M. B., and Henle, W., in "Aerobiology," p. 284, Washington, D. C., American Association for the Advancement of Science, 1942.
242. Wells, W. F., Wells, M. W., and Wilder, T. S., *Am. J Hyg.*, **35**, 97 (1942).
243. Wells, W. F., Wells, M. W., and Wilder, T. S., in "Aerobiology," p. 206, Washington, D. C., American Association for the Advancement of Science, 1942.
243a. Wells, W. F., and Zappasodi, P., *Science*, **96**, 277 (1942).
244. Weyrauch, F., and Rzymkowski, J., *Z. Hyg. Infektionskrank.*, **120**, 444 (1938).
245. Wheeler, S. M., Foley, G. E., and Jones, T. D., *Science*, **94**, 445 (1941).
246. Wheeler, S. M., and Jones, T. D., in "Aerobiology," p. 237, Washington, D. C., American Association for the Advancement of Science, 1942.
247. Whisler, B., *Iowa State College J. Science*, **14**, 215 (1940).
248. White, E., *Lancet*, **1**, 941 (1936).
249. Whytlaw-Gray, R., and Patterson, H. S., "Smoke: A Study of Aerial Disperse Systems," London, Edward Arnold and Co., 1932.
250. Williamson, A. E., and Gotaas, H. B., *Ind. Med.* **11**. (*Ind. Hyg. Sect.* **3**), 40 (1942).
251. Winslow, C.- E. A., and Browne, W. W., *Monthly Weather Rev.*, **42**, 452 (1914).
252. Winslow, C.- E. A., and Robinson, E. A., *J. Infectious Diseases*, **7**, 17 (1910).
253. Wodehouse, R. P., "Pollen Grains," New York, McGraw-Hill Book Co., 1935.
254. Wodehouse, R. P., in "Aerobiology," p. 8, Washington, D. C., American Association for the Advancement of Science, 1942.
255. Wright, H. D., Shone, H. R., and Tucker, J. R., *J. Path. & Bact.*, **52**, 111 (1941).
256. Yaglou, C. P., and Wilson, V., in "Aerobiology," p. 129, Washington, D. C., American Association for the Advancement of Science, 1942.
257. Yang, H. S., *Am. J. Dis Children*, **61**, 262 (1941).
258. Ziesché, H., *Z. Hyg. Infektionskrank.*, **57**, 50 (1907).
259. ZoBell, C. E., in "Aerobiology," p. 55, Washington, D. C., American Association for the Advancement of Science, 1942.

Physical Chemistry of Lipides *

MONA SPIEGEL-ADOLF, M.D.

Professor of Colloid Chemistry, Temple University, School of Medicine, Philadelphia, Pa.

Definition and Classification.

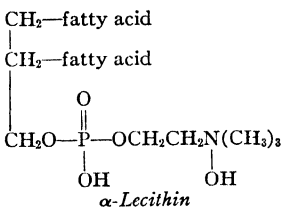
Peters and Van Slyke,[77] following the definition of Maclean,[68] describe lipides or lipins as "substances of a fat-like nature, yielding on hydrolysis fatty acids or derivatives of fatty acids, and containing in the molecule either nitrogen or nitrogen and phosphorus." The lipides can be divided into two groups: (1) phosphatides, which include lecithin, cephalin and sphingomyelin, in the first two of which N : P = 1, in the latter N : P = 2; and (2) cerebrosides, which do not contain phosphorus (but are distinguished by a carbohydrate group) include among others phrenosin and kerasin.

The different solubilities of these substances in organic solvents is summarized in Table 1 (Bull [12]):

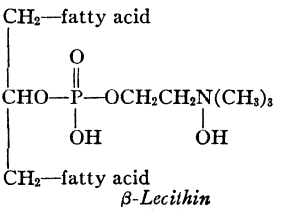
Table 1. Solubilities of lipides in organic solvents

	Ether	Alcohol	Acetone
Lecithin	S	S	X
Cephalin	S	X	X
Sphingomyelin	X	H	X
Cerebroside	X	H	H

Structural Formulas. Most of the biological and physicochemical work done on lipides has been concentrated on the phosphatides, especially on lecithin and cephalin. General formulas of these substances follow:

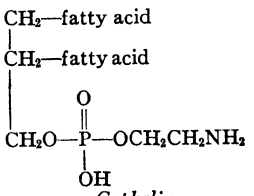


α-Lecithin
optically active

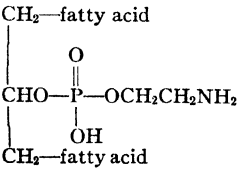


β-Lecithin
optically inactive

Until recently the following formulas were generally accepted for cephalin:



α-Cephalin



β-Cephalin

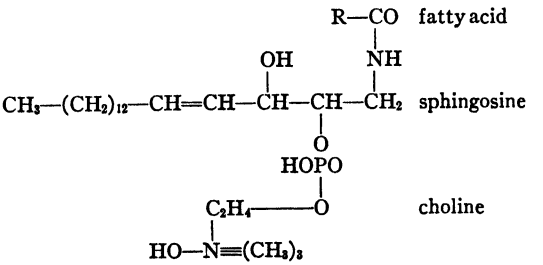
* For complete treatise on this field see Bloor, "Biochemistry of the Fatty Acids and their Compounds, the Lipids," New York, Reinhold Pub. Corp., 1943.

In 1941 Folch and Schneider [85] showed that over half of the nitrogen of cephalin is not in ethanolamine but in a hydroxyamino acid which Folch [84] later found to be serine. Chargaff and co-workers,[22] using isotope dilution methods and direct isolation,[21] were able to confirm these results. Folch and Schneider [85] found that intact cephalin reacts with nitrous acid and with ninhydrin in a manner which indicates the existence of both free NH_2 and COOH groups. They assume therefore "that the hydroxyamino acid in cephalin is attached to the rest of the phosphatide molecule by ester linkage through the hydroxyl group."

Folch [86] claims that cephalin is not a single substance but consists of a mixture of inositol, phosphatidyl-serine and of a more unsaturated compound, which does not contain any amino acids. Chargaff, Ziff and Rittenberg,[23] using nitrogen isotopes, found that phosphatides from beef brain and lung contain considerable amount of amino acid nitrogen. There is much less in liver and heart, and egg-yolk phosphatides are free of amino acids. Storing of phosphatides seems to affect the content of amino acid nitrogen. It seems interesting in this respect that, according to Chaikoff and co-workers,[39] using radioactive phosphorus, different tissues (liver, brain, excised nerve) are able to form phospholipides *in vitro*.

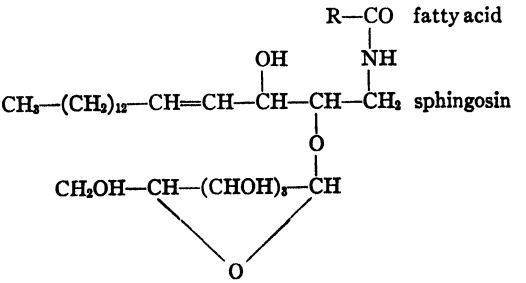
According to Klenk and Schuwirth [55] who based these statements on recent findings of Yokoyama and Suzuki,[110] lecithins and cephalins occurring in nature are not chemical individuals but mixtures of various substances. They differ from each other partly in their fatty acids and partly in the glycerin phosphoric acids (α and β forms).

Levene [68] has given a tentative formula for sphingomyelin



A plurality of sphingomyelins is assumed.[14, 115]

According to Thierfelder and Klenk,[115] there are four chemically well-defined cerebrosides: cerebron (phrenosin), kerasin, nervon and oxynervon, the structure of which correspond to the general formula



At least four different fatty acids have been isolated, which according to more recent reports (Chibnall, Piper and Williams [24]) are themselves only mixtures. Still more recently Klenk [54] has described the isolation of crystallizing neuroaminic acid from the cerebroside fraction of brain lipides. The nitrogen is present as amino acid nitrogen only.

Klenk [58] calls attention to the fact that cerebrosides and sphingomyelin contain fatty acids with 24 carbon atoms, whereas lecithin and cephalin are characterized by fatty acids with 20 and 22 carbon atoms. He surmises that warm-blooded animals can derive the longer fatty-acid chains from sugar and that the shorter ones result through β-oxidation of the former.

Reaction of Phosphatides with Water

The biological importance of the *lipides* seems to consist in the fact that, unlike the fats, *they are wetted by water*. When the phospholipides and cerebrosides in a substance first come into contact with water, part of the water enters the lipide mass, which starts to swell in an irregular manner. These shapes have been called "myeline forms" because of their similarity to medullated nerve fibers. Leathes [61] has explained this behavior by the formation of monomolecular films (see page 1130). When dissolved in water, phosphatides form colloidal solutions, lecithin and cephalin showing a more hydrophilic, and sphingomyelin a more hydrophobic character.

Schmitt and Palmer [86] reported that this behavior of the phosphatides is reflected in their x-ray diffraction patterns. The highly solvatized cephalin and lecithin show spacings that are considerably larger than those of the dry substances, the increment being due to the presence of water between the molecular leaflets at the polar interfaces. In cephalin and lecithin spacings as high as 120 Å have been observed, while the maximal values observed in wet sphingomyelin are about 80 Å. Cerebrosides do not emulsify in water, but show strong swelling when heated in the presence of water.

According to Hattori,[45] the smallest particles of lecithin sols show intensive Brownian movement in the ultramicroscope.

Preparation and Properties of Colloidal Solutions

Colloidal solutions of *lecithin* can be prepared by the method of Porges and Neubauer,[78] or Keeser,[52] who add small portions of alcoholic lecithin solutions to water and remove the non-aqueous solvent by boiling. Or lecithin and cephalin can be thoroughly ground and emulsified with distilled water. Even when 1 per cent (Keeser [52]) lecithin sols are kept in closed "Pyrex" bottles and mould infection is avoided, there is a marked change in appearance within the first days after the preparation of the solution.[100] The sols, which are first translucent and of a reddish-yellow hue, become more opaque, while much of the color fades. During 18 days' observation the electrical conductivity rose from 1.69 to 2.33×10^{-5} mho, the pH dropped from 3.86 to 3.71, while the viscosity showed a droped from 14051/5 to 1380 sec. The interferometric (-refractive) * values remain unchanged. The initial pH values of an emulsified lecithin sol are similar to those reached by the Keeser lecithin sol after 18 days. Fischgold and Chain [32] have reported spontaneous changes in lecithin even in the absence of air. Unlike lecithin, cephalin does not show signs of ageing.[104] This checks with findings of Bülow and Page,[14] according to which no hydrolytic cleavage of fatty acids could be observed in cephalin upon prolonged contact with air. The electrical conductivity of a 1 per cent Keeser lecithin sol is 1.69×10^{-5} mho, pH 3.86×10^{-3}. Corresponding values for a 1 per cent cephalin sol are 3.10×10^{-4} mho and pH 5.86 respectively. These latter values could be reduced by subsequent dialysis to 9.3×10^{-5} mho and pH 5.33.[104]

Further attempts at purification by electrodialysis, which is successful in the case of proteins, have not proved to be equally useful in the case of lecithin and cephalin. Even when an increase of pH was practically avoided (pH 3.73) by the use of a positively charged membrane on the side of the anode, lecithin and cephalin were completely precipitated, during electrodialysis, on the positively charged membrane.

* Measured with an Zeiss interferometer, which is based upon experiments of Rayleigh [*Proc. Roy. Soc. London*, 59, 203 (1896)].

The redissolved cephalin was more acid than the original solution (pH 4.03 against pH 5.86).[104]

Isoelectric Point

The ageing of lecithin shows its influence in the determination of the isoelectric point and, according to Chain and Kemp,[18] may be partly responsible for the large discrepancies in the reported values. Price[79] reports the isoelectric point of aged lecithin as pH 2.7, and assumes, since this value checks satisfactorily with others reported by Fujii,[40] Price and Lewis,[80] and Remesow,[82] that these authors too have worked with aged material containing adsorbed fatty acids. For freshly prepared lecithin Chain and Kemp,[18] Bull and Frampton [13] report values of pH 6.7, while the one calculated from the values of the dissociation constants of choline and glycerophosphoric acid and similar phosphoric acid derivatives is pH 7.5. Chain and Kemp [18] determined the isoelectric point of sphingomyelin as approximately pH 7 and found this value very sensitive to impurities.

Dielectric Properties of Phosphatides

Kuhn, Hausser, and Brydowna [59] have measured the dielectric constant of egg-yolk lecithin (prepared according to Levene and Rolf [64]) in alcohol and in benzene over a temperature range of 20° to 45°. Lecithin was dielectrically active only in alcohol in which it is unimolecular in betaine form. Both cephalin and lecithin are inactive in benzene, in which the latter polymerizes, reaching a molecular weight of about 3000. Studies of Hausser [46, 47] on sphingomyelin show a molecular resonance effect of the substance. The observed strong, anomalous dispersion is independent of viscosity; therefore the two long, weakly polar groups of the molecule are at rest and the strongly polar group (cholin phosphoric acid) vibrates like a pendulum with a low frequency. The author suggests the importance of these findings for a possible explanation of the behavior of brain and nerve substance to the specific action of short-wave therapy.

Acid and Alkali-binding Capacity

Studies on the ampholytic nature of phospholipids have been reported by Jukes [61] and Fischgold and Chain.[32] According to the latter, "the phospholipins, lecithin, lysolecithin, cephalin, sphingomyelin can bind in acid reaction 1 equivalent of H-ions. In alkaline reaction the phospholipins, which possess a primary amino group, give up 1 equivalent of H-ions. The quaternary bases are not capable of this reaction. The phospholipins containing a quaternary NH_4 base can exist only as cations or zwitter-ions, whereas the colamine phospholipins can exist as zwitter-ions, neutral molecules, anions, or cations. The ratio of zwitter-ions to neutral molecules is determined by the two acidity constants." A partial or even total replacement of colamine by serine should not fundamentally alter this concept.

Studies [104] of the *reactions of cephalin with acid (HCl)* show that a 1 per cent cephalin (human brain) solution becomes more opaque with increasing HCl concentration, until finally precipitation sets in. In 0.02*N* HCl the solution becomes heterogeneous, though it is still stable. From 0.01*N* upward centrifuging becomes effective in producing a precipitation of cephalin. Conductivity, pH and viscosity measurements in different HCl-cephalin mixtures give the following results: cephalin neutralizes a certain amount of acid, which increases with increasing HCl concentration to a maximum value; 1 g. of cephalin combines with $8.6 - 8.8 \times 10^{-8}$ ml *N* HCl. According to what has been said above, one molecule of cephalin in non-aqueous solutions combines with one molecule of acid. As Levene and West [65] found for brain cephalin having a molecular weight of 823.7, the normality of a 1 per cent solution ought to be 0.0124. The lower values found in these experiments suggest

that not all amino acid groups are freely accessible to the acid. This may be explained by the formation of colloidal aggregates in aqueous solutions.

The behavior of cephalin with HCl is very similar to the behavior observed in proteins with acid,[78] except that the computations based on pH and conductivity show that the ionization of the cephalin-HCl complex appears to decrease gradually with increasing HCl concentration.

These results seem to suggest the existence of a compound of cephalin and HCl in solution. However, Thierfelder and Klenk [115] reviewing the results of the literature on HCl·precipitates of cephalin, came to a different conclusion. In order to elucidate this question, special experiments were undertaken.[104]

(1) A solution containing 0.25 per cent cephalin and 0.025*N* HCl shows partial precipitation. A chloride determination in the filtrate after destruction of cephalin with HNO_3, shows that it contains only 90 per cent of the original amount of chlorine. This suggests that about 0.0025*N* HCl has been precipitated with a part of the cephalin.

(2) The conductivity was measured in a 0.25 per cent cephalin sol, in a 0.025*N* HCl solution, and in a solution containing 0.025*N* HCl and 0.25 per cent cephalin. The sum of the conductivities of pure cephalin and pure HCl was larger than the conductivity observed in the solution containing both substances. This difference in conductivities was 10 times greater than the conductivity of the pure cephalin solution. Calculated in terms of HCl normality, the apparent loss in conductivity corresponded to an amount of 0.0025*N*. These results seem to suggest that HCl not only combines with cephalin in solutions, but becomes part of the precipitate when precipitation occurs. By repeated centrifuging and replacement of the aqueous acid layer by pure water, most of the acid can be eliminated. This is due probably to hydrolysis of the lipide-acid compound. The acid cephalin flocculates, when redissolved in water, settles at an acid concentration which primarily did not interfere with stability in centrifugation. But neutralization with alkali restores the original behavior. These findings about the formation of the cephalin-HCl compounds are perhaps not entirely devoid of biological interest; *i.e.*, Peters and Man [76] claim that part of the chloride of the serum is combined with lipides.

The results of the viscosity determinations made with an Ostwald viscosimeter at a constant temperature of 30° show with increasing concentration of HCl a steady drop in viscosity, which only begins to rise again at a stage preceding precipitation.[109] The initial drop becomes less distinct if the mixtures stand for 24 hours. These observations on the changes of cephalin viscosities upon addition of acids are very similar to those reported by Handovsky and Wagner,[43] on lecithin under comparable conditions. But the behavior of cephalin is in this respect markedly different from the behavior observed in proteins; for with the latter increasing HCl concentrations produce at first a rise in viscosity and ionization.

Similar studies were made on lecithin of various origins (egg and human brain).[104] The results of these experiments show that both samples of lecithin bind HCl, but that according to their chemical differences (as indicated by their different pH) the binding capacity reaches its maximal value at a different HCl concentration. The fact that lecithin in aqueous solutions binds HCl is rather unexpected, since, according to Jukes,[51] and Grün and Limpächer,[42] lecithin is an internally neutralized compound and has no free amino-group. But systematic investigation [103] of this salt-binding capacity of lecithin sols has given evidence of the sorption power of lecithin. Furthermore, former experiences on globulins [92] have shown that acid binding can be supplemented by acid sorption.

Attempts were made to study the *alkali-binding capacity of cephalin.* Since the reaction of NaOH with cephalin may lead to some chemical changes in the cephalin, attempts were made to ascertain whether the highest NaOH concentrations used produced a splitting of cephalin.[104] As a breakdown of the cephalin molecule by

alkali ought to liberate fatty acids, conductivity methods could be used in order to trace an increase in conducting material.

Samples containing 1 per cent cephalin and 0.02*N* or 0.05*N* NaOH were exactly neutralized after 24 hours. The conductivities of these solutions were nearly identical with the sum of the conductivities of controls containing either pure cephalin or pure alkali. A slight decrease in the conductivities of the mixtures was explained by the inactivating effect of cephalin on the salt. A splitting of cephalin in the alkali concentrations used in our investigations can therefore be disregarded. The optical changes in cephalin on addition of alkali proved to be reversible on neutralization.

By using a method similar to the one used in the computation of the acid-binding capacity, it can be shown that the NaOH-binding capacity exceeds by far the combination power of cephalin for HCl. Apparently in the presence of excess alkali a maximal value could also be reached.

Viscosity measurements of cephalin at varying NaOH concentrations show a great similarity in behavior to proteins, especially globulins.[97] Here too, with increasing amounts of NaOH, the viscosity increases and drops again on further addition of alkali. The maximum viscosity is reached at a much lower concentration of alkali than the maximal binding capacity. Yet all these changes occur in the first hours after mixing cephalin and alkali. After several hours of standing, an increase of alkali has uniformly produced a corresponding decrease of viscosity.

The fact that cephalin, though combining with acid like a protein, shows no increase in viscosity like lecithin, may be of some interest to the biologist. The non-swelling of nervous substance in acids (Bauer;[11] Spiegel[90]) has been attributed by Hooker and Fischer[50] to reactions of lecithin prevailing over the reactions of the proteins. This explanation gains new support from the results on cephalin reported here. Thus according to Singer[89] brain contains more cephalin than lecithin. Recent studies of Spiegel and Spiegel-Adolf[91] have shown that in dead and living brains the polarization indices (a convenient measure of the swelling and the permeability of tissues) show a marked decrease on the addition of alkali, whereas they are unchanged or only minimally lowered in the presence of acids. According to Jukes,[51] lecithin does not bind alkalies.

According to W. and B. Tonnis[116] results may be different in separate studies of white and gray matter of the brain. According to those authors, sphingomyelin exceeds in swelling processes all cerebrosides especially in an acid medium. Sphingomyelin is reported to form 8.9 per cent of the white and only 0.1 per cent of the gray matter.

Reaction of Phosphatides with Neutral Salts

Like proteins, phosphatides (unpurified lecithin sols) are precipitated by *neutral salts;* the effectiveness of anions and cations follows roughly the series of Hofmeister and Pauli.[115] But there seem to exist two significant exceptions.

(1) Already Koch[56] and Porges and Neubauer[78] have reported the great sensitiveness of phosphatide sols to alkaline-earth salts. The small amount of the latter inducing precipitation comes within the range of effectiveness observed in hydrophobic colloids. The high values of the quotient $Q = \frac{C K_2 A^*}{C K_1 A}$ derived from this behavior accounts, according to Bungenberg de Jong and Saubert,[15] for the fact that univalent salts and bivalent salts have an antagonistic effect in suspended quartz particles coated with lecithin films of different degree of purification. Contrary to Koch,[56] Bungenberg de Jong and Saubert[15] do not explain the antagonistic effect by a special phosphatide reaction but by interaction of the ions of the medium.

* K_1A indicates a salt of which smaller concentrations are necessary to reverse the charge of a phosphatide than of K_2A. The concentrations producing a reversal of the charge are designated as $^C K_1 A$ and $^C K_2 A$.

(2) *Reactions of phosphatides with neutral salts* were studied with special reference to biological problems. Measurements of viscosity, interferometry and conductivity indicate that, as with proteins, lecithin in colloid solution combines with neutral salt.[103] But contrary to observations made on proteins, bromides,[99] in some instances, have twice the effect of chlorides; this becomes particularly marked in the observed decrease in viscosity. These findings, which are perhaps of interest to the pharmacologist, were explained by the action of free bromine. This hypothesis could be substantiated by showing that the specific effect of bromides depends upon the iodine values of the colloid lecithin. Reduction of these values by spontaneous oxidations, irradiation or iodination brings about parallel changes in the specific effects of bromides, which disappear completely in maximal iodination.[101] In acidified lecithin and cephalin the Hofmeister series is valid.

Quantitatively, the salt-binding capacity of lecithin sol [103] depends upon several factors, *e.g.* the time the salt is in contact with the colloid. Special experiments showed that these changes do not depend upon the ageing of the lecithin sol. The amount of salt bound to lecithin sol depends upon the concentration of both salt and lecithin. If the concentration of lecithin sol is kept constant and the amount of KCl varied, then the graphical presentation suggests the existence of a salt-binding maximum in a surplus of salt. The amount of salt bound to lecithin sol depends upon the degree of, dispersity of the latter. Bungenberg de Jong, Verberg and Westerkamp [17] have shown that even small amounts of impurities prevent the formation of a translucent sol. With a different preparation (Merck, 90 per cent pure) the lecithin sols became opaque and milky and the salt-binding capacity of the lecithin sol dropped to about 1/10 of the original value.

In order to elucidate the type of salt fixation to lecithin sol, two different series of experiments were undertaken.[103] In aqueous solutions lecithin is negatively charged. Electrophoresis experiments were made using the apparatus of Landsteiner and Pauli.[60] The KCl solution was varied. Up to 0.1*N* KCl no change in the anodic movement of the lecithin sol could be detected whereas, according to Koch and co-workers,[57] the salts with bivalent cations bring about a positive charge on the phosphatides. In systematic investigations on globulins [97] it could be shown that globulin causes different decreases of conductivity in isonormal solutions of KCl, NaCl and LiCl. This has been explained by the different velocities of the cations and as a proof that part of the cations had been bound to the globulin. Similar experiments made with lecithin sol gave analogous results.

The results of both the electrophoresis and the conductivity measurements seem to indicate that both salt ions are fixed by the lecithin, and that the new compound shows little if any ionization. Under this assumption 1 per cent lecithin allowed to stand for 24 hours in contact with 0.05*N* KCl binds approximately 18 per cent of the salt.

In another series of experiments the influence of some narcotics on the salt-binding capacity of lecithin was investigated.[103] Although the presence of ethyl alcohol, ether, chloroform, and chloral hydrate seems to decrease the salt-binding capacity of lecithin, quantitative measurements were attempted only in experiments with homologous alcohols, as these have been studied especially to correlate their chemical and biological behavior.[48] The results show that with increasing length of chain the effect of alcohols in decreasing the salt-binding capacity of lecithin becomes more marked. In the presence of amyl alcohol the salt-binding capacity practically disappears. At the same time the opacity of the lecithin sol increases, as well as its sensitiveness to salt precipitation. Similar behavior was observed by Freundlich and Rona [88] when colloidal ferric hydroxide was treated with narcotics.

Effects of different dielectric constants and changes in the sorption capacity of colloid lecithin brought about by various alcohols have been considered as a possible explanation. Koch and McLean [58] found no evidence that anesthetics or hypnotics

produce changes in the state of aggregation of lecithin or cephalin. Nevertheless, the most probable explanation of the reported changes of the salt-binding capacity of lecithin in the presence of homologous alcohols seems to be an assumed change in the degree of dispersion of the lecithin sol. The relation between salt-binding capacity and degree of dispersion has been mentioned above. If we may assume that egg lecithin behaves physicochemically like human lecithin, then the diminished ability of the lipides in the cellular surface films to fix salt ions should diminish the reactivity of the cell to stimuli. Changes in ion concentrations must reach a higher threshold in order to act upon the lipides of the cell surface or its interior. In this connection it seems of interest that there is a certain parallel between the narcotic effect of homologous alcohols and their power to lower the salt-binding capacity of lecithin.

Reaction of Lipides with Colloids and Bi-colloids

Since lecithin and cephalin in aqueous solutions are colloids, their *reactions with other colloids and especially bio-colloids* are of particular interest. Both lecithin and cephalin are able to protect colloidal gold against the flocculating effects of neutral salts.[102] Amounts of a 1 per cent cephalin sol greater than 0.1 cc protected 1 cc of the red gold sol when 0.2 cc of normal potassium chloride was added; 0.5 cc of the cephalin failed to protect completely, since the red gold changed to red-violet in color. Smaller quantities showed still less protective action.

Using similar quantities of gold sol and normal potassium chloride, human lecithin showed protective action in quantities of 0.3 cc and above of a 1 per cent lecithin sol; some protective action in quantities of 0.2 cc; little or none in smaller amounts. These experiments demonstrated that cephalin did not precipitate the colloidal gold sol, that presence of cephalin in appreciable concentration protected the colloidal gold against precipitation by potassium chloride, and that the protective action of cephalin toward colloidal gold is greater than is the protective action of human lecithin from the same brain tissue. The protective power of human lecithin is greater than that of egg lecithin (Merck). Both lecithin sols were prepared by the method of Keeser.[52] In its behavior toward colloidal gold, cephalin shows, therefore, a closer resemblance to gelatin than to the proteins of the blood and of the egg white, which precipitate gold, unless in the presence of electrolytes. Use has been already made of the protective power of phosphatides for industrial purposes. Eichberg[28] reports soybean lecithin as an effective stabilizer for leaded gasoline, preventing cloud formation, color change and aluminum corrosion (1 — 10 pounds of lecithin stabilizing 42,000 gallons of gasoline). The use of lecithin in chocolate confectionery is based, according to Erb,[29] upon moisture absorption by the phosphatides. The addition of 0.2 per cent lecithin will protect against increases in moisture content up to 1.5 per cent. Lecithin has also been used in the homogenization of milk,[71] and soybean lecithin may be used as a partial substitute for egg yolk in ice-cream and as a stabilizer of vitamin A in foods. (Stanley[112])

Because of the close connections between *lipides and proteins* in the human body, reactions between these two substances have been studied extensively. Porges and Neubauer,[78] and Arnd and Hafner[6] assumed the formation of lecithin-protein compounds; Handovsky and Wagner[43] report only negative results. In order to explain this discrepancy special investigations were started.[100]

The reaction between *lecithin and serum protcins* used by the former investigators largely depends upon the pH of the solution, the presence of neutral salts, and the age of the lecithin sol. The first point has been extensively investigated by Rona and Deutsch,[83] but underestimation of the important role of salt concentration in the relation between lecithin sol and proteins seems to be the reason for contradictory opinions in this matter.

Electrolyte-free solutions of serum-albumin and pseudoglobulin flocculate lecithin sol made from egg or brain lecithin. Fresh samples of 1 per cent lecithin sol are

flocculated to an increasing extent by serum-albumin from 0.4 per cent upward, whereas pseudoglobulin became effective at a concentration of 0.2 per cent. An aged lecithin sol is about 10 times more sensitive to serum-albumin and about 50 times more sensitive to pseudoglobulin flocculation than a fresh one.

According to Freundlich,[87] the ageing of lecithin sol consists of changes by which it partially loses its hydrophilic properties and grows more hydrophobic. In good accord with this, the amounts of protein necessary for the flocculation of aged lecithin sol are within the range of the corresponding values for gold sol.[95] But it seems obvious that the changes reported in the isoelectric point of lecithin upon ageing are the cause of the increased sensitiveness of aged lecithin toward serum proteins. As with most negative sols, the flocculation of lecithin by protein is inhibited by the presence of neutral salts. Since contradictory findings [6] have been described as to the formation of protein-lipide compounds, systematic studies in this direction were begun.[100]

The findings suggest the following interpretation. As long as the salt concentration is sufficient to avoid the faintest signs of flocculation, no changes in the refractive indices are observed. Different salts seem to behave in the same way. If the salt concentration is lowered to a point at which the solution becomes unstable, then the interferometric value of the solution is lower than the sum of the corresponding values for the separate components. The results are the same whether the unstability is due to an initial low salt concentration, or whether, through the ageing of the colloid, the salt concentration becomes insufficient. These findings explain the observations of Handovsky and Wagner [43] as well as those of Arnd and Hafner,[6] though they do not necessarily confirm the conclusions of the latter. The decrease in refraction must be explained primarily by a decrease of soluble material. The differences in the flocculating effects of serum-albumin and pseudoglobulin on lecithin sol persist in the presence of neutral salts. Larger quantities of the latter are necessary to prevent flocculation by pseudoglobulin than by serum-albumin. It can be shown that certain serum-albumin-neutral salt combinations can protect lecithin sol against subsequently added pseudoglobulin, although the salt present is not alone able to prevent the flocculating effect of the latter. Such experiments are possible in cases in which there is a gap between the salt concentration capable of inhibiting the flocculating effect of serum-albumin and the salt concentration having the same influence on pseudoglobulin. For this purpose, only lecithin sol of a certain age can be used. In freshly prepared samples, the differences are too slight; in very old sols, the protective power of salts, even against albumin, sometimes vanishes. For these experiments, according to Keeser, one could use only lecithin sols 14-21 days old, fresh emulsions of lecithin in distilled water, or lecithin sols derived from an older, perhaps oxidized sample of lecithin.

These results seem to show that under certain conditions flocculating concentrations of pseudoglobulin fail to produce this effect in the presence of serum-albumin and neutral salt. Though the neutral salt at the same concentration is, in itself, not sufficient to prevent the flocculating effect of pseudoglobulin, its presence is necessary to give the serum-albumin protective power. A non-flocculating concentration of the latter in an electrolyte-free medium is unable to inhibit the precipitation of lecithin sol by pseudoglobulin. The negative result of interferometric investigation made in order to trace some reaction between electrolyte-free serum-albumin and pseudoglobulin is in accordance with these findings. For this reason, the above-mentioned influence of the neutral salt-serum-albumin compound on pseudoglobulin cannot be directly compared with observations made by Pauli and Singer,[75] who found that albumin prevented, to some extent, the flocculation of certain colloid dyes by serum-albumin in the absence of salt.

Next to the salt concentration, the protective power of lecithin depends upon the ratio of serum-albumin to pseudoglobulin. If the latter falls below ⅔-½, precipitation

begins. It seems, therefore, that the behavior of the total protein mixture toward the lecithin sol depends on the relation of its components. The similarity of this behavior to that of serum under various pathological conditions is somewhat striking. There, too, the so-called albumin/globulin quotient (Hoffmann[49]) seems to determine, in many cases, the properties of the serum and other biological fluids. The reactions of spinal fluid toward gold sol in particular seem to depend largely on this property according to Schmitt.[87] An increase of globulin has been described in nearly all luetic changes. The above-mentioned results may be of some value for a further understanding of luetic reactions using flocculation methods (*e.g.* the Meinicke and Lange tests for syphilis).

There exists a marked, although possibly only quantitative, difference in the behavior of cephalin and lecithin (both derived from the same human brains) toward serum proteins.

Neither serum albumin in concentrations up to 1 per cent nor pseudoglobulin in concentrations up to 2.5 per cent cause any precipitation in fresh samples of cephalin.[102] In aged samples of cephalin no apparent change was called forth by serum albumin, although 1-2.5 per cent pseudoglobulin caused an increasing cloudiness. Even an interferometer did not show a change in the degree of dispersion of a cephalin sol upon the addition of serum-albumin, such as has been found in lecithin-protein mixtures when flocculation had been prevented by a low amount of salt. A 0.25 per cent cephalin solution gave an interferometer reading of 75 units, a 0.25 per cent serum-albumin sol a reading of 112 units and a sol containing 0.25 per cent cephalin and 0.25 per cent serum-albumin a reading of 189 interferometer units, this latter being within experimental error of the calculated 187 which should have been obtained had the values been additive.

The above-mentioned differences in the behavior of cephalin and lecithin can be used as tests for the protective power of cephalin on lecithin against the precipitating effect of protein. Lecithin of human brain was used in these experiments. In spite of the fact that cephalin sols mixed with lecithin sols show added values in the interferometer and therefore do not indicate a change of dispersion,* nevertheless the influence of cephalin on lecithin becomes very marked when protein is present.

Cephalin apparently protects against the precipitating effect both of serum-albumin and pseudoglobulin. In a 0.6 per cent serum-albumin sol containing 0.5 per cent of lecithin there was an immediate precipitation when 0.025 per cent or less of cephalin was present. If, however, the cephalin concentration equaled 0.05 per cent or more, no precipitation of the albumin-lecithin system took place. The importance of using cephalin-free samples of lecithin for physical-chemical tests is evident. On one hand these findings may explain the discrepancies in the literature about the flocculating effect of proteins on lecithin sols; on the other hand, they show that there exists a mechanism besides the action of neutral salts that is able to prevent flocculation of lecithin by proteins. These phenomena also may be interpreted as instances of cumulative protection as described by Alexander.[4]

In connection with the problem of blood coagulation, Chargaff and co-workers[19] have studied the reactions of phosphatides particularly with basic proteins. They report noteworthy differences in the behavior of cephalin and lecithin. Cephalin forms salt-like compounds with salmine, a protamine, over a pH range of 2-11, lecithin only at pH 10-11. Sphingomyelin is not precipitated by addition of salmine. Cephalin forms precipitates with thymus proteins at pH 2-7, lecithin only at pH 7-8. It is pointed out that the ability of the basic proteins to precipitate cephalin at a physiological pH parallels their inhibitory effect on blood coagulation in a marked way. Cephalin forms a compound with egg-albumin at pH 2-4; no analogous phenomenon is observed in lecithin. Similar behavior has been noted in histon, which

*According to Bull,[12] the wide difference in isoelectric points of the two lipides is suggestive of coacervate formation over a considerable range of pH.

apparently reacts with cephalin only at a pH less than 4.0. Only cephalin is able to disrupt the heme-globin linkage in oxy- and CO-hemoglobin by forming cephalin-globin compounds. These differences are supposed to indicate different biological rôles of cephalin and lecithin in the organism.[20] Chargaff [19] gives an extensive description of the compound formed by cephalin and salmine. The dried precipitates are characterized by a P : N ratio of 1 : 4 or 1 : 5; they swell in contact with water to a rubber-like consistency; they can be recrystallized from ethyl alcohol, etc. A further elucidation of the structure of such compounds is given by Schmitt and Palmer [86] in collaboration with Chargaff; they studied x-ray diffraction patterns of precipitates of cephalin with basic proteins. Depending on the protein used an increase of 10-15 Å above the spacing of cephalin was found. Therefore, the existence of only one more layer of protein can be assumed between each double layer of cephalin.

Schmitt and Palmer [86] assume that "the protein molecules are spread out flat at the polar interfaces of the cephalin layers and the binding of the negative groups of the cephalin by the positive groups of the basic proteins on both sides of the protein monolayers causes the collapse of the highly solvated cephalin system and the expulsion of water." Approximately 70-80 Å of water can be expelled in this way, indicating the extensive changes brought forth in lipide systems by the introduction of basic proteins. Reactions of lecithin with casein have been studied by Parsons,[72] who devised a colorimetric method to determine the amounts of lipide and protein entering into reaction. He explains the latter as a precipitation of colloids with opposite charge, the amount of which depends on the pH of the medium.

Fujii [40] reports a dehydrating effect of egg albumin upon lecithin by which the solubility of the latter in ether is increased. Bamberger [9] studied the influence of lecithin upon gelatin gels. He found an increase of the swelling capacity of 12-15 per cent independently of the pH, while the bound water, measured by the method of Rübner, is diminished. The behavior of cholesterol sols is antagonistic to lecithin, as it decreases the swelling capacity by 8-10 per cent but is apparently without influence on the amount of bound water.

It is impossible to give here even a short review of all the papers dealing with the relation of protein and lipide in serum (see globulins).[97] But it seems worthwhile to note the fact that, according to Went and Kuthy,[118] some of the changes brought forth in serum proteins by lipide extraction may be caused by partial denaturation of proteins.

There exist some references to a direct *influence* of *lipides* upon *heat-denaturation* of *proteins*. Bancroft and Rutzler [10] report that heat-coagulation of egg-albumin can be prevented through previous extraction with peroxide-free ether, which removes a lecithin-like substance. Pauli and Omar [74] have been able to show that through simultaneous removal of CO_2 the pH is shifted from pH 6.28 to pH 8.44, at which no heat coagulation occurs. Nevertheless, experiments with dialyzed cephalin ($K = 9.3 \times 10^{-5}$mho, pH = 5.33 in a 1 per cent solution) showed that, like gelatin, it was able to prevent heat coagulation in serum-alubumin of twice its own concentration, and failed to do so when the relation of cephalin to serum-albumin was increased to 1 : 3. In all these experiments the lipide protein mixtures and the controls were heated for fifteen minutes in a boiling-water bath.[102]

Baker, Harrison, Miller and Wexler [8] report an inhibition by phospholipides of the action of synthetic detergents on bacteria. They discuss the possibility that the phospholipides act by preventing a denaturation of the proteins of the bacterial membranes through the detergents.

Finally the effect of cephalin on *cholesterol sol* was studied. In order to have comparable data for human lecithin all experiments were made with this substance. The sol of cholesterol (Merck) was made according to the method of Keeser. After filtration the milky solution was fairly stable.[102]

In preliminary tests the reactions of the cholesterol sol toward neutral salts, serum-albumin and pseudoglobulin were investigated. The observations were in harmony with results of former investigators. In series of experiments it was shown that cephalin as well as lecithin is able to protect cholesterol sol against flocculation by salts or proteins. In a system containing 0.5 per cent cholesterol and 0.1 cc of a normal potassium chloride, cephalin in 0.0016 per cent nearly inhibited the flocculation of the cholesterol. When the cephalin concentration was 0.0024 per cent or greater, the flocculation was completely inhibited. Flocculation was likewise inhibited in a similar system when lecithin in concentration of 0.2 per cent or more was used instead of cephalin.

In a system containing 0.5 per cent of cholesterol and 0.05 per cent serum-albumin, cephalin in concentration of 0.01 per cent or more inhibited flocculation, whereas definite flocculation took place in a similar system when lecithin was present to the extent of 0.05 per cent; but no flocculation of the cholesterol-protein system took place when the lecithin concentration was raised to 0.1 per cent.

These findings may have to be considered in all cases in which the role of protective colloids has been thought important for the stability of cholesterol, *e.g.,* in the formation of concretions under pathological conditions.[98]

Chargaff and Ziff [19] have given evidence of salt formation between cephalin and the basic dyes methylene blue and thionine. No similar behavior was observed in lecithin. *Reactions of lipides with dyes* are of interest for the evaluation of histochemical reactions. According to Shapiro,[88] only cerebrosides but not phosphatides respond to the usual dyeing method, since the results became uniformly negative upon extraction with acetone. The author explains this behavior by the fact that phosphatides are bound to proteins. Extracted lipides, according to Diamare,[26] are not changed in their birefringence by treatment with "nilblau" or Sudan III after the usual fixation methods.

No theory has as yet been advanced as to the nature of the special structure of the cephalin molecule to which protective power of the lipide is due. Some evidence for such formulation might be obtained, however, by the results of the following experiments. It was ascertained that the protective power of cephalin on colloidal gold is practically unchanged after the solution has been boiled. This is in agreement with the results reported on the protective power of cephalin against the heat coagulation of protein, and proves that the protective power of cephalin is heat-stable within a certain pH range. On the contrary, the protective power of cephalin on cholesterol is practically completely destroyed by intensive irradiation with ultraviolet light (quartz mercury lamp, Hanovia, 220 V.d.c., four hours at 20 cm distance, 0.25 per cent cephalin sol). Thus, for example, a 0.5 per cent cholesterol sol containing 0.1 cc of normal potassium chloride was not protected, even when a quantity of irradiated cephalin as great as 0.8 per cent was added to the system.[102]

The *irradiated* cephalin sol does not show gold sol flocculation as observed in proteins under similar conditions, nor hemolysis of erythrocytes as has been described by Fabre and Simonnet [80] in irradiated lecithin, through the formation of lysolecithin-like substances. At all events, according to Magistris,[69] a sample of lecithin free from lysolecithin is not hemolytic. As with proteins, heat and short-wave light have a different effect on cephalin. An analysis of the effects of irradiation may, therefore, be helpful for the understanding of the protective power of cephalin.

Monomolecular Films of Phosphatides

Leathes [61] has observed the spreading of lecithin as *monomolecular films* on water, and explains the film formation by the fact that lecithin contains two polar groups. The hydrophobic paraffin chains try to escape from the surface of the water, while the hydrophilic glycerophosphocholic groups are anchored in the surface. The

film is formed by the cohesive powers between the hydrophobic groups standing nearly vertical to the surface of the water. By measuring, according to the method of Adam,[1] the surface of a monomolecular film taken up by a certain quantity of a certain substance, it is possible to compute the space occupied by one molecule of the substance. In this manner Leathes was able to show that the paraffin chains in lecithin are less closely packed than in fatty acids. This seems to facilitate the exchange of water and other substances and to indicate the biological importance of lecithin films. Alexander and Theorell [2] have included cephalin in similar studies and found (Alexander, Theorell and Aborg [3]) that cephalin films are stabilized by calcium ions.

Phosphatides in Artificial Membranes

The importance of lipides for polarization effects in biological membranes has been stressed by various authors.[67] Experiments on artificial membranes have been made in order to elucidate this phenomenon.[105] The artificial membranes consist of a sintered glass plate on which various colloids can be adsorbed. The polarization is determined by the polarization index, computed from conductivity data obtained at various frequencies and by measurements of potential differences.[111] It could be shown that the existence of polarization depends upon the presence of lipides (lecithin), although the synergistic action of another colloid (protein or starch) is necessary. The role of the latter substances seems to be more of a mechanical nature, *e.g.*, that of decreasing the size of the pores in the sintered glass plates.[105]

Such artificial composite membranes containing lecithin show a number of reactions observed in biological membranes only, *e.g.*, the specific effects of bromides, increase of polarization in the presence of magnesium salts, increased polarizability upon substitution of ethyl alcohol (used as solvent for lecithin) by a higher homologous alcohol of greater narcotic effects, etc.

Molecular Orientation of Phosphatides

Although cephalin and lecithin look amorphous in solid form, their particles acquire a certain degree of orientation upon swelling.[62] Under the influence of the water the particles become moveable against each other and can be oriented by surface tensions.[5]

Studies in birefringence of lecithin have been made by H. Ambronn.[5] According to Göthlin [41] the oleic acid radical of the latter is responsible for this behavior.

X-ray diffraction patterns of lecithin, cephalin, and sphingomyelin from mammalian central nervous system have been reported by Schmitt, Bear and Clark [85] and Schmitt and Palmer.[86] According to these authors, all three phosphatides show diffraction line or lines corresponding to spacings between 4.22-4.65 Å. Lecithin and cephalin show also spacings of about 46 Å; sphingomyelin has a diffraction line corresponding to spacings between 63-67Å. Schmitt and Palmer [86] take about 46 Å as the characteristic spacing of lecithin and cephalin.

Our own observations 107-109 have concentrated upon x-ray diffraction patterns of various lecithins and cephalins from the human nervous system, from soybean phosphatides, and egg-lecithin upon the influence of different steps in purification and upon experimental factors able to influence the x-ray diffraction patterns of these substances. Comparative studies of x-ray diffraction patterns of lipides show similarities and differences. All lipides under investigation show a diffraction ring or rings corresponding to spacings of approximately 4.5-4.8 Å. This checks with the above-mentioned findings of Schmitt, Bear and Clark,[85] although our technique was not chosen to give complete x-ray diffraction patterns. Three of our lipides show also faint rings corresponding to spacings of 2.90-3.1 Å.

Differences in the diffraction patterns seem to exist between lipides of different origin and within lipide individuals of the same group. In egg lecithin and in all

soybean phosphatides the main ring (4.35-4.51 Å) is diffuse, approximately 2 mm wide. In cephalin and lecithin from human brain there are several fairly sharp rings instead of a diffuse one, which correspond roughly in spacing to the latter.

Within the various groups no differences can be found between lecithin and cephalin from soybean in respect to the main spacing. But in the group of human brain lipides such differences are manifest. Purified human lecithin contains two more rings than the corresponding fraction of cephalin. Diffraction rings corresponding to spacings between 2.9-3.1 Å are observed in two types of egg lecithin, in crude, but not in purified lecithin from human brain, and in one cephalin from soybean. This seems to suggest varying effectiveness in fractionation.

Both the similarities and the differences between the members of the human brain and the soybean group seem significant. Already Schmitt and co-workers [85, 86] had reported that they had noticed "that dry samples of lecithin and cephalin give frequently two long spacings rather than one." They explain these lines either by the presence of mixtures of carbon chains of varying length and possibly varying angles of tilt (Malkin [70]) or by the presence of groups which are not clearly defined.

Our own investigations [109] have shown that the spacings of 4.17 and 4.6 Å observed in dried lecithin also occur in oleic acid upon freezing. As unsaturated acids are essential constituents of both lecithin and cephalin, this fact suggests that the above spacings too are characteristic of the phosphatides and worthy of further study. Attempts were therefore made to produce changes in the lipides which might influence their x-ray diffraction patterns. A few preliminary experiments were made with the crude sample of human lecithin. Drying over P_2O_5 to weight constancy does not, apparently, change the x-ray diffraction pattern in human lecithin in which sharp lines already exist in the spacings between 4.5-4.8 Å. But a sharpening of the diffuse band upon drying could be observed in soybean lecithin. Heating for two hours at 105° changes the inner of the three sharp rings of human lecithin. Further purification of the crude lecithin sample brings out a fourth ring in the pattern.

Since in a number of fluid fatty acids, freezing produces a separation of the diffuse diffraction band into several sharp lines,[110] similar experiments were tried on soybean lecithins of viscous consistency.[109] A separation into at least four separate lines (3.67, 3.86, 4.17, 4.6 Å) was achieved. But this method failed for obvious reasons in the case of solid cephalins. Therefore another procedure, known to produce dehydration and increase of crystallinity at least in proteins, has been tried. Previous experiences in proteins have shown that sharpening of the so-called backbone reflection is brought forth by different kinds of denaturation. A sample of euglobulin, which had lost its salt solubility by extensive purification, showed sharpening and splitting of the backbone reflection. The latter phenomenon has been described in heat denaturation of proteins by Astbury and Lomax [7] and has been confirmed by Spiegel-Adolf and Henny,[106] who have been able to show parallelism between x-ray diffraction pattern changes and reversibility of protein denaturation.[94, 96] Experiments were therefore started to study possible heat changes in lipides and to correlate them to the x-ray diffraction patterns of these substances. For obvious reasons soybean lipides were chosen for these experiments, since the x-ray diffraction patterns of these substances are originally diffuse rings. Nevertheless, we have duplicated the experiments to be reported here with lipides from human brain.

A technique of heat denaturation in lipides [107, 108] was developed. Since maximal heat precipitation occurs in proteins at isoelectric reaction and in the presence of neutral salts, cephalin was investigated under similar conditions. The aqueous solutions of cephalin from soybean and from human brain have a pH of 6.82-6.95 and 5.86 respectively. According to Thierfelder and Klenk,[115] the isoelectric point of cephalin is pH 2.7. This corresponds rather closely to the reaction of 1 per cent cephalin from human brain containing 0.01*N* HCl ($cH = 2.28 \times 10^{-3}$, pH = 2.64).

Such a solution does not show any changes upon heating, nor does a 1 per cent cephalin solution containing 0.15*N* KCl. A 1 per cent cephalin solution containing both 0.01*N* HCl and 0.15*N* KCl is stable at room temperature for at least 48 hours, but shows very marked precipitation upon boiling in a water bath. No precipitation occurs when the neutral salt is added after boiling and cooling of the acidified cephalin solution. The importance of the sequence of salt addition and heating for the occurrence of coagulation in proteins has already been stressed.[94] Lecithin from soybean shows upon heating a behavior similar to cephalin. These observations on cephalin and lecithin apparently show the existence of heat precipitation in both cephalin and lecithin under specific conditions. Since it has been found that this can be induced in protein material already heat-denatured,[98] heat precipitation alone cannot be accepted as a sufficient proof for the existence of heat denaturation. It seemed, therefore, of sufficient interest to examine the x-ray diffraction of lipides subjected to heat precipitation.[107, 108] To obtain x-ray diffraction patterns of the heat-precipitated lipides it is necessary to remove the salt and acid added to induce heat precipitation. When practically all the electrolytes are removed, heat-precipitated cephalin remains for the most part insoluble in the aqueous medium, while the heat-precipitated lecithin spontaneously goes into colloidal solution. Both preparations were dried in vacuum desiccaters over P_2O_5. In order to get comparable materials, lecithin and cephalin that had been boiled in the absence of electrolytes were treated in the same way. The x-ray diffraction patterns indicate differences in the behavior of heat-precipitated cephalin and lecithin. Heat-precipitated cephalin shows a sharpening of the otherwise diffuse ring which is similar to the so-called "backbone sharpening" in proteins.[7] It would, therefore, seem probable that cephalin is not only precipitated but also denatured (dehydrated) by heat. Cephalin that was simply boiled in the electrolyte-free solution showed the same diffuse ring as the genuine sample, except that the faint ring corresponding to a spacing of 2.90 Å became still fainter in the heat-precipitated sample. No changes could be observed in the diffraction pattern of heat-precipitated lecithin; it is identical with the patterns of boiled and ordinary lecithin.

Since heat-denatured serum-albumin regains its water-solubility and original x-ray diffraction pattern after treatment with diluted alkali,[94, 106] similar experiments were started on heat-precipitated cephalin. The precipitate dissolved very easily and after subsequent neutralization, dialysis and desiccation, the x-ray diffraction pattern once more showed a diffuse ring, similar to the pattern of the original preparation. Even the faint ring at 2.90 Å again became more visible. Heat denaturation of proteins had previously been interpreted as a condensation of the terminal positive and negative groups under ring formation.[94, 96] The isoelectric reaction conditions are optimal for an equal number of positive and negative terminal groups. A decrease of hydration induced by the addition of neutral salts ought to facilitate reactions between different protein molecules. It is interesting that similar conditions are necessary for heat precipitation of monophosphatides. Heat denaturation is observed only in cephalin, which like protein is an ampholyte, *i.e.*, it combines with both acid and alkali. It is absent in lecithin, which according to Jukes,[51] has no ampholytic properties in aqueous solution. Heat denaturation seems to have an effect, at least on cephalin, that makes its x-ray diffraction pattern approach that of brain cephalin. It would, therefore, seem correct to explain the differences seen in the x-ray diffraction patterns of the soybean and human brain lipide group as due to the more drastic treatment to which the latter lipides have been subjected. The differences reported in the pattern of crude and pure lecithin from human brain seem to point in the same direction. Hansteen-Cranner[44] claims that most lipides become denatured upon extraction. Our findings confirm the existence of lipide denaturation, which has been already conjectured on the basis of changed protective action in irradiated cephalin. Conversely, the differences in denaturation of lipides, as assumed from differences of their x-ray diffraction patterns, may be mostly of a quantitative nature.

Physicochemistry of Cerebrosides

There exist only very few *physicochemical* and *colloid-chemical* investigations of *cerebrosides*. Both the monograph of Thierfelder and Klenk [115] and the review of Klenk and Schuwirth [55] contain a chapter devoted to the physical chemistry of phosphatides; but an analogous presentation of cerebrosides is lacking. Nevertheless, a few items pertaining to physicochemical aspects of cerebrosides could be collected, besides the determinations of optical rotation and melting point used for the characterization of the single individuals.

Occurrence of myelin forms and liquid crystals (Rinne [81]) has been reported before. The cerebrosides crystallize well. It seems worth mentioning that there exist characteristic differences in the birefringence of the crystals of individual cerebrosides. Rosenheim [84] has based on this behavior a test of purity of either cerebron (phrenosin) or kerasin. X-ray diffraction studies on dry phrenosin and kerasin have been reported by Schmitt and Palmer.[86] According to these, kerasin and phrenosin show, besides diffraction lines corresponding roughly to the ones of lecithin and cephalin at 4.2 Å, others corresponding to spacings at 67 and 48 Å respectively. The x-ray diffraction pattern of kerasin is very similar to the one of sphingomyelin.

Chibnall, Piper and Williams [24] have used x-ray diffraction studies to ascertain the nature of fatty acids in phrenosin and kerasin. They found that three phrenosins exist in brain and at least three kerasins which are compounds of different fatty acids. Similar investigations have been made by Crowfoot.[25]

Spectrographical studies of kerasin have been reported by Dworacek and Pesta.[27] According to them, the graphical presentation of the log of the extinction coefficients of kerasin shows a maximum between 2675-2700 Å and a minimum between 2475-2525 Å. The authors have made use of their findings for tracing amounts of kerasin in the serum of guinea pigs after intraperitoneal injections with this cerebroside.

Teunissen [113] has used colloid-chemical methods to determine the purity of cerebroside and sphingomyelin from beef brain. In an earlier paper Bungenberg de Jong and Teunissen [16] established by electrophoretic measurements the sequence in which salts with positive ions up to a valency of six reverse the electric charge of negative colloids. This sequence of salts is characteristic for certin ionogenetic groups in colloids (phosphate, carboxyl, and sulfate groups). The same authors also described a way to determine the number of ionogenetic groups (charge density) for 1 g of dry colloid by the method of reciprocal hexol numbers (hexol nitrate is a complex 6-valent cobalt compound). Teunissen found that cerebrosides, which according to their structure are devoid of ionogenetic groups, nevertheless show not an infinite hexol number but a definite one, indicating the presence of a small amount of ionogenetic groups (61,000 g corresponding to 1 equivalent). Upon spontaneous oxidation there was a further decrease of the reciprocal hexol number, which could be brought back to the original value by recrystallization. The sequence of the salts reversing the electric charges clearly indicated that the ionogenetic groups in question were carboxyl groups.

Sphingomyelin, which because of its ampholytic character should have an infinite hexol number, gives a value of 76,000, indicating the presence of a small non-compensated negative charge. By comparison with the sequence of salts reversing the charge in egg lecithin the negative group could be identified as the phosphate group, due to a small cleavage in the positive cholin group.

Attempts to determine new physicochemical characteristics were further reported by Hausser,[46, 47] who also studied the dielectric constant of phrenosin. Turner and Watson [117] have studied surface films of phrenosin and kerasin besides sphingomyelin. For each substance the area occupied by the molecule in a closely packed surface film can be compressed to about 42 sq Å. These figures are in good agreement with the supposition that the limiting value is the area occupied by the paraffin chains, set side by side vertical to the water.

According to Lieb [66] and co-workers the carbohydrate group in cerebrosides is responsible for their "chromotropic reaction" (*i.e.*, changing the color of an aqueous thionine solution).

References

1. Adam, N. K., *Proc. Roy. Soc. London (A)* **101**, 452, 516 (1922).
2. Alexander, A. E., and Theorell, T., *Trans. Faraday Soc.*, **35**, 727 (1939).
3. —, —, and Aborg, C. G., *Ibid.*, **35**, 1200 (1939).
4. Alexander, J., "Colloid Chemistry," Vol. I, p. 619, New York, Reinhold Publishing Corp., 1926.
5. Ambronn, H., and Frey, A., "Das Polarisationsmikroskop," Leipzig, 1926.
6. Arnd, O., and Hafner, E. A., *Biochem. Z.*, **167**, 440 (1926).
7. Astbury, W. J., and Lomax, R., *J. Chem. Soc.*, **1935**, 846.
8. Baker, Z., Harrison, R. W., Miller, B. F., and Wexler, R., *J. Expl. Med.*, **74**, 621 (1941).
9. Bamberger, P., *Biochem Z.*, **266**, 175 (1933).
10. Bancroft, W. D., and Rutzler, J. E., Jr., *J. Phys. Chem.*, **35**, 144 (1931).
11. Bauer, *J. Arb. Neurol, Inst. Univ. Wien*, **19**, 87 (1911).
12. Bull, H. H., Cold Spring Harbor Symposia, **8**, 63 (1940).
13. —, and Frampton, V. L., *J. Am. Chem. Soc.*, **58**, 594 (1936).
14. Bülow, M., and Page, J. H., *Hoppe-Seylers Z.*, **205**, 25 (1932).
15. Bungenberg de Jong, H. G., and Saubert, G. G. P., *Protoplasma*, **26**, 282 (1936).
16. —, and Teunissen, P. H., *Kolloidchem. Beih.*, **47**, 254 (1938).
17. —, Verberg, G., and Westerkamp, R. E., *Koll-Z.*, **71**, 184 (1935).
18. Chain, E., and Kemp, I., *Biochem. J.*, **28**, 2052 (1934).
19. Chargaff, E., and Ziff, M., *J. Biol. Chem.*, **131**, 25, (1939).
20. —, —, and Hogg, B. M., *J. Biol. Chem.*, **131**, 35 (1939).
21. —, —, *J. Biol. Chem.*, **140**, 927 (1941).
22. —, —, and Rittenberg, D., *Ibid.*, **138**, 439 (1941).
23. —, —, —, *Ibid.*, **144**, 343 (1942).
24. Chibnall, A. C., Piper, S. H., and Williams, E. F., *Biochem. J.*, **30**, 100 (1936).
25. Crowfoot, C. M., *J. Chem Soc.*, **1936**, 716.
26. Diamare, V., *Atti d. soc. toscan. di scienze natural*, **34**, 24 (1925).
27. Dworacek, E., and Pesta, H., *Wiener klin. Wochenschr.*, **52**, 332 (1939).
28. Eichberg, J., *Ind. Eng. Chem., New Edition*, **19**, 575 (1941).
29. Erb, J. H., *Ice Cream Trade J.*, **37**, 18, 72 (1941).
30. Fabre, R., and Simonnet, H., *Bull. Soc. Chim. Biol.*, **10**, 1306 (1928).
31. Fischgold, H., and Chain, E., *Biochem J.*, **28**, 2044 (1934).
32. —, —, *Proc. Roy, Soc. London, B*, **117**, 239 (1935).
33. Fishler, M. C., Taurog, A., Perlman, I., and Chaikoff, J. L., *J. Biol. Chem.*, **141**, 809 (1941).
34. Folch, J., *J. Biol. Chem.*, **139**, 473, (1941).
35. —, and Schneider, H. A., *J. Biol. Chem.*, **137**, 51 (1941).
36. —, *Fed. Proc.*, **1**, 110 (1942).
37. Freundlich, H., "Kapillarchemie," Leipzig, 1922.
38. —, and Rona, P., *Biochem. Z.*, **131**, 87 (1917).
39. Fries, B. A., Schachner, H., and Chaikoff, I. L., *J. Biol. Chem.*, **144**, 59 (1942).
40. Fujii, N., *J. of Biochem.* **3**, 393 (1924).
41. Gothlin, G. F., *Abh. d. k. Schwed. Akad. d. Wiss.*, **51** (1913).
42. Grün, A., and Limpächer, R., *Ber. chem. Ges.*, **60**, 151 (1927).
43. Handovsky, H., and Wagner, R., *Biochem. Z.*, **31**, 32 (1911).
44. Hansteen-Cranner, B., *Planta*, **2**, 438 (1926).
45. Hattori, K., *Biochem. Z.*, **119**, 45 (1921).
46. Hausser, I., *Sitz. Ber. Heidelberg. Akad. Wiss. Math. Nat.*, Klasse No. 6, 1935.
47. —, Kuhn, R., and Giral, F., *Naturwiss.*, **23**, 639 (1935).
48. Henderson, V. E., *Physiol. Rev.*, **10**, 171 (1930).
49. Hoffmann, F., *Arch. exp. Path. Pharm.*, **16**, 133 (1883).
50. Hooker, M. O., and Fischer, H. M., *Koll.-Z.*, **10**, 283 (1912).
51. Jukes T. H., *J. Biol. Chem.*, **107**, 783 (1934).
52. Keeser, E., *Biochem. Z.*, **154**, 321 (1924).
53. Klenk, E., *Angew. Chem.*, **1934**, 271.
54. —, *Z. physiol. Chem.*, **268**, 50 (1941).
55. —, and Schuwirth, K., *Ann. Rev. Biochem.*, **6**, 115 (1937).
56. Koch, W., *Hoppe-Seyler's Z.*, **63**, 432 (1909).
57. —, and Pike, F. H., *J. Pharm.*, **2**, 245 (1910-11).
58. —, and McLean, F. C., *J. Pharm.*, **2**, 249 (1910-11).
59. Kuhn, R., Hausser, I., and Brydowna, W., *Ber.*, **68B**, 2386-8 (1935).
60. Landsteiner, K., and Pauli, W., *Verh. d. Kong. Int. Med.*, **25**, 271 (1908).
61. Leathes, J. B., *J. Physiol.*, **58**, Proc. VI (1923).
62. Lehmann, O., *Ergeb. d. Physiol.*, **16**, 255 (1917).
63. Levene, P. A., *J. Biol. Chem.*, **24**, 69 (1916).
64. —, and Rolf, J. P., *Ibid.*, **72**, 587 (1927).
65. —, and West, C. J., *Ibid.*, **24**, 41 (1916).
66. Lieb, H. *Wiener Klin. Woch.*, **55**, 465, (1942).
67. Loeb, J., and Beutner, R., *Science*, **37**, 677 (1913); *Biochem. Z.*, **51**, 288 (1913).
68. Maclean, H., and Maclean, I. S., "Lecithin and allied substances," 2nd ed., Longmans, Green and Co., London, 1927.
69. Magistris, H., *Biochem. Z.*, **210**, 85 (1929).
70. Malkin, T., *Trans. Faraday Soc.*, **29**, 977 (1933).
71. Pässler, J., Ger. Pat. 697,440, Sept. 19, 1940 (Cl. 53c, 5); *Chem. Abs.*, **35**, 6688 (1941).
72. Parsons, T. R., *Biochem. J.*, **22**, 800 (1928).
73. Pauli, W., "Colloid Chemistry," J. Alexander, Vol. 2, p. 223 New York, Reinhold Publishing Corp., 1928.
74. —, and Omar, A. M., *Kolloid-Z.*, **68**, 203 (1934).
75. —, and Singer, L., *Biochem.-Z.*, **244**, 76 (1932).
76. Peters, J. P., and Man, E. B., *J. Biol. Chem.*, **107**, 23 (1937).
77. —, and Van Slyke, D. D., "Quantitative clinical chemistry," Baltimore, Williams and Wilkins Co., 1932.

78. Porges, O., and Neubauer, E., *Biochem. Z.*, **7**, 152 (1908).
79. Price, C. W., *Biochem. J.*, **29**, 1021 (1935).
80. —, and Lewis, *Trans. Faraday Soc.*, **29**, 774, 1187 (1933).
81. Rinne, F., *Kolloid-Z.*, **60**, 288 (1932).
82. Remesow, J., *Biochem. Z.*, **218**, 86 (1930).
83. Rona, P., and Deutsch, W., *Biochem. Z.*, **171**, 89 (1926).
84. Rosenheim, O., *Biochem. J.*, **8**, 121 (1914).
85. Schmitt, F. O., Bear, L., and Clark, G. L., *Radiol.*, **25**, 131 (1935).
86. —, and Palmer, K. J., Cold Spring Harbor Symposia, **8**, 94 (1940).
87. Schmitt, W., *Koll.-Z.*, **41**, 261, 334 (1927).
88. Shapiro, S., *Arch. Pathol. Lab. Med.*, **3**, 661 (1927).
89. Singer, E., *Biochem. Z.*, **179**, 432 (1926).
90. Spiegel, E., *Pflügers Arch. ges. Physiol.*, **192**, 225 (1921).
91. —, and Spiegel-Adolf, M., *Am. J. Psychiat.*, **92**, 1145 (1936).
92. Spiegel-Adolf, M., *Kolloidchem. Beih.*, **18**, 275 (1923).
93. —, *Klin. Wochenschr.*, **5**, No. 28 (1926).
94. —, *Biochem. Z.*, **170**, 126 (1926).
95. —, *Biochem. Z.*, **180**, 395 (1927).
96. —, "Colloid Chem.," J. Alexander, Vol. II, p. 303 New York, Reinhold Pub. Corp., 1928.
97. —, "The Globulins," Dresden-Leipzig, 1930.
98. —, *Biochem. Z.*, **252**, 37 (1932).
99. —, *Klin. Wochenschr.*, **11**, 185 (1932).
100. —, *Biochem. J.*, **26**, 2183 (1932).
101. —, *Biochem. J.*, **29**, 2413 (1935).
102. —, *J. Am. Chem. Soc.*, **57**, 1431 (1935).
103. —, *Proc. Soc. Exp. Biol. Med.*, **35**, 263 (1936).
104. —, *Biochem. J.*, **30**, 1536 (1936).
105. —, *J. Gen. Physiol.*, **20**, 695 (1937).
106. —, and Henny, G. C., *J. Phys. Chem.*, **45**, 931, 1941; **46**, 581 (1942).
107. —, —, *Proc. Am. Soc. Biol. Chem., J. biol. Chem.*, **140**, CXXIII (1941).
108. —, —, *Radiologica Clinica*, **11**, 154 (1942).
109. —, —, Unpublished work.
110. —, —, *Fed. Proc.*, **1**, 135 (1942).
111. —, and Spiegel, E., *Proc. Soc. Exp. Biol. Med.*, **32**, 139 (1934).
112. Stanley, J., *Food Industries*, **14**, 69, (1942).
113. Teunissen, P. H., *Rec. Trav. Chim., Pays-Bas*, **57**, 863 (1938).
114. Thanhauser, S. L., and Bluski, J., *Z. physiol. Chem.*, **253**, 217 (1938).
115. Thierfelder, H., and Klenk, E., "Die Cerebroside und Phosphatide," Berlin, 1930.
116. Tönnis, W., and B., 3rd Annual Meeting of the German Soc. of Neurol. and Psychiat., Munich, 20-22. Sept. 1937; *Z. ges. Neurol.*, **161**, 123 (1938).
117. Turner, K., and Watson, M. M., *Biochem. J.*, **24**, 113 (1930).
118. Went, St., and Kuthy, A. von, *Z. Immunitaetsf.*, **82**, 392 (1934).
119. Yokoyama, Y., and Suzuki, B., *Proc. Imp. Acad. (Tokyo)*, **8**, 183 (1932).

Physicochemical Mechanisms in Neuropsychiatric Disorders

E. A. Spiegel, M.D.
Professor of Experimental Neurology

and

M. Spiegel-Adolf, M.D.
Professor of Colloid Chemistry

Temple Universiay • School of Medicine, Philadelphia, Pa.

Nervous and mental disorders may more or less easily be understood as due to deviations from normal functioning of the central nervous system, involving morphologic changes which are macro- or microscopically demonstrable in this organ. In a group of pathologic conditions, however, our present morphologic methods are unable to demonstrate structural alterations. In some of these functional conditions it may be sufficient to assume that they are purely 'psychogenic,' as for instance in hysterical manifestations which may be influenced by suggestion. In the majority of cases, however, the inability of morphologic methods to discover an "organic" basis of a neuropsychiatric disorder should not be accepted as a sufficient proof of psychogenic origin and of the absence of a structural or organic cause of the dis-

turbance. This leads one to the question whether changes in the physicochemical and particularly in the colloid-chemical state of parts of the central nervous system may be demonstrable in some neuropsychiatric disorders and may play a rôle in their genesis.

It is attempted here to discuss (1) what various types of physicochemical and especially colloid-chemical changes may be observed in and correlated to neuropsychiatric disorders; (2) to outline the conditions and factors causing or influencing such changes; and finally (3) to review the methods of investigation which may be of use in such studies.

Types of Changes Observed

Changes of the Cell Colloids. *State of hydration.* Brain swelling[146] may be found in various pathologic conditions: after brain trauma, in acute intoxications, (*e.g.*, uremia) and infections (*e.g.*, diphtheria), inflammatory and vascular processes, brain tumor (particularly gliomas[48, 85, 149, 155, 175]), delirium tremens, status epilepticus, catatonic states in schizophrenics. In a number of cases the underlying pathologic process remains undetermined. The importance of brain swelling lies first in the production of increased intracranial pressure due to the increased volume of the brain, particularly if the swelling affects the whole brain, and secondly in the production of local symptoms if the swelling occurs in certain areas only. The localized form may accompany circumscribed organic processes such as tumors, not only in their vicinity but also in distant parts of the central nervous system, thus complicating the clinical picture and rendering more difficult the local diagnosis. The fluctuating character of the signs and symptoms produced by localized brain swelling and their retrogression on application of dehydrating measures may help to differentiate them from those caused by the primary process.

It is important to distinguish true brain swelling, *i.e.*, increase of volume due to reversible water binding (hydration) of the cellular elements and their processes, from edema, *i.e.*, increase of fluid in the interstitial spaces of the tissue. Macroscopically, one finds in cerebral swelling that the convolutions are flattened and pressed against the skull so that the subarachnoid space is almost completely obliterated and contains scarcely any cerebrospinal fluid. Histologically, glial changes, particularly ameboid cells[8] were found, but they are not specific.[155] Scheinker[149] describes rarefaction and hydration of the ground substance, hydration with swelling of the myelin sheaths, axis-cylinders and glial cells and mild functional vascular disturbances (stasis and diapedesis); cerebral edema shows increase in the permeability of the vessel walls, distension of the perivascular spaces, transudation of serous fluid into the nerve tissue around the blood vessels with resulting liquefaction of the tissue, distension of the pericellular spaces and areolar and sieve-like appearance of the nerve tissue. Both conditions may frequently be combined in the same brain, and some authors even seem inclined to consider brain swelling as chiefly due to increase of extracellular water (see page 1146). The above-mentioned histological criteria are important, therefore, since they may help one to recognize how far true hydration, apart from edema, may be held responsible for brain swelling.

The swelling chiefly affects the white matter and much less the cortical gray. Due to the increased intracranial pressure the cortical gyri are usually flattened. The differences in the swelling of the gray and white matter is explained by W. Toennis[179] by the fact that the cortex contains only 0.1 per cent sphingomyelin, whereas the white matter contains 8.9 per cent of this substance, which is said to have the highest hydration capacity of all the phosphatides.

Very little is known about the genesis of brain swelling. A common denominator for the great variety of heterogeneous conditions in which brain swelling was found seems the rather acute destruction of nervous tissue within diffuse areas of the central nervous system. Such a view is corroborated by the appearance of the so-called

ameboid glia (Alzheimer), a type of glia cell which shows that noxious agents are present inducing a proliferation of glia cells as well as injuring these proliferating cells. The possible rôle of the destruction of white matter is indicated by the observation that the birefringence of the myelin sheaths decreases as early as 24-48 hours after severance of a myelinated nerve.[158] Since this decrease of the birefringence is reversible, at least in its initial stage, and a similar reversible decrease is observed in hydration of the myelin sheath,[159] it seems not improbable that a swelling of the myelin sheath takes place in the beginning of the Wallerian degeneration. Such a view is supported by the fact that the water content of degenerating nerves [113, 127] or central nerve tracts [95] is increased. Similar findings were reported for brain tissue in various pathologic conditions with degeneration of nervous tissue associated with decrease of lipoids such as general paresis,[93] dementia praecox,[96, 142] and pellagra.[97, 98] The breakdown of lipoids accompanying degeneration of nervous tissue seems to play an important rôle in the genesis of brain swelling. There appear increased amounts of hydrosoluble phosphorus in degenerating nerves [114] probably partly because of the disintegration of nucleoproteins and phosphatides that leads to the formation of phosphoric acid; the appearance of this acid is able to increase the state of hydration of proteins (*e.g.* of globulins [170]).

