UNIVERSAL
LIBRARY

OU_160180

UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No. 574.19 Accession No. 47606

Author 598 A Symposium on Aspects of Protein structure, 1963, Madras

Title Aspects of Protein structure. ed. by G.N. Ramachandran

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

1963

Aspects of Protein Structure

Aspects of Protein Structure

Proceedings of a Symposium held in Madras
14–18 January 1963 and organized by the
University of Madras

Edited by G. N. RAMACHANDRAN
Department of Physics, University of Madras, India

160180

ACADEMIC PRESS
London and New York · *1963*

ACADEMIC PRESS INC. (LONDON) LTD.
Berkeley Square House, Berkeley Square
London, W.1

U.S. Edition published by
ACADEMIC PRESS INC.
111 Fifth Avenue
New York 3, New York

Copyright © 1963 by ACADEMIC PRESS INC. (LONDON) LTD.

All rights reserved

No part of this book may be reproduced in any form, by photostat, microfilm, or any other means, without written permission from the publishers

Library of Congress Catalog Card Number: 58-4033

Printed in Great Britain by
Spottiswoode, Ballantyne and Company Limited
London and Colchester

List of Contributors

S. AKABORI, *Osaka University, Osaka, Japan* (p. 69)

P. AMBUJABAI, *Marine Biological Station, Porto Novo, India* (p. 343)

N. S. ANDREEVA, *Institute of Biophysics, Moscow, U.S.S.R.* (p. 137)

J. BELLO, *Roswell Park Memorial Institute, Buffalo, New York, U.S.A.* (p. 13)

J. P. BARGETZI, *University of Washington, Seattle, Washington, U.S.A.* (p. 319)

A. BERGER, *The Weizmann Institute of Science, Rehovoth, Israel* (p. 205)

P. R. BLAKEY, *Institute of Technology, Bradford, Yorkshire, England* (p. 103)

E. R. BLOUT, *Harvard University Medical School, Boston, Massachusetts, U.S.A.* (p. 241)

S. M. BOSE, *Central Leather Research Institute, Madras, India* (p. 357)

W. L. BRAGG, *The Royal Institution, London, England* (p. 1)

J. B. ČHATTERJEÁ, *School of Tropical Medicine, Calcutta, India* (p. 301)

YU. N. CHIRGADZE, *Institute of Biophysics, Moscow, U.S.S.R.* (p. 137)

M. CHVAPIL, *Institute of Industrial Hygiene and Occupational Diseases, Prague, Czechoslovakia* (p. 351)

B. ČMUCHALOVÁ, *Institute of Industrial Hygiene and Occupational Diseases, Prague, Czechoslovakia* (p. 351)

N. N. DAS GUPTA, *Saha Institute of Nuclear Physics, Calcutta, India* (p. 301)

F. E. DEJARNETTE, *Roswell Park Memorial Institute, Buffalo, New York, U.S.A.* (p. 13)

J. T. EDSALL, *Harvard University, Cambridge, Massachusetts, U.S.A.* (pp. 179, 371)

P. GANGULY, *Saha Institute of Nuclear Physics, Calcutta, India* (p. 301)

L. D. HAMILTON, *Sloan-Kettering Institute, New York, New York, U.S.A.* (p. 161)

F. HAPPEY, *Institute of Technology, Bradford, Yorkshire, England* (p. 103)

D. HARKER, *Roswell Park Memorial Institute, Buffalo, New York, U.S.A.* (p. 13)

A. J. HODGE, *California Institute of Technology, Pasadena, California, U.S.A.* (p. 289)

J. HURYCH, *Institute of Industrial Hygiene and Occupational Diseases, Prague, Czechoslovakia* (p. 351)

M. KAKUDO, *Osaka University, Osaka, Japan* (p. 69)

M. KALYANI, *Marine Biological Station, Porto Novo, India* (p. 343)

G. KARTHA, *Roswell Park Memorial Institute, Buffalo, New York, U.S.A.* (p. 13)

E. KATCHALSKI, *The Weizmann Institute of Science, Rehovoth, Israel* (p. 205)
Y. KATSUBE, *Osaka University, Osaka, Japan* (p. 69)
V. KOBRLE, *Institute of Industrial Hygiene and Occupational Diseases, Prague, Czechoslovakia* (p. 351)
K. KÜHN, *Anorganisch-Chemisches Institut der Universität, Heidelberg, West Germany* (p. 279)
J. KURTZ, *The Weizmann Institute of Science, Rehovoth, Israel* (p. 205)
M. I. MILLIONOVA, *Institute of Biophysics, Moscow, U.S.S.R.* (p. 137)
D. N. MISRA, *Saha Institute of Nuclear Physics, Calcutta, India* (p. 301)
T. MIYAZAWA, *Osaka University, Osaka, Japan* (p. 257)
S. MOORE, *The Rockefeller Institute, New York, New York, U.S.A.* (p. 309)
H. NEURATH, *University of Washington, Seattle, Washington, U.S.A.* (p. 319)
S. OCHOA, *New York University School of Medicine, New York, New York, U.S.A.* (p. 147)
J. A. PETRUSKA, *California Institute of Technology, Pasadena, California, U.S.A.* (p. 289)
D. C. PHILLIPS, *The Royal Institution, London, England* (p. 57)
G. N. RAMACHANDRAN, *University of Madras, Madras, India* (pp. 39, 121)
C. RAMAKRISHNAN, *University of Madras, Madras, India* (p. 121)
S. SAKAKIBARA, *Osaka University, Osaka, Japan* (p. 69)
K. S. V. SAMPATH KUMAR, *University of Washington, Seattle, Washington, U.S.A.* (p. 319)
A. B. SANYAL, *Saha Institute of Nuclear Physics, Calcutta, India* (p. 301)
Y. SASADA, *Osaka University, Osaka, Japan* (p. 69)
V. SASISEKHARAN, *University of Madras, Madras, India* (p. 121)
R. V. SESHAIYA, *Marine Biological Station, Porto Novo, India* (p. 343)
U. SHMUELI, *The Weizmann Institute of Science, Rehovoth, Israel* (p. 81)
G. SWANBECK, *Karolinska Institutet, Stockholm, Sweden* (p. 93)
W. TRAUB, *The Weizmann Institute of Science, Rehovoth, Israel* (p. 81)
K. A. WALSH, *University of Washington, Seattle, U.S.A.* (p. 319)
T. WEIS-FOGH, *Zoophysiological Laboratory B, Copenhagen, Denmark* (p. 337)
M. H. F. WILKINS, *King's College, London, England* (p. 23)
RALPH W. G. WYCKOFF, *University of Arizona, Tucson, Arizona, U.S.A.* (p. 267)

Preface

An International Symposium on Protein Structure and Crystallography was organized by the Department of Physics, University of Madras, during January, 1963. This volume is a report of the proceedings of the Symposium on Protein Structure, which formed a part of this Conference. The papers dealt with various aspects of protein structure, including in particular X-ray diffraction, optical, electron microsopic and chemical studies, with two papers dealing with the genetic code between nucleic acids and proteins. There was also a discussion of the strategy of protein research at the end of the Symposium. This session was chaired by Professor J. T. Edsall, who has kindly prepared a short report of the discussion for inclusion in this volume.

Professor Lawrence Bragg had kindly agreed to preside over the Symposium, but was prevented from so doing owing to illness. His Presidential Address, which was read in his absence, is included in this volume with his kind permission.

The Symposium was made possible by grants provided by the University of Madras, the University Grants Commission and the Council of Scientific and Industrial Research, Government of India. The Organizing Committee is deeply grateful to these agencies for the generous support of the Symposium. The organizers would also like to acknowledge the continuous support and encouragement given to them by Dr. A. L. Mudaliar, Vice-Chancellor, University of Madras. The Editor wishes to thank the Academic Press for their considerable help, in various ways, in providing preprints and for speedy publication of this volume. His thanks are also due to Dr. R. Srinivasan and Mr. C. Ramakrishnan for their assistance in reading the proofs and recording the discussions.

G. N. RAMACHANDRAN
Department of Physics
University of Madras

May 1963

Contents

List of Contributors v
Preface vii
Presidential Address. X-Ray Analysis of Biological Molecules. *By* Sir Lawrence Bragg 1

SECTION I—X-Ray Diffraction Studies

The Structure of Ribonuclease II; the Positions of the Heavy Atoms in Five "Dyed" Crystals. *By* G. Kartha, J. Bello, D. Harker and F. E. Dejarnette 13
X-Ray Diffraction Studies of the Molecular Configuration of Nucleic Acids. *By* M. H. F. Wilkins 23
The Triple Helical Structure of Collagen. *By* G. N. Ramachandran 39
Protein Crystallography in Cambridge and in London. *By* D. C. Phillips 57
X-Ray Analysis of *p*-Bromocarbobenzoxy-glycyl-L-prolyl-L-leucyl-glycine and its Related Peptides. *By* M. Kakudo, Y. Sasada, Y. Katsube, S. Sakakibara and S. Akabori . 69
Structure of Poly-L-proline I. *By* W. Traub and U. Shmueli . 81
The Molecular Structure of α-Keratin. *By* G. Swanbeck . . 93
Structure and Stability of Proteins and Allied Materials. *By* P. R. Blakey and F. Happey 103
Stereochemistry of Polypeptide and Polysaccharide Chain Conformations. *By* G. N. Ramachandran, C. Ramakrishnan and V. Sasisekharan 121
Structural Investigations of Polymers Related to Collagen. *By* N. S. Andreeva, M. I. Millionova and Yu. N. Chirgadze . 137

SECTION II—Genetic Code

Synthetic Polynucleotides and the Genetic Code. *By* S. Ochoa . 147
Proteins, Coding and Medicine. *By* L. D. Hamilton . . . 161

SECTION III—Optical Studies

Ultra-violet Spectra and Structure of Proteins. *By* J. T. Edsall 179
Behaviour in Solution of Polypeptides Related to Collagen. *By* E. Katchalski, A. Berger and J. Kurtz 205
Optical Rotatory Dispersion and the Structure of Polypeptides and Proteins in Solution. *By* E. R. Blout 241
Infra-red Studies of the Conformations of Polypeptides and Proteins. *By* T. Miyazawa 257

SECTION IV—Electron Microscopic Studies

The Electron Microscope in Research on Proteins. *By* R. W. G. Wyckoff 267
Investigations of the Renaturation of Tropocollagen. *By* K. Kühn 279
Recent Studies with the Electron Microscope on Ordered Aggregates of the Tropocollagen Macromolecule. *By* A. J. Hodge and J. A. Petruska 289
Electron Microscopic and Sedimentation Studies on Human Haemoglobins A, F and E. *By* N. N. Das Gupta, D. N. Misra, P. Ganguly, A. B. Sanyal and J. B. Chatterjea . . . 301

SECTION V—Chemical Studies

A Discussion of Methods that have Proved Useful in Research on Ribonuclease. *By* S. Moore 309
Identification and Sequence of Amino Acid Residues around the Thiol of the Active Site of Carboxypeptidase A. *By* K. S. V. Sampath Kumar, K. A. Walsh, J.-P. Bargetzi and H. Neurath 319
Resilin, a Rubber-like Protein, and its Significance. *By* T. Weis-Fogh 337
Amino Acid Composition of Ichthylepidin from Fish Scales. *By* R. V. Seshaiya, P. Ambujabai and M. Kalyani . . . 343
The Study of Early Phases in Collagen Biosynthesis. *By* M. Chvapil, J. Hurych, V. Kobrle and B. Čmuchalová . . 351
Methods for Determining the Nature of Linkages of Certain Constituents of the Carbohydrate Moiety to the Protein Core in Skin Mucoid. *By* S. M. Bose 357

SECTION VI—General Discussion

Strategy of Protein Research. *Edited by* J. T. Edsall . . 371
Author Index 373

PRESIDENTIAL ADDRESS

X-Ray Analysis of Biological Molecules

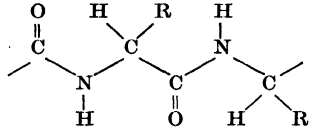
W. L. BRAGG

The Royal Institution, London, England

The X-ray analysis of biological molecules is a fascinating new development, which has taken place during the last ten years, and which holds out promise of opening up very important new scientific fields.

It has two sharply contrasted aspects. In the first line of attack, full use has been made of the known chemical constitution of the molecules as revealed by the investigations of biochemistry. Information supplied by electron microscopy has also been an invaluable aid. X-Ray results have in the first place given a hint as to the nature of the structure; possible models have then been constructed with the aid of the known chemical composition and by using the laws of stereochemistry and the detailed information about bond lengths and bond angles determined by the X-ray analysis of simpler compounds. Any plausible structure has then been tested by calculating how it would diffract X-rays and comparing these calculations with the observed X-ray diffraction effects. This is indeed the classical "trial and error" method of X-ray analysis when dealing with complicated molecules.

We may list the following as the successes of this line of attack. In the first place, there is Pauling's prediction of the nature of the polypeptide chain, the Pauling-Corey α-helix. From his fundamental studies of the nature and stability of the chemical bond, he predicted that the stable state of the polypeptide chain is helical in form. The amino acid residues which characterize the chain are linked, as had long been known, in the series



where R represents the group which specifies the amino acid. In Pauling's α-helix the CO of one turn is linked by a hydrogen bond to the NH of an adjacent turn of the spiral (Fig. 1). Pauling's helix was rapidly confirmed by studies of X-ray diffraction by natural protein chains in hair, and

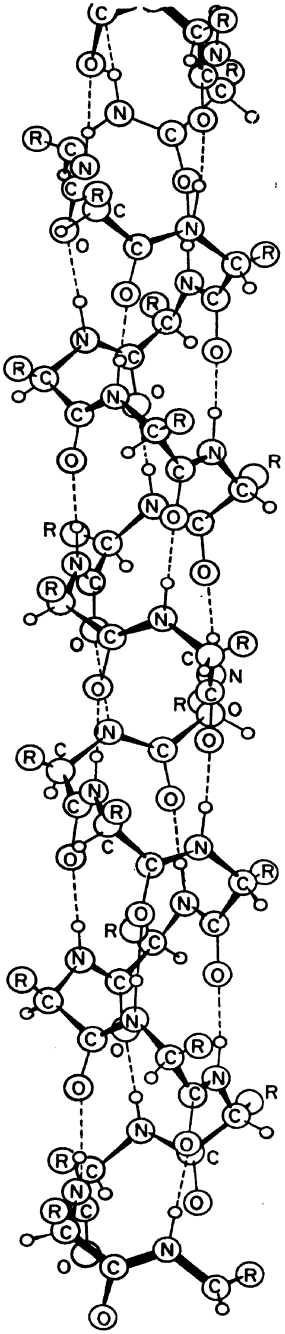


FIG. 1. The α-helix of Pauling.

synthetic protein chains. Two new aspects are noteworthy. In the first place, the helix is "irrational". There is not an integral number of amino-acid residues in each turn. In the second place, it led X-ray crystallographers to study the nature of diffraction by a helical structure. An analysis of helical diffraction by Cochran, Crick and Vand has had an immense influence on further studies of biological structures.

This analysis, for instance, played a vital role in the prediction of the structure of *nucleic acid* by Crick and Watson some ten years ago, a structure which has been fully confirmed by subsequent profound analysis. Wilkins had obtained excellent diffraction photographs with deoxyribonucleic acid (DNA), and Crick and Watson, realizing these must be ascribed to a spiral structure, proposed their famous double-helix structure for DNA (Fig. 2), which explains in such a fascinating way how hereditary characters are passed on from one generation to the next. The discovery of the DNA structure has been one of the major scientific advances of recent years. It has stimulated a vast amount of scientific work, particularly in America, and our knowledge of the hereditary principle has advanced very rapidly indeed. For instance, already the code according to which the nucleic acid determines the protein, for which it is the pattern, is becoming known.

Then again, Watson first showed that the rod-like viruses have a helical structure, and Franklin and Klug have developed this discovery. A virus has a structure of apparently identical protein molecules arranged in a geometrical way, which encloses a nucleic acid chain which determines the pattern of the virus. The nucleic acid, passing into the host body, is able to use the life processes of its victim to build this protecting envelope of protein. In the globular viruses the protein molecules are grouped in a form which reminds one of a fruit like a raspberry. The crystallographer, familiar with the symmetry forms he finds in crystals, has to readjust his conceptions because these regular forms only have a point-group symmetry. For instance, in a common form of virus there are sixty protein molecules in a structure which has two-fold, three-fold, and five-fold symmetry axes.

I need not remind you that the helical structure of collagen (Fig. 3) is another triumph of this new line of work, because the investigations of Professor Ramachandran are so famous. The structure of muscle has been attacked by a combination of X-ray and electron-microscope methods and the way in which contraction takes place by the sliding past each other of interleaved rods has been elucidated.

With the exception of virus, these bodies do not form the regular three-dimensional patterns characteristic of a crystal. The X-ray patterns are obtained from specimens which only have some feature of regularity, such as nearly parallel rods or chains, and are necessarily

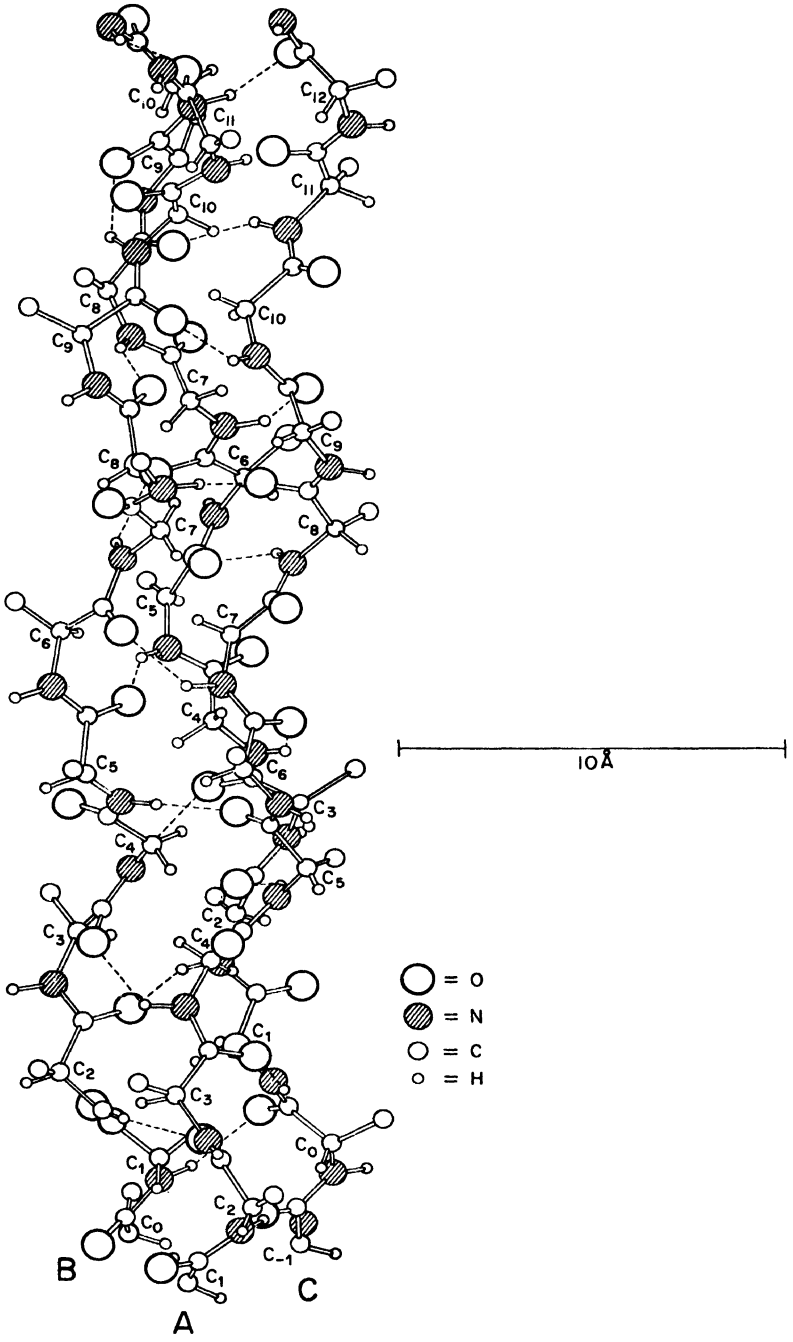


FIG. 3. The triple helical structure of collagen due to Ramachandran.

limited in the data they provide. They are used to determine the general scheme of the structures.

Passing now to the globular protein structures, the picture is very different. The globular protein molecules have molecular weights in the range of 12,000 to 1,000,000. Although it contains thousands of atoms, each molecule has a definite structure like the simpler molecules of organic chemistry. They form perfect crystals. Diffraction photographs show spots which indicate a regularity of structure out to a resolution of 1·5 Å or less. They provide therefore ideal material for X-ray analysis.

The story of their successful X-ray analysis is a romantic one. For a long time crystallographers viewed their elegant diffraction patterns with much the same feelings that an archaeologist must have in looking at the literary records of some old civilization without a clue as to how to interpret them and read their story. May I remind you of the nature of the difficulty. The classical approach of X-ray analysis is one of "trial and error" as has already been mentioned. Each X-ray diffraction is a measure of a periodic element in the regular crystalline pattern of electron density. If we know both the amplitude of each of these elements, and its phase referred to a point of the crystal lattice which is chosen as origin, the structure is solved. The periodic elements can be added together in a three-dimensional Fourier series, and the result is a map of the density everywhere in the crystal. The atoms appear as condensations of density which represent the cluster of electrons in them. The primary difficulty of X-ray analysis is that, whereas the amplitude of the periodic element can be measured by the intensity of the corresponding X-ray diffraction, there is no direct way of measuring the phase. The only criterion we can apply is that, if the phases have been attributed correctly, the result will represent the atoms we know are there; if the phases are wrong there will be a meaningless jumble of density distribution. The classical X-ray approach has therefore been to guess a probable structure, calculate how it would diffract X-rays, and compare the calculations with the observed strength of the X-ray diffractions. If an encouraging degree of correspondence is achieved, the structure is put through a process called "refinement". The phases of the supposed structure are calculated, a Fourier series is summed, and with good fortune it indicates adjustments to the supposed position of the atoms which improve the accuracy of the structure. The cycle is gone through several times, till finally the crystallographer's checks show that his structure must be close to the truth.

It will readily be understood that the complexity and difficulty of this process increases very rapidly indeed with the number of atoms in the molecule. The researcher is of course greatly helped in making his guesses by his knowledge of simpler molecules, and by the chemists'

views on the stereochemistry of the molecule he is investigating. Nevertheless the highest point reached by such methods has been the solution of molecules containing one or two hundred atoms. The solution of the structure of vitamin B_{12} by Mrs. Hodgkin and her colleagues in Oxford, for instance, is a landmark in X-ray crystallography.

In the present state of knowledge, such a procedure would be quite hopeless in the case of a protein molecule with its thousands of atoms. It may be that, when a number of these molecules have been analysed, we may learn so much about the principles which govern their structures that we can make an intelligent guess as to the probable structure of a new form. At the start, however, such knowledge is not available, and to make a series of guesses as to how the thousands of atoms are placed would be unthinkably difficult.

When I came to the Cavendish Laboratory in 1938, I found there M. F. Perutz who had obtained very fine diffraction pictures with the protein haemoglobin, in which I was greatly interested. I asked the Medical Research Council, then under the direction of Sir Edward Mellanby, to finance a small research team to investigate proteins by X-ray analysis. I was frank about the outlook. It was like multiplying a zero probability that success would be achieved by an infinity of importance if the structure came out; the result of this mathematical operation was anyone's guess. Fortunately he enthusiastically supported the venture. This small beginning with two or three workers in one room has now grown under the direction of Perutz and Kendrew into the Medical Research Council's Laboratory for Molecular Biology in Cambridge, a premier institution of its kind in the world. Its researches initiated the work on nucleic acid, virus and muscle, and their efforts have now been crowned in the last few years by the successful complete solution of a protein structure, after an attack which has lasted for twenty-five years.

The difficulties in solving so complex a structure seemed insuperable, but fortunately Nature has given us an unexpected bonus which removes the phase difficulty. The molecules are so large (30 Å–100 Å across) that, as Perutz discovered, one can attach heavy atoms or heavy-atom complexes to definite points of the molecule without disturbing the crystalline arrangement. Protein crystals are fragile associations of molecules, with often about half the space in the crystal occupied by mother-liquor. They have to be kept in this liquor or else they collapse. Two conditions are necessary. One must find a heavy atom which can be attached to a definite chemical feature on the outside of the molecule such as a sulphur atom, and also the bulge which it causes must be in a place where there is room for it in the gaps between the molecules; it must not be at a point where the molecules are in contact as it would then cause an alteration of the crystal lattice. The process of analysis consists in comparing

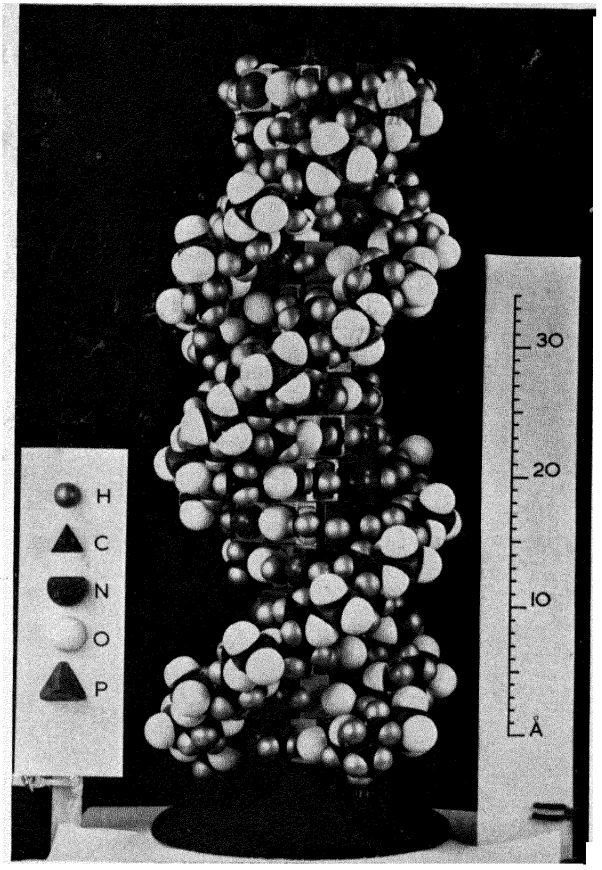


FIG. 2. The Watson-Crick double helical structure of DNA (*Courtesy of Dr. Wilkins*).

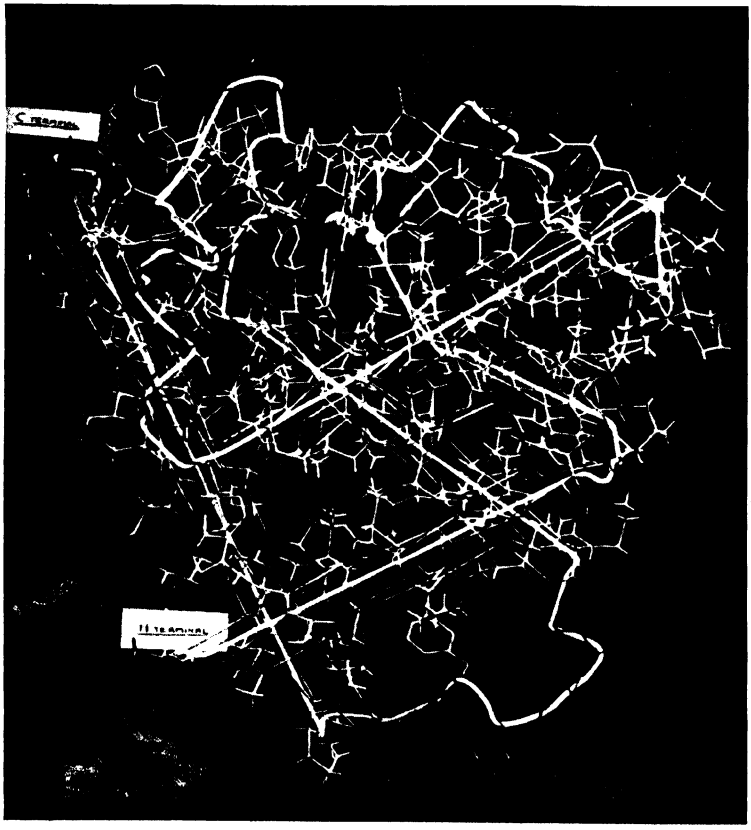
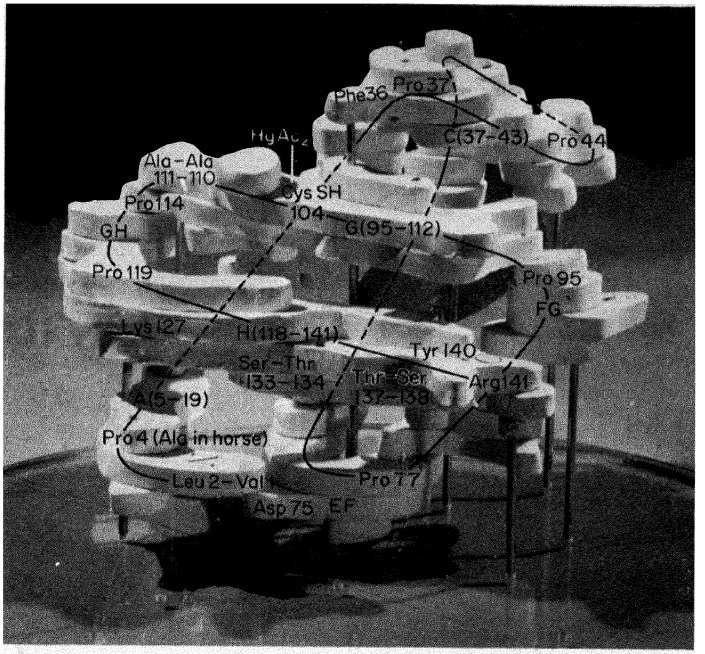
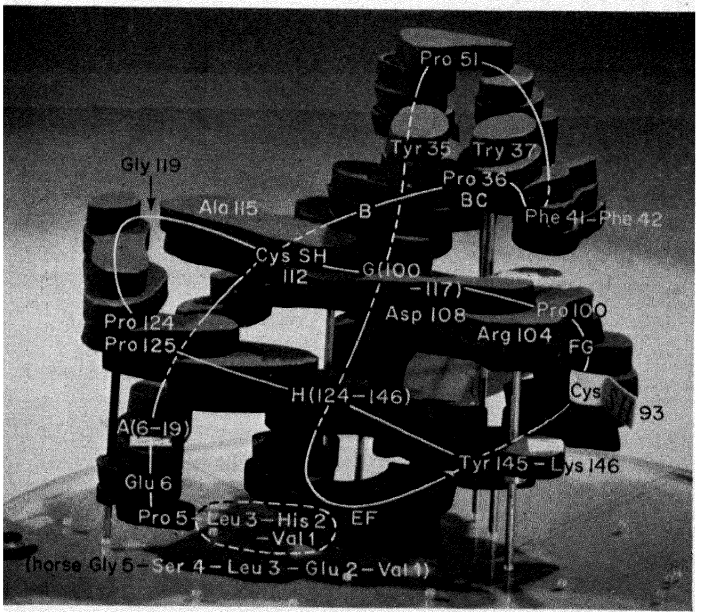


FIG. 4. The chain configuration in the structure of myoglobin, solved by Kendrew.



(a)



(b)

Fig. 5. The chain configuration (at low resolution) in haemoglobin, studied by Perutz. (a) The α-chain; (b) the β-chain.

quantitatively the diffraction by the native protein and that by the protein with a heavy-atom attachment. At first sight it seems strange that one heavy atom can modify the diffraction due to thousands of light atoms such as carbon, nitrogen and oxygen. It does so for the following reason. The resultant diffracted amplitude is due to contributions, which are called the f factors, from all the atoms in the unit structure. These amplitudes, with phases depending on the atomic positions, are added together in a vector diagram in the way familiar in the treatment of optical diffraction. Since the phases range over all values, the net result of the thousands of light atoms is proportional to the square-root of their number, as in the famous "drunkard's walk" problem. On the other hand, the heavy atom is at a definite place, and its vector is simply its f factor. So a single heavy atom like mercury, with an f factor of about 80, produces a contribution comparable to that of, for instance, 2500 light atoms with an f factor of 6, 7 or 8 [$\sqrt{(2500)} \times 7 = 350$]. Hence, measurable alterations of diffraction are produced.

Though the structure of the protein is initially unknown, it is always possible to find the positions of the heavy atoms in the unit cell. A statistical survey of the alterations it makes in the diffraction yields the necessary information, by methods familiar to X-ray crystallographers.

The position is therefore as follows. We know the amplitude and phase F_H of the vector representing the contribution of the heavy atom, the phase being measured with reference to some origin we have chosen in the crystal lattice. We know the amplitude, but not the phase, due to the native protein F_P and that of the protein with heavy atom, F_{P+H}. Since F_{P+H} is the vector resultant of F_P and F_H, then their vectors must form a closed triangle in our diagram, and this requirement tells us their phases. In practice the results with a single heavy atom are ambiguous. Two atoms clear up many of the ambiguities, but at least three are desirable to clear up most of them and provide cross-checks. An investigator therefore tries to find three types of heavy-atom attachment, at definite and different points, which satisfy the condition of not altering the crystal dimensions. If he is successful in doing this he can proceed directly to the solution of the structure, without any element of guesswork or trial and error.

If the way to the solution is direct, however, it is at the same time extremely complex. The first protein to be solved was myoglobin, by Kendrew (Fig. 4). This molecule has a molecular weight of about 17,000 and contains 2500 atoms. Its function is to store oxygen in muscle. The native proteins and four derivatives with heavy atoms or combinations of heavy atoms were measured. In the first place the resolution was taken to 2 Å, which meant measuring 10,000 diffractions for each type of crystal (in a later extension to 1½ Å, 20,000 have been measured). The

measurements must be corrected for absorption and geometric factors, and the results scaled to each other. The solution of 10,000 vector diagrams then gives the phases. These must be expressed as a Fourier series with 10,000 terms, and the series summed at about $100 \times 100 \times 50$ points inside the unit cell. The process would of course be impossibly lengthy without the aid of the electronic computer. It takes a small team some months to plot a density map with the figures turned out by the computer, and this plot has then to be interpreted. In the first interpretation, Kendrew built a large-scale model on the floor of the laboratory with vertical rods bearing coloured tags to represent density—several miles of rods were required. It was then possible to identify such features as α-helices and the haem group in the model and construct a new version on a smaller scale.

Almost the complete structure of myoglobin has now been determined. Haemoglobin, studied by Perutz (Fig. 5), is not yet determined to so high a resolution, but it is clear that it is composed of four units each of which is very closely related to, but not identical with, the myoglobin molecule. Other proteins are in various stages of analysis. Lysozyme is being studied in the Davy Farady Laboratory, by Corey in Pasadena, and by Dickerson in Illinois, and a successful start on lactoglobulin has been made in the Davy Faraday Laboratory. Chymotrypsinogen is being studied by Kraut in Seattle, and chymotrypsin by Blow in Cambridge.

I confess to a feeling of awe when I look at the model of the myoglobin structure and consider that its atomic architecture has been determined by X-ray analysis. Simultaneous studies of the amino acid sequence by Edmundson have been necessary to identify many of the residues; the X-ray results at this resolution cannot distinguish for instance between O, NH or CH_2. The structure has several runs of α-helix, and the Pauling-Corey model lends precision to the atomic positions in the helix. It remains true, however, that the structure has been determined directly without any preconceived ideas of its nature; indeed, none was available. Crystallographers assess the complexity of a structure by the number of parameters the values of which determine the atomic positions. At one stroke X-ray analysis has passed from structures with two or three hundred parameters to structures with many thousands. As an exercise, I plotted recently the logarithms of the number of parameters of determined structures against the years. In 1913 crystals with one parameter were hailed as striking examples of the success of X-ray methods. The curve rises almost linearly to the 1950's when the number is measured in hundreds. Then there is a steep rise to thousands, representing the success of direct methods. If we extrapolate we ought to be measuring structures with a million parameters in 1965. This is not so wild a prophecy as it might seem, for it is conceivable that similar detailed knowledge about

a virus structure might be forthcoming in the not too distant future.

I have dwelt on the technical side of the application of X-ray analysis. It is hardly necessary to stress the biological significance of this new knowledge. The way the protein molecules function, the way they are formed by nucleic acid acting as a pattern, their reaction to antibodies, to viruses, to hormones and vitamins, are all subjects which, we must anticipate, will now be studied with far greater effect because the molecular architecture of these bodies is known. A new field of science has been opened up.

SECTION I
X-Ray Diffraction Studies

The Structure of Ribonuclease II: The Positions of the Heavy Atoms in Five "Dyed" Crystals

G. KARTHA, J. BELLO, D. HARKER AND F. E. DEJARNETTE

Roswell Park Memorial Institute, Buffalo, New York, U.S.A.

ABSTRACT

Crystals of ribonuclease II have been successfully "dyed" with heavy atom compounds to produce five crystals isomorphous with the undyed crystal. X-Ray diffraction data obtained from these crystals have been used to locate the heavy atoms in each isomorph. A preliminary three-dimensional electron density function at 4 Å resolution has been computed for the undyed crystal, and is being interpreted. Further work at finer resolution is under way, and will be reported later.

The group of enzymes having ribonuclease activity is of great biochemical interest. Most of the research on structure and function in this group has been carried out on bovine pancreatic ribonuclease. In our laboratory we are investigating the structure of crystalline form II (King *et al.*, 1956) of bovine pancreatic ribonuclease. This enzyme contains a single polypeptide chain of 124 residues internally cross-linked by four disulphide bridges. The primary structure of ribonuclease has been determined in the laboratories of Stein and Moore at the Rockefeller Institute, and of Anfinsen at the National Institutes of Health (several papers in *J. biol. Chem.* 1956–62).

Crystals of ribonuclease II were used in the "soaked" and "dyed" conditions as sources of X-ray diffraction data. Table I lists these crystals. The intensities of the diffracted beams were measured on the Eulerian Cradle (Furnas and Harker, 1955), using stationary crystal-stationary counter techniques and were converted to relative $|F(hkl)|^2$ values in the usual way. The radiation was obtained from an X-ray tube with a copper target operated at 20 mA and 40 kVp. The "monochromatization" of the radiation was accomplished by a nickel–cobalt balanced filter pair. Data were collected to a resolution of 2 Å for the centrosymmetric $h0l$ zone and to 4 Å for the complete three-dimensional set, corresponding to over 600 and 1000 reflections, respectively. This was done for the free

protein, as well as the five derivatives. In the case of the three-dimensional data, both hkl and $\bar{h}\bar{k}\bar{l}$ reflection intensities were measured.

The data from the various derivatives were scaled to that of the free protein so that similar plots of $\langle|F|^2\rangle$ against scattering angle 2θ, as well

TABLE I. Ribonuclease II, space group $P2_1$, $Z = 2$. Soaked in 2-methyl-2,4-pentanediol

Crystal	Dye	a	b	c	β
Free protein STD II	0	30·19 Å	38·24 Å	53·06 Å	105·79°
Ems II 13	1 *cis*-Diglycine Pt 8:1†	29·71 Å	38·14 Å	53·05 Å	105·98°
Ems-1-33	2 *cis*-Diglycine Pt 4:1	29·73 Å	38·14 Å	52·98 Å	105·93°
SS-1-28C	3 $Pt(en)_3Cl_4$	30·15 Å	38·20 Å	53·10 Å	105·80°
TF-1-15	4 $Pt(NH_3)_2(NO_3)_2$	30·24 Å	38·14 Å	53·14 Å	106·35°
SS II $26B_1$	5 (UO S Sal + Pt)‡	30·22 Å	38·53 Å	52·81 Å	105·50°

† *cis*-Bis-(glycine)-platinum (II) or *cis*-$Pt(glycine)_2$.
‡ Potassium salt of uranyl-sulphosalicylic acid complex plus $Pt(en)_3Cl_4$.

as the crystal orientation angles ϕ and χ, were obtained. Scaled structure amplitudes were thus obtained for the free protein and the five derivatives which we shall denote as $|F_P|$, $|F_{P+H_1}|$, $F_{P+H_2}|, \ldots |F_{P+H_5}|$. We shall further denote by $|\bar{F}|$ the reflection which is inverse to the reflection $|F|$. Obviously, these two amplitudes should be the same in the absence of any effects due to anomalous scattering or experimental errors.

DETERMINATION OF HEAVY-ATOM POSITIONS

Fourier maps using various combinations of the structure amplitudes F_P and F_{P+H_i}. were made to reveal the positions of the heavy atoms in each of the dyed crystals. Three main methods were used to obtain the initial heavy-atom positions.

(1) Using the $h0l$ data, maps were made with $|\Delta_{is}|^2 = ||F_{P+H_i}| - |F_P||^2$ as coefficients. For a centrosymmetric set of reflections, it is easily seen that this will give the HH vectors between the heavy atoms in the derivative in question. This map gives the x and z co-ordinates of the heavy atom in those cases where the $|\Delta_{is}|^2$ Patterson can be easily and unambiguously interpreted as due to a few heavy atoms. This was indeed possible in derivatives 1, 2 and 3.

(2) Three-dimensional maps with coefficients $||F_{P+H_i}| - |F_P||^2 = \Delta_{is}^2$ were made with all data to 4 Å, where $P + H_i$ indicates the ith kind of

dyed crystal and P the undyed soaked crystal. These maps give (Blow, 1958) main positive peaks in regions corresponding to the ends of vectors between heavy atoms in the ith derivative, i.e. HH vectors in the ith derivative. In particular, the Harker section at $y = \frac{1}{2}$ should give the vectors $2x_i$, $2z_i$ for the heavy atoms. In cases where there is accidental correspondence of y parameters of the heavy atoms, this leads to non-Harker peaks of higher weight at $y = \frac{1}{2}$, as well as a similar peak at $y = 0$. Thus, comparison of the $|\Delta_{is}|^2$ Patterson sections at $y = 0$ and $y = \frac{1}{2}$ leads to identification of the Harker and non-Harker peaks and estimation of the x and z parameters of the heavy atoms. Maps of these sections are shown in Fig. 1. These parameters were compared with those obtained from the $h0l$ projections with centrosymmetric data. Further, inspection of the rest of the three-dimensional map for prominent peaks suggested the relative y parameters of the heavy atoms in any given derivative.

(3) Three-dimensional maps were computed for the various derivatives using the square of the anomalous scattering differences (Rossmann, 1961) between the direct and inverse reflections as coefficients, i.e. $|\Delta_{an}|^2 = ||\overline{F}_{P+Hi}| - |F_{P+Hi}||^2$. It can be shown that the anomalous difference Δ_{an} is proportional to $F_{H_i} \sin\alpha$ where F_{H_i} is the modulus of the structure amplitude of the heavy-atom group H_i and α an angle related to the free protein phase and, thus, varying fairly randomly from reflection to reflection. Hence, the anomalous Patterson has coefficients $F_{H_i}^2 \sin^2\alpha$ and the map will be the convolutions of the Fourier transforms of $F_{H_i}^2$ and $\sin^2\alpha$. Obviously, the main feature of the transform of $\sin^2\alpha$, which is always positive and varies randomly, will be a huge sharp positive peak at the origin with no other prominent peaks. Hence, the convolution of the $F_{H_i}^2$ and $\sin^2\alpha$ transforms will have a close resemblance to the $F_{H_i}^2$ transform, except that the peaks are more diffuse. This means that the main positive peaks in the anomalous difference Patterson will closely correspond to the self-Patterson of the anomalously scattering atoms in the derivative, which are usually also the heavy atoms we are trying to locate. This map has the great advantage that we use only data from the heavy-atom derivative, without any comparison with the undyed parent protein and thus avoids errors due to incorrect scaling and lack of isomorphism. However, the anomalous scattering effect being rather small and of the same order of magnitude as the uncertainties due to absorption and the experimental errors in intensity measurements, the $|\Delta_{an}|^2$ Patterson functions were used only as a check in confirming the heavy-atom positions derived from the earlier two methods.

All three methods gave consistent heavy atom positions in derivatives 1, 2 and 3. Two main sites were located in 1 and 2 and one site in derivative 3. Both in 1 and 2 the heavy atoms occupied the same two sites—as is to be expected—although with slightly different occupancies. Unfortunately,

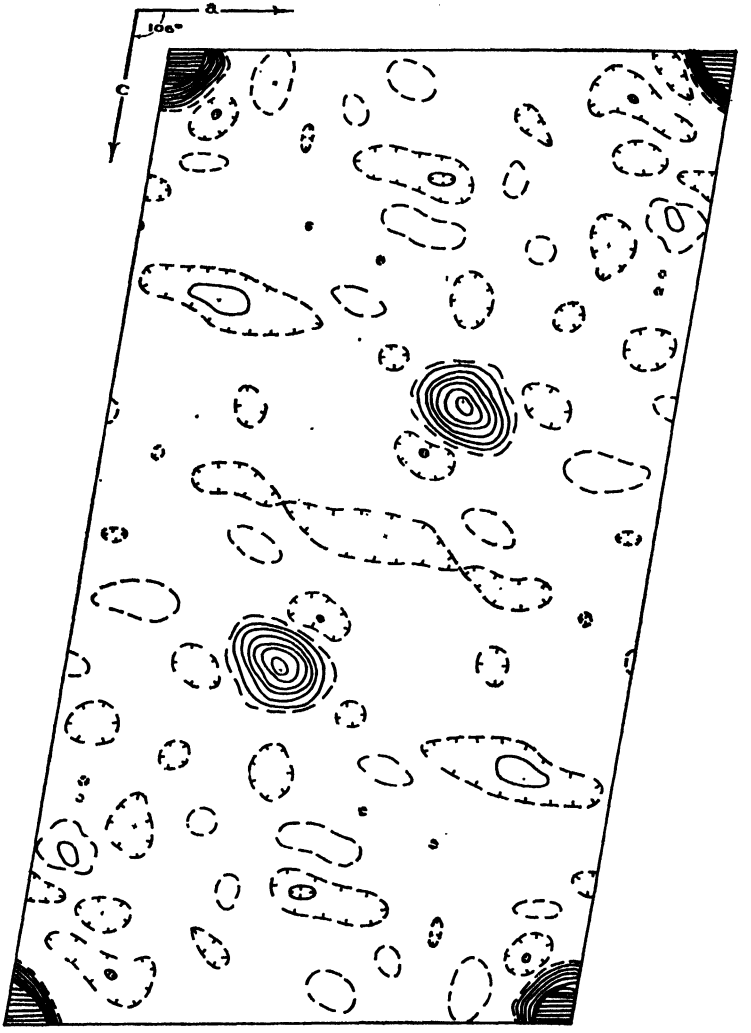


FIG. 1 (a). Sections through the "$|\Delta_{is}|^2$ Patterson" between crystals 1 and 0 at $y = 0$.

the y parameters of the two heavy-atom sites accidentally coincided, thus giving double weight peaks in the $y = 0$ and $y = \frac{1}{2}$ sections of both the $|\Delta_{is}|^2$ and $|\Delta_{an}|^2$ three-dimensional Pattersons.

REFINEMENT OF HEAVY-ATOM POSITIONS

From the x and z parameters of the heavy atoms, sets of signs for the centric reflections $F_P(h0l)$ were derived from the computed values and

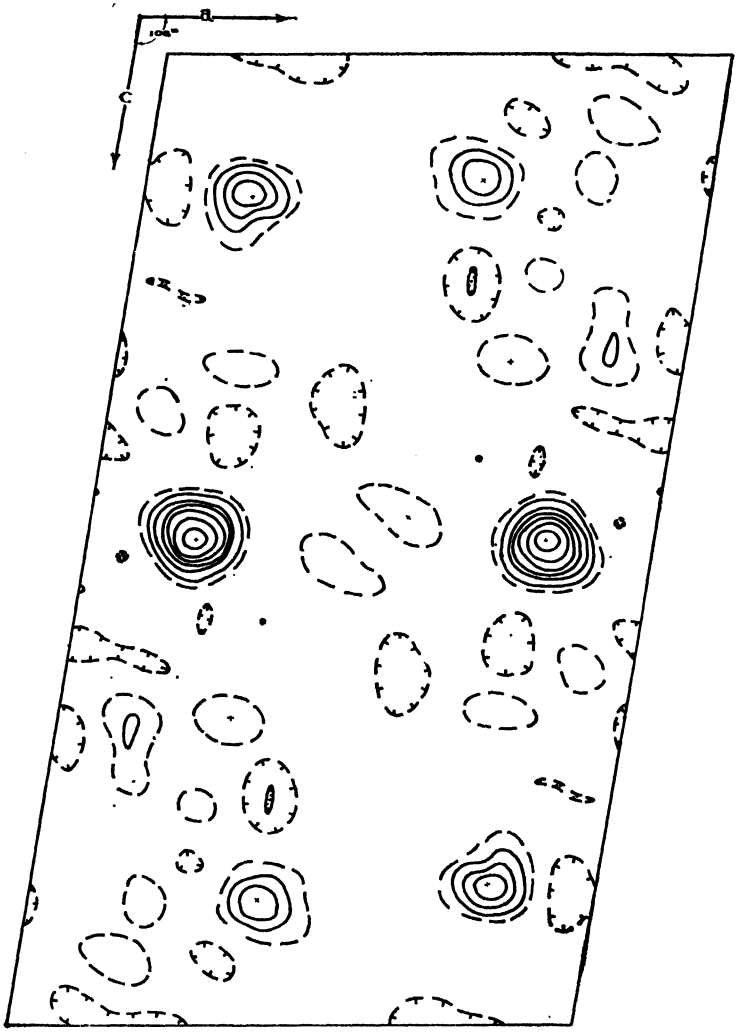


Fig. 1 (b). Sections through the "$|\Delta_{1s}|^2$ Patterson" between crystals 1 and 0 at $y = 15/30$.

the signs of $F_{H_i}(h0l)$. With these signs it was possible to compute projections on (010) of the electron density difference functions $\rho_{P+H_i} - \rho_P$. These should be equal to ρ_{H_i}, the projection of the heavy-atom densities on the (010) plane, if no incorrect F's or signs are involved. From these maps refined x_i and z_i parameters and occupancies, as well as minor additional sites, could be deduced and the new values used for the next cycle of $F_P(h0l)$ sign determination, and so on. Starting with the heavy-atom positions in the first three derivatives, obtained by the methods

described in earlier sections, this was done in a series of cycles of heavy-atom position refinement using data from all the five derivatives. Table II shows how well the $F_P(h0l)$ signs determined from the different heavy-atom derivatives agree.

TABLE II. Agreement between the signs determined from the heavy-atom derivatives

Type of agreement		Number of cases
All 5 signs agree		70
4 signs agree	other uncertain	24
4 signs agree	other disagrees	9
3 signs agree	2 others uncertain	17
3 signs agree	1 uncertain, 1 disagrees	2
2 signs agree	3 uncertain	9
	Total	131

Total number of $h0l$ reflections within 4 Å 160

Figure 2 is the projection on (010) of the electron density difference between crystals 1 and 0 (see Table I) obtained using the signs of $F(h0l)$ finally selected.

Having thus obtained the x and z parameters, the relative y parameters in any given derivative were determined from the $|\Delta_{is}|^2$ and $\Delta_{an}|^2$ three-dimensional maps and refined by a least squares procedure which minimized $\{w|\Delta_{is}| - |F_{H_i}|\}^2$. Here F_{H_i} is the structure amplitude of the heavy-atom configuration in the ith derivative and w is a weighting factor to take account of the fact that $|\Delta_{is}| \leqslant |F_H|$ depending on the phase angles of F_P and F_H. Three-dimensional data to 4 Å were used in these refinements. The occupancies were also refined during this stage.

ABSOLUTE VALUE OF y PARAMETERS

In space group $P2_1$, the choice of the origin of the y co-ordinates locating the heavy-atom positions H_i is arbitrary, and hence it is necessary to relate the y values of the heavy atoms in the different derivatives with respect to a common origin to enable the protein phase angles to be evaluated. This was done using two types of three-dimensional correlation map. In the first type $||F_{P+H_i}| - |F_{P+H_j}||^2 = |\Delta_i - \Delta_j|^2$ (Rossman, 1960) values were used as coefficients in obtaining the map. These maps have main positive peaks at the positions corresponding to the self-Pattersons of the heavy-atom groups H_iH_i and H_jH_j in the first and

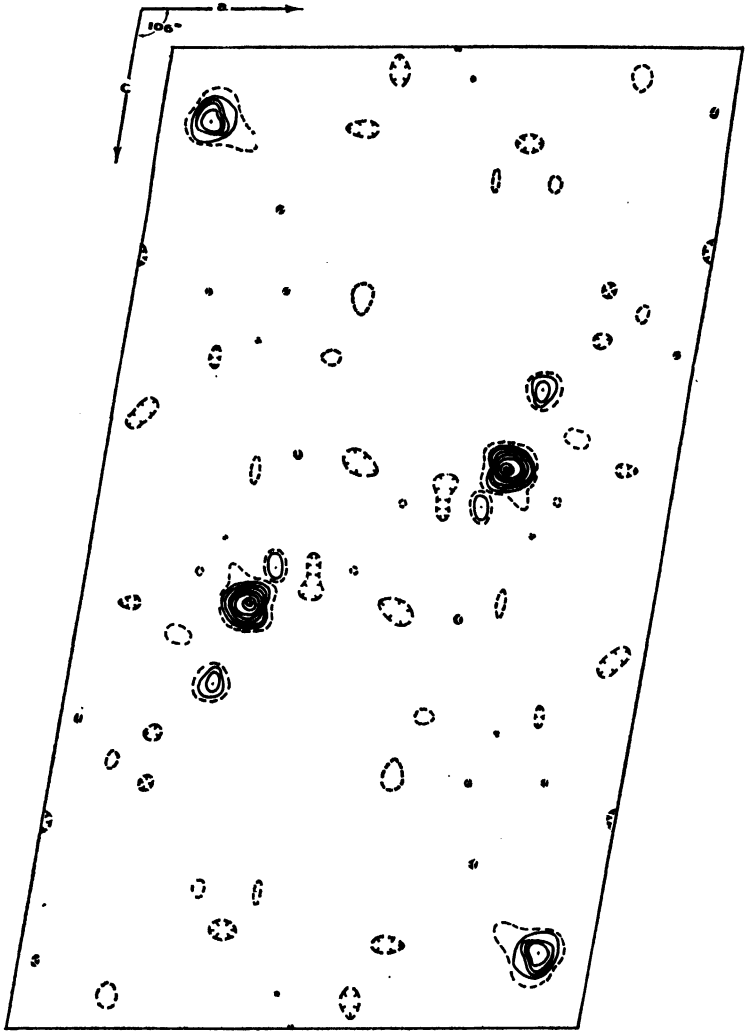


FIG. 2. Projection on (010) of the electron density difference between crystals 1 and 0.

second derivatives, as well as negative peaks H_iH_j corresponding to the image of the H_j group as seen from the H_i. These negative peak positions give the y parameters of the atoms in the second derivative with respect to the first. Evaluating these maps, varying the derivative H_j and keeping H_i throughout equal to H_1, we obtained the y parameters of the heavy atoms in the various derivatives with respect to the same origin as in derivative $H_i = H_1$.

Another correlation function (Kartha, unpublished), which was found to have much less background, was devised. It can be shown that a map with

$$\{|F_{P+H_1}| - |F_P|\}\{|F_{P+H_j}| - |F_P|\} = \Delta_i \Delta_j$$

as coefficients will have, as main peaks, only positive peaks corresponding to vectors H_iH_j; this correlates the y parameters of atoms in one derivative with those in the others with respect to the same origin.

Both types of function were used in placing the heavy-atom positions in all derivatives with respect to the same origin of co-ordinates as the protein. Their co-ordinates and occupancies are set forth in Table III.

TABLE III. Main heavy-atom sites and occupancies used in the first phase angle evaluations

	Site	Weight	x	y	z
Deriv. 1					
cis-Diglycine	A	130	0·096	0·030	0·073
Pt 8:1	B	170	0·709	0·000	0·434
	C	65	0·788	0·130	0·374
Deriv. 2					
cis-Diglycine	A	130	0·092	0·030	0·071
Pt 4:1	B	170	0·706	0·000	0·433
	C	45	0·775	0·130	0·347
Deriv. 3					
$Pt(en)_3Cl_4$	D	200	0·656	0·367	0·471
	E	50	0·185	0·180	0·062
Deriv. 4					
$Pt(NH_3)_2(NO_3)_2$	D	150	0·679	0·367	0·456
	B	100	0·693	0·000	0·427
	C	100	0·773	0·130	0·357
Deriv. 5					
$UO_2SSal.+Pt$	F	60	0·008	0·250	0·016
	E	50	0·173	0·180	0·054
	D	160	0·661	0·367	0·472
	C	50	0·767	0·129	0·357
	B	60	0·700	0·000	0·427

ELECTRON DENSITY MAPS

Using the best sets of signs for the $F_P(h0l)$, electron density projections on (010) were made of the free protein using data to various resolutions ranging from 4 Å to 2Å. However, the heavy overlap in projection precluded the possibility of detecting any recognizable feature of the molecule in the projection.