Another mechanism is indicated by the finding of Weil [185] that the benzene-soluble lipoids inhibit the swelling of proteins. Thus the diminution of lipoids in degenerative processes may favor the swelling of proteins. A further factor to be considered is the accumulation of cleavage products that may facilitate the swelling. According to Weil,[185] and de Crinis,[32] such a rôle may be played by urea, which increases the hydration of gelatin and fibrinogen.[42, 73] de Crinis, using the xanthydrol method, found increase of urea in the cerebral tissue in brain swelling due to various diseases. However, it is not proved that an increase of the water-binding power of the cell proteins is the decisive factor in the mechanism of brain swelling, since this process chiefly affects the white substance in which, according to Fraenkel & Linnert,[44] 74.8 per cent of the dry substance is lipoids. Increase of the permeability of the cells and their processes under the influence of toxic or other injurious agents may be an additional factor. It should also be emphasized that osmotic processes, apart from pure hydration, may play an important part in the mechanism of brain swelling. Such osmotic processes may come into action because of the accumulation of cleavage products in the tissue on the one hand, and to the decrease of water-binding substances (*e.g.*, proteins) in the blood plasma on the other.

Changes opposite to those observed in brain swelling, *i.e.*, shrinkage of the protoplasm due to a diminished water-binding capacity (synaeresis), may play an important part in normal and in pathologic or premature senesence of the central nervous system, as well as of the organism in general.[148a] As pointed out by Braunmuehl,[17a] such processes may cause, for example, the appearance of the so-called senile plaques or of the changes of the neurofibrils described by Alzheimer.

State of dispersion. Bancroft [8, 9] and his associates tried to establish a relationship between the state of dispersion of the brain colloids and mental disorders. He assumed that these colloids are slightly "overcoagulated" in the one group (manic-depressive psychoses) and overpeptized (over-dispersed) in the other (schizoid insanity). He supports his theory by experiments attempting to show that coagulating agents produce narcosis, whereas a peptizing agent such as sodium thiocyanate has an antagonistic effect to anesthetics. Henderson and Lucas,[74] however, failed to find a specific antagonism between sodium thiocyanate and several narcotic agents, and emphatically reject a coagulation theory of narcosis. Similarly Leake [102] was unable to confirm an antagonism between sodium thiocyanate and morphine. Harris and Katz [70] tested the effects of sodium amytal and sodium rhodanate, these being drugs which Bancroft recommended as coagulating and peptizing agents, respectively. While sodium amytal may have a transient ameliorating action on mental patients,

sodium rhodanate was without noticeable therapeutics effect. Contrasting effects between these two drugs having diagnostic value could not be found.

In view of the difficulty of obtaining direct information about the state of the brain colloids, serological studies are of interest; for they may perhaps shed some light also upon the colloids of the nervous system. Ludlum and Zozaya,[109] studying sera of mental cases, distinguish four groups of chemical patterns: (1) normal distribution of the different protein fractions, found in normal individuals, chronic demented cases, and diseases of purely psychological nature; (2) increase of the euglobulin fraction, found in acute psychoses produced by bacterial invasion and associated with disturbances of water interchange between blood and tissues; (3) increase of pseudoglobulin found in intoxications; and (4) increase in albumin observed in schizoid constitutionally inferior individuals, in whom Krause [100] reported globulin increase. In similar studies Frisch,[47] and Georgi,[53a] observed in epileptics, preceding the attack, an increase of serum protein, due to an increase of albumin, "the finely dispersed fraction" while the "coarsely dispersed" fractions showed normal values. In reviewing these studies one has to bear in mind that deviations in the proportions of the various protein fractions of the serum can be found in many conditions without mental or nervous disturbances, *e.g.*, increase of euglobulin in luetics,[128] increase in pseudoglobulin in experimental active immunization,[9a] increase in albumin * in experimental exposures to cold [84] and under the influence of high altitudes.[45a]

It seems, therefore, that other factors still unknown (perhaps in addition to the above-mentioned changes in the protein fractions) must be held responsible for the genesis of mental disturbances.

In a study of the precipitation of the plasma colloids by NaCl, Georgi found an abnormal lability of the plasma colloids (without change of the sedimentation rate of the red blood corpuscles) preceding spontaneous epileptiform seizures or during hyperventilation. He assumes that this abnormal lability is due to changes of ion-equilibrium as observed in hyperventilation. Since similar changes may appear in conditions associated with disturbances in the balance of the autonomic nervous system, it seems rather improbable that this abnormal lability of the plasma colloids is specific for convulsive disorders. Furthermore, it is not proved that the increased lability of the plasma colloids is able to influence the convulsive reactivity of the central nervous system. The changes in the state of the plasma colloids and those in the excitability of the central nervous system appearing in hyperventilation may have a common cause, such as disturbances in ion-equilibrium, but do not necessarily influence each other.

Changes of the Cellular Surface Films. The existence of a semipermeable membrane or surface film in nerve cells and fibers is indicated by the observation that the intact nerve cell or fiber is the seat of an electromotive force, a "current of injury" flowing from the intact to an injured part. According to the membrane theory, the cellular surface is semipermeable, allowing only some ions (*e.g.*, cations) to leave the interior of the cell,[18, 132] so that a Helmholtz electric double-layer (*e.g.*, cations outside, anions inside) exists on the surface. Various stimuli induce excitation by changing the ion concentration on these surface films.[129] These changes in ion concentration act upon the cell membrane by transitorily increasing its permeability [79, 107] and thus abolishing its potential in the stimulated region, so that an "action current" may be demonstrated flowing from a resting to an excited region. The local action currents between the depolarized region and its normal neighborhood act as stimuli to the next segments, and thus the nerve impulse is propagated by depolarization of successive segments of the nerve. In agreement with this

* According to E. Schiff and E. Roser [*Monatschr. f. Kinderheilk*, **19.** 15, (1920)] in newborn infants during the first days of life the albumin content of serum runs up to 90 per cent of the total protein.

theory, Cole and Curtis [27] observed on single unmyelinated fibers of the squid, that the electrical conductance increases during the passage of an impulse. Spiegel and Henny [160a] found that the convulsions following faradic stimulation of the brain are associated with decrease of its impedance (experiments on cats under superficial ether anesthesia). A similar change could also be demonstrated following single induction shocks applied to isolated, surviving frog's brains, where the disturbing influence of the circulation was eliminated, while such a change failed to appear in control experiments on dead brains.

Thus, the state of the surface films of the nerve cells and their fibers and their reactivity to the depolarizing influence of stimuli may be of decisive importance in determining the excitability and the conduction of impulses. It hardly needs to be emphasized that the state of the cellular surface and that of the cellular colloids are closely interrelated: on the one hand the permeability largely depends on the state of the colloids on the cell surface, and on the other hand changes of the permeability will affect the colloids in the interior of the cell.

Changes of the permeability of the ganglion cells seem to play an important part in the pathogenesis of convulsive disorders. Hypothetical views in this direction were expressed by various authors.[58, 116, 157, 177] Inferences drawn from observations on serum colloids (Georgi) upon the state of the nerve cell membranes can be of hypothetical nature only. Since the effects of injection of dyes (*e.g.*, fuchsin,[177]) into the circulation primarily depend upon the permeability of the capillaries, the results of injection experiments hardly permit one to evaluate the state of the cellular surface films. Spiegel and Spiegel-Adolf,[162, 167, 168] therefore, tried to gain somewhat more direct information by taking advantage of the fact that polarization phenomena developing on semipermeable cell surface membranes depend upon the density of these membranes and may serve as an indicator of their permeability (see p. 1150). They demonstrated that in a number of apparently quite heterogeneous conditions causing convulsive disorders, decrease of the density of the cellular surface films represents a common basic mechanism. Such a reversible decrease could be demonstrated for conditions inducing hydration, *e.g.*, swelling of the brain *in vitro* (in hypotonic solutions, in alkali), or *in vivo* (intravenous injection of distilled water,* $NaHCO_3$, and NaOH respectively, alkalosis produced by hyperventilation). Such an effect can easily be understood if one bears in mind that swelling loosens the structure of a membrane by intake of water between its molecules or micelles.[89] The increase of the convulsive reactivity in conditions associated with swelling of the brain may partly be due to increase of intracranial pressure (see p. 1137). A decrease of the polarizability, however, could be found when swelling of the brain was induced by intravenous injection of distilled water with the skull trephined and the dura opened, so that increase of the intracranial pressure was prevented. When the skull remains closed, *e.g.*, in the convulsions appearing after administration of excessive amounts of water,[148] one has apparently a combined effect of hydration (plus increase of interstitial water) and increase of intracranial pressure upon the cell membranes. Similar conditions may prevail in certain cases of epilepsy, as shown by the beneficial effect of measures subtracting water from the body-tissue reservoirs,[38] and by the fluctuations of weight indicating retention of water before seizures [49] in a group of epileptics. The importance of water balance in some epileptics was strikingly demonstrated by McQuarrie and Peeler,[117] who could induce convulsions in epileptic patients, but not in non-epileptic control subjects, by administering antidiuretic posterior pituitary extract (pitressin) during periods when the mineral intake was low while the water intake was unrestricted, so that a dilution of the extracellular body fluids resulted. Phasic disturbances of the water balance were also described in periodic psychoses.[180]

A second group of epileptogenous agents, which impair the cellular surface mem-

* See page 1143.

branes in the cortex as well as in subcortical ganglia, is represented by those producing anoxemia, *e.g.*, experimental asphyxia. Brain anemia produced in most of our experiments an initial drop of the polarizability. When the anemia lasts for several minutes, the density of the membranes may again increase, probably because of coagulation necrosis of the cells.

The impairment of the cell membranes produced by increase of intracranial pressure probably also belongs in this group, since the increase of pressure is associated with cerebral anemia. In a number of instances, *e.g.*, in various conditions associated with brain swelling (see p. 1137) there is probably the combined effect of hydration and increased intracranial pressure, due to the increase of brain volume with the skull closed. Anesthetics and hypnotics may act in the opposite direction as the epileptogenous agents, *i.e.*, by rendering the cellular surface films more dense as could be demonstrated for ether or chloroform, intravenous injection of dial, or intraperitoneal injection of phenobarbital (see p. 1143).

It should be emphasized that decrease of the density of the cell membranes, as produced by hydrating measures, or by impairment of the oxygen supply to the brain, represents only one of the mechanisms by which epileptogenous agents act. A second mechanism is the local increase of ion concentration; if it is large enough it may induce convulsions, although the cell surfaces are normal. This is exemplified by artificial acidosis which may be associated with convulsions,[86, 148] although according to our measurements it impairs the cell membranes only slightly, if at all. A similar mechanism may play a rôle in the genesis of hyperreflexia and of convulsions observed in dogs following intravenous injections of concentrated NaCl solutions.[128a, 130a]

Measurements of the polarizability in the exposed brain of patients during operation seem to indicate that the experimental results obtained may be applicable also to the human brain. We found, for instance, low values in a case of vascular hypertension with increased intracranial pressure, and also in two epileptics. But the material is too scant as yet to permit definite conclusions. The fact that convulsions can be induced in epileptics but not in normal subjects by procedures that will result in dilution of the extracellular body fluids is interpreted tentatively by McQuarrie [115] as indicating the existence of an inherent defect in the brain cell membranes in this disorder.

In dealing with patients who are repeatedly subjected to convulsions, one has to bear in mind that the repeated abnormal discharges of the brain cells and the associated vasomotor reactions may by themselves injure the brain cells and their membranes. In a study of the effects of convulsions induced by injection of metrazol, it was found that one single series of seizures of from 5 to 30 minutes' duration was able to diminish the polarizability of the cerebral hemispheres in guinea pigs.[164, 165] After hypoglycemic shock with convulsive seizures lasting several hours the effect was still more marked.

An impairment of the cells of the human central nervous system following metrazol and insulin convulsions may be inferred from studies of the spinal fluid. Spiegel-Adolf and Freed [174] found in the spinal fluid of schizophrenics the ratio of the interferometric values of non-electrolytes to electrolytes (1NE : 1E) higher after than before such convulsions; they interpreted these findings as indicating an increase of organic cleavage products in the spinal fluid following the convulsions. Similar changes were found following electrically induced convulsions in dogs.[174a]

In earlier observations of Spiegel-Adolf [178] abnormally high values of this ratio were found in 60 epileptics, representing the only demonstrable pathologic change of the spinal fluid in this disease. In agreement with the assumption that this increase of the 1NE:1E ratio was due to an impairment of the cells of the central nervous system secondary to the convulsions, this ratio was the higher the more severe and/or the longer the convulsions were.

Conditions Influencing the State of the Colloids and of the Surface Layers of the Nerve Cells

Changes of Hydrogen-ion Concentration. In reviewing the effect of disturbances of the acid-base equilibrium we must bear in mind that in the central nervous system we are dealing with effects on protein-lipoid mixtures, while in other organs the effects are preponderantly protein reactions. While alkalies influence the swelling of proteins and lipoids in a similar manner, causing up to certain concentrations a steep increase in viscosity and hydration, the action of acids upon the state of hydration of proteins and that of certain lipoids is antagonistic: up to certain concentrations they increase the viscosity and state of hydration of proteins but decrease those of lecithin and cephalin;[68, 144, 169] the water-binding capacity of sphingomyelin is, however, pronounced in an acid milieu, according to Toennis.[179] Thus one can easily understand that alkalies distinctly facilitate the hydration of brain tissue, as could be shown *in vitro* by gravimetric methods [11] and *in vitro* and *in vivo* by the associated decrease of the polarizability,[162, 168] whereas acids further swelling only slightly or not at all.[11, 12, 181] Bauer even assumed that acids inhibit swelling; his results were, however, at least partly due to post-mortem changes.

Recent experiments in which the pH was varied by phosphate buffer mixtures [152] showed a minimal increase of volume of rabbit's brain at the neutral point (pH about 7.40); the volume increase reached a maximum in the alkaline range (pH 7.6-8.3) within 2-3 hours, in the acid range (pH 6.00-7.30) beyond the 6th hour. After the 6th hour neutralizing processes in the tissue induced a retrogression of the volume increase. Using the diminution of the birefringence of the myelin sheaths as an indicator of the swelling of nerves, Spiegel [159] studied the effect of acids not only post mortem but also on surviving frogs' nerves. He found that addition of acids to diluted Ringer solution furthers the swelling only in concentrations that already impair the excitability. When acidosis was produced *intravitam* in frogs by endogenous production of acids (Fischer's [40] method of uranyl nitrate intoxication), or by oral administration of acids in rabbits, the study of the birefringence failed to show a swelling of the myelin sheaths. In agreement with these findings, the decrease of the polarizability under the influence of acids was much less marked than that induced by alkalies [162, 168] and production of acidosis *in vivo* failed to induce a definite decrease of the polarizability.[168] The swelling of the brain and decrease of its polarizability produced by alkali may be inhibited by immersion of the brain in acids; partial restoration of the former values may even be produced, particularly by acetic acid and tricarballylic acid. Also *in vivo* the changes induced by an alkaline shift are reversible.

Effect of Other Ions. Swelling of brain tissue may be inhibited by cations in the order Ca, Ba, Sr, > Na, Li, > K, Cs, > Rb; and by anions in the order citrate > tartrate > oxalate, sulfate > acetate > CNS > Br > NO_3 > Cl > I.[65] Similar findings were reported by Ludlum *et al.*,[108] who observed also a shrinking effect of aluminum and magnesium chloride. In agreement with Haldi's experiments on the effect of ions on water absorption of brain tissue * Spiegel and Spiegel-Adolf found that sodium inhibits the decrease of the polarizability of brain tissue * more than does potassium, whereas lithium was more effective than sodium. Calcium inhibited the decrease of the polarizability as compared with sodium, and increased the density of the cell membranes in a swollen brain. Magnesium, however, had no inhibitory effect upon the decrease of the polarizability. In the group of the trivalent cations, lanthanum had a distinct inhibitory effect upon the decrease of the polarizability as compared with sodium, but aluminum had no such effect. After previous swelling of the brain its polarizability increased in NaBr solution more than it did in equivalent NaCl solution, contrary to the Hofmeister series, but in agreement with Haldi's experi-

* As to the effect of various ions on the polarizability of artificial membranes, see 171.

ments. The effect of bromides may be explained by observations of Spiegel-Adolf [172] showing that the viscosity and refraction of lecithin-sols are more depressed by addition of bromides than by equal concentrations of chlorides. In accordance with Hofmeister's series, the polarizability decreased in NaI and NaCNS more than it did in equivalent NaCl solution.

Intravenous injection of hypotonic solutions or distilled water produce cerebral edema [183, 184] and eventually also hydration of the nerve cells [39] associated with decrease of the polarizability.[162, 168]

Anesthetics and Hypnotics. It is rather generally assumed that anesthetics and hypnotics act by changing the physicochemical state of the cells, particularly of their lipoids.[128, 188] Also the hypnotic efficiency of the barbiturates seems to be related to their reaction with lipoids.[178] An important part in the mechanism of anesthesia is played by changes of permeability of the cell membranes. Such changes were studied by numerous investigators.[52, 59, 60, 79, 107, 131, 186] A majority of them found diminution of the permeability under the action of the anesthetics on plant cells, eggs, and various tissues (skin, muscles, and peripheral nerves), as long as no irreversible injury to the cells with subsequent increase of the permeability was produced. Using polarizability as a measure of permeability of cell membranes, Spiegel and Spiegel-Adolf [163] extended such studies to the action of anesthetics and hypnotics upon the central nervous system *in vivo*. They found that ether, chloroform, chloral hydrate, and barbiturates increase the polarizability, indicating an increase in the density of the cellular surface films. This action is not limited to certain regions, since "cortex anesthetics" [124] such as ether, chloroform, and chloral hydrate have an effect upon the brain stem too, and a "brain stem anesthetic," such as dial, also acts upon the cerebral hemispheres. Because of the densifying action of anesthetics the cell membranes will become more able to resist the depolarizing effect of stimuli, which is an essential part of the excitation process (see page 1139), and thus the excitation threshold will be raised.

The finer mechanism by which anesthetics act upon the cell membrane escapes direct observation. The myelin sheath is, however, an object which permits one to study the effect of anesthetics upon surviving and still excitable nerves under the polarization microscope. It could be shown that the birefringence of the myelin sheaths of surviving frogs' nerves is diminished or abolished under the influence of anesthetics.[160] This is, at least under physiologic conditions, a reversible process; if one stops the anesthesia, the birefringence returns nearly simultaneously with the excitability. By using various concentrations, *e.g.*, of ethyl or methyl alcohol, it could be demonstrated that the effect of the anesthetics upon the myelin sheaths is antagonistic to that of water. It was inferred that narcotics try to displace the water that is in loose combination with myelin. As was pointed out by Winterstein, such a displacement of water may be explained by adsorption of the anesthetics to the cell colloids, which become surrounded by the narcotic. Such an adsorption to the lipoids will diminish their surface tension.[15, 23, 180] This seems an important factor in the mechanism of the above-mentioned changes in the birefringence of the myelin sheaths. The optical axes of the myelin sheath normally lie perpendicular to its surface and quickly regain this position after the anesthetic is withdrawn, as shown by the return of the birefringence. In view of the relationship of the optic axes to the surface of the myelin sheath, it does not seem far fetched to relate their arrangement to the action of surface-tension forces. Conversely, it may be inferred that anesthetics (in decreasing the surface tension of the lipoids) affect the intermicellar forces which are related to the surface tension. It seems of interest in this connection that Handovsky [66] explains the changes in the x-ray diffraction pattern observed under anesthesia as due to changes in the arrangement of fatty acid molecules.

Winterstein [186] assumes that lipoids are no more important than other cell colloids in the mechanism of anesthesia, while others [26, 107] emphasize the fundamental rôle

of the lipoids. It may, therefore, be mentioned that the index of polarization which we found changed under the action of the anesthetics, namely, the difference in conductivity at low and at high frequencies, could be imitated in studies on artificial biocolloid membranes,[171] but only by those membranes which contain lipoids (kephalin, lecithin) in fine dispersion.

The concept formed as to the action of the anesthetics on the lipoids of the surface layers of the nerve cells depends, of course, upon the notions one has regarding the structure of the cellular surface. Clowes assumes the existence of a balanced mixture of water-in-oil and oil-in-water emulsions. Anesthetics such as ether, chloroform, chloral hydrate, barbital, etc. when shaken with water and olive oil, give rise to droplets of water in oil (probably by lowering the surface tension of the oil more than that of water), and do not produce droplets of oil in water.[78] Water-in-oil formations also occur, according to these authors, when analgesic and antipyretic drugs (acetanilid, acetphenetidin, or acetylsalicylic acid) are used. Thus, Hirschfelder[77] assumes that the predominance of the water-in-oil emulsion seems to correspond to the "anesthetic and analgesic phase" in the surface layer of the nerve cells. In accordance with this view emulsions rich in the water-in-oil type are relatively impermeable to water and salts, as indicated by their low electric conductivity. Thus, the densifying effect of anesthetics may be well explained on the basis of this theory. It also gives an understanding of the magnesium-calcium antagonism discovered by Meltzer and Auer;[120] the anesthetizing magnesium salts tend to increase the water-in-oil phase of the emulsion in the cellular surfaces, while antagonistic calcium salts restore the normal proportion of the oil-in-water phase. Further experiments of Hirschfelder indicate that monovalent cations, particularly potassium, antagonize in a similar way the narcotic action of magnesium.

If we assume that the surface of the cells is formed by a continuous layer of lipoids such as lecithin and/or kephalin,[69, 94, 107, 111] we may also understand the effect of anesthetics by their action upon the lipoids. We must bear in mind that such a layer of phosphatids would also permit the passage of water and salts under normal conditions, since these lipoids are able to take up water and to combine with cations (Koch), the lipoid particles probably being hydrated, *i.e.*, surrounded by a layer of water. If the anesthetic becomes adsorbed on the lipoid particles, it will displace this water and a decreased permeability for water and electrolytes will result. The above-mentioned calcium-magnesium antagonism, however, can better be explained by the concept of Clowes and Hirschfelder.

The action of the anesthetics will, of course, not be limited to the cell surfaces; they will also affect the state of the colloids in the interior of the cell. Winterstein assumes that the enveloping of the cell colloids by the adsorbed anesthetic inhibits the intake and discharge of metabolic products, and the action of ferments. Thus, the effect of anesthetics upon the metabolism of the cell, *e.g.*, the inhibition of the oxidative processes (oxidation of glucose, lactic acid, and pyruvic acid[29a, 144a]), may be explained by the same basic reaction with the cell colloids as the increase in density of the cell surface.

Deficient Oxygen Supply. Disturbances of circulation associated with reduced oxygen supply may influence the state of hydration of the brain colloids, since products of metabolism such as lactic acid accumulate in the brain tissue and induce a change in its pH. While in anoxia due to reduced cerebral blood flow the cortical pH shifts in an acid direction, in anoxia resulting from breathing nitrogen there occurs a shift in an alkaline direction.[176] As we pointed out, alkaline reaction is conducive to hydration of brain tissue and increased permeability of the cell membranes, while acid reaction favors swelling of brain tissue to a lesser extent, if at all. Thus, one may understand that anoxic anoxia decreases the density of the cell membranes as indicated by measurements of the polarizability,[168] while the changes observed in cerebral anemia are more complex. In the majority of our experiments, at least

an initial diminution of the polarizability was observed; yet the polarization index of the cell membranes usually rose again during the occlusion of the cerebral vessels. One may perhaps infer that the local acidosis eventually induces a coagulation necrosis of the nerve cells. Histological observations seem to point in a similar direction, since Gildea and Cobb [58] report that shrinkage of the nerve cells and homogeneity of the protoplasm is the most frequent change following ligature of the cerebral vessels. For peripheral nerves the importance of oxygen supply for the state of polarization of the surface membranes may be inferred from the observation [54] that the membrane potential falls progressively during anoxia.

Neuropsychiatric disorders may be caused not only by insufficient oxygen supply of the central nervous system, but also by deficiencies in the ability of the cells to utilize the oxygen supplied. Deficiencies in oxidative processes in the cerebrum seem to play an important part in schizophrenia. This is indicated by the low content in neutral sulfur [96] and in catalytic iron; Freeman [45] found the latter deficient, particularly in the deepest layers of the frontal cortex in catatonics. The disturbance seems not to be limited to the cerebral cortex, however, as shown by the relatively poor reaction of the respiratory center to administration of CO_2 in the inspired air,[62] and the low oxygen consumption, and nitrogen and carbohydrate metabolism [188] of some schizophrenics.

Endocrines. *In vitro* experiments of Haldi [65] and associates indicate that thyroxin does not alter the swelling ability of the cerebral hemispheres, while suprarenine bitartrate solutions seem to have a slight inhibitory influence. Page,[134] however, doubts whether these observations may be applied to conditions *in vivo*. In thymectomized dog's brain swelling was observed,[92] which Liesegang [106] assumed was due to the associated acidosis. Such an explanation, however, seems improbable in view of the above-mentioned experiments on the influence of acidosis on the state of hydration of the brain. The hypoglycemia induced by insulin injection seems to be associated with hydration of the cells of the central nervous system.[190]

Vitamins. The importance of hydrogen-ion concentration and of oxidation processes for the state of the brain colloids may also shed light upon the action of some vitamins. Vitamin B_1 deficiency seems to be associated with defective ability of brain tissue to oxidize lactic acid particularly.[76, 139, 140] Vitamin A and its precursor carotene may play a role in the oxidative processes in medullated nerves, but the results of various investigators [43, 125] are controversial. Vitamin C seems to act as an oxygen carrier and to promote tissue oxidation. Wortis [187] points out that the high respiration rate of brain tissue probably has some connection with its relatively high ascorbic acid content. In experimental scurvy ascorbic acid diminishes in the brain; this may induce cellular anoxemia which may partly account for convulsions in late scurvy (Wortis). Changes in the chemical composition of the cerebral hemispheres and of the cerebellum were found in scurvy by Palladin and Ssawron,[135] in that the water and total protein content was increased and the phosphorus content decreased. The importance of vitamin D for the state of the central nervous system is probably due to its influence upon the calcium metabolism, the calcium content of brain tissue decreasing 25-35 per cent of its normal value in rickets (Wortis).

Methods of Investigation and Their Application

Gravimetric Methods (Comparison of Brain Volume and Skull Capacity). The volume of the brain may be ascertained by determination of its weight and its specific gravity, and compared with the capacity of the skull, measured by filling the skull with water.[146] According to Reichardt, in normal adults the brain is 10-16 per cent smaller than the skull capacity; in children the brain volume is relatively larger. If an increase of the brain volume in relation to the skull is found, one must of course ascertain whether it is due to hyperemia, inflammatory processes, tumor, edema, or true swelling.

If one wants to ascertain the ability of the dead central nervous system or parts of it to swell in various fluids, the increase of its weight after immersion for some time in the test fluid may be used as an indicator.[11] Weil first dries the brain tissue in a vacuum desiccator at 60–80°—a procedure which does not change its water-binding power—and determines the difference in weight of the fresh and of the dried substance. This gives the water content of the organ. Then he immerses the dried substance in distilled water. After swelling it is weighed and again dried in the vacuum desiccator. The difference of weight before and after drying indicates the water-binding power.

Chemical Methods. While the determination of the difference in the weight of the fresh and of the dried organ yields information regarding its total water content, it fails to tell what part of the water is intracellular and what is extracellular (in interstitial spaces). This question becomes of importance if one wants to know whether an increase of the water content of the brain is due to true swelling (increase of intracellular water) or to edema (increase of interstitial water).

Determination of the electrolyte content, particularly of the sodium, potassium, chloride and phosphate content, may help to solve this problem and also yield information regarding the selective impermeability of the cellular surface films. Perfusion experiments on isolated muscles,[126] as well as experiments *in vivo*,[72] indicate that muscles easily give off their sodium and chloride but retain nearly all their potassium and phosphate. The view seems, therefore, not unjustified that sodium and chloride are contained chiefly in the intercellular fluids, and potassium and phosphate in the intracellular fluids, and that the normal cellular membranes have a high selective impermeability to these ions. Such a concept was extended to the body in general.[51, 71, 138] Since electrolytes are largely responsible for maintaining the osmotic pressure of the body fluids and consequently also for the distribution of the water, changes in the volume of the intra- and extracellular fluids may be estimated in the following way, based on the assumption that all brain chloride is extracellular.[29, 71]

The volume of the extracellular water is calculated by dividing the total chloride by the concentration of chloride in the extracellular water; the latter is estimated by determination of the chloride concentration in the serum. The difference between the total water and the extracellular water gives the volume of the intracellular water. The product of the volume of the extracellular water and the concentration of potassium (or sodium) in the extracellular water (estimated from their concentration in the serum) gives the amount of the extracellular potassium (or sodium), and the difference between total and extracellular potassium gives the intracellular potassium. The latter value divided by the volume of the intracellular water equals the concentration of potassium in the intracellular water.

By using this principle, Yannet [190] arrived for instance at the conclusion that hypoglycemia first induces hydration and later shrinking of the cells in the central nervous system. In using the determination of the electrolytes for a study of the distribution of water in the central nervous system, one is of course faced with the problem whether the experiments on other organs, particularly on muscle, may be applied to the central nervous system. Amberson and associates [4] found that cerebrum, cerebellum, and spinal cord retain their chloride tenaciously (contrary to peripheral nerves and other organs), if the plasma chloride is reduced to as low as 6 per cent of normal. They conclude that in the central nervous system a considerable fraction of chloride is intracellular. According to Yannet,[189] however, the brain cells adjust themselves to the depletion of extracellular electrolytes by releasing intracellular potassium in quantities proportional to the decrease in the concentration of extracellular sodium. He considers his data following depletion of extracellular electrolytes consistent with the view that all sodium and probably all chloride in the brain is diffusible. He explains the contradictory findings of Amberson and as-

sociates by their failure to allow adequate time for equilibrium to occur between brain and plasma before removing tissue for analysis. In agreement with the concept that in the central nervous system potassium is chiefly intracellular are the findings in cerebrospinal fluid, which has a composition similar to that of interstitial fluid; it contains rather little potassium (average, according to Mestrezat,[121] 20.7 mg per cent) compared with its relatively high sodium content (average 322 mg per cent).

Optical Methods. Such methods are of particular interest, as they permit the study of the more intimate structure of the nerve cells and fibers in living or surviving preparations without the use of fixation or of stains, *e.g.*, examination with the ultramicroscope or under polarized light. After ultramicroscopic studies Auerbach [6] and Marinesco [112] asserted that the axoplasma does not contain continuous fibrils, and that the finding of neurofibrils is an artifact produced by fixation methods. At the most, very small particles may appear under the ultramicroscope in the axoplasma arranged in longitudinal direction. Under the flocculating action of certain electrolytes a neurofibril structure may, however, appear.[87] Thus one may regard the living substance of nerve cells and fibers as consisting of a sol with rod-like particles and a tendency to longitudinal orientation of these particles.[137] On quite fresh material Peterfi was also unable to see these particles under the ultramicroscope; they may, however, appear and orient themselves in the longitudinal direction and form fibrils, if the preparation stands for some time, which reminds one of ageing sols. In the state of degeneration of the axis cylinder the Tyndall phenomenon may be seen (Marinesco). The use of the ultramiscroscope in the study of pathologic swelling conditions in fresh brain tissue was recommended by Ludlum, Taft, and Nugent,[108] who observed by this means that less hydrous material has more definite edges and a more granular appearance than "wetter matter." The influence of various agents, *e.g.* the coagulating effect of anesthetics upon the colloids of the axoplasma, may be also demonstrated by the ultramicroscope (Auerbach). In the nerve cells (spinal ganglia) granules could be seen under the ultramicroscope; these are, however, not identical with the Nissl bodies (Marinesco). Peterfi was also unable to find Nissl granules.

Dark-field illumination may further be useful for a study of the distribution of minerals in the elements of the nervous system under normal and pathologic conditions on preparations obtained by microincineration.[2, 105, 151] The results of microincineration may be supplemented by quantitative measurements with the aid of emission spectra. Alexander and Myerson determined the iron by the increase of the intensity of the iron line produced by the addition of a certain amount of iron to a weighed amount of the sample. By relating the intensity of the iron lines to those of the other elements studied, they were able by reference to calibration curves to estimate the amount of the element present in the sample.*

The study in polarized light centers chiefly about the myelin sheath in myelinated nerves.

The myelin sheath has a positive birefringence, the optical axes lying perpendicular to the longitudinal axis of the nerve. On a transverse section they show a radial arrangement. Thus the optical axes upon which the microscopic picture depends are perpendicular to the longitudinal axis of the nerve, simulating a negative one-axis birefringence in the direction of the longitudinal axis.[5, 35] The axoplasma has a positive one-axis birefringence, as can be seen in unmyelinated nerves or after extraction of the myelin of a myelinated nerve.

The birefringence of the myelin sheaths is due to their glycerophosphatides.[61]

* The absorption spectrum of non-medullated nerves (the claw nerves of Homarus and Limulus) examined with the spectromicroscope presents two distinct bands characteristic of hemochromogen compounds; the hemochromogen present in these nerves does not react to reagents as does cytochrome. Medullated nerves (sciatic of green frogs) show two wide bands of the carotinoid pigments superimposed upon the general absorption of the phosphatids.[125]

Under the influence of hydrating agents and of lipoid-soluble anesthetics, the birefringence is decreased.[159, 160] These are, at least in the beginning, reversible processes. Of practical importance is the observation that the birefringence of the myelin sheath diminishes in the beginning of Waller's degeneration,[7, 153, 154, 158, 182] since this change permits one to demonstrate injuries to the myelin sheath at an early stage. This diminution of the birefringence is at first also reversible; it precedes the chemical breakdown of the myelin.

The assumption that the birefringence of the myelin is due to a crystalline structure[61] is supported by the x-ray diffraction pattern of myelinated nerves (see page 1149). In order to explain the typical arrangement of the optical axes of the myelin sheath perpendicular to its longitudinal axis, it seems necessary to believe that radial forces arranging the submicroscopic crystals exist in the myelin sheath, and that these forces are at least partly related to the surface tension. As was pointed out elsewhere,[161] the reversibility of the decrease of the birefringence produced by hydration or by anesthetics can hardly be explained without the action of surface-tension forces. The decrease of the birefringence of the myelin sheath following severance of the axis cylinder remains unexplained, if one does not take into account the mutual relationship ("zygiosis") of surface tension in the axoplasma and myelin sheath. A possible relationship between birefringence and electrostatic forces has been suggested by Keller.

While the disintegration of the axoplasma following severance of a nerve may be explained by the loss of its connection with the cell body and its nucleus, the breakdown of the myelin sheath can not be understood on this basis, since it consists of segments separated by Ranvier's rings. In view of this fact the following physicochemical explanation has been offered.[158] The differentiation of the neuron into a cell body and processes indicates the existence of differences in surface tension between the various parts of the cell. Cutting the nerve abolishes the action of these differences of surface tension, and the plasma of the axis cylinder tends to return to the spherical form. The relationship between the axis cylinder and the myelin that normally forms a cylindrical sheath around it may be compared to that of two immiscible fluids, A and B. According to Freundlich,[46] a drop of B will spread on the surface of A, if A (*i.e.*, the axoplasma in our case) has a greater surface tension than B (*i.e.*, the myelin). If the axoplasma disintegrates into numerous spheres, this relationship of the surface tensions is disturbed, and the myelin also tends to return to the spherical shape. These physical changes are first manifested in the decrease of the birefringence and precede the chemical disintegration of the myelin which is demonstrable by the Marchi method. It should be emphasized that these chemical changes cannot be explained by the above-mentioned physical mechanisms alone. One might surmise that they are caused by products (enzymes?) liberated during the breakdown of the axoplasma. These problems, however, require further study.

Microscopic staining methods may be utilized for a study of electrical potentials in the elements of the nervous system. Keller[91] groups the histologic stains, according to the direction in which they are carried by the electric current, into anodic and cathodic stains, and uses them for determination of the morphologic distribution of potentials. He finds, for instance, that the membrane of the nucleus and tigroid bodies are negative, the elements appearing as neurofibrils are electro-positive; the inner surface of the myelin sheaths is negative; their Lantermann-Schmidt-Golgi structures are positive; neuroglia and collagen fibers negative.

X-ray Diffraction. The results of studies under the ultramicroscope and in polarized light were supplemented by the x-ray diffraction method.[16, 66, 75, 90, 147, 150] By using an exposure time of only eight minutes, Boehm studied surviving nerves in a moist chamber. From a comparison of the diffraction patterns of various peripheral nerves (myelinated nerves with oriented lipoid, unmyelinated nerves with and with-

out oriented lipoid, nerves without definite lipoid), Boehm concluded that in the myelin sheaths the geometrical longer axis of the liquid crystalline structures is perpendicularly oriented to the direction of the nerve. This seems in agreement with the above-mentioned results of the studies in polarized light. It should here be recalled that polysaccharides and proteins behave differently, the longer axis lying parallel to the fiber axis; conditions in the myelin sheaths can be imitated, according to Thiessen by silk threads drawn through soap solutions. The oriented material in myelin sheaths is described as mixtures of cephalin and cerebrosides, with an average molecular spacing of approximately 4.8 Å. The amount of connective tissue varied considerably in the different peripheral nerves. In agreement with the results of ultramicroscopic studies, Boehm asserts that neurofibrils do not exist in the living nerve, and that the axis cylinder is a tube filled with an amorphous and viscous jelly. The water in the connective tissue of nerves is partly bound within the micelles. The effect of stimulation is controversial. While Boehm was unable to detect changes in the x-ray diffraction pattern of peripheral nerves even on strong faradic stimulation, Handovsky reports changes observed in the spinal cord of frogs after electric stimulation or strychnine poisoning. Both authors agree that narcosis modifies the x-ray diffraction pattern; this change is explained by action of the narcotics upon the lipoids (change in the arrangement of the fatty acid molecules, Handovsky).

Schmitt, Bear and Clark, using x-ray beams of longer wave-length, further elucidated these problems. They found in dried nerve fibers spacings which closely corresponded to those of extracted lipoids, lecithin and cholesterol. But the equatorial sickles at 11.5 Å, visible in fresh nerve, cannot be reproduced by the extracts. The dimensions of the fundamental aggregate of the liquid crystals of the fresh myelin sheath are given as $4.7 \times 9.4 \times 171$ Å. In human brains Reynolds, Corrigan and Hayden found some differences between gray and white matter of the cerebrum which became more appreciable upon drying. Orientation was found in nerve trunks but not in the white matter, where it could be detected only by studies of birefringence in polarized light.

Measurement of the Conductance and Polarizability

Measurements of the conductance of the exposed brain *in vivo* were performed by neurosurgeons,[63, 104, 122, 186] using a Wheatstone bridge arrangement. Since the conductance of normal brain tissue (cerebrum 1.8×10^{-3} mho, cerebellum 1.53×10^{-3} mho, with Meyer's electrodes) is distinctly different from that of tumors (about $3.3 - 5.0 \times 10^{-3}$ mho), this method may be used as a help in determining the seat of a tumor during operation. The conductance of the brain *in vivo* depends to a large extent upon the variations of its blood supply, since the conductance of blood (6.6×10^{-3} mho with Meyer's electrodes) is about 4 times that of brain tissue.

In order to obtain a sharp minimum on the Wheatstone bridge, it is necessary to use alternating currents of sinusoidal wave form, such as are generated by vacuum-tube oscillators (see [87, 88, 119]), and to balance not only the resistance but also the capacitance of the tissue (see p. 1150). The resistors used should be free of capacitance and inductance, and proper shielding and grounding of the bridge is essential particularly if higher frequencies are used (see [64]).

In order to obtain more complete information about the state of the tissue, particularly of the cells, the determination of the conductance alone is not sufficient. As was already mentioned, fluctuations in the blood supply occurring *in vivo* produce considerable alterations of the conductance. In post-mortem measurements (Crile, Hosmer, and Rowland [28] on rabbits) this disturbing factor is more or less eliminated. It is, however, still difficult to interpret differences of conductance obtained under various conditions, since such differences may be due to changes in the ion concentration of the intra- as well as of the extracellular fluids, or of the permea-

bility of semipermeable cell membranes. Some information about the condition of the latter may be obtained by measurement of the polarization phenomena appearing if electrical currents flow through the tissue.

Since the cell surfaces are impermeable for certain ions, forming Helmholtz electric double-layers,[101] a current flowing through the tissue will produce accumulation of ions on the semipermeable membranes and a counter electromotive force is formed. This becomes apparent as the polarization current if one interrupts the original current. Thus polarizable tissues act like condensers, showing a certain capacity (polarization capacity [60]). If one measures the conductance of a tissue, one deals not simply with an "ohmic" resistance but with an impedance (Z) consisting of a true ohmic resistance (R) and a capacitive reactance (X) *: $Z = \sqrt{R^2 + X^2}$ Therefore, in order to obtain absolute balance on the bridge, one has to compensate not only the resistance, but also the capacitance of the tissue which creates a phase difference between current and voltage. This may be done by a variable condenser connected in series or in parallel with the balancing resistance. The capacitance necessary to balance the bridge may be used as a measure of the polarizability of the tissue and of the state of the cellular surface films.[60] Or one may determine the ratio of the reactance to resistance which represents the tangent of the phase angle (ϕ): $\tan \phi = \frac{X}{R}$ [18, 26a, 110]. If one uses high frequencies, the cell membrane offers rather little impedance to the current, and the ohmic resistance of the cell interior is mainly measured.[60, 79, 119]

In these measurements of the polarization capacity or of the tangent of the phase angle it is necessary to obtain absolute balance on the bridge; and since each determination requires some time, it is rather difficult to record by this method quick changes of polarizability following experimental procedures *in vivo*.

Spiegel and Spiegel-Adolf [162, 165] therefore used another polarization phenomenon as a measure of polarizability in brain tissue, *viz.*, the fact that the conductance increases with increasing frequency of the alternating current, in brain tissue as well as in other organs. They determine the conductance at a certain high (K_h) and at a certain low frequency (K_l); the difference, expressed in percentage of the conductance at the low frequency, is called the polarization index: $\Delta = \frac{K_h - K_l}{K_l} \times 100$.

It was demonstrated by quantitative studies on artificial membranes [171] and on frogs' skin [166] (comparison of changes of Δ and of the dialysis of substances through these membranes) that measurements of Δ can be used as an indicator of permeability changes. A drop (rise) of Δ indicates lowered (increased) polarizability and thus increased (decreased) permeability of the cell surfaces.

Upon a similar principle is based a method to measure the thickness of the plasma membranes of single cells. The difference between the impedance at low frequencies and at frequencies over one million has been determined.[118, 141] This method, however, has not yet been applied to the central nervous system.

Hydrogen-ion Concentration. A method for recording the pH of the cerebral cortex *in vivo* was described by Dusser de Barenne, McCulloch and Nims.[84, 180] The authors use a glass electrode of the MacInnes type in conjunction with the microvoltmeter described by Burr, Lane and Nims.[22] A marked alkaline shift of the cortex could be produced by intravenous injection of sodium bicarbonate as well as by hyperventilation; a marked acid shift by venous injection of a dilute acid, by hypoventilation, or by ether anesthesia.

A low (high) pH is associated with low (high) electric activity, as shown by the electrocorticogram and also by decreased (increased) excitability of the cortex. It seems that the variations in excitability with pH are sufficient to account for the changes in the electrocorticogram. Electrical stimulation of a focus producing after-

* For measurement of the impedance see Horton.[82, 83]

discharge, as seen in the electrocorticogram, results in acidity not only in the stimulated area but also in distant foci that are functionally related to the stimulated area, so that after-discharge appears in them also.

If a "central" epileptoid discharge of the cortex is produced by intravenous injection of camphor, the record of the cortical pH shows an alkaline wave followed by an acid swing independent of the record of the pH of the blood, which presents a steady, slow acid shift.[88]

Microinjection may be used for the study of intracellular hydrogen-ion concentration, as was shown by Chambers and Pollock,[24] on the giant nerve cells of the medulla of *Lophius* (goose fish). Similarly, as in other cells, a hydrogen-ion gradient between cytoplasma (colorimetric pH about 6.9) and nucleus (pH 7.6 or above) was found.

Electrical Activity. A detailed description of the technique and application of electroencephalography is beyond the scope of this paper. We limit ourselves therefore to some general considerations of its relationship to physicochemical problems, and refer the reader for further details to the numerous excellent reviews in this field.[14, 31, 56, 86, 99]

Since the cell membranes are the seat of electrical potentials, the fluctuations of potentials that may be led off the various parts of the central nervous system may be considered from a physicochemical point of view as due to changes in the state of the membranes of cells or cell groups. Conversely, it may be expected that a study of the electrogram may serve as an indicator of the physicochemical state of the cells concerned. However, we must be rather careful in drawing such conclusions; we must, for instance, bear in mind that a reduction of electric activity may be the result of an increase in density of the cell membranes, or of an increase of permeability to such an extent that potential differences can no longer develop. In general, it may perhaps be stated that the easier the cell membranes respond to the depolarizing effect of stimuli, or the higher the state of "spontaneous" excitation of the cells, the more pronounced will be the fluctuations of potential; this will be particularly seen as an increase in frequency of the potential waves of the electrogram, in some instances as an increase of their amplitude. There is occasionally an inverse relation between frequency and amplitude, fast oscillations being associated with small, and slow oscillations with large amplitude. Conversely, the less the cell membranes will respond to stimuli, or the more their spontaneous activity is depressed, the smaller will be the electric activity and the slower waves will predominate. The most pronounced example of increased electric activity is seen in convulsions in experimental animals [41] or in patients.[57] Slow waves may be found in many conditions of depressed activity: in normal sleep,[19, 80] or during the influence of hypnotics; [20, 156] in various forms of impairment of the cells of the central nervous system, as in anoxemia; [21, 108] in increased intracranial pressure; and in local injury of the brain substance, *e.g.*, due to tumors.[181]

It seems of interest that quite a considerable degree of spontaneous oscillations may be picked up, not only from the cortex,[1, 10] but also from the thalamus,[156] when consciousness is so diminished that its outward manifestations are abolished. Thus some degree of neural activity may still be maintained in the prosencephalic as well as in the diencephalic centers upon which consciousness depends, even if the state of consciousness is so far depresssed that spontaneous movements cease.

In studying the effects of a certain agent, *e.g.*, of certain ions or drugs, we must bear in mind that opposite effects may be obtained, depending on the dosage or the duration of the action of a substance. In general, on application of small doses or in initial stages of an intoxication, we observe an increase of the electric activity; while on prolonged action or immediately on application of large doses, the electric discharges may be depressed, as could for instance be demonstrated for potassium chloride, CO_2, and acetylcholine.[17, 55]

A consideration of these quantitative factors may explain some controversial observations recorded in the literature.

References

1. Adrian, E. D., and Matthews, B. H., *J. Physiol.*, **81**, 440 (1934); *Brain*, **57**, 355 (1934).
2. Alexander, L., and Myerson, A., *Trans. Am. Neurol. Assoc.*, p. 109 (1937); *Arch. Neurol.*, **39** (1938).
3. Alzheimer, A., in F. Nissl, "Beitraege z. Frage nach d. Beziehung zw. Klinischem Verlauf u. anat. Befund b. Nerven u. Geisteskrankh," Springer, Berlin, vol. I, pt. 2, 1914.
4. Amberson, W. R., Nash, T. P., Mulder, A. G., and Binns, D., *Am. J. Physiol.*, **122**, 224 (1938).
5. Ambronn, "Bericht. d. saechs. Akademie Mathem. physikal. Klasse," Leipzig, 1890.
6. Auerbach, L., *Neurol. Cbl.*, 994 (1908): *Arch. mikr. Anat.*, **81**, I, 168 (1912); *Pflueger's Arch. ges. Physiol.*, **222**, 493 (1929).
7. Baldi, F., *Rev. de path. nerv.*, **35**, 158 (1930).
8. Bancroft, W. D., and Richter, G. H., *J. Phys. Chem.*, **35**, 215, 1606 (1931).
9. —, and Rutzler, S. E., *ibid.*, **35**, 1185 (1931); **36**, 2011 (1932).
9a. Banzhaf, E. J., and Gibson, R. B., *J. Biol. Chem.*, **3**, 253 (1907).
10. Bartley, S. H., and Bishop, G. H., *Am. J. Physiol.*, **103**, 173 (1933).
11. Bauer, J., *Arb. Neurol. Inst. Univ. Wien*, **19**, 87 (1911).
12. —, and Ames, T., *Z. ges. Neur.*, **5**, 698 (1912).
13. Bernstein, J., "Elektrobiologie," Braunschweig (F. Vieweg), 1912.
14. Bertrand, J., Delay, J., and Guillain, J., "L'électroencéphalogramme," Paris, Masson et Cie., 1939.
15. Berzeller, *Biochem. Z.*, **66**, 225 (1914).
16. Boehm, G., *Kolloid-Z.*, **62**, 22 (1933).
17. Bonnet, V., and Bremer, F., *Compt. rend. soc. de biol.*, **126**, 1271 (1937); **127**, 806, 812 (1938).
17a. Braunmuehl, A., *Z. ges. Neurol.*, **142**, 1 (1932).
18. Brazier, M., *Lancet*, **2**, 742 (1933); *J. Inst. Elect. Eng.*, **73**, 203 (1933); *Endocrinology*, **20**, 621 (1936).
19. Bremer, F., *Compt. rend. soc. de biol.*, **118**, 1235 (1935); **122**, 464 (1936).
20. —, *Ibid.*, **121**, 861 (1936).
21. —, and Thomas, J., *Ibid.*, **123**, 1256 (1936).
22. Burr, H. S., Lane, C. T., and Nims, L. F., *Yale J. Biol. Med.*, **9**, 65 (1936).
23. Calugareanu, *Biochem. Z.*, **29**, 96 (1910).
24. Chambers, R., and Pollock, H., *J. Gen. Physiol.*, **10**, 739 (1927).
25. Clark, G. L., "Applied X-rays," McGraw-Hill Book Co., New York, 1940.
26. Clowes, G. H., *Proc. Soc. Exp. Biol. Med.*, **11**, 8 (1913). *J. Physic. Chem.* 20, 407, (1916).
26a. Cole, K. S., in *Cold Spring Harbor Symposium Quantit. Biol.*, 1933.
27. —, and Curtis, H. J., *J. Gen. Physiol.*, **22**, 649 (1939).
28. Crile, G. W., Hosmer, H. R., and Rowland, A. F., *Am. J. Physiol.*, **60**, 59 (1922).
29. Darrow, D. C., and Yannet, H., *J. Clin. Invest.*, **14**, 266 (1935).
29a. Davies, D. R., and Quastel, J. H., *Biochem. J.*, **26**, 1672 (1932).
30. Davis, H., Davis, P. A., Loomis, A. L., Harvey, E. N., and Hobart, G., *J. Neurophysiol.*, **1**, 24 (1938).
31. —, *Ann. Rev. Physiol.*, **1**, 345 (1939).
32. de Crinis, M., *Z. ges. Neurol.*, **162**, 646 (1938).
33. Dusser de Barenne, J. G., Marshall, C. S., McCulloch, W. S., and Nims, L. F., *Am. J. Physiol.*, **124**, 631 (1938).
34. —, McCulloch, W. St., and Nims, L. F., *J. Cell. Comp. Physiol.*, **10**, 277 (1937); *Proc. Soc. Exp. Biol. Med.*, **36**, 462 (1937).
35. Ebner, V., "Untersuchungen ueb. d. Ursachen d. Anisotropie," Leipzig, 1882.
36. Elias, H., *Z. ges. Exp. Med.*, **7**, 1 (1919).
37. Ettisch, G., and Jochims, J., *Pflueger's Arch. ges. Physiol.*, **215**, 519, 675 (1927).
38. Fay, T., *Am. J. Psychiat.*, **85**, 783 (1929); *Arch. Neur. Psych.*, **23**, 920 (1930); *J. Nerv. Ment. Dis.*, **71**, 481 (1930).
39. Ferraro, A., *J. Nerv. Ment. Dis.*, **71**, 129 (1930).
40. Fischer, M. H., "Das Oedem," Dresden, 1910.
41. —, *Pflueger's Arch. ges. Physiol.*, **230**, 161 (1932); *Med. Klin.*, **29**, 15 (1933).
42. —, and Sykes, A., *Kolloid Z.*, **16**, 129 (1915).
43. Franke, W., *Z. physiol. Chem.*, **212**, 234 (1932).
44. Fraenkel, S., and Linnert, K., *Bioch. Z.*, **26**, 50 (1910).
45. Freeman, W., *Arch. Neur. Psychiat.*, **24**, 300 (1930); *J. Am. Med. Assn.*, **97**, 293 (1931).
45a. Frenkel-Tissot, H. C., *Schweiz. Med. Wochenschr.*, **3**, 613, 635 (1922).
46. Freundlich, H., "Kapillarchemie," 1919.
47. Frisch, F., "Das vegetative System der Epileptiker, pp. 57 ff., Berlin, J. Springer, 1928.
48. Fuenfgeld, E. D., *Z. Nervenheilk.*, **114**, 209 (1930).
49. —, *Arch. Neur. Psychiat.*, **23**, 915 (1930).
50. Gamble, J. L., "Chemical Anatomy, Physiology and Pathology of Extracellular Fluid," Dept. of Pediatrics, Harvard Med. Sch., 1939.
51. —, Ross, G. S., and Tisdall, F. T., *J. Biol. Chem.*, **57**, 633 (1923).
52. Gellhorn, E., and Weidling, K., *Arch. ges. Physiol.*, **210**, 492 (1925).
53. Georgi, F., *Z. ges. Neur.*, **106**, 751 (1926).
53a. —, and Fischer, O., "Humoralpathol. d. Nervenkrankheiten" in "Handb. d. Neurol." (Bumke-Foerster) vol. 7$_1$, p. 40 (1935).
54. Gerard, R. W., *Am. J. Physiol.*, **92**, 498 (1930).
55. —, *Trans. Am Neurol. Assoc.*, p. 55, (1936).
56. Gibbs, F. A., and Gibbs, E. L., "Atlas of Electroencephalography," Boston, 1941.
57. —, —, and Lennox, W. G., *Brain*, **60**, 377 (1937).
58. Gildea, E. F., and Cobb, S., *Arch. Neur. Psych.*, **23**, 876 (1930).
59. Gildemeister, M., *Ber. d. ges. Physiol.*, **2**, 182 (1920).
60. —, "Handb. d. norm. und pathol. Physiol." ed. by Bethe, Vol. **8**, p. 657, Berlin, Springer, 1928.
61. Goethlin, G., *Kungl. Svenska Vetenskaps Akad. Handlingar*, **51**, fasc. 1, 1913.
62. Golla, F. L., *Proc Roy. Soc. Med.*, (Sect. on Psychiatry) **22**, 31 (1929).
63. Grant, F. C., *J. Am. Med. Assoc.*, p. 2169, (1923).
64. Hague, B., "Alternating Current Bridge Methods," London, 1932.
65. Haldi, H., Rauth, J., Larkin, J., and Wright, P., *Am. J. Physiol.*, **80**, 631 (1927).
66. Handovsky, H., *Kolloid Z.*, **62**, 21 (1933).

67. —, and Thiessen, P. A., *Nachr. Ges. Wiss. Goettingen*, (May 23, 1930).
68. —, and Wagner, R., *Biochem. Z.*, **31**, 32 (1911).
69. Hansteen-Cranner, *Jahrb. f. wissensch. Botanik.*, **47**, 374 (1910).
70. Harris, M. M., and Katz, S. E. *Am. J. Psychiat.*, **12**, 1065 (1933).
71. Harrison, H. E., Darrow, D. C., and Yannet, H., *J. Biol. Chem.*, **113**, 515 (1936).
72. Hastings, B., and Eichelberger, L., *J. Biol. Chem.*, **117**, 73 (1937).
73. Heim, F., *Biochem. Z.*, **291**, 88 (1937).
74. Henderson, V. E., and Lucas, G. H., *J. Pharm. Exper. Ther.*, **44**, 253 (1932).
75. Herzog, R. O., and Jancke, W., *Festschrift d. Kaiser-Wilhelm Ges.* p. 118, Berlin, 1921.
76. Himwich, H., Fazikas, *et al., Proc. Soc. Exp. Biol. Med.*, **30**, 903 (1933).
77. Hirschfelder, A. D., *J. Pharm. Exp. Ther.*, **37**, 399 (1929).
78. —, and Serles, E. R., *Ibid.*, **29**, 441 (1926).
79. Hoeber, R., "Physik. Chemie d. Zelle a Gewebe," 6th ed., 1926.
80. Hoff, H., and Poetzl, O., *Z. ges. Neurol.*, **124**, 200 (1930).
81. Hooker, M. O., and Fischer, M. H., *Z. Chemie & Ind.* d. Kolloide, **10**, 283 (1912).
82. Horton, J. W., and Hertz, S., *Endocrinology*, **20**, 831 (1936).
83. Horton, J. W., and van Ravenswaay, A. C., *J. Franklin Inst.*, **220**, 557 (1935).
84. Howe, P. E., and Sanderson, F. S., *J. Biol. Chem.*, **62**, 767 (1925).
85. Jaburek, L., *Arch. Psychiat.*, **104**, 518 (1936).
86. Jasper, H. H., *Psychol. Bull.*, **34**, 411 (1937).
87. Jones, G., and Bollinger, G. M., *J. Am. Chem. Soc.*, **51**, 2407 (1929); **53**, 1207 (1931).
88. Jones, J., and Joseph, R. C., *J. Am. Chem. Soc.*, **50**, 1049 (1928).
89. Katz, J. R., *Erg. d. exakt Naturwiss.*, **3**, 316 (1924).
90. —, and Spiegel, E. A., cited by Katz, J. R., "Die Roentgenspektrographie als Untersuchungsmethode, Handb. biol. Arbeitsmethoden," Ed. by E. Abderhalden, Div. II, Part 3, Fasc. 6, page 3690, 1934.
91. Keller, R., *Biochem. Z.*, **128**, 409 (1922).
92. Klose, H., and Vogt, H., *Bruns. Beitr. z. Klin. Chirurgie*, **69** (1910).
93. Koch, M. L., *Arch. Neur. Psych.*, **7**, 488 (1922).
94. Koch, W. J., *J. Pharm. Exp. Ther.*, **2**, 253 (1910).
95. Koch, W., and Goodson, W. H., *Am. J. Physiol.*, **15**, 272 (1906).
96. —, and Mann, S., *Arch. Neur. Psychiat.*, **4**, 174 (1909).
97. Koch, W., and Riddle, O., *Am. J. Physiol.*, **47**, 124 (1918); *J. Comp. Neurol.*, **31**, 83 (1919).
98. Koch, M. L., and Voegtlin, C., *Hygienic Lab. Bull.* **103**, Government Printing Office, Washington, D. C., 1916.
99. Kornmueller, A. E., "Bioelektr. Erscheinungen d. Hirnrindenfelder," Thieme, Leipzig, 1937.
100. Krause, B., *Zschr. ges. Neur.*, **35**, 542 (1917).
101. Krueger, F., *Zschr. physik. Chemie*, **45**, 1 (1903).
102. Leake, C. D., *J. Am. Med. Assn.*, **100**, 682 (1933).
103. Lennox, W. G., Gibbs, F. A., and Gibbs, E. L., *J. Neurol. Psychiat.*, **1**, 211 (1938).
104. Lichotzky, *Cbl. f. Chirurg.*, **53**, 452 (1926).
105. Liesegang, R., *Biochem. Z.*, **28**, 413 (1910).
106. —, *Z. allg. Physiol.*, **11**, 347 (1911).
107. Lillie, R. S., "Protoplasmic Action and Nervous Action," Univ. of Chicago Press, 1923.
108. Ludlum, S., Taft, A. E., and Nugent, R. L., *Arch. Neurol.*, **23**, 1121 (1930).
109. —, and Zozaya, J., *J. Biol. Chem.*, **110**, No. 3 (Aug., 1935); *J. Physiol. Chem.*, **42**, No. 2 (Feb., 1938); *Trans. Am. Neurol. Assn.*, p. 110 (1938).
110. Lullies, H., *Pflueger's Arch. ges. Phys.*, **225**, 69 (1930), and in "Handb. d. Biol. Arbeitsmeth." by Abderhalden, Abt V, p. 1271 (1936).
111. MacDougall, D. T., *Proc. Am. Phil. Soc.*, **60**, 15 (1921); **61**, 33 (1922).
112. Marinesco, G., *Rivista speriment. di freniatria*, **50**, fasc. I-II, p. 1 (1926).
113. May, R. M., *Bull. Soc. Chimie Biol.*, **12**, 934 (1930).
114. May, R. M., *Encéphale*, **25**, 447 (1930).
115. McQuarrie, J., in "Chemistry and Medicine," ed. by M. B. Visscher, University of Minnesota Press, Minneapolis, 1940.
116. —, Manchester, R. C., and Husted, C., *Am. J. Dis. Child.*, **43**, 1519 (1932).
117. —, and Peeler, D. B., *J. Clin. Invest.*, **10**, 915 (1931).
118. McClendon, J. F., *Protoplasma*, **3**, 7 (1927).
119. —, and Hemingway, A., *Am. J. Physiol.*, **94**, 77 (1930).
120. Meltzer, S. J., and Auer, J., *Am. J. Physiol.*, **14**, 366 (1906); **15**, 387 (1906); **16**, 233 (1906).
121. Mestrezat, "Le liquide céphalorachidien," Paris, Maloine, 1912.
122. Meyer, A. W., *Cbl. Chirurg.*, **48**, II, p. 1825 (1921).
123. Meyer, H. H., *Arch. f. Exp. Path. Pharm.*, **42** (1899).
124. Molitor, H., and Pick, F. P., "Schlaf," ed. by Sarason, 1929.
125. Monaghan, B. R., and Schmitt, F. O., *Proc. Soc. Exp. Biol*, **28**, 705 (1931); *J. Biol. Chem.*, **96**, 387 (1932).
126. Mond, R., and Netter, H., *Pflueger's Arch. ges. Physiol.*, **230**, 42 (1932).
127. Mott, F. W., and Halliburton, W. D., *Phil. Trans. Roy. Soc. London*, **B**, **194**, 437 (1901).
128. Mueller, R., and Hough, W. H., *Wien. Klin. Wschr.*, **24**, 167 (1911).
128a. Muenzer, E., *Arch. Exp. Path.*, **41**, 74 (1898).
129. Nernst, W., *Pflueger's Arch.*, **122**, 275 (1908); **123**, 454 (1908).
130. Nims, L. F., McCulloch, W. S., and Dusser de Barenne, J. G., *Am. J. Physiol.*, **126**, 594 (1939).
130a. Novi, *Pflueger's Arch. ges. Physiol.*, **48**, 320 (1891).
131. Osterhout, W., *Science*, **35**, 112 (1912); **37**, 111 (1913).
132. Ostwald, W., *Z. physik. Chem.*, **6**, Heft 1 (1890).
133. Overton, Studien ub. d. Narkose, Jena, 1901.
134. Page, I. H., "Chemistry of the Brain," C. C. Thomas, Springfield, Ill., 1937.
135. Palladin, A., and Ssawron, E., *Biochem. Z.*, **200**, 244 (1928).
136. Pauli, W. E., and Redwitz, E., *Dtsch. Z. f. Chirurg.*, **193**, 343 (1925).
137. Peterfi, T., in "Handb. d. norm. und pathol. Physiol.," ed. by Bethe, **9**, 79 (1929).
138. Peters, J. P., "Body Water," Springfield, Ill., C. C. Thomas, 1936.
139. Peters, R. A., and Sinclair, *Biochem. J.*, **27**, 1677 (1933).
140. Peters, R. A., and Thompson, R. H., *Biochem. J.*, **28**, 916 (1934).
141. Philippson, *Bull. Acad. Roy. Belg.*, **7**, 389 (1921).
142. Pighini, G., *Neurolog. Bulletin, Columbia Univ.*, **2**, 395 (1919).
143. Pike, F. H., Osnato, M., and Notkin, J., *Assoc. Res. Nerv. Ment. Dis.*, **7**, 192 (1931).
144. Porges O., and Neubauer E., *Biochem. Z.*, **7**, 152 (1908).
144a. Quastel, J. H., and Wheatley, A. H., *Biochem. J.*, **25**, 117 (1931); **28**, 916, (1934); *Proc. Roy. Soc.*, **112, B**, 60 (1932).

145. Reichardt, M., "Die Untersuchung des gesunden u. kranken Gehirns mittels d. Wage," Jena, 1906.
146. Reichardt, M., *Z. ges. Neur. Ergebnisse*, **3**, 1 (1911); *Allg. Z. Psychiatrie*, **75**, 34 (1919).
147. Reynolds, L., Corrigan, K. E., and Hayden, H., *Am. J. Roentg.*, **43**, 81 (1940).
148. Rowntree, L. G., *J. Pharm. Exp. Ther.*, **29**, 135 (1926).
148a. Ružicka, V., *Arch. f. mikr. Anat.*, **101**, 459 (1924).
149. Scheinker, I., *Arch. Neurol. Psych.*, **45**, 117 (1941).
150. Schmitt, F. O., Bear, L., and Clark, G. L., *Radiology*, **25**, 131 (1935).
151. Scott, G. H., *Am. J. Anat.*, **53**, 243 (1933); *Protoplasma*, **20**, 133, 1933.
152. Selbach, H., *Z. ges. Neur.*, **162**, 145 (1938).
153. Setterfield, H. E., and Baird, T. T., *Stain Technol.*, **11**, 41 (1936).
154. —, and Sutton, T. S., *Anat. Rec.*, **61**, 397 (1935).
155. Spatz, H., *Arch. Psych.*, **88**, 790 (1929).
156. Spiegel, E., *Am. J. Physiol.*, **118**, 569 (1937).
157. —, *Am. J. Psychiat.*, **87**, 595 (1931).
158. —, *Beitr. z. pathol. Anat. u. z. allg. Pathol.*, **70**, 215 (1922).
159. —, *Pflueger's Arch. ges. Physiol.*, **192**, 225 (1921).
160. —, *Ibid.*, **192**, 240 (1921).
160a. Spiegel, E., and Henny, G., *Proceed. Soc. f. Exp. Biol. & Med.*, **51**, 382 (1942).
161. —, *Z. wissensch. Mikroscopie*, **41**, 383 (1924).
162. —, and Spiegel-Adolf, M., *Am. J. Psychiat.*, **92**, 1145 (1936).
163. —, —, *Arch. int. Pharmacodyn.*, **58**, 419 (1938).
164. —, —, *Proc. Soc. Exp. Biol.*, **42**, 834 (1939).
165. —, —, *J. Nerv. Ment. Dis.*, **93**, 750 (1941).
166. —, —, *Am. J. Physiol.*, **133**, 459 (1941).
167. —, —, *Revue neurol.*, **42**, 469 (1935).
168. —, —, *Proc. Soc. Exp. Biol.*, **34**, 799 (1936); *J. Nerv. Ment. Dis.*, **90**, 188 (1939).
169. Spiegel-Adolf, M., *Biochem. J.*, **30**, 1536 (1936).
170. —, "Globuline," Th. Steinkopff, Dresden, 1930.
171. —, *J. Gen. Physiol.*, **20**, 695 (1937).
172. —, *Klin. Wschr.*, **11**, 485 (1932); *Biochem. J.*, **26**, 2183 (1932).
173. —, *Proc. Soc. Exp. Biol. Med.*, **37**, 92 (1937).
174. —, and Freed, H., *Ibid.*, **40**, 398 (1939); *Confinia Neurologica*, **2**, 228 (1939).
174a. Spiegel-Adolf, M., Ashkenaz, E. W., Lee, A. J. and Spiegel E., *Proceed. Feder. Soc.*, 1944, in press.
175. Stengel, E., *Jahrb. Psych. Neur.*, **45**, 187 (1927).
176. Stone, W. E., Marshall, C., and Nims, L. F., *Am. J. Physiol.*, **132**, 770 (1941).
177. Syz, H. C., *J. Pharm. Exp. Ther.* **21**, 263 (1923); **30**, 1 (1926).
178. Tabern, D. L., and Shelberg, E. F., *J. Am. Chem. Soc.*, **55**, 328 (1933).
179. Toennis, W., *Z. ges. Neurol.*, **161**, 114 (1938).
180. Traube, *Pflueger's Arch. ges. Physiol.*, **153**, 299.
181. Walter, W. G., *Lancet*, **231**, 305 (1936); *Proc. Roy. Soc. Med.*, **30**, 579 (1937).
182. Weaver, H. M., *J. Lab. Clin. Med.*, **26**, 1295 (1941).
183. Weed, L. H., *Am. J. Anat.*, **31**, 253 (1923).
184. Weed, L. H., and Wegefarth, P., *J. Pharm. Exp. Ther.*, **13**, 317 (1919).
185. Weil, A., *Z. ges. Neurol.*, **55**, 49 (1920).
186. Winterstein, H., "Die Narkose," 2nd ed., Berlin, J. Springer, 1926.
187. Wortis, S. B., *Bull. Neurol. Inst. New York*, **4**, 588 (1936).
188. Wuth, O., in Bumke's "Handb. d. Geisteskrankh," Section III, **3**, 154 (1928).
189. Yannet, H., *Am. J. Physiol.*, **128**, 683 (1940).
190. Yannet, H., *Am. J. Dis. Child.*, **58**, 918 (1939).

Psychiatry

SEYMOUR DEWITT LUDLUM, M.D.

Gladwyne Research Laboratory, Gladwyne, Pa.

A definite physiological pathology of mental disease emerges when the chemical proportions of the blood proteins are plotted against the relative reactivity of the various physiological systems. These have a definite relationship, and the mental symptom may, without scientific faltering, be considered as simply another symptom of a definite physiological state. Terms that indicate the phenomenon as a disease entity, such as dementia praecox or manic depressive insanity, are quite unnecessary when it may be considered that such and such an amount of dynamic and proper conditioning of mind emanates from the proper optimum physiological, colloidal, and chemical states. Therefore, the psychological diagnosis under the present nomenclature in psychiatry, to be properly applied, must have the physiological chemical basis of symptoms to establish the structure of the psychosis.

A reversibility, in either direction, of the protein pattern from that to which the individual is accustomed, and at the same time a change in the physiology, in either inhibition or excitation, result in changes in brain function. These are dramatic

when observed for the first time, and become more and more interesting as the principle is seen developing in the various phases of mental disease.

The example of a person with repeated attacks of insanity characterized by confused manic episodes is given to show what takes place. In intervals of the disease this person exhibits a normal protein pattern of approximately albumin 65 per cent, pseudoglobulin 17 per cent, euglobulin 18 per cent, or figures within the normal band, and with normal physiological reactivities in temperature, Schilling hemogram, sedimentation, and the aggregation-dispersion phenomenon. He will run along in good physical and mental equilibrium for a year or more; then suddenly the albumin fraction will jump to 80 per cent, the pseudoglobulin will not change very much, but the euglobulin will become very low, and simultaneously all of the physiological symptoms enumerated above will recede to subnormal positions (Fig. 1, B). This reversibility of the chemical and physiological we have observed and charted a number of times, and each time the mental disorder of a confused manic state appears.

Another interesting point is that here is a rather sudden reversion to a type of immaturity (Fig. 1, A). This type of immaturity *per se* is very common. Constitutionally inferior persons (A) always have the chemical imbalance of high albumin and low euglobulin, as well as a subnormal physiology: subnormal temperature, leucopenia, with lymphocytosis, slow sedimentation, and the dispersion phenomenon in the blood drop. When living in this milieu they do very well, until subjected to excessive strain of environment or to infection, or to the adolescent, climateric, or involutional states; then there is a moderate increase in the euglobulin and a slight reduction in the serum albumin (Fig. 1, C). It is almost as though they could not stand the rise in the euglobulin fraction, the effort of reactivity being too much, quite as though they were allergic to any increase in euglobulin. They can live in the chemical and physiological state to which they were born, but any stimulus requiring more of them results in a mental and physiological disorder. It is in this group that the cases that we call dementia praecox are inducted; but we are really dealing with a constitutional state.

Charles Féré quotes J. Hunter [1] ("Oeuvres Complètes") as maintaining not the existence of a hereditary malady, but only a hereditary disposition to its contraction; and Féré [2] himself maintained that the predisposition is constituted by a state of degeneration, often characterized by gross anomalies of development.

W. Horsley Gantt,[3] in writing about Pavlovian work, says that the important things in evaluating the conditioned reflexes are the constitution of the individual and the strength of the stimulus. Constitution would be the hereditary disposition. In fact in all processes of disease or in the study of inhibition or excitation in any living thing, the fundamental factors are the constitution of the organism and the characteristics of the stimulus.

In this essay it is the purpose to show that the constitution is characterized by a ratio between the serum albumin and the serum euglobulin of the blood, and that proportional to their ratio is a correlated integration of reactive qualities of physiology, in temperature, hemogram, sedimentation, blood pressure, capillary tonus, and aggregation and dispersion in the colloidal sense as seen in the fresh blood viewed under the ultramicroscope, described later.