The heavy-atom positions and occupancies were used in calculating

the heavy-atom amplitudes F_{H_i} in the five derivatives. These, together with $|F_P|$ and $|F_{P+H_i}|$ were used in solving the phase circle diagrams (Harker, 1956) for the five derivatives, each derivative yielding two possible solutions for the free protein phase angle. From the resulting ten phase angles, five angles, one from each derivative, which had the minimum scatter around the average of the five, were picked out. The process was repeated, using only data from the first three derivatives. The average angle was accepted as the protein phase angle. In a few cases where four angles agreed very closely, but the fifth was very far out, this last one was left out in obtaining the angle to be used as the protein phase. These protein phase angles were used in evaluating a three-dimensional electron density function at 4 Å resolution, using about 1050 reflections within the corresponding limiting sphere. The map was evaluated at intervals of $a/30$, $b/30$, $c/60$. Electron density contours were drawn at arbitrary intervals of 0, 200 and 400 units. The map has been plotted on a stack of transparent sheets and is in the process of being studied to see whether it is possible to trace the course of the polypeptide chains and positions of sulphur bridges, and to see whether these results can be correlated with what is known of the amino acid sequence of the protein, as determined by chemical methods—and also with the 2 Å map obtained by Avey *et al.* (1962). We intend to recalculate the electron density function using more sophisticated methods of phase angle evaluation and weighting of Fourier terms to reduce false details in the maps that might have been introduced by various errors in the experimental and computational procedures. Results of these studies will be published elsewhere.

Acknowledgements

Thanks are due to Dr. David Harris for help in some of the data processing involving the IBM 1620, in particular, for the program for evaluating the phase angles of the three-dimensional data to 4 Å. Most of the remaining computations were done on the IBM 7090 at Bell Aero-systems at Niagara Falls.

We are also grateful to Mrs. Theresa Falzone, Miss Elsa Swyers, Miss Frances Holland, Miss Sylvia Scapa and Mr. Stefan Kon for valuable assistance in the preparation of heavy-atom dyes and ribonuclease crystals.

We are happy to acknowledge our debt to The Dean Langmuir Foundation, The National Science Foundation and the National Institutes of Health, without whose continued financial support this series of researches could not have been carried out.

References

Avey, H. P., Carlisle, H. C. and Shukla, P. D. (1962). *Brookhaven Symp. Biol.* **15**, 199.

Blow, D. M. (1958). *Proc. roy. Soc.* **A247**, 302.

Furnas, T. C. and Harker, D. (1955). *Rev. sci. Instrum.* **26**, 449.

Harker, D. (1956). *Acta cryst.* **9**, 1.

King, M. V., Magdoff, B. S., Adelman, M. B. and Harker, D. (1956). *Acta cryst.* **9**, 460.

Rossmann, M. G. (1960). *Acta cryst.* **13**, 221.

Rossmann, M. G. (1961). *Computing Methods and Phase Problem*, p. 264. Pergamon Press, Oxford.

Discussion

G. N. RAMACHANDRAN: The method of determining the heavy-atom position mentioned by you seems to work satisfactorily, as we all know, both in centrosymmetric as well as in non-centrosymmetric cases. While it is easy to understand the reason for this in the centrosymmetric case, it is not so apparent for the other case. I would like to mention that we have been able to show, by an analysis of the problem from a Fourier synthesis approach, that the method should theoretically work even in the non-centrosymmetric case.

D. HARKER: I suppose this kind of work has also been done in connection with myoglobin.

D. C. PHILLIPS: The method was used by D. M. Blow and developed by Rossmann; it was used extensively by Perutz and his co-workers in the work on haemoglobin, particularly in finding the relative y co-ordinates of the heavy atoms. I would like to know if a similar procedure was adopted in the work on ribonuclease.

G. KARTHA: The x and z parameters for the heavy atoms in the five different derivatives were refined using the centrosymmetric $h0l$ data, to a resolution of 2 Å. The relative y parameters in any single derivative were obtained from Pattersons and refined by Fourier and least squares. The relative y parameters in the different derivatives were obtained by the use of correlation Patterson functions using data from the derivative that proved to be the best one and the other four derivatives and the free protein. Both functions of type $|\Delta_1 - \Delta_2|^2$ as well as $\Delta_1 \Delta_2$ were used, where $\Delta_1 = |F_{P+H_1}| - |F_P|$, etc.

S. MOORE: What is your current approximation of the maximum amount of α-helix in the ribonuclease molecule?

D. HARKER: We do not need to make any approximation as we shall get it by direct structure determination. But I suppose there could be as much as 20%, which would be in agreement with the physicochemical work.

W. TRAUB: I have the impression from the electron density map that the various heavy atom sites have different occupancies. Could you please tell me if this is the case and if so how you estimated the occupancies at the various sites?

D. HARKER: It is true. In the final analysis it was estimated by least squares methods by adjusting the occupancy as one of the parameters. For the compound *cis*-diglycine Pt 4:1, we got an R value of 35% for the centrosymmetric data. I might say, in passing, that a similar computation made with Carlisle's data gave a value of 85% and I am pretty sure that Carlisle's structure needs improvement.

D. C. PHILLIPS: There is a good deal of evidence, for example from the work on lysozyme by Corey and others at the California Institute of Technology, that heavy-atom crystals may not be "isomorphous" with the native crystals, even when the cell dimensions are practically unchanged.

X-Ray Diffraction Studies of the Molecular Configuration of Nucleic Acids

M. H. F. Wilkins

Medical Research Council Biophysics Research Unit, Biophysics Department, King's College, University of London, England

ABSTRACT

The *A*, *B* and *C* configurations of DNA are briefly described, together with the main features of the corresponding X-ray diffraction patterns. Recent data concerning the *A* configuration are given. Three-dimensional Fourier synthesis techniques have been used with the data from the crystalline *B* form, and base-pairing schemes alternative to the Watson-Crick scheme have to a large extent been eliminated. Recent studies on micro-crystalline fibres of amino acid transfer RNA are described. The diffraction patterns of this type of RNA are well defined and bear a close resemblance to the *A*-type pattern of DNA. The molecular structure of the RNA and its relation to the *A* configuration of DNA is described. The diffuse patterns obtained from amorphous virus and ribosome RNA are shown to arise from the same type of helical structure as is found in transfer RNA. An account is also given of polarization microscope observations on liquid crystal forms of transfer RNA.

1. Introduction

That nucleic acids might have a very regular molecular structure was first indicated, in 1938, by X-ray diffraction patterns of DNA (deoxyribonucleic acid), which were obtained by Astbury, pioneer in so many fields of biomolecular structure. Since then, progress in nucleic acid X-ray structure analysis has consisted largely of slow but continual improvements of the diffraction patterns and of increasingly effective exploitation of the diffraction data. By combining interpretation of the diffraction patterns with molecular model building (in particular that of Watson and Crick), DNA molecular structure, with its revolutionary implications, has been discovered. In a similar way, the helical structures of RNA (ribonucleic acid) and of many synthetic polynucleotides have been elucidated.

Because nucleic acids are of great importance, it is very desirable to place the determinations of their structures on as firm a base as possible. It is not easy to judge the reliability of the structure determinations by

the usual criteria of X-ray crystallography. The nucleic acid diffraction data are somewhat limited and need to be combined with much stereochemical study; therefore there has been a gap between X-ray structure analysis of nucleic acids and conventional studies of crystalline substances. This gap is, however, narrowing—DNA is now giving reflections from spacings as small as 1·1 Å, and RNA crystals many microns across have been obtained. The possibility of obtaining single crystals is now in view. Moreover, the perfection of microcrystals of DNA is comparable with that of crystals of substances of low molecular weight: this is in spite of DNA molecules being of enormous size, some approaching a millimetre in length.

2. Nucleic Acids and their Biological Functions

There is considerable interest in nucleic acids, both because their functions are very important and because of the remarkable relation discovered for DNA between molecular structure and biological function. DNA contains, encoded in its molecular structure, the genetic information that is passed from one generation to the next. The molecules are threadlike and consist of two polynucleotide chains twisted round each other and joined together by hydrogen bonds. Each nucleotide consists of a purine or pyrimidine base, a deoxyribose ring and a phosphate group. There are four different bases, and the complicated sequence of these along the chains forms the genetic message. A triplet of bases corresponds to an amino acid, the sequence of triplets determining the sequence of amino acids in the polypeptide chain of a protein. The determining of protein structure by DNA base-sequence is the most fundamental step by means of which genes determine the inheritable constitution of living things.

Growth and reproduction require replication of the genetic material. DNA molecular structure is specially arranged to permit such replication to take place. Each base in DNA is hydrogen-bonded to a base in the opposite chain (Fig. 1). The base-pairs so formed are planar and have the special feature (Fig. 1) that the glycosidic bonds linking the bases to deoxyribose are the same distance apart in both pairs, and are all inclined at the same angle to a line joining equivalent atoms at the ends of the glycosidic bonds. Hence, all glycosidic bonds are equivalent, irrespective of the base to which they are attached. Therefore, if the base-pairs are stacked on each other, a regular helical molecule can be built. In such a structure there is no restriction on base sequence except that base-pairing results in the sequence in one chain being complementary to that in the other. If we ignore the differences between the bases, there are two symmetry elements in such a helical molecule, the screw or helix

axis and a series of dyad axes, each in the plane of a base-pair and relating the two glycosidic bonds in the pair, and, in addition, the deoxyribose and phosphate groups. These dyad axes operate on the whole deoxyribose-phosphate chain, and as a result the sequence of

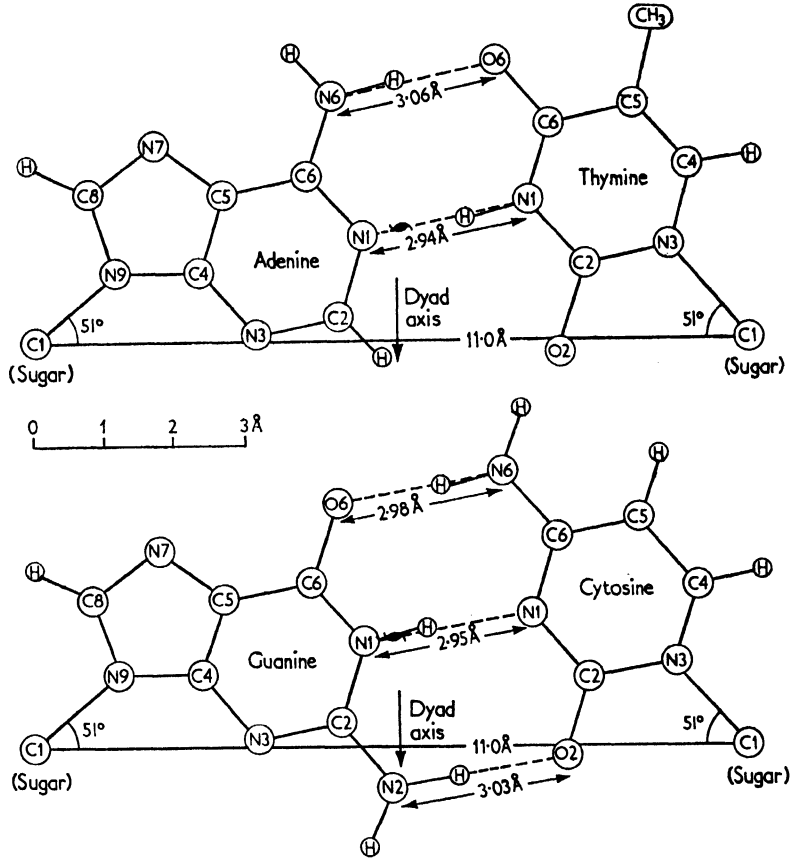


FIG. 1. Watson-Crick base-pairing scheme for DNA (refined by Spencer). In the *B* configuration of DNA, the helix axis is in the position shown on the dyad axis relating the C1—N9 and C1—N3 glycosidic bonds.

atoms runs in opposite directions in the two chains (such chains are said to be anti-parallel).

The regularity of structure of the molecule is the key to the mechanism of its replication. In essence the process is as follows. First, the two chains separate; next, each chain acts as a template on to which individual nucleotides are attached by base-pair hydrogen bonding. The spatial relationship of the glycosidic bonds in the base pair is exactly specified

during this process (e.g. by an enzyme). As a result, the hydrogen bonding is highly specific and the nucleotides are arranged in the sequence determined by the parent chain. The nucleotides are then joined together to form a polynucleotide chain. The complete DNA molecule so formed is identical to the parent molecule. After one replication, two daughter molecules are formed, each containing one chain of the parent molecule.

The base-pairing mechanism is a powerful means of determining interaction between polynucleotide chains. It might well be said that such interactions are the most fundamental macromolecular interaction in living things, for, apart from replication of DNA, the same process appears to operate when the genetic message is transmitted from DNA to the polynucleotide chain of messenger-RNA, and possibly at other stages in protein synthesis. (RNA is structurally very similar to DNA but contains an OH group attached to the C2 atom of the sugar ring.) Messenger-RNA carries the genetic base sequence to the seat of protein synthesis. The next step required in protein synthesis is attachment of the amino acids to the nucleotide triplets that code for them in the messenger-RNA. It is known that each amino acid is linked to a transfer-RNA molecule specific for that amino acid. It is likely that, once more, base-pairing is operative and that a nucleotide triplet in the transfer-RNA molecule is attached by base-pairing to the coding triplet in the messenger-RNA. In this way the amino acid is connected to the relevant coding triplet. The amino acids are thus arranged in the correct sequence along the messenger-RNA molecule and may then be joined together to form the polypeptide chain of the protein.

It is clear from these considerations that the primary structural problem in the field of nucleic acids is to define the base-pairs. The best way to do this is to determine the structure of DNA itself. In general, too, the structure of DNA and of the various types of RNA provide a basis for understanding their function.

3. The Molecular Configuration of DNA

Three well defined conformations of the DNA molecular double-helix have been observed. These are known as the *A*, *B* and *C* configurations. The *A* has eleven nucleotide pairs per helix turn, the pitch is 28 Å and the plane of the base-pairs is tilted 20° from perpendicular to the helix axis (Fig. 2). The *B* form has ten nucleotide pairs per turn of 34 Å pitch and base-pairs roughly perpendicular to the helix axis (Fig. 3). This form is found *in vivo*. The *C* form is very similar to the *B* except that there are $9\frac{1}{3}$ nucleotide pairs per helix turn and the base-pairs are tilted about 6°. These configurations have been studied by obtaining diffraction photographs of fibres consisting of DNA molecules oriented parallel to the

TABLE I. Summary of various forms of DNA in fibres

Configuration of molecule	Number of nucleotide pairs per turn of helix	Inclination of base pairs in molecule	Salt	Relative humidity and condition necessary	Crystal class	Crystallinity	Molecular positions	Unit-cell dimensions a (Å)	b (Å)	c (Å)	$\beta°$
A	11·0	20°	Na K Rb	75%	mono-clinic	crystalline	$\begin{cases} 0, 0, 0 \\ \frac{1}{2}, \frac{1}{2}, 0 \end{cases}$	22·24	40·62	28·15	97·0
B	10	~0°	Li	66% 3% LiCl in fibre	ortho-rhombic	crystalline	$\begin{cases} 0, 0, \frac{1}{8} \\ \frac{1}{2}, \frac{1}{2}, -\frac{1}{8} \end{cases}$	22·5	30·9	33·7	—
			Li	75–90%	ortho-rhombic	semi-crystalline	$\begin{cases} 0, 0, \frac{1}{8} \\ \frac{1}{2}, \frac{1}{2}, -\frac{1}{8} \end{cases}$	24·4	38·5	33·6	—
			Li Na K Rb	92%	hexagonal	semi-crystalline	$\begin{cases} 0, 0, 0 \\ \frac{1}{3}, \frac{2}{3}, \frac{1}{6} \\ \frac{2}{3}, \frac{1}{3}, -\frac{1}{6} \end{cases}$	46	—	34·6	—
B_2	9·9	0°?	Na	75% under tension	tetragonal	semi-crystalline	$\begin{cases} 0, 0, 0 \\ \frac{1}{2}, \frac{1}{2}, \frac{1}{4} \end{cases}$	27·4	—	33·8	—
C	9·3	−5°	Li	44% no LiCl	ortho-rhombic	semi-crystalline	$\begin{cases} 0, 0, \frac{1}{8} \\ \frac{1}{2}, \frac{1}{2}, -\frac{1}{8} \end{cases}$	20·1	31·9	30·9	—
			Li	44% in some specimens only. No LiCl in fibre	hexagonal	semi-crystalline	$\begin{cases} 0, 0, 0 \text{ or } \frac{1}{2} \\ \frac{1}{3}, \frac{2}{3}, \frac{1}{6} \text{ or } -\frac{1}{3} \\ \frac{2}{3}, \frac{1}{3}, -\frac{1}{6} \text{ or } \frac{1}{3} \end{cases}$	35·0	—	30·9	—

fibre axis. Depending on the alkali metal neutralizing the charge of the phosphate groups, and on the amount of water and salts in the fibre, the various configurations may be obtained in a number of crystalline and semi-crystalline forms (Table I).

The Fourier transform method has been used to determine the *A*, *B* and *C* configurations. The general form of the molecule, the number of nucleotides per helix turn, the helix pitch, etc., were deduced from the X-ray patterns. Molecular models were built and adjusted until agreement was reached between calculated and observed intensities. In all cases the structure of the base-pairs and the symmetry of the helical molecule were defined by the structural hypothesis of Watson and Crick.

The most accurate intensity data have been obtained from patterns of the crystalline *A* and *B* forms. Data from semi-crystalline forms are not so accurate and do not give three-dimensional information.

A. *Difficulties in the structural analysis of DNA*

There are several difficulties in the X-ray study of DNA that are not normally present in structural analyses of crystalline compounds.

(1) Because single crystals are not available and fibre patterns have to be used, a proportion of the reflections overlap and their intensities cannot be obtained separately. This difficulty is minimized by using a finely collimated X-ray beam (Fig. 4).

(2) In the fibre, the microcrystals are disoriented about the fibre (or *c*) axis. As a result, reflections some distance from the meridian appear weak, and are not observed at wider angles. The data therefore lack resolution in directions at right angles to *c*. In the meridional direction, reflections extend to 1·7 Å (Fig. 5) and an isolated reflection has been observed on the meridian at 1·1 Å for the crystalline *B* form.

(3) Lack of precise parallelism of the microcrystals causes the reflections to be extended into arcs. This is an important factor limiting the maximum angle at which any diffraction is observed.

(4) Although crystalline imperfection contributes to the difficulty of X-ray study of DNA, it is not the most important factor limiting the diffraction data. The intensity data indicate that the temperature factor ($B = 4$ Å^2) is the same as that found in most organic crystals. If single crystals of DNA could be obtained, it appears that the intensity data would not be much inferior to those obtained from crystals of organic compounds of quite low molecular weight.

(5) It should be emphasized that the success of the structure analysis of DNA has depended largely on knowledge of all the covalent links within DNA and of the stereochemistry of the component parts. Also helpful was direct evidence that the bases were hydrogen-bonded and indirect evidence that they occurred in pairs. But initially, a hypothesis

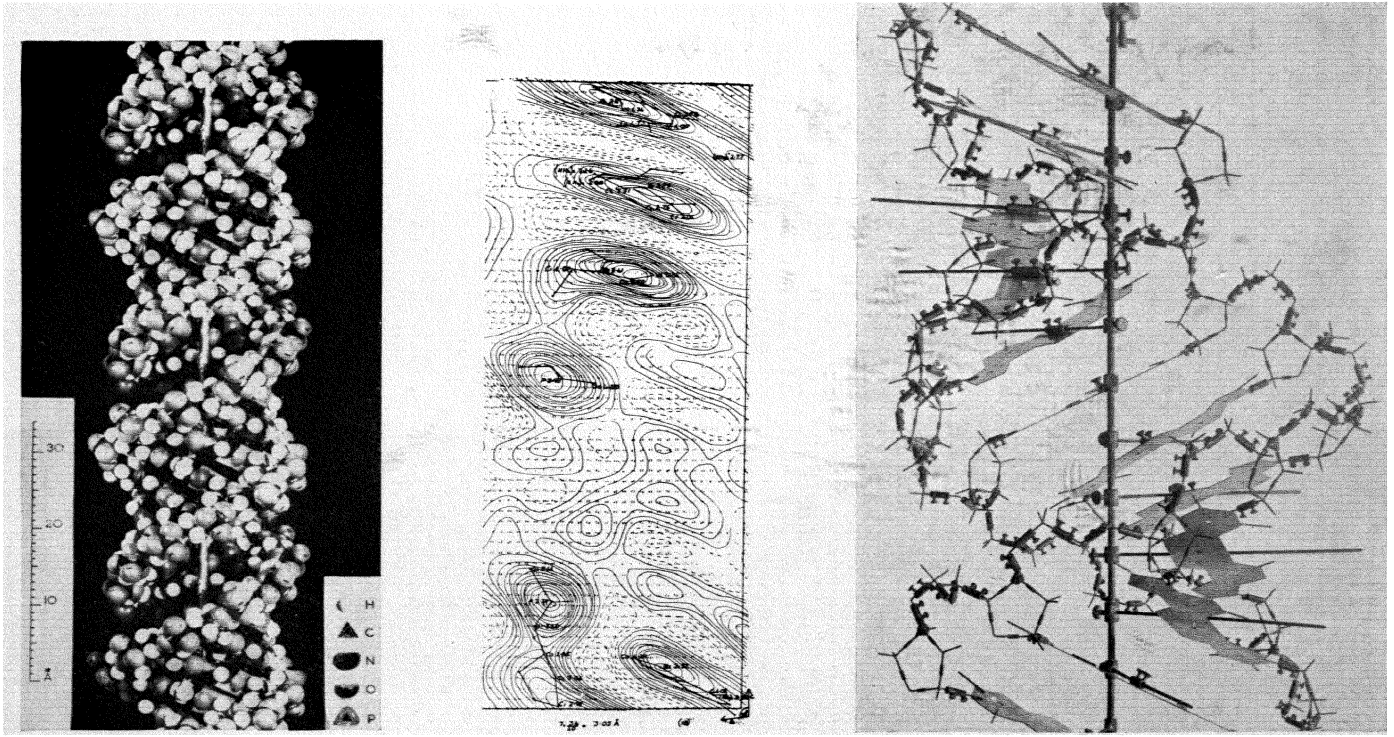


FIG. 2. *Right:* Molecular model of the *A* configuration of DNA. The helix axis is marked by a vertical rod. Thin wires lie along the covalent bonds except in the bases which are represented by plates; centres of atoms are at the intersections of the straight portions of wire. *Centre:* Electron density F_o Fourier map of a section of the molecule. The plane of the section passes close to the helix axis and is parallel to the plane of the photograph of the model. The section is on the same scale as the wire model. *Left:* View (on a smaller scale) of a space-filling model of the structure. The radii of the spherical parts in the model correspond to van der Waals radii.

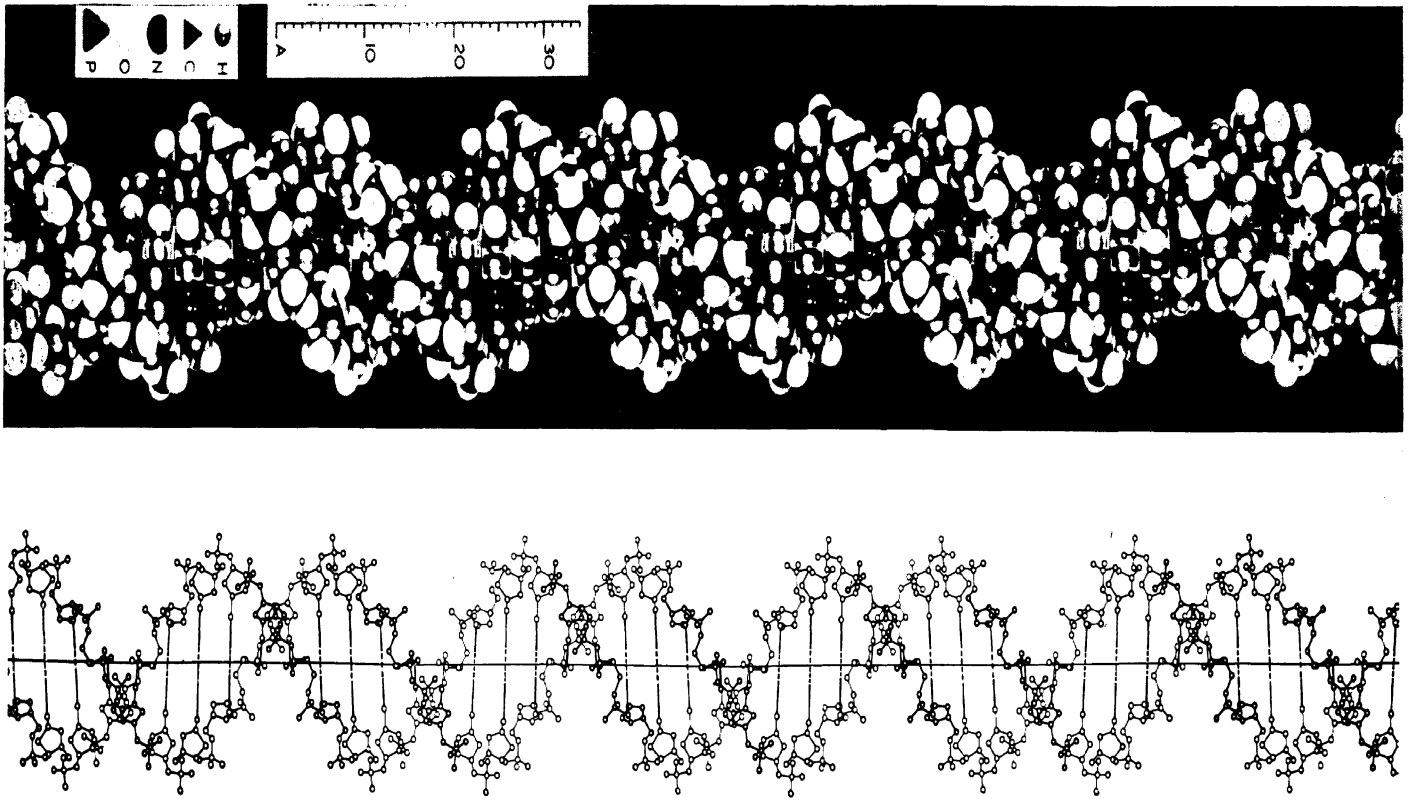


FIG. 3. *Lower:* Diagram of a projection of the *B* configuration of DNA. *Upper:* Corresponding view of a space-filling molecular model.

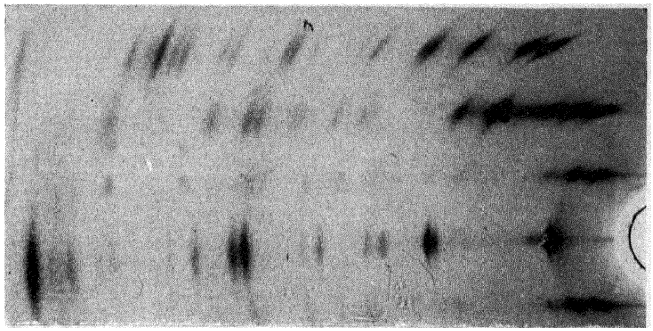


FIG. 4. Central portion of the diffraction pattern of an oriented fibre of microcrystalline DNA in the *B* configuration. Many single reflections are resolved.

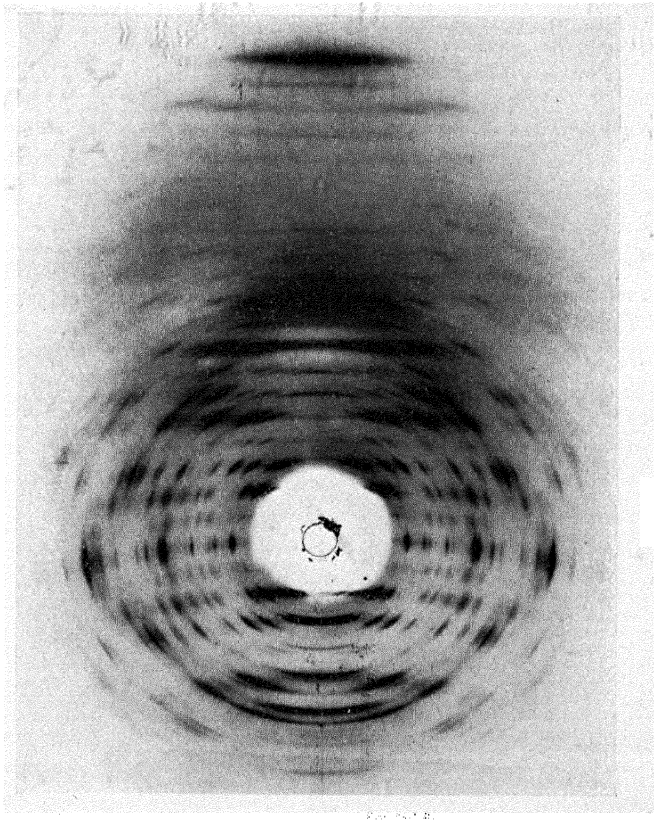


FIG. 5. Diffraction pattern of an oriented fibre of microcrystalline DNA in the *B* configuration. The fibre is in a vertical plane and is set at such an angle with respect to the X-ray beam that the (0020) reflection (1·68 Å) is recorded on the meridian at the top.

FIG. 6. Transfer-RNA viewed in the polarizing microscope with crossed nicols. The uniform areas probably correspond to single crystals. The field of view is about 30 μ wide.

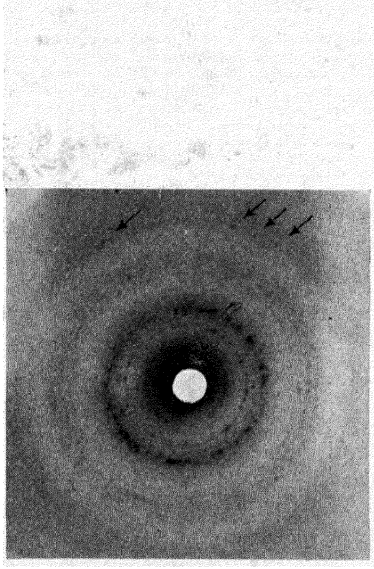


FIG. 7. Diffraction photograph of microcrystalline transfer-RNA, showing spots corresponding to reflections from single crystals. The arrows point to reflections from planes ~ 6 Å apart.

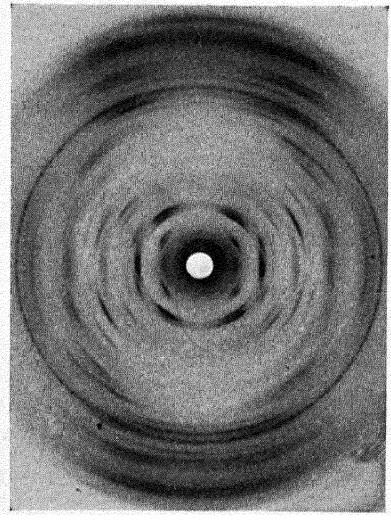


FIG. 8. *A*-Type pattern of DNA. Double-orientation in the fibre is indicated by differences of intensities of reflections in neighbouring quadrants.

(due to Watson and Crick) was required to specify which hydrogen bonds were involved in the pairing. However, one-third of the scattering matter in crystals of DNA is water and ions and there is little stereochemical guidance in deciding the positions of water molecules and of ions. This difficulty has been dealt with by using lithium ions, because they scatter little, and by regarding the water as a medium of uniform electron density filling the space between DNA molecules. If some of the water molecules occupy fixed crystallographic positions these should become evident on Fourier synthesis maps. Signs of such water molecules have been found.

4. Recent Developments in X-Ray Study of Nucleic Acids

(a) *The possibility of studying single crystals instead of oriented fibres*

Recent study of transfer RNA has shown that slow crystallization gives highly birefringent regions which are optically homogeneous and many μ across (Fig. 6). These regions appear to be single crystals. This is confirmed by X-ray study; well defined spots on the diffraction photographs correspond to reflections from single crystals (Fig. 7). Thorough study of this kind of crystal growth has not yet been made for DNA. There have, however, been several indications that large crystals of DNA might be grown. Some fibres of DNA show double-orientation (Fig. 8). With such a fibre it should be possible to obtain a series of diffraction patterns for different orientations of the fibre about its axis. In this way separate intensity values could be obtained for reflections that overlap when singly oriented fibres are used.

(b) *Use of Fourier synthesis methods in study of DNA*

A stage is reached in studying DNA structure by the Fourier transform method when adjustments in the model cease to give appreciable increase of agreement between calculated and observed intensities. F_o Fourier syntheses may be calculated using phases derived from the molecular model. Use of such syntheses and of difference syntheses has several advantages over the Fourier transform method as applied to DNA. Provided that the molecular model corresponds fairly closely to the real structure, need for refinement of the model will be indicated. The nature of the refinement will be given directly and the effects of refining one part of the model, e.g. the base-pair, may be studied separately from that of other parts. The presence of water molecules and ions in fixed positions should also be indicated.

Some Fourier synthesis studies by Dr. S. Arnott of DNA structure will now be very briefly described. Because the DNA diffraction data are of low resolution and have various special features, it is very desirable to

test that the Fourier synthesis method works satisfactorily when applied to DNA. That the method is satisfactory is indicated by F_o–F_c syntheses calculated using a molecular model with the base-pair displaced in its plane 1·5 Å from the position believed to be correct. The section of the synthesis in the plane of a base-pair (Fig. 9) clearly indicates the need for moving the base-pair in the expected direction. F_o–F_c syntheses were then used to find whether the Watson-Crick base-pairing scheme or an alternative Hoogsteen scheme is more nearly compatible with the

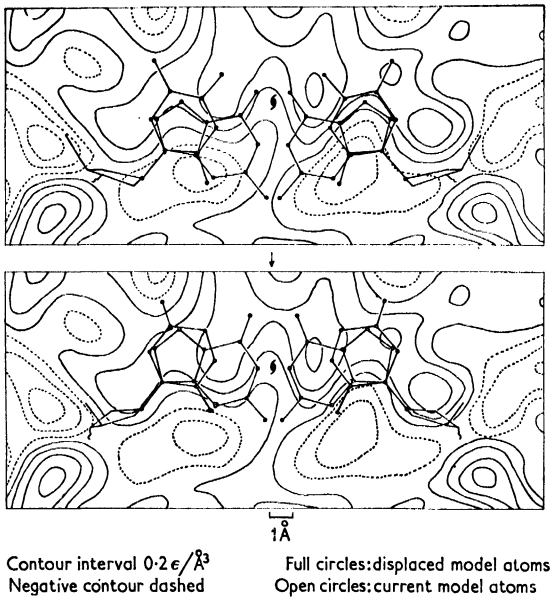


FIG. 9. Fourier difference synthesis for *B* configuration of DNA with Watson-Crick base-pairs displaced 1·5 Å from the correct position. The top diagram shows the base-pair in the displaced position; the correct position is indicated in the lower diagram.

X-ray data. The F_o map for the Watson-Crick scheme is shown in Fig. 10. The F_o–F_c map shows a fairly uniform electron density within the area of the base-pair. Regions of moderate height occur at the edge of the base-pair region. It may well be that the positive regions correspond to water molecules hydrogen-bonded to the bases. In contrast the corresponding F_o–F_c map for the Hoogsteen-type DNA model (Fig. 11) has a large positive region extending into the centre of the base-pair. This indicates that the Hoogsteen base-pair has insufficient scattering matter in the centre of the base-pair. It is evident that the Watson-Crick base-pair is to be preferred. There are, in fact, several other reasons why the Hoogsteen pair is not likely to apply to DNA.

Dr. D. A. Marvin has also used the Fourier synthesis method. He found, on Fourier sections, peaks that may correspond to water molecules in fixed positions (Fig. 12). A difficulty in such a study is that the resolution of the data is only just sufficient to resolve the oxygen atoms of water molecules.

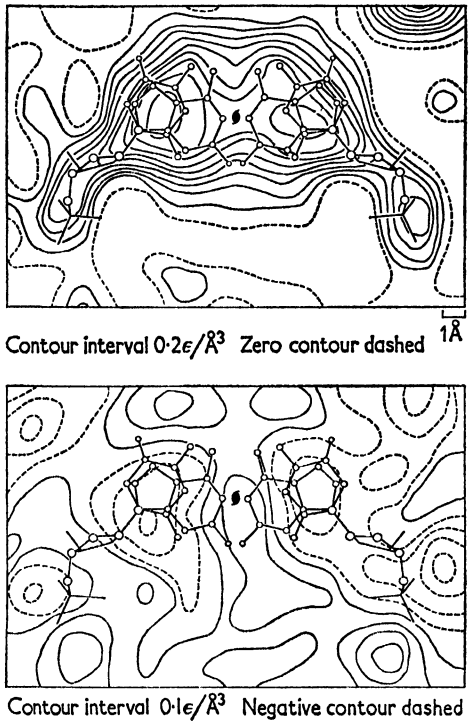


FIG. 10. Top: F_o Fourier synthesis for DNA *B* model with Watson-Crick base-pairs and which is believed to be close to the correct structure. Bottom: Corresponding difference synthesis. The section is in the plane of a base-pair.

(c) *Application of least-squares methods in studying X-ray data from DNA*

In refining a model of a large molecule it may be useful to fix the stereochemistry of part of the molecule and to move that part as a whole during refinement. This procedure was used during the Fourier transform study of DNA structure: the base-pair, deoxyribose and phosphate parts of DNA were regarded as rigid bodies. A least-squares analysis of this kind is being made by Drs. S. Arnott and C. L. Coulter using an I.B.M. 7090 computer, three translations and three rotations being applied to the three rigid groups. After such refinement, it is hoped that a new molecular model can be built with base, deoxyribose and phosphate

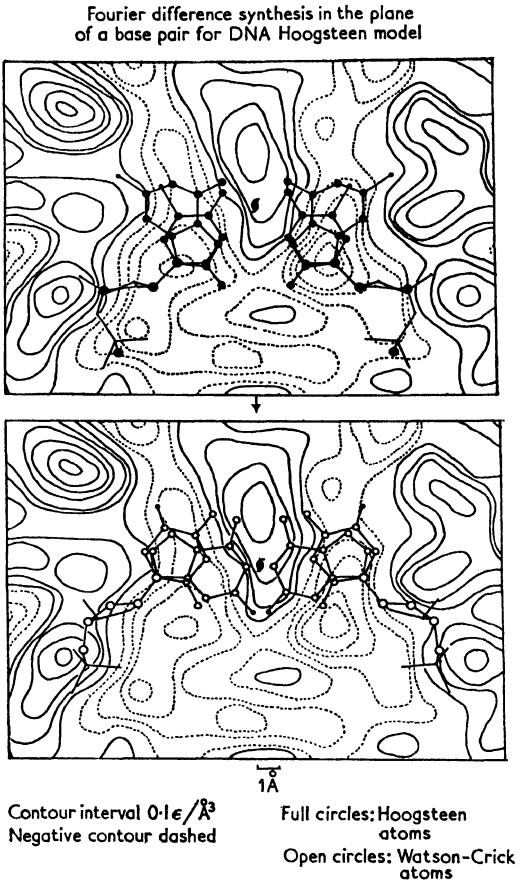


FIG. 11. Fourier difference synthesis for DNA model with Hoogsteen base-pairs. The top diagram shows the Hoogsteen base-pairs, the lower diagram shows Watson-Crick pairs superimposed on the Hoogsteen differences synthesis.

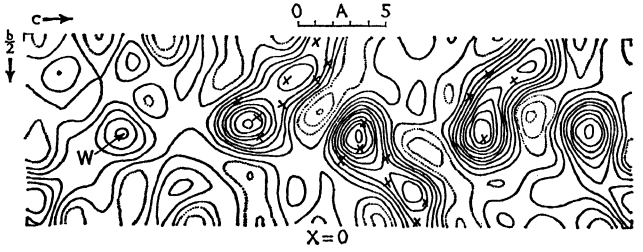


FIG. 12. F_o Fourier synthesis section in a plane parallel to the helix axis (vertical) and through phosphate groups of two neighbouring DNA molecules in *B* configuration. The X marks show positions of atoms in the molecular model. The peak at W and others adjacent to it may correspond to water molecules in fixed positions.

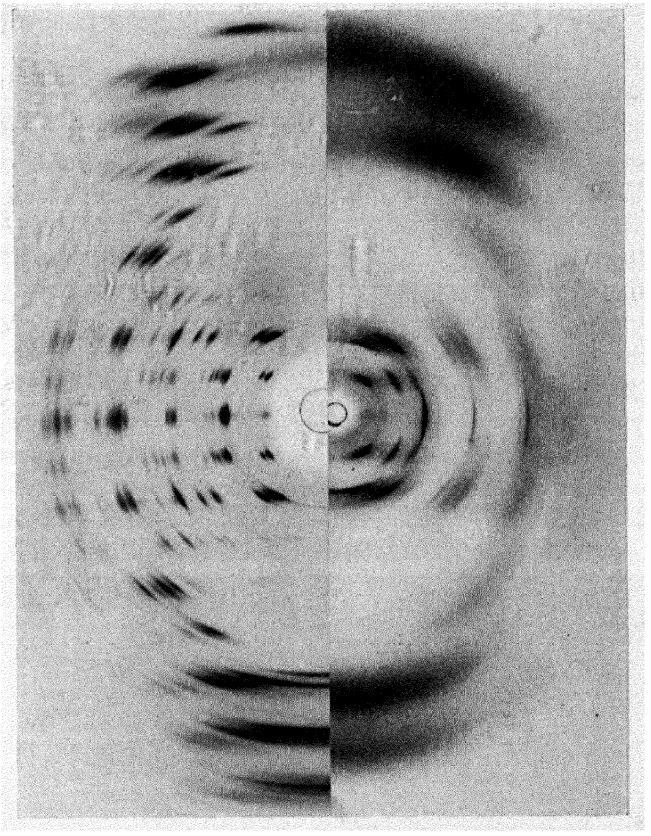


FIG. 13. Similarity of the diffraction patterns of oriented fibres of crystalline RNA (*right*) and DNA (*left*). Sharp reflections are visible in the central regions of the RNA pattern.

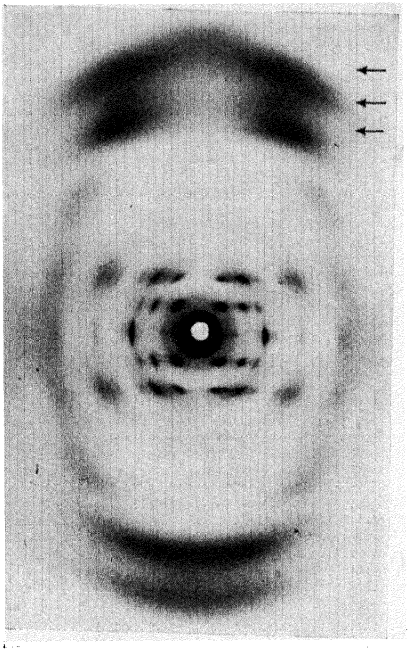


FIG. 14. Diffraction pattern of oriented fibre of transfer-RNA, showing resolution of the higher layer lines (indicated by arrows).

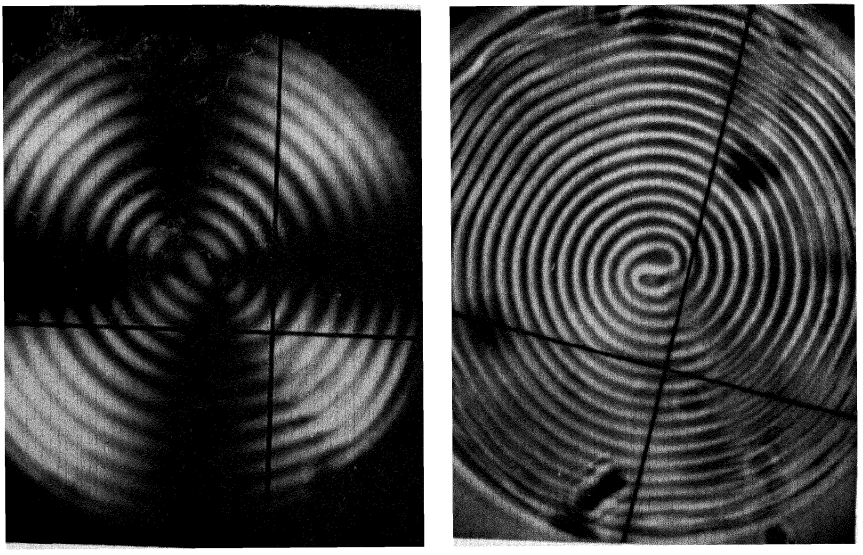


FIG. 16. Liquid-crystalline spherulites (about 50 μ diameter) of transfer-RNA, showing spiral forms due to varying refractive index. *Left:* Crossed nicols. *Right:* Unpolarized light.

parts joined in stereochemically reasonable fashion and in approximately the positions indicated by the least-squares analysis. It is expected that refinement will move few atoms more than a small fraction of 1 Å.

5. The Helical Structure of RNA

For many years well-defined X-ray diffraction patterns could not be obtained from RNA of any kind and no satisfactory interpretation had been given of the diffuse patterns available. Recently much progress has been made as a result of the discovery that suitably prepared transfer RNA can be crystallized by slow drying. Although the RNA molecules contain only about eighty nucleotides, it has been possible to orient the molecules in microcrystalline fibres. The diffraction patterns given by such fibres are of the same quality as those from semi-crystalline DNA. They are well-oriented and show sharp reflections in the central region (Fig. 13). In the outer regions the reflections become diffuse, this apparently being due to the helical molecules being disordered (like those in semi-crystalline DNA) by random screwing about the helix axes. The patterns are sufficiently well-oriented and defined for it to be clear that they are essentially the same as the *A*-type DNA fibre pattern (Fig. 13). It is only in the best RNA patterns that separation of the higher layer lines is obvious (Fig. 14). It is clear that the RNA has a helical structure essentially the same as that of DNA. However, the distribution of intensities on the higher layer lines, and in the central region of the pattern, is somewhat different from that in the DNA pattern. These differences appear to be due mainly to a slight difference of position of the phosphate group in RNA and DNA. The presence of the hydroxyl group in the ribose does not produce steric difficulties.

The structure of the helical part of the RNA molecule is clearly defined but little is known about other aspects of the structure. The positions of the ends of the polynucleotide chain are not known. By considering the molecular weight and the sedimentation rate it appears likely that each molecule consists of one polynucleotide chain folded-back on itself (like a hairpin). One half of the chain is twisted round the other and joined to it by Watson-Crick hydrogen bonds, thus forming one continuous helix of about $3\frac{1}{2}$ turns (Fig. 15). The base sequences must be complementary in the two parts of the chain which form the double-helix. As in the DNA, the sequence of atoms in the two parts runs in opposite directions along the helix, i.e. the chains are anti-parallel. The ends of the polynucleotide chain might both be at one end of the molecule or elsewhere (Fig. 15). If the ends are at one end of the molecule, there need be only one folded part of the chain joining one side of the helix to

the other. At such a fold, a length of chain including three or more nucleotides is required to join the two sides of the helix. Whereas the bases in the helical region are hydrogen-bonded in pairs, the three bases at the folded end have free hydrogen-bonding sites. It is possible that these three bases form Watson-Crick pairs with the three bases in the coding triplet in messenger RNA. In this way the transfer RNA could be connected to the coding triplet. The amino acid is known to be attached to one end of the transfer RNA polynucleotide chain.

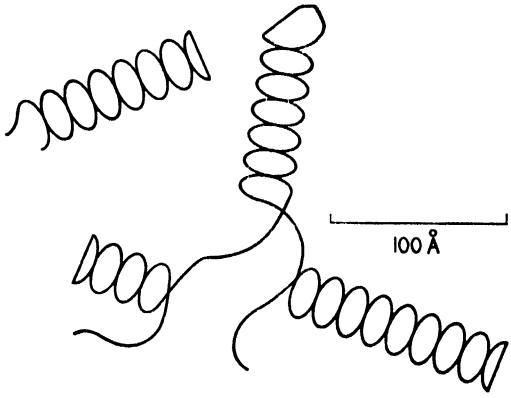


FIG. 15. Diagram showing the form of RNA molecules. At the top left is a transfer RNA molecule; the other part of the diagram indicates the probable form of virus or ribosome RNA molecules.

The deduction of complementary sequences within each molecule suggests that the molecule might be self-replicating. To replicate, the molecule would unfold and a second polynucleotide chain would form on the parent chain. If the two halves of each chain contained complementary sequences, both chains would be identical in their helical regions. After separating, each chain would fold into a double helix. However, the three or more unpaired bases in the folded ends of the parent and daughter molecules would not necessarily be identical.

(a) *Structure of the helical regions in virus and ribosome RNA*

The crystalline diffraction patterns from transfer RNA make clear the nature of the diffuse patterns obtained from virus and ribosome RNA. There is no doubt that they are essentially the same as the crystalline pattern, although there are minor variations: every feature of the diffuse patterns can be accounted for in terms of the crystalline pattern, and the small though characteristic changes with humidity are the same. One can be confident, therefore, that all types of RNA so far studied contain helical regions with a configuration essentially similar to that of *A* DNA.

The base compositions of virus and ribosome RNA are such that at least 10% of the material cannot exist in a DNA-like structure with complementary sequences. In agreement with this, solution studies indicate that the helical content of tobacco mosaic virus RNA is 88% and that of ribosome RNA is 77%. These studies also indicate that the molecules consist of single polynucleotide chains and that, where parts of the chain contain approximately complementary sequences, these parts may fold back on themselves to form double-helices, in which the two parts of the chain are anti-parallel (Fig. 15). The X-ray results show that the chains are anti-parallel. This provides further evidence that the helices in virus and ribosome RNA are formed by folding back of the chain.

6. Polarizing Microscope Studies of Transfer-RNA and its Liquid Crystal Forms

Solutions of transfer-RNA, on drying, show in the polarizing microscope a number of striking phenomena which corroborate the conclusions drawn from X-ray studies.

A dried layer formed at an interface with air is negatively birefringent. This indicates that the molecules are rod-shaped, negatively birefringent, and orient, as one would expect, parallel to the interface. Similarly, the sheared material in a drawn fibre of the kind used in the X-ray experiments has negative birefringence. The value of the birefringence is 0·07, which is consistent with the fibres consisting entirely of helical molecules like those of DNA in the *A* configuration, the high birefringence being due mainly to the bases lying roughly perpendicular to the length of the molecule.

The rod shape and negative birefringence of the molecules are confirmed by the properties of various liquid crystal forms which develop in concentrated solutions of transfer-RNA on standing. As expected, tactoids are negatively birefringent and, in spherulites, the direction of the greater refractive index is radial. Furthermore, the small axial ratio of the tactoids ($\sim 1{\cdot}5:1$), and the very small difference of refractive index between ordered and disordered regions, both indicate that the axial ratio of the molecules is not large. In agreement with this, the structure proposed has an axial ratio of 5:1.

In addition to simple spherulites there is a remarkable type of spherulite which, viewed from most angles, shows a double-spiral due to varying refractive index (Fig. 16). Similar bodies were found in solutions of α-helical poly-γ-benzyl-L-glutamate and their nature elucidated. A similar form was found in gels of DNA.

A new type of spherulite is found in the RNA, the spiral form being

replaced by concentric circles. It appears that both types of spherulite contain similar structures, except that the spiral type has a radial line of disinclination, whereas the circular type has a line of disinclination along two radii which form a diameter.

The variation of refractive index in these bodies is due to the varying direction of the molecules. The molecules lie parallel in layers with the length of the molecule in the plane of the layers. The direction of the molecules in one layer is, however, inclined at a small angle to that in the neighbouring layer. As a result, the direction of the molecule rotates on passing through a succession of layers, so that the whole assembly is twisted. The inclination of the molecules in one layer to those in the next shows that the packing forces between the molecules must be asymmetric and therefore that the molecules are asymmetric. A helical molecular structure has the required symmetry.

(a) *Can X-ray diffraction analysis provide the base sequence of transfer RNA?*

Since the biological specificity of nucleic acids is determined by the base sequences in them, the determination of these sequences is the most fundamental problem of nucleic acid research today. The number of bases in a DNA molecule is too large for a determination of base sequence by X-ray diffraction to be feasible. However, in transfer-RNA the number of bases is not too large. The size of the crystals observed by means of X-ray diffraction and polarization microscopy suggests that it should not be too difficult to grow crystals large enough for single-crystal X-ray analysis. If a pure preparation of transfer RNA-specific for one amino acid were used, one might expect that the crystals would be free of the disorder at present found and would be as perfect as those in microcrystalline DNA. If such crystals were obtained, and provided that all the molecules had identical structure and base sequence, a complete analysis of the primary structure of the molecule could be made, including the sequence of the bases and the form of the fold at the end of the helix.

Acknowledgements

The X-ray work described here has been done in the main by a group in our laboratory, now consisting of Drs. S. Arnott, W. Fuller, M. Spencer and myself. DNA and RNA were prepared respectively by our collaborators Dr. L. D. Hamilton of the Sloan-Kettering Institute, New York, and Dr. G. L. Brown of our Unit. Acknowledgements to all the workers who have helped us in the various researches are in the references below; here we mention specially Sir John Randall's encouragement, computing facilities and free time provided by the University of London Computer Unit and I.B.M. United Kingdom Ltd., and help with the photography by Mrs. F. Collier and Mr. Z. Gabor.

Some Key References to Various Aspects of the Structure of Nucleic Acids

The Watson-Crick model for DNA and replication hypothesis: Crick, F. H. C. and Watson, J. D. (1954). *Proc. roy. Soc.* A **223**, 80; Watson, J. D. and Crick, F. H. C. (1953). *Nature, Lond.* **171**, 737, 964.

Early X-ray studies of nucleic acids: Astbury, W. T. (1947). *Symp. Soc. Exp. Biol., I. Nucleic Acids*, p. 66. Cambridge University Press; Wilkins, M. H. F., Stokes, A. R. and Wilson, H. R. (1953). *Nature, Lond.* **171**, 738.

A and *B* configurations of DNA: Franklin, R. E. and Gosling, R. G. (1953). *Acta cryst.* **6**, 673; Langridge, R., Seeds, W. E., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F. and Hamilton, L. D. (1957). *J. biophys. biochem. Cytol.* **3**, 767.

Detailed account of the *B* configuration: Langridge, R., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F. and Hamilton, L. D. (1960). *J. mol. Biol.* **2**, 19, 38.

X-Ray study of RNA: Spencer, M., Fuller, W., Wilkins, M. H. F. and Brown, G. L. (1962). *Nature, Lond.* **194**, 1014.

The Genetic Code: Crick, F. H. C. (1962). *Sci. Amer.* October, p. 66.

Biochemical background: Davidson, J. N. (1960). *The Biochemistry of Nucleic Acids*, 4th Ed. Methuen, London.

Discussion of base-pairing schemes proposed as alternatives to that of Watson and Crick: Wilkins, M. H. F. (1960). In *Biological Structure and Function*, Vol. 1. Academic Press, New York.

Spiral forms in liquid crystals: Conmar Robinson (1961). *Tetrahedron* **13**, 219.

Discussion

O. SIDDIQI (Tata Institute of Fundamental Research, Bombay): I should like to know which fraction of the RNA do the pictures correspond to.

M. H. F. WILKINS: Transfer RNA.

O. SIDDIQI: In that case, it seems this RNA is largely hydrogen bonded. If so, I suppose one should expect a regularity of the base ratios, well-known for DNA.

M. H. F. WILKINS: Yes, you are quite right. It is a necessary consequence of this sort of structure and you have an approximately equal ratio. In other types of RNA which have only part of the molecule helical, the base ratio is not 1:1.

E. KATCHALSKI: We are working at present on the separation of the various fractions of transfer RNA. The relatively purified fractions obtained so far possess the formula t-RNA–amino acid–polypeptide. I wonder therefore whether such t-RNA–polypeptidyl derivatives are suitable for precise X-ray analysis.

M. H. F. WILKINS: How big is the peptide?

E. KATCHALSKI: Say, four to five hundred angstroms.

M. H. F. WILKINS: It will be interesting to try this. There are two difficulties, in trying to determine base sequences in RNA. One is to produce a fractionated transfer RNA which is specific for only one amino acid. Secondly, owing to the degeneracy in the coding, more than one sequence may be specific for one amino acid. However, if we can get RNA molecules which have a large proportion of the base sequence the same, and if they crystallize well, then X-ray studies may establish the common part of the sequence.

S. OCHOA: I would like to know whether the last slide shown by you was obtained with fractionated transfer RNA.

M. H. F. WILKINS: It was a mixture. So far we have obtained the best patterns only with the mixtures. With what we have tried so far, the crystalline perfection of RNA crystals is not as good as DNA.