This study has been made in the field of insanity, and what can be shown in resumé are these facts:

I. That there are immature constitutional states characterized in the blood serum by a high percentage of albumin fraction and a low euglobulin fraction, and depending upon their ratio there will be found varying non-reactivities in the above-mentioned physiological characteristics (A in Fig. 1).

II. That those persons who have good proportional protein patterns and adequate physiology but who suddenly go insane, as in the manic depressive type, either depressed or maniacal, suddenly become throw-backs; their serum albumin goes up,

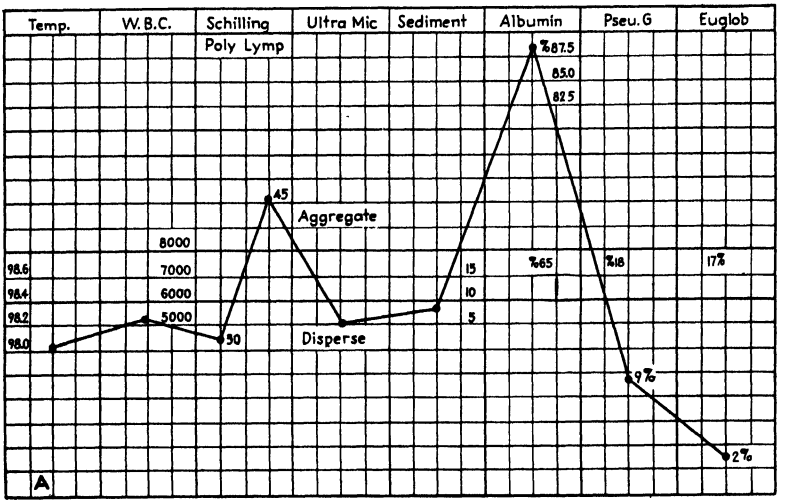


FIGURE 1a.

their euglobulin down, and their physiology becomes non-adequate and non-reactive in temperature, white cell count, Schilling index, blood pressure, capillaries, aggregation and dispersion, thus producing a picture identical with that shown in A of Fig. 1. This is not a hereditary deficiency in protein pattern, as in the first instance, but there is an instability of the molecular constitution which controls the differential protein level. This unstable type reverts to immaturity (B, in Fig. 1).

III. Those persons who by nature present high albumin and low euglobulin fractions all the time, as in A, become insane when any effort is made to throw them into more activity or into less activity, a divergence from the level in which they are accustomed to live. They can remain comfortably in a state of immaturity (C in Fig. 1).

So in any instance or level of pattern, with any departure from their optimal concentrations or their level of equilibriated physiology, the higher functions of the

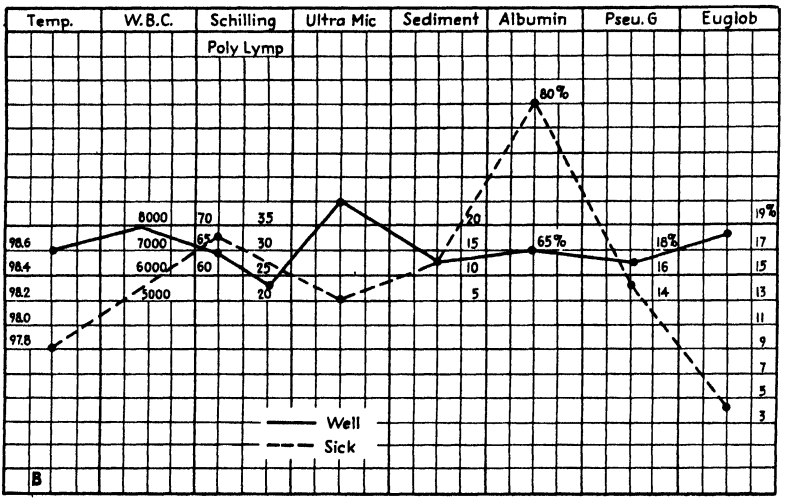


FIGURE 1b.

brain exhibit a disorder of function. Insanity is, therefore, a delicate portrayal of the characteristics of the forces of heredity and constitution, as well as of the development of a disease. The stimulus of insanity may be infection, adolescence, pregnancy; or it may have origins in strains of a psychological nature affecting the chemistry of the milieu.

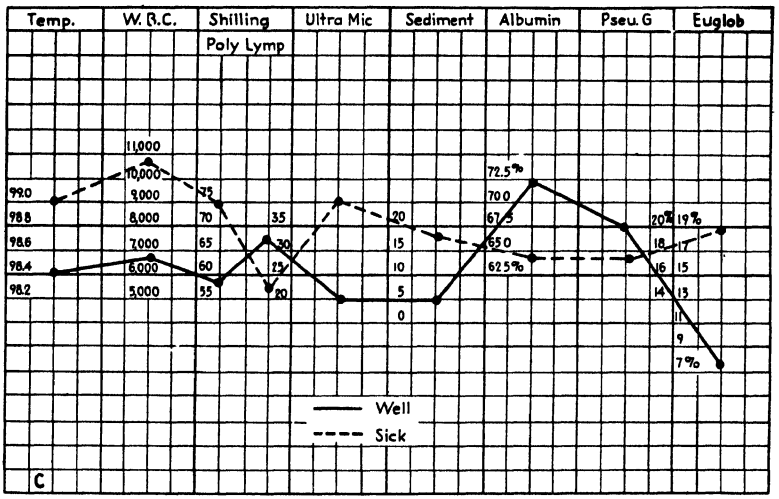


FIGURE 1c.

This is a very simple explanation of insanity, but not much imagination is required to realize that when there is this change in all the physiological characteristics mentioned, the additional symptom of mental aberration would result.

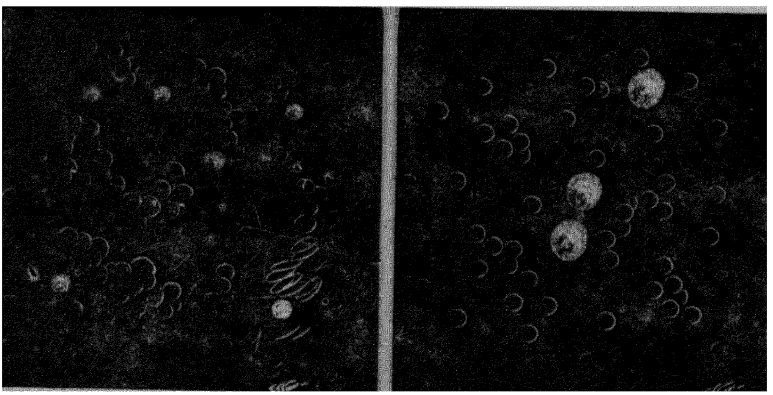
Freud[4] even says: "The neurotic breaks down at the same difficulties that are successfully overcome by normal people, in the same proportion as they break down in their equilibriated physiology. The neuroses have no peculiar content that belongs exclusively to them."

And Barcroft[5] sums up in his book, "Brain and its Environment": "It would seem therefore that in the case of oxygen concentration there is simply a certain optimal concentration in the blood plasma, that to which we are accustomed, and if there is any gross departure from this on either side the higher functions of the brain suffer, leaving the lower ones either to run riot or to flag as the case may be." And in writing about CO_2 concentration effects,[6] he says that both increased and decreased concentrations of CO_2 produced effects, and again the optimum in the natural fluid environment of the nervous system is necessary. Anoxia would be a natural expectation as a result of the physiological non-reactivity described in the preceding instances.

The word *optimum* brings to mind the state of the fluid environment of optimum colloidality. The essential necessity for normal functioning is so aptly stated by Alexander[7] in Vol. I of this series: "Figuratively speaking, if all the chemical substances comprising our organism were in true or crystalloid dispersion, reactions would proceed so rapidly that we would, so to say, live ten years in ten minutes. . . . Every organism is dependent upon the coordination of its chemical reactions *in point of time*, and this leisurely procedure depends largely on *degree of dispersion*, which keeps chemical reaction velocities within certain speed limits through its regulation of free surface and kinetic activity. Life lies between lysis and coagulation. The colloidal zone is, as it were, a vital metronome tolling off the tempo of life."

There are evidences of this sort in the blood of the insane person, seen by examining a drop of fresh blood in an ultramicroscope. The position of the red blood cells, one to the other, is one of aggregation or dispersion. This is also seen in the chylomicrons in the blood serum. It could be explained in the way Francis Schmitt [8] explains the phenomena of the growth processes of cells; and he has done it experimentally with blood cells; that is, he has produced aggregation with Ca or histone and dispersion with cephalon. With Ca or histone there is dehydration and the charge is reduced, whereas with cephalon there is hydration and the charge is increased. Charge and hydration are the fundamental factors in colloid chemistry, according to Kruyt.[9]

These phenomena, as seen ultramicroscopically in the blood of sick people, are what is meant by aggregation and dispersion under the groupings of the physiological symptoms mentioned heretofore as changing in proportion to the ratio of albumin to euglobulin, so that undoubtedly there is a distinct colloidal phenomenon present as part of these physiological changes, as shown in Fig. 2. This is a phase of the under-



AGGREGATED R.B.C.	DISPERSED R.B.C.
SHRUNKEN W.B.C.	SWOLLEN W.B.C.
FIBRIN INCREASED	FIBRIN ABSENT
LEUCOCYTOSIS	LEUCOPENIA

FIGURE 2. Phenomena of Aggregation and Dispersion. Fresh Blood Drop under Ultramicroscope.

lying physico-chemical state in health and disease which warrants much more study.

Suranyi has noted the same exhibitions by red blood cells, and they have been noted by other authors in the study of lipoids, cholesterin producing aggregation just as calcium and histone, and lecithin acting like cephalon.

Gildea [10] has shown an increase of cholesterin in the manic depressive cases, which cases from our point of view would be the unstable group described as Fig. 1, B, and a lessening in schizophrenia, which is to be expected in the cases here described as nonreactors.

Bancroft [11] was perfectly right in considering that aggregation and dispersion in the colloid sense have a bearing on the physiology of insanity. But there is more than that fact alone. There is the influence of the protein state, the lipoid state, the pH, and other factors affecting the colloid state. Another interesting phenomenon to be seen by ultramicroscopic examination of the serum, is a clearness or an opalescence, which varies with the proportion of albumin to euglobulin. All very much as Michaelis says happens to an aqueous soap solution when an amount of NaCl is added to bring about a strong opacity; and if there is then added a small amount of

$CaCl_2$, the opacity becomes less. This is good colloid chemistry and is important in studying the physico-chemical phenomena as they appear.

	Lipoids	
	Cholesterin	Lecithin
L. Suranyi, *Msch. Ungarischer Medizinер*, Budapest, 1928	Lowered negative charge R B C Aggregation Leucocytosis	Increased negative charge R B C Dispersion Leucopenia
Chem. Abs. **25**, 1931; *Migolie-Seli Kwai Med. J.*, **49**	Sedimentation +	Sedimentation –
Lewis	Emulsion Water in Oil	Emulsion Oil in Water
E. F. Gildea, *Arch. Neurol. & Psych.* **43**, (May, 1940)	Increased in Manic Depressive Lessened in Schizophrenia	

That we are dealing with a set of phenomena which includes the subject of immunity is borne out by Table 1.

Table 1. Averages of Physicochemical Observations of the Blood Serum in 2216 Mental and Nervous Cases

Group	Number of Cases	% of Group Cases	Type % of Cases for Total	Viscosity	Specific Gravity	Total Protein	Albumin	Pseudoglobulin	Euglobulin	% Albumin	% Pseudoglobulin	% Euglobulin	Osmotic Pressure
1	490	22.2		1.174	1.0274	7.11	4.42	1.34	1.30	62.1	18.8	18.1	
1A	304	13.7	62.0	1.725	1.0271	7.06	4.40	1.34	1.32	62.4	18.9	18.7	
1B	90	4.1	18.4	1.830	1.0288	7.88	4.90	1.48	1.50	62.2	18.7	19.1	
1C	96	4.4	19.6	1.718	1.0270	8.40	4.01	1.24	1.15	62.6	19.4	18.0	
2	722	32.5		1,821	1.0280	7.12	4.11	1.00	2.01	57.7	14.0	28.3	
2A	354	15.9	49.0	1.811	1.0278	7.20	4.25	1.00	1.95	59.1	13.9	27.0	
2B	230	10.4	31.9	1.992	1.0294	7.98	4.42	1.12	2.44	55.5	14.0	30.5	
2C	138	6.2	19.1	1.698	1.0267	6.20	3.66	0.90	1.64	59.0	14.5	26.4	
3	184	8.3		1.784	1.0279	7.09	3.90	2.05	1.14	55.1	28.9	16.0	
3A	82	3.7	44.6	1.748	1.0274	7.21	4.03	2.18	1.00	55.9	30.3	13.8	
3B	45	2.0	24.5	1.882	1.0290	7.82	4.05	2.18	1.59	51.8	27.8	20.4	
3C	57	2.6	30.9	1.741	1.0263	6.24	3.64	1.80	0.80	58.3	28.8	12.9	
4	820	37.0		1.714	1.0270	7.07	4.95	1.06	1.06	70.0	15.0	15.0	
4A	442	19.9	53.9	1.716	1.0272	7.12	4.98	1.08	1.06	69.9	15.2	14.9	
4B	170	7.7	20.7	1.764	1.0276	7.82	5.41	1.11	1.30	69.2	14.2	16.6	
4C	208	9.4	25.4	1.688	1.0266	6.29	4.46	1.00	0.83	70.9	15.9	13.2	
Total	2216	100	100	1.765	1.0275	7.09	4.35	1.36	1.37	61.4	19.2	19.4	

Group 1: Normal group.
Percentages: albumin 60-65; pseudoglobulin 15-20; euglobulin 15-20.
Group 2: Percentage of euglobulin 20 or more.
Group 3: Percentage of pseudoglobulin 22 or more.
Group 4: Percentage of albumin 65 or more.

These findings show that these mental cases fall into four definite groupings.

Group 2 is characterized by high euglobulin percentage and correlates immunologically with reaction to infection. Group 3 is characterized by high pseudoglobulin percentage and correlates with reaction to poisons, such as toxins from diseased organs like cirrhosis. These two groups, 2 and 3, are termed *hyper-reactor.* in the immunological sense.

Group 4 is characterized by high albumin percentages and correlates immunologically with the *non-reactor* type.

Group 1 contains those cases not falling within the other characterizations, but showing disturbance of the physico-chemical characteristics, specific gravity and viscosity, and it is termed *normal protein.*

More and more cases have been examined at the Laboratory during the period following the publication of 500 cases.[12] The methods used were based on the Kjeldahl, as described in the first publication.[13] The findings of the later cases have been assembled, and the four groups are always found to exist.

From these tables inferences can be drawn relative to the proteins. The square blocks around euglobulin in one group, pseudoglobulin in another group, and albumin in still another group, indicate how the cases fall. The averages can be considered as very instructive, because the averages in large groups of figures do not show important variation; but as these tables of averages do show these distinct groupings, it becomes evident that there were a great many cases with excessive values to make the averages sufficiently high to produce distinctive groups. To demonstrate characteristic patterns, frank cases must be studied clinically and with physiological systems.

The albuminous group are the non-reactors, and with them mental degeneration comes on much more easily. Here most of the precocious dementias appear, people who cannot stand the strain, who cannot go through with adolescence or the climacteric. The highly albuminous person does not necessarily go crazy, but when strain comes, he does not adequately meet it physiologically or psychologically. These are the cases described in A and C of Figure 1.

The globulinous groups are the hyper-reactors, immunologically, those of toxic origin, poisons, liver, degenerations, infections. All of these cases show a globulin increase. There is a probable increase in antibodies in the globulin fraction and an untoward increase in the excitor chemical molecules, so that their work of growth, resistance, and repair are overdone, and the result is a poisonous disintegration or an allergy.

We are brought face to face with this fact: infection can bring on a mental disease; but an intercurrent infectious disease like boils, pneumonia, or erysipelas can cure a mental disease. Some years ago abscesses were induced as a curative agent. It is no uncommon thing to observe mentally sick people recover during an attack of boils. This shows a relation to the principles of immunity, very much like treating infection with a non-specific protein. Other people who become mentally ill following an infection are probably sensitive to some molecules in the increased globulin reaction to the infection. They are possibly allergic to any increase of their own euglobulin. It is just as injurious to them as too much activator substance is to plants or protozoa.

There is something in this albumin-euglobulin relationship that is very akin to the stimulator-inhibitor system postulated by Dr. J. B. Murphy in cancer research and by Dr. V. du Vigneaud in the egg white (avidin) and egg yolk (biotin) relationship. With an increase in the albumin proportion there is an increase in the inhibitory mechanism; and, on the other hand, with an increase in the euglobulin fraction there is an increase in the excitatory mechanism.

An examination of the principles of immunity reveals a number of points of interest, for there is a common ground in the position of the euglobulin and the physiologic reactivity, and recovery.

Why, in the acutely insane, do the physiologic reactivity and adequacy become depressed and the euglobulins go lower, and why, in those prone to dement, are they low to begin with? It is as though that formulation represents a regression to immaturity and a degeneracy.

Zinsser [14] says in his "Immunity Principles": "The capacity to form antibodies is weak or absent in very young animals and only becomes fully developed as the individual seems to be fairly well established. In man, too, the production of antibodies appears to be less marked in infants than in older children." He quotes Lewis and Wells,[15] saying: "The blood serum of young animals is deficient in globulins which, as we shall see, are closely associated if not identical with antibodies, and at first they apparently lack the capacity to form these proteins."

Zinsser also says that Locke [16] notes significant variations in time required for the individual rabbits to return to normal temperature following immersion in cold water. The speed with which this warming up process takes place, runs parallel with the ability of the rabbit to resist intravenous infection with small number of virulent pneumococci.

The fact that we find proteins in arrangement similar to those found in other diseases, and that the warming up process in some rabbits is delayed, brings immunity reactions and physiological reactions closer together, and it could almost be said that these acutely insane with the non-reactive physiologic symptoms and the high albumin and low euglobulin have reverted to an immature stage in physical life.[17] This applies to those cases which are non-reactive and not, of course, to those cases where there is gross infection with increased temperature and white cell count, and with aggregation and rapid sedimentation.

This correlation theme can be carried much further: with subjects such as time, the relation of chemical time and physiological and diuranal time, as has been suggested by Carrel, and du Noüy, and Hoagland; and with specific substances of growth like the auxins, or biotin, or avid albumin (avidin): or the values of different factors like pH and salts; the colloid osmotic pressure and its relation to capillary pressure; and to the relation of oxygen to cerebral function. But to avoid overstatement, this essay simply draws attention to the concurrent evidence that changes in the albumin-euglobulin ratio have a relation to physiological reactivity, which is a chemical relation to a physiological function, and then to the possible relation to immunity, and a most interesting relation to cerebral function and psychology.

The psychiatric terms, dementia praecox, schizophrenia, manic depressive insanity, are of doubtful accuracy in the diagnosis of mental disease. With the protein estimation and physiological reactivity as a basis from which to evaluate the psychopathologic reactions, a better understanding of mental cases can be obtained than by using the ordinarily recognized classifications of mental disease; and a much wider avenue of therapeutic approach is opened. In fact, by establishing the relation between proteins and physiology we indicate that each individual has a normal pattern which is peculiarly his own, and from which he departs in sickness and to which he returns in health.

There is a certain optimal relation of protein and physiologic reactivity to which an individual becomes accustomed; if there is any gross departure from this level either way, there is a disturbance of the higher functions of the brain.

References

1. Hunter, J., "Oeuvres complètes," trad. Richelot, t. I, p. 226.
2. Féré, Ch., "La Famille Néuropathique," p. 8, Paris, 1898.
3. Gantt, W. H., *Am. J. Psych.*, 98, No. 4 (Jan., 1942).
4. Freud, S., "Collected Works."
5. Barcroft, J., "The Brain and Its Environment," p. 101, Yale Univ. Press, 1938.
6. *Ibid.*, p. 103.
7. Alexander, J., "Colloid Chemistry," Vol. I, pp. 25-26.
8. Schmitt, F., "Growth," Sup. to Vol. 5, 1941.
9. Kruyt, H. R., "Colloids," John Wiley & Sons, 1927.

10. Gildea, E. F., *Archives Neur. and Psych.*, **43**, 932-947 (May, 1940).
11. Bancroft, W. D., *Am. J. Pharm.*, **103**, No. 11 (1931).
12. Zozaya, J., Gladwyne Research Laboratory, *J. Phy. Chem.*, **42**, No. 2 (Feb., 1938).
13. *Ibid.*, *J. Biol. Chem.*, 110 No. 3 110 (Aug., 1935).
14. Zinsser, H., "Immunity," p. 123, The Macmillan Co., 1939.
15. Lewis, J. H., and Wells, H. G., *J. Am. Med. Ass.*, **78**, 863 (1922).
16. Locke, see Zinsser, H., "Immunity," p. 125.
17. Molholm, H. B., "Hyposensitivity to Foreign Protein in Schizophrenic Patients," *Psych. Quar.*, **16**, No. 3 (July, 1942).

Changes in the Surrounding Medium Produced by Free-living Cells

H. S. JENNINGS

University of California, Los Angeles

Single free-living cells, or Protozoa, in many cases modify the immediately surrounding medium by means of emanations from the cell. The unicellular organism is thus enveloped in an aura, or a modified zone. Such modified zones play striking rôles in the natural history of the organisms.

The existence of such an aura is known mainly from the reactions of other individuals that approach the cell in question. As the chemical or physical nature of the cell emanation is known in relatively few cases, the phenomena present a set of problems for physicochemical investigation.

Since the free-living cells which show these phenomena live in water, the cell emanations are as a rule in watery solution. Many of the organisms are complex in structure and function, having organs (or "organelles") of motility, of nervous conduction, of nutrition and excretion, and of reproduction. They resemble miniature machines, rather than undifferentiated bits of colloid material. Yet each is organized as a single mass of cytoplasm with a central dominating nucleus (or in some types with an additional minute reserve nucleus or "micronucleus"). They are therefore commonly considered to be single cells.

1. The existence of an aura or modified zone about Protozoan cells is shown in a most striking way by the assembling of many individuals in a small area (Figure 1). These phenomena have been most studied in species of *Paramecium*.

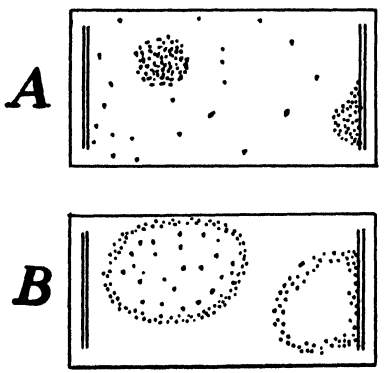


FIGURE 1. Spontaneous groups formed by Paramecia. A and B, successive stages as the groups grow larger. (After Jennings 1906)

If water in which individuals of *Paramecium* are scattered is placed on a plate under a cover glass, so that the organisms are swimming about in a thin layer of water, the phenomena are shown conspicuously. After a time the individuals are seen to

have begun to collect more densely in a small area. The collection enlarges and becomes more densely crowded, so that in time a close aggregation is formed in this area, while the remainder of the water is left with few or no individuals. Similar aggregations, a little less sharply defined, are formed when the organisms are placed in test tubes or in vessels of any form (Figure 2).

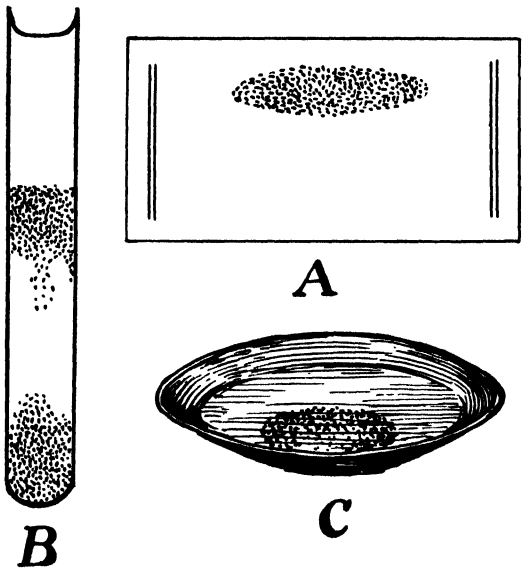


FIGURE 2. Spontaneous groups formed by Paramecium, under different conditions. A, on a glass slide, under a cover glass. B, in a test tube. C, in the bottom of a watch glass. (After Jennings 1906)

In these aggregations the individuals are seen to swim about freely. But on coming to an invisible outer boundary of the area, each individual turns back and remains in the area. This boundary at which the individuals turn back is definite and sharply defined. Individuals that are outside this area are swimming about freely in all directions. When the course of such an individual carries it to the outer boundary of the area of aggregation, it crosses this boundary freely and without reaction. But after coming thus within the area it reacts and turns back at coming to the outer boundary, so that it remains within the area. Thus every individual that enters the area remains, so that after a time a dense aggregation is formed. The course of any single individual entering the area, and remaining, is like that shown in Figure 3.

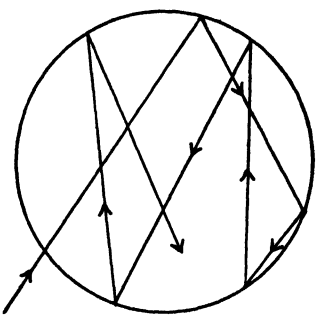


FIGURE 3. Path followed by a single Paramecium in entering and remaining in a region of weak acid, as in a spontaneously formed group. (After Jennings 1906)

Observation shows that these aggregations begin by the chance meeting in a certain spot of two or three individuals. This commonly results from the fact that the organisms react to a small solid or to a roughened surface by settling against it

and remaining quiet. Two or more individuals thus settle in a certain region. This region then becomes an "attractive" area, into which other individuals enter by chance, and remain in the way just described.

In the case of *Paramecium,* tests made with sensitive indicators show that fluid of the area in which the organisms collect has a faintly acid reaction, while the remainder of the fluid in which the animals live is commonly slightly alkaline. It is at the boundary of the acid region, where the swimming individuals would pass into the alkaline fluid, that reaction occurs, the individual turning back into the acid.

By introducing with a fine pipette a bubble of carbon dioxide, or a minute drop of water impregnated with carbon dioxide, an exactly similar aggregation can be produced, the animals gathering and behaving just as in the spontaneous aggregations (Figure 4). The carbon dioxide solution is of course faintly acid (carbonic

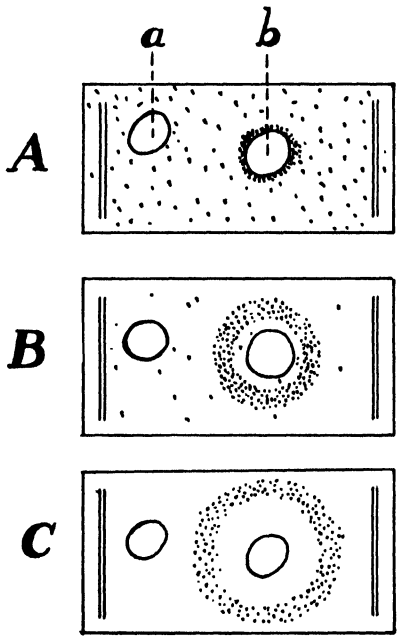


FIGURE 4. Collection of Paramecium about a bubble of CO_2, at *b*. At *a* is a bubble of air; the infusoria do not collect about it. A, B and C, successive stages. In B and C the infusoria collect in a ring at a little distance from the bubble of CO_2. (After Jennings 1906)

acid). The animals enter it freely and refuse to leave it, just as they do in the spontaneously formed groups.

Paramecium like other organisms, forms carbon dioxide in the respiratory processes, and this must diffuse into the surrounding water. This appears to be the secret of the spontaneous aggregations. As the individual gives off carbon dioxide, it becomes surrounded by an aura of that substance, having acid properties. Another individual swims by chance into this region and reacts in the way described, refusing to leave the acid. Its own carbon dioxide production is added to that of the original individual, so that the acid area becomes larger and more potent. Other individuals enter, and the area of aggregation continues to grow.

Experimentation shows that the aggregations in the case of *Paramecium* are due to the properties that are common to all acids, resulting from the presence of hydrogen ions. For weak acids of any kind induce aggregation, just as does the carbon dioxide solution.

If the acidity of the solution is sufficiently increased, the stronger acid repels the organisms. They refuse to enter it, just as they refuse to enter alkaline regions, so that an area of stronger acidity is left empty. This phenomenon is seen also in

the spontaneous aggregations. When such an aggregation has become very dense, the organisms begin to avoid its central parts. They remain in an outer ring in which the acidity is of optimum degree. Here they swim about, reacting negatively at both the inner and the outer boundary of the ring (Figure 5). Such ring-shaped aggregations are readily produced by weak acids of any kind.

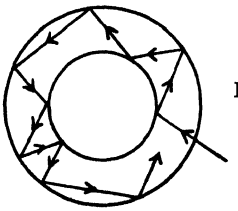


FIGURE 5. Path of the single Paramecium in a ring about an acid region. (After Jennings 1906)

The entrance to the acid region results merely from the natural lively roving movements of the animals. They do not "orient" and swim directly toward the acid area, but enter it in their roving course, and then remain.

Thus the single *Paramecium* cell is enveloped in a faintly acid aura. When a number of individuals come close together, their auras blend till a large faintly acid area is produced, in which a dense aggregation is formed.

Some other infusoria form aggregations in the same manner. *Colpidium colpoda, Cyclidium glaucoma* and *Chilomonas paramecium* collect in acid areas and form spontaneous aggregations, just as do species of *Paramecium.* When individuals of any of the above-mentioned species are mingled with those of *Paramecia,* all aggregate together in the same region.

But it is notable that certain other species of Protozoa belonging to the same group as *Paramecium* (the ciliate infusoria) form spontaneous aggregations that are not due to acid. *Oxytricha aeruginosa* and *Loxocephalus granulosus* collect in the same way as does *Paramecium,* but the areas in which they gather are not acid in reaction, and animals of these species do not aggregate in weakly acid solutions. If a mixture of individuals of *Paramecium* and of one of these species (*Oxytricha,* for example) is placed on a glass plate and covered as before described, both species form spontaneous aggregations, but the two aggregations are separate. Individuals of either species pass freely across the area of aggregation of the other species, not reacting at the boundary or remaining within the area.

Thus *Oxytricha* and *Loxocephalus* give off certain substances that form an aura about them, just as do *Paramecium, Colpidium, Cyclidium* and *Chilomonas,* but the substances are different in the two cases. The nature of the non-acid substances has not been investigated. It appears probable that a variety of such substance may be discovered when the natural history of the different species of infusoria is more fully examined.

2. There is evidence that some of the free-living Protozoa give off into the water substances that promote the reproductive processes of other individuals of the species. The picture here is far from clear, the phenomena being much less marked than in the class of cases just considered.

The fundamental observation is this: If several individuals are present in a small space, in some species each individual multiplies more rapidly than do single individuals kept separately. It appears, therefore, that the individuals in such cases probably secrete materials that pass into the water and stimulate other individuals to reproduce. The phenomenon was described for certain ciliate infusoria by Robertson in 1921; he called it the allelocatalytic reaction. Investigations have been carried out on many different species since Robertson's work, some seeming to confirm his views, others to contravene them.

In most cultures of infusoria the nutritional conditions are extremely complex, owing to the presence of bacteria, on which many species feed. This makes interpretation of the observed phenomena difficult. Investigators working with bacteria-free cultures have in a few cases confirmed the greater rate of reproduction when a number of individuals are present (Mast and Pace, 1938, in *Chilomonas;* Kidder, 1941, in *Tetrahymena*). In *Chilomonas* a substance is seemingly produced by the organisms, which in low concentrations increases the rate of reproduction, but decreases it when more concentrated.

In many cases the presence of excretory products in the water seemingly decreases the rate of multiplication. Yet cases are known in which the addition to the nutritive medium of old culture fluid, presumably containing excretory products, results in increase of reproduction. These and similar facts greatly complicate the situation as to allelocatalytic action.

Knowledge on this matter is thus in an unsatisfactory condition. The phenomena, if real, are very inconspicuous as compared with those shown in the aggregations of infusoria described in the previous section. Summaries of the data on the matter, and discussions of their significance will be found in the two chapters contributed respectively by R. P. Hall and O. W. Richards to the book "Protozoa in Biological Research" (1941), edited by Calkins and Summers.

3. Very conspicuous again are the phenomena shown in the sexual aggregations of infusoria. The organisms produce materials which cause individuals of different sex type to cling together in groups, and finally to unite in copulation or conjugation. Many remarkable phenomena result. The basic phenomena are as follows:

In some species of Protozoa there are two sex types, as there are two sexes in higher animals and plants. In other species or varieties there are four, eight, or even a larger number of sex types. When all the individuals in a region belong to the same sex type, the individuals remain scattered, swimming about freely (as in the upper left photograph, Figure 7). But when individuals of two different sex types (of the same species or variety) are mingled, the individuals cling together, forming clots or masses containing many individuals. Photographs of such clots, in comparison with the scattered individuals of single sex types are shown in Figures 6 and 7. The former shows the phenomena in certain minute flagellate infusoria; Figure 7 shows them in the large ciliate infusoria. In superficial features at least the phenomena resemble the agglutination of bacteria or of blood cells.

Details of the clotting may be illustrated from *Paramecium bursaria* (Figure 7), in which the cells are large enough so that the behavior of individuals can be ob-

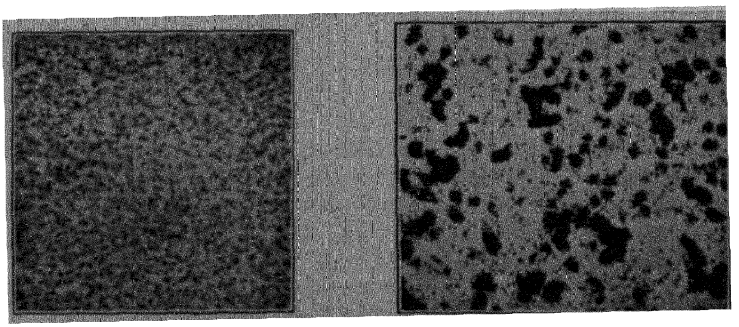


FIGURE 6. Clots formed by *Chlamydomonas eugametos.* At the left is shown the condition when individuals of but one sex type are present; they remain scattered. At the right, wo sex types are present; by their adhesion they have formed large groups or clots. After Moewus 1938)

served. The cell bodies are covered with fine cilia, by the motion of which the organisms swim about in the water.

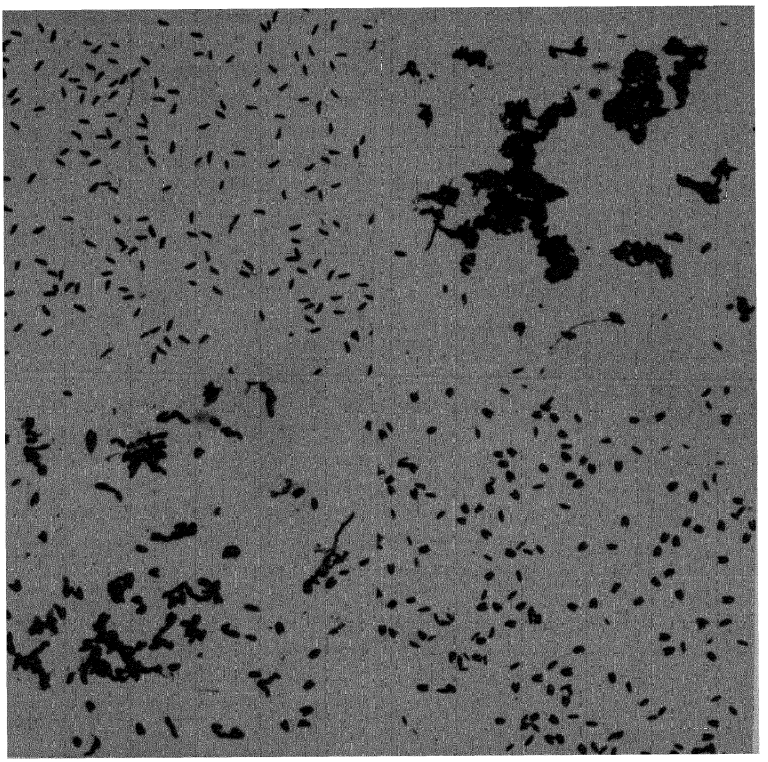


FIGURE 7. Photographs of the mating behavior in *Paramecium bursaria*. Upper left, individuals of but one mating type; they remain scattered. Upper right, mixture of two mating types, six minutes after the two were mingled. The individuals have adhered to form large clots. Lower left, mixture of two sex types about five hours later. The large clots have broken down into smaller ones, chains, and pairs. Lower right, mixture of two sex types 24 hours later. Most of the individuals are in pairs. (After Jennings 1939)

In this species there are in some varieties four different sex types, in other varieties eight. The individuals of the different sex types do not obviously differ in form or structure, but behave differently in the sexual reactions. When many individuals of two or more different sex types are placed together in a small amount of water, they at first continue their usual roving movements. These movements bring some of the individuals into accidental contact. If the two individuals that touch belong to different sex types, they stick together as if their bodies were covered with a strong adhesive material. Any parts of the cell body that thus come into contact adhere. The two individuals may thus adhere by their aboral or oral surfaces, by their rear ends or anterior ends, or in any other way. Thereupon they begin to move in an irregular way consequent upon the divergent action of the cilia in the two individuals. A third and fourth individual may become attached to the two; all flounder about irregularly. The groups continue to grow by the attachment of other individuals, till large clots are formed, such as are seen in Figure 7. Figure 8 shows a number of observed cases of irregular adhesions in early stages of the phenomena when

the individuals are few. Where two individuals adhere, often one of them drags the other backward or sidewise.

At the beginning it is the cilia that stick together. The cilia that are in contact on the two individuals are immobilized, while the free cilia continue to vibrate. Later the bodies come into contact by their surfaces, the union finally becoming intimate. In the large clots the individuals are in irregular contact by any parts of the body. It is always individuals of two different sex types that adhere; individuals of the same sex type do not.

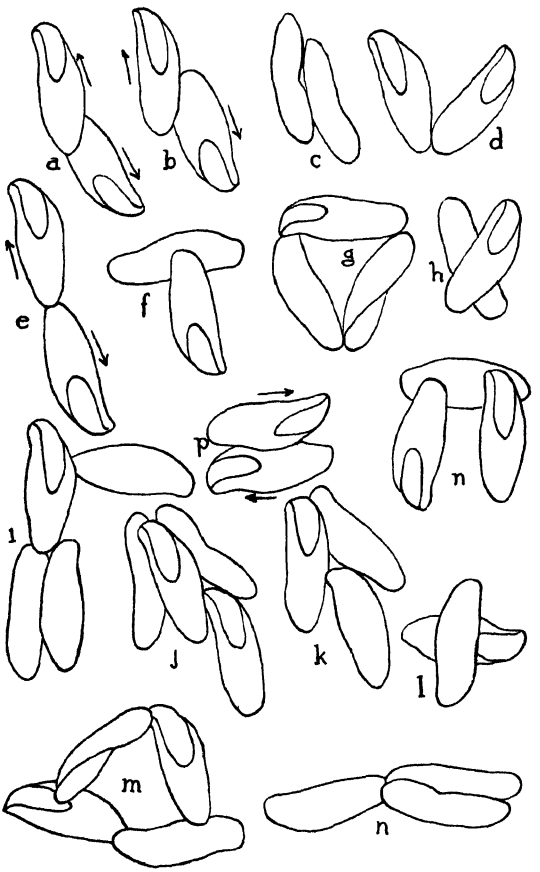


FIGURE 8. Paramecium bursaria. Irregular adhesion of individuals, in the early stages of the clotting that occurs when members of different sex types are mingled together. The arrows present in certain figures show the divergent directions in which the cilia tend to carry the different individuals. (After Jennings 1939)

In the entire behavior, in the case of ciliate infusoria such as *Paramecium*, it is evident that the coming together of the individuals is the accidental result of their natural roving movements. They do not orient at a distance and swim toward each other, or toward a group already formed. In some of the minute flagellate infusoria, on the other hand, there is said to be orientation at a distance, one cell swimming directly toward another or toward an aggregation already formed.

The large masses containing many adhering individuals usually remain united for about a half hour, then begin to break up into smaller masses. This slow break-

ing up is hastened by the approach of evening. By the disintegration of the large groups smaller groups are formed, and the process continues until only pairs of individuals remain in contact. These become attached by their oral surfaces, partly fuse, and remain united for many hours. During this period of union they exchange halves of their nuclei; that is, they fertilize one another. Then they separate, and each multiplies by fission. Such a sexual union, in which the two cells finally separate and each continues to multiply, is known as conjugation. The lower right photograph in Figure 7 shows the conditions after the clots have disintegrated and most of the organisms are in conjugating pairs.

In the Flagellates the phenomena are in the main similar to those in the Ciliata, but there are certain important differences. The flagellates are much more minute than the ciliates, so that the behavior of the individuals is not readily observable. But upon mixture of two different sex types of a species, large clots are formed by the adhesion of many individuals, just as occurs in the ciliates. This is shown in Figure 6. The large clots in the course of time break up into smaller clusters, and this continues (as in the ciliates) until only pairs of two individuals remain united. In the flagellates the two cells of the pair unite completely to form a single cell (the zygote). Such complete sexual union into one cell is known as copulation. The single cell or zygote later divides, producing many cells.

The phenomena above described suggest that in the flagellate and ciliate infusoria the cells of different sex types produce characteristic substances ("sex stuffs"), and that these sex stuffs cause the adhesion between cells of diverse sex types. The existence of such sex stuffs was demonstrated for the flagellates by Moewus, in 1933. In the species *Chlamydomonas cugametos,* Moewus removed the cells from dense cultures of the organisms, by centrifuging, or by filtration, leaving only the fluid. When this was done with cultures that were throughout of one sex type, and with the fluid so obtained individuals of the other sex type were treated, these individuals were caused to clot, just as occurs when individuals of the other sex type are added. In such clots, however, there is no final complete union of the cells, such as occurs when the two sets of cells belong to different sex types.

The sex stuffs have other effects. In a culture of *Chlamydomonas* of which all cells are of the same sex type, there is a certain period of development in which the cells are immobile and incapable of sexual union. If while in this condition they are treated with a filtrate from active functional cells that are of their own sex type (and that are capable of sexual union), this filtrate causes them to become likewise active and capable of sexual union.

Since the adhesion and aggregation of the organisms are thus shown to be due to their production of sex stuffs, the nature and properties of the sex stuffs are partly discoverable by the phenomena to which they give rise. These phenomena are varied and complex. Some of the chief features of the nature and action of the sex stuffs are the following:

(a) The adhesion and clotting that result from the presence of the sex stuffs is a reciprocal reaction between two different stuffs. In the clots of *Paramecium* the animals stick together as if covered with glue. But this is not due to a general adhesiveness of the cell bodies. The *paramecia* do not adhere indiscriminately to objects with which they come in contact. They do not adhere to inanimate objects, nor to individuals of their own sex type, nor to immature individuals, nor to individuals of another variety or species. Two different sex types or sex stuffs (belonging to the same variety or species) must be present in mature condition, for adhesion to occur.

(b) The sex stuffs exist as many diverse kinds. The number of differently acting sex stuffs in any species is necessarily the same as the number of differently acting sex types. In *Paramecium bursaria,* therefore, there are at least sixteen different sex stuffs. In Variety I of this species there are four sex types each differ-

ing in its adhesiveness from the other three. In Variety II eight sex types exist, each reacting differently from any of the other seven, or from any of the four of Variety I. In Variety III there are again four different sex types, none of them adhering to any of the sex types of Variety I or Variety II. Thus each of the sixteen different sex types in the three varieties of *Paramecium bursaria* shows different phenomena of adhesion from any of the other fifteen. In *Paramecium aurelia* three varieties are known, each with two sex types, so that the number of known different sex types is six. In the flagellate *Chlamydomonas* two diverse groups of sex types are known. In one of the groups there are six diverse interacting sex types, in the other group eight. In certain other flagellates and ciliates numbers of diverse sex types are known.

(c) The nature and properties of the sex stuffs have been directly studied mainly by Moewus (1938, 1940) in the flagellates. In *Chlamydomonas* there are in any variety two sex types, which Moewus calls male and female: we may follow him in these designations. In *Chlamydomonas eugametos* there is an inactive stage, from which the organisms are roused to activity by light. When first aroused to activity they are not capable of sexual union. But if subjected to light (blue or violet rays) for fifteen to twenty minutes, they become so; at this time active sex stuffs are presumably present. To make them capable of sexual union, male cells require longer subjection to light than do female cells.

In the stage when the cells are active (but still not capable of sexual union) the cells may be filtered out of the fluid in which they are found. In this way is procured a filtrate that is free from cells. Such a filtrate may be prepared from a culture of either male or female cells.

The filtrate thus obtained is inactive at first. But if subjected to light (blue) it produces the active sex stuffs (which induce adhesion and clotting as above described). When subjected continuously to light it first (after about twenty minutes) produces the female sex stuff. If the action of the light is continued (to about thirty minutes) the filtrate produces the male sex stuffs. The two sex stuffs thus produced from the filtrate have the characteristic effects of sex stuffs above described. Subjected for a still longer time to light the filtrate becomes inactive. The changes just described occur in the same way whether the filtrate is prepared from male or from female cells.

Thus four stages may be distinguished in the material of the filtrates: (1) a preparatory stage, inactive: (2) the female sex stuff: (3) the male sex stuff: (4) a final inactive condition.

If a filtrate is procured from active functional female cells (female sex stuff), this when subjected to blue light transforms into the male sex stuff. But the female stuff cannot be obtained from the male sex stuff.

By making mixtures of the preparatory stuff (1) and the inactive end stuff (4), the two sex stuffs can be produced. The female sex stuff is produced by mixing three parts of the preparatory stuff (1) with one part of the end stuff (4). The male sex stuff is derived by mixing one part of the preparatory stuff (1) with three parts of the end stuff (4).

It is clear, therefore, that the male and female sex stuffs (as well as the preparatory and end stuffs) are related substances: light transforms one into the other.

In the course of the experiments above described, Moewus tried the effects of various organic compounds, to see if similar phenomena could be observed. From the fact that the stuffs are transformed by blue and violet light, he was led to suspect that they are probably related to yellow or orange-yellow organic compounds. He was therefore led to experiment with such compounds (carotinoids) derived from the carrot, *Daucus carota*. He made watery extracts from the carrot, and found that one of these had an action similar to that of the female sex stuff. Moreover, when it had stood a while in the light, it acted like the male sex stuff.

This led Moewus to experiment with a large number of known carotinoids. Most were ineffective. But two of them acted like the sex stuffs. One of these, *cis*-crocetin dimethylester, has the properties of the preparatory stuff (1), while the related *trans*-crocetin dimethylester has the properties of the final inactive stuff (4). Mixtures of three parts of the *cis* compound with one part of the *trans* compound have the properties of the female sex stuff, while one part of the *cis* material with three parts of the *trans* material yields the properties of the male sex stuff.

It appears, therefore, that the male and female sex stuffs of *Chlamydomonas eugametos* are closely related materials belonging to the carotinoid group of compounds: they are presumably identical with mixtures of the *cis* and *trans* forms of crocetin dimethylester.

This work was extended by Moewus to include the sex stuffs in certain other varieties or species of *Chlamydomonas,* the sex types of which interact in accordance with certain rules. In these varieties eight different sex stuffs may be distinguished. They turn out to be different percentage combinations of the *cis* and *trans* forms of crocetin dimethylester. We may take from Sonneborn (1941, p. 677) a table showing the percentages of the *cis* and *trans* materials in these eight sex stuffs (designated in the table by the letters G to O).

	G	H	J	K	L	M	N	O
Percentage *cis*	95	85	75	65	35	25	15	5
Percentage *trans*	5	15	25	35	65	75	85	95

In this series of sex types, individuals or varieties having the same sex stuff do not unite, nor do those differing in the percentage combination by less than 20 per cent. But any individual or variety unites with others that differ from it by 20 per cent or more in its combination of *cis* and *trans* substances.

Moewus calculates that the latter are effective in producing copulation in a dilution of one part in 33 billion.

No such investigation has been made of the chemical nature and physical properties of the sex stuffs in the ciliates. But the similarity of the phenomena in this group to those in the Flagellata strongly suggests that they are brought about in a similar manner.

(d) Certain pecularities of the phenomena in the ciliates suggest a possible difference in the nature of the sex stuffs in the two cases. In the flagellates the sex stuffs are water-soluble materials (Moewus, 1938, p. 765), which are obtainable free from the organisms by filtering or centrifuging. They are thus present in solution in the culture medium. With this agrees the method of reaction of individuals of one sex type to those of the other sex type, as described by Pascher (1931) in *Chlamydomonas.* The active cells of one sex type orient about a quiet cell of the other sex type and from a distance swim directly toward it. Such reaction at a distance and orientation requires that the effective substance shall be dissolved in the water surrounding the cells. But in the ciliates the phenomena suggest that the adhesion and aggregation are surface phenomena, the sex stuffs being active only at the surface of the cells, and not in solution in the surrounding water. For in these organisms there is no reaction at a distance; the cells of one sex type do not orient and swim from a distance toward a cell of another sex type. The adhesion occurs only when the animals come in contact through their roving movements. These natural roving movements occur equally when cells of only one sex type are present. Such cells come in contact as frequently as do cells of different sex type, but when the cells are of the same sex type they do not adhere. It is desirable that investigation be directed upon the question whether in the ciliates the sex stuffs are dissolved in the surrounding fluid, or occur only on the cell surface.

(e) In the life history of the ciliates there is a period of immaturity, during which the organisms do not produce sex stuffs, and there is no adhesion, aggregation

or conjugation. This period of immaturity lasts for many cell generations. An individual cell that has just completed conjugation divides into many cells by repeated fissions, producing a large population known as a clone. All members of the clone are for a long time immature, showing no sexual reactions of any sort. In *Paramecium bursaria* immaturity continues for a period of from one to several months. During this time fission takes place at the rate of about one a day, so that millions of immature individuals are produced. Then comes a period during which maturity slowly develops. Now the cells presumably produce the sex stuffs, since they show the selective adhesion, aggregation and conjugation above described. The period of immaturity is one in which the cells neither produce sex stuffs, nor react to sex stuffs from other cells. For if to a clone of immature individuals we add a clone of highly reactive mature individuals of different sex type, there is no adhesion, clotting or conjugation between the two.

(f) In some of the ciliates (if not all) there is a daily periodicity in the occurrence of sex reactions, and therefore in the production of sex stuffs. In most varieties of *Paramecium bursaria* and *Paramecium aurelia* adhesion and clotting occur only at certain periods of the day. In Varieties I and II of the former the adhesion and other sex phenomena normally occur only during the daylight, from about nine A.M. to five P.M. In Variety 2 of *Paramecium aurelia* they occur only at night, from about six P.M. to seven A.M. In Variety 3 of the same species they occur only from about one A.M. to one P.M. In some other varieties there appears to be no periodicity; adhesion, aggregation and conjugation occur whenever different sex types are mixed. This is the case in Variety I of *Paramecium aurelia* and Variety III of *Paramecium bursaria.*

The periodicity in the sex reactions, and thus in the production of sex stuffs is due in the long run to the alternation of day and night. Yet the phenomena are not immediate direct effects of light and darkness. If the cultures are kept in continuous light or continuous darkness, the usual periodicity persists for some days. In this way *Paramecium bursaria* (Variety I or II), which normally reacts only in daylight, may be caused to react at the usual time though in total darkness. The periodicity of day and night has become a periodicity of organic processes, which become partly independent of the external periodicity. Yet if the periodicity of light and darkness is totally suspended for a long time, the periodicity of the sexual reactions is in time broken up and disappears. Sonneborn (1938) shows that in *Paramecium aurelia* (Variety 3) after five days of continuous light the sexual reactions are entirely lost, while after five days of continuous darkness the adhesion and clotting will occur at any time the two sex types are mixed.

(g) When the members of one clone have become mature, in *Paramecium bursaria* all are found to belong to the same sex type. As the clone continues to multiply, all the descendants by ordinary fission continue to remain of the same sex type, so that in ordinary fission the sex type is strictly inherited. This is the rule also in *Paramecium aurelia,* and seemingly in most species.

Even in the immature condition there are inherited differences between clones, corresponding to the later differences in sex type. For all the members of a given immature clone later develop the same sex type, and hence must produce the same sex stuff; whereas in another clone a different sex type (and sex stuff) is produced.

(h) But there occur in a clone at long intervals nuclear processes of special type, which result in the differentiation of the members of the single clone into two different sex types, producing different sex stuffs. Such phenomena, varying in different species, are known to occur both in the flagellates and the ciliates. They need not be more fully described here.

(i) At sexual union (conjugation in the ciliates, copulation in the flagellates), followed by fission, inheritance of the sex types occurs in accordance with varied and complex rules.

For details as to the rules of inheritance, and in general as to the biological rôles of the sex types, the chapters contributed by Sonneborn (1941) and Jennings (1941) to the volume "Protozoa in Biological Research" (edited by Calkins and Summers) may be consulted. These chapters contain extensive lists of references to the papers setting forth the original investigations on these matters.

References

Calkins, G. N., and F. M. Summers, Editors, "Protozoa in biological research," New York, 1941.

Hall, R. P., "Food requirements and other factors influencing growth of Protozoa in pure cultures," in "Protozoa in biological research," pages 475-516, New York, 1941.

Jennings, H. S., "Behavior of the Lower Organisms," New York, 1906.

—, "Genetics of *Paramecium bursaria*. I. Mating types and groups, their interrelations and distribution; mating behavior and self sterility." *Genetics*, **24**, 202-233 (1939).

—, "Inheritance in Protozoa," in "Protozoa in biological research," pages 710-771, New York, 1941.

Kidder, G. W., "Growth studies on ciliates. V. The acceleration and inhibition of ciliate growth in biologically conditioned medium," *Physiological Zoology*, **14**, 209-226 (1941).

Mast, S. O., and Pace, D. M., "The effect of substances produced by *Chilomonas paramecium*," *Physiological Zoology*, **11**, 350-382 (1938).

Moewus, F., "Untersuchungen uber die Sexualität und Entwicklung von Chlorophyceen," *Arch. Protistenk.*, **80**, 469-526 (1933).

—, "Carotinoide als Sexualstoffe von Algen," *Jahrh. wiss. Bot.*, **86**, 753-783 (1938).

—, "Carotinoid. Derivate als Geschlechtshestimmende stoffe von Algen," *Biol. Zbl.*, **60**, 143-166 (1940).

Pascher, A., "Gruppenbildung und 'Geschlechtswechsel' bei den Gameten einer Chlamydomonadine (*Chlamydomonas paupera*)," *Jahrb. wiss. Bot.*, **75**, 551-580 (1931).

Richards, O. W., "The growth of the Protozoa," in "Protozoa in biological research," pages 527-564, New York, 1941.

Robertson, T. B., "Experimental studies on cellular multiplication. II. The influence of mutual contiguity upon reproductive rate and the part played therein by the 'X-substance' in bacterial infusions which stimulate the multiplication of infusoria," *Biochem. J*, **15**, 612-619 (1921).

Sonneborn, T. M., "Mating types in *Paramecium aurelia*: diverse conditions for mating in different stocks: occurrence, number and interrelations of the types," *Proc. Am. Phil. Soc*, **79**, 411-434 (1938).

—, "Sexuality in unicellular organisms," in "Protozoa in biological research," pages 666-709, New York, 1941.

The Primary Physicochemical Basis of Organic Evolution

C. R. PLUNKETT

Definitions are made by man, and ignored by nature. It is a common experience in all fields of science that while increasing knowledge, up to a certain point, tends to bring out sharp distinctions between things formerly confused, beyond that point further advances often tend to break down these sharp distinctions. The distinction between living and lifeless things seems to be a case of this kind. This distinction was built up slowly, over a long period, with the gradual decline of primitive "animism," the recognition of the features common to plants and animals, the discovery of microorganisms, the step-by-step disproof of "spontaneous generation," etc. This phase reached its culmination during the nineteenth century, with the establishment of an "absolute" distinction between living and non-living things, embodied in various epigrammatic definitions of "life," and elevated to the status of a metaphysical categorical difference beyond the scope of scientific elucidation.

But this laboriously built boundary wall had scarcely been completed, before it began to show cracks. The theory of evolution itself, though on the one hand it unified the biological sciences, on the other hand tended to raise anew, and in a quite different aspect, the question of the possible natural origin of living from lifeless things. So also the cell theory, while aiding in the unification of biology, tended to encourage the further pursuit of "life," tracked down from the whole organism to the cell, into the still smaller components of the cell. Later this pursuit was greatly advanced by the gene theory, which has carried it down, through the gene, close to the chemical molecule.

Meanwhile the boundary wall was also being undermined from the other side by the artificial synthesis of ever-increasingly complex "organic" compounds, by the isolation, crystallization, analysis, and synthesis of the "mysterious" enzymes, vitamins, and hormones, by the demonstration of the conservation or energy in the living body, and later the step-by-step disproof of the alleged "exception" of living things to the second law of thermodynamics. Also from the physicochemical side, the advance of colloid chemistry and the theory of catalysis pushed upward, through the chemical molecule and the colloidal micelle, toward a junction with cytology and genetics pushing downward through the chromosome and the gene. Then, from an unforeseen direction, appeared a new opening—the discovery of the filtrable viruses—which went far toward completing the breach in this boundary wall: so far, indeed, that there is no agreement at all, even among those who would still maintain this wall, as to which side of it the viruses belong.

More recent advances in several fields of biology and chemistry—virus study, cytochemistry and gene physiology, the chemistry and physiology of enzymes and nucleoproteins, the general theory of catalysis, etc.[1]—have still further reduced the sharpness of the distinction between living and lifeless things, and leave little further excuse for regarding this difference as "absolute," or treating it as a metaphysical question outside the scope of scientific investigation. Rather must it be treated as a special bio-physico-chemical problem, to be elucidated by the continued development of experimental methods and the correlation of results in the several pertinent fields of investigation.

The characteristic properties which are supposed to distinguish living from non-living things are variously enumerated, defined, and emphasized by different biologists. The more generally cited of these characteristics include: the chemical composition of living bodies, characterized by the presence of complex carbon compounds, especially proteins, and particularly of specific optically active stereoisomers (in contrast to the racemic mixtures generally produced by artificial syntheses); the physicochemical structure of living matter, a complex, unstable, polyphase colloidal system, to which the rather indefinite term "protoplasm" is applied; over and above this general ultramicroscopic structure, the specific, highly differentiated, microscopic and macroscopic structure of cells and multicellular organisms (often designated as "organization"); metabolism, in its broadest sense including all the chemical reactions occurring in living organisms; liberation of energy in various forms, including movements, heat, bioelectric currents, etc.; responsiveness (also called "irritability" or "reactivity"), involving, in some of the higher organisms at least, subjective ("mental") aspects correlated with its objective manifestations; growth (often specified as being "by intussusception," which is merely a consequence of the heterogeneous, largely liquid, structure of protoplasm) and reproduction, involving the phenomena of heredity and variation. It is notoriously difficult, if not impossible, to define any of these properties in such a way as to distinguish them sharply from comparable properties displayed—though usually in simpler and more generalized form—by at least some non-living systems. Moreover, these properties are obviously not independent of, or sharply separable from, each other. And as to what particular property, or combination of properties, is most essentially distinctive or diagnostic of "life," there is the widest possible divergence of opinions among workers in the various fields of biology and biochemistry.

Once we frankly recognize the difference between living and lifeless things as relative, rather than absolute, we may approach this question in a somewhat different way: we may regard the complex of phenomena which we call "life" as characterized not so much by any unique special property, or properties, as by certain general features displayed in especially high degree by many of the properties of those bodies recognized, by general consent, as "living." These general features—which are entirely relative, and all closely interrelated—may be designated as *complexity,*

specificity, diversity, and ***adaptation*** (though they might be equally well named, divided, and combined in many other ways).

It is fairly generally recognized that many, if not all, of the characteristic properties of living organisms, whether chemical or physical, structural or functional, are distinguished from comparable properties of some non-living systems, not by any strictly definable qualitative differences, but rather by their higher degree of complexity. The composition and structure, the chemical reactions and physical activities, the responses and the growth of organisms are all much more complex than those of any non-living bodies. Even so complex and apparently peculiar a phenomenon as reproduction may be resolved into component processes (division, specific growth, etc.) each of which has simpler counterparts in some non-living systems.

The high degree of specificity so characteristic of many properties of organisms—their chemical components and reactions, their structures, responses, growth, etc.—is obviously correlated with their complexity. Simpler things, *e.g.,* inorganic chemical substances, may be just as specific, or even more so, but their specificity is likely to be less impressive because it does not involve such complex qualities and activities. While some of the specificity of organisms is individual, the greater part of it is group specificity (species-specificity, etc.) common to a group of individuals related by descent, *i.e.,* it is transmissible, or hereditary, specificity.

Diversity is, on the one hand, a correlative of specificity (for the specificity of one thing implies its diversity from other things, and conversely); but on the other hand, considered genetically rather than statically, it implies a limitation of hereditary specificity: for any diversity among individuals related by descent indicates that this specificity is not absolute, but more or less modifiable. Some, though not all, of the diversity originating by modification of transmitted specificity is itself specifically transmitted in turn. This continual modification and diversification of hereditary specificity by new heritable diversity results in the complex gradation of resemblances and differences among organisms, which involves all their characteristics more or less correlatively (*e.g.,* morphological similarities are correlated with similarities of specific chemical components, etc.).

The word "adaptation" has fallen somewhat into disrepute among many biologists because it has sometimes been used with teleological or vitalistic implications; but it is a perfectly good word to designate one of the most striking and universal features of practically all the characteristics of living organisms. In its most general sense, adaptation means merely "fitness to exist" in a given environment, which implies simply a certain degree of stability within a certain range of conditions. In this general sense, any organism is less "adapted" than many inorganic bodies, which are more stable and continue to exist much longer. It is, evidently, the complexity, the specificity, and the diversity of the adaptation of an organism which make it so remarkable. Adaptation, in any individual organism, has two aspects: first, its inborn, inherited, structural and functional "adaptedness"; and second, its "adaptability," or ability to change individually to meet changing conditions. It is sometimes supposed that this second aspect, individual "adaptability," is primary—an aboriginal, essential attribute of "life"—giving rise, somehow, to all the diverse specific hereditary adaptations of organisms (*e.g.,* Lamarck's theory of the inheritance of the effects of use and disuse, etc., in which the original effects, in the individual, are taken for granted without explanation). Actually, the relation between these two aspects of adaptation is just the opposite: individual "adaptability" is not the antecedent but a consequence—indeed, but one manifestation—of specific hereditary "adaptedness": an organism responds in a way advantageous to itself (*i.e.,* "adapts" itself) to a particular change of conditions only because, and if, it is already provided with some inherited special adaptive mechanism whose function it is to respond in just that particular way to that particular change. In many cases, of

course, one special adaptive mechanism may cover a number of contingencies which are not exactly alike but have some elements in common; but, in general, an organism does not respond "usefully" to a situation which is entirely new, or extremely unusual, in the history of its race. The complexity, specificity, and diversity of the adaptations of organisms are practically coextensive with the total complexity, specificity, and diversity of all their properties; indeed, there is relatively little of anything in an organism which is not adaptive. This situation has a completely adequate causal explanation in the theory of natural selection, which rests solely on the universal characteristics of biotic reproduction: multiplication, heredity, and variation.

Multiplication, heredity, and variation, though traditionally treated as separate "factors" in organic evolution, are really only conceptually separable aspects of a single process. For "heredity" implies *specific multiplication*, whereas "variation" implies that this specificity is somewhat *modifiable*: we may, therefore, characterize biotic reproduction as "modifiably specific multiplication." However, not all variation (*i.e.*, all diversity among organisms related by descent) is significant for evolution, but only *heritable* variation, *i.e.*, such variation as, originating by modification of self-reproducing specificity, is itself specifically self-reproducing thereafter (subject, of course, to further modification of the same kind). This kind of variation is now called "mutation" (disregarding disputed questions of technical definition, occasioned by various complications in the higher organisms). We may say, therefore, that the primary basis of organic evolution is *mutably specific multiplication*.

If the properties of living things differ from those of non-living things only in degree (of complexity, etc.) we might expect to find in some non-living systems properties comparable in some degree to mutably specific multiplication, though presumably in simpler and more generalized form. In the field of *catalysis* we find many examples of extreme specificity and also of modifiable specificity.[2] We also find a form of catalysis involving self-increase, or multiplication: namely, *autocatalysis*. By autocatalysis, in general, is meant catalysis of a reaction by one of its own products; the substance thus catalyzing its own production is called an *autocatalyst*. Autocatalysis, in this general sense, is no different, in regard to the chemical mechanisms involved, from heterocatalysis (*i.e.*, non-autocatalytic catalysis); it just happens (so to speak) that some substances catalyze reactions of which they themselves are products. Thus, for instance, the hydrolysis of an ester (in originally neutral solution) is autocatalyzed by the free acids produced; but this is no different, in its chemical mechanism, from the heterocatalysis of this reaction and other hydrolytic reactions by any acid. Again, the oxidations of many metals, and of various organic substances, are autocatalyzed by some of their oxidation products (peroxides) which act as oxygen "carriers"; this is the same mechanism by which, in many oxidative reactions, "foreign" peroxides act as heterocatalysts. While the rate of a heterocatalytic reaction is usually either continuously decreasing or constant, depending on the relative concentrations or effective surface areas of catalyst and reactants and other conditions of the system, the rate of an autocatalytic reaction, *under the simplest conditions* (*i.e.*, in a completely homogeneous closed system), will increase continuously for some time, because of the increasing concentration of the autocatalyst; but under other conditions, *e.g.*, insolubility of the autocatalyst or some of the reactants, continuous removal of the autocatalyst from the reaction system, etc., the rate of an autocatalytic reaction may also be constant or even decrease continuously from the beginning; the form of the reaction velocity curve is therefore neither a sufficient nor a necessary criterion of an autocatalytic reaction, especially in a complex heterogeneous system.

Catalyst *specificity* is, of course, a relative matter, and there are various degrees of specificity in different cases, in general chemistry as well as in biochemistry. The specificity of a catalyst consists in its effectiveness in catalyzing selectively *one*

particular reaction (involving certain particular reactants and yielding certain particular products), as measured by the rate of this particular reaction *relative* to the rates of other (simultaneous or possible) "competing" reactions (whether involving some different reactants or yielding different products from the same reactants) when catalyzed by this particular catalyst, *as compared with* the rate of this particular reaction relative to those of the "competing" reactions when catalyzed by any other catalyst (or uncatalyzed). In the case of autocatalysis, a highly specific autocatalyst would be one which catalyzes very selectively a particular reaction, one of whose products is *identical* with the autocatalyst (not merely any substance in the same general class), which would otherwise occur much more slowly, if at all, whether uncatalyzed or catalyzed by any other catalyst. Most of the autocatalytic reactions well known in general chemistry are not very specific: *e.g.*, the acid hydrolysis of esters, the "autooxidation" of metals and of various organic substances, etc. These reactions may be catalyzed by other catalysts (*i.e.*, heterocatalysts) just as effectively as by their own autocatalysts, or even more so; and the autocatalysts are merely general catalysts for all reactions of a certain type, of which the autocatalytic reaction happens to be one. Although there are many well known and extensively investigated cases of highly specific *heterocatalysis* in general chemistry, I have been unable to find any well established cases (not involving biocatalysts) of very specific autocatalysis of a strictly "chemical" nature. On the rather ill-defined borderline between "chemical" and "physical" processes, however, there is a very extensive and well known class of phenomena which are highly specific and comparable in some respects, at least, to autocatalysis: namely, *crystallization.*[8]

A crystal accelerates, or even initiates, very selectively, the formation from certain particular substances of more material identical (in most cases) with itself, which would otherwise occur much more slowly, if at all. Often it is difficult, in some cases almost impossible, to initiate crystallization except by "inoculation" with a crystal of the particular substance concerned; but even when crystallization starts spontaneously (*i.e.*, without "inoculation") the first-formed, ultramicroscopic crystals must greatly accelerate the process of crystallization at their surfaces; otherwise, we should never get visible (even microscopic) crystals at all, but only apparently "amorphous" precipitates. The degree of *chemical* specificity of crystal growth varies in different cases (*e.g.*, the cases of isomorphous substances), though in most cases it is highly specific; but the *crystal structure* is always highly specific.

Crystals of different (non-isomorphous) substances may grow simultaneously in the same solution, each crystal selecting and combining the particular kinds of molecules or ions of which it is itself composed. In the case of non-electrolytes (such as many organic compounds), this consists merely in selectively adsorbing already existing molecules of its own kind and arranging them in a definite pattern, or "space lattice"; but in the case of strong electrolytes (*e.g.*, salts), it involves taking up ions of two or more different kinds, in definite proportions, and arranging them in definite spatial relations to each other. Just such selective combination, definite orientation, and intimate juxtaposition of specific molecules or ions are indispensable prerequisites of chemical synthesis, and especially characteristic of surface catalysis; but in crystallization these preliminaries are not followed by "chemical" combination, in the usual sense of the word, that is, the formation of new primary co-valence bonds, such as are especially (though not exclusively) characteristic of organic compounds (*e.g.*, such bonds as C-C, C-N, C-O-C, etc.). For this reason, crystallization is not usually regarded as a chemical process, and hence we do not speak of it as "catalysis."

In contrast to such strictly "chemical" bonds, which are both strong and specific, the forces binding the components of a crystal are of two kinds: the strong but non-specific electrostatic attractions between oppositely charged ions (in electro-

lytes); and the specific but usually relatively weak "secondary valence" forces, or residual external electric fields of molecules. The fact that practically all substances may crystallize, under proper conditions, indicates that there is always some specific attraction between molecules of the same kind, though its strength varies widely in different substances.

The same specific "iso-molecular" cohesive forces are undoubtedly involved also in the formation of other kinds of iso-molecular aggregates, such as colloidal particles, fibers, and gels, which are especially characteristic of many complex biochemical substances, *e.g.*, proteins and polysaccharides. Specific intermolecular attractions also exist between molecules of different kinds, but the strength of these "hetero-molecular" forces varies even more widely in different cases: they may be much weaker than iso-molecular cohesion, or apparently even non-existent, in many cases (*e.g.*, between hydrocarbons and water), while at the other extreme (in so-called "molecular compounds") their strength approaches that of primary chemical bonds. In specific heterocatalysis the specific attractive forces between catalyst and reactants are strong enough to distort the molecular electric fields of the reactants sufficiently to disrupt (or greatly weaken) their primary chemical bonds and thus permit the formation of new bonds: this constitutes chemical "synthesis" or "decomposition," according to the relative sizes of the original and the resultant molecules. (Of course the distinction between primary and secondary valences is not entirely sharp, and therefore the question whether the "intermediate compounds" of catalyst and reactants are "adsorption compounds" or truly "chemical" compounds may often be one of somewhat arbitrary definition.)

From the foregoing considerations, it would appear that *specific autocatalysis* must involve some sort of combination of the kind of forces involved in specific iso-molecular aggregation (crystallization, etc.) and the kind involved in specific catalysis, especially surface catalysis, which may include catalysis at the surfaces of relatively large discrete molecules; in other words, specifically iso-molecular attractive forces with some of the special characteristics of the hetero-molecular forces between catalyst and reactants which facilitate the disruption or redistribution of primary co-valence bonds, resulting in chemical recombinations between the reactant molecules.[4] Since these two kinds of forces (*viz.*, iso-molecular and catalytic) are both, separately, very common, and are moreover fundamentally similar, both being manifestations of residual external electric fields of molecules, we might expect their combination to be not extremely unusual; but the fact is that highly specific *auto*catalysis is apparently still unknown in indisputably non-living matter. This would seem to suggest that some very special feature of molecular composition or structure may be necessary for a specifically autocatalytic substance; *e.g.*, the presence of some particular atomic grouping in the molecule, or perhaps some peculiar feature of the structure of the molecule as a whole. It should be noted, however, that certain special conditions of the milieu may be equally necessary for the autocatalytic activity of such a substance, especially if the autocatalytic reaction is one by which the autocatalyst is synthesized from simpler substances. The conditions must be such as to permit this reaction to proceed predominantly in the synthetic direction; for if the reverse reaction predominated it would result in the self-accelerated decomposition of the autocatalyst and eventually (unless an equilibrium were reached) its complete loss.