The Triple Helical Structure of Collagen

G. N. RAMACHANDRAN

Department of Physics, University of Madras, Guindy, Madras, India

ABSTRACT

The triple helical configuration of polypeptide chains, first put forward from Madras in 1954, has been found to be the basis of the structure of collagen and a number of related polypeptides, namely, polyglycine II, poly-L-proline II and poly-L-hydroxyproline-A. The original prototype structure, with three-fold screw symmetry, is found for the polypeptides, while collagen itself takes a closely similar structure, with a further twisting of all the three chains about the central axis of the protofibril. Recent work at Madras has shown that the value of the twist for a height of three residues is close to 30°. Also, a value of 2·95 Å was found for the residue height in unstretched collagen, which is distinctly larger by about 0·1 Å than the previous determinations. Conforming to these data for the helical framework, a satisfactory structure has been worked out for collagen, with two hydrogen bonds for every three residues. This structure does not have any objectionable short contacts, and it is distinctly superior to the Collagen II structure of Rich and Crick, which contains only one hydrogen bond for three residues and is also much less densely packed. The latest structure agrees well with X-ray and infra-red data (better than the one-bonded structure) and it also explains why collagen has a coiled-coil structure, unlike the simpler polypeptides. It is only very slightly different from the earlier (1955) model of Ramachandran and Kartha, and is topologically identical with it. When the sequence Gly–Pro–Hypro occurs in one of the chains, only five out of the six hydrogen bonds can be formed for a height of three residues, and a satisfactory alternative structure has been worked out for this case also.

The evidence for the triple helical structure from other sources is briefly reviewed and the possibility of this structure occurring in other proteins is discussed. A table is given comparing the main properties of the alpha helix and the Madras triple helix, which now appear to be the two main helical chain configurations which occur in polypeptides and proteins.

1. INTRODUCTION

The prototype of the currently accepted collagen triple helix was first put forward in 1954 from this laboratory (Ramachandran and Kartha, 1954). It was a departure from all the previous proposals in that it consisted of three helical polypeptide chains standing side by side, which were stabilized only by *inter*-chain hydrogen bonds. The structure was worked out essentially from the observation that collagen contained one-third the number of amino acid residues in it as glycine. This should

have a counterpart in the structure in that one-third the number of sites in the chains must be such that no β-carbon atom can occur in them. It is readily seen that such a condition is impossible to achieve in a single helical structure, but that it is possible, in principle, to do so in a triple helical structure. When these ideas were further worked out, it was found that a reasonably satisfactory structure could be postulated, in which each of the chains had a three-fold screw symmetry and the three chains were also related to each other by a three-fold screw. If the amino acid residues were assumed to have the naturally occurring L-configuration, then the individual helices were left-handed (symmetry 3_2) and the three chains were also related by the same symmetry 3_2.

At about the same time, various attempts were made to analyse the observed X-ray diffraction pattern in terms of a helical arrangement in the structure. Several possibilities were suggested by Cohen and Bear (1953) and Cowan *et al.* (1953). Following the postulation of the triple helical structure mentioned above, Ramachandran and Ambady (1954) re-examined the X-ray diffraction pattern and showed that the individual chains should have ten residues in three turns, or $3\frac{1}{3}$ residues per turn, instead of three residues per turn as was assumed in the prototype structure. This only necessitated a small alteration in the earlier structure and this was carried out by Ramachandran and Kartha (1955*a*, *b*). In fact, it was found that the hydrogen bonds were more satisfactory in this structure than in the earlier one with three residues per turn. The main difference from the prototype structure was that the three chains now formed coiled-coils and were all twisted around a common central axis. In fact, such a second coiling is demanded whenever the number of units per turn in the individual chains is not integral in a multiple-chain structure and there are systematic linkages between the different chains. The prototype structure and the coiled-coil structure are shown schematically in Fig. 1. The latter may be obtained from the prototype structure by twisting it about the common axis. In Fig. 1 (b), the twist for three residues is taken to be 30°, which is the latest value (see below).

The structure of Ramachandran and Kartha (1955*a*, *b*) contained *two* systematic hydrogen bonds of the type NH . . . O for every three residues, while the third NH-group was not involved in forming an internal hydrogen bond as this group was pointing out from the triple helical protofibril. However, this nitrogen could readily form a part of the five-membered ring of an imino-acid residue. Thus, the structure could readily accommodate a large amount of proline and hydroxyproline residues, which is characteristic of collagen. Apart from this good agreement with all the chemical information on collagen available at that time, the structure also agreed well with infra-red dichroism and X-ray diffraction data.

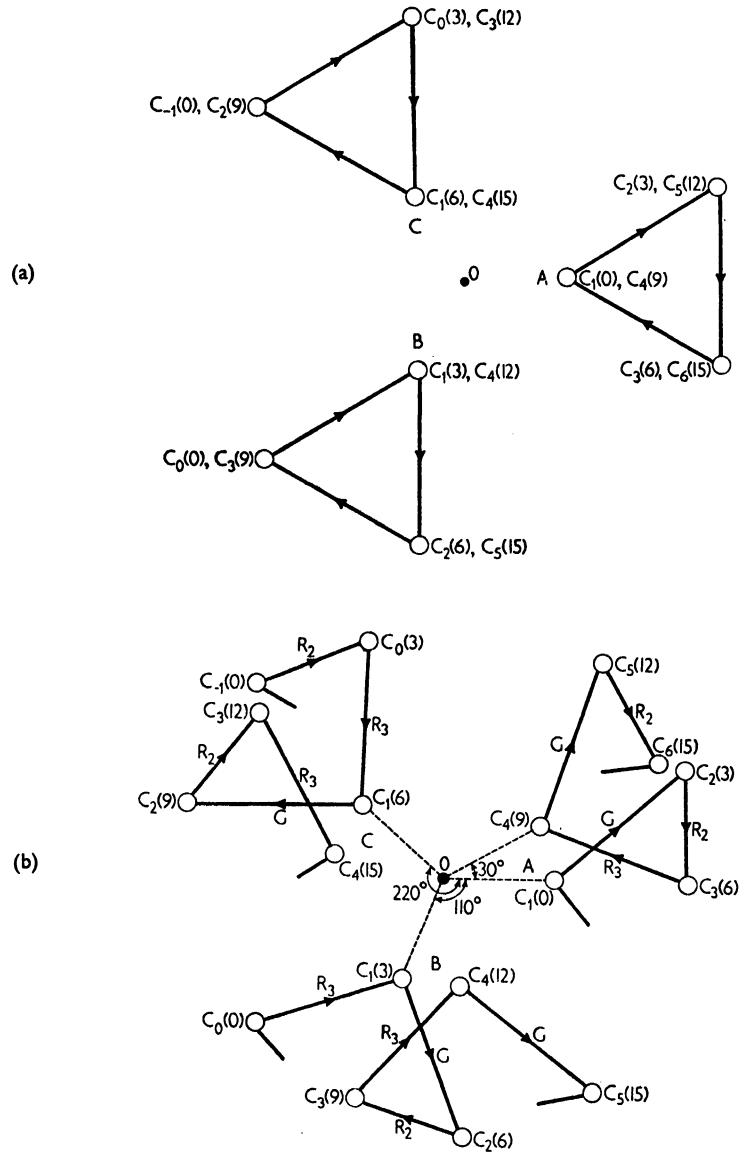


FIG. 1. The Madras triple helix, projected along the axis. (a) Uncoiled structure, with the three helices standing side by side, each having a three-fold screw axis. (b) Coiled-coil structure, in which each of the chains is twisted about the central axis through an angle of 30° for a height of three residues. The diagram is schematic. The circles represent the α-carbon atoms and the lines joining them the peptide groups. The numbers represent the heights of the various α-carbon atoms. The residue height is taken to be 3 Å, for convenience.

Soon after this structure was published, Rich and Crick (1955) suggested that there might be only one systematic hydrogen bond for every three residues instead of two, as in the above structure. This criticism was based on certain stereochemical criteria considered to be valid at that time. However, it was felt by the author that a structure with two sets of systematic hydrogen bonds should be preferable to one with only one set, and therefore a more detailed examination of the stereochemical criteria was made recently at Madras (Sasisekharan, 1962, see also Fuller, 1959). These recent studies indicated that contact distances, much less than those assumed by Rich and Crick, actually occurred in various organic structures and so the objections raised by them earlier are not valid. In view of these facts, the two-bonded structure, as originally put forward by Ramachandran and Kartha (1955*a*, *b*), was further refined in this laboratory and accurate co-ordinates of the atoms were worked out (Ramachandran and Sasisekharan, 1961*a*, *b*; Ramachandran *et al.*, 1962). While doing this, the helical parameters of the collagen structure were also refined (Lakshmanan *et al.*, 1962). It was found that the number of residues per turn was not exactly $3\frac{1}{3}$ ($= 10/3$) as was earlier supposed to be the case, but that it was about 3·28. In the same way, the resolved height per residue along the fibre axis was found to be 2·95 Å in unstretched collagen, about 0·1 Å more than the value accepted earlier.

In the following sections, we shall briefly summarize the main aspects of the latest structure of collagen and discuss its properties and applications.

2. Helical Parameters of the Collagen Structure

In view of the fact that the value of 10/3 determined by Ramachandran and Ambady (1954) for the number of residues per turn in the collagen structure has been accepted by all the later workers, it would be worth while mentioning the data that have led to a non-rational value of about 3·28 for this quantity. As mentioned earlier, the coiled-coil structure may be considered to be derived from the prototype structure by giving it a twist about the central axis through O. If we denote the value of this twist for three residues (expressed as a fraction of a full turn) by f, and if n is the number of residues per turn in the structure, then it is clear that the following relation should hold:

$$f = (n-3)/n \tag{1a}$$

or

$$n = 3/(1-f) \tag{1b}$$

Now, there seems to be no reason from stereochemistry, or otherwise, why f should be a rational fraction of a full turn, and consequently it is obvious that n also need not be rational. In fact, even a long time ago, some deviations from the rational value 10/3 were observed in the author's laboratory while trying to fit the ratios of the different layer spacings

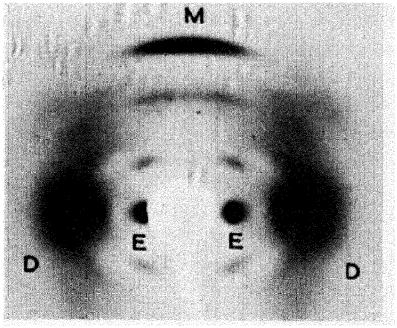


FIG. 2. Photograph of an inclined beam X-ray diffraction pattern of stretched collagen taken in a cylindrical camera. E = Equatorial spot; M = Meridional arc; D = diffuse blobs on the equator.

with a suitable value of n; but they were considered to be due to errors of observation. However, the accurate measurements of Lakshmanan *et al.* (1962) showed that the deviations were significant and they led to a systematically lower value of n, the average being 3·28. This means that the twist for three residues is, in degrees, $= (0{\cdot}28/3{\cdot}28) \times 360° = 30{\cdot}7°$, definitely less than $360°/10 = 36°$, obtained from the earlier value of $n = 10/3$. Since the probable error in f is certainly more than 1°, the value has been rounded off to 30° in all the later work. However, it must be remembered that the rational ratio $30°/360° = 1/12$ has no significance.

Similarly, the resolved height per residue (h) of the structure also had to be revised. This height is measured by finding the spacing of the meridional arc (approximately 3 Å) in the diffraction pattern of collagen. Now, unlike the diffraction pattern of a crystal, the pattern of a fibre like collagen which does not have a well defined unit cell will have scattering matter in reciprocal space not only along the intersection of the fibre axis with this layer, but also over a large area around this point. This is the origin of the extended meridional arc which is observed (M of Fig. 2). Consequently, if it is required to obtain the true meridional spacing, it is necessary to take a diffraction photograph with the fibre inclined at the equi-inclination angle for this layer and measure the spacing corresponding to the *lower edge* of the meridional arc M. This was exactly what was done by Lakshmanan *et al.* (1962), and they obtained values of $h = 2{\cdot}95$ Å for unstretched collagen and $h = 3{\cdot}05$ Å for stretched collagen.

It may be mentioned that when the pattern was recorded with a normal beam and the spacing corresponding to the middle of the arc so obtained was measured, then a value of 2·85 Å was obtained for the unstretched specimen, agreeing well with the earlier data. This shows that the larger value of 2·95 Å reported by these authors was not due to any differences between the specimens used by earlier workers and by them, but that it was due to the difference in interpretation. Since the latest approach is obviously the correct one from the theoretical point of view, it is clear that a valid structure for collagen must be based on a helical framework with $n = 3{\cdot}28$, and $h = 2{\cdot}95$ Å (for unstretched collagen), i.e. with a twist for three residues of 30°.

For a further elaboration of the ideas regarding non-integral helices and their interpretation and of the lattice structures occurring in fibres, see two recent papers by the author (Ramachandran, 1960, 1962).

3. The Standard Structure of Collagen

As is shown in Fig. 1 (b), each of the three chains of the helical framework of the collagen structure is twisted about a common axis through O, the amount of this twist being 30° (anti-clockwise) for every three

residues. In order that there may be a symmetrical relationship between the three chains, it is necessary that the other two chains must be obtained by the successive operations of a clockwise rotation of 110° and a translation of one residue height. (In Fig. 1, the residue height is taken to be 3 Å for convenience; it is strictly 2·95 Å in unstretched collagen.) Thus, starting from the α-carbon atom C_1 of chain *A* at level 0, successive applications of this operation lead to the atom C_1 of chain *B* at level 3 Å, the atom C_1 of chain *C* at level 6 Å, and finally to the atom C_4 of chain *A*

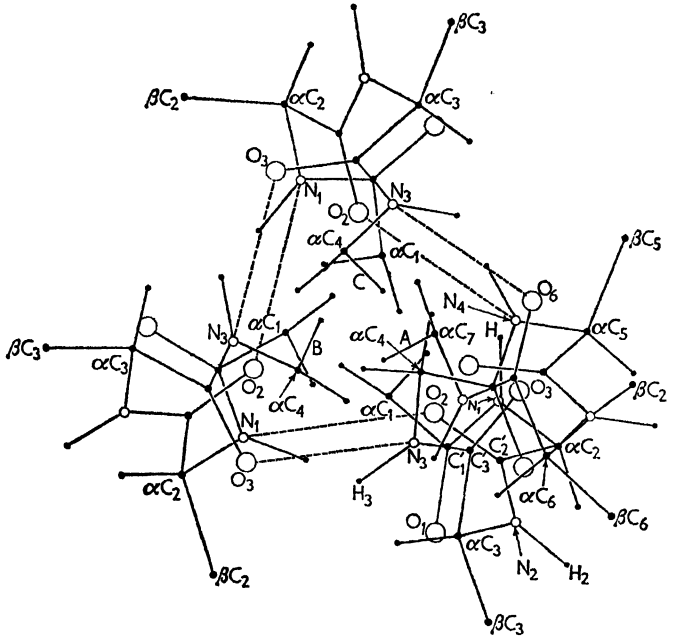


FIG. 3. Basal projection of the standard structure of collagen. Only the β-carbon atoms of the side-chains are shown. Note the occurrence of *two* hydrogen bonds for every three residues.

itself at level 9 Å. Thus, every third residue in each chain occupies an equivalent position, so far as the backbone is concerned. The three residues, marked G(R_1), R_2 and R_3, are however, not equivalent. In fact, in the actual structure of collagen (see below) it is impossible to have a β-carbon atom attached to any of the C_1-atoms, and hence the corresponding residues cannot have side-chains, and must be glycine residues.

Conforming to the above helical framework, a stereochemically satisfactory structure has been worked out (Ramachandran *et al.*, 1962), which is shown in Fig. 3 for a height of about 9 Å in projection. The atoms αC_1, αC_4, etc., occur on the inside of the triple helix and it will be

impossible to have a side group attached to them. The corresponding residues must therefore be necessarily glycine residues. All the other α-carbon atoms can have side-chains attached to them—in fact the β-carbon atoms are also marked in Fig. 3. The structure contains two hydrogen bonds for every three residues, of the types $N_1H_1(B) \ldots O_2(A)$ and $N_3H_3(B) \ldots O_3(C)$. The cylindrical co-ordinates of all the atoms in this "Standard Structure" for a height of 9 Å are given in Table I. The systematic arrangement of the hydrogen bonds is indicated in Table II.

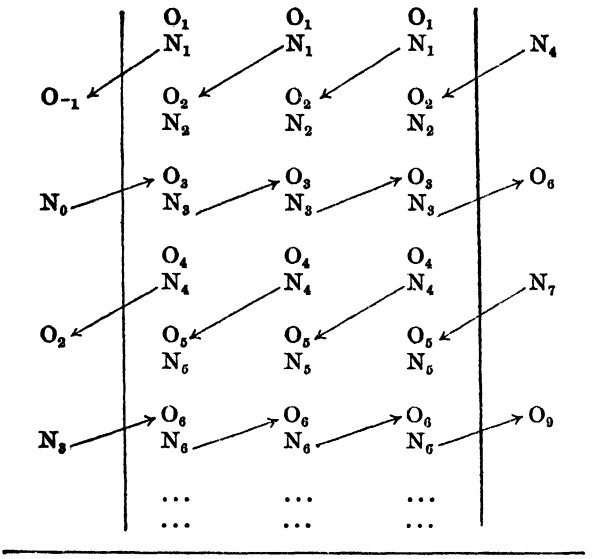
TABLE I. Cylindrical co-ordinates of the atoms in the backbone and the β-carbon atoms in the standard structure of the collagen protofibril for a height of three residues†

		Chain A		Chain B		Chain C	
Atom	r (Å)	ϕ (°)	z (Å)	ϕ (°)	z (Å)	ϕ (°)	z (Å)
αC_1	1·15	0·0	0·00	250·0	2·95	140·0	5·90
H_1'	1·10	45·7	−0·48	295·7	2·43	185·7	5·38
H_1''	0·63	311·8	0·50	201·8	3·45	91·8	6·40
C_1'	2·27	12·0	1·00	262·0	3·95	152·0	6·90
O_1	3·26	356·4	1·15	246·4	4·10	136·4	7·05
N_1	2·43	38·8	1·70	288·8	4·65	178·8	7·60
H_1	2·15	63·1	1·71	313·1	4·66	203·1	7·61
αC_2	3·52	35·4	2·72	285·4	5·67	175·4	8·62
H_2'	4·28	27·4	2·36	277·4	5·31	167·4	8·26
βC_2	4·34	54·0	3·07	304·0	6·02	194·0	8·97
C_2'	2·93	23·0	3·96	273·0	6·91	163·0	9·86
O_2	1·75	17·5	4·16	267·5	7·11	157·5	10·06
N_2	3·79	15·8	4·82	265·8	7·77	155·8	10·72
H_2	4·78	18·1	4·70	268·1	7·65	158·1	10·60
αC_3	3·47	2·0	6·00	252·0	8·95	142·0	11·90
H_3'	3·28	346·1	5·70	236·0	8·65	126·0	11·60
βC_3	4·78	0·1	6·79	250·1	9·74	140·1	12·69
C_3'	2·50	18·7	6·86	268·7	9·81	158·7	12·76
O_3	2·68	45·5	6·50	295·5	9·45	185·5	12·40
N_3	1·95	1·8	7·87	251·8	10·82	141·8	13·77
H_3	2·29	335·6	8·04	225·6	10·99	115·6	13·94

† The co-ordinates of the other atoms can be obtained from the relations (for any atom of index $n+3$ from that of index n in each of the chains A, B, C):

$$r_{n+3} = r_n,\ \phi_{n+3} = \phi_n + 30°,\ z_{n+3} = z_n + 8{\cdot}85 \text{ Å}.$$

TABLE II. Scheme of hydrogen bonds for a height of six residues



A perspective drawing of the structure is shown in Fig. 4. In this, only the α-carbon atoms are marked by symbols, and only the β-carbon atoms of the side-chains are shown wherever they occur, the remaining atoms of the side-chains being omitted. It will be seen from this that the structure is quite well packed and that the diameter of the protofibril is of the order of 12 Å.

The various bond distances, bond angles and interatomic contacts of this structure have been listed by Ramachandran *et al.* (1962), and these are all found to be within the permitted limits. The hydrogen bonds are both about 3 Å long, and the occurrence of such long hydrogen bonds is substantiated by the infra-red spectrum of collagen. The NH stretching frequency occurs at 3330 cm^{-1}, which is clearly larger than the usual value for other proteins. This value corresponds to a hydrogen-bond length of the order of 3 Å, according to the correlation worked out by Nakomoto *et al.* (1955). The observed infra-red dichroism of the NH and CO bands are also in agreement with what is predicted for the structure (Ramachandran *et al.*, 1962).

As mentioned earlier, side groups can be attached to any of the α-carbon atoms other than those at every third position (namely C_1, C_4, . . .). However, the pyrrolidine rings of proline and hydroxyproline

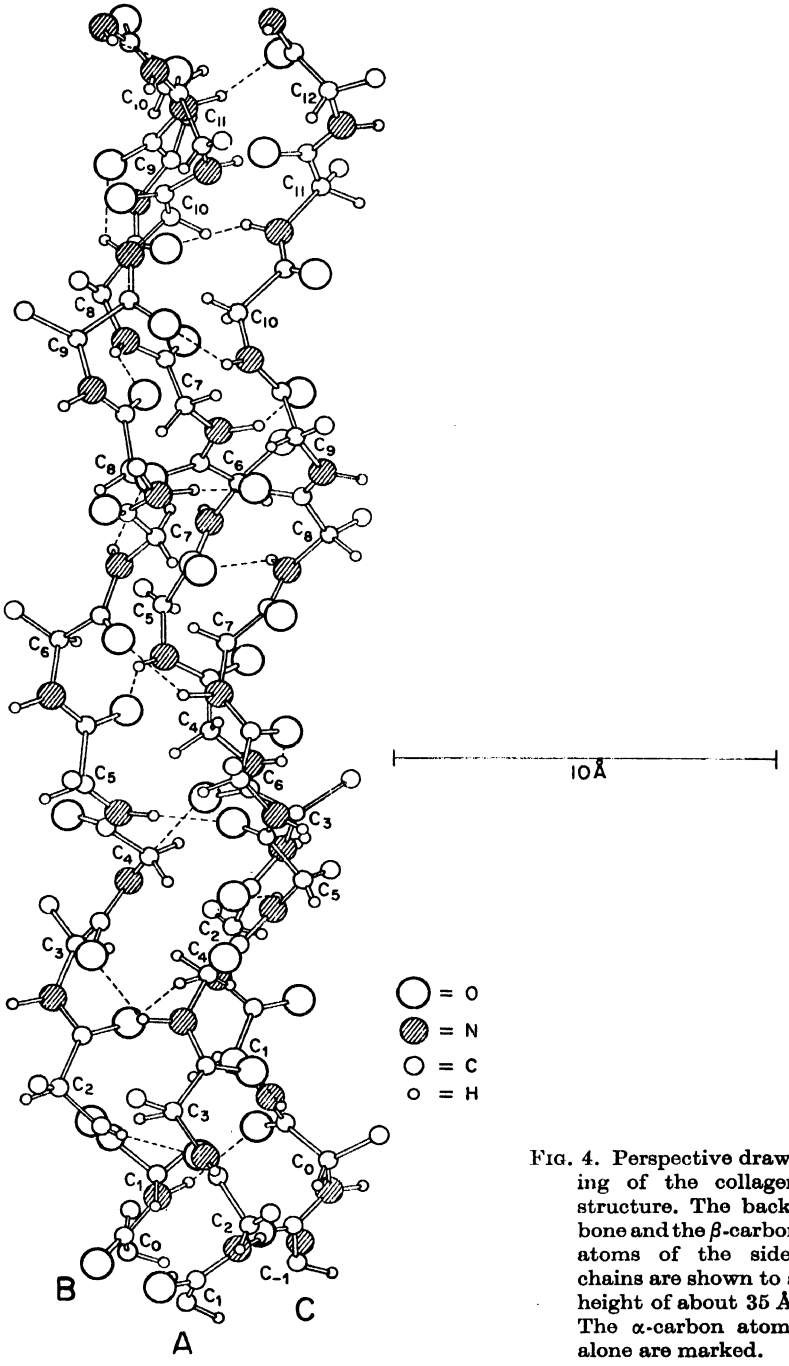


FIG. 4. Perspective drawing of the collagen structure. The backbone and the β-carbon atoms of the side-chains are shown to a height of about 35 Å. The α-carbon atoms alone are marked.

residues can occur without distortion of the chains only at the third position (counting the glycine residue as the first, i.e. at C_3, C_6, etc.). In this case, since the group N_2H_2 is not normally involved in a hydrogen bond stabilizing the triple-chain protofibril, no such hydrogen bond need be broken. When however a pyrrolidine ring occurs in the second position, then the hydrogen bond $N_1H_1 \ldots O_2$ cannot be formed. A very small distortion of the corresponding chain is needed, and five hydrogen

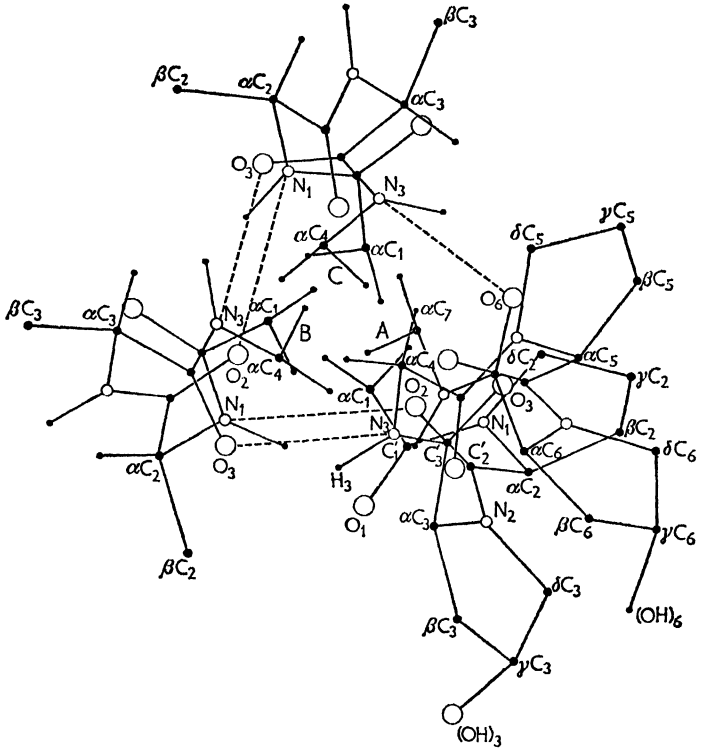


FIG. 5. Basal projection of the alternative structure having the sequence Gly–Pro–Hypro in only one chain.

bonds can then be formed instead of six, as in the standard structure.† A projection of such a structure with a sequence –Gly–Pro–Hypro– in one of the three chains is shown in Fig. 5. Even if all the three chains have the sequence –Gly–Pro–Hypro– locally, the standard structure need only be distorted by a small amount to accommodate these (Ramachandran *et al.*, 1962). Although the distortion introduced is quite small

† It may be mentioned that, in contrast to the collagen structure, proline cannot be accommodated in an α-helix without appreciable distortion and a complete change in direction of the helix will occur at sites where prolines exist.

(of the order of 0·2 Å), only three systematic hydrogen bonds can be formed for a height of three residues in such a case, because the imino-acid residues do not have NH groups for forming hydrogen bonds.

Thus, the standard structure (Figs. 3 and 4) should be considered to be the normal configuration for the collagen structure, which however undergoes a small distortion in local regions where the imino acid residues occur together. However, no distortion is introduced by the occurrence of any of the other residues.

In view of these latest developments, the two modifications of the Ramachandran-Kartha structure suggested earlier (namely the so-called Collagen I and II, Rich and Crick, 1955, 1958, 1961; Cowan *et al.*, 1955; Burge *et al.*, 1958) need not be discussed in detail. In both of these, there is only one hydrogen bond for every three residues in each chain, as against two in the latest structure, so that the latter is distinctly superior. In addition, the two-bonded structure has the advantage that the packing of the three helices is denser in it. Thus, the distance between the centres of the individual chains is 4·4 Å in the standard two-bonded structure, while it is of the order of 4·7–4·8 Å in the one-bonded structure, in spite of the fact that the hydrogen bonds in the latter are as short as 2·8 Å. It has been found very recently (Wallwork, 1962) that the average length of an NH...O hydrogen bond is 2·95 Å (average of 69 values). If this as well as the fact that the infra-red data indicate a length of about 3·0 Å are taken into account, then the packing distance in the one-bonded structure will be larger than 4·8 Å, which seems to be impossible from X-ray data.

Again, the occurrence of *two* systematic hydrogen bonds for every three residues gives a reason why the collagen structure has a coiled-coil structure and why the three chains are twisted around a common axis with a twist of 30° for three residues. If the twist is appreciably either more or less than this value, then only one systematic hydrogen bond can be formed. On the other hand, there is no such condition for a one-bonded structure. The occurrence of a twist of the right magnitude in the actual structure of collagen is therefore a strong case for the existence of two systematic hydrogen bonds for every three residues, as in the Madras structure.

The X-ray diffraction pattern also provides strong evidence for the occurrence of the well packed two-bonded structure. Apart from the general agreement between the calculated Fourier transform of the structure and the observed intensity distribution in the various layer lines, one particular feature requires special mention. This is the set of two diffuse blobs (D) observed on the equator of the diffraction pattern (Fig. 2). This is readily seen to arise from the occurrence of nearly parallel peptide planes at the same level in the structure. The photograph of a

model of the structure made of such planar units shows this particularly well—see, for example, A, B, C in Fig. 6. The interatomic vectors between similar atoms in two adjacent planes will all lie nearly on the equator and they will also be nearly equal in length, being centred around the value of 4·4 Å mentioned earlier. This will lead to a concentration of scattering power in reciprocal space corresponding to a spacing of around 4·4 Å on the equator and this is the origin of the diffuse blobs. The agreement of the observed spacing of the centre of this blob with that for the two-bonded structure, and its distinct disagreement with the larger value of 4·7 to 4·8 Å for the one-bonded structure again indicates that the latest structure is definitely superior. In fact, this distance is 4·77 Å in the recently published one-bonded structure of Rich and Crick (1961), which is definitely too large.

Thus, X-ray, infra-red and other evidence all favour the two-bonded structure in preference to the other possibilities that have been suggested earlier and we may therefore conclude that the bulk of the protofibril in collagen has this particular configuration. In those regions where proline and hydroxyproline residues occur together, there will be only one hydrogen bond for every three residues, but the chains themselves need only move about a little to accommodate them.

A more detailed discussion of the points mentioned in this section will be found in the papers by Ramachandran *et al.* (1962) and Lakshmanan *et al.* (1962).

4. Evidence for the Triple Helical Protofibril

We shall now consider the evidence for the occurrence of a triple-helical protofibril in collagen from sources other than X-ray diffraction. These are broadly of two types: (a) those which provide dimensions of the cross-section of the protofibril and its mass per unit length, and (b) those which provide an indication that there are three sub-units in the protofibrils.

Evidence of the first type has been obtained essentially from ultracentrifuge and light-scattering studies and from electron microscopy. In fact, it has come to light that one may even talk of a collagen macromolecule (the tropocollagen particle of Schmitt and co-workers, Schmitt *et al.*, 1953; Gross *et al.*, 1954), consisting of a triple helical protofibril of length about 2900 Å. The most systematic study using physical chemical studies are those of Boedtker and Doty (1955), who find that the collagen molecule in solution is a rigid rod of length 2900 Å and diameter 14 Å, with a molecular weight of about 300,000. The value of the cross-sectional diameter and of the mass per unit length, namely 100/Å, are both in excellent agreement with our triple helical structure. In fact, the dimen-

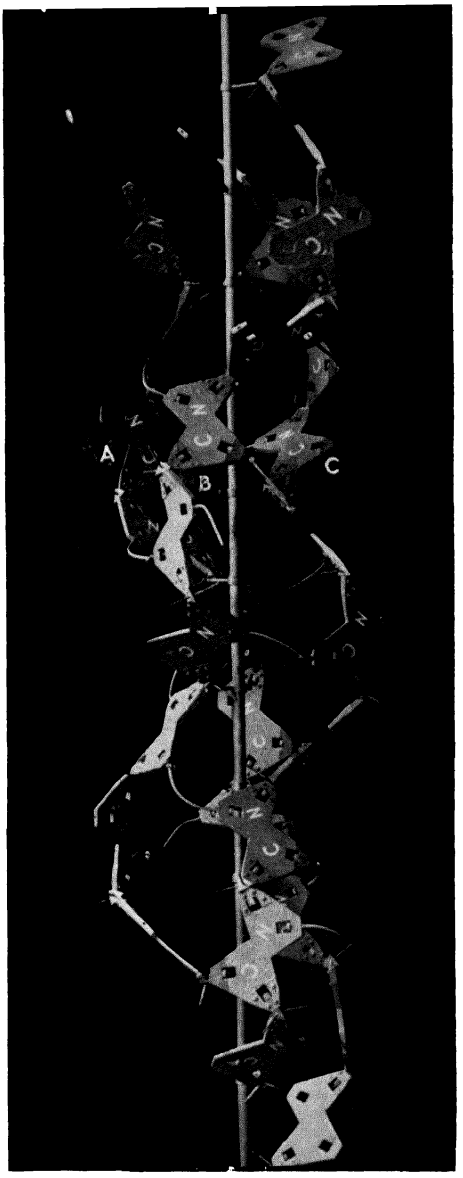


FIG. 6. Photograph of a model of the collagen chain, with planar peptide groups. Note the approximate parallelism of the three groups in different chains at the same level.

sions of the molecular particles have also been directly confirmed by electron microscopic observation (Hall and Doty, 1958).

During 1957–8, evidence began to accumulate that the tropocollagen molecule can be split into two sub-units on denaturation, one of which (α) had twice the molecular weight of the other (β) (Orekhovich and Shpikiter, 1957, 1958). This has been confirmed by various workers and, very recently, Piez *et al.* (1961) have observed two α-type components (α_1, α_2) and two β-type components (β_1, β_2) the latter having twice the molecular weight as the former. Variations in the contents of the minor amino acids were also observed between these, and it could be established that $\beta_1 = \alpha_1 + \alpha_2$ and $\beta_2 = 2\alpha_1$. It is generally believed that the normal undenatured molecule is $(\alpha + \beta)$, a small quantity of which (called γ) has also been observed in the denatured solution (Grassmann *et al.*, 1961). Thus, there is clear evidence that the collagen protofibril consists of three component parts.

Another evidence that the protofibril may have three peptide chains in it is available from the end-group determinations of the peptides derived from collagen. Thus, Grassmann *et al.* (1960) found that all the peptides obtained by tryptic digestion had glycine as the N-terminal residue, but that they had either arginine, or lysine, or both, as C-terminal residues. When both occurred, there were either 1Lys + 2Arg or 2Lys + 1Arg. In a number of cases, the molecular weight obtained by end-group determination was one-third the molecular weight obtained from the amino-acid analysis. All these go in support of the triple-chain nature of the protofibril.

5. Occurrence of the Triple Helix in Other Proteins and Polypeptides

After the prototype triple helix was put forward from Madras in 1954, a number of polypeptides have been found to have this structure. These are polymers of amino acids which are characteristic of collagen like glycine, proline and hydroxyproline, but unlike in collagen the polypeptide chains in these have only a simple three-fold screw symmetry in their structure and do not form coiled coils. The structures of the first two were worked out at the same time as Ramachandran and Kartha (1955*a*, *b*) modified their prototype structure into a coiled-coil structure. Thus, Cowan and McGavin (1955) showed that poly-L-proline had a hexagonal unit cell, with individual chains following a left-handed screw. Crick and Rich (1955) found that the X-ray pattern of one of the modifications of polyglycine, namely polyglycine II, could be fitted by a chain configuration having a three-fold screw symmetry. In 1959, Sasisekharan (1959) showed that poly-L-hydroxyproline also had a

structure with a left-handed three-fold screw symmetry. In fact, three helices occurred together as a triple helix in a unit cell of this polypeptide. Thus, just like the α-helix which is found to occur in the structure of a variety of polypeptides, the Madras helix also is one of the standard configurations possible for polypeptides.

It is interesting to note that the polypeptide $(Gly–Pro–Hypro)_n$ has been found recently to yield an X-ray powder pattern very similar to that

TABLE III. Properties of the α-helix and the triple helix

α-Helix	Triple helix
X-Ray pattern	
Equator: 9·5 Å Meridian: 5·1 Å (arc) and 1·5 Å	Equator: 10–12 Å, 5–6 Å Meridian: 3 Å Layer lines: At 10 Å and 4 Å
Infra-red dichroism	
Both NH stretching (3300 cm^{-1}) and CO stretching (1650 cm^{-1}) have parallel dichroism	Both bands have perpendicular dichroism
Optical rotation	
Specific rotation of the order of 0° to +50°	Specific rotation of the order of −350°
Birefringence	
Of the order of +0·01 or more	Very small, positive, ~ +0·005
Mechanical properties	
Can be extended by about 100%	Rather inextensible, can be extended at most by about 10%

of collagen, with a sharp ring at about 3 Å and a diffuse ring of spacing about 4·5 Å (Andreeva, private communication). There is therefore little doubt that it also has a triple helical structure like collagen, but whether it is a coiled-coil structure or not can only be decided when a good oriented fibre photograph is obtained.

Coming to proteins, there seems to be strong reasons to suppose that elastin has a triple helical structure. Firstly, it has one-third of its residues as glycine and its proline content is also about the same as for collagen. Secondly, the intensity distribution in its X-ray pattern is very close to that of shrunk collagen. Although the latter is not a conclusive argument, it may be mentioned that a faint trace of the 3 Å arc was

obtained in the author's laboratory by Ambady (unpublished) by stretching an autoclaved elastin fibre. All these, along with the fact that elastin and collagen often occur associated together in the tissues, strongly suggest that elastin may also have a collagen-like structure (Ramachandran and Santhanam, 1957).

Thus, recent studies on collagen have led to the postulation of a new type of chain configuration in proteins and polypeptides in general. In order to be of assistance in identifying this chain configuration in future studies, some of its main properties are given along with those of the α-helix in Table III.

References

Boedtker, H. and Doty, P. (1955). *J. Amer. chem. Soc.* **77**, 248.

Burge, R. E., Cowan, P. M. and McGavin, S. (1958). In *Recent Advances in Gelatin and Glue Research* (G. Stainsby, ed.), p. 25. Pergamon Press, London.

Cohen, C. and Bear, R. S. (1953). *J. Amer. chem. Soc.* **75**, 2783.

Cowan, P. M. and McGavin, S. (1955). *Nature, Lond.* **176**, 501.

Cowan, P. M., North, A. C. T. and Randall, J. T. (1953). In *Nature and Structure of Collagen*, p. 241 (J. T. Randall, ed.). Butterworth, London.

Cowan, P. M., McGavin, S. and North, A. C. T. (1955). *Nature, Lond.* **176**, 1062.

Crick, F. H. C. and Rich, A. (1955). *Nature, Lond.* **176**, 780.

Fuller, W. (1959). *J. phys. Chem.* **63**, 1705.

Grassmann, W., Hannig, K. and Schleyer, M. (1960). *Hoppe-Seyl. Z.* **322**, 71.

Grassmann, W., Hannig, K. and Engel, J. (1961). *Hoppe-Seyl. Z.* **324**, 284.

Gross, J., Highberger, J. H. and Schmitt, F. O. (1954). *Proc. nat. Acad. Sci., Wash.* **40**, 679.

Hall, C. E. and Doty, P. (1958). *J. Amer. chem. Soc.* **80**, 1269.

Lakshmanan, B. R., Ramakrishnan, C., Sasisekharan, V. and Thathachari, Y. T. (1962). In *Collagen* (N. Ramanathan, ed.), p. 117. Wiley, New York.

Nakamoto, K., Marghoshes, M. and Rundle, R. E. (1955). *J. Amer. chem. Soc.* **77**, 6480.

Orekhovich, V. N. and Shpikiter, V. O. (1957). *C. R. Acad. Sci., U.R.S.S.* **115**, 137.

Orekhovich, V. N. and Shpikiter, V. O. (1958). In *Recent Advances in Gelatin and Glue Research* (G. Stainsby, ed.), p. 87. Pergamon Press, London.

Piez, K. A., Lewis, M. S., Martin, G. R. and Gross, J. (1961). *Biochim. biophys. Acta* **53**, 596.

Ramachandran, G. N. (1960). *Proc. Indian Acad. Sci.* **A52**, 240.

Ramachandran, G. N. (1962). In *Collagen* (N. Ramanathan, ed.), p. 3. Wiley, New York.

Ramachandran, G. N. and Ambady, G. K. (1954). *Current Sci. (India)* **23**, 349.

Ramachandran, G. N. and Kartha, G. (1954). *Nature, Lond.* **174**, 269.

Ramachandran, G. N. and Kartha, G. (1955*a*). *Nature, Lond.* **176**, 593.

Ramachandran, G. N. and Kartha, G. (1955*b*). *Proc. Indian Acad. Sci.* **A42**, 215.

Ramachandran, G. N. and Santhanam, M. S. (1957). *Proc. Indian Acad. Sci.* **A45**, 124.

Ramachandran, G. N. and Sasisekharan, V. (1961*a*). *Nature, Lond.* **190**, 1004.

Ramachandran, G. N. and Sasisekharan, V. (1961*b*). *Current Sci. (India)* **30**, 127.

Ramachandran, G. N., Sasisekharan, V. and Thathachari, Y. T. (1962). In *Collagen* (N. Ramanathan, ed.), p. 81. Wiley, New York.
Rich, A. and Crick, F. H. C. (1955). *Nature, Lond.* **176**, 915.
Rich, A. and Crick, F. H. C. (1958). In *Recent Advances in Gelatin and Glue Research* (G. Stainsby, ed.), p. 20. Pergamon Press, London.
Rich, A. and Crick, F. H. C. (1961). *J. mol. Biol.* **3**, 483.
Schmitt, F. O., Gross, J. and Highberger, J. H. (1953). *Proc. nat. Acad. Sci., Wash.* **39**, 459.
Sasisekharan, V. (1959). *Acta cryst.* **12**, 903.
Sasisekharan, V. (1962). In *Collagen* (N. Ramanathan, ed.), p. 39. Wiley, New York.
Wallwork, S. C. (1962). *Acta cryst.* **15**, 758.

Discussion

D. HARKER: Chemical evidence indicates the proline plus hydroxyproline is about 25% of collagen. Would not this allow more hydrogen bonding than your remarks indicate?

G. N. RAMACHANDRAN: I believe that there is local enrichment of proline and hydroxyproline and that they do not occur throughout the chain. Grassmann's work supports this view since he has found that there are regions rich in proline and hydroxyproline and others rich in polar amino acids.

A. ELLIOTT (King's College, London): Is the structure of collagen impossible if the sense of one chain is reversed?

G. N. RAMACHANDRAN: Yes, we find that we do not get a satisfactory structure.

A. ELLIOTT: I do not think that one can infer from the frequency of the NH stretching band that the hydrogen bond is weaker than in other polypeptides. Other factors could affect the frequency, for it seems that the bond at *ca.* 3300 cm^{-1}. (3330 cm^{-1}. in collagen) is not the fundamental unperturbed hydrogen-bonded NH stretching mode.

G. N. RAMACHANDRAN: That is true. But I might still mention that, quite apart from infra-red data, there is evidence that the hydrogen bond lengths from the nitrogen of an NH group to a carbonyl oxygen atom should be of the order of 2·95 Å (S. C. Wallwork (1962). *Acta cryst.* **15**, 758). This agrees well with our structure and not so in the case of the one-bonded structure.

D. HARKER: Kendrew has observed in the structure of myoglobin that for the α-helix the bond distance is of the order of 2·85 Å.

M. CHVAPIL: Under normal conditions collagen (in the native state) is in an extended state. What are the forces which stabilize this extension so far as the thermal shrinkage phenomenon is concerned? How do you explain according to your model the fact that the force preventing thermal contraction is steadily increasing with age.

G. N. RAMACHANDRAN: The shrinkage can be very easily understood with our model. When collagen is heated beyond a particular temperature, the stabilizing hydrogen bonds are broken and the whole framework crumples down. As regards the question of ageing, we have done experiments of a reverse nature. Collagen fibre treated with neutral salt solutions (by which process the fibre shrinks) shows certain variations in the X-ray diffraction pattern which are in reverse order to that obtained during ageing (G. N. Ramachandran (1958). *Recent Advances in Gelatin and Glue Research* (G. Stainsby, ed.), p. 32. Pergamon Press, London; M. S. Santhanam (1959). *Proc. Indian Acad. Sci.* **A49**, 251).

F. HAPPEY: What mechanism does Prof. Ramachandran anticipate will explain

the lattice extension of 7–10% in the fibre direction in collagen; in particular what compression forces are likely to be involved in the cross-section of the molecular chain assembly in the unit cell.

G. N. RAMACHANDRAN: Although the residue height (h) in collagen is only about 3 Å, which is definitely less than in the fully extended chain, only a limited extension is possible, because the three chains are coupled together. If the rotation angle (see Ramachandran *et al.*, this volume, p. 121) ϕ is close to 120° then h does not vary by more than 10% for a wide range of ϕ' (see G. N. Ramachandran, V. Sasisekharan and Y. T. Thathachari (1962). In *Collagen* (N. Ramanathan, ed.), p. 81. Wiley, New York, regarding this).

M. H. F. WILKINS: Purely from energy considerations, it is clear that an addition of a hydrogen bond will suffice for bringing a number of atoms 0·3 Å closer than the equilibrium (van der Waals) distances. But the point is that we do not know just where this equilibrium comes.

G. N. RAMACHANDRAN: Can we not get it empirically by an analysis of the known crystal structures? Actually, we did this and found that in hydrogen-bonded structures, the distances can be 0·3–0·4 Å less than what is normally found otherwise.

M. H. F. WILKINS: I agree with you generally, but it would be better if we have more experimental data on this.

Protein Crystallography in Cambridge and in London

D. C. PHILLIPS

Medical Research Council External Staff,
The Royal Institution, London, England

ABSTRACT

The Medical Research Council Laboratory for Molecular Biology and the Davy-Faraday Research Laboratory, The Royal Institution, London, have been closely associated in the X-ray analysis of protein structures for the past eight years. Current work includes refinement of the structures of haemoglobin and myoglobin at higher resolution and in the oxygenated and reduced states; a low-resolution study of hen egg-white lysozyme; preliminary studies of α-chymotrypsin and of β-lactoglobulin; the development of automatic apparatus for making single-crystal X-ray measurements; and the investigation of new methods of protein structure analysis. Progress in all of these fields will be reviewed.

INTRODUCTION

Since Sir Lawrence Bragg, who was so unfortunately prevented from presiding over this symposium, moved from Cambridge to the Royal Institution in 1954, the work of the Davy-Faraday Research laboratory has been closely linked with that of the Medical Research Council laboratory for Molecular Biology in Cambridge. Besides work in Cambridge, Drs. M. F. Perutz and J. C. Kendrew have been Readers in the Davy-Faraday laboratory and, of course, Sir Lawrence has inspired us all. It is a great honour for me to have this opportunity of telling you something about the recent work in the two laboratories and, following Sir Lawrence Bragg's Presidential address, which reviewed all but the latest results, I shall restrict myself for the most part to unpublished developments.

In an organization such as ours there is, of course, a continual interchange of ideas and nearly everyone contributes something to every project. The names given are those of the workers associated most closely with the particular researches, without regard to whether they work in Cambridge or in London.

APPARATUS DEVELOPMENT

(U. W. Arndt, F. B. Jones, E. L. McGandy and D. C. Phillips)

Once the initial difficulties of finding heavy atom derivatives or some other way of solving the structures have been overcome the progress of

protein crystallography depends increasingly on the availability of apparatus for making large numbers of intensity measurements and of high-speed electronic digital computers for doing the calculations. A great deal of work has been done therefore on the development of automatic single crystal diffractometers. The laboratory version of the linear diffractometer (Arndt and Phillips, 1961) has been in almost constant use for nearly three years, and two additional instruments of this type are now in operation. In addition, two automatic three-circle diffractometers are nearing completion. These instruments set themselves automatically to angles fed in on punched tape and, like the linear diffractometer, they provide punched-tape output of the intensity measurements suitable for direct input to a computer. The Cambridge instrument (Arndt and McGandy, 1963) uses Moiré fringe gratings on all three circles, while the London instrument (Arndt and Jones, 1963) uses stepping motors associated with a control system that can operate two diffractometers at once. The usefulness of these instruments, and of the photographic methods with high intensity X-ray tubes which were developed earlier, will be apparent in the accounts of protein structure analyses that follow.

Myoglobin

(J. C. Kendrew; C. C. F. Blake, C. Branden, C. L. Coulter, A. B. Edmundson, D. C. Phillips, Helen Scouloudi, Violet C. Shore, L. Stryer and H. C. Watson)

The structure of sperm-whale myoglobin at 2 Å resolution has already been described in some detail (Kendrew *et al.*, 1960, 1961) and, during the past two years, we have been concerned with improving the resolution of the electron density map by including virtually all the observable reflections which extend to spacings of about 1·4 Å. The new measurements were made by means of the linear diffractometer rather than photographically and, because of the change in technique, it was necessary to re-investigate the effects on the crystals of prolonged irradiation.

Radiation damage

Photographically one records the intensity of each reflection in a reciprocal lattice level averaged over the exposure time of the photograph. The diffractometer, on the other hand, measures the intensities in sequence, starting at high angles and scanning the densely populated rows in a zig-zag motion to the centre of the diffraction pattern. If each reflection is measured once, about 50 reflections are measured in one hour and a complete level is measured with about 24 hours' exposure of the crystal to the X-rays. Clearly the state of the crystal at the end of this period may differ significantly from that obtaining at the beginning.

This danger was investigated in an experiment in which the $h0l$ reflection intensities were measured seven times over from the same crystal during a total exposure of nearly 300 hours, corresponding to a dose of about 50 Mrads (Blake and Phillips, 1962). The results showed that during exposure to X-rays an increasing part of the crystal is made amorphous and a further fraction is somewhat disordered, while the remainder is more or less unchanged. Each quantum of Cu $K\alpha$ radiation that is absorbed appears to damage severely about 150 molecules.

When the individual reflection intensities are corrected for these general effects, some progressive variations appear which may be interpretable in terms of specific, radiation-induced changes to the structure. But the successive sets of measurements can be made to agree quite well. Thus the agreement indices, for intensities, between the first and subsequent sets of measurements were:

run	2	3	4	5	6	7
	5·7	7·1	7·8	10·2	11·2	17·3%

This result suggested that irradiation damage is not very serious over the first 30 hours' exposure and it was decided to limit the irradiation of each crystal to this amount or less.

The first set of measurements was also compared with those measurements made photographically, to check the consistency of measurements made by the two methods. For this purpose the reflections were divided into six equal groups with increasing $\sin\theta$ values, and the two sets of measurements were scaled together by an overall scaling factor. The variation in agreement index and scaling factor in the different $\sin\theta$ ranges is shown in Table I.

TABLE I. Comparison of diffractometer and photographic data: $h0l$ reflections

Group	$\Sigma F_p/\Sigma F_d$	R(%)
1	1·00	5·1
2	0·98	5·2
3	1·00	5·5
4	0·99	6·8
5	0·99	7·2
6	0·97	10·3
all	—	6·4

The increase in R with $\sin\theta$ is determined mainly by the decrease in average intensity. The agreement of the two techniques, in fact, is encouragingly good and it was confirmed by measurements of other crystals.

The reflections from sperm-whale myoglobin crystals were measured out to about 1·4 Å spacing where the diffraction pattern fades out. Only reflections in group 6 of the 2 Å data were remeasured for scaling and a new crystal was used for each of twenty-four levels up the c-axis and for eight levels up the b-axis. In all, some 50,000 measurements were made in less than three months. The data were processed by means of the University of London "Mercury" computer using programmes developed by Dr. A. C. T. North, and the levels were scaled together using reflections in common rows (Rollet and Sparks, 1960). All the levels were then merged together giving some 10,500 significant measurements of independent reflections, 3500 of which overlapped with the 2 Å data and could be used in scaling to them. The myoglobin data now comprise some 17,000 reflections with significant intensities which are being used in a final refinement of the structure.

1·4 Å Refinement of the structure

Instead of continuing to use the method of isomorphous replacement, which involves the collection of data from a number of different isomorphous derivatives, we have reverted to the conventional method of successive refinement. Dr. Kendrew described the method and our first results in his Nobel lecture (Kendrew, 1963).

"From a study of the 2 Å Fourier synthesis, spatial co-ordinates were assigned to about three-quarters of the atoms in the molecule. Owing to the limited resolution of this synthesis, the accuracy with which atoms could be located was a good deal less than is desirable, but this imprecision was compensated for by their number, a good deal higher in proportion to the size of the structure than is generally necessary for the success of the refinement method. This method consists in calculating the phases of all the reflections from the co-ordinates of the atoms that have already been located. A Fourier synthesis is then computed using *observed* amplitudes and *calculated* phases. This synthesis necessarily shows all the atoms that have been used for calculating phases, but it also reveals the positions of additional atoms, by peaks of reduced density, and indicates minor errors in the positions of the atoms already located. The corrected atomic co-ordinates and the additional atoms are used in the next cycle of refinement.

"We have so far carried out two cycles of refinement, including 825 atoms in the first and 925 atoms in the second (myoglobin contains in all 1260 atoms excluding hydrogen; in addition there are some 400 atoms of

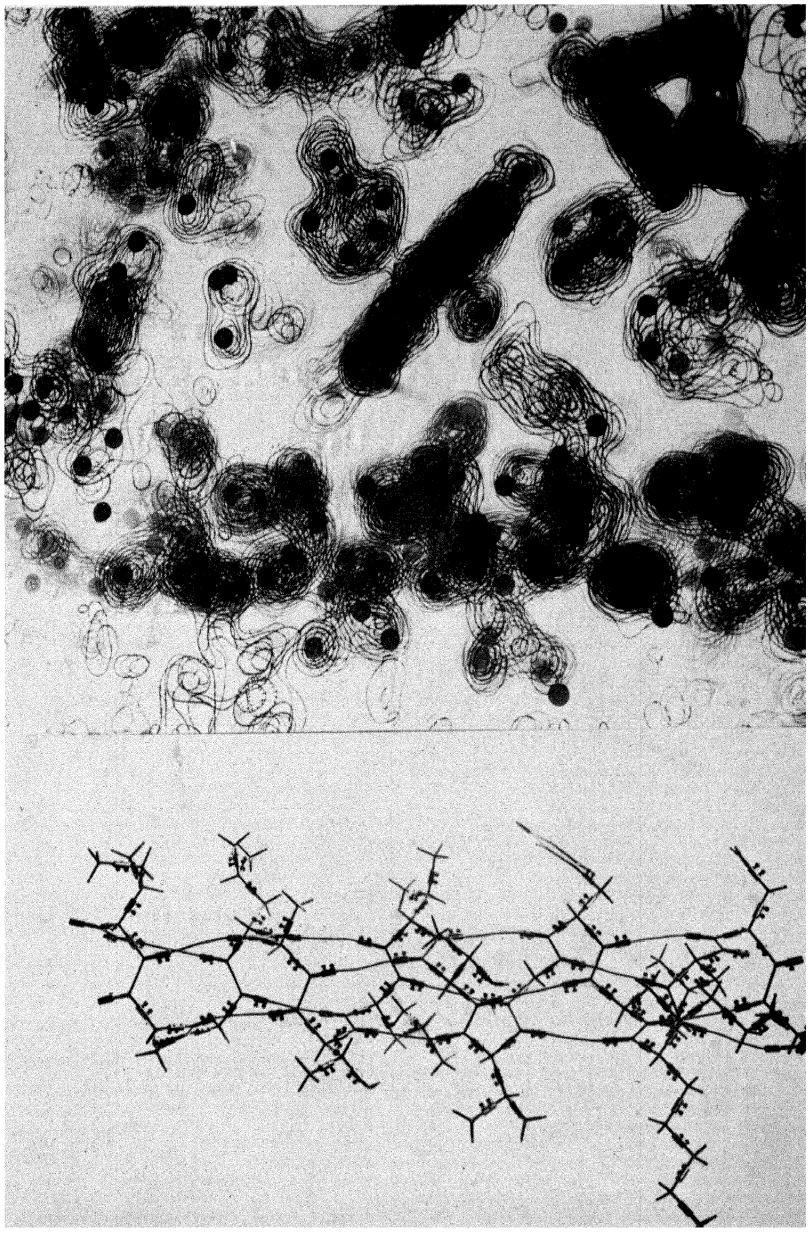


FIG. 1. (a) Part of the 1·4 Å Fourier synthesis of sperm-whale myoglobin. *Centre:* the haem group (edge-on), showing haem-linked and distal histidines, and water molecule attached to the iron atom. *Top right:* a helix end-on. *Bottom:* a helix seen longitudinally, together with several side chains. (b) Model of α-helix with side-chains corresponding to those in (a).