Predominance of the synthetic reaction requires either that all the component substances, from which the autocatalyst synthesizes itself, be maintained in sufficient concentration, in the immediate milieu, to force the reaction in the synthetic direction (by "mass action"), or else that the synthetic reaction be practically irreversible under the conditions of the system. The latter alternative implies that some product of the reaction (either the autocatalyst itself or some "by-product" of its synthesis) be continuously removed from the phase in which the reaction occurs; *e.g.*, if the

synthesis involves combination of two or more molecules with elimination of the elements of water (so that the reverse reaction is hydrolytic), the autocatalytic synthesis must either proceed in a non-aqueous phase, from which the water is continuously removed by some means (chemical or physical) as fast as it is formed, or else the autocatalyst itself must escape continuously, as it is formed, from the aqueous phase, *e.g.*, by precipitation or diffusion. In a heterocatalytic reaction, the reaction can be kept going in the desired direction by the continuous removal of the products, as formed, from contact with the catalyst, but this is not possible in an autocatalytic reaction, for the autocatalyst cannot be removed from contact with *itself*. In addition, other special conditions of the reaction system (temperature, pH, other ionic activities, oxidation-reduction potential, etc.) will presumably be necessary for the autocatalytic reaction, as in other catalytic processes. The more complex the autocatalytic substance and the reaction by which it synthesizes itself may be, the more complexly and specifically "adapted" to this reaction, presumably, must be its milieu. (In this connection it is significant that all the synthetic reactions in a living organism cease, and decomposition proceeds irreversibly, as soon as the organism is "dead," *i.e.*, upon the failure of any of the complex of conditions necessary for the synthetic processes.)

Modification of catalyst specificity (however effected) consists in increasing or decreasing the effectiveness of the catalyst on one particular reaction, while leaving unchanged, or changing to a different extent or in the opposite direction, its effectiveness on other "competing" reactions: thus the modification may increase or decrease the original specificity of the catalyst or alter it qualitatively (*i.e.*, render the catalyst specific for a different reaction from the one for which it was originally specific). Many cases of such catalyst modification are well known in general chemistry. Changes of specificity of biocatalysts also can be effected artificially; and it is not improbable that such modifications may also occur naturally, in the living organism, and may play important roles in various biological processes.[2] Such modification, when a *hetero*catalyst is concerned, does not constitute "mutation," as this term is used in biology; for, since the heterocatalyst is not self-reproducing, *i.e.*, does not multiply itself, as an autocatalyst does, its modifications also are not self-reproducing, unless, of course, the modification should convert the heterocatalyst into an autocatalyst. Modification of the specificity of an *auto*catalyst, however, would be essentially equivalent to mutation, *provided* the modified catalyst is also specifically autocatalytic *and* specifically different, in respect of its autocatalytic reaction, from the original one.

These qualifications may be illustrated by consideration of an analogous case in the field of crystallization. Suppose a slowly evaporating solution, saturated with respect to a number of different solutes (in the form of ions or molecules or both), in which a number of different kinds of crystals are growing. If we speak of crystal growth as an "autocatalytic" process, simply for the purpose of the analogy, each kind of crystal is specifically "autocatalytic" and specifically different, in respect of its "autocatalytic reaction," *i.e.*, its growth, from every other kind (the specificity referred to is specificity of crystal structure, not necessarily of chemical composition). Now suppose we remove one particular crystal from the solution, and change its composition or constitution in any way, *e.g.*, by chemical substitution or isomeric change, by addition or subtraction of "water of crystallization," etc., and then return the resulting product, whatever its form, to the solution. Now, in regard to the "autocatalytic" (growth) properties of this "modified" material, we may distinguish three cases: (1) It may be non-crystalline, in which case it will be incapable of specific growth, *i.e.*, it will no longer be "specifically autocatalytic." (2) It may be crystalline and isomorphous with the original crystal, in which case it will still be "specifically autocatalytic" (capable of specific growth), but will not be specifically different in this respect from the original crystal; the new material formed, as the

crystal grows, will be identical in the original and the "modified" crystal, the proportions of the two components depending solely on the composition of the solution. (3) It may be crystalline and non-isomorphous with the original crystal, in which case it will be "specifically autocatalytic" and also specifically different in this respect from the original crystal, for it will now have a different specific growth form and will take up different substances from the solution in its subsequent growth: this last case is analogous to mutation. Of course, if some constituent of this "mutant" crystal is not present in this solution, or other conditions (of temperature, concentration, etc.) render this particular crystal form unstable in this solution, it cannot grow here but will, on the contrary, dissolve and may eventually be lost entirely: in biological terms, it is "inviable" under these particular conditions because it is not "adapted" to them, although it might be "viable" (capable of existence and growth) under other conditions.

Now let us imagine a substance of the "organic" type, in the purely chemical sense, made up of several different radicals united by primary co-valence bonds (*e.g.*, a polypeptide, composed of several amino acids), in a solution containing all these radicals in the form of separate molecules (or ions). Let this imaginary substance be endowed with two properties: first, it adsorbs specifically molecules (or ions) identical with its own constituent radicals, in a spatial pattern identical with its own molecular structure, as a salt crystal takes up ions identical with its own constituents in a pattern identical with its own structure; secondly, it catalyzes the *chemical* combination of these adsorbed molecules (or ions) with one another by the same kind of linkages by which its own constituent radicals are united, *e.g.*, if a polypeptide, by peptide linkages between the successive amino acids. The product of this chemical reaction would thus be a molecule identical with that of the original substance, which is therefore a *specific autocatalyst.* This substance would thus continue to increase so long as the conditions (of concentration, etc., as previously discussed) are such as to permit this reaction to proceed in the synthetic direction, as a crystal increases so long as the concentration of its constituents, and other conditions, in the surrounding solution are such as to permit its growth. It need not be assumed, however, that this increase would necessarily be in the usual form of crystal growth, *i.e.*, a massive aggregate: if, for instance, the molecule of the autocatalyst is "flat," with only two (opposite) surfaces at which the specific adsorption can occur, its autocatalyzed increase might take the form of a linear filament, of unimolecular thickness, growing in a line perpendicular to the plane of this molecule; this filament, moreover, might be subject to disruption, *e.g.*, by convection currents, or its own electrostatic charges, etc., so that the autocatalyst might remain colloidally, or even molecularly, dispersed.

Now suppose some portion, let us say a single molecule, of this autocatalytic substance to undergo a chemical change of any kind (addition, subtraction, substitution, or intramolecular rearrangement). In regard to the effect of this "modification" on the autocatalytic properties of the substance, we have the same three possibilities as in our previous analogy of the crystal: (1) Its autocatalytic property may be destroyed entirely. (2) Its specific autocatalysis may not be altered at all. For instance, if the change consists in the addition of an extra radical to one end of the molecule, the property of specific adsorption may not extend to this new radical, and the molecule may continue to autocatalyze the synthesis of the original substance without this addition; or some relatively slight chemical change in one of the constituent radicals may not destroy its power to adsorb selectively the molecules or ions corresponding to its original form; etc. In this case the modification will not be perpetuated. (3) It may be specifically autocatalytic in the new form, which is specifically different in this respect from the original one. For instance, an intramolecular rearrangement of the constituent radicals may result in a corresponding change in the spatial pattern of the molecules or ions adsorbed, and

consequently of the new autocatalyst molecules synthesized; or an added or substituted radical may share the power of specific adsorption of molecules or ions of its own kind, resulting in their incorporation in the newly synthesized autocatalyst molecules; or the change might be of the nature of polymerization, perpetuated in the subsequent autocatalysis, etc. In this case the modification will be self-perpetuating, provided the surrounding conditions, as previously discussed, permit this. This last type of modification constitutes "mutation," and an autocatalytic substance capable of this kind of change would be a *mutably specific autocatalyst.* It is to be noted that this term does not imply that *every* change in such a substance must be of this kind; most changes would be more likely to be of the other two kinds mentioned above, especially the first.

No indisputably non-living substance or material with the property of mutably specific autocatalysis is known. We say "indisputably" non-living because it is precisely in the disputed border zone between the "living" and the "non-living" that we find the clearest examples of mutably specific autocatalysts, namely, the *viruses.* The propriety of speaking of a virus, even when it is crystallizable, as a "substance" may be questioned, because crystallization is not a sufficient criterion of a pure substance, in the strict chemical sense; but of course the same objection may be raised in the case of any natural protein, or indeed any material whose chemical constitution has not been completely determined. As there is, on the other hand, no evidence that a virus is not a single substance, we may speak of it, noncommittally, as a "substance or material," with the understanding that we are speaking of a specific kind of material with certain definite constant properties—chemical, biological, and crystallographic—by which it can always be identified and which survive unaltered such manipulations as isolation, crystallization, colloidal dispersion, and unlimited transfers in series. In any case, this question does not affect the propriety of speaking of a virus as an "autocatalyst," because many catalysts, of course, are not pure substances but "mixtures"—or, better, definite physicochemical combinations—of two or more substances. That a virus is specifically autocatalytic is unquestionable: each virus multiplies specifically (under proper conditions, of course) and is specifically different from other viruses; it differs chemically from the normal components of the medium in which it multiplies, and its multiplication must therefore involve chemical synthesis of the substance or substances of which it is composed, which does not occur unless the virus is initially present. The evidence for mutability of viruses is less abundant, but seems to be quite conclusive in some cases: a particular "strain" of virus, with certain specifically self-reproducing characteristics, gives rise to a new "strain" with different specifically self-reproducing characteristics. The viruses, or at least those in which such mutations have been demonstrated, are therefore mutably specific autocatalysts.

Although the viruses are the only mutably specific autocatalysts which have, as yet, been chemically "isolated," the evidence from genetics and cytochemistry demonstrates quite conclusively that all organisms contain such substances (or materials); indeed, in the higher organisms at least, there are thousands of specifically different ones in each cell—namely, the *genes.* The evidence that the genes are specifically autocatalytic is essentially similar to that in the case of the viruses, though somewhat less direct, in that the genes have not yet been chemically isolated. As they have not been isolated, there is no *direct* evidence that the different genes differ from each other *chemically,* though in view of the large number of specifically different genes any other interpretation seems highly improbable; but this is not essential to their being specific autocatalysts, as specifically different catalysts, in many cases, differ from each other "physically" rather than "chemically." The material ("chromatin") comprising, or containing, the genes differs chemically from the surrounding substances, and multiplies by synthesis of its own components *in situ.* The

individual genes (as identified by their phenotypic effects) are specifically self-reproducing; and the loss of a particular gene entails the permanent absence, not only of the specific phenotypic effects by which it is identified, but also of the particular bit of chromatin material at the point where this gene, when present, is located; that is, this particular gene is necessary for the reproduction of this particular bit of chromatin material, which may be considered to be the gene, although it may also contain additional material not constituting part of the gene: this is largely a question of definition. The evidence that each gene is specifically autocatalytic (under proper conditions, of course) appears, therefore, to be logically complete. The evidence for mutability of the genes (or many of them, at least) is, of course, abundant: the present concept of "mutation" is based primarily on mutations of individual genes, though other kinds of heritable variations may also be included in this term. As mutably specific autocatalysis is *the* essential property of the gene *qua* gene (of which its other distinctive properties may be "by-products"), we shall here call any substance or material having this property a "genic" substance or material.

Genic substances or materials, as thus defined, are present: in the *chromosomes* of all organisms having definite chromosomes; in *bacteria* and similar organisms which have no definitely distinguishable chromosomes, but contain "chromatin" material; in the *viruses* (including the bacteriophages); probably in *plastids,* or at least some of them; and possibly also in other cytoplasmic "formed bodies" or in the "general cytoplasm." What have all these materials in common, aside from the property of mutably specific autocatalysis? There is one feature of *chemical composition* common to all of them and especially characteristic of those materials (chromatin and viruses) which are most certainly genic—namely, the presence of *nucleoproteins.*

It has long been known that chromosomes are composed entirely, or almost entirely, of nucleoproteins, or of nucleoproteins and their components (nucleic acids and proteins). More recent advances in our knowledge of the finer structure and chemistry of the chromosomes have all served to identify the nucleoproteins still more closely with the specifically genic material.[5] The essential part of a chromosome is the *chromonema,* a very slender unbranched filament containing a single linear series of genes. Each chromosome consists, in different kinds of cells and at different stages, either of a single chromonema surrounded by an "envelope" of non-genic material, or of a bundle of two to many parallel chromonemata with envelope material surrounding the whole bundle and also the individual chromonemata, or sub-bundles of them, within it. Each chromonema is made up of a linear series of alternating segments of chromatic ("euchromatin") and achromatic material. Each of the euchromatin segments of each chromonema contains (or is) a single gene, which is specifically different from all (or most) of the other genes of the same chromonema (the genes in corresponding positions in the several chromonemata of one chromosome are usually identical, having been produced by the specific self-reproduction of a single gene). The intervening achromatic segments apparently consist of non-genic material, probably produced by the adjacent genes. The genic (chromatic) material is composed of nucleoproteins. The achromatic material is composed chiefly, if not entirely, of simple proteins of the "protamine" type, with a longitudinal fibrous structure, probably forming a continuous filament extending throughout the length of the chromonema, to which the individual genes are firmly attached, perhaps by some kind of chemical bonds (*e.g.,* between the acidic nucleoproteins and the basic protamines). The "envelope" material of the chromosomes is composed chiefly, or perhaps entirely, of basic simple proteins of the "histone" type. The large proportions of protamines and histones found, by gross chemical analysis, in condensed chromosomal material, such as sperm nuclei, may therefore have no significance in regard to the composition of the genes them-

selves, being merely "skeletal" and "envelope" materials of the chromosomes. The protein components of the nucleoproteins composing the genic material are, in part at least, "higher" proteins of the "globulin" type, and may perhaps also contain some other prosthetic groups besides the nuclei acid (*e.g.*, iron is generally found in, or closely associated with, nucleoproteins). Some parts of chromosomes, and some whole chromosomes, consist of a different kind of chromatin material, called "heterochromatin," which does not have the typical "banded" structure of alternate chromatic and achromatic segments, but is chromatic throughout. This heterochromatin is also composed of nucleoproteins, with a larger proportion of nucleic acid than the euchromatin regions. The heterochromatin regions are also genic (specifically self-reproducing and mutable), but less differentiated than the euchromatin regions, *i.e.*, containing fewer specifically different genes in proportion to the volume of chromatin.

Bacteria—which are, of course, specifically self-producing and mutable—differ from most other organisms in having no definitely distinguishable chromosomes, but they contain an unusually high proportion (as compared with most ordinary cells, taking the cell as a whole) of "chromatin," *i.e.*, material staining with stains which are more or less specific for nucleic acids. Gross chemical analysis reveals a correspondingly high proportion of nucleoproteins in the bacteria. Relatively little is yet known about the fine structure or composition of bacterial chromatin. Other simple organisms, such as the blue-green algae, which likewise lack definite chromosomes, also always contain chromatin (sometimes forming more or less chromosome-like bodies), but little is yet known of its fine structure or composition in these cases either.

All viruses which have been isolated and analyzed are nucleoproteins, the protein components being "higher" proteins of the "globulin" type, as is probably true also of chromosomal genes, or at least some of them, as previously indicated. If the viruses are "pure substances"—as suggested, though of course not proved, by their crystallization—they are the only pure genic substances which have yet been isolated. In any case, they are the "purest" genic materials yet available for direct chemical analysis. The chemical constitution of any virus is, of course, not yet fully determined, or even very nearly so, any more than is that of any other natural protein.

There is a good deal of apparently conclusive evidence that the plastids of plant cells, or at least some of them, are specifically self-reproducing and mutable. There is also good evidence that they contain nucleoproteins.

The question of the presence of genic materials in the cytoplasm (apart from plastids) is a very complex and quite unsettled one, involving both the problems of "cytoplasmic inheritance" and of "self-differentiation." Apart from the cases of plastid inheritance, the positive evidence for "cytoplasmic inheritance," properly so called, *i.e.*, the persistent hereditary transmission of specifically self-reproducing cytoplasmic components, is meager and generally inconclusive; on the other hand, the whole body of genetic data indicates that cytoplasmic inheritance must be, at best, exceptional. It is still possible to evade this conclusion by appealing from the demonstrable evidence to as yet unverifiable hypotheses; but, for the present at least, we must regard cytoplasmic inheritance as doubtful or exceptional. Should its occurrence, whether general or exceptional, be clearly demonstrated, it must imply the existence in the cytoplasm of some kind of "genic" substances or materials, in the sense in which we are here using the term. The problem of "self-differentiation," *i.e.*, the maintenance and continuation of the differentiation of ontogenetically differentiated cells, tissues, or organs, after their removal from the conditions which may have initiated this differentiation, is one of a different kind: here the facts are, in general, well established, but there are several proposed explanations of them. Some of these explanations postulate specifically self-reproducing cytoplasmic components—whether cytoplasmically inherited or, more probably, originally

produced by the action of the chromosomal genes—while others assume only continuing interaction of the chromosomal genes with differentiated (but not self-reproducing) cytoplasmic components which, originally produced by the action of these genes, condition their subsequent activities. Whatever may be the eventual answers to these questions, we do find nucleoproteins, in greatly varying amounts, in the cytoplasm of many, if not all, kinds of cells. It seems quite significant that the quantity of nucleoproteins in the cytoplasm is greatest in cells which are growing rapidly, or actively engaged in other synthetic processes such as the production of special secretory materials. The concentration of nucleoproteins in the cytoplasm, moreover, is greatest in the immediate vicinity of the nucleus, suggesting their ultimate derivation from the activities of the chromosomal genes.

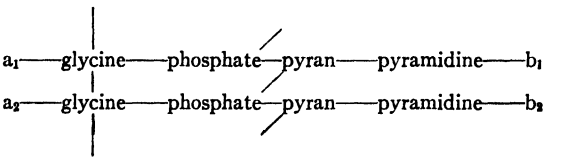
On the basis of all the present evidence, it thus appears very probable that all known "genic" substances or materials are, or contain, nucleoproteins. This does not, of course, imply that all nucleoproteins are "genic," or that the manifestation of "genic" properties by any nucleoprotein (or material containing it) is independent of surrounding substances and conditions.

The chemical constitution, *i.e.*, molecular structure, of nucleoproteins, like that of other proteins, is still very incompletely known. Nucleoproteins are compounds of simple proteins (polypeptides) with nucleic acids (polynucleotides), though it is not unlikely that some of them may also contain other prosthetic groups, *e.g.*, ferriferous groups. A considerable part of the protein is very loosely bound to the rest of the nucleoprotein, apparently in a salt-like combination, but it is doubtful that this loosely bound protein is really an integral or constant part of the nucleoprotein molecule. Another protein fraction is more firmly bound to the nucleic acid, probably by linkages of the peptide or ester types, *i.e.*, by combination of hydroxyl groups with one another or with amino or imino groups, with elimination of the constituents of water; such linkages might involve the free carboxyl or amino groups, or the internal imino groups, of the protein, and the phosphoric acid or nitrogenous basic groups (or possibly the sugar groups) of the nucleic acid.

Only two different nucleic acids (ribonucleic acid and thymonucleic acid) are yet known; but in view of the relatively few sources from which nucleic acids have been isolated and analyzed, there may well be others: within the general type of structure common to the nucleic acids there is the possibility of numerous variations of composition and constitution. On the whole, however, it is improbable that differences in the nucleic acids can account for more than a very small fraction of the differences between the thousands of specifically different genes in each organism (at least in the higher organisms) and the many millions which must exist in all organisms. On the other hand, the possibilities of differences of composition and constitution between the protein components are practically infinite (of the order of 10^{1000} or even more); and we know that there are specific differences between the corresponding proteins of even closely related organisms, as well as many different proteins in each organism. These considerations, in conjunction with the fact that apparently only *nucleo*proteins are "genic," suggest the obvious and temptingly simple hypothesis that the nucleic acid is responsible for the general "genic" property, while the protein components are responsible for the specific diversity among genes. But this may well be a gross over-simplification of the actual relations: it is quite probable that at least some part, perhaps some particular groups, of the protein molecule, as well as the nucleic acid, or some essential part of its molecule, and the form of combination between them, may all be equally essential for the "genic" property; variations in the nucleic acids and in the linkages between them and the proteins may also contribute to the specific diversity among genes.

One of the marked chemical properties of nucleic acid is its capacity for polymerization, which may play an important role in the structure of the nucleoprotein molecule. For instance, we might picture a long chain of polymerized nucleic acid

as forming the "backbone" of the nucleoprotein molecule, to which various amino-acids or polypeptides, and perhaps other groups, are attached as "side-chains." On the other hand, the "backbone" of the molecule might be a polypeptide chain, with the nucleic acid, and perhaps other prosthetic groups, attached to it; or it might be a chain of alternate polypeptide and polynucleotide segments; etc. Another possibility is a longitudinal axis composed of parallel polypeptide and polynucleotide chains, linked together all along their lengths to form a series of condensed heterocyclic rings, to which various side-chains are attached. A normal straight-chain polypeptide may be regarded as a series of glycine residues united by peptide linkages, with the various other amino acid residues (minus their carboxyl and α carbons) attached to the α carbons of the glycine as "side-chains." Similarly a polynucleotide is a series of mononucleotides united by ester (or possibly other) linkages: each mononucleotide may be regarded as made up, fundamentally, of a phosphate radical, a pyran (or perhaps a furan) ring, from which the pentose and hexose sugars are derivable, and a pyramidine ring, from which the purine as well as the pyramidine bases are derivable. If we assume (merely for the sake of a simple picture) that these polypeptide and polynucleotide chains are united through the glycine and phosphate groups, we might have a fundamentally repetitive chain molecule something like this:



with its main axis running longitudinally, *i.e.*, vertically in the plane of the page, and the side-chains (a_1, a_2, . . . , b_1, b_2, . . .) extending laterally, *i.e.*, horizontally in the plane of the page: the "a"s representing the various amino acid residues (minus their first two carbons), to which, in turn, other groups may be attached, *e.g.*, through secondary carboxyl or amino groups; the "b"s the various groups substituted in the pyramidine ring to form the several pyramidine and purine bases, to which, in turn, other groups, *e.g.*, amino acids, etc., may also be attached; other variations are possible by various substitutions in the pyran (or furan) ring to form different sugars, by variations in some of the linkages, etc. Of course this particular picture of the molecular structure is entirely speculative and is not intended to be suggested as necessarily the most probable one. It is intended, rather, merely to suggest a type of molecular structure with certain features appropriate to a "genic" substance capable of indefinitely continued evolution: namely, the possibility of unlimited increase in complexity, while retaining its essential "genic" property throughout, by continued repetition (by polymerization) of an essentially constant fundamental group (which may be assumed to be responsible for the essential "genic" property), and at the same time the possibility of unlimited specific differentiation, by changes (mutations) in the side-chains (or other substituent groups), consisting in replacements, rearrangements, additions, or subtractions of these groups or parts of them. This principle of repetition with differentiation is found at every "level" of biotic organization: the repetition of fundamentally similar but specifically differentiated genes (together with non-genic material, presumably produced by the genes) to form the chromosome; the repetition of fundamentally similar but specifically differentiated chromosomes (together with non-chromosomal material, presumably produced originally by the chromosomes) to form the cell; the repetition of fundamentally similar but specifically differentiated cells, together with non-cellular material produced by the cells, to form the multi-

cellular organism; and again in the formation of "colonies" and "societies" of multicellular organisms.

The longitudinal axis and the side-chains of this hypothetical "genic" molecule lie in the same plane, so that the molecule is essentially "flat," like those of benzene and many other organic substances, and its autocatalytic multiplication may be assumed to occur by the specific adsorption (and subsequent combination) of its constituent radicals at the "flat" surfaces of the molecule. This would tend to produce a filament of identical molecules (a sort of "linear crystal") extending in a direction perpendicular to the plane of the original molecule; transverse divisions of this filament, *i.e.*, divisions in planes parallel to that of the original molecule, whether autogenous or determined by some external force, would result in numerical increase of discrete particles, either multimolecular or unimolecular, of the "genic" substance. In particular cases, *e.g.*, in the chromosomes, one of the "flat" surfaces of the "genic" molecule may be attached (whether by chemical combination or by adsorption) to the surface of some other material, so that its autocatalytic growth can occur at only one surface; and there may be special conditions causing the separation of these molecules as fast as they are formed, so that the "genic" particles remain unimolecular; but it is not necessary to assume that this is generally true.

The longitudinal dimension of this hypothetical "genic" molecule may be called the "dimension of polymerization," its lateral dimension the "dimension of mutation," and its perpendicular dimension the "dimension of reproduction." Since the reproduction of chromosomal genes takes place in a direction perpendicular to the longitudinal axis of the chromonema, this axis must be in or parallel to the plane of the "genic" molecule, and it seems natural to suppose that it coincides with or parallels the longitudinal dimension of this molecule. From the viewpoint of those who would regard the whole chromonema as a chemically united continuum, in which the individual "genes" are merely locally differentiated regions, the whole chromonema would be a single polymeric "molecule," of the general type postulated above, with the individual "genes" corresponding to its "side-chains," or groups of them. In this view, the achromatic segments of the chromonema are not "non-genic," but integral parts of the genic "molecule," presumably its less differentiated regions; in which case the "backbone" of the molecule must presumably be a polypeptide chain, with the nucleic acids only in the "side-chains."

All the facts of mutation, crossing over, and translocation, however, appear to indicate a very considerable degree of independence for the individual genes. The only genetic evidence apparently contradicting this is that of "position effects" (modification of the effects of one gene by its neighboring genes in the same chromosome); these effects are, however, on the average, relatively slight as compared with those of intragenic mutation, and, though fairly frequent, may still be exceptional rather than general. They are probably comparable to the modification of a specific catalyst by alteration of the "carrier" without change in the specific catalytic group itself. The attachment of the individual genes to the "skeleton" of the chromonema may very well be of a chemical or "semi-chemical" nature: the "loosely bound" protein fraction of nucleoproteins probably represents the "skeletal" and "envelope" components of the chromosomes.

It has been remarked that some chemists consider "the whole ocean and everything in it a single molecule, and fishing a chemical reaction because it breaks up this molecule:" if we are to swallow the ocean, we need not strain at a chromonema. Without presuming, therefore, to define "a molecule," we shall, for purposes of the present discussion, consider "a gene," that is, a chromosomal gene, to be the nucleoprotein material of a single euchromatin segment of a single chromonema (excluding the simple protein "skeletal" and "envelope" material). If we consider this genic material to be made up of polymeric molecules of the general type postulated above, with the longitudinal dimension ("dimension of polymerization") of these molecules

parallel to the axis of the chromonema, and assume that the length of each euchromatin segment is that of a single such molecule, then each of these molecules would be made up of between 200 and 1000 of the "fundamental units" (glycine-mononeucleotide with "side-chains"); assuming an average of from one to three amino acid residues to each mononeucleotide, such a molecule would have a molecular weight of between 100,000 and 800,000 and a volume of between 60 and 500 $(m\mu)^3$, corresponding to a "diameter" of 5-10 $m\mu$ on the assumption that the molecule is spherical and compact. The molecular volume so computed is that of such a molecule in the crystalline state, if not hydrated or combined with inorganic ions, etc. In a largely aqueous system (such as the living cell), however, nucleoproteins and similar substances are heavily hydrated and combined with inorganic ions, forming colloidal particles or gels containing a very large proportion of water: nucleic acid, for instance, forms quite a firm gel containing 95 per cent of water, and muscle fibers, which are much more "solid" than chromosomes, contain about 75 per cent of water; the water content of chromosomes is probably somewhere between these limits. The apparent volumes of the "genic" molecules, therefore, when measured either as discrete particles in solution or as components of a gel, would be several times as large as those computed above and their "diameters" perhaps twice as great.

The earlier estimates[6] of the size of individual genes (in *Drosophila melanogaster*), computed by dividing the total volume of the sperm nucleus (which is simply a compact mass of condensed chromosomes, comprising one haploid set of chromosomes), about 0.25 μ^3, by the estimated minimum number of genes in one haploid genome, about 2000-3000, gave a maximum volume per gene of about 100,000 $(m\mu)^3$, or a "diameter" of 50-60 $m\mu$. These figures, being based on a *minimum* number of genes and the *maximum* volume occupied by them, *i.e.*, the total volume of the chromosomes, are absolute *maximum* values, and never pretended to give any indication of the most probable actual sizes of the genes, although they have been widely misinterpreted in this sense.

It is now evident, however, that the actual size of the individual genes must be much smaller than this. In the first place, it now appears that the number of genes in one haploid genome (of *Drosophila melanogaster*) is probably between 5,000 and 10,000, exclusive of the heterochromatin, in which the number of individual genes cannot be estimated; and there are at least two chromonemata in each chromosome in the sperm nucleus: so the number of euchromatin genes in the sperm nucleus is at least 10,000-20,000. Moreover, the aggregate volume of these genes—after subtracting from the volume of the chromosomes the volumes of the heterochromatin, of the "envelope" material, and of the achromatic "skeletons" of the chromonemata, which probably extend right *through* the euchromatic segments—can hardly be more than 10 per cent of the total volume of the chromosomes, and may be much less. If this genic material contains, like "protoplasm" in general, about 80 per cent of water, the volume of the organic substance of the genes is not more than 2 per cent of the total chromosome volume. This would give not over 5,000,000 $(m\mu)^3$ for the aggregate volume of the "dry" euchromatin genes in the sperm nucleus, or 250-500 $(m\mu)^3$ for each gene, which is within the range of the volumes calculated above for the hypothetical nucleoprotein molecule. Of course these calculations are full of assumptions, and therefore prove nothing; but they suggest that it is quite possible that each individual gene, in the sperm nucleus, *may* be a single molecule, without unduly stretching the usual meaning of the word. Certain genetic data have been interpreted as positive evidence that the individual genes are, in fact, single molecules, at least in the sperm cells of *Drosophila melanogaster;* on the other hand, other genetic evidence, from different organisms, has been interpreted as indicating a "compound," presumably multimolecular, structure of the genes. In neither case, however, does the evidence adduced appear to be conclusive, even for the particular cases directly concerned, as other interpretations of the genetic data are always

possible. Nor is it at all necessary to assume that all genes are alike in this respect, or even that any one gene is so at all times. It is quite conceivable, for instance, that a particular gene might be unimolecular at the time of mitotic division, but multimolecular during the intermitotic stage; or that individual genes might be unimolecular throughout the mitotic cycle in rapidly dividing cells, but become multimolecular in more slowly dividing cells or in mature somatic cells; etc. The only point which here concerns us is that no evidence at present available compels us to assume that "a gene" is *necessarily* a multimolecular structure.

There is one very general and characteristic biochemical phenomenon, which (so far as I know) has never been satisfactorily explained on any other basis, which might very plausibly be explained on the basis of the unimolecular composition of individual genes: *viz.*, the universal presence in living organisms and their products of specific optically asymmetric substances, in contrast to the racemic mixtures of such substances invariably found in materials of abiotic origin, whether natural or artificial. It is generally agreed that the synthesis of optically asymmetric substances in or by organisms is due to the specific catalytic activity of biocatalysts which are themselves optically asymmetric. But the production of these asymmetric biocataylsts must, in turn, be conditioned by the preexistence of other asymmetric substances, and so on. This course of reasoning, carried to its logical conclusion, points to optically asymmetric genes, or "genic" substances, as the original starting points of these series of asymmetric syntheses. But if the individual gene is, *at any stage of the life cycle,* or has been *at any time in its evolutionary history,* a single molecule, then it must be optically asymmetric (provided it contains any asymmetric carbon atoms at all), for a single molecule cannot be a racemic "mixture." This argument, for what it may be worth, does not require that existing genes be unimolecular at all stages, or even necessarily at any stage, of the cell cycle or the life cycle of presently existing organisms, but merely that they have been unimolecular at some time in their evolutionary history: in other words, that the "original" genes were single molecules. The "original" genic materials, however produced, may have been racemic mixtures, but if each individual genic molecule reproduces itself exactly (in respect to optical asymmetry), then each presently existing type of gene must (at least in the great majority of cases) be derived ultimately from a single unimolecular "ancestor," which is quite probable in view of the vast amount of "chance" elimination in the course of evolution. There may be some other equally plausible explanation of this highly distinctive biochemical characteristic, but if so I am unaware of it.

If, then, we assume that "a gene" *may* be a single molecule, *i.e.*, that a gene *qua* gene is not necessarily a multimolecular structure, the question naturally arises: how complex must such a "genic" molecule be? Or, conversely, how simple *may* it be? Of course we have as yet no direct evidence bearing on this question. On the assumptions made above, a single gene (in *Drosophila melanogaster*), if it is unimolecular, would have a molecular weight of between 100,000 and 800,000, and would consist of between (roughly) 10,000 and 100,000 atoms. These values, of course, represent the maximum sizes of the genes (in this particular organism) and the actual sizes of the gene molecules, or of many of them, may be very much smaller, for all we know. Even if the genes of this organism, and other existing organisms, should be single molecules of such huge size, this does not prove that all "genic" molecules need be so large, or anywhere near it. These genes are, of course, products of a long evolution, in the course of which their chemical, as well as physical, complexity may well have increased greatly above that of their "original" ancestors, just as the structural complexity of chromosomes, cells, and organisms has done. If the genic molecule is supposed to have an essentially polymeric structure, as suggested above, it may be supposed that each of the "fundamental units" making up this "chain molecule" should itself have "genic" properties, and that the simplest

possible gene need therefore be no larger than a single one of these units, *e.g.*, a single mononucleotide combined with one or two amino-acid residues. However, this does not necessarily follow, for the "genic" property might require not merely a certain combination of chemical constituents, but also certain physical characteristics dependent on the size of the molecule: for instance, the isomolecular adsorption process, which we have supposed to be an essential step in autocatalytic synthesis, might require a certain minimum molecular weight and surface area, which would thus determine the size of the "smallest possible" genic molecule, quite aside from the degree of strictly "chemical" complexity which may be essential to the "genic" property.

In existing organisms, the production of complex molecules, such as those presumably composing the genes (as well as other complex biochemical substances), undoubtedly involves a long series of synthetic processes, regulated by a system of specific biocatalysts, which are themselves, in turn, produced by complex interactions between a specific system of genes and a specific cytoplasmic milieu, this whole complex "reaction system" being a product of many millions of years of "adaptive" evolution. The specific autocatalytic synthesis of each individual gene is but the final step of such a complex synthetic process. The "original" genes, however, must be presumed to have been produced, in some way, without benefit of such a pre-existent "adaptive" reaction system.

From the purely mathematical standpoint of the theory of probability, the "spontaneous" formation of any structure, however complex, by the purely fortuitous concatenation of all its constituent elements, in the proper proportions and the proper mutual relations, is always "possible," given sufficient time and the co-existence of all these elements in sufficient concentrations. However, the probability of the "spontaneous" origin, at one step, of a molecule of 100,000 or even 10,000 atoms, of at least 5 different kinds, in certain definite proportions and specific stereochemical relations, from its elementary constituents or from inorganic or very simple organic compounds—while mathematically far greater than the probability of the "spontaneous generation," from the same substances, of a man or an amoeba or even a bacterium—is so remote as to be practically beyond the bounds of credibility. This difficulty would be greatly reduced, of course, if the "original" genes were much smaller molecules, of say 50 to 100 atoms each, comparable, for instance, to the single "fundamental units" postulated above, which might be formed by the combination of substances such as CO, HCN, NH_3, C_2H_2, C_2H_4, H_3PO_4, etc., perhaps through several intermediate stages. If such molecules were formed "spontaneously," *i.e.*, by abiotic and non-autocatalytic processes, in sufficient concentrations, they might also polymerize to form much larger molecules, since spontaneous polymerization is characteristic of many organic substances (as well as some inorganic ones.)

The problem, thus posed, of the origin of the "first" genes, is merely a more specific formulation, in terms of the gene theory, of the long recognized dilemma which confronts any naturalistic explanation of the "origin of life": which came first—the complex organic compounds (proteins, etc.) which constitute the essential chemical basis of "life" as we know it, or the complex physicochemical structure of the living organism which appears to be the essential basis of the "natural" synthesis of these compounds? Obviously, on any naturalistic explanation, neither could have come "first," in anything like their present complex forms: both the chemical components and the special structure of "living matter," as we know it, must have evolved *pari passu* from much simpler antecedents. Moreover, as remains of living things, apparently not essentially different from some of those still existing, are found as far back in geological time as it would be physically possible for any such remains to be preserved, it seems probable that these first—presumably molecular or ultramicroscopic—antecedents originated under the very different physical and chemical conditions of very early geological periods. In this case, it would seem that they must

have been quite different, chemically as well as physically, from the components and structures of all existing organisms; and that the chemical, as well as the structural, characteristics of living things, as they now exist, have evolved *pari passu* with the geological evolution of the earth.

Without speculating further on the origin of the "first" gene (or "genic substance"), or on its chemical composition or structure, let us take it as "given," and proceed to examine the *necessary* consequences of its essential property *qua* gene, *viz.*, the property of mutably specific autocatalysis, independently of any (other) special chemical or physical properties. We may postulate, to start with, a single "genic substance," even a single "molecule" of such a "substance," assuming nothing about its composition, structure, or properties except that it is, by definition, capable of mutably specific autocatalysis under certain (unknown) conditions, which we must also assume as "given."

We cannot properly speak of a substance as being first specifically autocatalytic but not mutable and later "becoming" mutable: for if it is capable of changing, at any subsequent time, into something different in any way but also specifically autocatalytic, then the original substance is, by definition, mutable. We may, however, for the sake of simplicity, examine first the relatively simple consequences of the property of specific autocatalysis, before considering the far greater complexities introduced by mutability. The "genic substance," no matter how little of it there may be originally—even a single "molecule"—will tend (under proper conditions, of course), by virtue of being autocatalytic, to increase in amount; its rate of increase, moreover, will be (at first) an ever accelerating one, for the more of the genic substance exists the more of it will be formed in any given period of time (other things being equal), *i.e.*, the rate of an autocatalytic process *tends* to be proportional to the amount of the autocatalyst present at any given time (assuming that all of the autocatalyst is "active," etc.).

This increase, however, can occur only at the expense of some other substance or substances, and these must be of certain special kinds: for instance, they must contain the particular chemical elements constituting the genic substance, and presumably they must contain these elements in certain particular combinations, which the genic substance can either incorporate directly into its own molecules, or decompose into fragments which it can incorporate. The total existing quantity of such substances (which we may call its "food" substances) will therefore determine an unsurpassable upper limit to the increase of the genic substance. This limit, of course, may be a variable one, as the quantity of the "food" substances may be increased or decreased by other natural processes, but this quantity must have some finite upper limit, which will determine the absolute limit of increase of the genic substance, remembering that we are assuming for the moment that the genic substance does not mutate.

But there may be other factors which restrict the increase of the genic substance even below this theoretically possible limit. For instance, the genic substance may not be able to increase, or even to exist, in all places where its "food" substances exist: it may be limited by various external conditions, such as temperature, physical state of the milieu, etc.; thus the upper limit of its increase would be determined not by the total existing quantity of its "food" substances, but by the quantity available in those environments in which it can exist and "grow." Its ability to utilize the available "food" substances may be further limited by factors such as the physical form of its growth; *e.g.*, whether it forms compact massive aggregates (like most crystals), two-dimensional laminae, linear filaments, or discrete molecules in solution. Any form which tends to limit the surface of contact with the environment or the facility of dispersal throughout the appropriate environment would thereby tend also to limit the utilization of the available "food" substances.

Another factor, consequent on the autocatalytic process itself, which may tend to inhibit this process, is the accumulation of "by-products": unless the autocatalytic

reaction is an extremely simple one (simple additive combination), it will form other products besides the autocatalyst itself, and such "by-products" usually tend to inhibit the reaction which produces them.

Besides all these factors tending to limit the increase of the autocatalyst, we must consider also those tending actually to decrease it. These will include not only external conditions which may act, in one way or another, to destroy some of the autocatalytic substance, but also, in all probability, a certain amount of spontaneous decomposition. All chemical reactions are, to a greater or less extent, reversible; and, roughly speaking, the more complex a substance is the more unstable it is likely to be. Therefore the genic substance, at the same time that it is forming more of itself by autocatalytic synthesis, will probably also be losing some of itself by decomposition, and its *net* rate of increase, at any given time, will be the *difference* between the rates of these two opposite processes. The rate of decomposition will tend to increase as the amount of the genic substance increases; this is also true, *at first*, of the rate of formation of this substance (because it is autocatalytic); but as the various limiting factors discussed above come into play to inhibit its further increase, the rate of decomposition will gradually overtake the rate of formation, until the two become equal, and the *net* rate of increase becomes zero, *i.e.*, the genic substance will have reached a quantitative *equilibrium* with its environment. This equilibrium may be shifted in either direction by changes in any of the environmental factors which affect either the rate of formation or the rate of decomposition, or both, of the autocatalyst. Thus this substance may increase or decrease *in quantity* from time to time; or, in the case of an extreme change in some essential environmental condition, it may even disappear entirely.

We have now to consider the consequences of the second, and more especially distinctive, aspect of the defining property of the gene—mutability. The meaning of this term has been discussed previously. As there emphasized, it does not imply that *every* change in a genic substance is of the special kind which we call mutation. It seems probable that *most* changes in its composition or structure would result in loss of its essential property of specific autocatalysis: such changes would merely constitute decomposition of the genic substance, which has already been considered. Perhaps other changes would not affect its autocatalytic activity at all: these would be of no significance, since they would not be perpetuated. The term "mutability" implies only that *some* changes, however infrequent, are of the special kind called "mutation": *i.e.*, change of the original genic substance into a different one, which is also *specifically* autocatalytic and specifically *different*, in respect to its autocatalytic reaction, from the original one. We need not here consider the causes or the intrinsic nature of such changes, but merely their consequences.

As the immediate result of a single mutation, we have two *different* genic substances, each tending to increase itself specifically. These two genic substances—the "old" one and the "new" one—may differ in their specific rates of formation or decomposition (or both) under the same conditions, in the kinds of "food" substances they require, in the kinds of environmental conditions most favorable to them, etc. Let us consider first the consequences of a mutation which does not alter the "food" or other environmental requirements of the genic substance. The effective differences between the "old" and the "new" genic substances then reduce, for our purposes, to differences in their specific rates of formation and decomposition. Obviously, the one having the greater *net* rate of increase, under the given conditions, will increase more rapidly than the other. Moreover, since the ultimate limiting factor—the amount of available "food" material—is (by hypothesis) common to the two genic substances, the total possible quantity of these two substances together will be the same as that of either of them if it alone were present. In this sense, either one can increase only at the expense of the other: thus these two genic substances are in *direct, unmitigated competition.* When both genic substances have reached quanti-

tative equilibrium with the environment, their relative amounts will be a simple mathematical function of their respective specific rates of formation and decomposition, regardless of which one was present first.

Thus if the original genic substance had already reached equilibrium before the mutation, the new one must, if it increases at all, partly replace the old one; and if the new genic substance happens to be the one with the greater net rate of increase it will replace the greater part of the old one. As other mutations of this kind occur, the genic substances with greater net rates of increase will increasingly replace, to a large extent, those with smaller net rates of increase. This does not necessarily mean that those with the smaller net rates of increase will be entirely eliminated, but as their equilibrium quantities become smaller they will run an ever greater risk of complete elimination by "accidental" factors—that is, by local and temporary environmental changes of an unfavorable or directly destructive nature. This probability of *chance* elimination, which is obviously greatest for those genic substances existing in smallest quantity, must be especially great for *new* genic substances, *i.e.*, those newly produced by mutation, which exist at first as single molecules. Therefore, once any of the genic substances have reached quantitative equilibrium with the environment (or nearly so), any new genic substance has little chance of ever becoming established at all unless it has an advantage over its competitors because of a greater net rate of increase.

Thus, regardless of what kinds of mutations may occur, or of the relative frequencies of different mutations, only those new genic substances which have this advantage will, in general, replace any of the old ones to any appreciable extent; and the greater this advantage, the greater the extent to which such new genic substances will replace old ones, especially those with the smallest net rates of increase. The general result of such replacement of genic substances with smaller net rates of increase by those with greater net rates of increase, augmented by the probability of complete elimination of those genic substances (whether new or old) with the smallest net rates of increase, must be a practically *continuous increase in the average net rate of increase* of the "population" of competing genic substances considered as a whole. This is, in its simplest form, the principle of *natural selection.* It should be especially noted that this principle involves no new assumption of any kind; it follows directly and necessarily from the property of mutably specific autocatalysis, together with the limitation of the total "population" by environmental factors (especially the available "food" supply). What we have been calling "rate of formation" and "rate of decomposition" are usually called, in speaking of organisms, "reproduction rate" and "death rate" respectively; and the difference between these two rates, which we have called "the net rate of increase" is essentially what is meant by such terms as "survival value" or "fitness." If a population is in complete equilibrium, there is no actual net increase (or decrease) of any of its component types, the relative "fitness" of which is then manifested solely by their relative numbers; but such complete equilibrium can scarcely ever exist for any considerable time in any actual population of organisms, for (even in the absence of any general environmental changes) the occurrence of new mutations and the chance elimination of the rarest types must cause continual shifts, in the general direction of replacement of less "fit" types by more "fit" types, resulting in a practically *continuous increase in average fitness* of the population as a whole. This is (roughly stated) the principle of natural selection, in terms more commonly used by biologists; it differs from our previous statement, as will be seen, only in the replacement of "the net rate of increase" by the term "fitness," which means essentially the same thing.

Returning to our hypothetical primitive "genic substances," we have next to consider the consequences of mutations which alter not merely the rates of formation and decomposition of the genic substance but its relations to its environment, espe-

cially with respect to those environmental factors which chiefly limit its possible increase. Such would be, for instance, a mutation which would enable the new genic substance to utilize in its autocatalytic synthesis "food" substances which the original genic substance had not been able to use, or to exist and multiply under some environmental conditions where the original one could not do so. It is obvious that an immediate consequence of such a mutation would be the mitigation of competition between the new genic substance and the old one: the more they differ in these respects—that is, the less their "food" and other environmental requirements coincide—the less they will compete with each other. This reduction of competition will not only enhance the chances of the new genic substance to establish and multiply itself, but will also raise the limit to which the two genic substances (the old and the new one) together can ultimately increase, and will thus *increase the average "fitness"* (ability to increase) of the "population" as a whole. We should therefore expect that, as successive mutations occur, those new genic substances will be most likely to survive and multiply which differ most, in these respects, from other existing genic substances: and there should thus be a general tendency for the whole "population" (of all genic substances) to become ever more diversified in respect to their environmental relations.

The limit of the decrease of direct competition between different genic substances would be, of course, its complete elimination: as, for instance, if two such substances could exist only in entirely different, mutually exclusive, environments, or had completely different "food" requirements. But even complete elimination of competition is by no means the limit of the advantageous effects of diversification, for different genic substances may also have positively beneficial effects on each other. The most obvious relation of this kind is when one genic substance can utilize as "food" substances some of the "by-products" of another genic substance; this relation will usually be mutually beneficial, as the accumulation of its own by-products would presumably usually be deleterious to the genic substance which produces them. But, beyond this, the relation may be a reciprocal one, *i.e.,* each of two different genic substances may be able to use some of the by-products of the other; or it may be a cyclical relation involving three or more different genic substances, *e.g.,* by-products of substance A may be used as "food" by substance B, whose own by-products are used by substance C, which in turn forms by-products useful to A; or the cycle may be still more complex.*

Reciprocal and cyclical food relations of this kind are, in fact, not exceptional but quite universal among actually existing organisms. We need not here discuss further the varied and complex ecological relations which are possible—and actually exist—among different kinds of organisms; our present point is simply that all these relations, which tend to increase the "fitness" of the population (of all organisms) as a whole, are made possible by *diversification* within this population, that is, between the various kinds of organisms. A second result, therefore, of the principle of natural selection is a *continuous increase of diversification* within the population, as this contributes greatly to the increase in "fitness" of the population as a whole.

We must now consider *how* this progressive diversification can be effected, and this necessitates, in the first place, some consideration of the intrinsic nature of the process of mutation. If, as we have assumed, the original "genic" material is a single "substance," *i.e.,* a chemical compound, any change in it must be, primarily, a chemical change, that is, a change in composition or structure of one (or more) of its molecules. At the same time, any *mutation* must be a change of such a kind as to preserve that special feature (or those special features) of its composition or structure upon which the special property of specific autocatalysis depends. Now, all the pos-

* A well-known case is found in Lichens in which a fungus and an alga live in symbiosis. J. A.

sible kinds of changes which a molecule may undergo are of four general types: (1) *loss* of some portion of the molecule (decomposition); (2) *addition* of something—more atoms, radicals, or other molecules—to the molecule (combination, of which polymerization is a special case); (3) *replacement* of some portion of the molecule by something else (substitution); (4) *rearrangement* of parts within the molecule (isomeric change). If the original genic molecule is as simple as possible, consistent with its essential property of mutable autocatalysis, the possibilities of mutation by loss alone must be extremely limited. The possibilities of mutation by replacement (substitution) or internal rearrangement might be somewhat more extensive, depending on just how complex the original genic molecule is, and just how much its composition or structure can be altered without destroying its essential "genic" property.

It would seem, however, that the possibilities of mutation by addition should be greatest of all, since the addition of something extra would seem, on the whole, less likely to destroy the essential peculiar feature (or features) of the original composition or structure than any of the other types of change. This should be especially true of polymerization, which (as previously suggested) may be of especial importance in the evolution of complex genic molecules. Every mutation by addition (including polymerization) would increase the possibilities of further mutations by replacement (substitution) or internal rearrangement—that is, the more complex the structure of the genic molecule becomes, the greater become the possibilities of diversification among genic substances by subsequent mutations. To put this in another way: if the first genic substances were the simplest possible ones, then the *continuous increase of diversification* among genic substances must involve also an *increase in average complexity* of their constitution, though this does not necessarily mean that the simpler ones will be entirely replaced by more complex ones. We know that living organisms contain extremely complex chemical compounds having many thousands of atoms to the molecule; and the genes, or some of them, *may* be among the most complex of these organic compounds. This need not mean, however, that the "original" genes, or even all now existing genes, must be so complex; for, as we have just seen, we should expect the progressive diversification of the genes, resulting from natural selection, to involve also an increase in complexity of many (not necessarily all) of them. For this very reason, therefore, the extreme complexity of chemical constitution characteristic of many components of existing organisms may itself be a result of a long-continued process of evolution (by mutation and natural selection) from originally much simpler substances.

The possibilities of continued diversification are not limited to the complexity, however great this may be, attainable within the structure of a single molecule. Since even the largest organic molecules are still far below microscopic size, there would seem to be some factor (or factors) limiting their possible size—whether it be inevitable chemical instability beyond a certain size, some feature of the mechanism of their synthesis, or something else. Whatever the nature of this limitation may be, it is evident that the possibilities of greater complexity, and therefore of greater diversification, would be enormously enhanced if the evolution of the gene should overstep the boundary of molecular dimensions, and advance from a molecular to a supra-molecular structure. One way in which this might occur is by the permanent association of several or many genic molecules in a definite multimolecular aggregate: such aggregates may not only attain greater size and structural complexity than single molecules, but have also the possibility of much greater diversification, by independent mutations of the several genic molecules comprising the aggregate, resulting in an internally differentiated aggregate of genic molecules having special chemical and spatial relations to one another; the potential complexity and diversification of such genic aggregates would seem to be almost unlimited. The individual

genes (in the conventional sense) of the higher organisms, or some of them, *may* be such multimolecular aggregates of genic molecules; and the chromosomes are essentially genic aggregates (though probably containing also some non-genic substances): thus there may be multimolecular genic aggregates of several successive "orders."

Another way in which supra-molecular complexity may be attained is by the permanent association of a gene with non-genic substances, *i.e.*, substances which are not themselves autocatalytic but are produced, directly or indirectly, by the activity of the gene. These non-genic substances would presumably be, in the first instance, by-products of the autocatalytic action of the gene. In most cases by-products have an inhibiting effect on the reaction which produces them, but in some special cases the opposite is true. Thus it might happen that some by-product of a particular gene would react with some substance in the environment in such a way as to convert a substance which was not directly utilizable (as "food") by the gene into something that is utilizable, or to "protect" the gene from some deleterious substance or condition, etc. It is evident that any gene which happened to form such "useful" by-products would thereby acquire an "advantage" over other genes—by being able to use a wider variety of "food" substances, or to occupy a greater range of environments, or to survive more unfavorable vicissitudes, etc.—and would thus have a better chance to survive and multiply. We might expect, therefore, in the course of time, by successive mutations and the operation of natural selection, a continuous increase in such "useful" by-products and in their diversity and complexity. In the case of an internally differentiated aggregate of genes, the possibilities of such "useful" by-product systems are obviously still more diverse and complex—by reactions of one gene with by-products of another, reactions between by-products of different genes, between these by-products or their secondary products and environmental substances, etc.

By a combination of these processes—the aggregation and differentiation of genes, and the elaboration of a complex envelope of non-genic substances produced (directly or indirectly) by the activity of the genes—the complex structure of the *cell,* typical of most existing organisms, may have been gradually evolved. *Pari passu* with the structure of the cell must evolve the process of *cell division,* by which the cell reproduces itself *as a whole,* maintaining the quantitative and spatial relationships of its various genes and non-genic components. Some of the stages in this evolution of the cell and the process of cell division seem to be represented by the simplest existing microscopic organisms—the bacteria and the blue-green algae. Along with increasing complexity of chemical composition, structure, and reproductive processes must also go increasing complexity and diversity of metabolism, physical activities, and responses, all these being, primarily, means of enabling the organism to utilize different kinds of food substances and inhabit different kinds of environments, thus mitigating the severity of competition and thereby enhancing the increase of the "population" as a whole.

Just as the individual genes, *i.e.*, the molecules or unit particles of genic substances or materials, are limited to ultramicroscopic dimensions, the individual cells are limited (with some exceptions) to microscopic dimensions. In this case, the principal limiting factor seems to be the quantitative relation between the volume and the surface area of the cell. But just as the limitations of size, and therefore of complexity and diversity, of the individual gene have been overcome by the evolution of the polygenic cell, so the limitations of the individual cell are, in turn, surmounted by the evolution of the *multicellular* organism. This evolution is a repetition on a higher level (so to speak) of the same kinds of processes as those involved in the evolution of the cell from the single gene: the aggregation of cells in definite spatial relations, differentiation (structural and functional) among these cells, and the elaboration

of non-cellular substances about and between the cells. This further increase in structural complexity is again accompanied by increasing complexity and diversity of the processes of nutrition, growth, and reproduction, and of the activities and responses of the organisms—all of which, again, enhance the possibilities of their "making a living" in different ways and under diverse conditions. This same kind of evolutionary process is repeated again, to some extent, on a still higher level, in the evolution of "colonies" and "societies" among some kinds of multicellular organisms.

The argument of the preceding paragraphs may be summarized, briefly, as follows. Given a substance or body with the property of *mutably specific autocatalysis,* the limitation to its increase necessarily imposed by a finite environment results in the operation of the principle of *natural selection* upon its various mutant "descendants." From this follow: (1) continuous increase in average "fitness"—ability to survive and multiply—of the "population" (of all these descendants) as a whole; (2) continuous increase of diversification within this population, because this contributes greatly to the increase in "fitness" of the population as a whole; (3) increase in average complexity of the members of this population, because this makes possible continued increase of diversity among them. These results of natural selection—"fitness" (adaptation), diversity, and complexity—are, together with hereditary specificity (which is a direct consequence of specific autocatalysis), the most characteristic general features of living organisms as we know them.

At just what stage of complexity we should recognize such substances or bodies as "living" would seem to be a rather arbitrary matter. If we do not choose to recognize the simplest possible bodies (perhaps single molecules) which are mutably autocatalytic as "living," then we may say that "life" is a result of an evolutionary process identical, in essentials, with that by which the higher organisms have been derived from the lower, and by which populations of existing organisms are continuously modified at present. If we suppose that the property of mutable autocatalysis pre-requires a fairly high degree of complexity, at least in chemical constitution, then the first mutable autocatalyst may itself have been the product of a long process of "evolution"—in the general sense of gradual natural change—by which it had been built up little by little from simpler substances. This process, however, whatever its nature, must be something essentially different from what we call "organic evolution," the essential factors of which are mutation and natural selection. It is at the moment when the property of mutability first appears that the principle of natural selection begins to operate, and the nature of the "evolutionary" process changes radically to that characteristic of living organisms—the process of *organic evolution.* It might, then, appear most logical, at least in the light of our present knowledge, to draw the line between the "lifeless" and the "living" at this point—to regard as "living" any substance or body which is specifically autocatalytic and mutable. By this criterion, "life" would be coextensive with the process of organic evolution, of which all its complex and diverse manifestations are products—we might almost say, incidental by-products. And the primary physico-chemical basis of organic evolution—and hence, by this definition, of "life" also—is the phenomenon of *mutably specific autocatalysis.*

Notes

1. Recent advances in a number of these fields are reviewed in other chapters of this volume.
2. These are discussed by J. Alexander elsewhere in this volume.
3. This comparison and its possible bearing on biological problems were discussed many years ago by Troland (1917).
4. A more definite picture of a suggested autocatalytic mechanism of this sort was presented by Alexander and Bridges (1928).
5. The very brief summary of the fine structure and chemistry of the chromosomes which follows is based chiefly on the work of Bridges (1935) and the recent review by Schultz (1941).
6. Morgan (1922); Muller (1929); Plunkett (1930).

References

Alexander, J., and Bridges, C. B., "Some Physico-Chemical Aspects of Life, Mutation, and Evolution." In: J. Alexander, "Colloid Chemistry," Vol. 2, New York, (Reinhold Pub. Corp.), 1928.
Bridges, C. B., "Salivary Chromosome Maps": *J. Heredity*, **26**, 60-64 (1935).
Morgan, T. H., "On the Mechanism of Heredity": *Proc. Roy. Soc. London B*, **94**, 162-197 (1922).
Muller, H. J., "The Gene as the Basis of Life": *Proc. Int. Congr. of Plant Sciences*, **1**, 897-921 (1929).
Plunkett, C. R., "Outlines of Modern Biology," pp. 512-517, 654-667, New York, 1930.
Schultz, J., "The Evidence of the Nucleoprotein Nature of the Gene": *Cold Spring Harbor Symposia on Quantitative Biology*, **9**, 55-65 (1941).
Troland, L. T., "Biological Enigmas and the Theory of Enzyme Actions": *Am. Nat.*, **51**, 321-350 (1917).

Appendix

TABLE OF NUCLEAR PROPERTIES AS OF JAN. 1, 1943 *
From "Introduction to the Atomic Nucleus,"
by Robley D. Evans, Mass. Inst. of Tech.

Nuclei which have both stable and metastable states are listed on a single line. In such cases only the half-life, type and energy of the radiations are associated with the metastable (isomeric) state. The mass, abundance, and nuclear moments refer to the stable state.

Two lines are devoted to each nucleus which has two radioactive states (ground state and isomeric state).

Symbols used: Radioactive transactions are denoted as follows: $\alpha = \alpha$ ray; $\beta^- =$ negatron continuous β ray spectrum; $\beta^+ =$ positron β ray spectrum; $e^- =$ internal conversion electrons; $\gamma =$ nuclear gamma rays or isomeric transition; K = orbital electron capture.

Energies are given in MEV. For continuous β ray spectra the extrapolated upper limit of energy is given.

Nuclear moments. I = total nuclear angular momentum in $h/2\pi$ units; $\mu =$ magnetic dipole moment in nuclear magnetons, $eh/4\pi Mc$; q = electric quadrupole moment in $10^{-24}cm^2$ times the electronic charge. All masses are for the neutral atoms, on the physical scale $O^{16} = 16$.

* Just to show what results may be expected to emerge from the data here tabulated, the age of the potassium atoms in a sample of K obtained from the Pultusk meteorite has been determined by the use of the long-lived beta emitter (K_{40}) and found to be the same as telluric K, thus indicating that this meteorite is of the same age as the solar system (W. C. Schumb, R. D. Evans and W. U. Leaders, J. Am. Chem. Soc., **63**, 1203 (1941)). And the percentage of K in a sample may be determined from its radiation as registered by a Geiger counter (R. B. Barnes and R. J. Salley, Ind. Eng. Chem., Anal. Ed., **15**, 4-7 (1943)). Editor.

Element	Mass	Relative Abundance	Half Life	Radiation	Particles MEV	γ-Rays (MEV)	I	μ	q×10⁻²⁴
$_0n^1$	1.00893						1/2	−1.935	
$_1H^1$	1.008130	99.98					1/2	+2.7896	
$_1H^2$	2.014722	0.02					2/2	+0.855	+1.193
$_1H^3$	3.01705		~30y	β^-	0.01				
$_2He^3$	3.01711	$\sim 10^{-7}$							
$_2He^4$	4.00386	100					0		
$_2He^6$			0.8s	β^-	3.7				
$_3Li^6$	6.01684	7.9					2/2	+0.820	
$_3Li^7$	7.01818	92.1					3/2	+3.253	
$_3Li^8$	8.0251		0.88s	β^-, α	12(β)				
$_4Be^7$	7.01908		53d	K, γ		0.485			
$_4Be^8$	8.00765								
$_4Be^9$	9.01494	100					3/2	−1.176	
$_4Be^{10}$	10.01671		$\gg 10^3$y	β^-, γ	~0.5	<0.5			
$_5B^{10}$	10.01633	18.4					2/2	+0.598	
$_5B^{11}$	11.01295	81.6					3/2	+2.686	
$_5B^{12}$	12.019		0.022s	β^-	12				
$_6C^{10}$			8.8s		3.4				
$_6C^{11}$	11.01499		21.0m	β^+	0.98				
$_6C^{12}$	12.003861	98.9							
$_6C^{13}$	13.00766	1.1					1/2	+0.701	
$_6C^{14}$	14.00780		10^3–10^5y	β^-	0.090				
$_7N^{13}$	13.01005		9.93m	β^+, γ	0.92, 1.20	0.28			
$_7N^{14}$	14.00756	99.62					2/2	+0.402	
$_7N^{15}$	15.00495	0.38					1/2	+0.280	
$_7N^{16}$	16.011		8s	β^-	6.0?				
$_8O^{15}$	15.0078		126s	β^+	1.7				
$_8O^{16}$	16.00000	99.76					0		
$_8O^{17}$	17.00449	0.04							
$_8O^{18}$	18.00369	0.20							
$_8O^{19}$			31s	β^-					
$_9F^{17}$	17.0076		70s	β^+	2.1				
$_9F^{18}$	18.0056		112m	β^+	0.7				
$_9F^{19}$	19.00452	100					1/2	+2.622	
$_9F^{20}$	20.0063		12s	β^-, γ	5.0	2.2			
$_{10}Ne^{19}$			20.3s	β^+	2.20				
$_{10}Ne^{20}$	19.99896	90.00							
$_{10}Ne^{21}$	20.99968	0.27							
$_{10}Ne^{22}$	21.99864	9.73							

$_{10}Ne^{23}$	23.0005		40s	β^-	4.1			
$_{11}Na^{21}$			23s					
$_{11}Na^{22}$	22.0002		3.0y	β^+	0.58	1.3		
$_{11}Na^{23}$	22.99680	100					3/2	+2.216
$_{11}Na^{24}$	23.9974		14.8h	β^-, γ	1.4	1.38, 2.73		
$_{12}Mg^{23}$			11.6s	β^+	2.82			
$_{12}Mg^{24}$	23.99189	77.4						
$_{12}Mg^{25}$	24.99277	11.5						
$_{12}Mg^{26}$	25.99062	11.1						
$_{12}Mg^{27}$	26.9921		10.2m	β^-, γ	1.8	1.02; 0.84; 0.64?		
$_{13}Al^{26}$	25.9929		7.0s	β^+	2.99			
$_{13}Al^{27}$	26.9916	100					5/2	+3.628
$_{13}Al^{28}$	27.9903		2.4m	β^-, γ	3.3	2.3 or 1.82		
$_{13}Al^{29}$	28.9904		6.7m	β^-	2.5			
$_{14}Si^{27}$	26.9931		4.92s	β^+	3.74 or 3.5			
$_{14}Si^{28}$	27.9866	89.6						
$_{14}Si^{29}$	28.9864	6.2						
$_{14}Si^{30}$	29.9832	4.2						
$_{14}Si^{31}$	30.9862		170m	β^-	1.8	No γ		
$_{15}P^{29}$			4.6s	β^+	3.6			
$_{15}P^{30}$	29.9882		2.55m	β^+	3.0			
$_{15}P^{31}$	30.9839	100						
$_{15}P^{32}$	31.9841		14.30d	β^-	1.69	No γ		
$_{16}S^{31}$			3.18s	β^+	3.9			
$_{16}S^{32}$	31.9823	95.1					0	
$_{16}S^{33}$	32.9818	0.74						
$_{16}S^{34}$	33.978	4.2						
$_{16}S^{35}$			88d	β^-	0.120			
$_{16}S^{36}$		0.016						
$_{17}Cl^{33}$			2.8s	β^+	4.13			
$_{17}Cl^{34}$	33.981		33m	β^+	2.5			
$_{17}Cl^{35}$	34.97903	75.4					5/2	+1.368
$_{17}Cl^{36}$			71y	β^+, K, β^-	0.7(β^-)			
$_{17}Cl^{37}$	36.97786	24.6					5/2	+1.136
$_{17}Cl^{38}$	37.981		37m	β^-, γ	1.1, 5.0	1.65, 2.15		
$_{18}A^{35}$			1.91s	β^+	4.38			
$_{18}A^{36}$	35.9785	0.307						
$_{18}A^{37}$			34d					
$_{18}A^{38}$	37.9751	0.061						
$_{18}A^{39}$			4m	β^-				
$_{18}A^{40}$	39.97564	99.632						
$_{18}A^{41}$	40.9770		110m	β^-, γ	1.5	1.37		

Element	Mass	Relative Abundance	Half Life	Radiation	Particles MEV	γ-Rays (MEV)	I	μ	q×10⁻²⁴
$_{19}K^{38}$			7.7m	β^+, γ	2.3				
$_{19}K^{39}$	38.97518	93.4					3/2	+0.391	
$_{19}K^{40}$	39.975	0.011	1.39×10^{10}y		0.7 to 1.3		8/2	−1.290	
$_{19}K^{41}$	40.9739	6.6					3/2	+0.215	
$_{19}K^{42}$			12.4h	β^-	3.5	No γ			
$_{19}K^{43,\ 44}$			18m	β^-					
$_{20}Ca^{39}$			4.5m	β^+					
$_{20}Ca^{40}$	39.9745	96.97							
$_{20}Ca^{41}$			8.5d	K, γ, e^-		1.1			
$_{20}Ca^{42}$	41.9711	0.64							
$_{20}Ca^{43}$	42.9723	0.145							
$_{20}Ca^{44}$		2.06							
$_{20}Ca^{45}$			180d	β^-, γ	0.2	0.7			
$_{20}Ca^{46}$		0.0033							
$_{20}Ca^{48}$		0.185							
$_{20}Ca^{49}$			2.5h	β^-, γ	2.3	0.8			
$_{20}Ca^{49}$			30m						
$_{21}Sc^{42}$			13.5d	β^+	1.4 -				
$_{21}Sc^{43}$			4h	β^+	0.4, 1.4	1.0			
$_{21}Sc^{44}$			52h	e^-, γ		0.268			
$_{21}Sc^{44}$			4.1h	β^+	1.45				
$_{21}Sc^{45}$	44.9701	100					7/2	+4.6	
$_{21}Sc^{46}$			85d	β^-, γ, K	0.26	1.25			
$_{21}Sc^{47}$			63h	β^-, γ	1.1				
$_{21}Sc^{48}$			44h	β^-, γ	0.640	1.35			
$_{21}Sc^{49}$			57m	β^-	1.8	No γ			
$_{22}Ti^{45}$			3h	β^+	1.2				
$_{22}Ti^{46}$	45.9678	7.94							
$_{22}Ti^{47}$		7.75							
$_{22}Ti^{48}$	47.9651	73.45							
$_{22}Ti^{49}$	48.9664	5.52							
$_{22}Ti^{50}$	49.963	5.34							
$_{22}Ti^{51}$			2.9m	β^-, γ	0.36	1.0			
$_{22}Ti^{51}$			72d	β^-, γ					
$_{23}V^{47\ or\ 49}$			600d	K	no β^+ or e^-	No γ			
$_{23}V^{48}$			16d	β^+, K, γ	1.0	1.05			
$_{23}V^{49\ or\ 47}$			33m	β^+	1.9				
$_{23}V^{50}$			3.7h	β^+					
$_{23}V^{51}$	50.9587	100							

$_{23}V^{52}$			3.9m	β^-	2.05			
$_{24}Cr^{49}$			41.9m	β^+	1.45	0.2, 1.6		
$_{24}Cr^{50}$		4.49						
$_{24}Cr^{51}$			26.5d	K, γ, e^-		0.5, 1		
$_{24}Cr^{52}$	51.9582	83.78						
$_{24}Cr^{53}$	52.9572	9.43						
$_{24}Cr^{54}$	53.960	2.30						
$_{24}Cr^{55}$			1.6–2.3h					
$_{25}Mn^{51}$			46m	β^+	2.0			
$_{25}Mn^{52}$			21m	β^+, γ	2.2	1.2		
$_{25}Mn^{52}$			6.5d	β^+, γ, K	0.77	1.0		
$_{25}Mn^{54}$			310d	K, γ		0.845		
$_{25}Mn^{55}$	54.965	100					5/2	+3
$_{25}Mn^{56}$			2.59h	β^-, γ	.7, 1.09, 2.88	.846, 1.81, 2.13		
$_{26}Fe^{53}$			8.9m	β^+				
$_{26}Fe^{54}$	53.960	6.04						
$_{26}Fe^{55}$			~4y	K, e^-		.08		
$_{26}Fe^{56}$	55.9572	91.54						
$_{26}Fe^{57}$	56.9609	2.11						
$_{26}Fe^{58}$		0.28						
$_{26}Fe^{59}$			47d	β^-, γ	0.257, 0.460	1.30, 1.10		
$_{27}Co^{55}$			18.2h	β^+, γ	1.50	0.16, 0.31, 0.8 1.2		
$_{27}Co^{56}$			80d	β^+, γ, K	1.4	0.846, 1.27, 2.6 3.3		
$_{27}Co^{57}$			210d	K, γ, e^-, β^+	0.26	0.119, 0.131		
$_{27}Co^{58}$			72d	β^+, K, γ	<0.5	.813		
$_{27}Co^{59}$		100					7/2	(+2)–(+3)
$_{27}Co^{60}$			5.3y	β^-, γ	0.31	1.10, 1.30		
$_{27}Co^{60}$			10.7m	e^-, β^-, γ	>1	.06		
$_{28}Ni^{57}$			36h	β^+	0.67			
$_{28}Ni^{58}$	57.959	67.4						
$_{28}Ni^{60}$		26.7						
$_{28}Ni^{61}$		1.2						
$_{28}Ni^{62}$		3.8						
$_{28}Ni^{63}$			2.6h	β^-, γ	1.9	0.280, 0.65, 0.93		
$_{28}Ni^{64}$	63.955	0.88						
$_{29}Cu^{58,\ 60}$			81s	β^+				
$_{29}Cu^{58,\ 60}$			7.9m	β^+				
$_{29}Cu^{61}$			3.4h	β^+, K	0.9	No γ		
$_{29}Cu^{62}$			10.5m	β^+	2.6			

Element	Mass	Relative Abundance	Half Life	Radiation	Particles MEV	γ-Rays (MEV)	I	μ	q×10⁻²⁴
$_{29}Cu^{63}$	62.956	68					3/2	+2.5	−0.1
$_{29}Cu^{64}$	63.957		12.8h	β^-, β^+, K	0.58β^-, 0.66β^+	No γ			
$_{29}Cu^{65}$	64.955	32					3/2	+2.6	−0.1
$_{29}Cu^{66}$			5m	β^-	2.9				
$_{30}Zn^{63}$			38m	β^+	2.32				
$_{30}Zn^{64}$	63.956	50.9							
$_{30}Zn^{65}$			250d	β^+, K, γ, e^-	0.4β^+	1.14			
$_{30}Zn^{66}$	65.952	27.3							
$_{30}Zn^{67}$		3.9					3/2	−1.7	
$_{30}Zn^{68}$	67.956	17.4							
$_{30}Zn^{69}$			13.8h	γ, e^-		0.440			
$_{30}Zn^{69}$			57m	β^-	1.0	No γ			
$_{30}Zn^{70}$	69.954	0.5							
$_{31}Ga^{64}$			48m	β^+					
$_{31}Ga^{65}$			15m	K, e^-		0.054, 0.117			
$_{31}Ga^{66}$			9.4h	β^+	3.1				
$_{31}Ga^{67}$			83h	K, γ, e^-		0.0925, 0.18, 0.30			
$_{31}Ga^{68}$			68m	β^+	1.9				
$_{31}Ga^{69}$	68.955	61.2					3/2	+2.11	+0.20
$_{31}Ga^{70}$			18.5m	β^-, γ, e^-	1.7	0.054, 0.117			
$_{31}Ga^{71}$	70.953	38.8					3/2	+2.69	+0.13
$_{31}Ga^{72}$			14h	β^-, γ	2.6	1.0			
$_{32}Ge^{70}$		21.2							
$_{32}Ge^{71}$			40h	β^+	1.2				
$_{32}Ge^{72}$			11d	e^-, K	0.6	No γ			
$_{32}Ge^{67, 69, 71}$			195d						
$_{32}Ge^{72}$		27.3							
$_{32}Ge^{73}$		7.9							
$_{32}Ge^{74}$		37.1							
$_{32}Ge^{75}$			89m	β^-	1.2				
$_{32}Ge^{76}$		6.5							
$_{32}Ge^{77}$			12h	β^-	1.9				
$_{33}As^{71}$			50h	β^+, γ					
$_{33}As^{71}$			88m	β^+, γ					
$_{33}As^{72}$			26h	β^+					
$_{33}As^{74}$			16d	β^-, β^+, γ	1.3, 0.9	0.582			
$_{33}As^{75}$	74.934	100					3/2	+1.5	+0.3