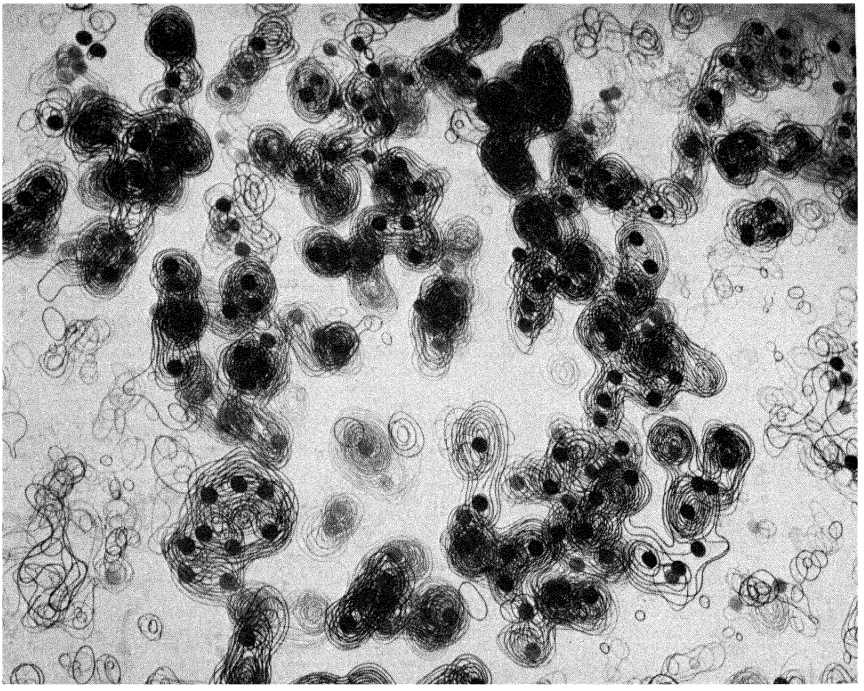


FIG. 2. Part of the 1·4 Å Fourier synthesis of sperm-whale myloglobin. *Left centre:* a tryptophan residue; *to the left:* a liquid region between two molecules.

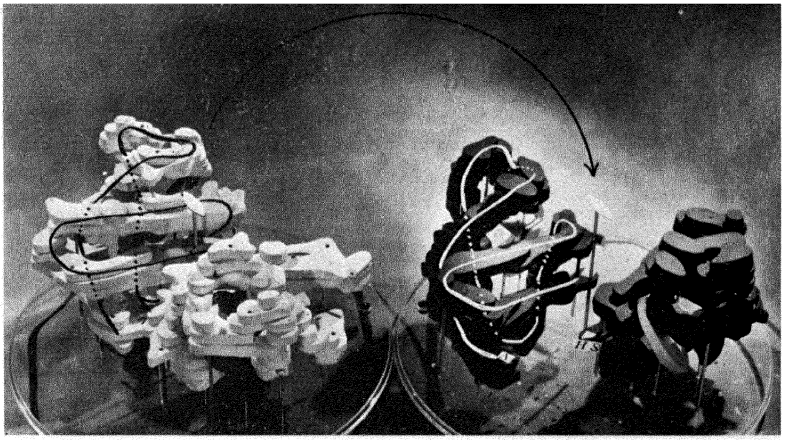


FIG. 3. Two pairs of chains, in the horse oxyhaemoglobin molecule, symmetrically related by the dyad axis. The arrow shows how one pair is placed over the other to assemble the complete molecule.

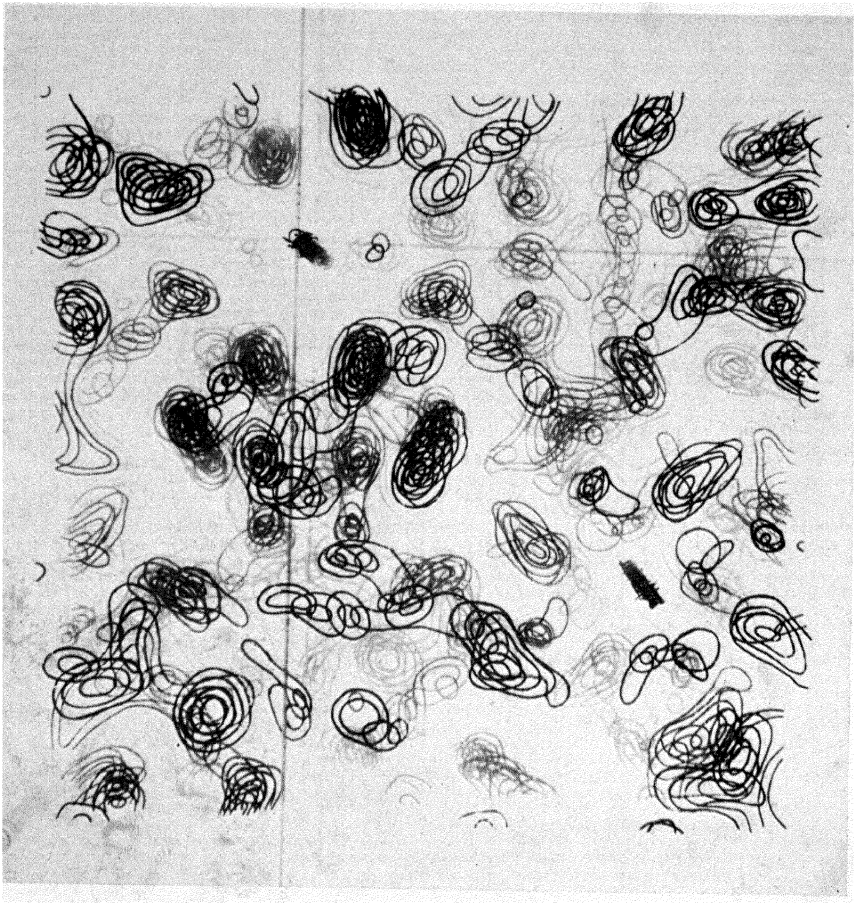


FIG. 5. Electron density distribution in tetragonal lysozyme at 6 Å resolution, view parallel to the *c*-axis. The horizontal and vertical lines represent the two-fold rotation axes and intersect the two-fold screw axis, upper left of centre. The four-fold screw axis is lower right of centre. Contour interval about 0·07 electrons/Å^3 with the lowest near 0·6 electrons/Å^3.

liquid and salt solution, a proportion of which are bound to fixed sites on the surface of the molecule). One or two further cycles of refinement will probably be necessary, but in the meantime the 1·4 Å Fourier synthesis based on the second cycle is much better resolved than the 2 Å synthesis. In many cases neighbouring covalently bonded atoms are just resolved, the background between groups of atoms is much cleaner than before and, finally, many of the disturbances found in the region of the heavy-atom sites in the 2 Å synthesis have disappeared. Figures 1 and 2 give some impression of this synthesis.

Chemical studies

Meanwhile Dr. Edmundson has greatly advanced his study of the amino acid sequence of myoglobin; in particular, he has characterized a large number of chymotryptic peptides in addition to the tryptic ones previously described (Edmundson and Hirs, 1961) and he is now analysing the three components liberated by the action of cyanogen bromide at the two methionine residues. Taking the results of the X-ray and chemical studies together, some 120 amino acid residues are now known with almost complete certainty and many of the remaining 30 with fair probability. There is little doubt that the residual ambiguities will shortly be resolved and that the positions of all the atoms in the structure will be known with reasonable accuracy, with the exception of a few long side chains (such as lysine) which appear not to occupy defined positions in the crystal structure.

General nature of the structure

About 118 of the 151 amino acid residues in the molecule make up eight segments of right-handed α-helix, of lengths ranging from seven to twenty-four residues. These segments are joined by two sharp corners and five non-helical segments (of 1–8 residues). There is also a non-helical tail of five residues at the carbonyl end of the chain. The whole structure is extremely compact with next to no water inside the molecule and a very small volume of internal empty space. The haem group is almost normal to the surface of the molecule with one edge, that containing the polar propionic acid groups, at the surface and the rest buried deeply within.

Nearly all the side chains containing polar groups are on the surface while the interior of the molecule is almost entirely made up of non-polar residues, generally close-packed and in Van der Waals' contact with their neighbours. The Van der Waals' forces between these non-polar residues in the interior of the molecule are clearly most important for maintaining the integrity of the whole structure.

The interactions of the haem group require special consideration

since they must be responsible for the characteristic function of myoglobin. The fifth co-ordination position of the iron atom is occupied by a ring nitrogen atom of a histidine residue. On the other (distal) side of the iron atom, occupying its sixth co-ordination position, is a water molecule, as would be expected in ferrimyoglobin. Beyond the water molecule, in a position suitable for hydrogen-bond formation, is a second histidine residue. For the rest the environment of the haem group is almost entirely non-polar.

Future work

It is hoped that further study of this structure and the structures of oxy- and reduced myoglobin will reveal the nature of the oxygenation reaction in precise structural terms. In this connection a study of azide myoglobin, which exhibits some structural differences from met-myoglobin has already been begun and there are indications that a reorientation of the haem group is involved.

Haemoglobin

(M. F. Perutz; L. Goaman, E. L. McGandy, Hilary Muirhead and J. Prothero)

The structure of horse haemoglobin at 5·5 Å resolution has already been described (Cullis *et al.*, 1961, 1962).

In agreement with the chemical evidence, the electron density maps show four haem groups and four separate chains which are identical in pairs;† these chains are very similar in structure and each bears a strong resemblance to that in sperm-whale myoglobin. The haemoglobin molecule is assembled by first matching each chain with its symmetrically related partner, then inverting one pair (white) and placing it over the other pair (black) as shown in Fig. 3 to form a compact spheroidal molecule with the haem groups arranged in separate pockets on the surface. The problem now is to increase the resolution of this image and, by this and other studies, to determine the nature of the oxygenation reaction.

2 Å Resolution analysis of horse oxyhaemoglobin

This project is still in the data collection phase. So far 2 Å data of the native protein have been collected using the linear diffractometer, and heavy-atom containing crystals are now being measured.

Human reduced haemoglobin

The exciting new work on this structure (Muirhead and Perutz, 1963) was described by Dr. Perutz in his Nobel lecture (Perutz, 1963).

"The oxygen-free form of haemoglobin, somewhat inappropriately called reduced haemoglobin, has long been known to differ from oxy-

† See Fig. 5, "X-Ray Analysis of Biological Molecules", p. 1.

haemoglobin in its solubility, crystal structure and other properties, which suggested that the explanation should be sought in a structural re-arrangement between the two forms. Unfortunately, reduced haemoglobin of horse crystallizes in a form unsuitable for detailed X-ray analysis, so that human haemoglobin, which is more amenable, is being studied instead. (In view of the close similarity between the amino acid sequences of the two species it seems unlikely that the structure of human oxyhaemoglobin, which has not been determined, differs from

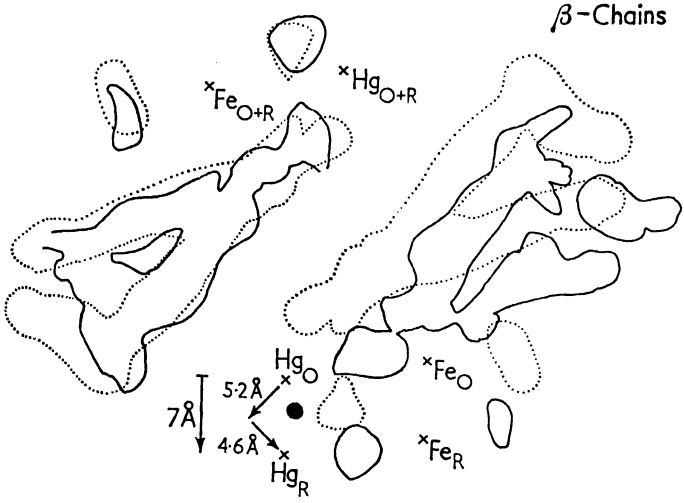


FIG. 4. Broken line: outline of the helical regions G and H together with the positions of the iron and mercury atoms (attached to cysteines) and sections through helices E and F in horse oxyhaemoglobin (O). Full line: the same for human reduced haemoglobin (R). One member of each pair has been superimposed so as to bring the iron and mercury positions into coincidence.

that of horse oxyhaemoglobin in the same manner as the human reduced form—but the point remains to be investigated.)

"So far the structure of human reduced haemoglobin is based on the X-ray analysis of only two isomorphous heavy-atom derivatives, as compared to the six used for horse oxyhaemoglobin and, furthermore, these two particular derivatives are insufficient to decide the ambiguity in the majority of phase angles. Hence, the electron density maps have been calculated by a mathematical approximation and the results are not yet as well defined as those for horse oxyhaemoglobin.

"Despite these imperfections, several features stand out clearly. The molecule is made up of four sub-units which appear to be very similar in structure to those found in horse oxyhaemoglobin, but there is a striking

re-arrangement of the two black sub-units, involving an increase in the distance between symmetrically related features by over 7 Å (Fig. 4). The relative arrangement of the white subunits is affected to a much lesser extent, if at all."

It seems likely that the oxygenation reaction in haemoglobin will eventually find its explanation in terms of the structural changes of which these new results are the first indications. But it may be necessary to solve the structure of at least one form in atomic detail—and that will take some time.

Hen Egg-white Lysozyme

(C. C. F. Blake, Ruth H. Fenn, D. F. Koenig, A. C. T. North, D. C. Phillips and R. J. Poljak)

A 6 Å resolution Fourier map of the electron density in tetragonal crystals of hen-egg lysozyme has been described recently (Blake *et al.*, 1962) and we are now working to improve the accuracy of the analysis and to increase its resolution. The present map, which is shown in Fig. 5, was obtained in the standard way using three heavy atom derivatives. These crystals, which contained mercuri-iodide, chloropalladite ($PdCl_4$) and *o*-mercuryhydroxytoluene-*p*-sulphonic acid (MHTS), were prepared by diffusing the heavy-atom groups into previously grown crystals. To a first approximation the $PdCl_4$ and MHTS groups occupy single sites in general positions, but the mercuri-iodide is in a special position on a two-fold rotation axis and at this resolution it is found to be best represented by two isotropic "atoms" about 6 Å apart. Use of the mercuri-iodide derivative was further complicated by the fact that the heavy-atom content diminished during exposure to X-rays so that the site occupancy had to be made a function of exposure time in the phase calculations.

The intensities again were measured by means of the linear diffractometer and all the calculations were done on the University of London "Mercury" computer. The phases were determined systematically by the phase-probability method and anomalous dispersion effects were taken into account, both in the phase determination and in order to distinguish between the enantiomorphous space groups. In the first calculations the overall mean "figure-of-merit" was 0·86.

The heavy-atom parameters have now been refined using various difference Fourier maps from which it appeared that $PdCl_4$ and MHTS were doubly substituted, each derivative having a second site with weight about 20% of the main site. Furthermore some refinement of the mercuri-iodide parameters was possible; the difference maps showed that the two "sites" on the two-fold axis are not equally occupied.

These changes have led to an increase in the overall mean figure-of-merit to 0·89 but they have not materially altered the Fourier map.

We are now collecting data for an analysis of the structure at 2 Å resolution and at the same time we are comparing our results with those obtained by Stanford, Marsh and Corey (1962) and by Dickerson *et al.* (1962).

β-Lactoglobulin

(D. W. Green; P. D. Baker, J. Copola, R. H. Simmons; and R. Aschaffenburg, National Institute for Research in Dairying, University of Reading)

Sub-units

In solution at neutral pH the molecular weight of cow β-lactoglobulin is 36,000. In several crystal forms this is also the asymmetric unit, but there are three in which the asymmetric unit is a molecular subunit of weight 18,000. The first to be found was a Cd derivative:

Lattice S: $a = 36{\cdot}4$; $b = 127{\cdot}6$; $c = 36{\cdot}4$ Å; $\beta = 98°\ 12'$; $P2_1$; 2×36K.

Although, strictly, the asymmetric unit is 36K, the (010) projection shows strong pseudo-symmetry ($B22_12$), with an asymmetric unit of 18K. More recently, both *A* and *B* genetic variants of lactoglobulin have been crystallized in these two forms:

Lattice Y: $a = 55{\cdot}7$; $b = 67{\cdot}2$; $c = 81{\cdot}7$Å; $B22_12$; 8×18K.
Lattice Z: $a = 54{\cdot}3$; $c = 117{\cdot}0$Å; $P3_12$; 6×18K.

α-Helices

The radial distribution of intensities from a salt-free crystal form:

Lattice W: $a = 36{\cdot}4$; $b = 68{\cdot}2$; $c = 72{\cdot}4$ Å; $\beta = 92°\ 12'$; $P2_1$; 2×36K.

has a well-marked 10 Å peak, suggesting the presence of helical folding. A quantitative comparison with haemoglobin has been made, using $(\mathrm{grad}\rho)^2\ \mathrm{d}V$ summed for intensities multiplied by a smoothing function peaked at 10 Å (Crick, 1953).

Apparently the same proportion of the chain is helically folded as in haemoglobin, but it is possible that the absolute scale of intensities, on which the calculation depends, may be in error by up to 30%.

Isomorphous replacement

Simple mercurials have been stoichiometrically linked to the sulphydryl groups (1 per 18K) in lattice Y and lattice Z. Refinement of the Hg parameters is proceeding satisfactorily. HgI_4 also combines at one major site per 18K, and possibly other minor sites.

Buffalo lactoglobulin

Crystals of buffalo lactoglobulin supplied by Dr. A. Sen of the Bose Institute, Calcutta, are closely related to a form of cow lactoglobulin crystallized in similar conditions:

Buffalo: $a = 35{\cdot}9$; $b = 127{\cdot}7$; $c = 35{\cdot}9$ Å; $\beta = 106°\ 17'$; $P2_1$.
Cow lattice R: $a = 36{\cdot}1$; $b = 127{\cdot}5$; $c = 36{\cdot}0$ Å; $\beta = 106°\ 05'$; $P2_1$.

The molecular shapes must be similar, and a more detailed structural homology is suggested by similar reflection intensities.

α-Chymotrypsin

(D. M. Blow, B. Jeffery and M. G. Rossmann)

α-Chymotrypsin crystals are monoclinic, $a = 49$, $b = 67$, $c = 66$ Å; $\beta = 102°$; Space group $P2_1$ with four molecules per unit cell, i.e. two molecules per asymmetric unit.

Heavy-atom derivatives have been prepared containing $PtCl_4$, PtI_4 and $PtBr_4$, all of which occupy the same two sites per molecule, and a preliminary Fourier synthesis has been calculated using phases determined chiefly by the PtI_4 derivative which shows the greatest anomalous dispersion effects. This map is now being studied and refined.

New methods of phase determination

It is intended to improve the phase determination for this structure by making use of the fact that there are two molecules per asymmetric unit and using a new method of phase determination (Rossman and Blow, 1962, 1963). This method has not yet been proved completely, but encouraging results have been obtained in a study of insulin. It would be of enormous value since many large protein molecules appear to be built up from subunits.

Acknowledgements

I am indebted to all my colleagues for their help in preparing this review and for permission to quote unpublished results, and I would like to thank Dr. M. F. Perutz and Dr. J. C. Kendrew in particular for allowing me to read and quote from their Nobel lectures.

References

Arndt, U. W. and Jones, F. B. (1963). To be published.
Arndt, U. W. and McGandy, E. L. (1963). To be published.
Arndt, U. W. and Phillips, D. C. (1961). *Acta Cryst.* **14**, 807.
Blake, C. C. F. and Phillips, D. C. (1962). *Biological Effects of Ionising Radiation at the Molecular Level.* International Atomic Energy Agency, Vienna.

Blake, C. C. F., Fenn, R. H., North, A. C. T., Phillips, D. C. and Poljak, R. J. (1962). *Nature, Lond.* **196**, 1173.
Crick, F. H. C. (1953). *Acta cryst.* **6**, 600.
Cullis, A. F., Muirhead, H., North, A. C. T., Perutz, M. F. and Rossmann, M. G. (1961). *Proc. roy. Soc.* **A265**, 14.
Cullis, A. F., Muirhead, H., North, A. C. T., Perutz, M. F. and Rossmann, M. G. (1962). *Proc. roy. Soc.* **A265**, 161.
Dickerson, R. E., Reddy, J. M., Pinkerton, M. and Steinhauf, L. K. (1962). *Nature, Lond.* **196**, 1178.
Edmundson, A. B. and Hirs, C. H. W. (1961). *Nature, Lond.* **190**, 663.
Kendrew, J. C. (1963). *Nobel Lectures, Yearbook 1962*. Elsevier, Amsterdam.
Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. and Shore, V. C. (1960). *Nature, Lond.* **185**, 422.
Kendrew, J. C., Watson, H. C., Strandberg, B. E., Dickerson, R. E., Phillips, D. C. and Shore, V. C. (1961). *Nature, Lond.* **190**, 666.
Muirhead, H. and Perutz, M. F. (1963). To be published.
Perutz, M. F. (1963). *Nobel Lectures, Yearbook 1962*. Elsevier, Amsterdam.
Rollett, J. S. and Sparks, R. A. (1960). *Acta cryst.* **13**, 273.
Rossmann, M. G. and Blow, D. M. (1962). *Acta cryst.* **15**, 24.
Rossmann, M. G. and Blow, D. M. (1963). *Acta cryst.* **16**, 39.
Stanford, R. H., Marsh, R. E. and Corey, R. B. (1962). *Nature, Lond.* **196**, 1176.

Discussion

G. J. S. RAO (Indian Institute of Science, Bangalore): Does the change in orientation of the haem group after oxygenation support the "crevice theory" of Pauling?

D. C. PHILLIPS: There seems to be a definite change in orientation after oxygenation, but it is too early to discuss it in detail. The haem group can certainly be described as lying in a crevice.

G. J. S. RAO: Benesch has shown that the reactivity of SH groups of haemoglobin towards iodoacetamide increases on oxygenation suggesting a change in conformation in solutions also.

D. C. PHILLIPS: I understand this is so, but it is really too early to say anything about the reaction mechanism in atomic detail from the structure.

G. KARTHA: I understand that Dr. Dickerson and others are working on the structure of lysozyme. Have you been able to compare your results with theirs?

D. C. PHILLIPS: As I said, there are two other groups besides ours working on lysozyme. Dickerson and Steinrauf in Illinois are working on the triclinic form and they have calculated a 6 Å Fourier. However, the determination of their heavy atom position is not yet well established enough for them to publish the results. We have not yet compared our results with theirs. On the other hand, the map produced in California by Stanford, Marsh and Corey has interesting correspondence with ours.

X-Ray Analysis of *p*-Bromocarbobenzoxy-glycyl-L-prolyl-L-leucyl-glycine and its Related Peptides

M. KAKUDO, Y. SASADA, Y. KATSUBE,
S. SAKAKIBARA AND S. AKABORI

Institute for Protein Research, Osaka University, Osaka, Japan

ABSTRACT

The X-ray investigation on the crystals of *p*-bromocarbobenzoxy-glycyl-L-prolyl-L-leucyl-glycine and its related oligopeptides have been carried out to elucidate the relationships between their molecular configurations and the substrate specificity to collagenase. A number of bromo substituted carbobenzoxy peptides were also synthesized for the sake of the X-ray work. The crystallographic data of a series of these peptides were examined, and the molecular configurations are discussed on the basis of the data. In particular, a detailed structure analysis has been carried out on the tetrapeptides. Although the structure refinement is not in a final stage yet, the approximate feature of the molecule has been derived here. The whole figure of the molecule is approximately described as a folded chain like the 2_n-form of Bragg. It is an interesting fact that this folded form of the peptide chain proposed as a protein model has appeared in a lower peptide like this substance. The folding of the main chain of the peptide is produced by forming a couple of intramolecular hydrogen bonds, and seems to be an intermediate form between the helical structure and 2_n-form.

The cell dimensions of *p*- and *o*-bromocarbobenzoxy-glycyl-L-prolyl-L-leucyl-glycyl-L-proline are also discussed. One of the modifications, δ-form seems to be somewhat different from the other modifications and the bromo derivatives in its molecular configuration.

1. INTRODUCTION

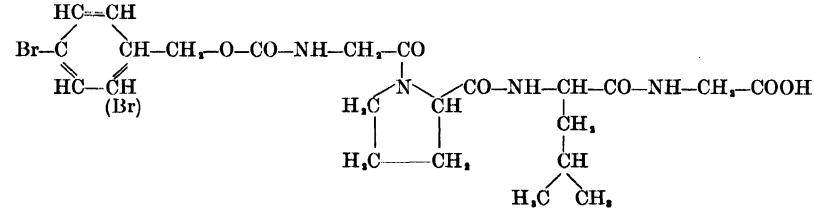
It is well-known that carbobenzoxy-glycyl-L-prolyl-L-leucyl-glycyl-L-proline was first synthesized as a substrate to collagenase in a crystalline form which shows a high degree of substrate specificity (Nagai and Noda, 1959; Nagai *et al.*, 1960). It has been recognized that the specificity of this enzymatic reaction primarily depends upon the sequence of amino acid residues such as –Pro-Leu-Gly-Pro–. The present interest is whether the reaction is also influenced by a spatial configuration of the peptide chain. If both the sequence of residues and a particular configuration of the peptide are essential conditions for the substrate specificity to collagenase, the structure of a characteristic part of the collagen molecule

itself may be scrutinized with reference to the structure of this peptide. Recently, some of the lower peptides, obtained by step by step removal of residues from its *C*-terminal end, were synthesized systematically to compare approximate figures of molecular configurations (Sakakibara and Nagai, 1960), and the respective bromine-substituted carbobenzoxy peptides were also synthesized.

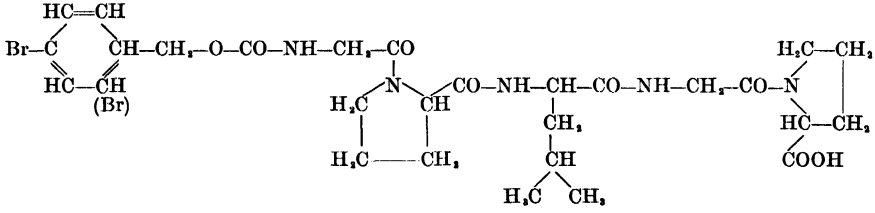
It has been reported that the configuration of the pentapeptide having the substrate specificity seems to be somewhat different from the other lower peptides when their crystallographic data are compared (Sasada *et al.*, 1961; Sasada and Kakudo, 1961).

The simplest way to facilitate X-ray analysis of the structure is to introduce a bromine atom as a heavy atom into the substances, because the heavy atom gives, more or less, some clue to the phase determination of the reflections.

The present accounts deal with the synthesis of *p*- and *o*-bromo-substituted tetrapeptides,



p- and *o*-bromopentapeptides,



and the determination of the unit-cell dimensions and space group of these crystals. A short report at an intermediate stage of the structure analysis of the carbobenzoxytetrapeptide is also included.

2. Synthesis of Peptides

Several bromo-substituted peptide derivatives were synthesized according to the procedure described in the earlier reports (Nagai and Noda, 1959; Nagai *et al.*, 1960; Sakakibara and Nagai, 1960) except that

p- or *o*-bromocarbobenzoxy chloride was used in place of carbobenzoxy chloride. Methods using *p*-nitrophenyl esters were also tested. The synthetic routes are as shown below:

o-Br-Z-Cl (I) + H-Gly-OH → *o*-Br-Z-Gly-OH (II)

II + NpOH $\xrightarrow{\text{DCCD}}$ *o*-Br-Z-Gly-ONp (III)

III + H-Pro-Leu-Gly-OEt → *o*-Br-Z-Gly-Pro-Leu-Gly-OEt (IV)

IV → *o*-Br-Z-Gly-Pro-Gly-OH (V)

IV → *o*-Br-Z-Gly-Pro-Leu-Gly-$NHNH_2$ (VI)

o-Br-Z-Gly-Pro-Leu-GlyN_3(← VI) + H-Pro-OEt →

o-Br-Z-Gly-Pro-Leu-Gly-Pro-OEt (VII)

VII → *o*-Br-Z-Gly-Pro-Leu-Gly-Pro-OH (VIII)

p-Br-Z-Cl (IX) + H-Gly-OEt → *p*-Br-Z-Gly-OEt (X)

X → *p*-Br-Z-Gly-$NHNH_2$ (XI)

p-Br-Z-Gly-N_3 (← XI) + H-Pro-Leu-Gly-OEt →

p-Br-Z-Gly-Pro-Leu-Gly-OEt (XII)

XII → *p*-Br-Z-Gly-Pro-Leu-Gly-OH (XIII)

XIII → *p*-Br-Z-Gly-Pro-Leu-Gly-$NHNH_2$ (XIV)

p-Br-Z-Gly-Pro-Leu-GlyN_3 (← XIV) + H-Pro-OEt →

p-Br-Z-Gly-Pro-Leu-Gly-Pro-OEt (XV)

XV → *p*-Br-Z-Gly-Pro-Leu-Gly-Pro-OH (XVI)

o-Br-Z- = *o*-bromocarbobenzoxy

p-Br-Z- = *p*-bromocarbobenzoxy

Np = *p*-nitrophenyl

It was confirmed by means of micro analysis that the final products of these syntheses were in exact agreement with their respective chemical formulae.

3. X-Ray Measurement and Crystallographic Data

These substances were recrystallized from ethyl acetate solution. Crystal habits are as follows:

p-Bromotetrapeptide: needle-shaped crystal elongated along the *a*-axis.

o-Bromotetrapeptide: needle-shaped crystal elongated along the *a*-axis.

p-Bromopentapeptide: cube-like crystal.

o-Bromopentapeptide: cube-like crystal.