$_{33}As^{76}$			26.8h	β^-, γ, β^+, K	1.1, 1.7, 2.7β^-, 0.7, 2.6β^+	3.2, 2.2, 1.5			
$_{33}As$			90d	e^-, γ, K		0.052			
$_{33}As^{78}$			65m	β^-, γ	1.4	0.27			
$_{34}Se^{74}$		0.9							
$_{34}Se^{75}$			48d	K, γ, e^-		0.50			
$_{34}Se^{75\text{ or }76}$			160d			8 γ rays <0.3			
$_{34}Se^{76}$		9.5							
$_{34}Se^{77}$		8.3							
$_{34}Se^{78}$	77.938	24.0							
$_{34}Se^{79,\ 81}$			57m	e^-		0.098			
$_{34}Se^{79,\ 81}$			19m	β^-	1.5				
$_{34}Se^{80}$	79.942	48.0							
$_{34}Se^{82}$		9.3							
$_{34}Se^{83}$			30m	β^-					
$_{35}Br^{78}$			6.4m	β^+, e^-, γ	2.3β^+	0.046, 0.108			
$_{35}Br^{79}$	78.929	50.6					3/2	+2.6	
$_{35}Br^{80}$			4.4h	e^-, γ		0.49, 0.037			
$_{35}Br^{80}$			18m	β^-, γ	2.0	<0.5			
$_{35}Br^{81}$	80.930	49.4					3/2	+2.6	
$_{35}Br^{82}$			34h	β^-, γ	0.465	0.547, 0.787, 1.35			
$_{35}Br^{83}$			140m	β^-	1.05	No γ			
$_{36}Kr^{78}$	77.945	0.35							
$_{36}Kr^{79,\ 81}$			34h	β^+	0.5				
$_{36}Kr^{79,\ 81}$			13s	e^-, γ		0.187			
$_{36}Kr^{79,\ 81}$			55s	e^-, γ		0.127			
$_{36}Kr^{80}$		2.01							
$_{36}Kr^{82}$	81.939	11.53							
$_{36}Kr^{83}$		11.53	113m	e^-		0.046, 0.029	9/2	−1.0	+0.04
$_{36}Kr^{84}$	83.938	57.11							
$_{36}Kr^{86}$	85.939	17.47							
$_{36}Kr^{87}$			74m	β^-					
$_{36}Kr^{87}$			4.5h	β^-					
$_{36}Kr^{88}$			3h	β^-	2.3				
$_{36}Kr^{89}$			~2m	β^-					
$_{36}Kr>^{90}$			<0.5m	β^-					
$_{37}Rb^{82}$			20m						
$_{37}Rb^{84}$			6.5h						
$_{37}Rb$			42m						
$_{37}Rb$			200h						
$_{37}Rb^{85}$		72.3					5/2	+1.340	

Element	Mass	Relative Abundance	Half Life	Radiation	Particles MEV	γ-Rays (MEV)	I	μ	q×10⁻²⁴
$_{37}Rb^{86}$			19.5d	β^-	1.60				
$_{37}Rb^{87}$		27.7	6×10^{10}y	β^-	0.13		3/2	+2.733	
$_{37}Rb^{88}$			18m	β^-	5.1				
$_{37}Rb^{89}$			15m	β^-, γ	3.8				
$_{37}Rb>^{90}$			80s	β^-					
$_{38}Sr^{84}$		0.56							
$_{38}Sr^{85}$			65d	K, γ		0.8			
$_{38}Sr^{85}$			70m	e^-, γ		0.170			
$_{38}Sr^{86}$		9.86							
$_{38}Sr^{87}$		7.02	2.7h	e^-, γ		0.386	9/2	−1.1	
$_{38}Sr^{88}$		82.56							
$_{38}Sr^{89}$			55d	β^-	1.32	No γ			
$_{38}Sr>^{90}$			7m	β^-					
$_{38}Sr>^{90}$			6h	β^-					
$_{39}Y^{86,88}$			105d	K, γ		0.908, 1.89, 3.0			
$_{39}Y^{87}$			14h	e^-, γ		0.5			
$_{39}Y^{87}$			80h	K		No γ			
$_{39}Y^{88}$			2.0h	β^+	1.2				
$_{39}Y^{89}$		100							
$_{39}Y^{90}$			60h	β^-	2.6				
$_{39}Y>^{90}$			3.3h						
$_{40}Zr^{89}$			78h	β^+	1.0	No γ			
$_{40}Zr^{89}$			4.5m	e^-, γ, K					
$_{40}Zr$			18m	β^-					
$_{40}Zr$			90m	β^-	~1.5				
$_{40}Zr$			70h	β^-	1.17				
$_{40}Zr^{90}$		48							
$_{40}Zr^{91}$		11.5							
$_{40}Zr^{92}$		22							
$_{40}Zr^{93}$			63d	β^-	~0.25				
$_{40}Zr^{94}$		17							
$_{40}Zr^{95}$			17.0h	β^-	1				
$_{40}Zr^{96}$		1.5							
$_{40}Zr^{97}$			6m	β^-	~1.9				
$_{41}Cb$			4m						
$_{41}Cb$			12m						
$_{41}Cb$			38m						
$_{41}Cb$			21h						

$_{41}Cb$			96h					
$_{41}Cb^{92}$			11d	β^-	1.38			
$_{41}Cb^{93}$	92.926	100	~55d	e^-		~0.15	9/2	+3.7
$_{41}Cb^{94}$			6.6m	β^-, γ	1.4	0.4		
$_{41}Cb^{95}$			75m	β^-	1			
$_{42}Mo^{92}$		15.5						
$_{42}Mo^{93}$			7h					
$_{42}Mo^{91, 93}$			17m	β^+	2.65			
$_{42}Mo^{94}$		8.7						
$_{42}Mo^{95}$	94.945	16.3						
$_{42}Mo^{96}$	95.946	16.8						
$_{42}Mo^{97}$	96.945	8.7						
$_{42}Mo^{98}$	97.944	25.4						
$_{42}Mo^{99}$			67h	β^-, γ	1.5	0.4		
$_{42}Mo^{100}$	99.945	8.6						
$_{42}Mo^{101}$			19m	β^-	1.8			
$_{43}{}^{96}$			2.7h	β^+				
$_{43}{}^{99}$		100	6.6h	e^-, γ		0.136		
$_{43}{}^{101}$			9m	β^-	1.1			
$_{43}$			90d	K, e^-		0.096		
$_{43}$			62d	K, γ				
$_{43}$			110h	K, e^-, γ	0.6	0.05, 0.5		
$_{43}$			55m	β^-, γ	2.5			
$_{43}$			36.5h	β^-				
$_{43}$			18s	β^-				
$_{43}$			~2d	K				
$_{44}Ru^{95}$			20m					
$_{44}Ru^{96}$	95.946	5						
$_{44}Ru^{98}$		?						
$_{44}Ru^{99}$	98.944	12						
$_{44}Ru^{100}$		14						
$_{44}Ru^{101}$		22						
$_{44}Ru^{102}$		30						
$_{44}Ru^{103}$			4h	β^-				
$_{44}Ru^{104}$		17						
$_{44}Ru^{105}$			20h	β^-				
$_{44}Ru$			39h	β^-				
$_{44}Ru$			11d					
$_{44}Ru$			90m					
$_{45}Rh^{101}$		0.08?						
$_{45}Rh^{103}$		99.92						
$_{45}Rh^{104}$			4.2m	e^-		0.055, 0.080		

Element	Mass	Relative Abundance	Half Life	Radiation	Particles MEV	γ-Rays (MEV)	I	μ	q×10⁻²⁴
$_{45}Rh^{104}$			44s	β^-	2.3				
$_{45}Rh^{105}$			46d	β^-					
$_{45}Rh$			1.1h	β^-					
$_{45}Rh$			3h						
$_{45}Rh$			10.7h						
$_{45}Rh$			3d						
$_{46}Pd^{102}$		0.8							
$_{46}Pd^{104}$	103.946	9.3							
$_{46}Pd^{105}$	104.945	22.6							
$_{46}Pd^{106}$	105.945	27.2							
$_{46}Pd^{107, 109}$			13h	β^-	1.03				
$_{46}Pd^{108}$	107.943	26.8							
$_{46}Pd^{110}$	109.942	13.5							
$_{46}Pd^{111}$			17m	β^-					
$_{47}Ag^{102}$			73m						
$_{47}Ag^{104}$			16.3m						
$_{47}Ag^{105}$			45d	K		0.29, 0.35, 0.42, 0.50, 0.62			
$_{47}Ag^{106}$			24.5m	β^+	2.04	No γ			
$_{47}Ag^{106}$			8.2d	K, e^-, γ	1.2	1.06, 1.63			
$_{47}Ag^{107}$	106.948	52.5	40s (or Ag^{109})	e^-		0.086	1/2	−0.1	
$_{47}Ag^{108}$	107.947		2.3m	β^-	2.8				
$_{47}Ag^{109}$		47.5					1/2	−0.2	
$_{47}Ag^{110}$			22s	β^-, γ	2.8				
$_{47}Ag^{108, 110}$			225d	β^-		0.65; 0.93; 1.51			
$_{47}Ag^{111}$			7.5d	β^-		No γ			
$_{47}Ag^{112}$			3.2h	β^-, γ	2.2				
$_{48}Cd^{106}$		1.4							
$_{48}Cd^{107, 109}$			6.7h	K, γ		0.8, 0.092			
$_{48}Cd^{107, 109}$			~90d	K					
$_{48}Cd^{108}$		1.0							
$_{48}Cd^{109}$			33m	β^+					
$_{48}Cd^{110}$		12.8							
$_{48}Cd^{111}$		13.0					1/2	−0.65	
$_{48}Cd^{112}$		24.2							
$_{48}Cd^{113}$		12.3					1/2	−0.65	
$_{48}Cd^{114}$		28.0							
$_{48}Cd^{115}$			2.5d	β^-, γ	1.11	0.55			
$_{48}Cd^{116}$		7.3							

$_{48}Cd^{117}$			3.75h	β^-					
$_{48}Cd$			50m	e^-					
$_{49}In^{110}$			65m	β^+	2.0				
$_{49}In^{111}$			20m	β^+, γ, e^-	$1.7\beta^+$	0.16			
$_{49}In^{112}$			2.7d	K, γ, e^-		0.17, 0.25			
$_{49}In^{113}$		4.5	105m	γ, e^-		0.39	9/2	+5.49	
$_{49}In^{114}$			48d	e^-		0.19			
$_{49}In^{114}$			72s	β^-	1.98				
$_{49}In^{115}$		95.5	4.1h	e^-, γ		0.34	9/2	+5.49	+0.84
$_{49}In^{116}$			13s	β^-	2.8	No γ			
$_{49}In^{116}$			54m	β^-, γ	0.85				
						0.43;1.1;1.3;2.3			
$_{49}In^{117}$			117m	β^-, γ, e^-	1.73				
$_{50}Sn^{112}$		1.1							
$_{50}Sn^{113}$			70–105d	K, e^-, γ		0.085			
$_{50}Sn^{114}$		0.8							
$_{50}Sn^{115}$		0.4							
$_{50}Sn^{116}$	115.942	15.5							
$_{50}Sn^{117}$		9.1					1/2	−0.89	
$_{50}Sn^{118}$	117.939	22.5							
$_{50}Sn^{119}$	118.938	9.8					1/2	−0.89	
$_{50}Sn^{120}$	119.94	28.5							
$_{50}Sn^{<119}$			25m	β^-					
$_{50}Sn^{<119}$			3h	β^-					
$_{50}Sn^{<119}$			13d	β^-					
$_{50}Sn^{122}$	121.944	5.5							
$_{50}Sn^{124}$	123.943	6.8							
$_{50}Sn^{125}$			9m	β^-					
$_{50}Sn$			40m	β^-					
$_{50}Sn$			26h	β^-					
$_{50}Sn$			10d	β^-					
$_{50}Sn$			400d						
$_{51}Sb$			3.5m	β^-					
$_{51}Sb^{116,\ 118}$			3.6m	β^+					
$_{51}Sb^{120}$			17m	β^+	1.53				
$_{51}Sb^{121}$		56					5/2	+3.7	
$_{51}Sb^{122}$			2.8d	β^-	0.81, 176	0.96			
$_{51}Sb^{123}$		44					7/2	+2.8	
$_{51}Sb^{124}$			60d	β^-K, γ	1.53	1.75; 0.07			
$_{51}Sb^{<126}$			3h	β^-					
$_{51}Sb^{<126}$			~45d						
$_{51}Sb^{<126}$			~2y						

Element	Mass	Relative Abundance	Half Life	Radiation	Particles MEV	γ-Rays (MEV)	I	μ	q×10⁻²⁴
$_{51}Sb^{127}$			80h	β^-					
$_{51}Sb^{128}$			4.2h	β^-					
$_{51}Sb>^{131}$			<10m	β^-					
$_{52}Te^{120}$		<0.1							
$_{52}Te^{121}$			125d	K, e^- γ		0.5			
$_{52}Te^{122}$		2.9							
$_{52}Te^{123}$		1.6							
$_{52}Te^{124}$		4.5							
$_{52}Te^{125}$		6.0							
$_{52}Te^{126}$	125.937	19.0							
$_{52}Te^{127}$			90d	e^-		0.086			
$_{52}Te^{127}$			9.3h	β^-					
$_{52}Te^{128}$	127.936	32.8							
$_{52}Te^{129}$			32d	e^-		~0.10			
$_{52}Te^{129}$			72m	β^-					
$_{52}Te^{130}$		33.1							
$_{52}Te^{131}$			30h	e^-		0.17			
$_{52}Te^{131}$			25m	β^-					
$_{52}Te>^{131}$			43m	β^-					
$_{52}Te>^{131}$			60m	β^-					
$_{52}Te>^{131}$			77h	β^-					
$_{52}Te>^{131}$			~15m	β^-					
$_{53}I^{124}$			4.0d	β^+					
$_{53}I^{126}$			13.0d	β^-, γ	1.1	0.5			
$_{53}I^{127}$	126.933	100					5/2	+2.8	−0.8?
$_{53}I^{128}$			25m	β^-, γ	1.2, 2.1	0.4			
$_{53}I^{130}$			12.6h	β^-, γ	0.61; 1.03	0.417, 0.535, 0.670, 0.740			
$_{53}I^{131}$			8.0d	β^-, e^-, γ	0.595	0.367, 0.080			
$_{53}I>^{131}$			2.4h	β^-					
$_{53}I>^{131}$			54m	β^-					
$_{53}I>^{131}$			6.6h	β^-					
$_{53}I>^{131}$			22h	β^-					
$_{54}Xe^{124}$		0.094							
$_{54}Xe^{126}$		0.088							
$_{54}Xe^{127}$			75s	e^-, γ		0.175, 0.125			
$_{54}Xe^{127}$			34d	e^-, γ		0.9			
$_{54}Xe^{128}$		1.90							
$_{54}Xe>^{131}$			4.3d	e^-		0.083			

$_{54}Xe>^{131}$			9.4h						
$_{54}Xe^{129}$	128.946	26.23					1/2	−0.9	
$_{54}Xe^{130}$		4.07							
$_{54}Xe^{131}$		21.17					3/2	+0.7	0±0.03
$_{54}Xe^{132}$	131.946	26.96							
$_{54}Xe^{133}$			5.4d						
$_{54}Xe^{134}$		10.54							
$_{54}Xe^{135}$			9.6h						
$_{54}Xe^{136}$		8.95							
$_{54}Xe^{137}$			68m						
$_{54}Xe^{139}$			<0.5m	β^-					
$_{54}Xe>^{139}$			17m	β^-					
$_{55}Cs^{133}$		100					7/2	+2.558	
$_{55}Cs^{134}$			3h	β^-	1				
$_{55}Cs^{134}$			1.7y	β^-, γ	0.9				
$_{55}Cs^{139}$			7m	β^-					
$_{55}Cs$			40s	β^-					
$_{55}Cs>^{139}$			33m	β^-	2.6				
$_{56}Ba^{130}$		0.101							
$_{56}Ba^{132}$		0.097							
$_{56}Ba^{133}$			30h	e^-, γ		0.276			
$_{56}Ba^{122}$			340h	K, e^-, γ		0.017			
$_{56}Ba^{134}$		2.42							
$_{56}Ba^{135}$		6.59					3/2	+0.8363	
$_{56}Ba^{136}$		7.81							
$_{56}Ba^{137}$		11.32					3/2	0.9354	
$_{56}Ba^{138}$	137.916	71.66							
$_{56}Ba^{139}$			86m	β^-	1	0.6			
$_{56}Ba$			3m						
$_{56}Ba>^{140}$			14m	β^-					
$_{56}Ba>^{140}$			<1m	β^-					
$_{57}La^{137}$			175h,	K, γ		0,88			
$_{57}La^{138}$			40h,	β^-, γ		2.00			
$_{57}La^{139}$	138.955	100					7/2	+2.8	
$_{57}La^{140}$			31h	β^-	0.8				
$_{57}La>^{140}$			2.5h	β^-					
$_{57}La>^{140}$			<30m	β^-					
$_{58}Ce^{136}$		0.002							
$_{58}Ce^{138}$		0.002							
$_{58}Ce^{139}$			2.1m	β^+					
$_{58}Ce^{140}$		90							
$_{58}Ce^{141,\ 143}$			15d		0.12				

Element	Mass	Relative Abundance	Half Life	Radiation	Particles MEV	γ-Rays (MEV)	I	μ	q×10⁻²⁴
$_{58}Ce^{142}$		10							
$_{59}Pr^{140}$			3.4m	β^+	2.40				
$_{59}Pr^{141}$		100							
$_{59}Pr^{142}$			19.3h	β^-	2.14	1.9			
$_{60}Nd^{142}$		25.95							
$_{60}Nd^{143}$		13.0							
$_{60}Nd^{144}$		27.6							
$_{60}Nd^{145}$		9.2							
$_{60}Nd^{146}$	145.960	16.5							
$_{60}Nd^{147}$			84h	β^-					
$_{60}Nd^{148}$	147.961	6.8							
$_{60}Nd^{149}$			2.0h	β^-					
$_{60}Nd^{150}$	149.967	5.95							
$_{60}Nd^{151}$			21m	β^-					
$_{61}$			12.5h	β^-					
$_{62}Sm^{144}$		3							
$_{62}Sm^{147}$		17							
$_{62}Sm^{148}$		14	6.3×10^{11}y	α	2.4				
$_{62}Sm^{149}$		15							
$_{62}Sm^{150}$		5							
$_{62}Sm^{152}$		26							
$_{62}Sm^{154}$		20							
$_{62}Sm$			21m	β^-					
$_{62}Sm$			46h	β^-					
$_{63}Eu^{150}$			27h	β^+					
$_{63}Eu^{151}$		49.1					5/2	+3.4	~+1.2
$_{63}Eu^{152,\ 154}$			9.2h	β^-, γ, e^-	1.88	0.123 0.725			
$_{63}Eu^{152,\ 154}$			>1y	β^-, γ	0.8				
$_{63}Eu^{152,\ 154}$			12m						
$_{63}Eu^{152,\ 154}$			105m						
$_{63}Eu^{153}$		50.9					5/2	+1.5	~+2.5
$_{64}Gd^{152}$		0.2							
$_{64}Gd^{154}$		1.5							
$_{64}Gd^{155}$	154.977	20.7							
$_{64}Gd^{156}$	155.976	22.6							
$_{64}Gd^{157}$	156.976	16.7							
$_{64}Gd^{158}$		22.6							
$_{64}Gd^{159,\ 161}$			8h						

$_{64}Gd^{160}$		15.7							
$_{65}Tb^{159}$		100							
$_{65}Tb^{160}$			3.9h	β^-					
$_{66}Dy^{158}$		0.1							
$_{66}Dy^{160}$		1.5							
$_{66}Dy^{161}$		21.6							
$_{66}Dy^{162}$		24.6							
$_{66}Dy^{163}$		24.6							
$_{66}Dy^{164}$		27.6							
$_{66}Dy^{165}$			2.5h	β^-	1.2	1.1			
$_{66}Dy?$			2.2m	β^+					
$_{67}Ho^{164}$			47m	β^-					
$_{67}Ho^{165}$	164.98	100							
$_{67}Ho^{166}$			35h	β^-	1.6				
$_{68}Er^{162}$		0.25							
$_{68}Er^{164}$		2.0							
$_{68}Er^{165}$			1.1m	β^+					
$_{68}Er^{166}$		35.2							
$_{68}Er^{167}$		23.5							
$_{68}Er^{168}$		29.3							
$_{68}Er^{169,\ 171}$			7m						
$_{68}Er^{169,\ 171}$			12h	β^-					
$_{68}Er^{170}$		9.8							
$_{69}Tm^{169}$		100							
$_{69}Tm^{170}$			105d						
$_{70}Yb^{168}$		0.06							
$_{70}Yb^{170}$		2							
$_{70}Yb^{171}$		8.8					1/2	+0.45	
$_{70}Yb^{172}$		23.5							
$_{70}Yb^{173}$		16.7					5/2	−0.65	+3.9
$_{70}Yb^{174}$		37.2							
$_{70}Yb^{175,\ 177}$			3.5h						
$_{70}Yb^{176}$		11.8							
$_{70}Yb?$			41h						
$_{71}Lu^{175}$		97.5					7/2	+2.6	+5.9
$_{71}Lu^{176}$		2.5	7×10^{10}y	β^-	0.22		$\geqslant 7$	+3.8	(+6)−(+8)
$_{71}Lu^{176,\ 177}$			4h						
$_{71}Lu^{176,\ 177}$			6d						
$_{72}Hf^{172}?$		<0.1							
$_{72}Hf^{174}$		0.3							
$_{72}Hf^{176}$		5							
$_{72}Hf^{177}$		19							

Element	Mass	Relative Abundance	Half Life	Radiation	Particles MEV	γ-Rays (MEV)	I	μ	q×10⁻²⁴
$_{72}Hf^{178}$		28							
$_{72}Hf^{179}$		18							
$_{72}Hf^{180}$		30							
$_{72}Hf^{181}$			55d	β^-					
$_{73}Ta^{180}$			14–21m						
$_{73}Ta^{180}$			8.2h	K, e^-, γ, β^-?	<0.5e^-?				
$_{73}Ta^{181}$	180.928	100							
$_{73}Ta^{182}$			97d	β^-					
$_{74}W^{180}$		~0.2							
$_{74}W^{182}$		22.6							
$_{74}W^{183}$		17.3							
$_{74}W^{184}$	184.00	30.1							
$_{74}W^{185}$			77d	β^-, γ	0.4–0.5				
$_{74}W^{186}$		29.8							
$_{74}W^{187}$			23h	β^-, γ	0.5, 1.4	0.9			
$_{75}Re$			41–55m	β^+					
$_{75}Re$			13m						
$_{75}Re$			>40d	K?, γ					
$_{75}Re^{185}$		38.2					5/2	+3.3	+2.8
$_{75}Re^{186}$			90h	β^-	1.05	No γ			
$_{75}Re^{187}$	186.981	61.8					5/2	3.3	+2.6
$_{75}Re^{188}$			18h	β^-	2.5				
$_{76}Os^{184}$		0.018							
$_{76}Os^{186}$		1.59							
$_{76}Os^{187}$		1.64							
$_{76}Os^{188}$		13.3							
$_{76}Os^{189}$		16.1							
$_{76}Os^{190}$	190.038	26.4							
$_{76}Os^{191}$			32h	β^-	1.5				
$_{76}Os^{192}$	192.038	41.0							
$_{76}Os^{193}$			17d		0.35				
$_{77}Ir^{191}$	191.040	38.5							
$_{77}Ir^{192,\ 194}$			1.5m	β^-					
$_{77}Ir^{192,\ 194}$			19h	β^-	2.2				
$_{77}Ir^{192,\ 194}$			60d	β^-					
$_{77}Ir^{193}$	193.041	61.5							
$_{78}Pt^{192}$		0.8							
$_{78}Pt^{194}$	194.040	30.2							
$_{78}Pt^{195}$	195.040	35.3					1/2	+0.6	

$_{78}Pt^{196}$	196.039	26.6							
$_{78}Pt^{197}$			18h	β^-					
$_{78}Pt^{197}$			3.3d	β^-					
$_{78}Pt^{198}$	198.044	7.2							
$_{78}Pt^{199}$			31m	β^-					
$_{79}Au^{196}$			13h	β^-					
$_{79}Au^{196, 197}$			4–5d	β^-, γ	0.36	0.41			
$_{79}Au^{197}$	197.039	100					3/2	+0.195	
$_{79}Au^{198}$			2.7d	β^-. γ	0.78	0.42			
$_{79}Au^{199}$			3.3d	β^-, γ					
$_{80}Hg^{196}$		0.15							
$_{80}Hg$			43m	K, e^-, γ	<0.4	<0.25			
$_{80}Hg^{197}$			25h	e^-, γ		0.161; 0.130			
$_{80}Hg^{197}$			65h	K, e^-, γ		0.075			
$_{80}Hg^{198}$		10.11							
$_{80}Hg^{199}$		17.03					1/2	+0.547	
$_{80}Hg^{200}$	200.028	23.26							
$_{80}Hg^{201}$		13.17					3/2	−0.607	+05
$_{80}Hg^{202}$		29.56							
$_{80}Hg^{204}$		6.72							
$_{81}Tl^{200}$			4m						
$_{81}Tl^{200}$			3.8h						
$_{81}Tl^{203}$	203.057	29.1					1/2	+1.4	
$_{81}Tl^{204}$			4.23m	β^-	1.6	No γ			
$_{81}Tl^{205}$	205.057	70.9					1/2	+1.4	
$_{81}Tl^{206}$			1–2y	β^-					
$_{81}AcC''^{207}$			4.76m	β^-	1.47	<0.5			
$_{81}ThC''^{208}$			3.1m	β^-	1.795				
$_{81}RaC''^{210}$			1.32m	β^-	1.95				
$_{82}Pb^{204}$	204.058	1.48							
$_{82}Pb^{206}$		23.59	52h	e^-, γ		~0.5			
$_{82}Pb^{207}$		22.64					1/2	+0.6	
$_{82}Pb^{208}$	208.057	52.29							
$_{82}Pb^{209}$			3.0h	β^-					
$_{82}RaD^{210}$			22.2y	β^-	0.025				
$_{82}AcB^{211}$			36.1m	β^-	1.00				
$_{82}ThB^{212}$			10.6h	β^-	0.36				
$_{82}RaB^{214}$			26.8m	β^-	0.65 (0.241; (11.5%); 0.294 (25.8%); 0.350 (45.0%))				

Element	Mass	Relative Abundance	Half Life	Radiation	Particles MEV	γ-Rays (MEV)	I	μ	q×10⁻²⁴
$_{83}Bi^{209}$	209.055	100					9/2	+3.6	−0.4
$_{83}RaE^{210}$			4.975d	β^-	1.17				
$_{83}AcC^{211}$			2.16m	α	6.610, 261				
$_{83}ThC^{212}$			60.6m	α, β	α = 6.082, etc. β = 2.25				
$_{83}RaC^{214}$			19.72m	α, β (99.96%)	α = 5.505, 5.444 β = 3.13 { 0.607; 0.766; 0.933; 1.120; 1.238; 1.379; 1.761; 2.198				
$_{84}Po^{210}$			139.5d	α	5.298				
$_{84}AcC'^{211}$			2×10^{-3}s	α	7.435				
$_{84}ThC'^{212}$			4×10^{-8}s	α	8.7759 etc.				
$_{84}RaC'^{214}$			2×10^{-4}s	α	7.6802 etc.				
$_{84}AcA^{215}$			2×10^{-3}s	α	7.366				
$_{84}ThA^{216}$			0.145s	α	6.7744				
$_{84}RaA^{218}$			3.05m	α	5.9981	no γ			
$_{85}$ ^{211}			7.5h	α, K	6				
$_{86}An^{219}$			3.92s	α	6.825, 6.560, 6.435				
$_{86}Tn^{220}$			54.50s	α	6.2818				
$_{86}Rn^{222}$			3.823d	α	5.4860	no γ			
$_{87}AcK^{223}$			21m	β^-	1.2				
$_{88}AcX^{223}$			11.2d	α	5.718, 5.606, 5.532				
$_{88}ThX^{224}$			3.64d	α	5.6813				
$_{88}Ra^{226}$	226.10		1600y	α	4.792, 4.611	0.184 (1.2%)			
$_{88}MsTh_1^{228}$			6.7y	β^-	0.053				
$_{89}Ac^{227}$			13.4y	α, β^-	α=5.13 β=0.22				
$_{89}MsTh_2^{228}$			6.13h	β^-	1.6				
$_{90}RdAc^{227}$			18.9d	α	6.050, 5.989, 5.765, 5.718 etc.				
$_{90}RaTh^{228}$			1.90y	α	5.419, 5.334				
$_{90}Io^{230}$			82,250y	α	4.59				
$_{90}AcY^{231}$			24.64h	β^-					
$_{90}Th^{232}$			1.389×10^{10}y	α	4.08				
$_{90}Th^{233}$	232.12		26m	β^-					
$_{90}UX_1^{234}$			24.1d	β^-, e^-, γ	0.228	0.802			

$_{91}Pa^{231}$			3.2×10^{4}y	α	5.02				
$_{91}Pa^{233}$			25d	β^-, γ, e^-	0.23	0.30; <0·1			
$_{91}UX_2{}^{234}$			1.14m	β^-	2.33				
$_{91}UZ^{234}$			6.7h	β^-	0.9				
$_{92}UII^{234}$		0.006	2.7×10^{5}y	α	4.68				
$_{92}AcU^{235}$	235.12	0.71	7.1×10^{8}y	α	4.40				
$_{92}U^{237}$			~7d	β^-, γ	0.26				
$_{92}UI^{238}$	238.14	99.28	4.53×10^{9}y	α	4.13				
$_{92}U^{239}$			23m	β^-					
$_{92}{}^{239}$			2.3d	β^-	0.47	0.22, 0.27			

Index to Authors

A

Abbe, 165
Abbott, A. D., 119
Abbott, Lynn D. F., Jr., 597
Abderhalden, E., 544, 1153
Abel, J. J., 596
Abell, 982, 985
Abitz, W., 233, 664
Aborg, C. G., 1131, 1135
Abraham, E. P., 704
Abramson, H. A., 409, 410, 638, 663, 930, 949, 1115
Adair, 1013
Adair, G. S., 416, 433
Adair, M. E., 416, 433
Adam, N. K., 23, 26, 29, 30, 58, 72, 115, 120, 249, 252, 1131, 1135
Adams, G. D., 387
Adams, P. D., 701, 704
Adamkiewicz, 926
Addison, O. C., 120
Adinoff, B., 60, 102
Adkins, Homer, 551, 597
Adler, E., 699, 704, 705
Adrian, Sir E. D., 4, 12, 579, 597, 764, 766, 768, 1152
Aebersold, P. C., 379
Agardh, J. G., 618, 622, 628, 661
Agid, R., 704
Agliardi, N., 467, 471
Aguilhon, 558, 597
Akesson, A., 410, 411
Albers, H., 705
Albers, V. U., 604, 609, 610
Albright, F., 713, 736, 743, 752, 991, 994
Aldrich, 750, 752
Alexander, A. E., 97, 1131, 1135
Alexander, Jerome, 1, 10, 11, 12, 279, 373, 457, 458, 471, 545, 557, 571, 590, 600, 682, 684, 808, 815, 818, 820, 830, 850, 851, 858, 859, 880, 916, 1049
Alexander, L., 1147, 1152
Alexander-Jackson, Eleanor G., 569
Alexander, W., 740, 752
Allard, H. A., 787, 807
Allen, 1054
Allen, C. E., 628, 661
Allen, E., 597
Allen, Edgar, 815, 1014, 1015
Allison, F. E., 702, 704
Allison, V. D., 1115
Alsberg, C. A., 667, 682
Altberg, W. J., 370
Alvarez-Tostado, C., 409, 410
Alzheimer, A., 1138
Amberson, W. R., 1146, 1152
Ambronn, H., 624, 631, 634, 660, 661, 662, 1131, 1135, 1152
Ames, T., 1152
Amontons, 247, 252
Anderson, 38, 981, 985, 1022
Anderson, C. D., 9
Anderson, D. B., 634, 648, 654, **661, 662**
Anderson, H., 512
Anderson, T. F., 101, 186, 220, 222, 223, 225, 226, 228, 229, 233, 235, 797, 798, 800, 808
Andersson, K. J. I., 421, 422, 432, 657
Andervont, 1003, 1008, 1010, 1012, 1024, 1027
Andrade, E. N. da C., 240, 252, 360, 370
Andrews, A. C., 957, 970, 977, 1031, 1034, 1035
Andrewes, C. H., 1115, 1118
von Angerer, C., 1118
Ansbacher, S., 702, 704, 706, 707, 734, 738
Anson, M. L., 563, 597, 799, 807
Anthony, 304, 308
Antopol, W., 692, 705
Antvors, E., 524, 525
Appell, E. A., 1117
Aquilonias, L., 844, 846, 849
Archibald, W. J., 412, 415, 432
Ardenne, M. von, 162, 165, 166, 168, 169, 170, 174, 175, 176, 178, 181, 186, 188, 195, 212, 228, 233, 234, 370, 489, 511
Armangué, M., 978
Armbruster, 443
Arminio, J., 756, 762
Armstrong, Charles, 586
Armstrong, E. C., 512
Armstrong, H. G., 1083
Armstrong, Joanne L., 807
Arnd, O., 1126, 1135
Arnold, A., 684, 704, 706
Arnold, F. A., 598
Arnold, L., 1115
Arnold, W., 557, 598, 602, 609
Arnow, L. Earle, 582, 598
Arrhenius, S., 240, 252, 473, 567, 767, 768, 771, 772, 773, 776, 778, 780, 783, 970, 971, 977
Arthus, 982, 985
Asbach, H. R., 371
Ascham, J. K., 725, 737
Ascherson, F. M., 880
Aschoff, L., 1077, 1082
Astbury, W. T., 145, 233, 529, 530, 544, 562, 646, 662, 836, 839, 841, 848, 883, 897, 900, 1132, 1136
Astrup, T., 955
Astwood, E. B., 744, 752
Atkin, 536, 544
Atkinson, W. B., 705
Aub, J. C., 713, 736, 991, 994
Auer, J., 736, 981, 985, 1144, 1153
Auerbach, L., 1147, 1152
Auhagen, E., 686, 704, 721, 737
Austin, 443
Avery, O. T., 977
Axelrod, A. E., 556, 683, 696, 697, 701, 704
Avogadro, 413
Avrami, M., 518
Ayers, W. C., 761
Azuma, R., 887, 899

B

Baas-Becking, L. G. M., 604, 609
Babers, F. H., 977
Babinet, 895
Bachman, C. H., 164, 165, 183, 234
Bachmann, W. E., 742
Bachrach, 471
Bacmeister, A., 1082
Badger, 1017
Baeyer, E., 899
Bagg, 1016
Baier, W. E., 659, 662
Bailey, 530, 544
Bailey, A. J., 648, 661, 662
Bailey, E. D., 432
Bailey, I. W., 648, 661, 662
Bailey, K. C., 308, 898, 900
Baird, T. T., 1154
Baker, 308
Baker, A. H., 1115, 1118
Baker, R. F., 130, 170, 176, 182, 234
Baker, Walter C., 1095
Baker, W. O., 252, 272, 280, 326
Baker, Z., 1117, 1129, 1135
Balareff, 455, 456
Baldi, F., 1152
Baldwin, 667, 682
Baldwin, G. C., 387
Ball, 1038
Ball, E. G., 695, 698, 699, 704, 705
Ballard, A., 474, 475, 504, 512
Ballentine, R., 852, 859
Ballif, L., 696, 704
Balls, W. L., 629, 630, 662
Baly, E. C. C., 608, 609
Bamberger, P., 1129
Banca, M. C., 166, 235
Bancroft, 257
Bancroft, D., 336
Bancroft, Wilder D., 475, 496, 512, 1129, 1135, 1138, 1152, 1158, 1162
Banerji, G. G., 685, 686, 691, 704
Bang, F., 1026, 1049
Banga, I., 687, 697, 704, 721, 737
Banzhaf, E. J., 977, 1152
Barcroft, Sir J., 986, 993, 994, 1157, 1161
Bardy, H., 928, 949
Barenberg, L. H., 1115, 1116
Barger, C., 718, 737
Barger, G., 926
Barker, 742, 752
Barnes, 1024
Barnes, C. R., 626, 649, 662
Barnes, U. E., 1117
Barnes, W. H., 861, 863
Barnes, H. T., 598
Barnes, R. B., 174, 209, 210, 211, 234, 1198
Barrett, 441, 443
Barron, C., 964, 978
Barron, E. S. G., 688, 689, 690, 691, 700, 703, 704
Barrows, F. L., 648, 662
Barsoum, G. S., 922, 923, 949
Barth, L. G., 851, 854, 859
Barth, T. F. W., 326
Bartley, S. H., 1152
Basinski, D. H., 703, 706
Bass, L. W., 836, 849
Basset, J., 332, 337
Bates, 681, 682

Bateson, Sir William, 584, 809
Bauer, 515, 1124, 1135, 1142
Bauer, H., 823, 824, 828, 829, 848
Bauer, J. H., 412, 418, 419, 423, 427, 428, 430, 432, 433
Bauer, W., 713, 736, 991, 994
Baumann, C. A., 695, 704
Baumann, E., 600
Baumberger, J. P., 955
Baver, L. D., 193
Bawden, F. C., 789, 801, 802, 807, 848
Bawn, 682
Bayer, A. von, 892
Bayliss, Sir W. M., 596
Beach, J. Y., 130
Beadle, G. W., 582, 599, 820, 823, 850
Beams, 418, 420, 426, 427, 428, 430, 432, 433, 434
Bear, 543, 544, 674, 675, 678, 679, 683
Bear, L., 1131, 1136, 1149, 1154
Bear, R. S., 231, 234, 868, 875
Beard, D., 411, 433
Beard, J. W., 411, 430, 433, 434, 807, 1000, 1029, 1032, 1049, 1050
Beattie, J. M., 938, 949
Beauvallet, U., 704
Bechhold, H., 791, 807
Becke, 444, 456
Becker, M., 692, 704
Beckmann, C. O., 433
Beeck, O., 130
Beer, de, 752
Beeman, Norvil, 100
Beevers, C. A., 141, 145
Behrens, H., 628, 662
Behrens, J. W., 622, 662
Behrens, M., 843, 844, 849
Beidl, 982, 985
Beijerinck, U. W., 785, 807
Beischer, D., 234
Beisele, J., 830, 848
Bell, 544, 682
Bell, D. J., 12, 560, 598
Bell, F. O., 836, 841, 848
Belling, J., 816, 822, 827, 831, 832, 833, 839, 847, 848
Belonischkin, B., 1082
Belynski, S. W., 370
Bendich, A., 955
Bendien, W. M., 636, 666
Bender, J. F., 166, 235
Bennett, 456
Bensley, R. R., 662
Berberian, D., 941, 949
Berenblum, 1001, 1026
Berg, R. L., 727, 737
Berg, W. F., 489, 512
Bergel, F., 689, 704
Berger, E., 968, 978
Bergmann, 968
Bergmann, L., 338, 370
Bergmann, U., 536, 544, 578
Berger, C. A., 831, 848
Berger, Hans, 765
Berkeley, Bishop, 12
Berkeley, E. E., 653, 662
Berkman, S., 550, 696, 704
Berluga, R. Y. S., 370
Bernal, J. D., 95, 272, 280, 493, 512, 646, 662, 790, 796, 805, 807, 840, 848
Bernard, Claude, 750, 918, 985, 994
Bernard, H., 977
Bernfeld, 680, 682, 683
Bernheim, F., 587, 696, 705
Bernheim, F., 741, 752
Bernini, Gian, 1951
Bernstein, 884
Bernstein, J., 1152
Bernstein, T. B., 1115
Bernthsen, 600
Bert, Paul, 1083
Berthold, G., 620, 662
Bertram, S., 234
Bertrand, J., 1152
Berzelius, 546, 610
Berzeller, 1152
Bessey, O. A., 701, 706, 730, 732, 733, 734, 738
Best, R. J., 790, 791, 807
Betts, R. H., 1117
Betz, U. D., 119
Beudant, 444
Beutner, R., 1135
Bevan, E. J., 630, 662
Biancani, E., 368, 370, 371
Biancani, H., 371
Biasotti, A., 752, 989, 994
Bichat, 42
Bielenberg, W., 467, 471
Bier, O., 921, 922, 923, 925, 949
Bigg, E., 1117
Bikerman, J. J., 87, 88
Billequin, U., 617
Billitzer, J., 977
Bills, C. E., 734, 738
Biltz, W., 977
Binkley, S. B., 955
Binns, D., 1152
Birch, F., 328, 337
Birch, T. W., 685, 704, 725, 737
Birkhäuser, H., 692, 706, 707
Birkofer, L., 707
Birnie, 441, 443
Bischoff, 1037, 1038
Biscoe, J., 234, 326, 418, 433, 808
Bishop, G. H., 1152
Bittner, J. J., 1008, 1009, 1011, 1012, 1040
Bjerknes, C. A., 370
Bjerknes, V., 370
Björnstahl, Y., 433
Black, R. J., 926, 949
Black, S. A., 420, 432
Blackfan, K. D., 1115
Blackett, P. M. S., 375
Blackman, F. F., 602, 609
Blair, W. G. Scott, 510, 512
Blake, 456
Blake, J. T., 286, 308
Blakeslee, A. F., 598, 648
Blakeslee, A. S., 2, 12
Blatt, U. L., 1115
Blinc, 682
Blinks, L. R., 607, 609, 610
Bliss, S. W., 988, 994
Blix, G., 406, 410
Block, 544
Block, R. J., 848
Blodgett, K. B., 87
Blondlot, 42
Blonquist, 456
Bloom, W., 938, 941, 950
Bloomfield, A. L., 1115
Bloomquist, 440, 443
Bloor, 1119
Blum, H. F., 588, 598, 589, 1115
Blum, W., 12
Blumberg, H., 372
Blumenthal, H. T., 999, 1038
Blumfield, 1053
Bluski, J., 1136
Bobolev, V., 370
Bodian, D., 844, 849
Bodine, J. H., 588, 598
von Boedecker, 588, 598
Boeder, P., 796, 807
Boehm, G., 1148, 1149, 1152
Boehm, J., 897, 900
Boehm, R. U., 637, 662
Boelter, L. U. K., 1115
Boell, E. J., 598
Boersch, H., 162, 166, 172, 181, 234
Boestad, G., 412, 433
Bogert, C. M., 581, 598
Boisbaudran, du, 444, 456
Boivin, A., 977
Bolam, T. R., 514, 518, 762, 784
Bolduan, O. E., 104
Boll, Franz, 753
Bolling, D., 926, 949
Bollinger, G. M., 1153
Bollman, J. L., 735, 738
Boltzmann, 294
Bondy, 347, 349, 350, 351, 357, 359, 363, 365, 373
Bondy, C., 1088
Bonner, J., 638, 662
Bonnet, V., 1152
Booher, Lela E., 707, 722, 730, 733, 737, 738
Boothroyd, E. R., 823, 850
Boothby, W., 1090
Boppel, 668, 682
Bordet, C., 950
Bordet, J., 569, 969, 977
Borelli, P., 662
Born, U., 512
Borries, B. von, 162, 166, 169, 170, 174, 182, 188, 212, 222, 234, 808
Borrel, A., 938, 949
Bose, Sir J. C., 609
Bott, P. A., 950
Bouchard, J., 365, 370
Boulanger, P., 705
Bourdillon, J., 409, 410
Bourdillon, R. B., 1115
Boutaric, A., 365, 370
Boveri, 590, 598
Bovie, 332
Bowman, 142
Bowman, F. B., 919, 949
Bowman, F. H., 620, 628, 662
Bowman, K. L., 979
Boyd, E., 101, 244
Boyd, G. E., 86, 101, 102
Boyd, T. A., 252
Boyd, Wm. C., 957, 959, 963, 964, 975, 976, 977, 978
Boyle, R. W., 347, 357, 359, 363, 367, 370, 1088
Boyle, Robert, 444, 456
Bozeman, M. L., 829, 848, 849
Bozler, E., 895, 896, 900
Bracconot, H., 610, 662
Brachet, J., 598, 843, 844, 845, 847, 848
Bradfield, Robert, 192
Bradford, 518
Bradley, A. J., 326
Brady, A. P., 104, 110
Brady, G. S., 234
Bragg, Sir W. H., 127, 141, 142, 145, 280, 326, 631, 662
Brainard, D. H., 1117
Brand, 544
Brandt, K., 849
Brandt, O., 356, 357, 360, 361, 363, 370, 371
Branfoot, U. H., 619, 662
Braun, A. C., 586, 600
Braunmuehl, A., 1138, 1152
Brawn, C. E. H., 10, 12
Bray, 443
Brazier, U., 1152
Brecht, K., 704
Bredig, G., 556, 598
Breedis, C., 1049
Bremer, F., 1152
Breslauer, F., 928, 949
Bresler, 293
Bretschneider, H., 467, 471
de Bretteville, A., 270, 276
Breusch, F. L., 736
Bridges, C. B., 11, 572, 598, 600, 808, 811, 815, 820, 821, 823, 824, 830, 831, 848, 858, 859, 1049, 1196, 1197
Bridgman, W. B., 252
Bridgman, P. W., 327, 336, 337
Briggs, 1033
Brimhall, 682
Brinkhous, K. M., 955, 956, 957
Brittain, J., 410
Brobecker, W. M., 379
Brockway, 130
Broda, E. E., 760, 761

Brode, Wallace R., 450
Broginart, 617
de Broglie, 121, 515
Brohult, S., 356, 371, 434, 567
Brömel, H., 699, 706
Brönsted, J. N., 257, 265
Brooks, G. L., 1115
Brown, A. N., 709, 736
Brown, E. P., 371
Brown, F., 646, 662
Brown, F. E., 99, 100
Brown, H. W., 1118
Brown, N. A., 599
Brown, R. U., 648, 662
Brown, Robert, 175, 554
Brown, W. A., 1115
Brown, W. G., 471
Browne, W. W., 1118
Bruce, A. N., 928, 949
Brüche, E., 154, 234
Brücke, 894
Brues, A. M., 598
Brühl, R., 1078
Brunauer, 97, 99, 443
Brunswig, H., 252
Bryan, 1012
Bryan, W. R., 430, 433
Buchanan, K. S., 701, 702, 706
Buchbinder, L., 1115, 1117
Bucher, 518
Büchner, 860
Buchner, H., 930, 949
Buchtal, F., 893, 895, 899
Buck, J. B., 848
Buckley, 450, 453, 455, 456
Bücks, K., 371
Budgett, H. M., 4
Buerger, U. J., 142, 145, 276
Buffon, 593
Bugbee, 750, 752
Bull, H. B., 373
Bull, H. H., 1119, 1122, 1128, 1135
Bullock, 1019
Bullock, E. R., 473, 511, 512
Bullock, L. T., 978
Bullowa, J. G. M., 12, 569
Bülow, U., 1121, 1135
Bumke, 1152, 1154
Bunge, 707
Bungenberg de Jong, H. G., 8, 12 1124, 1125, 1135
Bunn, 449, 454, 455, 456
Burankova, K., 518
Burch, C. R., 257, 264, 265, 266
Burden, S. S., 411
Burge, W. E., 599
Burger, F. J., 359, 360, 373
Burgin, C. J., 581, 599
Burk, D., 702, 704, 1001, 1049
Burk, R. E., 285
Burky, E. L., 1115
Burmaster, C. F., 955
Burns, 1008, 1014, 1024
Burr, H. S., 1150, 1152
Burrows, G., 265, 266
Burrows, H., 941, 942, 949, 1017, 1026
Burton, A. C., 772, 784
Burton, C. J., 174, 209, 210, 211, 234, 649, 662
Burton, E. F., 388, 410
Burwasser, F. G., 646, 663
Burwell, J. T., 326
Bury, C. R., 115, 120
Busch, H., 154, 155, 234
Busse, W. F., 326
Butendijk, F., 766, 784
Butler, U. R., 637, 662
Butler, R. E., 728, 738
Butschli, 867
Butt, C., 434
Butt, H. R., 734, 738
Butterworth, J., 629, 662
Butterworth, J. S., 1019, 1049
Buttolph, L. J., 1115
Brydowna, W., 1122, 1135
Byer, J., 692, 704

C

Calbrick, C. J., 154, 234
Caldenius, C., 524, 525
Caldwell, 682
Caldwell, K. S., 265
Calkins, Gary N., 599, 1173
Callan, H. G., 823, 848
Callier, 507
Callow, E. H., 860, 863
Calmette, A., 977
Calugarenu, 1152
Cambell, E., 785
Cameron, A. E., 511
Cameron, A. T., 748, 752
Cameron, F. K., 634, 662
Campbell, Dan H., 571, 599
Camps, F. E., 1116
Cannan, R. Keith, 389, 401, 410, 1049
Cannon, W. B., 745, 752, 984, 985, 986, 988, 989, 994
Carlson, A. J., 985, 1051
Carlson, J. G., 822, 848
Carman, 440
Carman, E. F., 101
Carothers, W. H., 49, 280, 283, 285, 663, 653, 662
Carpenter, H. C. H., 334, 337
Carrel, A., 1161
Cary, A., 85
Caspersson, 833, 836, 837, 838, 839, 840, 841, 843, 844, 845, 846, 847, 849
Castellani, 912
Castleman, R. A., 1115
Caston, 304, 308
Catcheside, D. G., 825, 828, 829, 848
Cauchoix, 1013
Cawood, W., 360, 372
Cekada, E. B., 954
Cerecedo, L. R., 687, 705
Cerovska, J., 371
Chadwick, James, 9
Chaffee, E., 409, 411
Chaikoff, J. L., 380, 1135
Chain, E., 926, 927, 928, 949, 1121, 1122, 1135
Chambers, L. A., 220, 235, 342, 343, 344, 347, 351, 370, 371, 807
Chambers, Robert, 864, 875, 882, 930, 949, 1151, 1152
Chamot, 654
Champy, 1014, 1021, 1033
Chandler, C. A., 959, 978
Chanutin, A., 955
Chao, 544
Chapin, C. V., 1100, 1115
Chapple, C. C., 1115
Chardonnet, Hilaire de, 624, 630, 662
Chargaff, E., 409, 410, 955, 957, 1120, 1128, 1129, 1130, 1135
Charles, A. F., 956
Charleston, V. D., 1117
Chase, A. M., 756, 760, 761
Chatterji, A. C., 514, 518
Chaussé, P., 1115
Chen, T., 978
Cheney, L. L., 1116
Cheng, L. H., 979
Cheng, V. C., 100
Cherches, K. A., 518
Chesley, F. G., 273
Chibnall, A. C., 535, 536, 537, 538, 544, 1120, 1134, 1135
Chick, H., 693, 704
Child, C. M., 853, 854, 855, 856, 859
Chiles, J. A., Jr., 419, 433, 434
Cholnoky, L. von, 471
Chope, H. D., 1100, 1115
Chow, B. F., 411, 434, 967, 977
Choy, F., 1096
Christian, W., 557, 600, 693, 694, 697, 707, 722, 737
Christiansen, 515, 518
Christensen, J. J., 1115
Church, C. F., 707
Churchill, E. D., 917, 949
Cimiotti, J. G., 1116
Clair, H. W., St., 342, 344, 345, 360, 371
Clamann, H. J., 886, 887, 899
Clapson, 308
Clapeyron, 84, 436
Clark, 440, 441, 443
Clark, A. J., 1096
Clark, C. R., 252
Clark, E. L., 930, 938, 941, 949
Clark, E. R., 930, 938, 941, 949
Clarke, F. H., 409, 410
Clark, G. L., 7, 49, 99, 100, 146, 147, 151, 632, 635, 1131, 1136, 1149, 1152, 1154
Clark, R. K., 387
Clarke, Beverly L., 457, 469, 470, 471
Claude, A., 429, 433, 836, 849, 1000, 1032, 1049
Claus, 1021
Claus, B., 352, 353, 354, 369
Claus, P. E., 428, 433
Cleveland, 597
Clifford, A. T., 634, 662
Cloudman, 1008
Clowes, G. H. A., 58, 1144, 1152
Clutton, R. F., 978
Cobb, S., 1145, 1152
Coblentz, W. W., 1115
Cochrane, 121, 130
Cochrane, W., 235
Cockcroft, 375, 380
Cocoa, 979, 985
Coghill, R. D., 409, 410
Cohausen, J. H., 736
Cohen, 441, 443
Cohen, A., 598
Cohen, L. H., 785
Cohen, S. S., 800, 801, 807, 843, 849, 955
Cohn, E. J., 409, 410, 955
Cohnheim, J., 919, 928, 929, 930, 949, 1041
Cole, A. G., 410
Cole, K. S., 867, 868, 869, 875, 1010, 1140, 1152
Colebrook, L., 1115
Coleman, A. P., 525
Collie, B., 112, 119
Collins, K. R., 978
Collip, J. B., 743, 752
Colvin, M. G., 1115
Committee on Apparatus, etc., 1115
Committee on Pollen Survey, 1115
Compton, A. H., 145
Compton, J., 641, 646, 648, 657, 661, 662, 663
Conant, J. B., 332, 337
Cone, W. M. II., 350, 371
Conklin, E. G., 592, 598
Conn, J. B., 977
Cook, A. H., 598
Cook, J. W., 590, 598, 1017, 1022
Cooper, 1021
Cooper, G. R., 433, 807
Copeland, 23, 31, 40, 59, 60, 67
Copeland, L. E., 102
Copley, A. L., 955
Coppen, F. M. V., 512
Corey, 233
Corey, R. B., 796, 808
Cori, 560, 680, 681, 682, 683, 1007
Correns, C., 808
Corrigan, K. E., 1149, 1154
Corson, D. R., 379
Corran, H. S., 699, 704
Correns, C., 620, 621, 644, 662
Costenan, 518
Cottrell, C. L., 895, 896, 900
Cottrell, F., 454, 456
Couette, 57, 354

Council on Foods, A. M. A., 719, 737
Cowan, S. L., 763, 784
Cowdry, E. V., 875, 1021
Cowgill, G. R., 706
Cowper, Wm., 1111
Cox, E. G., 145
Cox, Herald R., 569
Cox, R. P., 733, 738
Cradle Society report, 1115
Cragie, J., 807
Cralley, L. J., 1117
Cramer, 1013, 1019, 1021
Crile, G. W., 1149, 1152
de Crinis, M., 1138, 1152
Crittenden, E. D., 471
Croft-Hill, A., 598
Cromwell, H. W., 978
Crookes, Sir Wm., 9
Cross, C. F., 630, 662
Crowe, M. O'L., 467, 471
Crowell, M. F., 709, 717, 736
Crowfoot, C. M., 1135
Crowfoot, D., 272, 280, 535, 544
Crozier, W. J., 767, 769, 770, 771, 784
Cruickshank, R., 1115
Crum, W., 629, 663
Cullen, 444, 456
Culpepper, 614
Cunningham, E., 2, 12
Cunningham, J. T., 593, 596, 598
Curie, 445, 456
Curie, I., 377, 380
Curran, H. R., 1115
Curry, V., 850
Curtis, 1019
Curtis, H. J., 1140, 1152
Curtis, J. M., 740, 752
Cusick, P. L., 728, 738
Czapek, F., 630, 663

D

Daguerre, 472
Dahl, A. O., 1115
Dahlstrom, Roy, 100
Dahr, 514, 518
Dainty, M., 900
Dakin, H. D., 538, 544, 588, 598, 963, 978
Dale, Sir Henry H., 579, 745, 746, 752, 978, 981, 985
Dale, M. L., 1115
Dalla Valle, J. M., 1115
Dalton, John, 2
Danelli, J. F., 868, 875, 881, 882, 918, 949, 1096
Dam, H., 734, 738
Daneel, R., 820, 849
Dangers, H. W., 352, 371
Daniels, F., 602, 609
Daniewski, W., 371
Danilor, V. I., 371
Dankov, P. D., 493, 494, 512
Dann, W. J., 697, 704, 705
Danzer, M., 933, 949
Darling, S., 955
Darlington, C. D., 822, 823, 827, 830, 847, 849
Darrow, D. C., 1152
Dart, 308
Darwin, Charles, 584, 591, 593
Das, N. B., 705
Dauphine, A., 638, 663
Dauvillier, A., 146, 151
Davenport, C. B., 588, 598
Davey, Wheeler P., 122, 454, 456
Davidson, C. S., 956
Davies, A. W., 733, 738
Davies, D. G., 120
Davies, D. R., 1152
Davies, E. C. H., 49, 99
Davies, P. A., 332, 337
Davis, 308, 456
Davis, Alice R., 332
Davis, C. C., 286
Davis, D. O., 326
Davis, H., 1152
Davis, J. W., 551, 597
Davis, P. A., 1152
Davison, H., 1096
Davisson, C. J., 154, 234
Dawson, H. M., 569, 598, 599
Dean, H. Trendly, 583, 598
Dean, R. S., 360, 371
Debye, P., 359, 631, 663, 790
Deelman, 1026
Delafosse, 456
Delage, Y., 663
Delaunay, A., 933, 949
De Ganis, G. F., 594, 598
De Geer, Ebba Hult, 524, 526
De Geer, Gerard, 523, 524, 525, 526
Deinert, F., 598
Delay, J., 1152
Delbrück, M., 571, 599, 830, 842, 849
Demerec, M., 585, 598, 824, 825, 826, 832, 848, 849
Demann, W., 371
Deming, 443
Denis, W., 949
Dennis, E. W., 941, 949
Derjaguin, B., 252
Derksen, 682
Dersch, H., 512
Dervichian, 48
Desai, W. B., 518
Desch, 518
Desnuelle, P., 705
Desreux, V., 410
Detwiler, R. S., 757, 758, 761
Detwiler, S. B., Jr., 257, 265, 266
Deuel, H. J., Jr., 685, 705
Deuticke, H. T., 896, 900
Deutsch, W., 1126, 1136
De Vault, D. C., 609
Devaux, H., 876, 879, 880, 882
Devenish, E. A., 1115
Devine, J. W., 1116
Dewan, J. G., 705
De Witt, 443
De Witt, J. B., 702, 706
De Witt, T. W., 602, 609
Diamare, V., 1130, 1135
Dick, G. F., 1116
Dick, M., 950
Dickens, 1001
Dickinson, S., 899
Dickey, 518
Dickey, P. A., 4
Dickinson, 544
Dickson, 657
Dienes, L., 569, 598
Dietz, A., 939, 950
van Dijck, W. J. D., 266
Dingle, J. H., 959, 978
Dinsmore, R. P., 286
Dippel, L., 620, 663
Dittmar, C., 511
Dixon, H. M., 930, 949
Dixon, M., 724, 737
Dmochowski, 1004, 1005
Dobriner, 1003
Dobzhansky, Th., 585, 591, 598, 823, 824, 825, 849
Döderlein, 1026
Doerr, 979, 985
Dognon, A., 368, 370, 371
Doisy, E. A., 955
Dolland, 614
Donnan, F. G., 399, 415, 432, 480, 709, 748, 762, 904
Dore, W. H., 635, 666
Dorée, C., 630, 662
Dorfman, A., 696, 704, 706
Dorp, D. A. van, 686, 688, 707
Dörsing, K., 357, 363, 371
Doss, K. S. G., 120
Dosse, J., 156, 157, 158, 161, 162, 164, 175, 234
Dougherty, R. H., 1117
Douglas, S. R., 1115
Douglass, A. E., 519, 520, 522, 526
Dounce, A., 836, 837, 844, 849
Dow, R. B., 241, 252, 332, 337, 807
Doyle, M. E., 1117
Dragstedt, C. A., 923, 950, 979
Dragstedt, L. R., 1055, 1059
Drinker, C. K., 907, 916, 917, 941, 943, 949
Drinker, P., 1115
Duane, W., 145
Dubinin, N. P., 585, 598, 823, 847, 849
Du Bois, D., 956
Dubois-Reymond, 884
Dubuisson, M., 888, 889, 890, 896, 900
Dudley, H. W., 963, 978
Dudley, S. F., 1115
Duggar, B. M., 602, 609, 610, 807, 1115
Duke-Elder, W. S., 757, 761
Dumpert, 682
Duncan, J. T., 572, 574, 599
Dunn, L. C., 816
Dunham, G. C., 1115
Dunkin, G. W., 1115
Dunning, 1019
Dunnoyer, L., 265, 266
Durham, O. C., 1116
Duran-Reynals, F., 434, 928, 949, 1000, 1001, 1034, 1049
Durham, H. E., 938, 949
Durr, H., 512
Dushman, S., 265, 266
Dusser de Barenne, J. G., 1150, 1152, 1153
Dustin, A. P., 598
Duthie, E. S., 926, 927, 928, 949
Dworacek, E., 1134, 1135
Dye, W. B., 119
Dyer, Thistleton, 592
van Dyke, H. B., 411, 434

E

Eagle, H., 955, 965, 978
Eagleton, A. J., 1116
Eaken, R. E., 684, 705
Ebbecke, U., 920, 949
Eberhart, 450, 456
Ebner, von, 894, 1152
Ecker, F. E., 703, 704
Eckerson, S. H., 211, 234, 638, 641, 649, 650, 654, 660, 661, 663
Eckert, 456
Edgerton, H. E., 1116, 1117
Edlbacher, S., 692, 704, 836
Edsall, J. T., 894, 895, 898, 900
Edward, D. G., 1116, 1118
Eggersten, 456
Eggleston, L. V., 691, 705
Eggert, J., 507, 508, 512
Egloff, Gustav, 550
Ehrenfest, 24
Ehret, L., 369, 373
Ehrlenmeyer, H., 968, 978
Eichberg, J., 1126, 1135
Eichelberger, L., 1153
Eichoff, H. J., 602, 609
Eijkman, C., 683, 704
Eilers, 668, 682
Einstein, A., 236, 598, 1088, 1090
Eisclberg, 996
Eisenberg, P., 978
Eitel, W., 190, 195, 234
Ekwall, P., 567
Elam, C. F., 334, 337
Elford, W. J., 433, 434, 796, 1116
Elias, H., 1152
Ellermann, 999, 1031
Elliott, F. A., 484, 511
Elliott, K. A. C., 725, 737
Elliott, S. D., 1117
Ellis, C., 285, 1116
Ellis, R. V., 1115
Ellsworth, R., 743, 752

Elvehjem, C. A., 556, 598, 683, 684, 685, 688, 689, 691, 701, 704, 706, 717, 737
Elvove, E., 598
Embden, 688
Embree, N. D., 262, 265
Emerson, R., 557, 598, 602, 609
Emmett, P. H., 97, 99, 434, 443
Emmons, C. W., 830, 849
van der Ende, J., 1116, 1117, 1118
Endell, K., 193, 195
Endres, H., 598
Engel, R. W., 702, 704
Engel, W., 646, 664
Engelbreth-Holm, 1023
Engelhardt, 544
Engelhardt, W. A., 898, 900
Engelmann, 897, 900
Erb, J. H., 1126, 1135
Erikson, C. A., 1116
Erickson, B. N., 955, 956
Erikson, M. H., 1116
Eriksson-Quensel, I.-B., 411, 421, 433, 434, 562, 567, 791, 799, 807
Erlenmeyer, H., 467, 471
Ermolenka, N., 518
Ernst, E., 890, 891, 892, 899
Erxleben, 1002
Essex, H. E., 989, 994
Ettisch, G., 1152
Euler, H. von, 614, 688, 693, 694, 695, 704, 705
Evans, A. C., 939, 949
Evans, C. H., 512
Evans, F. R., 1115
Evans, H. M., 409, 410, 684, 705, 744, 752, 815, 919, 949
Evans, Robley D., 377, 1198, *et seq.*
Evans, R. M., 511
Ewald, A., 756, 760, 761
Ewart, A. J., 628, 663
Ewell, R., 252
Ewing, D. T., 100
Ewing, E. H., 100
Ewing, J., 1049
Eymers, J. G., 602, 609
Eyring, Henry, 236, 252, 271, 280, 473, 511
Eyster, W. E., 582, 598

F

Faber, H. K., 1116
Fabre, R., 1130, 1135
Fahey, K. R., 959, 978
Fähraeus, R., 567
Fair, G. M., 1118
Falin, 1014
Famulener, W., 977
Fankuchen, I., 790, 796, 805, 797, 807, 840, 848
Fano, U., 828, 829, 849
Faraday, M., 562
Farr, Wanda K., 10, 12, 211, 234, 610, 638, 641, 643, 648, 649, 650, 655, 657, 658, 660, 661
Farrar, 1013
Farris, L. P., 356, 372
Fawcett, E. W., 265, 266
Faxen, O. H., 412, 433
Fay, T., 1152
Fazekas, J. F., 785, 1153
Feigl, F., 469, 471
Feinberg, S. M., 1115
Fekete, 1012, 1013
Feldberg, W., 921, 949, 984, 985
Feldman, A., 81, 100
Fell, H. B., 875
Fell, N., 409, 410
Felton, L. D., 978
Felty, A. R., 1115
Fenn, W. O., 763, 784, 887, 898, 899
Féré, Charles, 1155, 1161
Ferguson, John H., 950, 951, 954, 955, 956
Ferguson, R. H., 279
Fermi, E., 9
Fernholz, E., 735, 738
Ferraro, A., 1152
Ferry, R. M., 952
Fessard, A., 760, 784
Fessenden, 343, 344
Fetzer, 679, 682
Feulgen, R., 837, 849
Fevold, 744, 752
Fibiger, 1019
Fick, 884
Fidiam, J. F., 196, 234
Field, G. S., 370
Field, M. E., 907, 916, 917, 943, 949
Fieser, Louis F., 589, 598, 734, 738, 1017
Figge, F. H. J., 705, 706
Fildes, P., 705
Filmer, J. F., 598
Finch, G. I., 130
Finkelstein, H., 433, 434
Finland, M., 1116
Finn, S. R., 1115, 1116, 1118
Firket, 996
Fischer, 26, 607
Fischer, A., 956
Fischer, E., 892, 896, 899, 900
Fischer, Earl K., 101
Fischer, Emil, 546, 550, 598
Fischer, F. G., 705
Fischer, R., 603, 609, 610, 619, 631, 635, 636
Fischer, M. H., 1124, 1135, 1142
Fischer, O., 1152
Fischgold, H., 1121, 1122, 1135
Fisher, 750, 752
Fisher, E. A., 1117
Fisher, R. B., 685, 705
Fishler, M. C., 1135
Fizeau, 895
Flatters, A., 628, 663
Fleischer, 999
Fletcher, 864
Fletcher, E. S., 370
Flinn, F. B., 1117
Florkin, M., 956
Florence, R. T., 67, 101
Flory, P. J., 252, 283, 285, 291, 292, 293, 306
Flosdorf, E. W., 342, 371
Flugge, C., 1100, 1116
Foerster, 1152
Folch, J., 1120, 1135
Folch-Pi, J., 953, 956
Foley, G. E., 1118
Folin, O., 949, 988, 994
Food and Nutrition Board, 716, 728, 737
Foote, 456
Ford, T. F., 196, 234
Forster, F., 371
Foster, 683
Foster, F. I., 747, 752
Fothergill, L. D., 959, 978
Fort, M., 630, 663
Foucault, 388, 391, 895
Fourcray, 444, 456
Fourier, 132
Fourt, L., 53, 101, 978
Fourt, P. C., 101
Fowkes, F. M., 101
Fowler, 95
Fowler, R. H., 252, 978
Fox, A. L., 2, 12
Fracastorius, 1100
Fraenkel, S., 1138, 1152
Frampton, V. L., 791, 792, 794, 807, 1122, 1135
France, Wesley G., 443, 456, 496, 968, 1082
Francis, F., 279
Francis, T., 978
Francis, T., Jr., 1116
Franck, J., 603, 609
Frank, J. A., 852, 859
Franke, W., 1152
François, 782
Fraser, H. F., 696, 706
Fraunhofer, 166, 614
Fredericq, Henri, 579
Freed, H., 1141, 1154
Freed, S. C., 924, 949
Freeman, W., 1145, 1152
Fremy, E., 610, 617, 618, 619, 620, 650, 660, 663
French, 674, 676, 677, 681, 682
French, C. S., 602, 609
Frenkel, 479
Frenkel, Y., 293, 371
Frenkel-Tissot, H. C., 1152
Frenzel, H., 371
Fresnel, 166
Freund, E. H., 636, 653, 663
Freund, H., 356, 357, 360, 361, 363, 371
Freud, B. B., 100
Freud, S., 1157, 1161
Freudenberg, 674, 675, 668, 681, 682
Freudenberg, K., 285
Freundlich, H., 340, 341, 342, 345, 352, 354, 355, 367, 369, 371, 448, 459, 471, 484, 631, 638, 663, 788, 872, 972, 978, 1125, 1127, 1135, 1148, 1152
Frevel, L. K., 122, 130
Frey, A., 631, 632, 634, 662
Frey-Wyssling, A., 636, 649, 654, 661, 663, 674, 682, 840, 842, 849
Friedel, G., 276, 280, 309, 599
Friedewald, 1026
Friedgood, H. B., 744
Friedemann, T. E., 704
Fridericia, L. S., 732, 738
Friedman, I., 702, 705
Friedrich-Freksa, H., 807
Fries, E. F. B., 769, 784
Fricke, R., 518
Frisch, F., 1139, 1152
Fritz, 469
Frolowa, S. L., 831, 837, 849
Froman, D. K., 370
Frondel, 454, 455, 456
Frost, D. V., 693, 705
Frumkin, 42, 43
Fruton, J. S., 578
Fuenfgeld, E. D., 1152
Fuhr, I., 716, 737
Fuju, N., 1122, 1129, 1135
Fujimami, 1000, 1031
Fuld, 682
Fuller, 308, 679, 682
Fuller, C. S., 252
Fuller, H. C., 10, 12
Funk, C., 684, 693, 705, 706
Funk, I. C., 693, 705
Furbach, E., 370
Fürth, 544
Furth, J., 1002, 1010, 1019, 1020, 1024, 1049
Furth, von, 884, 896
Fuss, 614

G

Gabritchevsky, G., 930, 949
Gaddum, J. S., 922, 923, 949
Gaffron, H., 603, 609
Gagarin, J., 410
Gaidukov, N., 628, 663
Gaines, N., 342, 343, 347, 371
Galambos, R., 12
Galanin, M. D., 372
Gallay, 682
Galloway, I. A., 433
Galton, 345
Gamble, J. L., 709, 711, 736, 1152
Gans, David M., 100
Gantt, W. Horsley, 1154, 1161
Gard, S., 433, 807
Gardner, C. E., 1116
Gardner, Wm. U., 380, 597, 1013, 1015, 1038, 1049
Garden, N. B., 372

Gargarin, Jr., 410
Garner, M. E., 85
Garner, W. E., 499, 512
Gates, F. L., 1116
Gates, R. R., 823, 844, 849
Gaubert, 445, 449, 454, 456
Gaucher, L., 628, 663
Gauss, 89, 167, 295, 509
Gauthier, P. P., 1116
Gavin, G., 692, 702, 706, 725, 737
Gay, E. H., 831, 847, 849
Gay-Lussac, 610
Gehenio, P. M., 783, 785, 862, 863
Gehman, 438, 439, 443
Geiger, 377
Geis, J. R., 360, 371
Geitler, L., 822, 830, 831, 849
Gellhorn, E., 1152
Gemmill, C. L., 748, 752
Gengon, 569
Georgi, F., 1139, 1140, 1152
Gerard, R. W., 1152
Germer, Lester H., 6, 12, 120, 130
Gerngross, O., 233, 636, 663
Gersh, L., 843, 844, 849
Gershenson, 842
Geschickter, 1013
Gesell, R., 766, 784
Gettler, A. O., 720, 737
Gettner, H. H., 409, 410
Ghosh, B. N., 977, 978
Gibbs, 904
Gibbs, E. L., 766, 784, 1152, 1153
Gibbs, F. A., 766, 784, 1152, 1153
Gibbs, J. W., 79, 116, 444, 445, 454, 456
Gibbs, R. E., 279, 326
Gibbs, W. E., 1116
Gibson, R. B., 978, 1152
Gibson, W. H., 630, 663
Giddings, N. J., 337
Gilbert, 985
Gilbert, E. C., 100
Gildea, E. F., 1145, 1152, 1158, 1159, 1162
Gildemeister, M., 1152
Gilding, H. P., 921, 950
Giles, N. H., 829, 849
Gilliland, 443
Gillings, 342, 345, 354, 355, 371
Gilman, H., 609
Gilson, E., 623, 624, 663
Giral, F., 1135
Gizycki, 190
Glaser, O. C., 598
Glaser, W., 154, 156, 161, 162, 175, 234
Glasstone, S., 236, 252
Glazko, A. J., 956
Glazunov, 469
Gleisberg, W., 630, 663
Glenn, W. L., 941, 949
Gleria, J. di, 192
Gley, E., 752
Glick, D., 692, 705
Glick, P. A., 1116
Glover, J. A., 1116
Go, Y., 326
Goby, P., 146, 151
Godber, G. E., 1115
Goebel, W. F., 967, 977, 978
Goethe, W., 11
Goetz, A., 863
Goetz, S. Scott, 863
Gohr, H., 371
Goldblatt, 1056
Goldenberg, 679, 682
Goldenberg, N., 10, 12
Goldhahn, H., 467, 471
Goldschmidt, R., 739, 752
Goldschmidt, S., 921, 949
Goldschmidt, T., 820, 824, 849
Goldschmidt, V., 1072, 1082
Goldsmith, M., 663
Goldstein, 1022
Golgi, 1148
Golla, F. L., 1152
Goncharov, S. V., 646, 663
Gonzales, P., 978
Goodeve, C. F., 760, 761
Goodhart, R. S., 687, 705
Goodman, J. G., 684, 706
Goodner, K., 959, 960, 978
Goodpasture, E. W., 919, 949, 1054
Goodson, W. H., 1153
Goodwin, T. H., 145
Goodyear, 286
Goppelsroeder, F., 459, 471
Goranson, R. W., 572, 598
Gorbatschew, S. W., 360, 371
Gordon, A. H., 705
Gordon, U. H., 1116
Gordon, S., 36, 67, 68, 102
Gore, V., 518
Goreczky, L., 428, 433
Gotthardt, E., 190, 234
Gottlieb, 682
Gottschalk, H. V., 360, 371
Goudsmit, J., 687, 707, 721, 737
Gough, 287
Gotaas, H. B., 1118
Göthlin, G. F., 1131, 1135, 1152
Gould, B. S., 771, 772, 784, 1116
Gowen, J. W., 807, 831, 847, 849
van de Graaff, 376, 380
Gradis, H., 703, 704
Grafe, V., 630, 663
Grafton, E. H., 100
Graham, C. L., 478
Graham, Thomas, 458, 556, 561, 598, 611, 663
Gralén, N., 433, 434, 563, 567
Grand, C. G., 930, 949
Granick, S., 605, 609
Grant, F. C., 1152
Grant, R. T., 920, 921, 930, 949, 950
Gratia, A., 428, 433
Gravilescu, N., 685, 705
Gray, 457
Gray, F., 234
Gray, H., 986, 994
Green, 443, 1012
Green, A. A., 959, 978
Green, D. E., 548, 558, 684, 686, 699, 705, 746, 752
Green, H. N., 702, 705
Green, J. W., 648, 658, 661, 664
Green, L., 602, 609
Green, R. H., 223, 234, 807
Green, Robt. G., 569, 598
Greene, 998
Greene, D., 1115, 1116
Greenberg, B., 1115, 1116
Greenberg, D. M., 711, 736
Greenberg, L. D., 1096
Greenstein, 898, 900, 1003
Greenstein, J., 840, 849
Greenspan, L., 1115
Greep, R. O., 411, 434
Gregg, D. C., 977
Greninger, A. B., 335
Grew, N., 614, 665
Griese, A., 694, 707
Griffen, 831
Griffin, D. R., 12
Griffith, 450, 456, 569, 598
Grijn, 683
Grim, R. E., 192, 193, 234
Grindley, J., 120
Groen, J., 701, 706
Groetzinger, Gerhart, 101
Gross, 544
Gross, D., 567
Gross, S. T., 147, 151
Grossford, A., 950
Grossman, A. J., 120
Grossowicz, N., 705
Grote, 750, 752
Gruber, N., 687, 707
Gruetzmacher, J., 341, 342, 343, 354, 371
Grün, A., 1123, 1135
Guernsey, E. W., 1117
Guffits, J. J., 703, 704
Guggisberg, H., 807
Guilbert, H. R., 733, 738
Guillain, J., 1152
Guilliermond, A., 638, 663
Gulick, A., 836, 849
Gunderman, J., 120
Günther, G., 699, 704
Gurney, R. W., 477, 478, 489, 492, 511
Guth, Eugene, 286, 289, 293, 304, 306, 307, 308, 636, 663
Guthowsky, U., 619, 624, 663
Gutman, A. B., 410, 560
Gutman, E. B., 410, 560
Gye, 1000
Guyer, 1021
Guyer, M. F., 428, 433
Guzman, J. de, 240, 252
Guyot, 42
György, P., 684, 705, 729, 738

H

Haas, G., 186, 234
Haas, E., 699, 705
Haas, V. H., 587
Haber, Fritz, 456, 550, 552, 600
Hadidian, Z., 776, 777, 778, 780, 784, 785
Hadley, Philip B., 598
Hafner, E. A., 1126, 1135
Hägg, G., 326
Hague, B., 1152
Haig, C., 733, 738, 756, 761
Hakazawa, F., 917, 949
Halber, 1004
Haldane, J. S., 992, 994, 1083, 1091, 1092, 1093
Haldane, J. B. S., 820, 832, 849
Haldi, H., 1142, 1145, 1152
Hale, H. P., 875
Hall, 1002
Hall, C. E., 207, 230, 231, 234, 235, 489, 490, 511
Hall, J. L., 409, 410
Hall, R. P., 1166, 1173
Halliburton, W. D., 1153
Halliday, N., 702, 705
Halpert, A., 950
Ham, T. H., 956
Hamburger, 473
Hamilton, A., 1116
Hamilton, Jos. G., 373, 377, 379, 380
Hammersten, E., 836, 842, 843, 849, 850
Hammarsten, O., 956
Hammett, 1022
Hanan, 996
Hanawalt, J. D., 122, 130
Hance, R. T., 186
Handler, P., 697, 704
Handovsky, H., 1123, 1126, 1135, 1143, 1149, 1152, 1153
Hanes, C. S., 560, 598, 667, 674, 675, 680, 681, 682, 683
Hansen, 626
Hansmann, G. H., 713, 737
Hanson, A. U., 743, 752
Hanson, W. T., Jr., 511
Hansteen-Cranner, B., 630, 634, 664, 1133, 1135
Hanusek, T. F., 628, 664
Happ, 496
Harden, 560
Harding, 29
Hardy, Sir Wm. B., 13, 15, 252, 485, 511
Hare, 1022
Hare, R., 1116
Haring, I. M., 648, 661, 666
Harington, C. R., 718, 737, 978
Harisaki, Y., 372
Harker, 535

Harker, D., 136, 145, 187, 216, 217, 218, 219, 235
Harkins, Henry N., 100
Harkins, Wm. D., 12, 15, 23, 31, 36, 38, 39, 40, 43, 49, 51, 58, 59, 60, 62, 67, 68, 81, 83, 85, 88, 97, 99, 101, 102, 244, 485, 511, 572, 978
Harper, H. A., 685, 705
Harpuder, K., 692, 704
Harrelson, R. T., Jr., 702, 706
Harrer, C. J., 703, 705
Harries, E. H. R., 1116
Harris, B. L., 443
Harris, L. J., 685, 704, 705, 793
Harris, U., 649, 658, 661, 667, 899
Harris, U. U., 1138, 1153
Harris, P. L., 733, 738
Harris, T., 955, 978
Harris, T. N., 1116
Harrison, H. E., 1153
Harrison, Ross B., 573, 576, 598
Harrison, R. W., 1129, 1135
Harrison, W., 629, 630, 664
Harrow, B., 956
Hart, E. B., 598, 664, 702, 717, 737
Hart, G., 1116
Hartig, T., 664
Harting, P., 611, 664
Hartley, G. S., 107, 112, 115, 119, 120, 486
Hartman, 518
Hartmann, 345, 346, 371
Hartmann, H., 892, 899
Hartree, W., 884, 885, 886, 887, 895, 899
Hartung, M. C., 863
Harvey, E. Newton, 4, 12, 347, 369, 371, 428, 433, 868, 875, 877, 881, 882
Hastings, A. B., 701, 706, 708, 713, 736, 775, 779, 785, 956, 1153
Hassid, W. Z., 609, 681, 683
Hatch, T. F., 1115, 1116
Hatschek, Emil, 514, 880
Hattori, K., 1121, 1135
Haurowitz, F., 555, 598, 961, 962, 978
Hauser, E. A., 120, 663
Hausser, I., 1122, 1134, 1135
Haüy, 443, 444, 456
Havas, L., 598
Haver, F. E., Jr., 641, 663
Havighurst, R. J., 145
Hawes, R. C., 1116
Hawkins, 1001
Haworth, W. N., 285, 635, 637, 664, 681, 683
Hayasaka, E., 685, 705
Hayden, H., 1149, 1154
Heath, 683
Heath, C. W., 718, 737
Hearne-Creech, 1022
Heatley, N. G., 598
Hecht, Selig, 443, 753, 756, 758, 760, 761
Hedenius, A., 434, 563, 567
Hedges, E. S., 513, 518
Heegaard, E., 687, 705
Heering, 190, 206, 235
Hehre, E. J., 570, 598
Heidelberger, M., 433, 567, 569, 841, 849, 961, 963, 964, 974, 976, 977, 978
Heidemann, E., 356, 357, 360, 363, 370, 371
Heidenhain, R., 886, 899
Heidenreich, R. D., 170, 186, 216, 217, 234
Heilbrunn, L. V., 852, 859, 871, 874, 875, 1021
Heim, F., 1153
Heiman, 1013
Heiss, J. H., 252
Heitz, E., 844, 849
Heitz, E., 822, 823, 849
Heiwinkel, H., 705
Hektoen, L., 410, 963
Heller, H., 750
Helmholtz, 79, 472, 884, 885
Hemsbach, H. Meckel von, 1064, 1065, 1082
Henderson, V. E., 1135, 1138, 1153
Hendricks, S. B., 139, 145, 193, 195, 234
Henle, W., 807, 1116, 1118
Henle, Jacob, 1067
Hengstenberg, J., 211, 635, 664
Henkes, R. A., 326
Hennessy, D. J., 687, 705
Henny, G. C., 1132, 1136, 1140, 1154
Henriot, E., 166, 427, 433
Henry, D. C., 410
Henschen, G. E., 409, 411
Henze, M., 598
Heptner, M. A., 847, 849
Herbert, D., 686, 705
Herbst, C., 575
Herchel, J. F. W., 611, 664
Herlitzka, A., 603, 609
Hermann, 233, 278, 280
Hermann, K., 636, 663, 664
Hermance, H. W., 469, 470, 471
Hermans, J. J., 359, 360, 365, 366, 371
Herrick, J. F., 989, 994
Herriott, R. M., 410
Hersey, U. D., 252
Hertwig, O., 864
Hertz, S., 1153
Herzog, A., 628, 629, 630, 664
Herzog, Maximilian, 999
Herzog, R. O., 416, 433, 631, 634, 649, 664, 794, 795, 807, 1153
Hess, Kurt, 115, 120, 634, 646, 648, 649, 661, 664
Hesselvik, L., 409, 410
Hester, J. B., 598
Hesz, K., 234
Hettich, A., 143, 145
Heuser, E., 285, 648, 649, 658, 661, 664
Hevesy, G., 376, 377, 380
Hewett, 1017
Hewett, C. L., 590, 598
Heycock, 146, 151
Heymann, B., 1116
Heymann, E., 354, 371
Heyn, A. N. J., 638, 649, 662, 655
Heyroth, F. F., 1116
Hibbert, H., 285
Hickman, K. C. D., 253, 265, 266, 733, 738
Hieger, I., 590, 598, 1017, 1024, 1025, 1049
Higginbotham, 682
Higson, G. I., 507, 512
Hildebrand, J. H., 271, 280, 1099
Hildesheimer, A., 686, 706
Hill, A. Croft, 546
Hill, Sir A. V., 784, 884, 885, 886, 887, 895, 899
Hill, D. K., 884, 899
Hill, L., 1115
Hill, R., 604, 609
Hill, S. E., 785
Hillier, J., 130, 169, 170, 172, 174, 175, 176, 182, 186, 213, 214, 235
Himwich, H. E., 785, 1153
Hinshelwood, C. N., 512, 768, 785
Hippke, E., 1116
Hippocrates, 1100
Hirshfelder, A. D., 919, 951, 1144, 1153
Hirst, E. L., 10, 12, 682
Hirszfeld, 1004
Hisaw, 744, 752
Hitchcock, A. E., 598
Hitchcock, C. S., 326
Hixon, R. M., 667, 682, 683
Hoagland, Hudson, 762, 776, 777, 778, 780, 783, 1161
Hoagland, D. R., 763, 770, 772, 784, 785
Hobart, G., 12, 1152
Hock, C. W., 658, 664
Hodapp, E. L., 783, 784, 785, 863
Hodgson, M. B., 488, 512
Hoeber, R., 867, 875, 1153
Hoerr, G. W., 119
Hofbauer, 1014
Hofer, J. W., 706, 721, 737
Hoff, H., 1153
Hoff, van't, 104, 473, 476
Hoffman, E. J., 86, 119
Hoffmann, F., 1128, 1135
Hofmann, K., 729, 738
Hofman, R., 511
Hofmann, U., 193, 195
Hofmeister, F., 714, 1124
Hogan, 740, 752
Högberg, B., 705
Hogg, B. M., 1135
Hogness, T. R., 699, 705, 781
Holker, J., 518
Hollaender, A., 830, 849, 1115, 1116
Hollaender, L. L., 1117
Holm, E., 732, 738
Holmes, H. N., 512, 740, 752
Hooker, M. O., 1124, 1135
Holt, 298, 308
Holt, E., 956
Holtfreter, J., 576, 599
Holz, A. M., 835, 849
Hoog, H., 326
Hooke, R., 296, 612, 613, 614, 624, 659, 664
Hooker, C. W., 433
Hooker, S. B., 963, 964, 965, 975, 977, 978
Hoover, M. E., 824, 849
Hoover, S. R., 702, 704
Hopf, 328, 337
Hopkins, 926
Hopkins, S. J., 978
Hopwood, 456
Horecker, B. L., 699, 705
Horning, 1013
Horowitz, W. H., 687, 705
Horsfall, F. L., 959, 960, 978
Horton, J. W., 1150, 1153
Horsfall, F. L., 404, 411, 432, 433
Horwitt, W. K., 544, 696, 704, 956
Horwood, U. P., 1116
Hosmer, H. R., 1149, 1152
Hosoya, Y., 759, 761
Hotchkiss, R. D., 967, 978
Houck, R. C., 511
Hough, W. H., 1153
Houston, J., 687, 705
Houssay, B. A., 744, 748, 752, 989, 994
How, A. E., 975, 978
Howatt, D. D., 265, 266
Howe, P. E., 1153
Howell, O. R. 107, 119
Howell, W. H., 952, 953, 956
Howells, T. A., 658, 661, 664
Hoxton, L. G., 419, 433
Høyrup, M., 403, 411
Hoyle, 1004
Huber, C., 693, 705
Huckel, 790
Hudack, S. S., 918, 941, 945, 949
Huddleson, F., 1116
Huddleson, I. F., 976, 978
Hudson, H., 487, 511, 682
Huelschmann, H., 692, 705
Hueper, 1019
Huffman, H. M., 326
Huggins, 1013
Huggins, Charles, 590
Huggins, Maurice L., 95, 131,

142, 145, 233, 271, 280, 308, 326, 487, 497, 498, 499, 500, 502, 510, 512
Hughes, 518
Hughes, A. H., 62, 66
Hughes, E. H., 733, 738
Hughes, T. P., 432, 433
Hughes, W. L., 410
Huguenard, E., 427, 433
Hull, T. G., 1116
Hull, T. Z., 709, 736
Humbert, R. P., 195, 234
Humphery, E. R., 99
Humphrey, G. C., 705
Hungerbühler, J., 620, 664
Hunter, J., 1154, 1161
Huntington, 588
Huntington, Ellsworth, 522
Hurtley, W. H., 265
Husemann, 668, 682
Huskins, C. L., 822, 849
Husted, C., 1153
Hutchens, J. O., 883
Huxley, J. S., 739, 752
Huxley, Thos., 11

I

Ichikawa, 589, 600
Illig, R., 416, 433, 807
Inawashiro, R., 685, 705
Ingen-Housz, J., 625, 664
Ingram, 750, 752
Inman, O. L., 605, 609, 610
Inugami, K., 924, 950
Inutsuka, N., 929, 949
Ipatuff, U. N., 598
Isbell, H., 728, 738
Isherwood, 682
Ishikawa, 1017, 1029
Issayeff, 941, 949
Itallie, van, 682
Iterson, G. van, Jr., 646, 664
Ives, H. E., 512
Ivy, 1054
Iwanowski, Dm., 785, 807

J

Jabbynski, 513, 515, 516, 518
Jaburek, L., 1153
Jacobi, 682
Jacobi, A., 12
Jacobs, J., 978
Jacobs, U. H., 875
Jacobs, R. B., 331, 337
Jacobsohn, K., 789, 808
Jacobson, 581
Jacobson, L. O., 1116
Jaeger, 837
Jakus, U. A., 230, 231, 235
James, H. M., 286, 289, 304, 306, 307, 308
James, T. H., 473, 482, 483, 484, 485, 486, 502, 503, 510, 511, 512
Jameson, E., 409, 410
Jancke, W., 664, 1153
Janssen, Hans, 613
Jaques, L. B., 954, 956
Jassen, L. W., 807
Janssen, Zacharias, 613
Jassens, F. A., 826, 849
Jasper, H. H., 1153
Jausseran, G., 512
Jeans, 751
Jeans, P. C., 735, 738
Jeffries, Z., 335, 337
Jeghers, H., 756, 762
Jelley, E. E., 492, 512
Jennings, Herbert S., 574, 575, 594, 598, 1048, 1052, 1162, 1163, 1164, 1165, 1167, 1168
Jennings, L., 410
Jennison, U. W., 1099, 1116, 1117
Jenrette, W., 849
Jensen, 544
Jensen, C. O., 996, 1030, 1049
Jerchel, D., 570, 599
Jessop, G., 72
Jewell, W., 265, 266
Jirkovsky, R., 469, 471
Jobling, J. W., 956, 1000, 1032
Jochims, J., 1152
Johannson, H., 154, 234
Johnson, 119
Johnson, C. H., 373
Johnson, D. S., 623, 624, 664
Johnson, E., 728, 737
Johnson, L. R., 978
Johnson, M. L., 757, 761
Johnson, V., 1051
Johnson, W. A., 725, 737
Johnston, James F. W., 558, 598
Johnston, S. A., 104
Johnstone, O. P., 943, 951
Joliot, F., 377, 380
Jollos, V., 595, 598, 1148
Joly, 1022
Jones, 544
Jones, F. E., 120
Jones, G., 1153
Jones, H. B., 380
Jones, J., 1153
Jones, T. D., 921, 949, 1118
Jong, de, 142
Jordan, Hubert F., 100
Jorpes, J. E., 956
Joseph, R. C., 1153
Jost, W., 511, 512
Joule, 287, 308
Joyner, 658
Juday, C., 602, 609
Jukes, T. H., 956, 1122, 1123, 1133, 1135
Juliusburger, F., 366, 371
Junk, W., 1117
Jura, George, 97, 102
Jurin, 458
Just, E. E., 865

K

Kabat, E. A., 409, 410, 411, 567, 917, 960, 978, 979, 1049
Kabsch, W., 620, 664
Kadish, M. A., 922, 924, 927, 937, 950
Kahlbaum, 335
Kahler, 1012
Kalckar, H. M., 884, 899, 900
Kamen, U. D., 434, 609
Kamm, Oliver, 739, 740, 744, 747, 750, 752
Kamp, F., 896, 900
Kapitza, P., 334, 337
Kaplan, U. H., 956
Kapnick, I. R., 959, 978
Karrer, E., 326
Karrer, Joanne L., 807
Karström, H., 561, 598
Kassell, 544
Katz, 984, 985
Katz, J. R., 636, 674, 675, 682, 1153
Katz, S. E., 1138, 1153
Katzenelbogen, E., 705
Kaufmann, B. P., 823, 828, 829, 849
Kauffmann, G., 978
Kausche, G. A., 174, 188, 226, 234, 797, 798, 807
Kautsky, H., 603, 608, 610
Kauzmann, W., 252, 271, 280
Kearney, E. B., 235
Keeler, 599
Keeser, E., 1121, 1126, 1127, 1129, 1135
Keevil, N. L., 1116
Kehler, H., 186, 234
Keilin, D., 557, 563, 599
Keitt, G. W., 1116
Kekwick, R. A., 409, 410, 433, 567, 803, 804, 807
Kellaway, 983, 985
Keller, R., 1148, 1153
Kellogg, W. H., 1116
Kelly, E., 683, 706
Kelley, O. J., 203, 235
Kelvin, Lord, 287
Kemp, I., 1122, 1135
Kendall, E. C., 718, 737, 752
Kendall, F. E., 976, 977, 978
Kennaway, E. L., 1017, 1049, 1050
Kendall, James, 468, 471
Kenny, A., 1115
Kenrick, 42
Kensler, C. J., 1050
Kerejci, L. E., 434
Kerr, T., 648, 654, 661, 665, 669, 675, 676, 677, 678, 682
Kerst, D. W., 380, 387
Kersten, H., 373
Keschischian, K. H., 1116
Kettering, C. F., 610
Keys, A., 915, 916
Kharamonenko, 518
Kharasch, U. S., 610
Khariton, Y., 370
Khayyám, Omar, 597
Khvostova, V. V., 823, 850
Khouvine, V., 636, 664
Kidd, J. G., 1004, 1005, 1026, 1028, 1029, 1031, 1050
Kidder, G. W., 1166, 1173
Kiesel, 580
Kiessig, H., 646, 664
Kik, U. C., 701, 702, 706
Kimura, O., 354, 371
King, A. M., 85
King, C. G., 703, 705
King, H. H., 100
King, L. V., 359, 372
King, R. L., 428, 433
Kinnersley, H. W., 686, 705, 721, 737
Kinosita, 1003
Kirchof, M., 610, 664
Kirschbaum, A., 433
Kirchner, F., 130
Kirkwood, J. G., 48, 51, 77, 101
Kirseck, 190
Kistiakowsky, G. B., 977
Kitchen, D. K., 739
Klassen, K. P., 740, 752
Klebs, G. A., 620, 664
Klein, H., 711, 736
Klein, J. R., 696, 697, 701, 705
Klein, V., 372
Kleinberger, E., 569
Klemperer, O., 154, 235
Kleinzeller, A., 900
Klenk, E., 1120, 1121, 1122, 1134, 1135, 1136
Kligler, I. J., 705
Klose, H., 1153
Klug, H. P., 326
Knapp, E., 830, 849
Knappeis, G. G., 893, 895, 899
Knight, B. C. J. C., 705
Knight, C. A., 443, 594, 599, 798, 807, 848, 849
Knoll, M., 156, 175, 235
Knorr, H. V., 604, 609, 610
Knox, W. E., 686, 699, 705
Knudson, M., 265, 266
Koch, 190, 206, 235
Koch, E., 496
Koch, H. W., 387
Koch, M. B., 706
Koch, M. L., 1153
Koch, W., 1124, 1125, 1135, 1153
Kodani, M., 849
Koeniger, H., 1116
Kögl, Fritz, 597, 729, 738, 1002
Kohlhass, R., 279
Kohlransch, F., 410
Kohlschütter, V., 496, 512
Kohn, H. I., 602, 609, 696, 701, 705
Kolle, 985
Koller, L. R., 1116
Kolmer, 757
Kolnitz, H. von, 718, 737
Kolthoff, 456

Koltzoff, N. K., 572, 599, 831, 840, 849
Kon, S. K., 687, 705
König, W., 356, 363, 372
Kopak, M. J., 875, 877, 882, 883
Kornfeld, G., 510, 511, 512
Kornmueller, A. E., 1153
Koser, S. A., 696, 704, 706
Kossel, A., 453, 454, 456, 836, 839, 849
Kossel, H., 950
Kossikov, K. V., 823, 850
Kostytschew, S., 608, 610
Krabbe, G., 620, 664
Kraemer, Elmer O., 282, 285, 412, 416, 432, 433, 434, 471, 682
Krafft, F., 102, 119, 265
Krahl, U. E., 883
Kraissl, C. J., 1116
Krampitz, L. O., 691, 705
Krasnikow, A., 326
Krasser, F., 620, 664
Kratky, O., 633, 636, 649, 653, 664, 665
Krausche, G. A., 229, 234
Kraus, 1082
Kraus, F., 961, 978, 982, 985
Krause, B., 1139, 1153
Krause, F., 220, 234, 235
Krchma, L. C., 955
Krebs, H. A., 691, 705, 725, 737
Kregel, E. A., 210, 235, 649, 666
Krejci, L. E., 411, 567
Kremnev, L., 347, 350, 372
Kretschmer, M., 209, 210, 235, 649, 666
Kristnamurti, K., 120
Krogh, A., 901, 902, 904, 910, 914, 915, 916, 919, 920, 921, 928, 949
Kruetzkamp, J., 1116
Krüger, F., 367, 372, 1153
Krumm, J., 182, 234
Kruppke, E., 169, 235
Kruyt, H. R., 8, 12, 1156, 1161
Kuan-Han Sun, 308, 326
Kubo, 1057
Kubowitz, F., 721, 737
Kudar, H., 416, 433, 807
Kuhn, R., 570, 599, 692, 698, 705, 1122, 1135
Kuhn, W., 794, 796, 807
Kühne, W., 753, 755, 756, 758, 759, 760, 761
Kundt, A., 347, 356, 372, 1088
Kunifusa, J., 978
Kunitz, 843
Kunitz, M., 954
Küntzel, 233
Kusnetowsky, N., 920, 949
Küster, A., 507, 508, 512
Kuster, E., 518
Kúthy, A. von, 759, 762, 1129, 1136
Kützing, 622, 665
Kuwada, Y., 822, 849
Kyes, P., 978

L

Laass, 449, 457
Labrouste, 23
Lacassagne, A., 377, 1007, 1015, 1020, 1023, 1050
Lackman, D. B., 220, 235
La Cour, L., 823, 849
Lagsdin, J. B., 430, 434
Laidlaw, P. P., 1115
Laidler, K., 236, 252
Laidler, T. J., 372
Laing, M. E., 104, 114, 119, 120
Laki, K., 705
Lamarck, J. B., 592, 593, 1175
Lamarque, P., 146, 151
Lambert, R. H., 511
Lambrecht, R., 610
LaMer, V. K., 326
Lamm, O., 395, 410, 412, 416, 421, 433, 567, 795, 803, 804
Lampitt, L. H., 10, 12, 679, 682
Landis, Eugene M., 900, 916, 917, 918, 919, 929, 949
Landis, Q., 433
Landsteiner, K., 571, 957, 958, 963, 964, 965, 967, 978, 1025, 1131, 1135
Landström-Hydén, H., 844, 845, 849
Lane, C. T., 1150, 1152
Lane, R. L., 728, 738
Lange, B., 1116
Langenbeck, W., 689
Langdon-Brown, W., 744, 752
de Langhe, J. E., 511
Langmuir, I., 15, 17, 27, 37, 62, 77, 87, 98, 257, 265, 373, 380, 445, 448, 456, 473, 485, 511, 579, 790, 791, 796, 807, 876, 880, 883, 969, 971, 972, 973
Lankester, E. R., 576, 599
Lansing, W. D., 433
Lantermann, 1148
Larsen, Esper S., 335, 337
Larmor, 374
La Rosa, W., 410
Lasareff, P., 608, 610
Laselle, P. A., 12
Laschtschenko, 1116
Lash, 456
Laskowski, 1004
Lasnitzki, 1003
Lathrop, Abbie, 1006, 1050
Lathrop, A. E. C., 1050
Latimer, W. M., 95
La Tour, F. Dupré, 279
Lattes, 377, 380
Laue, M., 631, 665
Lauffer, M. A., 410, 433, 434, 785, 787, 788, 791, 792, 794, 795, 797, 798, 804, 807, 848
Laurens, H., 1116
Lauritsen, 376
Lavedan, 1014
Lavin, G. I., 807
Lawrence, A. S. C., 115, 120, 276, 280, 898, 900
Lawrence, Ernest O., 373, 376, 377, 380
Lawrence, John H., 373, 379, 380
Lazarow, A., 574, 599
Lea, M. Carey, 335, 336, 337, 474
Lea, D. E., 828, 849
Leake, Chauncey D., 1094, 1097, 1138, 1153
Leakey, S. A., 526
Lease, J. G., 683, 706
Leathes, 62
Leathes, J. B., 1121, 1130, 1131, 1135
Leaders, W. U., 1198
Lebel, H., 956
Leber, T., 930, 933, 949
Ledingham, J. C. G., 978
Leete, H. M., 1116
Leeuwenhoek, A. von, 614, 665
Legge, J. W., 703, 705
Lehmann, 347
Lehmann, F. E., 599
Lehmann, J. F., 370
Lehmann, O., 372, 1088, 1135
Lehmann-Facius, 1004
Leitch, A., 1026, 1050
Leitch, I., 715, 737
Leiter, J., 1050
Leitner, J., 1116
Lemberg, R. L., 599, 703, 705
Lembke, A., 223, 235
Lemon, H. F., 1117
Lemon, H. M., 1117
Leonard, 744, 752
Lenard, P., 169, 235
Lennette, E. H., 409, 410
Lennox, W. G., 1153
Lepeschkin, W. W., 867, 875
Lepkovsky, S., 684, 705
Lester, J. H., 628, 665
Leutscher, 409
Levine, M., 917
Levene, P. A., 434, 836, 849, 1120, 1122, 1135
Levin, L., 434
Levine, B. S., 629, 665, 978
Levine, S., 790, 791, 807
Levshin, V. L., 372
Lewis, 443, 981, 985, 1122, 1136
Lewis, C. M., 602, 609
Lewis, E. B., 825, 834, 849
Lewis, E. P., 356, 372
Lewis, G. N., 326
Lewis, J. H., 963, 978, 1159, 1162
Lewis, J. M., 733, 738
Lewis, P. A., 929, 949
Lewis, Sir Thomas, 901, 911, 915, 916, 919, 920, 949
Li, C. H., 409, 410
Li, C. P., 979
Lichotzky, 1153
Lichtwitz, L., 456, 1063, 1082
Lidén, R., 524
Lidwell, O. M., 1115
Lieb, H., 1135
Liebig, 610
Liesegang, R. E., 496, 505, 512, 513, 518, 1072, 1075, 1077, 1080, 1082, 1145, 1153
Lieser, 661
des Ligneris, M. J. A., 1000, 1022, 1024, 1025, 1026, 1049
Lillie, F. R., 852, 859
Lillie, R. S., 765, 785, 1153
Limpacher, R., 1123, 1135
Lin, K., 978
Lindahl, P. E., 598, 856, 859
Lindegren, C. C., 572
Lindegren, G., 599
Lindemann, R., 636, 663
Lindner, E., 924, 949
Lineweaver, 443
Ling, C. Y., 609
Link, H. F., 622, 628, 657, 665
Linke, F. W., 430, 431, 433
Linnert, K., 1138, 1152
Linquist, F. E., 430, 432
Lippincott, 1038
Lipmann, F., 609, 687, 688, 690, 691, 702, 705, 721, 737, 884, 888, 899
Lippmann, 207, 491
Lipschitz, W. A., 685, 688, 706
Lipschütz, 1013, 1026
Lipton, M. A., 688, 689, 691, 706
Lisbonne, M., 337
Little, C. C., 1008, 1009, 1011, 1012, 1013
Liu, S., 372
Livingston, 376
Livingston, P. C., 1096
Livizzari, 444, 456
Ljubimova, M. N., 544, 898, 900
Llewellyn, F. J., 145
Lloyd, 518
Lobashov, M. E., 842, 849
Locke, 1161, 1162
Lockwood, W. H., 703, 705
Lodge, Sir Oliver, 468, 469, 471
Loeb, Jacques, 638, 665, 1135
Loeb, Leo, 589, 995, 999, 1006, 1007, 1023, 1024, 1025, 1050
Loehwing, W. F., 580, 581, 590
Loewi, Otto, 578, 599, 745
Loewit, 981, 985
Lohmann, K., 686, 687, 688, 689, 706, 721, 737, 888, 896, 898, 899, 900
Lomax, R., 1132, 1135
Long, 1037, 1038
Long, A. P., 1116, 1117
Long, C. N. H., 945, 949
Long, F. A., 101
Longnecker, H. E., 692, 706

Longsworth, Lewis G., 387, 389, 395, 400, 410, 544, 937, 949
Loomis, A., 4, 12
Loomis, A. G., 196, 234
Loomis, A. L., 337, 340, 346, 351, 352, 363, 368, 370, 371, 373, 1152
Loos, H. O., 921, 949
Loosli, C. G., 1117
Lord, F. T., 938, 949
Lorentz, 155, 159
Loring, H. S., 801, 807, 808
Lottermoser, A., 105, 119, 486, 511, 628
Love, 443
Loveland, R. P., 511
Lovelock, J. E., 1115
Lowe, 682
Lowe, E. P., 1117
Lowell, James R., 592
Lowman, A., 467, 471
Lowry, O. H., 708, 736
Lu, G. D., 685, 706
Lubarsch, O., 1067, 1082
Lubimenko, V. N., 603, 609
Lubkin, V., 949
Lucas, C. C., 581, 599
Lucas, G. H., 1138, 1153
Lucké, B., 929, 950, 1033
Luckiesh, M., 1117
Lucretius, 6
Ludewig, S., 955
Ludford, R. J., 599
Ludlum, S. R., 1139, 1147, 1153, 1154
Lüdtke, Max, 661, 664
Luetscher, J. A., Jr., 409, 410
Lugg, 544
Lühdemann, R., 271, 280
Lullies, H., 1153
Lumière, A., 1051
Lumsden, 1004
Lundberg, H., 526
Lundgren, H. P., 409, 411, 426, 433, 562, 964, 977, 978
Lundsgaard, Einar, 888, 899
Lunt, L., 986, 994
Lüppo-Cramer, 482, 486, 487, 505, 511, 512
Luria, S. E., 223, 235
Lurie, M. B., 939, 941, 949, 1118
Lush, D., 1118
Lusk, G., 945, 949
Lüttgens, W., 721, 737
Lutton, E. S., 279
Luyet, B. J., 783, 784, 785, 859, 861, 862, 863
Lwoff, A., 696, 704
Lyman, C. M., 688, 690, 691, 704, 729, 738
Lynch, 1020, 1023
Lynch, Clara, 1008
Lyngbye, H. C., 614, 615, 621, 622, 628, 660, 665
Lyons, W. R., 409, 410
Lysholm, A., 416, 434
Lythgoe, R. J., 759, 761

M

Macallum, A. B., 707, 708, 736
MacArthur, 540, 541, 542, 544
MacArthur, J. W., 599
MacClement, W. D., 796, 807
MacCurdy, I. T., 919, 949
MacDonald, J. T., 512
MacDonald, K., 1117
MacDougall, D. T., 1153
MacDowell, E. E., 998, 1009, 1010
Macewan, D., 141, 145
Mach, E., 345, 372
Mache, H., 372
Macheboeuf, M. A., 337
Machemer, H., 635, 664
MacInnes, Duncan A., 387, 389, 400, 410, 544, 937, 949, 1150
MacKenzie, 1026
MacKenzie, K., 834, 849
MacKenzie, K. R., 379
MacLean, F. C., 956
Maclean, H., 1119, 1135
Maclean, I. S., 1119, 1135
MacLeod, J., 934, 950
MacNevin, 456
Macomber, M., 274, 275
Macrae, T. F., 693, 704
Macy, I. G., 714, 737
Madden, R. J., 696, 704
Maddox, R. L., 511
Magee, J. L., 602, 609
Magistris, H., 630, 663, 1130, 1135
Magne, H., 1115
Mahl, H., 164, 168, 183, 184, 185, 235
Mainzer, F., 956
Maisel, 740, 752
Major, C. B., 1117
Malkin, T., 1132, 1135
Maksimova, Ch., 373
Malisoff, W. M., 371
Malkiel, S., 976, 977, 978
Malkin, T., 279
Mallun, 1117
Malmberg, M., 693, 705
Malpighi, M., 614, 665
Man, E. B., 1135
Mann, F. C., 989, 994
Mann, P. J. G., 685, 697, 704, 706
Mann, S., 1153
Manchester, R. C., 1153
Mann, T., 557, 599
Mandelbaum, J., 758, 761
Manegold, E., 511
Mangin, L., 619, 620, 621, 624, 665
Mangun, G. H., 748, 752
Manning, W. M., 602, 609
Manwaring, W. H., 593, 599
Marangoni, 15
Marc, 445, 456
Marchal, Paul, 575
Marco Polo, 559
Margaria, R., 888, 890, 898
Marinesco, G., 1147, 1153
Marinesco, N., 349, 372
Mark, H., 211, 235, 280, 285, 308, 328, 337, 632, 633, 636, 653, 663, 664, 665
Mark, J. van der, 785
Markin, L., 978
Markham, R., 807
Markley, K. S., 257, 258, 265
Marquette, W., 659, 665
Marrack, J. R., 570, 599, 960, 965, 968, 978
Marriage, 518
Marsh, David F., 1094, 1097
Marshak Carlson, 828
Marshall, C., 1154
Marshall, C. F., 193, 197, 199, 204, 235
Martin, A. J. P., 693, 704
Martin, D. W., 1116
Martin, H. E., 119
Martin, R. A., 649, 667
Marton, L., 166, 169, 174, 175, 176, 192, 220, 235, 280
Marvel, C. S., 285
Marvin, H. M., 929, 949
Marwick, T. C., 646, 662
Marx, F., 961, 978
Masing, G., 369, 372
Masket, A. V., 433, 955
Mason, 518, 654
Mason, H. L., 728, 738
Mason, M., 412, 433
Massart, J., 950
Mast, S. O., 1166, 1173
Masterman, A. T., 1117
Matthews, A. P., 1117
Matthews, B. H. C., 12, 1152
Matthews, F. W., 861, 863
Matthews, J., 471
Matthews, J. E., 337
Matthews, J. Merrit, 630, 665
Mattill, H. A., 702, 705
Mattoon, R. W., 88, 101
Mautz, F. R., 748, 752
Maxwell, L. R., 130
May, 518
May, R. M., 1153
Mayer, A., 987, 994
Mayer, J. E., 512
Mayer, M., 977, 978
Mayneord, W. V., 1017, 1026, 1049
Maximow, A. A., 938, 950
Maxwell, Clerk, 896
Mazia, D., 837, 839, 840, 843, 849
McAdams, 443
McBain, J. W., 37, 102, 104, 114, 115, 119, 120, 270, 273, 277, 428, 433
McBain, M. E. L., 115, 119
McBurney, 456
McBurney, C. H., 729, 738
McCall, Robt., 630, 663
McCay, 1051, 1062
McCay, C. M., 709, 717, 736
McClellan, R. H., 919, 949
McClellan, V., 410
McClenahan, 682
McClendon, J. F., 1153
McClintock, 822, 823, 825, 827, 828, 833, 844, 846, 849, 850
McCollum, E. V., 685, 686, 705, 711, 719, 737
McCready, 681, 683
McCulloch, U. S., 599, 1153
McClure, 1060
McConnell, W. J., 1117
McCoy, G. W., 1117
McCulloch, W. St., 1150, 1152
McCutcheon, M., 929, 930, 949, 950
McFarlane, A. S., 410, 412, 427, 433, 803, 804, 807
McFarlan, R. L., 331, 337
McGath, T. B., 1118
McGlone, B., 921, 949
McIver, M. A., 988, 994
McIlhenry, E. W., 702, 706, 725, 737
McIlwain, H., 697, 706
McIntosh, J., 428, 432, 433
McKhann, C. F., 959, 978, 1115, 1116, 1117
McLaughlin, H. M., 100
McLean, F. C., 1125, 1135
McLean, J., 956
McKinley, J. B., 717, 737
McMaster, P. D., 918, 941, 945, 949
McMasters, 518
McMeekin, T. L., 410
McNally, J. G., 511
McNider, 1056
McQuarrie, I., 1140, 1141, 1153
McPherson, 298, 308
Meade, T. H., 265, 266
Meckel von Hemsbach, H., 1064, 1065, 1082
Mees, C. E. Kenneth, 511, 512
Mehl, J. W., 796, 807
Meidinger, W., 493, 512
Meischen, F., 839, 850
Meitina, R. A., 544, 900
Melchers, G., 799, 807
Melchionna, R. H., 1050
Meleney, F. L., 1116, 1117
Mellanby, J., 956
Mellar, J. W., 6
Mellors, R. C., 748, 752
Melnick, J. L., 409, 411, 689, 690, 706
Melroy, 1023
Meltzer, C., 929, 950
Meltzer, S. J., 712, 736, 929, 950, 1144, 1153
Melville, D. B., 729, 738
Melville, W. H., 285
Melzer, L., 705
Mendel, Gregor, 587, 808, 809, 816, 819
Mendel, Lafayette B., 599

Mendelssohn, K., 512
Mendhall, 518
Mensbrugghe, van der, 15
Menke, W., 604, 609, 848, 850
Menkin, M. F., 950
Menkin, V., 1096
Menkin, Valy, 917, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 943, 944, 945, 946, 948, 950
Meikeljohn, A. P., 685, 706
Merrill, R. C., 119, 120
Merritt, H. H., 785
Mesrobeanu, I., 977
Mestre, H., 605, 608, 609, 610
Mestrezat, 1147, 1153
Metchnikoff, E., 929, 950
Metz, C. W., 824, 831, 850
Meyen, F. J. F., 618, 665
Meyer, 570, 649, 650
Meyer, A., 625, 665
Meyer, A. W., 1153
Meyer, G., 511
Meyer, H. H., 1153
Meyer, K., 409, 411, 956, 978
Meyer, K. H., 115, 120, 271, 280, 285, 326, 328, 337, 886, 887, 888, 897, 899
Meyerhof, Otto, 560, 599, 688, 702, 713, 833, 884, 888, 890, 891, 892, 893, 896, 899, 900
Miall, M., 900
Miall, Stephen, 12, 599
Michaelis, 1158
Michalowsky, 1014
Mider, 1024
Miers, H. A., 6, 518
Miescher, 836
Mikhalev, 518
Milam, D. F., 1117
Miles, A. A., 433, 1115
Miller, 125
Miller, B. F., 1117, 1129, 1135
Miller, G. L., 411, 567, 799, 801, 807
Miller, H., 1116
Miller, R. F., 733, 738
Milligan, L. H., 354, 373
Mills, C. A., 952, 956
Mills, G. F., 120
Milstone, H., 956
Minaev, V. A., 629, 665
Minajeff, W., 629, 665
Minami, G., 924, 950
Minnibeck, 544
Minz, B., 704
Mirbel, de, 617
Mirsky, A. E., 563, 750, 836, 837, 838, 841, 846, 850
Mitchell, 62
Mittasch, Alwin, 600
Mittenzwei, H., 610
Moeller, 544
Moeller, T., 517
Moewus, F., 570, 599, 820, 850, 1166, 1170, 1171, 1173
Mohl, H. von, 616, 617, 618, 625, 650, 665
Möhle, W., 891, 893, 899
Molisch, H., 624, 630, 665
Molitor, H., 1153
Moir, H. C., 2, 12
Möllendorf, von, 938, 950
Monaghan, B. R., 1153
Mond, R., 1153
Moon, V., 933, 950
Mooney, 308
Mooney, R. C. L., 325, 326
Moore, D. H., 409, 410, 434, 955
Moore, J. H., 648, 662
Moore, M., 785
Moore, T., 733, 738
Moore, T. S., 95
Moore, R. A., 1050
Moore, W., 252
Morau, 996
Moravek, 518
Morawitz, 952
Morey, G. W., 326
Morgan, J. W., 62, 85, 100
Morgan, Sir Gilbert T., 552
Morgan, T. B., 1196, 1197
Morgan, Thos. H., 810, 811, 820, 821, 850
Moricard, 1013
Moriguchi, N., 372
Morris, 1038
Morozov, A. A., 372
Morrell, Jacques, 550
Morris, 438, 439, 443
Morrison, Philip, 386, 387
Morrow, M. B., 1117
Morse, 513, 515, 516, 518
Morse, J. K., 713, 736
Morton, 1024
Morton, G. A., 154, 225, 235
Mosenthal, H. de, 626, 627, 661, 665
Moser, F., 852, 859
Mott, F. W., 1153
Mott, N. F., 477, 478, 489, 492, 511
Mould, W. F., 1099
Moulton, S., 1117
Movus, 1038
Moxon, A. L., 559, 599
Moyse, A., 474
Mudd, S., 220, 222, 224, 235, 929, 950, 965, 978, 1118
Mueller, H., 789, 807
Mueller, R., 1153
Muenzer, E., 1153
Muir, C., 1115
Mulder, A. G., 1152
Mullen, 679, 682
Muller, A., 270, 273, 279
Müller, F., 938, 950
Müller, H., 362, 371, 372
Müller, H. C., 620, 665
Muller, H. J., 578, 584, 585, 806, 813, 820, 823, 825, 829, 830, 831, 832, 833, 834, 842, 845, 847, 850, 1197
Müller, H. O., 169, 182, 186, 211, 235
Mullins, L. J., 898, 900
Munchen, W., 950
del Mundo, Fé, 1115
Muntwyler, E., 748, 752
Muralt, A. von, 884, 892, 894, 895, 899, 900
Murphy, J. B., 998, 1000, 1001, 1005, 1031, 1049, 1050, 1160
Murray, 1006, 1008, 1011, 1012
Murray, R. C., 119
Mustard, R. A., 956
Mutzenbecker, P. von, 567
Muus, J., 701, 706, 956
Myers, L. M., 154, 235
Myers, R. J., 39, 51, 59, 100, 372
Myers, Wm. G., 456
Myerson, A., 1147, 1152
Mylon, E., 956
Myrback, 682

N

Nachmansohn, D., 579, 580, 598, 888, 899
Nagel, K., 469, 471
Nageli, 864, 867
Nägeli, Carl von, 617, 618, 619, 620, 621, 623, 624, 631, 632, 636, 653, 660, 665
Nagy, R., 1117
Naruse, N., 372
Nash, T. P., 1152
National Research Council, 122, 716, 728, 737
Naunyn, B., 1072, 1075, 1079, 1082
Navez, A. E., 768, 785
Naylor, 682
Neale, S. M., 636, 665, 682
Nebel, B. R., 830, 850
Neblette, C. B., 512
Neddermeyer, Seth, 9
Needham, D. M., 544, 706, 898, 899, 900
Needham, J., 577, 598, 599, 898, 899, 900
Negelein, E., 601, 602, 609, 699, 706
Nelson, 750, 1013
Nernst, W., 445, 456, 1153
Netter, H., 1153
Nettleship, A., 933, 934, 950
Neubauer, E., 1121, 1124, 1136, 1153
Neuberg, C., 560, 686, 706, 759, 761
Newcombe, H. B., 850
Neukirch, P., 950
Neumann, 518
Neurath, H., 413, 433, 791, 794, 795, 803, 807, 879, 883, 964, 978
Newburgh, 1062
Newman, L. H., 988, 994
Neville, 146, 151
Newell, J. M., 409, 410, 411
Newton, 381, 682
Nichols, J. B., 411, 427, 432, 434
Nickerson, R. F., 649, 665
Niemann, 536, 544
de Niépce, Nicèphore, 472
Nietz, A. H., 511
Nims, B., 955
Nims, L. F., 1150, 1152, 1153, 1154
Niggli, 445, 456
Nikitin, L. V., 372
Niles, G. E., 120
Nilsson, E., 524, 525
Nishikawa, S., 631, 665
Nissl, F., 1152
Noble, W. C., Jr., 1117
Noddack, W., 602, 609
Noetzel, W., 943, 950
Nolf, P., 956
Noll, D., 895, 899
Noll, F., 620, 665
Norberg, B., 843, 850
Nord, F. F., 609, 705
Nordbö, R., 952, 956
Nordlund, I., 347, 372
Nordlund, I., 347, 372
Nordmann, M., 930, 950
Norin, E., 524
Norman, A. G., 649, 665
Norrich, G. R. W., 285
Northrup, J. H., 409, 807, 954, 978
Norton, F. J., 789, 807
Notkin, J., 1153
du Noüy, P. L., 876, 1161
Noujdin, N. O., 847, 850
Novi, 1153
Novikoff, A., 857, 859
Nowinski, W. W., 599
Noyes, 445, 456
Noyes, W. A., Jr., 14
Nugent, R. L., 978, 1147, 1153
Nutini, L. G., 1117
Nutting, G. C., 59, 85, 101
Nutting, P. G., 507, 512

O

Obermeyer, 963
Ochoa, S., 687, 688, 689, 691, 695, 701, 706, 721, 722, 737
Ochs, I., 949
Odake, S., 693, 706
Oestreicher, A., 610
Offerman, C. A., 824, 845, 850
Ogg, 456
Ogston, 803
Ohlmeyer, F., 695, 706
Oka, S., 360, 372
Okell, C. C., 1117
Okuneff, N., 920, 950
Oldham, J. W. H., 271, 280
Oliver, 682
Olney, M. B., 1096

Olpin, A. R., 512
Olsen, O., 1117
Olson, A. R., 372
Olson, R. A., 373
Omar, A. M., 1129, 1135
Oncley, J. L., 410, 433
Ono, K., 631, 665
Ono, S., 372
Opie, E. L., 926, 938, 950
Oppenheimer, C., 706, 956
Ord, W. M., 12, 444, 456, 611, 665, 879
Orent, E. R., 711, 736
Ornstein, L. S., 603, 609, 696, 704
Orth, H., 609
Orth, O. S., 599
Osborn, Henry F., 593, 599
Oseen, C. W., 372
Osnato, M., 1153
Ost, H., 630, 665
Osterhout, W. J. V., 762, 763, 785, 1153
Ostrovski, E. P., 372
Ostwald, Wilhelm, 473, 514, 518, 1153
Ostwald, Wolfgang, 514, 518
O'Sullivan, C. M., 428, 433
Ott, E., 186, 235, 279
Otterson, 657
Otto, 981, 985
Overbeek, J. van, 583, 599
Overholt, R. H., 1117
Overholzer, 1014, 1015
Overton, 1153
Owen, 1022
Oxman, M. F., 411

P

Pace, D. M., 1166, 1173
Pachewitsch, 580
Pacsu, 679, 682
Padtberg, J. H., 988, 994
Page, 287
Page, I. N., 785
Page, I. H., 1145, 1153
Page, J. H., 1121, 1135
Paine, 456
Paine, C. G., 1117
Painter, T. S., 821, 824, 830, 831, 832, 845, 848, 850
Palladin, A., 1145, 1153
Palladin, V. I., 622, 665
Paletta, 1021
Palmer, 441, 443
Palmer, K. J., 1121, 1129, 1131, 1134, 1136
Paneth, 457
Panshin, I. B., 585, 599, 847, 850
Pappenheimer, A. M., 964, 965, 977, 978
Pappenheimer, A. M., Jr., 409, 411, 433
Paracelsus, 593
Parfentjev, I. A., 956, 960, 978, 979
Park, 1060
Park, Wm., 598
Parker, B. W., 1117
Parker, George H., 579, 599, 746, 752
Parker, R. C., 360, 372
Parkins, W. M., 950
Parks, G. S., 280, 324, 326
Parnas, J. K., 973, 978
Parrot, J. L., 950
Parsons, H. T., 683, 706
Pascher, A., 1171, 1173
Pässler, J., 1135
Passmore, R., 685, 706
Pasteur, L., 550, 1001, 1002, 1100, 1117
Patck, A. J., 756, 761
Patterson, 535
Patterson, A. L., 136, 145
Patterson, H. S., 360, 372, 1118
Patterson, J. T., 575, 832, 850
Pauli, W., 9, 1124, 1125, 1129, 1135
Pauli, W. E., 1153
Pauling, L., 145, 326, 512, 571, 599, 815, 841, 850, 961, 967, 968, 978, 1049
Paunow, P. G., 372
Pavlov, P., 1154
Payen, A., 616, 617, 619, 620, 650, 660, 665
Peacock, 1019
Pearson, E. B., 360, 372
Pearson, P. B., 581, 599
Peat, 683
Pecher, Charles, 378, 380
Peck, V. G., 216, 217, 234
Pecker, J. S., 433
Pederson, Kai O., 409, 411, 421, 431, 433, 434, 544, 564, 567, 721, 722, 737, 807, 836, 850, 964, 978
Peeler, D. B., 1140, 1153
Peissakhowitsch, S., 475, 511
Pennell, R. B., 976, 978
Pentimalli, 1001
Percival, G. H., 921, 950
Perkinson, J. R., Jr., 703, 706
Perlmann, G., 690, 705
Perlman, I., 1135
Perrin, F., 416, 433, 794, 795, 807
Perry, E. S., 475, 511
Perry, L. H., 120, 434
Perutz, 544
Pesta, H., 1134, 1135
Peterfi, T., 1147, 1153
Petering, H. G., 602, 609
Peters, H., 354, 373
Peters, J. P., 1119, 1135, 1136, 1153
Peters, R. A., 685, 686, 687, 688, 691, 705, 706, 721, 722, 737, 1153
Peterson, 308
Peterson, A. W., 956
Peterson, D. K., 941, 949
Petralia, S., 360, 372
Peugnet, H. B., 701, 706
Pfankuch, E., 226, 235, 798, 807
Pfeffer, 867
Pfeffer, W., 665
Pfeiffer, 1013
Pfeiffer, P., 95
Pfiffner, J. J., 740, 747, 752, 950
Pfluger, 864
Pfitzer, E., 620, 665
Phelps, E. B., 1115, 1117, 1118
Philippi, G. T., 883
Phillips, J. W. C., 279
Phillipson, 1153
Philpot, J. S., 395, 410, 421, 422, 433, 567
Phipps, J. W., 265, 266
Picard, R. G., 235
Piccard, Jean, 1082
Pick, 963
Pick, E. P., 1153
Pickarski, G., 235
Pickels, E. G., 411, 412, 418, 419, 423, 427, 428, 430, 432, 433, 434, 759, 760, 761, 808
Picken, L. E. R., 886, 887, 897, 899
Pickering, J. W., 956
Pierce, 513, 515, 516, 518
Pierce, G. W., 342, 372
Pighini, G., 1153
Pike, E. W., 589, 599
Pike, F. H., 589, 599, 1135, 1153
Pillemer, L., 703, 704
Pincus, G., 783, 785
Pincus, S., 1117
Piontelli, R., 372
Piper, S. H., 269, 279, 1120, 1134, 1136
Pirie, 848
Pirie, N. W., 433, 801, 802, 807
Pirie, W. M., 599
Pirquet, A., 366, 371
Pirson, A., 609
Pittman, M., 696, 706
Planck, 515
Plass, G. N., 164, 235
Plass, M., 699, 704
Platz, B. R., 725, 737
Platt, B. S., 685, 706
Plaut, F., 978
Pliny, C. P., 308
Ploetz, 682
Plough, H. H., 850
Plunkett, Chas. R., 591, 599, 1173, 1196, 1197
Pluzhnik, E. E., 371
Pobolev, V., 370
Pockels, Agnes, 35, 37
Podolsky, 376
Poetzl, O., 1153
Pohlman, R., 360, 372
Poisson, 89, 292
Pol, B. van der, 766, 785
Poland, 1053
Polevitsky, K., 220, 235
Pollack, H., 874, 875, 1151, 1152
Pollister, A. W., 822, 836, 839, 849
Polotskii, I. G., 372, 373
Polson, A., 567, 794, 807
Ponder, E., 868, 875, 934, 950
Pope, Alexander, 11
Popow, K., 372
Porges, O., 1121, 1124, 1136, 1153
Porter, B. H., 372
Porter, C. W., 372
Porter, E. F., 88, 965, 978
Pott, Percivall, 589
Potter, U. R., 685, 704, 706
Potter, J. S., 836, 849
Pouchet, 579
Poulter, T. C., 333, 337
Pourbaix, 1022
Powell, E. O., 1116, 1118
Powell, Richard E., 237, 252
Power, F. W., 587, 599
Powney, J., 120
Poyner, H., 830, 848
Prakke, 233
Pratt, E. F., 702, 706
Prebus, Albert F., 152, 176, 235, 489, 838
Pregl, 950
Presnell, A. K., 702, 706
Pressman, R., 1117
Preston, R. D., 649, 654, 665
Price, 1022
Price, W. C., 807, 808
Price, C. W., 1122, 1136
Priestley, J. G., 992, 994, 1083, 1091
Priestly, J., 625, 630, 665
Prince, H. E., 1117
Proctor, B. E., 1117
Protas, I. R., 372
Prokofyeva-Belgoskaia, A. A., 823, 833, 845, 847, 850
Prout, W., 610, 611, 650, 659, 666
Prytherch, H. F., 599
Puck, T. T., 1117
Pulvertaft, R. J. V., 1117
Punnett, R. C., 584, 809
Purdy, 1000
Purkinje, 756, 1056
Puschel, F., 105, 119
Putnoky, 997

Q

Quackenbush, F. W., 262, 266, 733, 738
Quastel, J. H., 691, 692, 697, 706, 1152, 1153
Querido, A., 696, 704
Quensel, O., 567
Quick, A. J., 956
Quilliam, J. P., 761

Quincke, G., 4, 12, 15
Quinton, 707

R

Raab, O., 588
Rabinowitch, B., 234
Rabinowitsch, A. J., 475, 493, 511
Rackemann, 984, 985
Radley, 667, 682
Radozewski, O. E., 234
Raff, R., 285
Rahn, O., 599
Rainey, G., 611, 666
Ralph, P. H., 956
Ralston, A. W., 86
Ramaiak, K. S., 518
Raman, Sir C. V., 518
Ramberg, E. G., 176, 215, 234, 235
Ramo, S., 164, 165, 234
Ramsdell, S. G., 920, 950
Ramsden, W., 880, 883
Randall, A., 1067, 1068, 1082
Randall, J. T., 326
Randolph, 984, 985
Randolph, A., 372
Ransmeier, J. C., 956
Ranson, 750, 752
Ranson, S. W., 993, 994
Raper, H. S., 582, 599
Raper, Kenneth B., 573, 574
Rapport, J. A., 824, 850
Raspail, 880
Rauch, K., 705
Ravenswaay, A. C., von, 1153
Rawlins, T. A., 788, 808
Rawson, A. J., 430, 432
Rayleigh, Lord, 15, 18, 35, 166, 348, 372, 1121
Reames, H. R., 704
Record, B. R., 409, 410
Redwitz, E., 1153
Reece, G. M., 1117
Reeds, Chester A., 519
Regendanz, P., 929, 950
Reggiani, M., 372
Rehner, 293, 308
Reichardt, M., 1145, 1154
Reichstein, T., 747, 752
Reifenstein, G. H., 937, 950
Reike, F. F., 602, 609
Reimann, 1022
Rein, Herbert, 664
Reinders, W., 454, 457, 473, 492, 511, 638, 666
Reinecke, F., 648, 666
Reinhard, 1039
Reinhold, H., 496
Reiss, R., 620, 666
Reissek, S., 629, 666
Reller, 1021
Reman, E. C., 609
Remec, B., 628, 666
Remesow, J., 1122, 1136
Remington, R. E., 718, 737
Remy, E., 979
Rentschler, H. C., 1117
Renwick, F. F., 482, 511
Reuther, A., 930, 950
Rex, R. O., 949
Reychler, A., 102, 119
Reymann, G. C., 959, 978
Reyniers, James A., 875, 883, 1108, 1117
Reynolds, L., 1149, 1154
Reynolds, Osborne, 366, 372, 1088
Reynolds, S. R. M., 746, 747, 752
Reznikoff, P., 875
Rheinboldt, H., 459, 471, 759, 761
Rhoades, H. E., 561, 599
Rhoades, M. M., 850
Rhoads, 1003, 1038
Rhodewald, G., 600
Ribbert, 1041
Richards, A. G., 186, 233
Richards, L. M., 551, 597
Richards, O. W., 1166, 1173
Richards, W. T., 346, 353, 368, 372
Richardson, 682
Richardson, E. G., 372
Richardson, R. J., 945, 950
Richter, 603
Richter, C. P., 750, 752
Richter, D., 746, 752
Richter, M. M., 1050
Richet, 979, 982, 985
Ricker, G., 929, 950
Rideal, Eric, 42, 62, 85
Riddle, O., 1153
Riecker, H. H., 990, 994
Riegel, E. R., 517
Riehl, N., 589, 599
Ries, H. E., 101
Rigdon, R. H., 928, 950
Righter, F. L., 512
Rigterink, 456
Riley, 535, 544
Riley, E. C., 1118
Riley, D. P., 805, 807
Rinehart, J. F., 1096
Rinde, H., 411, 414, 420, 434
Rinn, H. W., 122, 130
Rinne, F., 1134, 1136
Ritter, G. J., 646, 666
Rittenberg, D., 1135
Rittner, E. S., 373
Ritzan, G., 369, 372
Ritzel, 445
Rivers, T. M., 434, 808
Roberts, 1026
Roberts, D., 501, 503, 512
Roberts, E. A., 648, 661, 666
Roberts, L. E., 100
Roberts, R. M., 977
Robertson, E. B., 1118
Robertson, E. C., 1117
Robertson, J. Monteath, 145, 547, 548, 549, 550
Robertson, O. H., 1117
Robertson, T. B., 1165, 1173
Robeznieks, I., 705
Robinson, 471, 560
Robinson, F. W., 1117
Robinson, H. G. B., 107, 119
Robinson, J. R., 793, 795, 808
Robinson, W. O., 10, 12
Rocha e Silva, M., 921, 922, 923, 925, 950, 982, 985
Rochaix, A., 1117
Rodebush, W. H., 95
Roedig, A., 705
Roffo, 1019
Rogers, U., 598
Rogovin, S., 636, 663
Rogowski, F., 340, 341, 367, 368, 369, 371
Rohrmann, E., 729, 738
Rolf, J. P., 1122, 1135
Roll, A., 373
Romé de l'Isle, J. B. L., 444, 457
Rommel, O., 373
Rona, P., 1125, 1126, 1135, 1136
Rooks, R., 1117
Rooksby, H. P., 326
Rose, C. S., 684, 705
Rose, S. M., 857, 859
Rose, W. C., 719, 737
Roseberry, H. H., 713, 736
Rosenau, 981, 985
Rosenberg, H. R., 722, 724, 737
Rosenheim, O., 1134, 1136
Rosenstern, I., 1117
Rosenthal, Sanford M., 916
Roser, E., 1139
Roseveare, W. E., 252
Ross, A. F., 410, 433, 807, 808
Ross, F. E., 512
Ross, G. S., 736, 1152
Rossenbeck, H., 849
Rossem, A., van, 286
Rossiter, R. J., 701, 706
Rothen, A., 410, 411, 426, 433, 434, 964, 965, 978
Rothmund, Paul, 600, 609, 610
Rouget, 901
Rountree, L. G., 1154
Rous, Peyton, 590, 909, 916, 918, 921, 950, 1000, 1026, 1028, 1029, 1030, 1031, 1032, 1034, 1049, 1050
Rowe, 750, 752
Rowland, A. F., 1149, 1152
Rowntree, 1060
Rowntree, L. G., 750, 752
Royer, 455, 457
Rschevkin, S. N., 353, 372
Ruben, S., 434, 606, 609
Rubin, M. A., 785
Rubner, 1129
Rudall, 530
Ruddy, Sister M. V., 702, 706
Rudy, H., 698, 705, 978
Ruge, U., 649, 666
Rugiero, C. A., 939, 950
Ruhl-Koupal, J., 1115
Rundle, C., 1117
Rundle, R. E., 667, 676, 681, 682
Runge, G., 518
Runnstrom, J., 852, 855, 859
Ruska, E., 154, 156, 159, 162, 166, 169, 170, 175, 182, 186, 208, 209, 210, 211, 223, 234, 235
Ruska, E., 808
Ruska, H., 223, 226, 229, 235, 649, 666, 807, 808
Russel-Wells, B., 637, 666
Russell, Maj., 474, 511
Russow, E., 620, 666
Rutherford, Sir Ernest, 9, 169, 375, 555
Rutzler, J. E., 1129, 1135
Rutzler, S. E., 1152
Ružicka, V., 1154
Ryan, L. W., 100
Ryer, F. V., 276
Rytov, S. M., 372
Rzymkowski, J., 1118

S

Sachs, J., 625, 650, 666
Sacks, T., 884, 899
Sah, P. P. T., 979
St. Clair, H. W., 342, 344, 345, 360, 370
Sair, 679, 682
Sakmann, B. W., 789, 807
Salaman, M. H., 970, 978
Salley, R. J., 1198
Salmon, C. S., 119
Salcher, P., 345, 372
Salisbury, W. W., 372
Salomon, K., 706
Salmon, W. D., 684, 706
Samec, M., 667, 680, 682
Sammis, C. S., 112, 119
Samsa, 678, 682
Samuel, S., 929, 950
Samuels, 1038
Sanderson, E. S., 1153
Sanford, C. R., 265
Sanger, P. W., 1116
Saniger, E. B., 434
Santessen, L., 849
Sasaki, 1003
Sassure, N. T. de, 625, 666, 667
Sata, N., 372
Saubert, G. G. P., 1124, 1135
Sauer, L. W., 1116
Saum, A. M., 791, 795, 807
Saunders, F., 696, 704, 706
Sauter, E., 636, 666
Saville, W. B., 279
Sax, K., 828, 829, 850
Saxl, P., 896, 900
Saxton, 998
Saylor, C. P., 599
Sazerac, 558, 597
Scarlet, 614
Scarisbrick, R., 604, 609

Schaaf, 682
Schacht, H., 666
Schade, H., 457, 917, 918, 938, 950, 1051, 1070
Schaefer, V. L., 796, 807, 883
Schaeffer, V. J., 62, 187, 216, 217, 218, 219, 235
Schaffer, P. A., 512
Schär, B., 692, 707
Sharp, L. W., 666
Scheer, J. van der, 409, 411, 434, 978
Scheffer, W., 489, 512
Scheibe, A., 339, 373
Scheil, E., 371
Scheinker, L., 1137, 1154
Schemin, 1000, 1032
Schenck, E., 839, 849
Schenk, 982, 985
Schenken, 1008, 1013, 1024
Scherp, H. W., 430, 432
Scherrer, P., 457, 631, 663
Scherzer, O., 154, 235
Schiebel, H. M., 1116
Schiff, 690
Schiff, E., 1139
Schiff, L. I., 169, 175, 192, 235
Schifrin, A., 704
Schimper, 650
Schjelderup, H., 326
Schlaer, S., 756, 760, 761
Schleiden, M. J., 610, 614, 618, 621, 666
Schlenk, F., 705
Schlesinger, M., 430, 433, 434, 791, 807
Schlotman, 654
Schmid, G., 354, 355, 369, 373
Schmidlin, J., 467, 471
Schmidt, 1148
Schmidt, A., 952, 956
Schmidt, C. L. A., 711, 736, 848
Schmidt, E., 353, 354, 371
Schmidt, G., 434
Schmidt, W. J., 754, 759, 761, 840, 850
Schmidt, W. T., 894, 899
Schmitt, F. O., 230, 231, 233, 235, 373, 759, 762, 868, 869, 875, 895, 899, 1121, 1129, 1131, 1132, 1134, 1136, 1149, 1153, 1154, 1158, 1161
Schmitt, W., 1128, 1136
Schmitz, F., 620, 666
Schmitz, A., 954, 956
Schneider, H. A., 725, 737, 1120, 1135
Schoch, T. J., 668, 669, 675, 676, 677, 682
Schoen, 489, 490
Schoen, A. L., 476, 511
Schoepfle, G., 882
Schön, K., 471
Schönauer, W., 467, 471
Schönbein, 459
Schønheyder, F., 734, 738, 956
Schoon, 190, 206, 207, 234, 235
Schoon, T., 267, 269, 270, 279
Schorger, A. W., 630, 666
Schott, 1038
Schrader, F., 822, 823, 850
Schramm, G., 807
Schrieber, H., 830, 849
Schrodinger, 515
Schuler, L., 466, 471
Schulman, 35, 42
Schulman, J. H., 883
Schulman, J. H., 62, 65, 66, 67, 68
Schultes, H., 371
Schultz, 981, 985
Schultz, Fr. N., 580
Schultz, J., 1197
Schultz, Jack, 819, 823, 824, 825, 826, 831, 832, 833, 838, 843, 844, 845, 846, 847, 849, 850
Schulz, G. V., 285
Schulz, W., 899
Schulze, E., 622, 623, 666
Schumb, W. C., 354, 373, 1198
Schunck, E., 628, 666
Schuster, P., 686, 687, 688, 689, 706, 721, 737
Schusterius, C., 234
Schuwirth, K., 1120, 1135
Schuyl, J. W., 701, 706
Schwann, T., 610, 666
Schwerin, P., 555, 598, 978
Schwartz, W. P., 703, 704
Schwarz, G., 482, 486, 505, 511
Schweitzer, E., 617, 666
Scott, C. M., 921, 950
Scott, T. F. M., 434
Screenivasaya, M., 599
Scribner, B. F., 10, 12
Scudder, J., 406, 411
Sealock, R. R., 703, 706
Sealy, W. C., 433
Sears, G. R., 649, 666
Sears, G. S., 210, 235
Sebrell, W. H., 728, 738
Seeber, 457
Seegers, W. H., 956
Seemann, H. J., 373
Segré, E., 379
Seidl, K., 511
Seifriz, W., 632, 634
Selbach, H., 1154
Selbie, F. R., 431, 433
Seljakow, N., 326
Selwyn, E. W. H., 512
Semenoff, 512
Senarmont, 444, 457
Serber, R., 387
Serebrovsky, A. S., 825, 845, 850
Serles, E. R., 1153
Sesjulinski, B. M., 373
Sevag, M. G., 935, 950
Severinghaus, A. E., 433, 434, 744, 752
Severny, A. B., 371
Severson, 675, 682
Shabad, 1024, 1025
Shablykin, P. N., 373
Shafer, 607
Shafer, J., Jr., 610
Shafer, W., 151
Shaffer, P. A., 145
Shapiro, H., 434, 882
Shapiro, S., 1130, 1135
Sharp, D. G., 411, 433, 434, 1117
Sharp, E. R., 120
Sharp, L. W., 630, 666
Shaw, 654
Shaw, B. T., 195, 196, 202, 203, 207, 235
Shear, M. J., 1017, 1024, 1027, 1050
Shearer, G., 279
Shedlovsky, T., 393, 406, 407, 410, 411, 808, 937, 949
Shelberg, E. F., 1154
Shemyakin, 515, 518
Shepard, N. A., 286
Sheppard, S. E., 308, 472, 474, 475, 478, 484, 486, 487, 488, 492, 496, 497, 511, 512
Sherman, A., 473, 511
Sherman, H. C., 712, 715, 720, 736, 737
Sherman, W. C., 685, 706
Sherwin, 587, 599
Sherwin, C. P., 956
Shettles, L. B., 783, 785
Shi-Chang Shen, 898, 900
Shick, 985
Shields, 1017
Shils, M. E., 686, 706, 719, 737
Shimamura, T., 693, 706
Shimkin, 1012, 1020
Shipley, R. A., 409, 411
Shirai, 998
Shohl, A. T., 712, 736
Shone, H. R., 1118
Shope, 806, 1003, 1005, 1031, 1032
Shourie, K. L., 706
Shute, H. L., 120
Shuwirth, 1134
Shwachman, H., 1116
Sia, R. H. P., 569, 598, 599
Sibata, K., 354, 373
Sidoroff, B. N., 585, 598
Siebert, Florence B., 409, 411, 567
Siedentopf, H., 336
Siengalewicz, 919, 950
Signer, R., 412, 434, 836, 850
Silberstein, L., 677, 682
Silverberg, R. J., 1116
Silverman, Alexander, 308
Silverman, M., 686, 706, 1099
Sim, R. P., 1117
Simha, R., 794, 808
Simola, P. E., 686, 687, 706
Simmonnet, H., 1130, 1135
Simonds, 982, 983, 985
Sinclair, 1153
Sinclair, H. M., 687, 705, 706
Singer, E., 1124, 1136
Singer, R., 636, 667
Sinnott, 816
Sisson, W. A., 544, 641, 644, 646, 648, 649, 657, 658, 660, 663
Sivertz, V., 119
Sizer, I. W., 771, 772, 774, 784, 785
Sjogren, B., 417, 434, 567
Skarstrom, C., 420, 431, 433, 434
Skelton, H. P., 987, 994
Skow, R. K., 610
Slagle, F. B., 279
Slattery, M. K., 326
Slizynski, B. M., 834, 850
Slye, Maude, 1008, 1050
Smadel, J. E., 223, 234, 393, 411, 433, 434, 807, 808
Small, W. S., 1116
Smillie, W. G., 1100, 1115, 1117
Smirnova, 998
Smith, 443, 1015
Smith, E. C., 602, 609
Smith, F., 916, 918, 950
Smith, E. F., 586, 599
Smith, E. L., 434, 557, 599, 605, 609, 759, 760, 762
Smith, F. D., 349, 373
Smith, G. M., 597
Smith, H. P., 956, 957
Smith, J. C., 279
Smith, K. M., 796, 807
Smith, P. E., 744, 752
Smith, T. B., 518
Smith, Theobald, 981
Smith, W., 1115
Smittenberg, J., 326
Smolens, J., 936, 950
Smyth, C. P., 272, 280, 326
Smyth, D. H., 706
Snapper, 1055
Snell, E. E., 684, 705, 734, 738, 825
Snoddy, L. B., 433
Snyder, 308
Snyder, L. H., 587, 599
Snyder, R. L., 213, 214, 235
Sober, H. A., 691, 704, 706
Soddy, F., 2, 12
Sokoloff, S. J., 373
Soley, M. H., 379, 380
Sollner, Karl, 337, 340, 341, 346, 347, 349, 350, 351, 352, 353, 357, 359, 360, 361, 363, 365, 367, 369, 371, 373, 1088
Solomon, H. C., 785
Solotorovsky, M., 1117
Solowey, M., 1115, 1117
Solotorovsky, M., 1115
Solov'eva, L. R., 372
Sommer, 430, 433
Sommers, Sheldon C., 926, 927, 950
Sompkin, 740, 752
Sonneborn, T. M., 599, 1171, 1173
Soper, F. G., 501, 503, 512
Sorby, 604
Sörensen, C., 373
Sörensen, E., 370

Sørensen, S. L. P., 403, 411
Sorge, H., 759, 762
Soskin, S., 989, 994
Spaght, M., 280
Spangenberg, 457
Spangler, J. M., 1049
Spatz, H., 1154
Spear, E. B., 286, 291
Spedding, F. H., 676, 682
Spemann, H., 576, 599
Spencer, 1023
Spencer, Herbert, 2, 11, 591
Spencer, Leo, 630, 663
Spencer, W. V., 120
Spiegel, 1013
Spiegel, E., 1124, 1136, 1140, 1142, 1143, 1150, 1153, 1154
Spiegel-Adolf, M., 1119, 1124, 1132, 1136, 1140, 1141, 1142, 1143, 1150, 1154
Spierer, C., 634, 666
Spies, T. D., 696, 704
Spiro, 1082
Sponsler, O. L., 10, 12, 635, 646, 660
Spooner, E. T. C., 1118
Sprague, 682
Sprince, H., 1049
Spring, Walthere, 336
Sproul, 1000
Spychalski, R., 280
Sreenivasaya, M., 808
Ssawron, E., 1145, 1153
Stacey, M., 569
Stadtler, L. J., 825, 830, 832, 833, 834, 835, 850
Staehlin, S., 692, 707
Stakman, E. C., 1117
Stamm, A. J., 567
Stanley, J., 1126, 1136
Stanley, W. M., 226, 228, 229, 235, 410, 430, 433, 434, 457, 594, 599, 785, 786, 791, 794, 796, 798, 800, 802, 806, 808, 843, 847, 848, 849, 850
Stanfield, 581
Stapelfeldt, F., 663
Stare, F. J., 695, 704
Starling, E. H., 739, 741, 752, 903, 904, 906, 916, 917, 950
Staude, H., 482, 511
Staudinger, H., 285, 635, 636, 648, 666, 667, 668, 682
Stauff, J., 115, 120, 270, 280
Stauffer, J. F., 602, 609
Stearn, A. E., 252
Stearns, G., 717, 735, 737, 738
Stedman, E., 691, 706, 836
Steeger, A., 1117
Steele, C. C., 609
Steenbock, H., 262, 266, 598, 683, 705, 716, 717, 725, 736, 737, 738
Stehle, R. L., 751, 752
Stein, 544
Steinberg, B., 939, 950
Steinberg, R. A., 10, 586
Steiner, 1025
Stendel, R., 486, 511
Stengel, E., 1154
Stenhagen, E., 67, 406, 409, 411, 957
Steno, 443, 457
Sterling, A., 411
Stern, A. C., 1117
Stern, Curt, 835, 850
Stern, K. G., 409, 410, 411, 431, 433, 434, 689, 690, 706, 737, 844, 850
Steven, D., 762
Stevens, F. A., 1117
Stevenson, 1003
Stewart, 1022
Stewart, G. W., 561
Stewart, T. D., 1097
Stieglitz, E. J., 1051
Stiles, 945
Stiles, W., 609, 649, 667
Stimson, A. M., 1117
Stirn, F. E., 684, 706
Stockmeyer, 291, 308
Stokes, Sir George G., 236, 413, 414, 434, 791, 792
Stokes, J., Jr., 1116
Stokinger, A. E., 963, 978
Stoll, A., 603, 609
Stolz, E., 727, 737
Stone, W. E., 1154
Stone, W. R., 1118
Stoney, G. Johnstone, 9
Storks, K. H., 130
Stotz, E., 775, 779, 785
Straibel, H., 373
Strain, H. H., 459, 462, 463, 464, 466, 471, 609
Strang, P. M., 12
Stranski, 454, 457
Strasburger, E., 622, 624, 628, 638, 660, 667
Straub, F. B., 698, 706
Strauss, W., 1117
Street, 544
Strong, 1008, 1009, 1015, 1023
Strong, F. M., 704
Strong, J., 265, 266
Strutinski, L., 326
Stubbe, H., 798, 807
Stuber, C., 279
Studnitz, Z., 897, 900
Stumpf, P. K., 699, 705
Sturm, 1000
Sturtevant, A. H., 821, 823, 834, 850
Submarine Signal Co., 373
Subrahmanyan, V., 686, 705
Sugg, J. Y., 570, 598
Sullmann, H., 706
Sullivan, J. C., 1116
Sumner, Francis B., 579, 599
Summers, F. M., 1173
Sumner, J. B., 567
Sun, Kuan-Han, 308, 326
Suntzeff, 1016, 1023, 1024
Surangi, L., 1158, 1159
Sure, B., 701, 702, 706
Susich, G. von, 326
Suto-Nagy, G. J. de, 956
Sutton, 823
Sutton, T. S., 1154
Suzuki, B., 1120, 1136
Suzuki, U., 693, 706
Svedberg, The, 411, 412, 413, 414, 415, 416, 417, 420, 426, 427, 428, 431, 432, 433, 434, 488, 512, 544, 562, 563, 564, 567, 580, 605, 791, 795, 799, 807, 850, 952, 957
Svensson, H., 395, 409, 411, 422, 434
Swaminathan, M., 706
Swanson, C. P., 830, 850
Sweet, S. S., 486, 511
Swezy, O., 815
Swingle, K. F., 704
Swingle, W. W., 950
Sykes, A., 1152
Sylvester, 5
Sylvestri, F., 599
Syz, H. C., 1154
Szalay, A., 354, 355, 373
Szent-Gyorgi, A., 354, 373, 589, 599, 687, 697, 704, 785
Szymanska, R. M., 693, 706

T

Tabern, D. L., 1154
Tabulae Biologicae, 1117
Taft, A., 1147, 1153
Tage-Hansen, E., 734, 738
Tagnon, H. J., 956, 957
Takahashi, W. N., 788, 808
Tammann, G., 284, 329, 330, 337, 499, 457, 860, 863
Tang, S. P., 609
Tannenbaum, A., 1038, 1050
Tansley, K., 759, 762
Tanturi, H. R., 939, 950
Tapley, G. O., 1116
Tartar, H. V., 107, 115, 119, 350, 371
Tatum, E. L., 582, 599
Tatum, G. R., 360, 373
Taub, E., 586
Tauber, H., 599, 721, 737
Tauber, J., 721, 737
Taylor, 544, 831, 832, 845
Taylor, A. R., 411, 433, 434
Taylor, F. H. L., 956, 957
Taylor, G. B., 370
Taylor, Hugh S., 373, 380, 473, 511
Taylor, T. I., 350, 371
Taurog, A., 1135
Teague, O., 1117
Teague, R. S., 994
Teller, 97, 99, 443
Ten Broeck, C., 979
Tennenbaum, M., 691, 706
Tennyson, A., 3
Teorell, T., 409, 411
Terletzki, P., 667
Terzaghi, K. von, 192
Tesla, Nicola, 188, 341
Teunisson, P. H., 1134, 1135, 1136
Teverovskii, B. M., 371
Thanhauser, S. L., 1136
Thayer, S. A., 955
Theis, R. M., 702, 706
Theorell, B., 838, 843, 849
Theorell, H., 411, 567, 698, 700, 706
Theorell, T., 1131, 1135
Thierfelder, A., 598
Thierfelder, H., 1120, 1122, 1134, 1136
Tiselius, A., 952, 957, 960, 978, 979
Thiessen, P. A., 270, 279, 280, 634, 635, 667, 1149
Thoennes, G., 863
Thom, C., 586, 599
Thomas, A. W., 516
Thomas, B. G. H., 1117
Thomas, J., 1152
Thomas, J. C., 1115, 1117
Thomas, P. T., 847, 849
Thomas, S. B., 280, 326
Thomas, W. M., 119
Thompson, K. H. S., 685, 706
Thompson, K. W., 409, 411
Thompson, R. H., 1153
Thomson, D., 1117
Thomson, G. P., 234
Thomson, H. E., 285
Thomson, Sir J. J., 4, 9, 12, 121, 130
Thomson, R., 1117
Thomson, St. Clair, 1117
Thomson, Wm., 287
Thornhill, 443
Thornberry, H. H., 796, 807
Thovert, J., 421, 434
Tiebackx, F. W., 8, 12
Tilden, 682
Timofeef-Ressovsky, N. W., 834, 842, 850
Tisdall, F. E., 736
Tisdall, F. T., 1152
Tiselius, A., 387, 389, 390, 391, 392, 401, 404, 409, 411, 415, 421, 431, 433, 434, 467, 471, 567, 646, 667, 799
Titley, A. F., 104, 119
Titus, R., 512
Todd, A. R., 689, 704
Todd, C., 970, 979
Toepler, A., 389, 391, 421, 434, 461, 467
Tolman, R. C., 1117
Tönnis, B., 729, 738, 1124, 1136
Tönnis, W., 1124, 1136, 1137, 1142, 1154
Topley, W. W. C., 572, 574, 599
Torda, Clara, 579, 599
Torrès, 775
Tostado, C. Alvarez, 409, 410
Townsend, G. F., 581, 599
Toy, F. C., 511
Traube, 1154
Trecul, 617

Treffers, H. P., 434, 977, 978
Trethewie, 983, 985
Trexler, P. C., 1117
Trillat, A., 1117
Trillat, J. J., 279, 372
Trimble, 599
Trimble, H. C., 988, 994
Trivelli, A. P. H., 488, 489, 497, 511, 512
Troland, Leonard T., 600, 820, 850, 1197
Tröndle, A., 628, 667
Truesdail, J. H., 729, 738
Trurnit, H., 807
Tschermak, E., 808
Tschirch, A., 619, 620, 622, 624, 667
Tswett, M., 458, 460, 604
Tsutsui, 1017
Tuchman, J., 704
Tucker, J. R., 1118
Tumanski, S. S., 373
Tuorila, P., 362, 373
Tupper-Carey, 630
Turner, C. E., 1116, 1117
Turner, K., 1134, 1136
Twombley, 1016
Twort, 1026
Twort, C. C., 1115, 1118
Tyler, A., 852, 859
Tyndall, J., 562, 1118
Tyzzer, 1006, 1039

U

Ubbelohde, A. R., 271, 279, 280
Uber, 830
Ubisch, L. von, 856, 859
Uhlenhuth, P., 979, 999
Ulbricht, 893
Underhill, F. P., 916
Underwood, E. J., 598
Ungar, 482, 486, 511
Ungar, G., 929, 950
Ungerer, 518
Upchurch, S. E., 1116
Urazovskŭ, S. S., 373
Urban, F., 701, 706
Usher, Archbishop John, 591
Ussing, 544

V

Valentin, G., 616, 618, 621, 622, 628, 650, 660, 667
Valeton, 445, 446, 455, 457
Valkó, E., 326
Volotov, E. N., 823, 849, 850
Vance, A. W., 169, 182, 186, 235
Van der Mensbrugghe, 15
Van Hook, A., 513, 518
Van Niel, C. B., 558, 599, 603, 607, 609
Van Rysselberghe, P., 115, 120
Vanselow, W., 496, 512
Van Slyke, D. D., 535, 587, 599, 751, 752, 950, 1119, 1135
Vars, H. M., 950
Vasilin, 1021
Vauguelin, 444, 456
Vegard, L., 326
Veil, S., 518
Veldman, H., 687, 706, 707
Velich, 996
Vennesland, B., 686, 705
Venturello, G., 467, 471
Verberg, G., 1125, 1135
Vermeulen, D., 609
Verworn, 864
Verzär, F., 599, 701
Verzár, K., 759, 762
Vessie, P. R., 599
Vestin, R., 688, 705
Vetter, H., 693, 705
Vickery, 544
Victor, E., 761
Vigneaud, V. de, 597, 704, 729, 730, 738, 1160
Vigoureaux, P., 339, 373
Villari, 289, 291
Vilter, R. W., 696, 706
Vilter, S. P., 696, 706
Vincent, S., 752
Vinci, Leonardo da, 458
Virchow, 1043
Visscher, M. B., 370, 1012, 1088
Vitek, V., 609
Vladimirskii, V. V., 372
Vlodrop, C. van, 265
Vogt, W. W., 308
Voegtlin, C., 1037, 1153
Vold, Majorie J., 266, 274
Vold, Robert D., 266, 274
Volk, R., 978
Vogt, H., 1153
Von Pirquet, 985
Vorländer, 278
Vries, Hugo de, 584, 808

W

Waals, van der, 241, 303
de Waal, H. L., 1115
Waddell, J., 598, 717, 737
Waddington, C. H., 577, 599
Wagner, C., 496
Wagner, K. H., 733, 738
Wagner, R., 973, 978, 1123, 1126, 1135
Wagenschein, M., 356, 373
Wald, George, 702, 707, 731, 753, 755, 756, 757, 758, 760, 761, 762
Walker, 600
Walker, H. M., 1096
Walker, J. W., 1117, 1118
Walker, R. D., 443, 511
Wallen-Lawrence, 744, 752
Waller, 1148
Walter, J., 236, 252
Walter, W. G., 1154
Walters, W., 1118
Walton, 375, 380
Walton, E. T. S., 387
Walton, R. P., 667, 682
Wampler, Roy W., 100
Warburg, O., 557, 563, 600, 601, 602, 603, 606, 609, 693, 694, 697, 700, 707, 722, 737, 775, 785, 1001, 1050
Ward, 570
Ward, A. F. H., 107, 115, 119
Warner, C. R., 939, 940, 946, 950
Warner, E. D., 956
Warren, 1017
Warren, B. E., 234, 310, 326
Warrick, F. B., 560
Washburn, E. W., 257, 265
Wassink, E. C., 602, 609
Watanabe, S., 373
Waterman, H. I., 257, 258, 265
Waters, E. T., 956
Watson, C. C., 428, 434, 567
Watson, D. W., 409, 411
Watson, M. M., 1134, 1136
Wasserman, A., 979, 985
Waugh, D. F., 869, 875, 880, 883
Wayne, E. J., 950
Weakley, Chas. E., Jr., 337
Weatherford, H. L., 938, 951
Weaver, H. M., 1154
Weaver, W., 412, 433, 434
Webb, J. H., 477, 478, 489, 511, 512
Weber, H., 411
Weber, H. H., 894, 895, 896, 897, 898, 899, 900
Webley, D. M., 706
Weden, H., 951
Weed, A. J., 427, 433
Weed, L. H., 1154
Wegefarth, P., 1154
Weidenhagen, R., 705
Weidling, K., 1152
Weil, A., 1138, 1154
Weil, A. J., 371, 979, 982, 985
Weil-Malherbe, H., 688, 689, 690, 707
Weimarn, P. P. von, 10, 12, 286, 561, 600
Weinland, 456
Weinstock, H. H., 729, 738
Weiskotten, H. G., 950
Weismann, A., 592, 593
Weiser, H. J., 457
Weiss, 1022
Weiss, C., 938, 951
Weissberger, A., 495, 511, 512
Weith, A. J., 285
Welker, 1004
Wells, A. A., 1116
Wells, D., 1118
Wells, H. G., 497, 943, 951, 963, 979, 1161, 1162
Wells, M. W., 1118
Wells, P. V., 12
Wells, W. F., 1104, 1112, 1115, 1118
Wendt, G., 705
Went, St., 1129, 1136
Werge, J., 511
Wergin, W., 646, 648, 649, 661, 664, 667
Werkman, C. H., 609, 686, 691, 705, 706
Werner, 631
Werner, A. A., 745, 752
Wert, L. R. van, 331, 337
Wertheim, 680, 682
Wertheimer, D., 703, 704
West, 1002
West, C. J., 1135
West, J., 139, 145
Westenbrink, H. G. K., 687, 688, 707, 721, 737
Westerfeld, W. W., 686, 705, 727, 737
Westerkamp, R. E., 1125, 1135
Westphal, O., 705
Wexler, R., 1129, 1135
Weygand, F., 698, 705, 1118
Wheatley, A. H. M., 691, 706, 1153
Wheeler, E., 630, 667
Wheeler, S. M., 1118
Whetstone, R., 10, 12
Whipple, D. V., 707
Whisler, B., 1118
Whistler, R. L., 649, 667
Whitaker, D. M., 853, 859
Whitby, 308
White, 544
White, A., 409, 411
White, A. H., 6, 12, 130
White, E., 1118
White, J. F., 471
White, M. J. D., 822, 850
White, P. R., 586, 600
Whitney, M. E., 955
Whitney, W. R., 5, 445, 456
Whytlaw-Gray, R., 1118
Wickwire, G. C., 599
Wideröe, R., 387
Widgoff, L., 517
Wiegand, 308
Wieland, H., 759, 762
Wieland, T., 692, 705
Wiegner, G., 362, 373
Wieler, A., 649, 658, 659, 661, 667
Wiener, 788, 894, 896
Wiener, O., 421, 434
Wiesner, 864
Wiesner, J., 621, 622, 628, 638, 660, 667
Wigand, A., 667
Wightman, E. P., 511
Wylard, J., 721, 737
Wilander, O., 957
Wilder, R. M., 728, 738
Wilder, T. S., 1117, 1118
Wile, V. J., 235
Williams, 457, 598
Williams, E. F., 1120, 1134, 1135
Williams, I., 285
Williams, J. W., 409, 411, 426, 428, 433, 434, 567, 964, 977, 978

Williams, R. D., 728, 738
Williams, R. E., 756, 761
Williams, R. J., 3, 12, 729, 738
Williams, R. R., 684, 693, 705, 706, 707, 721, 728, 729, 737, 738
Williamson, A. E., 1118
Willis, Thomas, 750
Willstatter, Richard, 557, 603
Wilm, D., 193
Wilman, H., 130
Wilson, 38
Wilson, C. W., 659, 662
Wilson, D. A., 279, 1116
Wilson, E. B., 826, 864
Wilson, G. B., 823
Wilson, J., 572, 574, 599
Wilson, J. N., 460, 461, 471
Wilson, T. C., 331, 337
Wilson, V., 1115, 1118
Wiltshire, M. O. P., 746, 752
Winmill, T. F., 95
Winterstein, A., 471
Winterstein, H., 1143, 1154
Winslow, C.-E. A., 1118
Winternitz, M. C., 919, 949, 951, 956
Winters, M., 990, 994
Wintersteiner, O., 544, 747, 752
Winzler, R. J., 704
Wisansky, W. A., 706
Wislicenus, H., 496, 512
Wissler, A., 807
Wisselingh, C. van, 620, 630, 667
Wissler, Clark, 520
Withrow, L., 252
Wodehouse, R. P., 1118
Woelisch, E., 886, 887, 899, 952, 957
Woerdeman, N. W., 600
Woglom, W. H., 999, 1002, 1050
Wohl, K., 603, 609
Wohlbach, S. B., 1096
Wöhler, 610
Wolbach, S. B., 730, 732, 733, 734, 738
Wolf, E. P., 951
Wolf, M., 602, 609
Wolfe, 456
Wolff, 682
Wolff, H. G., 579, 599
Wolkow, M., 600
Wolman, 1022
Wong, S. Y., 951
Wood, H. G., 691, 707
Wood, J. E., 930, 949
Wood, L. A., 119
Wood, R. W., 337, 340, 346, 351, 352, 363, 368, 370, 371, 373
Woods, H. J., 538, 544
Woodside, M. R., 733, 738
Wooley, D. W., 583, 600, 704
Wooley, J. G., 728, 738
Woolley, 1013
Working, E. B., 957
Wormall, A., 978
Wortis, S. B., 1145, 1154
Wright, G. G., 978, 1022
Wright, H. D., 1118
Wright, K. A., 119
Wright, Sewell, 820, 830, 835, 850
Wrinch, D., 840, 850
Wu, Hsien, 372, 978, 979
Wulf, 515, 518
Wulff, P., 511
Wurmser, R., 603, 609
Wuth, O., 1154
Wybauw, L., 929, 951
Wyckoff, R W. G., 233, 326, 409, 411, 418, 430, 433, 434, 791, 794, 796, 798, 800, 807, 808
Wyk, A. J. A. van der, 115, 120
Wyman, J., 979
Wyman, P., 88

Y

Yaglou, C. P., 1118
Yagoda, H., 469, 471
Yamagawa, 589, 600, 1017, 1028
Yanagihara, A., 514, 518
Yang, H. H., 466, 471
Yang, H. S., 1118
Yannet, H., 1146, 1152, 1153, 1154
Yano, T., 511
Yasuda, S., 582, 600, 583
Yokoyama, Y., 1120, 1136
Yoshida, 1003
Young, 560, 582, 1003
Young, G. T., 10, 12
Young, T. F., 101
Young, W. T., 373
Yutzy, H., 504

Z

Zachariasen, W. H., 310, 326
Zacher, H., 663
Zappasodi, P., 1118
Zechmeister, L., 459, 464, 466, 471
Zellat, J., 1116
Zeller, E. A., 692, 707
Zerfas, L. G., 724, 737
Zezulinski, V. M., 373
Ziesché, H., 1118
Ziff, M., 409, 410, 955, 957, 1120, 1130, 1135
Zima, O., 721, 737
Zimmer, 842
Zimmerman, P. W., 598
Zimmermann, 922
Zinsser, Hans, 975, 979, 984, 1161, 1162
Zisman, 42
Zisman, W. A., 572, 598
Zo Bell, C. E., 1118
Zocher, H., 789, 808
Zollman, H., 58, 100
Zozaya, J., 1139, 1153
Zsigmondy, R., 10, 333, 336, 350, 580, 584, 880
Zuker, Fr., 192
Zweifach, B. W., 10, 12, 918, 951
Zwicky, F., 280
Zworykin, V. K., 130, 154, 169, 182, 186, 213, 214, 215, 235

Index to Subjects

A

Abbe theory of microscope, 165
Aberration, chromatic, 161
Aberration, diffraction, 165
Aberration, spherical, 161
Abnormalities, natural and induced, 584
Acacia, use in hypoproteinemic edema, 906
Accelerating lenses, 163
Accretion, growth by, 11
Acenaphthene, 586
Acetaldehyde, 560, 561
Acetobacter suboxidans, 558
Acetoin, 727
Acetone, 21
Acetyl aneurin, 692
Acetylcholine, 578, 579, 580, 691, 692, 745
Acetylcholine, release of, 898
Acetylene black, 206, 435, 438
Acetylene tetrabromide, 21
Achaete, gene, 824
Acid aura, of protozoans, 1165
Acid-amine mixtures, 74
Acide cellulique, 618
Acid-forming foods, 719, 720
Acridine, 588
Acridine, as sensitizer to light, 588
Acrylic ester polymers, 281
Action currents in nerves, 764
Action potentials in nerves, 588
Activated state, 238
Activators, chemical, 578
Activation, energy and volume changes of, 240
Activation, in polymerization, 283, 284
Active muscle, changes in, 883, 884, 885
Active muscle, heat in, 884, 885
Adamsonia fibers, 654
Adaptation, 1175
Adaptive enzymes, 561
Adenocarcinoma, 996, 997
Adenosine triphosphatase, 533
Adenosine triphosphate, 689, 672, 884, 885, 898
Adenylic acid, 884, 885
Adhesion, under pressure, 333
Adrenal cortical hormones, 701
Adrenal cortex, 740, 747
Adrenal medulla, 740
Adrenalin, 745
Adrenals, in age, 1060
Adrenotropic hormone, 744
Adsorbents, 461, 462
Adsorption, 97
Adsorption and antibody production, 962
Adsorption and crystallization, 443
Adsorption isotherm, 449, 971, 972
Adsorption isotherm, new, 98
Adsorption of gases on solids, 434
Adsorption, selective, 457
Adsorption, specific, 968
Adsorption theory of photographic development, 500
Adsorption, types of, 453
Aeneas, sculpture by Bernini, 1051
Aerobiology, 1100
Aero-emphysema, 1082, 1085
Aerosols, 109
Aerosols, coagulation of, by ultrasonics, 356, 360
Aerosols, from respiratory activities, 1101
Aerosols, infective, 1099
Aerosols, physical stability of, 1105
Aerosols, viability of, 1106
Aetiozymase, 688
Agate, 516
Agave fibers, 654
Age and cancer, 1036
Age and diet, 1062
Age and resistance, 1062
Ageing and wisdom, 1053
Ageing, nature of, 1051
Ageing of unicellular organisms, 1052
Ageing, progressive changes in, 1052
Agglutinates, 965
Agglutinin, egg, 852
Aggregation in colloids, 115
Aggregation of paraffin chain compounds, 266
Aggregations in gases, by ultrasonic waves, 356
Aggregations in liquids, by ultrasonics, 357, 364
Air-borne infections, experimental, 1113
Air, disinfection of, 1109, 1111, 1114
Air filters, to prevent infection, 1109
Air-locks, in electron microscope, 180
Air-raid shelters, air in, 1111
Air sampling, 1101
Akanthias, 1063
Alactoid contraction of muscle, 888
Albinism, 817
Albite, 541
Albumin-euglobulin ratio, 1160
Albuminocholia, 1080
Alcohol addicts, 727
Alcohol, and time sense, 782
Alcohol, hydrogen bonds in, 96
Alcohols, catalytic production of, 553
Alcohols, effects of on monolayers, 69
Aleuronat, 944
Alfalfa mosaic virus, 786, 806
Alga, sexual change in, 570
Algae, cellulose in, 637
Alkali metal emulsions, 350
"Alkaline" foods, 720
Alkalized catalysts, 552, 553
Alkaptonuria, 588
Allel, 816
Allelocatalytic reaction, 1165
Allelomorph, 816
Allergen, 979
Allergy, 979 *et seq.*
Allergy, heredity, 588
Allium cepa, 621
Allotropic forms, 6
Alloxazine mononucleotide, 699
Alumina, as carrier, 552
Alums, adsorption of dyes by, 447
Amandin, constants of, 566
Amalgamation of iron, 334
Ameba, and acid, 1164
Ameba cytoplasm, 876, 878
American dietaries, 728, 729
Amidol, 487
Amine, in ultracentrifuge, 432
d-amino acid oxidase, 698, 699
Amino-acid residues of proteins, 536, 537, 543
o-aminoazotoluene, as carcinogen, 1003
p-aminobenzoic acid, 702, 730, 739
2-Amino-5-diethylamino-toluene, 483
p-Aminodiethylaniline, 483
p-Aminophenol, 483
Ammonia, catalytic production of, 550, 551, 552
Ammonium alum, adsorption of dyes by, 447
Ammonium nitrate explosives, 249
Amoebae, 862
Amaeboid glia, 1138
Amontons' law, 247
Amorphous state, 308
Ampholytic nature of phosphatides, 1122
Amurotic idiocy, 588
n-Amyl alcohol, 553
n-Amyl bromide, 272
β-amylase, in soybeans, 681
Amylose, 667, 669, 671, 675, 680

Amylose, adsorption of, by cellulose, 668
Amylopectin, 667, 668, 671, 675, 680
Anaerobic glycolysis in cancer, 1001
Analysis, capillary, 459
Analysis, chemical, 457
Analysis, chromatographic, 459
Analysis, ultracentrifugal, 411, 432
Anaphase, 816
Anaphylactic reaction, in guinea pigs, 981
Anaphylactogen, 979
Anaphylatoxin, 983
Anaphylaxis, 979 *et seq.*
Anaphylaxis in various animals, 980
Anaphylaxis, relation to inflammation, 983
Anatase, 97
Anauxite, 193
Androgenesis, 581
Androgen control of prostatic cancer, 590
Anesthesia, Mg and Ca in, 712
Anesthetics, action of, 1143
Angle centrifuge, 420, 428
Aniline emulsions, 351
Animals and anaphylaxis, 980
Animal pole, 853, 856, 857
Anion-active colloidal electrolytes, 102
Anomalous viscosity and ultrasonic waves, 354
Anorthite, 541
Anoxia, in brain, 1144
Anthracene, 465
Anti-blacktongue factor, 693
Antibodies, 957
Antibodies, formation of, 570
Antibodies against cancer, 1004
Antibodies, chemical nature of, 959
Antibodies, in ultracentrifuge, 432
Antibodies, molecular weights of, 959
Antibodies, production of by adsorption, 962
Antibodies, reaction to non-antigens, 963
Antibodies, specificity of, 957
Antibodies, stereo factors in, 958
Antibodies, valence of, 961
Antibody-antigen reaction, 964
Antibody, electrophoresis of, 409
Antibody formation, 841
Antibody molecules, shape of, 964
Anticatalase, layers of, 88
Antidromic impulses, 929
Antigen-antibody films, 961
Antigen-antibody precipitate, composition of, 977
Antigen-antibody reaction, in electron microscope, 230, 231
Antigen-antibody reaction, conditions influencing, 975, 976
Antigen, as mould for antibody, 570, 571
Antigen molecules, shape of, 964
Antigens, 957, 962
Antigens, polyvalency of, 961
Antigens, specificity of, 964
Antigenicity, nature of, 962
Antigorite, 193
Anti-hemorrhagic vitamin, 734
Antilysin, 971
Antipneumococcus serum globulins, constants of, 565, 566
Antirachitic vitamin, 734
Anti-sera, in electron microscope, 224
Antisera, specific adsorption of, by bacteria, 574
Antitoxin, 959
Anti-typhoid serum, 225
Antiviral antibodies, 970
Ants, 595
Ants, sex in, 581
Aplastic anemia, blood in, 408
Apples, 583
Apples, internal cork of, 559
Apples, little leaf of, 559
Applied mathematics, 6
Arachidic acid, 47, 244
Arbacia, 923, 932
Arbacia eggs, 867, 869, 870, 876, 877
Arca, 564
Area of monolayers, 85
Areas of surfaces, determination of, 434
Arenicola, 563, 567
Arginine, 537, 543, 562
Argentine, varves in, 525
Armadillo, polyembryony in, 575
Arrow black, 435, 438
Arsenic sol, flocculation of, by iron oxide sol, 970
Arteries, in age, 1055
Arthus phenomenon, 982, 983
Artifacts, microscopic, 634
Artificial pearls, 518
Artificial silk, 612, 613, 614, 624
Arrhenius equation, 767, 768, 772,, 773, 776, 777, 778, 780, 781, 783
Asbestos, 201
Ascidian blood, 558
Ascophyllum, 638
Ascorbic acid, 703, 730
Ascorbic acid, effect on collagen, 730
Asparagine, 537, 538
Aspartic acid, 537, 538
Aspergilli, 586, 638
Asphyxia, 1141
Asphyxia in guinea pig anaphylaxis, 981
Association (A) polymerization, defined, 281
Astacus muscle fibers, 897
Asterias oocytes, 880
Atavism, 816
Atomic structure of solids, 122
Atta, 581
Attapulgite, 204, 205
Aucuba japonica, 638
Aucuba mosaic virus, 786
Autocatalysis, 858, 1176
Autocatalysis and cancer, 1044
Autonomic nervous system and homeostasis, 993
Autoradiography, 377, 379
Auxins, 583
Aviator's aero-emphysema, 1087
Aviators, conditioning of, 1091
Avidin, 1160
Avogadro constant, 413
Axon reflexes, 929
Axoplasma, 1148
p-azoxyanisole, 272

B

Bacillus anthrasis, 221
Bacillus delbrückü, 687, 690, 700
Bacillus influenzae, 696
Bacillus macerans, 675, 681
Bacillus megatharium, 221
Bacillus Phlei protein, constants of, 564
Bacillus prodigeosus, 806
Backcross, 816
Bacteria, 1183
Bacteria, effects of pressure on, 332
Bacteria, in air, 1105
Bacteria, in electron microscope, 220, 222
Bacteria, invasion by, 943
Bacteria, invasiveness of, 944
Bacteria, specific adsorption of antisera by, 574
Bacterial dissociation, 568
Bacterial flavoprotein, 700
Bacterial sensitivity, 1109
Bacteriodes fundiformis, 569
Bacteriophage, 428, 786, 806
Bacteriophage, in electron microscope, 223
Bacterium tumorfaciens, 1030
Bakelite, 281
Bakelite, etched, 214
Balances, for film study, 35
Ballas, 7
Banana starch, 669
Bananas, 583
Bar gene, 812, 824
Barium stearate, 127
Barium stearate films, 88
Barley globulin *a*, 564
Barley globulin, constants of, 565
Barley starch, 669
Basal metabolic rate, 1052
Base-exchange, 193
Base-forming foods, 719, 720
Bats, flight of, 3
Bauer equation, 515
Beer-time, 782
Bees, sex in, 581
Beetling of cellulose, 630
Beidellite, 193, 197, 198, 199
Belling's hypothesis, 832, 833, 838, 847
Bence-Jones proteins, constants of, 565
"Bends," 1082, *et seq.*
Benign tumor vs. cancer, 995, 998
Bentonite, 193, 197, 198, 200, 201, 202, 203
Bentonite, electric double refraction of, 789
Benzanthracene derivatives and cancer, 1017, 1023

Benzene, 465
Benzene, emulsion of, 346, 351, 352
Benzene, spreading of, 15, 19, 21, 22
Benzoic acid, elimination of, 587
Benzopurpurin, 111
Benzopurpurin dispersions, 355
1, 2-benzpyrene, 466
3, 4-benzpyrene (same as 1, 2-benzpyrene), 590
Benzpyrene and cancer, 1017
Benzpyrene cancer, 1004
Berger rhythm, 579, 765, 766
Beriberi, 683, 685
Beryllium, breakdown of, by cyclotron, 377
Betabacterium vermiforme, 570
Betatron, 380
Betatron, in medicine, 385
Beta vulgaris, 623.
Bier's spots, 921
Bile, 750
Bile pigments, 703
Bile salts, as detergents, 605
Bilharzia and cancer, 1019
Biocatalyst chains, 559
Biocatalysts, 556, 683
Biocatalysts, modification of, 568
Biochemical constitution of cancer tissues, 1001
Bioelectric effects, 579
Bioelectric potentials, 762
Biological action of flavoproteins, 700
Biological action of "gas," 1094, 1096
Biological activity of thiamine and cocarboxylase, 687
Biological effects of bivalent ions, 58
Biological effects of ultrasonics, 370
Biological fluids, surface tension of, 100
Biological polymers, 303
Biological rates, 770
Biological units, 9
Biophores, 593
Bios, 597
Biotin, 597, 684, 702, 729, 1160
Biotin, in cancer, 1002
Biotin, structure of, 730
Bipotential lenses, 163
Birefringence, intrinsic, 754
Birefringence of lecithin, 1131
Bivalent ions, biological effects of, 58
Black carbon, structure of, 128
Black phosphorus, 331
Black powder, 249
Black silver, 506
Blackman reaction, 602
Blacktongue, 693, 697
"Blast," effects of, 1095, 1097
Blastomere, 816
Blastula cells, coherence of, 574, 575
Blastula, divertula from, 575
Blastula, invagination of, 575
Bleeding, use of acacia in, 906
Blood, a fluid matrix, 986
Blood and sea water, 707
Blood, aeration of, 1085
Blood, capillary circulation of, 900
Blood, coagulation of, 951
Blood coagulation, effect of pressure on, 332
Blood, coenzyme content of, 696
Blood, electrophoretic study of, 406
Blood, gas emboli in, 1083
Blood, gases in, 1083, 1085
Blood, in age, 1061
Blood, in insanity, 1139
Blood, in psychiatry, 1154
Blood, maintenance of neutrality, 992
Blood pigments, 563
Blood plasma analysis, 518
Blood plasma, electrophoretic pattern of, 397
Blood, role of, 986
Blood sugar and brain waves, 774
Boarts, 6
Bone, cancellous, 747
Bone marrow, in inflammation, 936
Bone, salts in, 713
Bone, structure of, 583
Bones, as organs, 713
Bones, calcium storage in, 991
Bones, fixation of radio-elements by, 378
Bones, in age, 1060
Bordet-Gengou media, 569
Borna disease virus, 806
Borocetrin, 558
Boron, traces, 558, 559
Bostrychia, 638
Boundary friction, 247
Boundary lubrication, 247
Bragg formula, 127
Brain, and endocrines, 1145
Brain, and vitamins, 1145
Brain colloids and oxygen, 1144
Brain colloids, degree of dispersion, 1138
Brain, electric activity of, 1151
Brain, electrical behavior of, 1149
Brain, in age, 1056, 1057
Brain swelling, 1137
Brain, pH of, 1150
Brain volume and skull capacity, 1145
Brain waves, 765, 766, 772, 774, 778, 1151
Branched molecules, in starch, 667
Brachiopods, 595
Breast cancer, 1009, 1010
de Broglie equation, 515
Bromcamphocarbonic acid, 556
Bromides, action of, 1143
Bromoform, 21
Bronchospasm, in anaphylaxis, 982
Brown-Pearce carcinoma, 1005
Brown-Pearce tumor, 1031
Brownian motion, 175, 298, 414, 554, 871, 1089
Browning, of cauliflower, 559
Brucella antigen, 976
Brucite, 192
Brush heap structure, 115
Buckwheat, as sensitizer against light, 588
Bubbles, collapse of, 348
Bulk rubber, structure of, 293
Bubbles, critical size of, 1086
Bubbles, growth of, 1084, 1086
Buna, 281
Burns, treatment of, 916
Bush sickness, 559
Busycon, 806
Busycon hemocyanin, 226, 228
Butadiene polymers, 281, 303
Butanol in starch separations, 668, 675
Butterfly wing, 191
Butter-yellow, as carcinogen, 1003, 1018
iso-Butyl alcohol, 553
n-Butyl alcohol, 553
n-Butyl mercaptan, 3
Butyl bromide, 272
Butyl rubber, 303

C

Cadmium bromide, 333
Calciferol, 67, 734
Calcification, enzymes in, 560
Calcium carbonate, crystal forms of, 444
Calcium, effects of on cell walls, 866
Calcium, free, in cells, 852
Calcium, in blood coagulation, 952
Calcium, homeostasis of, 991
Calcium infarct, 1067
Calcium, in nutrition, 711
Calcium, in the egg, 852
Calcium oxalate, crystal forms of, 444
Calcium stearate, 47, 60
Calcium stearate films, 87
Calcosphaerites, 611
Calculi, 1063
Collier constant, 507
Calocarus, 566
Canal viscosimeters, 50
Canary pox virus, 806
Cancellous bone, 747
Cancer, 589
Cancer, a growth problem, 1042
Cancer and age, 1036
Cancer and antibodies, 1004
Cancer and autocatalysis, 1044
Cancer and bile acids, 1024
Cancer and cellular autonomy, 1046
Cancer and chronic irritations, 1014
Cancer and complement fixation, 1004
Cancer and diet, 1037
Cancer and differentiation, 1041
Cancer and dyes, 1003
Cancer and estrogens, 1013
Cancer and frogs, 1033
Cancer and hormones, 590, 1013, 1014
Cancer and liver extracts, 1025

Cancer and metazoan parasites, 1019
Cancer and phenanthracene derivatives, 1017
Cancer and physical agents, 1019
Cancer and polycyclic compounds, 1017
Cancer and radiation, 1019
Cancer and regeneration, 1043
Cancer and Roentgen rays, 1019, 1039
Cancer and somatic mutation, 1039
Cancer and tape-worm, 1019
Cancer and unicellular organisms, 1022, 1023
Cancer and viruses, 1030
Cancer and warts, 1027
Cancer and X-rays, 1019
Cancer, ascorbic acid in, 1003
Cancer, avidin in, 1003
Cancer, biotin, 1002
Cancer, carbohydrate metabolism in, 1001
Cancer, causes of, 1005, *et seq.*
Cancer, causes and nature of, 995
Cancer, cell division in, 1021, 1022
Cancer, gradual development of, 1025, 1026
Cancer, heredity in, 1006
Cancer, hormones in, 1006, 1015, 1016, 1025
Cancer, immune mechanisms in, 999
Cancer, in plants, 586
Cancer in plants, developed by a "chemical," 586
Cancer, irreversability of, 1047
Cancer, K isotopes in, 1003
Cancer, kinds of, 996
Cancer, mammary, 1009, 1010, 1025
Cancer, milk factor in, 1006
Cancer, ovarian influence in, 1007
Cancer, prostatic, 1013
Cancer, stereoisomers in, 1002
Cancer, stimulating factors in, 1014
Cancer, theories regarding, 1042
Cancer tissues, biochemical constitution of, 1001
Cancer, treatment by radio-isotopes, 377, 378
Cancer vs. benign tumor, 995, 998
Cancer vs. normal tissue, 999
Cancer-virus, adaptation of, 1001
Cancerogenic substances, mode of action, 1020
Capillaries, action of heat and cold on, 910
Capillaries, control of, 986
Capillaries, dilation of, 901
Capillaries, in inflammation, 917
Capillaries, total area of, 902
Capillary-active substances, and monolayers, 66
Capillary analysis, 459
Capillary blood pressure, 904
Capillary circulation, 900, 918
Capillary endothelium, 902
Capillary injury, effects of, 914, 919, 920
Capillary permeability, increased in inflammation, 920, 927, 947
Capillary pressure and tissue function, 912
Capillary wall, impermeability of, to plasma proteins, 906
Carbohydrate-fat transformation, 692
Carbohydrate metabolism, 684
Carbohydrate metabolism and pituitary, 744
Carbohydrates, in ultracentrifuge, 432
Carbohydrate storage, 988
Carbon blacks, 206, 435, 438
Carbon blacks, adsorption of N by, 435
Carbon, colloidal, 190, 204
Carbon disulfide, spreading of, 17, 19, 22
Carbon disulfide, under pressure, 332
Carbon, structure of, 128
Carbon tetrabromide, under pressure, 330
Carbon tetrachloride, 21
Carbonado, 6
Carbonyl ion, 214
Carboxylase, structure of, 721
Carcinogenetic hydrocarbons, 1017, 1018
Carcinoma, 590
Carcinoma, defined, 996
Carcinomatoids, 1027
Cardio-vascular system in age, 1054, 1055
Carotene, 733
Carotinoids, as "sex stuff," 1171
Carpogenesis, 581
Carriers, for catalysts, 551, 552, 557, 580
Cartilage, calcification of, 560
"Carriers" of infection, 1104
Cassion disease, 1082 *et seq.*
Castration, for prostate cancer, 746
Catalase, 557
Catalase, adsorption bands of, 550
Catalase, constants of, 565
Catalase, multilayers of, 88
Catalysis, 1176
Catalysis, in biology, 545
Catalysis, industrial importance of, 550
Catalysis, in general, 545
Catalysis, kinetic aspects of, 553
Catalysis, origin of term, 546
Catalyst action, nature of, 770
Catalyst carriers, 551
Catalyst chains, 771
Catalyst formation, 851
Catalyst modification, 1179
Catalyst modifiers, 551, 568, 575, 593
Catalyst poisons, 551
Catalyst promotors, 551
Catalyst systems, 781
Catalyst systems or chains, 884
Catalysts, abnormal, 587
Catalysts, adsorption of N by, 436
Catalysts and hormones, 745
Catalysts, formation of, 851, 857, 858
Catalysts, modified, 858
Catalysts, mode of function, 546
Catalysts, speed of action, 781
Catalysts, surfaces of, 123
Cataphoresis (see also electrophoresis), 387
Catatonia, 1145
Catatorulin effect, 687
Cation-active colloidal electrolytes, 102
Catlinite, 328
Cauliflower, browning of, 559
Cavitation, 348, 349, 353, 354, 364, 1084
Celanthrene red, 263
Celery, blanching of, 583
Celery, cracked stem of, 559
Cell theory, 610, 612
Cell surface and extraneous layers, 869
Cell walls, 865
Cell wall of bacteria, 220
Cellobiose, structure of, 637
Cells, thiamine and cocarboxylase in, 687
Cellular autonomy and cancer, 1046
Cells, and milieu, 1162
Cells, extraneous coats of, 865
Cells, polarization of, 764
Cells, surface tension of, 868
Cellulose, 10, 661
Cellulose, amorphous, 653
Cellulose chains, 635
Cellulose, colloid behavior of, 628
Cellulose, colloid binder in, 653
Cellulose, colloidal behavior of, 623
Cellulose, complexity of, 630, 631, 634, 636
Cellulose component of cotton fiber, 648
Cellulose, crystalline nature of, 631
Cellulose derivatives as polymers, 303
Cellulose, electron microscope view of, 649
Cellulose fibers, 208, 211
Cellulose formation, 643, 644, 647, 651, 652
Cellulose, identification of, 616, 617
Cellulose, in ultracentrifuge, 432
Cellulose, model for, 637
Cellulose nitrate silk, 624
Cellulose particles, ellipsoidal, 638, 639
Cellulose, polymerization degree of, 283
Cellulose, refractive indices of, 654
Cellulose, résumé of views on, 659
Cellulose rings, formed by plastids, 643, 646, 647
Cellulose "solutions," 628
Cellulose solutions, viscosity of, 657
Cellulose structure, 632, 634, 646, 648
Cellulose, substances included under, 626, 646, 648
Cellulose, ultramicroscopic examination of, 628
Cellulose, under polarized light, 631
Cellulose under X-rays, 631
Cellulose xanthate, cellulose particles in, 656
Cementing substances in fibers, 649
Centrifugal molecular stills, 258
Centrifugation, high-speed, 411
Centrosome, 816
Cephalin, 1119 *et seq.*
Cephalin, and acids, 1122
Cephalin, and alkalis, 1123
Cephalin, and radiation, 1130
Cephalin, colloidal solutions of, 1121
Cephalin, in blood coagulation, 953
Ceramium, 565, 566
Ceratozamia, 623
Cerebron, 1120

Cerebrosides, 1119, *et seq.*
Cerebrosides, physical chemistry of, 1134
Cesium bismuth iodide, microradiograph of, 151
Cesium iodide, 124
Cetyl alcohol, 67, 243
Cetyl pyridinium chloride, 111
Cetyltrimethylammonium bromide, 112
Chalcocite, 335
Chamberland filters, 580
Chameleon, color changes in, 579
Chara, 623
Character, in genetics, 816
Charge theory of multilayer potentials, 88
Charges on multilayers, 89
Chain polymers, 281
Chemical activators, 578, 739, 745
Chemical analysis, 457
Chemical bond, defined, 815
Chemical changes due to pressure, 332
Chemical effects of ultrasonics, 369
Chemical kinetics and time, 782
Chemical pacemakers, 762, 767
Chemoreception, 581
Chemotactic factor, 922
Chemotaxis, 930
Chiasmata, 816, 827
Chicken sarcoma, 1000
Chicken sarcoma, cause of, 1000
Chicken sarcoma of Rous, 1031
Chicken tumor virus, 806
Chili, varves in, 525
Chilomonas paramecium, 1165
Chimpanzee, elimination of poisons by, 587
Chironomus, 565
Chitin, 611
Chlamydomonas, 1166, 1169
Chlorella, 601, 602, 606, 607
Chlorine, in nutrition, 708
Chlorobutadiene polymers, 281
Chloroform, 21
Chlorofucine, 604
Chlorophyll, 556, 557, 600, 604
Chlorophyll, in ultracentrifuge, 432
Chloropicrin, 1095
Chloroplast, as oxygen producer, 625
Chloroplastin, 604
Chloroplasts, 604, 605, 621, 644, 650, 654, 659
Chlorosis of tomatoes, 558
Chocolate, lecithin in, 1126
Cholangioma, 1003
Cholecystitis, 1080
Cholesterin, in insanity, 1158
Cholesterol, 745
Cholesterol, in calculi, 1073, 1074, 1076
Cholesterol, in liver, 702
Cholesterol sol, 1129
Cholesterols, 465, 466
Cholinergic nerves, 692
Cholinesterase, 579, 580, 692, 745
Cholocystokinin, 741
Chondrus, 637, 638
Chorea, Huntington's, 588
Choriomeningitis, 586
Choriomeningitis virus, 586
Chromatic aberration, 161
Chromatium, 602
Chromatid, 816
Chromatin, 816
Chromatographic adsorption, 745
Chromatographic adsorbent, 462, 463
Chromatographic analysis, 459, 460
Chromatographic solvents, 462, 463
Chromatography, scope of, 464
Chromocenter, 816
Chromomere (defined), 816
Chromomeres, 822
Chromonema, 816, 1182
Chromosome map, 811, 816
Chromosomes, 816
Chromosomes and radiation, 828
Chromosomes, breakages of, 827
Chromosomes, chemistry of, 835
Chromosomes, differentiated structures of, 822
Chromosomes, lampbrush, 831
Chromosome rearrangements, 826
Chromosomes, separation of, 572
Chromosomes, submicroscopic structure of, 840
Chromotropic reaction, 1135
Chronic irritations and cancer, 1014
Chronology, based on tree rings and varves, 519
Chronology, Usher's, 591
Chrysene, 466
Chrysotile, 193, 201
Ciliata, 1169
Circulating hormones, 739
Circulation, capillary, 900
Citrus, die-back of, 559
Citric acid cycle, 725, 726, 736
Clapeyron equation, 436
Clay minerals, in electron microscope, 192
Clay, rheopexy in, 366
Clays, structure of, 192
Cleveland Hospital Clinic accident, 1095
Cliff dwellers, building dates, 520
Clivia nobilis, 631
Clone, defined, 573
Closterium, 634
Clot inhibition, 954
Clumping of infusoria, 1166
Clupein, 562
Coacervation, 8
Coagulation of blood, 951
Coagulation, sonic, 361, 362, 363, 364
Cobalt hydroxide, 517
Cobalt, traces, 559
Cocarboxylase, 721
Cocarboxylase, action of, 689, 690
Cocarboxylase as catalyst, 691
Cocarboxylase, formation of, 688
Cocarboxylase, structure of, 686
Cockroach heartbeat, 769
Cockroaches, 595
Coefficient of friction, 247
Coenzyme, 722, 723, 883
Coenzyme I, 693, 696, 724
Coenzyme II, 693, 696
Cohesive colloids, 10, 574, 585
Colchicine, production of chromosome mutations by, 586
Cold-blooded animals, 770
Collagen, 531, 541, 543, 730
Collagen and ascorbic acid, 730
Collagen, banded structure of, 838
Collagen, electron micrographs of, 230, 232
Collodion photographic plate, 473
Colloid binder in cellulose, 653
Colloid chemistry of viruses, 785
Colloid, origin of term, 561
Colloid osmotic pressure of plasma, 903, 906
Colloid principles in calculus formation, 1065
Colloid properties of cell membranes, 630
Colloidal aspects of antigen-antibody reaction, 966, 969
Colloidal carbon, 204
Colloidal electrolytes, 102
Colloidal electrolytes, non-hydrolyzable, 105
Colloidal gold, 229
Colloidal gold, in electron microscope, 174, 188
Colloidal particles in solutions, 114
Colloid protection, removal of, 955
Colloidal solutions of lecithin, 1121
Colloidal state, generality of, 561
Colloidality, optimum, 1157
Colloidality, zone of maximum, 553
Colloids and crystallization, 444
Colloids and lipides, 1126
Colloids, diffusion of, 556
Colloids, effect of, on crystallization, 611
Colloids, sonic coagulation of, 365
Colloids, viscosity of, 236
Colon bacteria, 224
Coloring fruits, 583
Colpidum, 1165
Colpoda, 862
Complement fixation, 970
Complexes, of rhodopsin, 759
Compound microscope, invention of, 613
Compressed vapor films, 26
Compressibility of monolayers, 46
Complexity of egg white, 403
Conalbumin, 401, 403, 404
Concanavalins, constants of, 565
Concentric layers in cotton fibers, 629, 630, 646, 648
Concretions, 1063
Condensation (C) polymerization, 281
Condensed films, 69
Condensed liquid phase, 30, 244
Conductance, of brain, 1149
Conductivity of colloidal electrolytes, 107

Confectionary, lecithin in, 1126
Cones and rods, 754
Confera aera, 618
Conic structures, in soaps, 279
Conjugated antigens, 958
Conjugation, 816
Connective tissue, in age, 1060
Constitutative enzymes, 561
Contagion, spread of, 1114
Contraction of muscle, 890
Contraction of myosin and phosphorylasis, 898
Convection, in electrophoresis, 392
Convulsive disorders, 1140
Copolymerization, 281
Corallina, 638
Corn starches, 669
Corpora amylacea, 1064
Cortin, 924
Cotton vs. cellulose, 628
Cotton fiber, cellulose portion of, 648
Cotton fiber, disintegration of, 654
Cotton fibers, refractive indices of, 654
Cotton fiber, growth of, 629, 630, 639, 640
Cotton, impurities in, 629
Coughing, and infection, 1102
Coulomb forces, 967
Copper in blood of horse-shoe crab, 558
Copper, in bloods, 558
Copper, in development of oysters, 580
Copper, in enzymes, 558
Copper in feathers, 558
Copper, in nutrition, 717
Copper, traces, 558, 559, 580, 717
Cori ester, 560
Corn, white bud of, 559
Corpuscular proteins, 529, 533
Corpus luteum, 741
Corticotropic hormone, 744
Corticosterones, 747
Cotton, 9
Cotton fibers, electron micrographs of, 209, 211
Cow sera, 566
Cozymase, 688
Cramps, in muscles, 710
Creaming of toluene dispersion by ultrasonics, 358
Creatine phosphate, 722
Creation, The, 591
Creep, 334
Cresyl blue, oxidation of, 780, 781
Cretinism, 588
Crickets, chirping of, 769
Cristobalite, 319
Critical radius, of bubbles, 1086
Critical state, in gases, 1090
Crocetin dimethyl ester, *cis* and *trans*, 1171
Crocin, as activator, 570
Crooked molecules, in monolayers, 64
Crossing-over, 810, 816, 826
Croton oil, 930
Crotoxin, constants of, 565
Crown gall, 586
Crystal forms of paraffin chain compounds, 267
Crystal growth, 860, 1177
Crystal growth, theories of, 444, 445
Crystal habit, changes in, 443, 444
Crystalline nature of starch, 673
Crystallites, 624
Crystallization, as a stepwise process, 561
Crystallization at rising temperatures, 861
Crystallization by ultrasonics, 369
Crystallization, effect of colloids on, 611
Crystallization, globular, 611
Crystallization, lethal effects of, 863
Crystallization phenomena, 561
Crystals, adsorption by, 968
Crystals and crystallization, 443
Crystals, forces in, 968
Cucumber mosaic virus, 786, 806
Cumingia, 871
Cuprammonium, as solvent for cellulose, 617, 623
Cuprammonium, for cellulose, 657, 658
Curd fibers, in soap, 273
Curly gene, 812
Curtin, 630, 634
Cyanidin, 583
Cyanine dyes, 485, 486
Cybotactic groups, 561
Cyclic distillation, 260
Cyclidium, 1165
Cyclostomata, 563
Cyclotron, 373, 380
Cyclotron, in medicine, 376, 378
p-Cymene, 21
Cystine, 537, 538
Cysts, in bones, 713
Cytochrome C, 563
Cytochrome C, constants of, 564
Cytochrome C reductase, 699
Cytochrome systems, 774, 775
"Cytoflav," 698
Cytolysis, general, 881, 882
Cytoplasm, 816
Cytoplasm and chromosomes, 594
Cytoplasm, surface activity of, 877
Cytoplasmic (defined), 816
Cytoplasmic inheritance, 1183

D

Daguerreotype, 472, 510
Dahlia, 623
Dahlite, 713
Dalmatian dogs, inheritance in, 588
Dalmatian dogs, spots on, 582
Daltonism, 2
Daniglacial period, 523
Danysz phenomenon, 969
Daphnia, 563, 566
Daphnia eggs, 881
Dasypus hybridus, 575
Dates, fixing, 519
Dauermodificationen, 1048
Debye-Hückel theory, 105, 790
Debye-Scherrer patterns, 641
Decahydronaphthalene, 552
Decalin emulsions, 352
Deficiency, in genetics, 810, 813
Degassing liquids by ultrasonic waves, 367
Dehydro-ascorbic acid, structure of, 731
7-Dehydrocholesterol, 465
Dehydrogenase systems, 695
Dehydrogenases, in egg, 856
Deletion, 810
Dementia praecox, 1138
Denaturation of viruses, 800
Denatured proteins, 530, 534
Dendritic growths, 496
Density, in photography, 509
Deposition of films, 91
Dermacentor variabilis, 569
Dermatitis venenata, 921
Dermatosomes, 621, 622, 624, 638
Desensitizing dyes, 487
7-deshydro-cholesterol, 734
Desiccation of organisms, 862
Desoxycholic acid and cancer, 1024
Desoxyribonucleoproteins in chromosomes, 833, 845
Desoxyribose, 836
Desquamation, and infection, 1101, 1104
Determination of particle size and weight, 412
Determination of surface areas, 434
Determinants, 593
Detergency, factors in, 118
Detergents, 102
Detonation velocities of explosives, 249
Detoxification mechanisms, 587
Deuteron beams, 375
Devaux effect, 876, 879, 880, 881, 882
Developing agents, 483, 490
Development, in photography, 472
Devitrification, 859, 861, 863
Dextran, as cement, 570
Dextrine, 616
Dextrin, Schardinger, 681, 682
Diabetes and inflammation, 945
Diabetes insipidus, 749
Diabetes mellitus, 748
Diakinesis, 816
Diaminodurene, 483
Diamond, 6
Diamond, as a macromolecule, 291
Diamond, as adsorbent, 968
Diamond, surface energy of, 102
Diaphorase, 699
Diastase, acting on starch, 555
Diathermic effects of ultrasonic waves, 369
1, 2, 5, 6-dibenzanthracene, 590

2, 3, 6, 7-Dibenzanthracene, 465
Dibromanthracene, 86
Dichroism of rhodopsin, 754
Dickite, 193, 194, 195, 203
Dicoumarin, 951, 957
Dictyostelium discoideum, 574
Die-back, of citrus, 559
Dielectric properties of phosphatides, 1122
Diet and cancer, 1037
Diet and life span, 1062
Differentiation, 576, 578, 816, 851, 852
Differentiation and cancer, 1041
Differentiation and the orderly course of life, 572
Differentiation, maintenance of, 858
Differential diffusion, 457
Diffraction aberration, 165
Diffusion, 250
Diffusion constant, 413
Diffusion constants of proteins, 564
Diffusion constants, of viruses, 795, 803
Diffusion, differential, 457
Diffusion gradient, 573
Diffusion rings, 470
Diffusion through capillaries, 908, 918, 920
Digestion hypothesis of antibody formation, 963
Digestive system, in ageing, 1053
Digitonin, and rhodopsin, 759, 760
Dihydroquinonate ion, 476
Dihydro-riboflavin, 722
Dihydroxyacetone phosphate, 560
Diisoamyl, 21
Dilatancy, 367
Dilatant systems and ultrasonics, 366
Dimers, defined, 281
Dimethylaminoazobenzene, as carcinogen, 1003
2, 4-dimethylpentanol, 553
Dinitrophenol, 774
Dinosaurs, 595
Dioctyl, 2 ethylhexanol, 109
Diphosphopyridine nucleotide, 693, 694, 722, 724
Diphtheria, and brain, 1137
Diphtheria antitoxin, 977
Diphtheria bacilli, 225
Diphtheria toxin, 960
Diploid, 817
Dipole bonds, 967
Direction of chemical change by catalysis, 545
Dismutation, 691
Dispersions of sediments, etc., by ultrasonic waves, 352
Dispersions of solids by ultrasonic waves, 353
Dissociation of bacteria, 568
Dissociation of proteins, 562
Distance activators, 739
Distillation, cyclic, 260
Distillation, vacuum, 253
Diver's palsy, 1082
Diving, 1091
Divining rod, 3
Dodecyl sulphonic acid, 105
Dogs, anaphylaxis in, 982
Dominant character, 809
Donnan effect, 415, 432
Donnan equilibrium, 399, 480, 709, 739, 748, 762, 904
Dopa, 582
Double bonds, *cis* and *trans* arrangements, 66
Double bonds, effects of, 44
Double refraction change in muscle, 894
Double refraction, of virus, 787, 788, 793
Double refraction, sonic, 360
Dowsers, 3
Drone bee, 581
Droplet infection, 1100, 1101
Dropping mercury electrode, 607
Drop-weight method, 100
Drosophila, 572, 582, 584, 585, 810, 811, 824, 1187
Dust, and infection, 1101, 1104
Dust control, 1109
Duplex films, transformation of, 21, 82
Duplex spreading, 15
Dwarfism in plants, 583
Dye solubility in detergents, 108
Dyes, adsorption of, 464, 465
Dyes, adsorption of, by crystals, 445
Dyes and cancer, 1003
Dyes, diffusion of, 922
Dyscrinisms, 740

E

Ear, in age, 1058
Eberthalla typhosa, 224, 225
Echinoderm egg, 870, 873
Ecitron, 581
Ectoderm, 853
Ectoderm, development of, 575
Ectromelia virus, 806
Edema, of brain, 1142, *et seq.*
Edemas, from diet, 710
Edestin, 535, 537
Edestin, constants of, 566
Edsall-Muralt sol, 895, 898
Egg albumin, constitution of, 537
Egg albumin molecule, size of, 806
Egg albumin, sedimentation of, 419
Egg-white injury, 729
Egg-white, poached, 535
Egg-white toxicity, 684
Egyptian cotton, concentric layers of, 629
Ehrlich mast-cells, 954
Eicosanol, 49
Einstein equation, 236
Einsteinian universe, 2
Eka-iodine, 379
Elaidic acid, 65
Elastic after-effects of pressure, 334
Elasticity of rubber, etc., 286
Elastomers, synthetic, 303
Electric currents, safe vs. dangerous, 772
Electric phenomena in biology, 579
Electrically driven ultracentrifuges, 420
Electrocardiograph, 579
Electrochemical effects of ultrasonics, 370
Electro-depositing metals and ultrasonic waves, 353
Electroencephalograms, 765, 766
Electroencephalography, 1151
Electrography, 469
Electromagnetic sound generators, 339, 343
Electron, 9
Electron microscope, 838
Electron microscope view of cell membranes, 649
Electron density summations, 142
Electron diffraction, 120
Electron diffraction by small crystals, 124
Electron diffraction of dyed sodium bromate, 452
Electron gun, 177
Electron lenses, 154
Electron lenses, aberrations in, 161
Electron lenses, theory of, 156
Electron micrographs of silver sols, 476, 477, 500
Electron microscope, 152, 273
Electron microscope, applications of, 188
Electron microscope as a diffraction camera, 121
Electron microscope, construction of, 175
Electron microscope, focal depth, 190
Electron microscope in microbiology, 220
Electron microscope, interpretation of results, 153, 154, 166
Electron microscope, observations on crystals by, 455
Electron microscope, preparation of specimens for, 185
Electron microscope, resolving power of, 172, 175
Electron microscope, scanning type, 213
Electron microscopic examinations, in photography, 489, 490, 491, 506
Electron optics, 154
Electron probe, 213
Electrons, acceleration of, 381, 382
Electrons, deflection of by material, 170
Electrons, high speed, 380
Electrostatic electron microscope, 184
Electrostatic objective lens, 163
Electrophoresis apparatus, 389
Electrophoresis of proteins, 387
Electrophoretic method, warning about use of, 401
Electrophoretic mobilities of serum components, 960
Electrophoretic study of blood, 406
Eledone, 566
Elements, isotopes of, 1198, *et seq.*
Elephantiasis, 943
Elimination curve, in molecular distillation, 260
Ellipsoidal cellulose particles, 638, 639
Elon, 483, 490

Eluants, in chromatography, 462, 463
Elution, 462, 851, 853
Embryonic development, 851, 1041
Emphysema, prevention of, 1090
Emulsification by sound waves, 346
Emulsions, oriented wedge theory of, 100
Emulsofor, 103
Enamel, mottled, 559
Enation mosaic virus, 786
Endocrine glands, 739
Endocrine system, 740, 741
Endocrine system, in age, 1058
Endocrines and brain, 1145
Endocrines, 596
Endocrinology, 751
Endoderm, 853
Endoderm, formation of, 575
Energy carriers, 603
Energy change of activation, 240
Energy of expansion of monolayers, 76
Enterogastrin, 741
Entropy change of activation, 241
Enzootic ataxia, 559
Enzyme pacemakers, 767, 771
Enzyme-polysaccharide combination, 569
Enzyme, "yellow," 557
Enzymes, 739
Enzymes, action on starch, 680
Enzymes, adaptive, 561
Enzymes and hormones, 741
Enzymes and vitamins, 683
Enzymes, assembly of, 722
Enzymes, colloidal dispersion of, 855, 857
Enzymes, constitutive, 561
Enzymes, dispersion of, 851, 857
Enzymes, effects of pressure on, 332
Enzymes, electrophoresis of, 410
Enzymes, formation of, 851, 857, 858
Enzymes in photosynthesis, 606
Enzymes, nature of, 781, 851
Enzymes, poisoning, 777
Enzymes, "synthetic," 556
Enzymes, ultramicroscopic view of, in action, 555
Enzymic action, speed of, 772, 774
Enzymic phosphorylation of thiamine, 688
Enzymic relations of vitamins, 702
Epidemics, spread of, 1114
Epigenesis, 593, 817
Epilepsy, 1139
Epistasis, 588, 817
Equatorial plate, 817
Equine encephalomyelitis virus, 786, 806
Equisitum, 623
Equivalence zone, 974
Ergosterol, 734
Erythema, 921
Erythrocruorin, constants of, 564, 565
Erythrocruorins, 563
Erythrocruorins, constants of, 566, 567
Escherichia coli, 569, 771
Eskimos, counting by, 5
Estrone, 747
Ethyl acetate, melting of, 329
Ethyl alcohol, 560
Ethyldichloroarsine, 1095, 1098
Ethyl iodide, 21
Ethyl margarate, 47
Ethyl-*p*-azoxybenzoate, 272
Ethyl monylate, 21
Ethyl palmitate, 47
Ethylene bromide, 21, 22
Ethylene gas in ripening fruits, 583
Ethylene polymers, 281
Eucolloidality, 553
Eugenics, 817
Euglena, 604, 862
Euglobulin, 957, 959, 960
Euglobulin, in insanity, 1155 *et. seq.*
Eupolymers defined, 281
Evolution, 584
Evolution, basis of, 591
Evolution, physicochemical basis of, 591, 1173
Excelsin, constants of, 566
Expanded liquid films, 23
Expanded liquid phase, 27
Expanded monolayers, 69
Explosives, detonation velocities of, 249
Extraneous coats of living cells, 865
Extraneous layers of cells and protoplasm, 869
Exudates, in inflammation, 922, 924, 925, 926, 935, 939, 940, 948
Eye color and hormones, 582
Eye, interior layers of, 754
Eyesight, in age, 1057

F

Factor W, 693
Fagopyrism, 588
Falling-film still, 257
Famine edema, 710
Faraday-Tyndall effect, 562
Fat from protein, 725
Fat-soluble vitamins, 731
Fat storage, 990
Fatty acid phases, 28
Fatty acids, crystals of, 269
Feldspars, 541
Fenn effect, 887
Ferric oxide, 123
Ferrous oxalate, 490
Fertilization, 851, 852
Fertilization membrane, 865
Fertilizin, 852
Fessenden oscillator, 343
Feulgen test, 831, 837, 843
Fiber structure, 628
Fibers, intercuticular substances in, 627
Fibrin, 951
Fibrinogen, 951
Fibroin, silk, 531
Fibrocytes, 938
Fibrous proteins, 529, 534
Film balances, 35
Film deposition, 91
Film lubrication, 246
Film potentials, 41, 90
Film pressures, 21
Film pressure, determination of, 35, 37
Films, effect of pressure on, 64
Films, formation, 14
Films, mixtures in, 69
Films of antigen-antibody, 961
Films of mixed acids, 72
Films, rubbing, 128
Filters and viruses, 796
Filters, permeability of, 580
Filtration, of cancerous substance, 999
Finely divided solids, determination of surface, 97
Fmiglacial period, 523, 525
First-aid in "gas" attack, 1097
Fish-liver oil, vacuum distillation of, 263
Fishes, color change in, 579
Fizeau-Foucault fringes, 895
Fixation, in inflammation, 941, 944, 947
Flagellates, 1169
Flat-fishes, color changes in, 579
Flat lying molecules, 49
Flavin-adenine-dinucleotide, 724
Flavobacterium, 569
Flavoproteins, 698, 699, 700
Flotation, principles of, 100
Fluorescence, in photosynthesis, 608
Fluorescent screens, in focusing, 178
Fluorine, in bone, 713
Fluorosis, 559, 583
Focus, in electron microscope, 170, 172
Focusing, in electron microscope, 176
Fogs, formation of, by ultrasonic waves, 351
Folic acid, 730
Food, benefits from, 683
Foods, acid- and base-forming, 719, 720
Foot-and-mouth disease virus, 806
Forces, long-range interparticle, 790
Formic acid and starch, 673
Formic acid, catalytic breakdown of, 551
Formica rufa, 581
Formulae, bases of, 6
Formvar, 187
Formvar films, 217, 219
Foucault-Toepler schlieren effect, 388, 389, 391
Fourier series, 133
Fourier summations, 136, 141
Fourier syntheses, 132, 139
Fowl leukemia, 1002
Fowl plague virus, 806
Fowls, elimination of benzoic acid by, 587
Fractionation of exudates, 935
Fraunhofer diffraction, 166

Free calcium in cells, 852
Free energy of spreading oils, 18
Freezing, 860, 862
Freezing, heat of, 84
"Frenching" of orange trees, 580
Fresnel diffraction, 166, 172
Friction, boundary, 247
Friction, coefficient of, 247
Frictional coefficient, 412
Frictional coefficient of viruses, 794, 803
Friedaender's bacillus, 569, 930
Fringed micells, 635
Fringed micellar hypothesis, 660
Frogs, cancer in, 1033
Fructofuranose-6-phosphate, 560
Fructofuranose-1, 6-diphosphate, 560
Fucus, 637, 853, 855
Fungi, mutations in, 586
Fusobacterium, 222

G

Galactose, 561, 622
Gallium, and moulds, 10, 558
Gallstones, 443, 1063 *et seq.*, 1068, 1069
Gallstones, prevalence of, 1070
Galton whistle, 345
Gamete, 817
Gas bubbles, liberation of by ultrasonics, 347
Gas current vibration generators, 339, 345
Gas emboli, in blood, 1083, 1085
Gas interchange between tissue and capillaries, 910
"Gas," in war, action of, 1094
Gaseous films, 25
Gases, adsorption of, by solids, 434
Gases, in ultracentrifuge, 432
Gases, influence of on stability of Hg sols, 350
Gasoline, lecithin in, 1126
Gastric juice, 750
Gastric juice, in age, 1053
Gastridium, 615, 650
Gastrula formation, 575
Gaussian distribution, 295, 297, 301
Gaussian image, 167
Gaussian probability, 509
Geeldikkop, 589
Gel, defined, 291
Gelatin, composition of, 543
Gelatin, constitution of, 538
Gelatin dynamite, 249
Gelatin, non-antigenicity of, 963
Gelatin, swelling of, 480
Gelation, theory of, 291
Gels, dispersion of, by ultrasonic waves, 352
Gene, action of, 815
Gene, as a colloid unit, 808
Gene, as autocatalytic catalyst, 820
Gene, defined, 817
Gene, nuclear function of, 841
Gene reproduction, 830
Gene, size of, 806
Generators for sonic and ultrasonic waves, 338
Genes, 1181
Genes and cancer, 1007, 1008, 1039
Genes as chemical units, 819
Genes, chemical nature of, 848, 1184, 1185
Genes, linear order of, 821
Genes, mutations in, 584
Genes, size of, 1187
Genic balance, 814
Genic loci, stability of, 834
Genic map, 811
Genic substances, 1182
Genom, 817
Genotype, 817
Geochronology, 524
Geological factors in evolution, 595
Geological formations, Liesegang rings in, 516
Geology, schools of, 501
Germ plasm, 817
Germicidal aerosols, 1110
Gerontology, 1051
Gibbs-Donnan equilibrium, 904
Gibbs law, 457
Gibbs theorem, 116
Gibbsite, 193
Gigartina, 637
Glacial clay deposits, 522
Glass beads, adsorption of N by, 441
Glass, under pressure, 333
Glasses containing hydrogen bridges, 323
Glasses of various oxides, 319
Glasses, silicate, 317
Glassy state, 308
Glassy sulfur, 335
Gliadin, constants of, 564
Glioma, 1137
Globular proteins, 529
Globulin antibody fractions, 960
Globulins, fractionation of, 959, 960
Globulins, in blood, 406
Globulins, in ovalbumin, 401
Glucone ogenesis in inflammation, 945, 948
Glucopyranose-1-phosphate, 560
Glucopyranose-6-phosphate, 560
Glucose glass, 323
Glucose storage, 988
Glucose, structure of, 637
Glucose, yeast fermentation of, 559, 560
Glutamic acid, 537, 538, 695
Glutamine, 537, 538
Glutamine, to eliminate poison, 587
Glyceraldehyde phosphate, 560
Glyceric acid-2-phosphate, 560
Glyceric acid-3-phosphate, 560
Glyceride fats, vacuum distillation of, 265
Glycerin, made by fermentation process, 560
Glycine, 543
Glycogen, 10, 560, 574, 668, 884, 885
Glycolosis, in cancer, 1001
Glyptals, 281
Goiter, 588
Goiter, and iodine, 559
Gold ruby glass, color changes in, 584
Gold sols, effect of ultrasonic waves on, 350
Golgi structures, 1148
Gonads, in age, 1058
Gonadotropic hormone, 744
Gonococcus, 569
Gongylonema neoplasticum, 1019
Goose fish, 1151
Gotiglacial period, 523, 525
Grain, in photography, 506
Graphite, as adsorbent, 968
Graphite dispersions, 353
Green plants as oxygen producers, 625
Grey silver, 506
Growth, 11
Growth energy of tumors, 998
Growth hormone, 744
Guanidine chloride, 562
Guinea pigs, anaphylaxis in, 981
Gun powder, 249
Gurney-Mott hypothesis of photographic development, 478, 489, 492, 497, 498
Gutathione, 731
Gynandromorph, 817
Gypsum dispersions, 353

H

Haas' flavoprotein, 699
Haber process, 550, 552
Habit vs. habitat, 591
Haemophilus parainfluenzae, 696
Halicystis, 615, 641, 642, 643, 650, 651, 653, 659, 660
Halloysite, 193, 195, 203
Haploid, defined, 817, 818
Hapten, 571
Haptens, 958, 964, 966
Haptogen membrane, 880
Harker summations, 136
Hartmann whistle, 361
Hassal bodies, 1064
Hay fever pollens, 1105
Hearing, in age, 1058
Heart, in age, 1054, 1055
Heart muscle flavoprotein, 699
Heart, sinoaricular node of, 765
Heat denaturation of lipides, 1132
Heat edema, 911, 912
Heat of freezing, 84
Heat of serological reactions, 966
Heat of spreading, 84
Heat, thermoelastic, 886
H-ion concentration and nervous tissue, 1142
Helical configuration in starch, 675
Helium, in conditioning aviators, 1091

Helmholtz double layers, 472, 1139
Helix pomatia, 562, 566, 567
Hematite dispersions, 353
Hemerythrins, 563
Hemicelluloses, 622
Hemocyanin, 226, 562, 563
Hemocyanin, split by ultrasonics, 356
Hemocyanin, ultracentrifugation of, 415, 424
Hemocyanins, constants of, 566, 567
Hemoglobin, sedimentation of, 418
Hemoglobin molecule, size of, 806
Hemoglobin, synthesis of, 717
Hemoglobins, 534, 535, 563
Hemoglobins, constants of, 565
Hemophilia, 588
Hemophilus influenzae, 569
Hemorrhage and splenic contraction, 986
Hemosiderin, 717
Heparin, 951, 954, 983
Hepatoma, of rats, 1003
Heptaldehyde, 21
Heptane, 21
Heptyl alcohol, 19, 22
n-Heptyl alcohol, 553
Heredity and constitution, in insanity, 1156
Heredity, in cancer, 1006
Herpes simplex virus, 806
Heteroauxin, 586
Heterochromatin, 826, 845, 846
Heterochromomeres, 845
Heterotypic division, 817
Heterozygote, 817
Hexadecane, spreading of, 18, 21
Hexadecyl alcohol, 78
Hexadecylamines, 113
Hexol numbers, 1134
Hexose monophosphate, 694
Hexose-6-monophosphate, 884
n-Hexyl alcohol, 553
High-power X-rays, from betatron, 385
High pressures, effects of, 327
High-pressure phases, 31
High pressure vapor films, 26
High vacuum distillation, 253
Hinny, 591
Histamine, effects of, 920
Histamine, effect of on skin, 915
Histamine, in anaphylaxis, 983, 984
Histamine, in inflammation, 922, 923, 930
Histidine, 537
Histiocytes, 938
Histones, 840
Hofmeister series, 714, 1124, 1125
Hofmeister series, in nervous tissue, 1142
"Holes" in liquids, 239
Holmes rib grass strain of mosaic virus, 594
Homarus, 566
Homeostasis, 985 *et seq.*
Homeostasis by regulating processes, 992
Homeostasis by overflow, 991
Homeosynapsis, 817
Homogentisic acid, 588
Homologous (in genetics), 817
Homozygous, 817
Homunculus, 593
Hoof deformation, from selenium, 559
Hooke's law, 296, 299
Hordein, constants of, 564
Horizontal float film balance, 37
Hormomer, defined, 741
Hormone action, mode of, 739
Hormones, 579
Hormones, adrenal cortical, 701
Hormones and cancer, 590
Hormones and enzymes, 741
Hormones and eye color, 582
Hormones and vitamins, 684
Hormones, electrophorens of, 409
Hormones, in cancer, 1006, 1015, 1016
Hormones, in evolution, 596
Hormones, isolation of, by electrophoresis, 407
Hormones, selective adsorption of, 745
Hormones, synthesis of, 742
Horror autotoxicus, 962
Horse antibodies, 959, 960
Horse eucephalomyelitis, 786, 806
Horse sera, 566
Horse-shoe crab blood, 558
Hospital beds, separation of, 1108
Hospital infections, 1113
H substance of Lewis, 920
Huggin's theory of photographic development, 497, 500
Human chromosomes, 815
Human time sense and temperature, 783
Hunger edema, 710
Huntington's chorea, 588
Hybrids, sterility in, 591
Hydranths, development of, 854
Hydrogen bond in monolayers, 95
Hydrogen bonds, 29, 241, 247
Hydrogen bonds in serological reactions, 967, 968
Hydrogen bridges in glasses, 323
Hydrogen bridges in low-molecular organic compounds, 325
Hydrogenation of hydromatic compounds, 551, 552
Hydrohumors, 579
Hydromatic compounds, hydrogenation of, 551, 552
Hydroquinone, 483, 490, 502
Hydroquinone monobenzyl ether as enzyme inhibitor, 582
Hydrostatic pressure, effects of, 327
ω-Hydroxydecanoic acid, 47, 50
Hydroxylamine, 483, 502
Hydroxyproline, 543
Hymenogeny, 880
Hypericism, 588
Hypericum, 589
Hyperphosphatemia, 743
Hypoproteinemic edema, use of acacia in, 906
Hypersensitiveness, 979, 980, 984
Hypochlorite, for war gas, 1097
Hypoglucemic reaction, 988
Hypoglycemic shock, 1141
Hypotonic edema, 710
Hypnotics, action of, 1143
Hysteresis, 334

I

Ice age, 524
Ice-cream, lecithin in, 1126
Ice in frozen organisms, 862
Ice, various kinds of, 330
Idaein, 583
Idants, 593
Identical twins, 817
Idioplasm, 619
Ids, 593, 660
Igepal, 103
Image contours, in electron microscope, 170
Immersion lenses, 163
Immune mechanisms in cancer, 999
Immunity and inflammation, 941
Immunology, 570, 957 *et seq.*
Immunology, electrophoresis studies in, 409
Impedance, in muscle, 896
Impurities, importance of, 10
Incense, to sterilize air, 1111
Incubation period, in anaphylaxis, 980
Indene polymers, 284, 303
Independent assortment, in genetics, 809
β-indole acetic acid, 586
Infants, protection against air-borne infection, 1108, 1114
Infarct, calcium, 1067
Infective aerosols, 1099
Infective aerosols, control of, 1107
Inflammation, 918 *et seq.*
Inflammation and bacterial invasion, 943
Inflammation and immunity, 941
Inflammation, cellular sequences in, 938
Inflammation, relation of anaphylaxis to, 983
Inflammatory exhudates, protein fractionation of, 935
Influenza infection, 1113
Influenza virus, 786, 806
Inhibition of clotting, 954
Inhibition zone, 974
Inhibitors, in photography, 487
Inorganic ions, adsorption of, 467
Inorganic polymers, 303
Inositol, 583, 730
Inositol phosphoric acid, 711
Insanities, classification of, 1161
Insanity, 1136
Insects, polyembryony in, 575
Insulin, 534, 535, 742, 748

Insulin, constants of, 565
Insulin, zinc in, 11
Insulin and carbohydrate storage, 988
Insulin shock, 1141
Intercellular cement, in capillaries, 901
Interchange (in genetics), 817
Intercranial pressure, 1140
Intercuticular substance in fibers, 627, 646
Interfaces, intermolecular forces at, 93
Interfacial conditions in photographic development, 492
Interfacial microtensiometer, 875
Interfacial tension, in cells, 875
Intermediate liquid phase, 28
Intermolecular forces at surfaces and interfaces, 93
Intermediate liquid phase, 27
Internal cork of apples, 559
Interpolymerization, 281
Interpretation of electron micrographs, 154, 166
Intersex, 815
Intestines, in age, 1054
Intrinsic birefringence, 754
Intussusception, growth by, 11
Invagination of blastula, 575, 576
Invasiveness of bacteria, 944
Inversion, 810, 813
Iodine, action of, on capillaries, 928
Iodine and starch, 675, 677
Iodine, in nutrition, 718
Iodine, pleochroism of, 676
Iodine, traces, 559
Iodized salt, 719
Iodoacetic acid, 888, 892, 896
Iodopsin, 756
Iron, absorption of, 718
Iron, amalgamation of, 334
Iron, from carbonyl, 214
Iron, in nutrition, 717
Iron oxide, 123
Iron oxide sol, flocculation of by arsenic sol, 970
Irreversible effects of high pressures, 327
Ischemia of tissues, 912
Isoalloxazine ring, 698, 700
Isoamylalcohol, 22
Isoantibodies, 963
Isoantigens, 963
Isoelectric point of phosphatides, 1122
Isomaltose, synthesis of, 546
Iso-molecular forces, 1178
Isolation, in evolution, 591
Ionization produced by neutrons, 376
Isopentane, 19, 21
Isoprene, polymerization of to rubber by pressure, 332
Isopropyl bromide, 272
Isotopes, 8
Isotope dilution method, 1120
Isotopes, table of, 1198 *et seq.*
Isotopes, in ultracentrifuge, 432

J

Jacobson's organ, 581
Jaundice, 406, 589
Jaundice, blood in, 408
Jelly, vegetable, 616
Jurin's law, 458

K

Kangaroo tendon, 232
Kangri cancer, 589
Kaolinite, 192, 193, 194, 195, 196, 197
Kerasin, 1120, 1134
Keratin, 531, 533, 539, 540
Keratin from wool, constitution of, 538, 539
Keratin-myosin group, 530, 532, 541
Kerr electro-optical effect, of virus, 788
Kidney resorption, 750
Kidney stones, 443, 1064, 1073
Kidney tumors in frogs, 1033
Kidneys, in age, 1056
Kinetic aspects of catalysis, 553
King crab blood, 558
King-snake, reaction of rattlesnake to, 581
"Knocking," by too rapid explosion, 249
Knowledge, relativity of, 2
Kolmer's droplets, 757
Kossel-Stranski theory, 454, 455
Koroseal, 281
Krebs' citric acid cycle, 725, 726
Kubanka wheat, 916
Kundt's dust figures, 356

L

Labile states and ultrasonics, 369
Lactalbumin, constants of, 564
Lactic acid, 884, 885
Lactic acid, in tissues, 685
Lactogenic hormone, 409
Lactoglobulin, 535, 538
Lactoglobulin, constants of, 565
Lactuca, 623
Lamarckism, 592, 596
Lambert-Beer law, 893
Lambs, swayback of, 559
Laminaria, 637, 638
Lampblack, 206
Lampbrush chromosomes, 831, 846
Lampetra, 563, 564, 565
Lange gold sol, 229
Lange test, 1128
Langenbeck cycle, 689
Langmuir's adsorption theory, 473
Langmuir pump, 257
Laplace equation, 164
Latent image, 475
Latent mosaic of potato, 786
Lattice flaws, 271
Lattices in clay minerals, 200
Lauric acid, 59
Lauryl pyridinium lauryl sulphate, 102
Lauryl sulfonic acid, 107
Lazy maize, 583
Lead azide, 444
Lead soaps, 267
Leaves, venation system in, 627
Lecithin, 878, 881, 882, 1119 *et seq.*
Lecithin, and colloidal gold, 1126
Lecithin, as protective colloid, 1126
Lecithin sols, ageing of, 1127
Lecithin, colloidal solutions of, 1121
Lecithin, isolectric point, 1122
Lenses, electron, 154
Lenses, electrostatic, 163
Lenses, from duplex films, 21
Lenses, magnetic, 155
Leptotene, 817
Leucine, 543
Leuconostoc mesenteroides, 570
Leucoplasts, 650
Leukemia, 377
Leukemia, blood in, 407
Leukemia, defined, 1031
Leukemia, of fowls, 1002
Leukocidin, 944
Leukocytes, migration into inflamed tissue, 929, 930, 933, 936
Leukocytosis and leukotaxine, 933
Leukocytosis-promoting factor, 936
Leukoderma, induced chemically, 582
Leukoprotease, 938
Leukotaxine, as permeability factor, 922, 923, 931, 933
Leukotaxine, isolation of, 924, 925
Leukotaxine, properties of, 924
Lewisite, 1095
Lichens, 597, 1193
Liesegang rings, 496, 513
Liesegang rings and calculi, 1072, 1075, 1077, 1080
Liesegang rings, theories regarding, 514, 515
Life and change of state, 861
Life and states of protoplasm, 859, 861
Life, criteria of, 1181
Life, definition of, 1173
Life of bubbles, 1088
Life, origin of, 1189
Light, and sex reactions, 1172
Light-dark field transition, 170
Light-sensitization, 588, 589
Light, sensitization against, 588, 589
Lignin and cellulose, 634
Lily bulb starch, 669
Limulus blood, 558
Limulus hemocyanin, 227, 228, 415, 421
Linear order of genes, 821
Linkage, in genetics, 809, 817

Lipacaic, 1059
Lipides and colloids, 1126
Lipides and dyes, 1130
Lipides, heat denaturation of, 1132
Lipides, physical chemistry of, 1119
Lipohumors, 579
Lippmann emulsions, 147, 207, 210, 491
Liquid crystals, 1134
Liquid paraffin chain compounds, 271
Liquid phase, condensed, 30
Liquid phase, expanded, 27
Liquid phase, intermediate, 28
Liquid surfaces, films, 12
Liquids, degassing of, 367
Liquids, "holes" in, 239
Liquids, viscosity of, 236, 239
Lithium chloride, effects of on eggs, 853, 856
Lithium tritide, 377
Lithium, under pressure, 335
Little leaf of apples, 559
Liver aldehyde oxidose, 699, 700
Liver and protein storage, 990
Liver, chemical changes in, 587
Liver, in age, 1054
Liver, oxidation by, 559
Liver substance and cancer, 1018
Local activators, 739
Locus, in genetics, 817
Lophius, 1151
Lorentz equation, 155, 159
Lorol, 1111
Louping ill virus, 806
Loxocephalus, 1165
Low temperatures and organisms, 783
Lubricating oils, viscosity of, 241
Lubrication, boundary, 247
Lubrication of surfaces, 249
Lucite, 281
Lumbricus, 563, 567
Luteotropic hormone, 744
Luteinizing hormone, isolation of, 407
Lymph, 907, 908, 913, 914
Lymphatic system, in age, 1060, 1061
Lymphatic blockade, 941
Lymphoprotease, 938
Lymphosarcoma, 377
Lymphosarcomatosis, 378
Lysine, 537, 538, 543, 562
Lysis, by phage, 224
Lysolecithin, 1130

M

Maassen filters, 580
MacInnes electrode, 1150
Mackerel, egg oil and cytoplasm, 876, 878, 881
Macromolecules, 9, 635
Macromolecules defined, 281
Macromolecules, formation of, 283
Macromolecules, split by ultrasonic waves, 354
Magnesium-calcium antagonism, 1144
Magnesium bentonite, 197, 200, 201
Magnesium, in nutrition, 711
Magnesium oxide, electron micrograph of, 185
Magnesium, traces, 558
Magnetostriction, 342
Magnetostriction sound generators, 343
Magnetostrictive sound generators, 339, 342
Magnetic lenses, 155
Maize, genetics of, 825
Maize, lazy, 583
Malignant tumors, 995
Malacostraca, 563
Maltase, 546
Mammalian hemoglobin, 563, 565
Mammary cancer, 377
Manganese, traces, 580
Manic depressive type, 1158, 1159
Mannose, 561
Marchi method, 1148
Margaric acid, 47, 85
Margaric nitrile, 29
Mathematics, use of, 5
Masks, to control infection, 1107
Matter, structure of, 1
Maturation division, 817
Mechanical stress, effects of, 327
Mechanoreceptors, 765
Medical use of ultrasonics, 369
Megatherium bacteriophage, 806
Meiosis, 817
Melanin, 582
Melanin formation, 702
Melanphores, 579
Melons, 583
Melting process, in paraffin chain compounds, 271
Membranes, permeability of, 580
Membranes, plant cell, 610
Meruric fulminate, behavior under pressure, 336
Mendel's first law, 808
Mendel's second law, 809
Mercury, action of on iron, 334
Mercury, emulsification of, 346, 349, 363
Mercury fulminate, 249
Mercury isotopes, separation of, 257
Mercury, spreading of liquids and solids on, 20
Mercury vapor, as photographic developer, 472
Mercuric sulfide, red and black forms of, 331
Mercury telluride, 333
Mesitylene, 21
Mesoderm, 853
Mesoderm, formation of, 575
Mesomorphic forms in paraffin chain compounds, 271, 279
Mesomorphic forms of sodium soaps, 273, 279
Mesomorphic state, 266, 309
Mesomorphic state, of virus, 786
Mesotron, 9
Metabolism, in cancer cells, 1001, 1002
Metabolism and electric potentials, 762
Metabolism, rate of, 858
Metabolism rate, and organization, 858
Metabolites as organizers, 854
Metakentrin, 407, 408
Metallic oxides, in electron microscope, 174, 185, 186
Metals, emulsions of, 346, 347, 349, 353
Metals, microradiographs of, 150
Metals, mixtures of, by ultrasonics, 369
Metals, under pressure, 334
Metamyelocytes, 934
Metaphase, 817
Metastases, 995
Metastases in cancer, 1028
Metazoa, cells of, 873
Metazoan parasites and cancer, 1019
Mating behavior, in infusoria, 1166
Medullated nerves, 1147
Meinicke test, 1128
Membrane effects of phosphatides, 1131
Meteorological factors in evolution, 595
Metathesis, 513
Methanol, catalytic production of, 552, 553
Methemoglobin, 557
Methionine, 537
Methyl iodide, 20, 21
β-Methylamyl alcohol, 553
2-methylbutanol, 553
β-Methylbutyl alcohol, 553
Methylcholanthrene, 590, 1021, 1023
Methylcholanthrene and cancer, 1017
Methylene blue, 111
Methylene blue, adsorption of, 968
Methylene iodide, 21
Methylene iodide, spreading of, 18, 19, 22
α-methylglucoside, 561
2-methyl-1, 4-naphthoquinone, 734
2-methylpentanol, 553
2-methyl-3-phytyl-1,4-naphthoquinone, 734
Metrazol, 1141
Miasmas, 1100
Mica dispersions, 353
Micas, 193
Micellae, 864
Micellar hypothesis, 619, 660
Micells, 9, 619, 621, 622, 632
Micells, fringed, 633, 653
Micells, substructures in, 636
Microbiology, electron microscope in, 220
Micro-chromatography, 467
Micrococcus lysodeikticus, 691
Microchemistry, 614, 659
Micrographs, interpretation of, 166
Microincineration, 1147
Micronex, 204, 205, 206, 207, 208, 435, 438
Microörganisms and nicotinic acid, 693, 696
Microörganisms, in air, 1101
Microradiography, 146
Micro-tensiometer, 875
Microscope, compound, invention of, 613

Microscope condenser, invention of, 614
Microscopic image, light vs. electron, 168
Microscopic techniques, development of, 611
Microsomes, 622, 624, 638
Migrating dipoles, 366
Milieu interne, 985
Milk factor, 1033, 1035
Milk factor, in cancer, 1006
Milk flavoprotein, 699, 700
Milk, lecithin in homogenization of, 1126
Milk, modification of, 6
Milk, ultrasonic treatment of, 343, 344, 351
Miller indices, 125
Minerals, in nutrition, 707
Minerals under pressure, 335
Mitochondria, nature of, 1000
Mitosis, 818
Mixed acids, films of, 72
Mixtures, 8
Mixtures in films, 69
Modification, 573, 593, 858
Modification of biocatalysts, 568
Modification, of catalysts, 1179
Modified catalysts, 858
Modifier gene, 818
Modifiers, 578
Modifiers, abnormal, 587
Molar frictional constants of proteins, 564
Molds, 1105
Molecular arrangement in films, 125
Molecular assemblage in proteins, 562
Molecular coalescence, 611
Molecular disintegration, 611
Molecular orientation, 13, 50
Molecular orientation of phosphatides, 1131
Molecular pattern of myosin, 897
Molecular stills, 253
Molecular structure, determination of, 131
Muscular tissue, colloidal behavior of, 896
Molecular weights of proteins, 564
Molecule, definition of, 1187
Molecules, packing of, 46
Molybdenum, traces, 559
o-Monobromtoluene, 21
Monoiodobenzene, 21
Monolayers, 24
Monolayers, interference by, 217
Monolayers, compressibility of, 46
Monolayers, energy on expansion of, 76
Monolayers, hydrogen bond in, 95
Monolayers, molecular arrangement in, 125
Monolayers, phase relations of, 22
Monolayers, ultramicroscopic examination of, 63
Monolayers, viscosity of, 50
Monolayers whose molecules lie flat, 49
Monomer, defined, 280
Monomolecular films of phosphatides, 1130
Monoploid, 818
Monozygotic twins, 573, 817
Montmorillonite, 192, 193, 197, 198, 202
Morpho cypris, 191
Morphological pacemakers, 767
Mosaic, 818
Mosaic eggs, 576
Mosaics, in maize, 832
Mosic gold suspensions, ultrasonic effects with, 359
Mosaic patches, genic, 739
Moss, frozen, 862
Mottled enamel, in teeth, 559
"Moulding" of antibodies, 570
Mt. Kilimandjaro, glaciation of, 525
Mt. Pelée, 333
Mouse anti-alopecia factor, 583
Mouse heart, speed of beat, 766
Mouse sera, 409
Moving boundary method, 391
Moving boundary method, in chromatography, 468
Moving boundary method in electrophoresis, 388
Mucopolysaccharides, 569
Mule, 591
Mulita, polyembryony in, 575
Multicellular organisms, 1193
Multilayer potentials, 88
Multilayers, 572
Multilayers from monolayers, 87
Multilayers, molecular arrangement in, 125
Musaphaga violacea, 558
Muscle, alactoid contraction of, 888
Muscle, activity of, 883
Muscle, and iodoacetic acid, 888
Muscle and rubber, 887
Muscle, change in double refraction of, 894
Muscle, chemical changes in, 884, 885
Muscle, contraction of, 890
Muscle, impedance in, 896
Muscle, irritability of, 712
Muscle, myosin in, 533
Muscle physiology, 713
Muscle, proteins in, 897
Muscle, pH change in, 887
Muscle, relaxation heat of, 887
Muscle, rigor in, 891
Muscle, tetanic isometric contraction of, 892
Muscle, thermoelastic properties of, 886
Muscle, transparency changes in, 892, 893
Muscle, volume changes in, 890, 891
Muscle, water shift in, 892
Muscular contraction, effect of on capillaries, 912
Mustard gas, 1095, 1098
Muscular cramps, 710
Muscular dystrophy, 733
Mutase system, in yeast, 560
Mutation, 568, 1193
Mutation and cancer, 1039
Mutation, chemicals in, 585
Mutation, defined, 818
Mutation, nature of, 841
Mutations, 584
Mutations, intragenic, 833
Mutual translocation, in genetics, 818
Myasthenia gravis, 579
Myelin, 1134
Myelin forms, 1134
Myelin sheaths, 1143, 1147
Myeloma, blood in, 408
Myogen, 897
Myogen A, constants of, 565
Myoglobin, 563
Myoglobin, constants of, 564
Myosin, 530, 531, 533, 539, 540, 896, 897
Myosin and potassium salts, 898
Myosin, as an enzyme, 898
Myosin, colloidal behavior of, 896
Myosin contraction and phosphorlasis, 898
Myosin, denatured, 894
Myosin from rabbit, constitution of, 538, 539
Myosin micelles, 883
Myosin sols, stability of, 898
Myosin threads, behavior of, 898
Myosin, undenatured, 895
Myristal sulfonic acid, 115
Myristic acid, 26, 27, 41, 43, 47, 70, 85, 243
Myxamoebae, 574, 862
Myxoma virus, 806

N

Nacrite, 193
Naphaquinones, 951
Naphthacene, 465, 466
Naphthalene, 465
α-naphthalene, 583
α-naphthalene acetamide, 583
Naphthalene acetic acid, 583
Naphthalene, electron contour maps of, 546, 549, 550
Natural colloids, 561
Natural selection, 591, 1192
Neat soap, 274
Negative mesotron, 9
Negros, leukoderma in, 582
Nematic state, 266
Nematodes, 862
Necrosin, 938, 939
Neisseria catarrhalis, 1106
Nematodes and cancer, 1019
Neon tube, rhythmic flashing of, 766
Neoprene, 303
Nephrops, 566
Nephrosis, blood in, 408
Nerve impulses, 764
Nerve impulses, transmission of, 580
Nerve model, Lillie's, 765
Nervon, 1120
Nervous influences in homeostasis, 993
Nervous system, in age, 1056, 1057
Nervous transmission, 579
Nervous tissue and short-wave therapy, 1122

Nervous tissue, electrical phenomona in, 1139
Nervous tissue, surface films of, 1139
Nerves, action currents in, 764
Neuberg ester, 560
Neuroaminic acid, 1120
Neurohumors, 579
Neurones, 579
Neurones, as small glands, 579
Neuropsychiatric disorders, 1136
Neurospora, 844
Neutrality, maintenance of, 992
Neutrino, 9
Neutron, 9
Neutron therapy, 377
Neutrons, ionization by, 376
New-born infants, iron in blood of, 717
Newfoundland, varves in, 526
Nickel phthalocyanine, 546, 548
Nicotinamide, 693, 724
Nicotinamide as coenzyme, 723
Nicotinamide ironporphyrin, 697
Nicotinic acid, 692, 693
Night-blindness, 683, 732, 756, 757
Ninhydrin, 954
Nissl bodies, 1147
Nitella, 765
Nitrates, synthetic, 550
Nitric acid, catalytic production of, 550
Nitrobenzene, 21
Nitrobenzene emulsions, 351
Nitrobenzimidazole, 487
Nitrocellulose, 249
Nitrocellulose silk, 624
Nitromannite, 249
o-Nitrotoluene, 21
Nitrogen, adsorption of, 435, 436
Nitrous fumes, 1095
Noise from collapse of steam bubbles, 349
Nonadecanoic acid, 47
Nonadecanol, 49
Non-disjunction, 815, 818
Non-gaseous monolayers, changes in, 32
Non-Newtonian viscosity, 52
Nontronite, 197, 199
Normality, deviations from, 584
Nubecula, 1064, 1065
Nuclear membrane, 871
Nuclear moments of atoms, 1198
Nuclear properties, of elements, 1198 *et seq.*
Nuclei, formation of, 860
Nucleic acids, electrophoretic study of, 409
Nucleohistone, 837
Nucleoli, 846
Nucleolus, 818, 843
Nucleoprotein, in viruses, 786, 798
Nucleoproteins, 1182, 1184
Nucleotides, 836
Nucleus, defined, 818
Nucleus formation, 329
Nujol, 63
Nutrition, minerals and vitamins in, 707
Nylon, 281, 303
Nylon fibers, 890

O

Obelia geniculata, 1022
Obesity and age, 1062, 1063
Objective lens, electrostatic, 163
Obligate parasites, 597
Octadecanol, 244
Octane, 20
Octadecane nitrile, 29, 47
Octadecanol, 32, 33, 49, 54
Octopus, 566
Octyl alcohol, 20, 21, 47
Oedogonium, 638
Oenothera, 825
Oil drops in protoplasm, 872, 876, 877, 882
Oil films, spreading of, 14
Oil fountain, due to ultrasonic waves, 341
Oils, emulsions of, by sound waves, 347
Oil-water interfaces, and adsorption of protein, 879
Oleic acid, 21, 65, 243
Oncotic pressure of plasma, 903, 906, 917
Ontogeny, 818
Oöcyte, 818
Ophiphagous snakes, 581
Opsonins, 930
Optical analysis, in ultracentrifuge, 421, 422
Optical theory of electron lenses, 156
Optical ultracentrifuges, 416, 432
Orange trees, "Frenching" of, 580
Orchidectomy, in cancer, 1013
Orchids, 597
Organic liquids, emulsification of, 346, 349, 351
Organic polymers, 324
Organizers, 576, 577, 853, 854, 856
Organization and metabolism rate, 858
Organization, in embryology, 853, 856
Organization, rate of, 858
Orientation in gaseous systems, by ultrasonics, 356
Orientations in liquids, by ultrasonics, 359
Orientation of molecules, 13
Oriented adsorption, 486
Orinthine, 587
Orthokinetic coagulation, 362, 364
Osmotic behavior of soap solutions, 104
Osmotic pressure of plasma proteins, 989
Osteoporosis, 731, 734
Osteosclerosis, 559, 588
Ostwald viscometer, 795
Otoliths, 1063
Ovalbumin, constants of, 565
Ovalbumin, electrophoretic pattern of, 399
Ovalbumin, purification of, 401
Ovaries, in age, 1058
Ovary, 741
Ovomucoid, 401, 403, 405
Oxidation speed, 780, 781
Oxide glasses, 319
Oxidized oils, 14
Oxidation by ultrasonics, 369
Oxygen-CO_2 balance, 992
Oxygen, inhalation of, 1091
Oxygen production by green plants, 625
Oxygen supply, in brain, 1144
Oxynervon, 1120
Oxytocic hormone, 750
Oxytrichia, 1165
Oysters, development of, 580
Oyster, sex reversal in, 570

P

Pacinian corpuscles, 765
Packing of molecules, 46
Palinurus, 566
Palmitic acid, 41, 47, 85, 243, 271
Palm tree, sex reversal in, 570
Palmitic acid, vacuum distillation of, 265
Paludina, 566, 567
Pancreas, 740
Pancreas and carbohydrate storage, 989
Pancreas, in age, 1054, 1059
Pancreatic juice, 750
Pandalus, 566
Pangen, 818
Pangenesis, 593
Pantothenic acid, 702, 729
Pantothenic acid in royal jelly, 581
Papilloma, 1026
Papilloma virus, 1029
Paracrystalline state, in virus, 786
Paraffin alcohols, 33
Paraffin chain compounds, aggregation of, 266
Paraffin oil, spreading of, 17
Parahormones, 596
Paramecia, 862
Paramecia and cancerogenic substances, 1023
Paramecium, 1162 *et seq.*
Paramecium caudata, 594
Parasympathetic nerves, 746
Parathyroids, 740, 743
Parathyroids and calcium metabolism, 991
Parathyroid and kidney calculi, 1067
Parathyroid, in age, 1060
Paresis, 1138
Paresis, brain waves in, 772, 773, 777
Parrot fever, 1104
Parthenogenesis, 818
Particle shape, 415, 416, 432
Particle size and weight, determination of, 412
Partition functions, 237
Pasteur effect, in cancer, 1001
Patterson summations, 136
Patterson-Harker projections, 139
Patterson-Harker section, 535

Pearlite, 215, 216, 217
Pearls, artificial, 518
Pears, 583
Peas, genetic tests with, 809
Pecan rosette, 559
Pectic acid, 610
Pectic material, 631
Pectic material in cotton fibers, 641
Pectic substances, 617, 624
Pectin, 610, 667
Pectine, 617
Pectose, 617, 620, 630
Peizoelectric sound generators, 339
Pellagra, 695, 696, 1138
Pellagrins, 696
Pelvetia, 638
Pentaerythritol, 143, 144
Pentadecylic acid, 24, 28, 41, 47, 63, 70, 80, 85
Pentnucleotide, 934, 937
Pepsin, activation of, 555
Pepsin, constants of, 565
Pepsin, electrophoretic study of, 409
Peptone shock, 984
Periodicity, in sex reactions, 1172
Permeability of capillaries, and leukotaxine, 927
Permeability of membranes, 580
Peregal, 103
Perikinetic coagulation, 362
Periodic precipitates, 513, 516
Peritonitis, blood in, 407
Pernicious anemia, treatment danger, 1018
Peroxidase, 557
Peritonsillar abscess, blood in, 407
Peroxidase, constants of, 565
Persimmons, 583
Petunia, Golden Rose strain, 582, 583, 591
Phallusia mamillata, blood of, 558
Phagocytosis, 929, 930, 948
Phase changes due to pressure, 328
Phase relations of monolayers, 22
Phase transitions, 22, 51
Phases, two dimensional, 23
pH change in muscle, 887
Penanthracene derivatives and cancer, 1017, 1023
Phenanthrene, 86, 466
Phenobarbital, 774, 1141
Phenotype, 818
Phenylacetic acid, elimination of, 587
p-Phenylene-diamine, 505
p-phenylenediamine, oxidation of, 779
p-Phenylenes, 483, 490
Phenylthiocarbamid, 2
Pheophytin, 605
Philosophy, and science, 1
Phlogosin, 930
Phorphyropsin, 756, 760
Phosgene, 1095
Phosphagen, 889
Phosphatides, 1119, *et seq.*
Phosphatides, and neutral salts, 1124
Phosphatides, dielectric properties of, 1122
Phosphatides, in membranes, 1131
Phosphatides, isoelectric point of, 1122
Phosphatides, molecular orientation of, 1131
Phosphatides, reaction with water, 1121
Phosphocreatinine, 884, 885
Phosphoglucomutase, 560
Phospholipides, ampholytic nature of, 1122
Phosphoratase, 560
Phosphopyruvic acid, 689
Phosphors, 589
Phosphorus, in nutrition, 711
Phosphorus, forms of, 331
Phosphorus metabolism, 743
Phosphorylase, 560, 681
Phosphorylation reactions, 606
Photographion, 893
Photographic emulsions and ultrasonic waves, 352
Photographic emulsions, silver bromide in, 207, 210
Photoelectric effect, 387
Photoelectrons, 497
Photographic development, 472
Photolysis, 500
Photometric constant, 507
Photon, 387
Photopigment solutions, 759
Photopigments, in retina, 755
Photoreception, anatomy of, 753
Photosynthesis, 600, 626
Photosynthesis, rate of, 607
Photosynthin, 604
Phrenosin, 1120, 1134
Phthalic acid crystals, 445
Phthalocyanine, 546, 548
Phthalocyanine, iron, 546
Phthalocyanine, nickel, 546, 548
Phycocyans, constants of, 565, 566
Phyllochlorin, 605
Phyllochlorins, 557
Phylogeny, 818
Physical changes in active muscle, 883
Physical development, in photography, 500
Physicochemical factors in insanity etc., 1136
Physiologic electroneutrality, 719
Physiological rhythms, 762
Phytin, 711
Phytomonas tumefaciens, 586
Picric acid, 249
Picric acid and ameba, 874
Pig sera, 566
Pilocarpine, 746
d-Pimeric acid, 44
Pine, of sheep, 559
Pineal gland, 740
Pineapples, 583
Pipestone, 328
Pisum sativum, 808
"Pitressin," 749, 1140
Pituitary and sugar metabolism, 989
Pituitary, anterior, 740, 744
Pituitary, in age, 1059
Pituitary, posterior, 740, 749, 751
Pituitrin, 901
Placenta, hormone formation in, 741
Plagioclase feldspars, 541
Plansomes, 864
Plant activators, 740
Plant cancer, 586
Plant diseases cured by trace elements, 558, 559
Plant cell membranes, 610
Plant products, recognition of, 610, 611
Plants, sex in, 580
Plantain-eaters, 558
Planorbis, 563, 566
Plastid, defined, 818
Plastid inheritance, 818
Plastids, 625, 648, 661
Plastids, formation of cellulose by, 644, 651, 652
Plasma analysis, by Liesegang ring formations, 518
Plasma membrane, 865, 867
Plasma, oncotic pressure of, 903, 906
Plasma thirst, 916
Plastic films, 52
Plastic flow, 333
Plastic solid phase, 244
Plasticity, 250
Plastids, 1182
"Plawm" trough, 37
Pleistocene in North America, 525
Pleochroism, 676
Pleuropneumonia organism, 806
Pneumococci, 944
Pneumococci antibodies, 960
Pneumonia, blood in, 407
Pneumonia infection, 1113
Pneumococci, kinds of, 569
Pneumococci, transformation of, 569
Pneumococcus, transformation of, 586
Poikilothermous animals, 770
Point mutations, 818
"Poison gas," action of, 1094, 1096
Poisons, 587
Poisons, for catalysts, 551, 552
Poisson distribution, 292
Polar body, 818
Polarizability, of brain, 1149
Polarized light, cellulose under, 631
Poles, in egg, 853, 857
Poliomyelitis infection, 1113
Poliomyelitis virus, 806
Pollen and hay fever, 1105
Pollination of Petunia, 582, 583
Pollopas, 281
Polycyclic compounds, 44
Polycyclic compounds, adsorption of, 466
Polycythemia vera, 378
Polydactylism, 588
Polyembryony, 575

Polyenes, adsorption of, 464
Polygonum fagopyrum, 588
Polyisobutylene, polymerization degree of, 283
Polyglycerol phthalate, polymerization degree of, 283
Polymerization, 280
Polymerization degree of various compounds, 283
Polymers, characterization of, 281
Polymers, organic, 324
Polypeptide chains, 530, 532
Polyphenol oxidase, 556
Polyploidy, 586
Polysaccharides, 569
Polysaccharides, electrophoretic study of, 409
Polystyrene films, 216
Polystyrene, polymerization degree of, 283
Polystyrenes, in ultracentrifuge, 432
Polythene, 281
Polytrichum, 648
Polyvinyl acetate, polymerization degree of, 283
Pondermotive effect of ultrasonics, 356, 359, 365
Pomera, 581
Popcorn starch, 669
Porcupine quill, 540, 542
Pore-size of filters, 796
Porphyra, 637
Position effect, 568, 585, 592, 824
Positive mesotron, 9
Positive organic ions in films, 71
Positron, 9
Posterior pituitary action, theory of, 751
Postglacial period, 523
Potassium acid soaps, 267
Potassium alum, adsorption of dyes by, 447
Potassium chlorate, stability under pressure, 333
Potassium decylate, 119
Potassium, determination of, 1198
Potassium dihydrogen phosphate, 139, 140
Potassium dodecylate, 119
Potassium, in muscle, 763
Potassium, in nutrition, 708
Potassium oleate, 104
Potassium salts and myosin, 898
Potassium-sodium balance, 763
Potassium soaps, 270
Potassium thiocyanate, in egg, 856
Potassium vs. sodium, 751
Potassium sulfate, adsorption of dyes by, 450
Potassium tellurite, reduction of, by bacteria, 225
Potassium trioxalatochromiate, 968
Potato starch, 669
Potential of multilayers, 88
Powder, black, 249
Powders, surface determination of, 97
Precancerous state, 1027
Precipitation zones, 973
Precipitin reaction, 972
Preparation of specimens for electron microscope, 185
Pressor-antidiuretic hormone, 750
Pressure, effect of, on mercuric fulminate, 336
Pressure, effects of, 327
Pressure-temperature relations in films, 80
Progesterone, 741, 747
Proline, 543
Promotors, 545
Promotors, for catalysts, 551
Propagation of polymerization, 284
Propanol, 553
Prophase, 818
Propyl alcohol, 21
n-Propyl alcohol, 553
Propylene glycol, to sterilize air, 1111
Prostate, cancer of, 476
Prostate gland, 741
Prostate, cancer of, 1013
Prostate, in age, 1059
Prostatic cancer, 590
Prosthetic groups, 556, 557, 563, 568, 569, 580, 590, 603, 698, 699, 851, 883, 927
Protamine and insulin, 11
Protamine block theory, 836
Protamines, 836
Protective action of gelatin, 484
Protective colloids, 638
Protection, removal of, 955
Protective colloids in urine, 1065
Protein, adsorption of, 879
Protein fractionation of inflammatory exhudates, 935
Protein molecules, in electron microscope, 226
Protein, sugar production from, 945, 948
Proteins, 529 *et seq.*
Proteins and cellulose, 638
Proteins as antigens, 964
Protein-chlorophyll complex, 603
Proteins, dissociation of, 562
Proteins, electrophoresis of, 387
Proteins, in muscle, 897
Proteins, in ultracentrifuge, 415, 432
Proteins of blood, in insanity, 1154
Proteins, molecular assemblage in, 562
Proteins of egg white, 403
Protein storage, 989
Protein structure, 530
Proteolysis, in inflammation, 926
Proteus, 697
Prothrombin, 951, 952
Prothrombin, activators of, 952
Protochlorophyll, 608
Protococcus, 604
Protoplast, 864
Protoplasm, action of salts on, 874
Protoplasm, consistency of, 871
Protoplasm, defined, 818
Protoplasm, diffusion in, 868
Protoplasm, H-ion concentration of, 873
Protoplasm, oil drops in, 872, 876, 877, 882
Protoplasm, physical properties of, 864
Protoplasm, proteins in, 873
Protoplasm, physical states of, 859, 862
Protoplasmic streaming, 628
Protoplasm, water-miscibility of, 872
Protoplasmic proteins, surface-chemical properties of, 875
Protoplasmic surface layer, repair and permeability of, 870
Protoplasmic surface layer, thickness of, 869
Protoplast, 614
Protozoa, and milieu, 1162
Provitamin D, 734
Pseudoglobulin, in insanity, 1155
Pseudoglobulins, 959
Pseudo rabies virus, 806
Psittacosis, 1104
Psittacosis virus, 806
Psoriasis, 921
Psychiatry, 1154
Psychophysiological time, 782
Pueblo Indians, 522
Pukall filters, 580
Pultusk, meteorite, 1198
Pure mathematics, 5
Purkinje cells, in age, 1056
Purkinje shift, 756, 757
Purple bacteria, 607
Putnam clay, 197
Pyleocystitis, 1064
Pyrene, 466
Pyridoxine, 702, 725
Pyrimadine, 690
Pyrogallol, 483
Pyrophyllite, 193
Pyruvate metabolism, 685, 690
Pyruvic acid, 560, 725, 884
Pyruvic acid oxidase, 699
enol-Pyruvic acid phosphate, 560

Q

Quartz, 203
Quartz peizoelectric oscillators, 339
Quartz vibrators, 339
Quebec, varves in, 525
Queen bees, 581
Quill, of porcupine, 540, 541, 542

R

Rabies virus, 806
Rabbit antibodies, 959, 960
Rabbit papilloma of Shope, 1032
Racemization of proteins, 963
Radar, aural, 3
Radiation absorbed by green leaves, 608
Radiation and cancer, 1019
Radiation and chromosomes, 828, 831, 841
Radiation and mutation, 584, 585
Radiation, to sterilize air, 1109, 1114
Radioactive isotopes, 1199, et seq.

Radioactive isotopes in biology, 376, 377
Radio-bismuth, 376
Radio-calcium, 378
Radioelements, properties of, 1199-1217
Radio-iodine, in medicine, 377, 379
Radio-lead, 376
Radio-phosphorus in medicine, 377
Radio-strontium in medicine, 378
Radio-zinc, 379
Raffinose, 561
Ragweed proteins, 409
Rambouillet sheep, 819
Ranvier's rings, 1148
Rana pipiens, and cancer, 1033
Rare earths, separation of, 468
Rat sera, 409
Rat tail tendon, 232
Rat-tail tendon, collagen from, 531
Rates of reaction, theory of, 236
Rattlesnake, reaction of, 581
Rayleigh criterion, 166
Reaction rates, theory of, 236
Recessive character, 809
Reciprocal (in genetics), 819
Recoil electrons, 385
Recombination (in genetics), 819
Recrystallization under pressure, 334
Red phosphorus, 331
Redox systems, 725
Reduction division, 819
Refractive vs. absorption method, in ultracentifuge, 421
Relativity, 2
Regeneration and cancer, 1043
Relativity, of all knowledge, 2
Relaxation heat of muscle, 887
Renal calculi, 1071
Replica process, 213, 215, 216
Reproduction of genes, 830
Residual valence fields, 453, 454
Resolving power, light vs. electron microscope, 168
Resolving power of electron microscope, 172, 175
Respiratory catalyst, 697
Respiratory mediators, 699
Respiratory proteins, 563
Retinal photopigments, 755
Retinenes, 755, 756
Retrograded starch, 673
Rheumatic fever, blood in, 407
Rheopectic systems and ultrasonics, 366
Rhizobium, 729
Rhizoclonium, 646
Rhodanilic acid, 968
Rhodopsin, complexes of, 759
Rhodopsin cycle, 732, 754, 755, 756, 757, 760, 761
Rhodopsin, dichroism of, 754
Rhodymenia, 637, 638
Rhythmic flashing of a neon tube, 766
Rhythms, physiological, 762, 767
Rhythmic precipitation, 513
Rheopexy, 366
Riboflavin, 690, 697, 698
Riboflavin as coenzyme, 722, 723
Riboflavin-adenine dinucleotide, 699
Rib-grass mosaic virus, 786, 798
Ribonucleic acid, 1184
Ribonuclease, 410
Ribonucleic acids, 843
Rice starch, 669
Ricin, 970
Rickettsia, 806
Rift valley fever virus, 806
Rift valley varves, 525
Rigor, in muscle, 891, 896
Ring worm, 921
Ringer solution, 784
Robinson ester, 560, 694
Rods and cones, 754
Root growth, toxic ions for, 630
Rosa, 818
Roses, chromosome number in, 818
Rosette, of pecans, 559
Rossia, 567
Rot of sugar beets, 559
Rotifers, 862
Rouget cells, 901
Rough forms of bacteria, 568
Rous chicken sarcoma, 590, 1000
Royal jelly and bees, 581
Rubbed films, 128
Rubber and muscle, 887
Rubber and rubberlike materials, elasticity of, 286
Rubber chains, free rotation in, 293
Rubber, carbon for, 205
Rubber hydrocarbon, crystallization of, 561
Rubber molecule, 289, 290
Rubber, polymerization degree of, 283
Rubber, structure of, 289, 293
Rubber, thermoelastic effects of, 287
Ruthenium red test, 638
Rutherford atom, 169

S

Saliva and mouth flora, 1053
Saliva particles, and infection, 1103
Saliva particles, size of, 1103
Sago starch, 669
St. Johnswort, 588
Saliva, 750
Salivary gland chromosomes, 813, 814, 822, 823, 831, 839, 845
Salt balance, 747
Salt, iodized, 719
Salt licks, 711
Sampling, from ultracentrifuge, 431
Sand drown of tobacco, 558
Saprolegnia, 638
Sarcoma, 590
Sarcoma, defined, 996
Sarcoma, fowl, 428
Sarcoma, of fowls, 1031
Sciara, 829
Scanning method, for schlieren, 395
Scanning microscope, 213
Schardinger enzyme, 681
Schardinger dextrin, 681, 682
Schiff base, 690
Schilling hemogram, 1155
Schistosoma and cancer, 1019
Schizophrenia, 1159
Schizophrenia and oxygen utilization, 1145
Schlieren, 421
Schlieren method, 391
Schlieren effect, 389, 391
Schlieren scanning method, 395, 397
Schlieren, Toepler, 461
Schwarz-Ungar hypothesis, 486
Schweitzer's reagent, discovery of, 617
Science and philosophy, 1
Scotch tape, for lifting films, 218
Scurvy, 683, 731
Sea food, iodine in, 718
Sea-urchin egg, 853, 855, 857
Sea water and blood, 708
Secretin, 741
Sedimentation constant, 413
Sedimentation constants of proteins, 564
Sedimentation, convection free, 412
Sedimentation equilibrium, 425
Sedimentation velocity method, 413
"Seed" vs. "blood," 578
Seeds, effect of pressure on germination of, 332
Selective adsorption, 457
Segregation, in genetics, 808
Segregation, to prevent infection, 1108, 1114
Selective adsorption, 851, 853
Selective adsorption of hormones, 745
Selenites, as enzyme inactivators, 775
Selenium, traces, 559
Selenium, vitreous, 318
Self-duplication, 572
Self-welding, 334
Sense cell thresholds, 764
Sense of time, 782, 783
Sense organs, 764
Senses, relativity of, 2
Sensitivity of vision, 756
Sensitization, in cancer, 1027
Sequoia tree calendar, 521, 522
Serine, 538
Serological reactions, heat of, 966
Serological reactions, duplex nature of, 966
Serumalbumins, constants of, 565
Serumglobulins, 565
Sex aggregations, in infusoria, 1166
Sex determination, 814

Sex glands, 740, 746
Sex hormones, 746
Sex, in plants, 580
Sex, in ants, 581
Sex in bees, 581
Sex-limited (in genetics), 819
Sex reactions, and light, 1172
Sex reversal, 570
"Sex stuffs," 1169
"Sex stuffs," chemical nature of, 1171
Sex types, multiple, 1166, 1170
Shark, 748
Sharks, urea in, 596
Sharples air-driven tubular centrifuge, 431
Shearing stress, chemical change by, 335, 336
Sheep, bush sickness of, 559
Sherman-Bourquin diet, 693
Shock, anaphylactic, 979, 980
Shope papilloma, 1003, 1005, 1029, 1044
Shope rabbit papilloma virus, 786, 806
Shortening, 10
Short-wave therapy, in nervous tissue, 1122
Sib (sibling), 819
Silica films, 216
Silica gel, thixotropic, 352
Silica, under pressure, 335
Silica, vitreous, 318
Silicate glasses, 317, 319
Silk, artificial, 612, 613, 614
Silk fibroin, 531
Silk-worm, spinning by, 613
Silver bromide, in photographic emulsions, 207, 210
Silver chromate rings, 516
Silver in developed photograph, form of, 487, 488
Silver oxide, 331
Silver sulfide specks as sensitivity centers, 488
Silver sols, electron micrographs of, 476, 477
Sinoauricular node, 765
Sirius, dense companion of, 327
Size and weight of particles, 412
Size of clay particles, 203
Size of viruses, etc., 806
Skin, in age, 1061
Skin pigment, formation of, 582
Skull capacity and brain volume, 1145
Skunk, odor of, 3
"Slides," for electron microscope, 187
Slime moulds, 573, 574
Slime moulds, differentiation of, 574
Small crystals, electron diffraction by, 124
Smallpox virus, 223
Smectic liquid crystals, 272
Smell, in rattlesnakes, 581, 582
Smell, sense of, 3
Smoke, to sterilize air, 1111
Smooth forms of bacteria, 568
Sneezing, and infection, 1102
Soap, hydrolysis of, 103
Soap solutions, 103
Soap, to combat "gas," 1098
Soaps, 267
Soaps, mesomorphic forms of, 277
Sodium arachidate, 278
Sodium amytal, 1138
Sodium cetyl sulphate, 67
Sodium chloride, octahedral crystals of, 444, 445, 446
Sodium chloride storage, 987
Sodium chlorate, instability under pressure, 333
Sodium dehydrocholate, 110
Sodium desoxycholate, 110
Sodium dodecyl sulphate, 557
Sodium hypochlorite, to sterilize air, 1111
Sodium, in nutrition, 708
Sodium laurate, 275
Sodium myristate, 277
Sodium oleate, 112
Sodium palmitate, transitions of, 273, 275
Sodium rhodanate, 1138
Sodium soaps, mesomorphic forms of, 273
Sodium salts, adsorption of dyes by, 452
Sodium stearate, crystal forms of, 276
Sodium stearate and ultrasonic waves, 354, 355
Sodium sulphite, in yeast fermentation by yeast, 560
Sodium taurocholate, 569
Sodium vs. potassium, 751
Soils and trace elements, 559
Solder, 8
Soldier ants, 581
Solid phase, 31
Solid surfaces, films on, 12
Solid, surface of, 248
Solidification, rheopectic, 366
Solidification, types of, 860
Solids, 22
Solids, atomic structure of, 122
Solids, dispersion of, by ultrasonic waves, 353
Sölnhofen slate, 366
Solutes and crystallization rate on freezing, 860
Solutions of photopigments, 759
Somatic (defined), 819
Somatotropic hormone, 744
Sonic double refraction, 360
Sonic vibration, action of on bacteria, 220, 221
Sonic waves, 337
Sorbitol, 558
Sorbose, 558
***Sorghum*, 847**
Sorghum starch, 669
Sorocarp, development of, 574
Sound, thermal effects of, 367
Sound waves, effects of, 346
Sound waves, making emulsions with, 346
Soybean lecithin, 1126
Soybeans, β-amylase in, 681
Spat, development of, 580
Specialization, 1
Species, described, 591
Species specificity, basis of, 570
Specific autocatalysis, 1176, 1178
Specificity, of catalysts, 1176
Sperm, action of, in egg, 852
Sperm, frozen, 783
Sperm, storage of, 783, 784
Spermatid, 819
Spermatozoa, frozen, 862
Sphalerite, 335
Spherical aberration, 161
Spherical viruses, 801, 802
Sphingomyelin, 1119 *et seq.*
Spierer lens, 634
Spindle (in cytology), 819
Spinning top centrifuges, 427
***Spirillum rubrum*, 602, 604**
***Spirographis*, 567**
***Spirogyra*, 604, 628, 638**
***Spiroptera neoplastica*, 1019**
Spleen, contraction of in hemorrhage, 986
Sport, in biology, 819
Spot tests, 469
Spreading, duplex, 15
Spreading, heat of, 84
Spreading of films, 14
Spreading, theories of, 14
Squeezed films, 64
Stability of genes, 834, 835
***Staphylococcus*, 697, 944**
Staphylococcus bacteriophage, 806
Staphalococcus aureus, 1106
Star gene, 825
Starch, 9, 616, 661
Starch, action of diastase on, 555
Starch, chemistry of, 667
Starch, crystallites in, 673
Starch formates, 673
Starch formation, 650
Starch, fractionization of, 667, 668
Starch gels, 679
Starch, hydrogen bonds in, 679
Starch-iodine complex, 675, 677
Starch, methylated, 668
Starch, molecularly dispersed, 672
Starch pastes, 679
Starch, swelling of, 678
Starch, synthetic, 681
Starch suspensions, 679
Starch, X-ray examination of, 674
Starches, actions of enzymes on, 680
Starches, iodine titration of, 669
Starches, waxy, 669
Static, 5
Statistical mechanical theory of reaction rates, 236
Steady states in the organism, 985
Steam bubbles, noise on collapse of, 349
Stearic acid, 47, 65, 128, 243
Stearic acid, vacuum distillation of, 265
Stearyl alcohol, 65, 70
Stearyl amine, 65

Stearoids, 747
Steatite dispersions, 353
Steel, carbon, 9
Stellaria media, 604
Stephanoptera, 607
Sterility in hybrids, 591
Sterols, vacuum distillation of, 263
Stereoscopic electron micrograph, 190
Stereoscopic micrographs, 168
Sterols, adsorption of, 465
Stilbestrol, (Diethylstilbestrol), 746, 747, 1013
Stilbestrol, in cancer, 1013
Stills, vacuum, 253, 257
Stokes' equation, 236
Stokes' law, 414
Stokes' law and aerosols, 1106
Stokes' law, and viruses, 791, 792
Stomach, in age, 1053, 1054
"Stones," 1063 *et seq.*
Storage, by inundation, 987
Storage by segregation, 988
Straub's heart muscle flavoprotein, 699
Strawberry gallbladder, 1068
Stream double refraction of viruses, 793
Streaming potential, 4
Streptococcus infection, 1113
Streptobacillus moniliformis, 569
Streptococci, mild local reaction to, 944
Streptococcus varians, 602, 604
Stress-strain curve in rubber, 299, 301
Stretch-spinning, by silk worm, 613
Strobotac, 420
Streptococci, electron micrographs of, 220
Structure birefringence, 754
Structure, of collagen, 231, 233
Structure of matter, 1
Structural units and levels, 8, 9
Styrene polymers, 281, 303
Suberin, 630, 634
Subneat phase in soap, 274
Subwaxy phase in soap, 274
Sucrose, 561
Succinic dehydrogenase, 701
Succus entericus, 750
Sugar beets, rot of, 559
Sugars, fermentation of, by yeasts, 560, 561
Sugar production from protein, 945, 948
Sulfur dispersions, 353
Sulfur glass, 335
Sulfur in nutrition, 719
Sulphur, vitreous, 318
Sun spots and tree rings, 519
Sun tan, 582
Supercontraction, of myosin, 897
Supercooled liquids, 308
Supercurd phase, in soap, 277
Superfemales, 815
Superliquid phase, 56
Supermales, 815
Supermicells, 634
Superwaxy phase in soap, 274
Sulfonamide, crystallization of, in body, 444
Suprasonics, 337
Surface activity of cytoplasm, 877
Surface areas, determination of, 434
Surface films, viscosity of, 243
Surface potential, 58
Surface properties of colloidal electrolytes, 116
Surface studies with electron microscope, 211
Surface tension in cells, 875, 876
Surface tension of cells, 868
Surfaces, films on, 12
Survival of fittest, 591
Suspending action, 118
Svedberg protein units, 605
Svedberg unit, 413
Swayback of lambs, 559
Sweat-bee, 148
Sweden, varves in, 524
Swedish time scale, 524, 525, 526
Swedish angle centrifuge, 429
Swelling of digits, 911
Sympathetic nerves, 746
Sympathin, 579, 745
Symplexes, 557, 603
Synapsis, 572, 813, 819
Syndactylism, 588
Synezesis, 819
Synthetic elastomers, 305
Syphilis, tests for, 1128
Systems of embryonic organization, 853, 854, 856

T

Tabby, 613
Tactoids, micro, 789
Taenia crassicollis and cancer, 1019
Tagged atoms, 606
Talc, 193
Talc dispersions, 353
Tapioca starch, 680
Tapioca, waxy starches as replacement for, 680
Tar cancer, 589
Tardigrades, 862
Taste, relativity of, 2
Taste, variation in, 2
Tea, yellows of, 559
Teart of Somerset, 559
Teeth, fluorine in, 559, 583
Teeth, in age, 1060
Telophase, 819
Temperaments, 587
Temperature and nuclei, 860
Temperature and physiological rhythms, 769
Temperature change, effect of on capillaries, 910
Temperature, maintenance of uniform, 993
Temperature, of small units, 547
Tendons, 232
Tergitol, 108
Termination of polymerization, 284
Termites, 597
Tertiary paraffin-chain compounds, 47
Tesla coil, 188, 341
Test cross, in genetics, 819
Testes, in age, 1058
Testis, 741
Testosterone and cancer, 1013
Tetanolysin, 971
Tetany, following parathyroid excision, 743
Tetany, from Ca deficiency, 712
Tetrahydro-*d*-pimeric acid, 44
Tetrahymena, 1166
Tetralin emulsions, 352
Tetraploid, 819
Textile fibers, in electron microscope, 208
Textile mills, static in, 5
Thawing of frozen cells, 862
Theory of antibody formation, 570, 571
Thermal effects of sound, 367
Thermal energy balance in photographic development, 495
Thermatomic black, 435, 438
Thermax, 206, 435, 438
Thermoelastic effects in rubber, 287, 294
Thermoelastic heat, 886
Thermoplastic polymers, 325
Thermostats, for film balance, 39
Theories of spreading, 14
Thiamin, 684, 686, 687, 688
Thiamin as coenzyme, 723
Thiamin deficiency, 685
Thiamin, phosphorylation of, 558
Thiamin pyrophosphate, 721
Thiamin pyrophosphate, non-diffusibility of, 721
Thiazole, 690
Thickness of film, in electron microscope, 169
Thiocol, 303
Thixotropic silica gel, 352
Thixotropic systems and ultrasonic waves, 352
Thixotropy, 246, 872
Threonine, 538
Thrombin, 951
Thrombokinase, 952
Thromboplastin, 952
Thymonucleohistone, 416
Thymonucleic acid, 1184
Thymus, 740
Thymus, effects of, 1060
Thyroid, adsorption of radio-elements by, 378, 379
Thyroid gland, 740
Thyroid gland, iodine in, 718
Thyroid, in age, 1059
Thyroid, treatment of, 377, 379
Thyroglobulin, 562
Thyroglobulin, constants of, 566
Thyrotropic hormone, 744
Thyroxin, 562, 718, 774
Tiglon, 591
Time, and alcohol, 782

Time and chemical kinetics, 782
Time, sense of, 782, 783
Tinea corporis, 921
Tiselius apparatus, 960
Tissue fluid, 907
Tissue function and capillary pressure, 912
Tissue pressure, 909
Tissue thirst, 916
Titanium dioxide, surface of, 97
Tobacco mosaic virus, 226, 229, 230, 785, 787, 795, 796, 797, 799, 800, 806
Tobacco necrosis virus, 786
Tobacco ring spot virus, 786, 806
Tobacco, sand drown of, 558
Tocopherols, 262, 733
Toepler schlieren, 461, 467
Toepler's schlieren, in ultracentrifuge, 421
Toluene, 21
Toluene emulsions, 351, 352
Tomato, absorption of radio-zinc by, 379
Tomato bushy stunt virus, 786, 802, 803, 805, 806
Tomato mosaic virus, 799
Tomatoes, 583
Tomatoes, effect of Mn on, 580
Toning photographs, 505
Touracou feathers, 558
Toxin-antitoxin, 969, 970, 971
Toxin-antitoxin reactions, 970
Toxin, in ultracentrifuge, 432
Toxin of staphylococci, 944
Trace elements, 558, 559
Trace elements and health, 559
Trachea of mosquito larva, 190
Tradescantia, 829, 830
"Training" of yeast, 561
Trait (in genetics), 819
Transformation of bacteria, 568, 569
Transition, light-dark field, 170
Translocation, 810, 819
Transmission electron microscope images, 166
Transmutation of elements, 375
Transplantation of tumors, 996
Transparency changes in muscle, 893
Tree-ring chronology, 519, 526
Tree rings and sun spots, 519
Tree rings, to fix dates, 519
Trehalose, 561
Tremor, 579
Tribulus, 589
Trichites, 496
Tridecylic acid, 47
Tridymite, 319
Triethylcetyl ammonium cetyl sulphate, 102
Trimers, defined, 281
Trinitrotoluene, 249
Triolein, 47
Triose phosphate, 560
Triphosphopyridine nucleotide, 694
Triploid, 819
Tritium, 377
Trophoplasm, 619
Tropins, 930
Troughs, for film study, 35
Tryptase, 951, 953
Tryptogen, 954
Tryptophan, 537, 538
Tubercle bacilli and cancer, 1030
Tubercle bacilli, and localization of dyes, 919
Tuberculosis, air-borne, 1112
Tuberculosis bacillus, electrophoretic study of, 409
Tuberculosis bacillus protein, constants of, 565
Tubularia, 854, 855
Tumor tissue, secretion of hormones by, 741
Tumors, malignant vs. benign, 995
Tumors, transplantation of, 996
Turacin, 558
Two-dimensional phases, 23
Typha, 623
Typhoid bacteria, 224
Tyrosinase, 556, 582
Tyrosinase, and *p*-amino-benzoic acid, 702

U

Ultracentrifugal analysis, 411, 432
Ultracentrifugal constants of proteins, 564 *et seq.*
Ultracentrifuges, 411
Ultracentrifugation of viruses, 791, 803
Ultramicroscopic activity in cells, 573
Ultramicroscopic examination of cellulose, 628
Ultramicroscopic examination of crystallization, 561
Ultramicroscopic examination of gold ruby glass, 584
Ultramicroscopic examination of monolayers, 63
Ultramicroscopic illumination, 880
Ultramicroscopic study of nerves, 1147
Ultramicroscopic view of enzyme actions, 555
Ultramicroscopic visibility, 634
Ultrasonic waves, 337
Ultrasonics, 337
Ultraviolet, for sterilization, 1109, 1110
Unicellular animals, ageing of, 1052
Units, structural, 9
Unipotential lenses, 163
Unit character, in genetics, 819
Uranyl nitrate intoxication, 1142
Urea, 333
Urea, in sharks, 596
Urea, inhibition of fixation by, 942
Urease, constants of, 566
Uremia, brain swelling in, 1137
Uric acid, crystal forms of, 444
Urine, protective colloids in, 1065
Uronic acid, 657

V

Vaccina, 410
Vaccinia virus, 223, 801, 806, 970
Vacuum distillation, 253
Vacuum type ultracentrifuges, 417
Valonia, 621, 638, 644, 646, 651, 652, 653, 654, 762
Valonia cytoplasm, 876
Vanadium catalysts, 550
Vanadium in ascidian blood, 558
Vanadium pentoxide dispersions, 353, 355, 360, 789
Vapor films, 25
Variation, in biology, 819
Variations, 584
Variety, in taxonomy, 819
Varves, to fix dates, 519, 522, 524
Vasopressin, 751
Vector structures, 136
Vegetable jelly, 616
Vegetal pole, 853, 856, 857
Venation system in leaf, 627
Vertical pull film balance, 36
Vesicular stomatitis virus, 806
Vestigial wing gene, 812
Vibrations and bubbles, 1088
Vicia faba, 630
Victrone, 281
Vinegar eels, 862, 863
Vinylacetate polymers, 281
Vinylchloride polymers, 281, 303
Vinylite, 281
Viosterol, 734
Virus, mosaic, 594
Virus, of choriomeningitis, 586
Virus, of papilloma, 1029
Virus protein crystals, 797, 802
Virus proteins, electrophoretic study of, 410
Virus, tobacco mosaic, 226
Virus, vaccine, 223
Virus, yellow fever, 425, 430
Viruses, 1181
Viruses, ageing of, 800
Viruses and cancer, 1030
Viruses and radiations, 801
Viruses and X-ray diffraction, 796, 803, 804
Viruses, anisometry of, 791, 799
Viruses, chemical derivatives of, 799
Viruses, chemical differences in, 798
Viruses, chemical nature of, 848
Viruses, colloid chemistry of, 785
Viruses, cultivation of, 597
Viruses, denaturation of, 800
Viruses, diffusion constants of, 795, 803
Viruses, dispersion of, 800
Viruses, effect of pressure on, 332
Viruses, electron micrographs of, 799, 805
Viruses, electrophoresis of, 799, 804
Viruses, failure of Stokes' law, 792
Viruses, filtration of, 796
Viruses, inactivation of, 801

Viruses, in ultracentrifuge, 432
Viruses, rod-shaped, 787, 797
Viruses, sedimentation velocity of, 803
Viruses, shape of, 787
Viruses, sizes of, 805, 806
Viruses, viscosity of, 794
Viruses, ultracentrifugation of, 791, 803
Viscosity, 51
Viscosity and ultrasonic waves, 354
Viscosity in cellulose solutions, 657
Viscosity, non-Newtonian, 52
Viscosity of colloids, 236
Viscosity of liquids, 236, 239
Viscosity of monolayers, 50
Viscosity of surface films, 243
Viscous flow, 239
Visual purple, 702
Vision, in age, 1057
Visual processes, 753
Visual purple, 732
Visual purple, in ultracentrifuge, 432
Vitality tests, 862
Vitamers, defined, 741
Vitamin A, 683, 702
Vitamin A and calculi, 1067
Vitamin A, distillation of, 263
Vitamin A_2, in vision, 756
Vitamin, anti-hemorrhagic, 734
Vitamin, anti-rachitic, 734
Vitamin B, 721
Vitamin B complex, 701
Vitamin D, 702, 734
Vitamin D_3, 465
Vitamin E, 702, 733
Vitamin E, vacuum distillation of, 264
Vitamin H, 597
Vitamin K, 734
Vitamin requisites, 728
Vitamins, 739, 740
Vitamins, and brain, 1145
Vitamins and enzymes, 683
Vitamins and hormones, 684
Vitamins B, 721, 730
Vitamins, enzymic relations of, 702
Vitamins, fat-soluble, 731
Vitamins, in catalytic systems, 721
Vitamins, in nutrition, 707, 720, 721
Vitamins, questions to be answered, 703
Vitelline membrane, 866
Vitreous state, 859, 862
Vitreous selenium, 318
Vitreous silica, 318
Vitreous solids, 308
Vitreous state, 308
Vitreous sulphur, 318
Vitrification, 859, 860, 862
Vitrification of sperm, 783
"Vitriol," origin of word, 308
Viviparus hemocyanin, 227, 228
Volume change of activation, 240

W

Wallerian degeneration in nerves, 1138, 1148
van der Waals, equation, 303
van der Waals forces, 241
War edema, 710
War gases, action of, 1094, 1096
War gases, burns from, 916
Ward-Meyer bacterium, 570
Water, bad effect of excess, 711
Water balance, 747, 749
Water, distribution in body, 709
Water, freezing of, under pressure, 330
Water lily fibers, 654
Water metabolism, 749, 750
Water of crystallization, squeezing out of, 333
Water, non-spreading of on organic liquids, 18
Water spray, and infection, 1105
Water storage, 987
Water, vitrification of, 861, 862
Warts and cancer, 1027
Wave lengths of sound waves, 338
Wave patterns, 358, 359
Waxy phase in soap, 274
Waxy starches, 669, 680
Weight and longevity, 1062
Weight and size of particles, 412
Weismannism, 592, 593
Werner's theory, 631
Wet plate, in photography, 473
Wetting, 118
Wetting agents, 108
Wheat starch, 669
White bud of corn, 559
Wilson cloud chamber, 376
Wilting coefficient, 916
Wing scales of butterfly, 191
Wood-ant, 581
Wood, distintegration of, 646
Wood, reconstruction of, 637
Wood's metal, emulsification of, 350
Worker bees, 581
Wound healing, 1021
Wound infection, 1113
Wurtzite, 335
Wyex, 435, 438

X

Xanhydrol method, 1138
Xenoplastic graft, 819
Xeroderma, 1028
Xeroderma pigmentosum, 1016
X-ray diagrams of cellulose, 642
X-ray diagrams of proteins, 531
X-ray diffraction pattern of nerves, 1148
X-ray diffraction patterns of phosphatides, 1131
X-rays and cancer, 1039
X-rays, production of, by the betatron, 381
X-rays, structure determination with, 131
Xylan, 668
Xylene emulsions, 351
o-Xylene, 21
p-Xylene, 21

Y

Yeast, carboxylase system in, 688
Yeast fermentation in alkaline media, 561
Yeast, fermentation of glucose by, 559, 560
Yeast flavoprotein, 699
Yeast fumarate hydrase, 699, 700
Yeast, training of, 561
Yellow enzyme, 557, 563, 697, 698
Yellow enzyme, constitution of, 698
Yellow fever virus, 425, 430
Yellows, of tea, 559
Young-Harden ester, 560
Young's modulus, 299, 300
Yucca fibers, 654

Z

Zea mays, 830
Zein, constants of, 565
Zeppelin, as airplane carrier, 1093
Zinc oxide, adsorption of N by, 439
Zinc oxide, surface area of, 440
Zinc, traces, 559
Zimmermann reaction, 922, 923
Zipper, analogy to catalyst, 546
Zone of maximum colloidality, 553, 1157
Zones of precipitation, 973
Zygiosis, 1148
Zygote, defined, 819
Zygotene, 819
Zymoplastic substance, 952