Crystallographic data of these substances were obtained from oscillation, Weissenberg and precession photographs about each of the three principal axes. Accurate measurement by the counting technique was made using the single crystal orienter of G.E. XRD-6. Space groups were derived from systematic absence of spectra. Densities were measured by the usual flotation method. The results are listed in Table I.

TABLE 1. Crystallographic data of bromocarbobenzoxy peptides

(a) *Carbobenzoxy-glycyl-*L*-prolyl-*L*-leucyl-glycine* ($C_{23}H_{32}O_7N_4$, mol. wt. 476·5)

	Unsubstituted	*p*-Bromo	*o*-Bromo
a	6·58	6·25	6·20 Å
b	13·62	14·29	14·24 Å
c	27·58	29·68	28·15 Å
V	2472	2651	2485 Å^3
D (measured)	1·30	1·40	1·45 g/cm^3
D (calculated)	1·28	1·39	1·48 g/cm^3
Z	4	4	4
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$

(b) *Carbobenzoxy-glycyl-*L*-prolyl-*L*-leucyl-glycyl-*L*-proline* ($C_{28}H_{39}O_8N_5$, mol. wt. 573·7)

	Unsubstituted			
	α-Modification	δ-Modification	*p*-Bromo	*o*-Bromo
a	13·54	13·89	15·36	13·83 Å
b	14·73	7·50	15·02	14·13 Å
c	10·28	16·03	21·21	11·03 Å
β	105·6	100·4	120·0	115·3°
V	1975	1643	4238	1944 Å^3
D (measured)	1·22	1·23	1·27	1·33 g/cm^3
Z	2	2	4	2
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Apparent mol. wt.	727	608	807	780

The cell dimension of *p*-bromo substituted tetrapeptide was observed to be expanding by about 0·67 and 2·10 Å along the *b*- and *c*-axes respectively. Thus, although the crystal is not strictly isomorphous with the original peptide, it may be reasonable to suppose that the molecular configuration of these two peptides are very similar to each other. If so, this change of the unit-cell dimensions shows that the direction of the Br—C bond is nearly parallel to the *c*-axis. On the other hand, the *o*-bromine atom produces smaller changes of the *b*- and *c*-axes by about 0·63 and 0·57 Å, respectively. This may suggest that the hydrogen atom at the ortho position in the tetrapeptide does not directly contact with its

neighbouring molecules except for the $-CH_2-$ group of its own molecule. In other words, there is a crevass between molecules around this region. The increase of observed density of the *o*-bromo compound from the unsubstituted one is 1·5 times the increase for the *p*-bromo compound. This is fairly compatible with the above mentioned argument.

As regards pentapeptides, at least four modifications of the crystal were found (Sasada and Kakudo, 1961). In particular, the δ-form which was found recently is rather related to the lower peptides on the basis of unit-cell dimensions, and is different from the other three forms. The unit cells of the bromo compounds seem to correspond to the α-form of unsubstituted peptides. These relations between the crystals in the pentapeptide group, however, are difficult to discuss further from the crystallographic data only.

4. Structure Analysis of *p*-Bromocarbobenzoxy-Gly-L-Pro-L-Leu-Gly

Oscillation and Weissenberg photographs about the *a*-axis were taken using nickel-filtered Cu $K\alpha$ radiation. The unit-cell dimensions and space group are shown in Table I. For the intensity measurements, a set of multiple-film equi-inclination Weissenberg photographs about the *a*-axis were prepared from very thin crystals using four sheets of film. Observed reflections at approximately 100 hr exposure were up to about 1·4 $Å^{-1}$ in ξ for $0kl$ and $1kl$, and for higher layer reflections it was necessary to expose for more than 150 hr. Some Weissenberg photographs were taken about the *b*-axis of needle crystals. Intensity measurements were made by the visual method and partially by the counting method. Intensities were corrected for Lorentz and polarization effects, and the temperature factor and scale factor were obtained by Wilson's method; $B = 10{\cdot}0$ $Å^2$.

Four molecules of *p*-bromocarbobenzoxy-Gly-Pro-Leu-glycine exist in a unit cell, and the atomic parameters for 35 atoms excluding hydrogen atoms are to be determined. As shown in Table I, the *a*-axes of these crystals are all short. It was therefore decided to start from two-dimensional structure analysis in the *a*-projection. The crystal analysis described here was carried out mainly on the *p*-bromotetrapeptide.

At the first stage of this analysis, the Patterson functions projected on the *a*-plane were calculated for every crystal. As regards the bromo-substituted peptide, the contribution of the heavy bromine atom to the Patterson function is expected to be appreciable, as estimated from $\sum f_{Br}^2 / \sum f_p^2 = 0{\cdot}83$, where p runs over the atoms in the peptide. However, in this case, the Patterson maps of the bromo compounds did not distinguish the positions of the bromine atoms. The sharpened Patterson

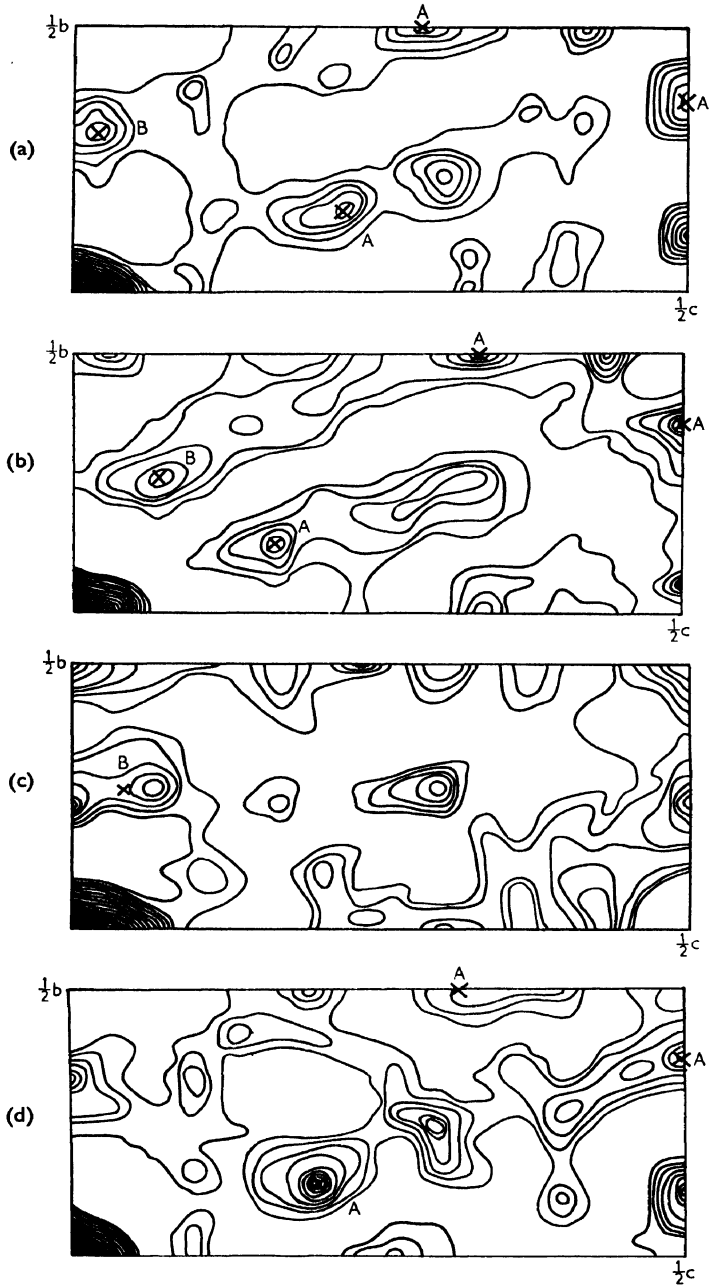


FIG. 1. Patterson maps projected on (100). (a) *p*-Bromocarbobenzoxytetrapeptide; (b) *o*-bromocarbobenzoxytetrapeptide; (c) carbobenzoxytetrapeptide; (d) sharpened one of *p*-bromocarbobenzoxytetrapeptide. The peaks marked by "A" and "B" in each figure correspond to the Br–Br vectors and benzene–benzene vector, respectively.

function and modified Patterson function were also tried, and as an example the Patterson function sharpened with a factor

$$\exp(-8{\cdot}0\sin^2\theta/\lambda^2)$$

is shown in Fig. 1 (d). In these maps, the Br–Br vectors showed up as comparatively distinct peaks, and in particular, the sharpened Patterson maps calculated with stepwise varying sharpening factors indicated more distinctly the Br–Br vectors as stepwise increasing peaks in height. However, to confirm these Br–Br vectors, the isomorphous replacement method was applied with the unsubstituted tetrapeptide and the *p*-bromotetrapeptide. If the relationship of the isomorphous replacement method exists between the peptide and its bromo peptides, although it is supposed to be only very approximately applicable, the following relation will be expected.

$$F_{\mathrm{Br}} = F_{p+\mathrm{Br}} - F_p, \qquad F_{\mathrm{Br}} = \pm|F_{p+\mathrm{Br}}| \mp |F_p|$$

where, the F_{Br}, F_p, $F_{p+\mathrm{Br}}$ are the structure factors of the bromine atom, peptide, bromo-substituted peptide respectively. Then, if $|F_{\mathrm{Br}}|$ is small enough compared with $|F_p|$, $(\Sigma f_{\mathrm{Br}}/\Sigma f_p = 0{\cdot}14)$, the following formulae are obtained. (This is, however, somewhat incompatible with the principle of the heavy-atom method.)

$$F_{\mathrm{Br}} = \pm(|F_{p+\mathrm{Br}}| - |F_p|), \qquad F_{\mathrm{Br}}^2 = (|F_{p+\mathrm{Br}}| - |F_p|)^2 = |\Delta F|^2$$

Therefore, the $|\Delta F|^2$ Patterson function should strongly emphasize the Br–Br vectors in the map. In fact, the $|\Delta F|^2$ Patterson map shows the peaks correctly corresponding to the Br–Br vectors, as shown in Fig. 2.

FIG. 2. Patterson map using $|\Delta F|^2$.

The initial co-ordinates of the bromine atom in *p*-bromotetrapeptide were thus determined.

The next stage is the Fourier analysis by the heavy-atom method based on the position of the bromine atom and by the isomorphous

replacement method. At first the signs of structure factors were determined by the heavy-atom method using the bromine atom in the peptide, and one of the Fourier maps calculated with these signs is given in Fig. 3. The Fourier map using the signs based on the isomorphous replacement method is also given in Fig. 4.

These Fourier maps did not give good resolution for the atomic positions of the molecule. It was, however, easy to locate the positions of

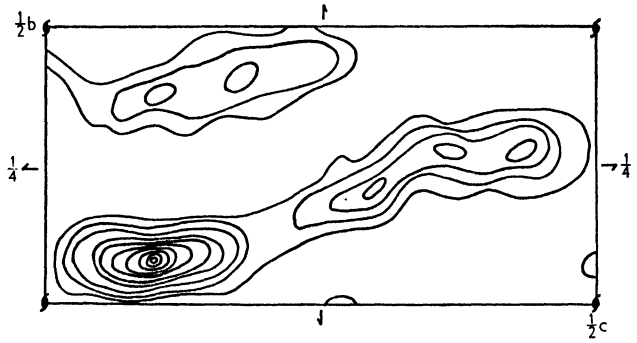


FIG. 3. Electron-density map by heavy-atom method. Contours are in arbitrary scale but equal intervals.

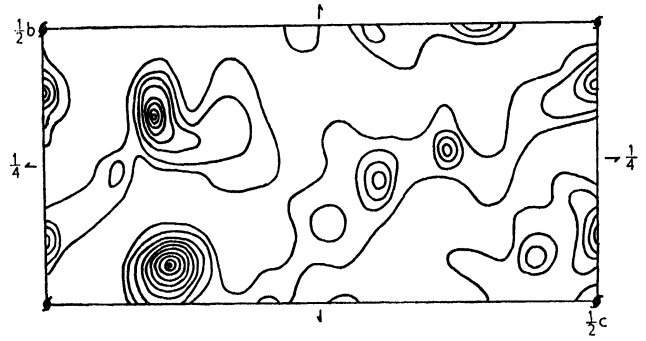


FIG. 4. Electron-density map by isomorphous replacement method. Contours are in arbitrary scale but equal intervals.

atoms given by a plausible model of the molecule on the density peaks. Structure factor refinements by the trial and error method and successive Fourier analysis were repeated until the R-factor was around 50%. Then, the least squares refinements using about 80 $F(0kl)$'s were tried and ten cycles were carried out for the most plausible model of the molecule. The minimum R-value at this refinement stage was about 23%. The final Fourier map is shown in Fig. 5, and the whole feature of the molecule is also drawn in the same illustration. The x-parameter of each

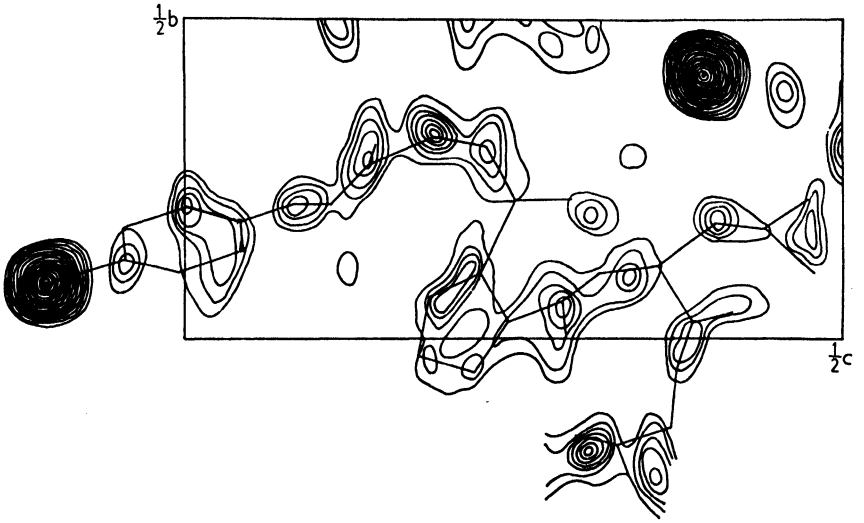


FIG. 5. Electron-density map projected on (100). The contour interval is 1 e/Å^2. The density of the base-line is 3 e/Å^2, and the contour lines in the minus region are neglected.

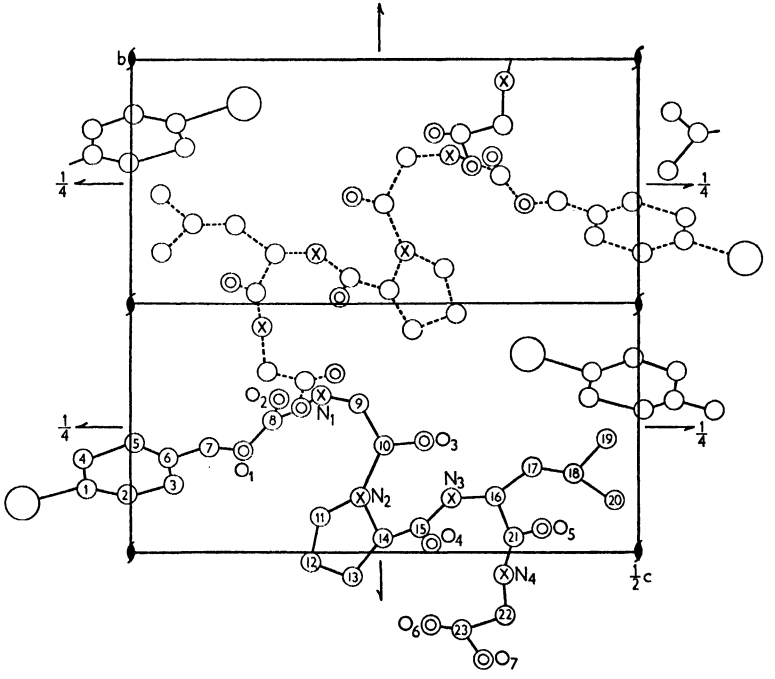


FIG. 6. Projection of the structure viewed along the *a*-axis.

atom was assumed from the trials with the aid of the model and from the Patterson map projected on the b-plane. The y and z parameters thus refined are listed in Table II.

TABLE II. Atomic parameters

Atom	y	z	Atom	y	z
C(1)	0·125	−0·045	C(21)	0·026	0·395
C(2)	0·096	−0·002	C(22)	−0·140	0·363
C(3)	0·135	0·034	C(23)	−0·172	0·320
C(4)	0·166	−0·040	—	—	—
C(5)	0·214	0·001	N(1)	0·314	0·183
C(6)	0·182	0·037	N(2)	0·117	0·229
C(7)	0·212	0·080	N(3)	0·103	0·315
C(8)	0·266	0·139	N(4)	−0·048	0·376
C(9)	0·307	0·226	—	—	—
C(10)	0·212	0·252	O(1)	0·213	0·108
C(11)	0·064	0·189	O(2)	0·309	0·145
C(12)	−0·031	0·177	O(3)	0·228	0·299
C(13)	−0·053	0·222	O(4)	0·008	0·290
C(14)	0·023	0·246	O(5)	0·042	0·423
C(15)	0·060	0·287	O(6)	−0·153	0·295
C(16)	0·114	0·365	O(7)	−0·218	0·335
C(17)	0·169	0·403	—	—	—
C(18)	0·164	0·442	Br	0·087	−0·103
C(19)	0·129	0·487	—	—	—
C(20)	0·223	0·484	—	—	—

5. DISCUSSION

A. *Bromopentapeptides*

There are four molecules of the p-bromo compound in the unit cell, while the unit cell of the o-bromo compound contains two molecules. As the space group $P2_1$ has only two equivalent positions, there exist two kinds of crystallographically independent molecules in the p-bromo compound. However, the reflections with odd l indices of the p-bromo compound were fairly weak. These facts suggest that the p-bromo compound is similar to the o-bromo compound not only in the molecular shapes, but also in the over-all nature of molecular packing.

B. p-*Bromotetrapeptide*

The bond distances and angles in this molecule cannot be precisely calculated because of the ambiguity of the x parameters. However, the approximate structure of the molecule derived here is considered to be

sufficient to discuss the whole feature of the peptide chain and the orientations of the amino acid residues. The principal chain starting from the carbobenzoxy group attached to the N-terminal end of the peptide is in a stretched form with *trans-trans* bondings up to the first glycine residue. Then, the chain is folded back so as to dispose N(1)—C(9) and C(10)—N(2) in a *gauche* form to C(9)—C(10). This seems to be due to a stabilization of the whole structure of the molecule and molecular packing. A twist occurs at the proline residue so as to give a ring formed by an intramolecular hydrogen bond between O(3) and N(3)H. It is a very interesting fact that the main chain around here is folded, and forms the intramolecular hydrogen bond just like the 2_n type of Bragg *et al.* (1950). However, this feature is somewhat different from those in collagen (Ramachandran and Kartha, 1955), and in di- or tripeptides containing a proline ring (Fridrichsons and Mathieson, 1962; Leung and Marsh, 1958). Following the proline residue, the chain is folded back again to give the same type of hydrogen bond. Another interesting fact is that the folding plane at the proline residue and the folding plane at the leucine residue are not in the same plane, but are at about 30 degrees to each other. This suggests that the peptide chain seems to be just in an intermediate state to conform to a helical structure.

As regards the molecular association in the crystal, *p*-bromocarbobenzoxy groups contact each other in an antiparallel manner throughout a screw axis along the *c*-axis. There are small cavities around the *ortho* positions and the proline rings, and this fact may give a reasonable explanation of the expansion of the unit cell by the *ortho* substitution of bromine atom.

Another evidence from the infra-red absorption spectra suggests the validity of the molecular configuration. The 3340 cm^{-1}, 3268 cm^{-1} and 3111 cm^{-1} absorption bands are assigned as the hydrogen bonded NH and OH stretching vibrations. In the Amide I and II regions the absorption bands at 1681, 1653, 1639, 1555 and 1522 cm^{-1} were observed, and in the low-frequency region the absorption band at 670 cm^{-1} was also observed. These absorption bands may also uphold the possibility of the existence of a folded structure in the peptide chain. A more detailed analysis of the tetrapeptide is in progress now.

Acknowledgements

The authors wish to express their sincere thanks to Miss K. Tanaka, Y. Ogawa and Mr. S. Orii, who have given so much assistance for the X-ray experiments. The calculations throughout in this structure analysis were carried out by NEAC-2203 electronic computer, and acknowledgement also is made to Miss Y. Ogawa for her co-operation in the computing. The infra-red absorption spectra were measured by Dr. K. Fukushima, and he also gave kind suggestions about the interpretation of the spectra. The authors express their sincere gratitude to him, too.

References

Bragg, W. L., Kendrew, J. C. and Perutz, M. F. (1950). *Proc. roy. Soc.* **A203**, 321.
Fridrichsons, J. and Mathieson, A. McL. (1962). *Acta cryst.* **15**, 569.
Leung, Y. C. and Marsh, R. E. (1958). *Acta cryst.* **11**, 17.
Nagai, Y. and Noda, H. (1959). *Biochim. biophys. Acta* **34**, 298.
Nagai, Y., Sakakibara, S., Noda, H. and Akabori, S. (1960). *Biochim. biophys. Acta* **37**, 567.
Ramachandran, G. N. and Kartha, G. (1955). *Proc. Indian Acad. Sci.* **A42**, 215.
Sakakibara, S. and Nagai, Y. (1960). *Bull. chem. Soc. Japan* **33**, 1537.
Sasada, Y. and Kakudo, M. (1961). *Acta cryst.* **14**, 1305.
Sasada, Y. Tanaka, K., Ogawa, Y. and Kakudo, M. (1961). *Acta cryst.* **14**, 326.

Discussion

G. N. RAMACHANDRAN: The occurrence of the internal hydrogen bond of the type $2n$ in this peptide is quite interesting, and I think until now it has never been found in either a simple peptide or a polypeptide, though it has long ago been predicted by Huggins and others; but about a year ago, Dr. Sasisekharan and I tried to analyse different types of helical structures and we did find that there is a definite possibility of its occurrence (G. N. Ramachandran, C. Ramakrishnan and V. Sasisekharan, this volume, p. 121). It is a particular type of configuration which, if continued, will have a two-fold symmetry and it is nice to see that it does occur in a peptide structure.

M. KAKUDO: We are going to continue a more detailed analysis.

C. RAMAKRISHNAN: What is the order of hydrogen bond lengths in these peptides?

M. KAKUDO: Precise bond length measurements were not possible since the x-co-ordinates were rather unreliable. The distance of the N(3)···O(3) and N(4)···O(4) contacts, which are expected to form internal hydrogen bonds, were both estimated to be less than 3 Å for the present model.

Structure of Poly-L-proline I

W. Traub and U. Shmueli

Department of X-ray Crystallography, Weizmann Institute of Science, Rehovoth, Israel

ABSTRACT

The structure of poly-L-proline I has been determined from X-ray diffraction photographs of powder and oriented specimens. The polymer forms a right-handed helix with a translation of 1·90 Å and a rotation of 108° per proline residue. The peptide groups are in the *cis* configuration. The atomic co-ordinates have been determined and the possibility of alternative structures eliminated by molecular-model studies and accurate geometrical constructions. Poly-L-proline I has the monoclinic space group $P2_1$ with pseudo-hexagonal unit-cell dimensions $a = b = 9{\cdot}05$ Å, $c = 19{\cdot}0$ Å and $\gamma = 120°$. A satisfactory mode of packing the polymer chains in this unit cell has been found. The structure has been confirmed by a comparison of calculated and observed intensities, which show good agreement.

1. Introduction

From the point of view of protein structure, the imino acid proline is of special interest. Lacking a hydrogen atom at the imide group, proline can participate in, at the most, one hydrogen bond as compared with the two made by each amino acid residue in an α-helix or extended β structure (Pauling and Corey, 1953). Furthermore, the geometry of the pyrrolidine ring prevents proline from fitting into an undistorted α-helix except at its amino-end (Lindley, 1955). For these reasons, proline is believed to be associated with corners in polypeptide chains and to have an important role concerning the tertiary structure of globular proteins, as has recently been borne out by studies of myoglobin (Kendrew *et al.*, 1961) and haemoglobin (Watson and Kendrew, 1961). It has also been pointed out that, if the proline peptide bond is in the *cis* configuration (i.e. the two α-carbons adjacent to the bond are *cis* to one another), a sharp bend, involving a reversal of direction, can be produced in a polypeptide chain (Edsall, 1954).

Poly-L-proline has been synthesized through the polymerization of *N*-carboxy-L-proline anhydride (Berger *et al.*, 1954). It does not form the α and β structures formed by several other synthetic polypeptides (Bamford *et al.*, 1956). However, it has been found to exist in two distinct forms, which exhibit markedly different optical rotations, infra-red

spectra (Blout and Fasman, 1958; Steinberg *et al.*, 1958) and X-ray diffraction patterns. The polymer obtained from the polymerization of the anhydride in pyridine, which is termed poly-L-proline I, exhibits a specific rotation $[\alpha]_D^{25} = +50°$ in acetic acid or water, whereas poly-L-proline II, obtained from form I by mutarotation in acid or water, shows $[\alpha]_D^{25} = -540°$ (Kurtz *et al.*, 1956). Reverse mutarotation of poly-L-proline II to poly-L-proline I in alcoholic media has also been observed (Steinberg *et al.*, 1958). Similar transitions from one form to another have been observed in several other poly-α-imino acids. These include poly-*O*-acetyl-L-hydroxyproline (Kurtz *et al.*, 1958), poly-L-pipecolic acid (Kurtz *et al.*, 1962), and polydehydro-L-proline.* However, the nature of these mutarotational changes has not yet been clearly established, as up to now the molecular architecture of only one of these structural forms, poly-L-proline II, has been elucidated (Cowan and McGavin, 1955; Sasisekharan, 1959).

Poly-L-proline II has been shown to possess a left-handed helical structure with peptide bonds in the *trans* configuration. The structure has three proline residues per turn of the helix and a repeat distance of 3·12 Å per residue along the longitudinal axis.

Extensive physico-chemical studies have been made of poly-L-proline in solution, and from these much has been inferred about the possible structure of poly-L-proline I. The chemical identity of poly-L-proline I and poly-L-proline II was established by the observation that both polymers yield L-proline quantitatively on acid hydrolysis (Kurtz *et al.*, 1956). Axial ratios determined by viscosity measurements have been taken to indicate that poly-L-proline I has an appreciably smaller length per residue than poly-L-proline II (Steinberg *et al.*, 1960). The catalysis of mutarotation and reverse mutarotation by acids (Steinberg, 1960) has led to the suggestion that *cis-trans* isomerization of the peptide bond is involved and that poly-L-proline I has *cis* peptide bonds (Kurtz *et al.*, 1956; Steinberg *et al.*, 1960). Furthermore, from an estimate of the intrinsic residue rotation of L-proline in the peptide chain as −300°, and calculations based on the Fitts and Kirkwood (1956*a*, 1956*b*) theory for the optical rotation of helical molecules, it has been concluded that poly-L-proline I has a right-handed helical structure (Harrington and Sela, 1958). It has been reported that the stereochemical plausibility of these ideas has been tested by Crick and Rich, who succeeded in building a right-handed helical model of poly-L-proline with peptide bonds exclusively in the *cis* configuration (Harrington and Sela, 1958).

An X-ray analysis of the molecular structure of poly-L-proline I seemed desirable for several reasons: to extend knowledge of the configuration of the proline residue in polypeptide chains; to elucidate the

* *See* Behaviour in Solution of Polypeptides Related to Collagen, p. 205.

nature of the structural changes accompanying mutarotation; and to test the validity of the theories and interpretations that led to the structural postulates outlined above. However, until recently a structure analysis of form I seemed to be precluded by the lack of adequate X-ray data.

2. Experimental Data

X-ray powder photographs of poly-L-proline I and poly-L-proline II are shown in Fig. 1 (a) and (b). In contrast to the rich highly crystalline pattern of form II, form I, when precipitated with ether from the polymerization medium, pyridine, shows only a few diffuse rings, hardly susceptible to detailed structural interpretation. Far more promising X-ray photographs were obtained by Sasisekharan (1960), who made an extensive study of complex formation and structural transformations in poly-L-proline when forms I and II were treated with various solvents. In particular, he found that poly-L-proline I made into a wet paste with glacial acetic acid gave a crystalline X-ray diffraction pattern apparently of a poly-L-proline–acetic-acid complex. When the paste was air-dried another new crystalline pattern appeared. He found that the specific rotation of the material giving the latter pattern was $[\alpha]_D^{27} = -100°$ approximately, and concluded that this was a new form of poly-L-proline, intermediate between forms I and II, to which he gave the name poly-L-proline IA. Sasisekharan was also able to obtain the IA pattern by treating form I with formic acid, *m*-cresol, ethyl alcohol, or formamide.

The wealth of detail in the IA pattern, and a rough correspondence which we noted between the spacings of its stronger lines and those of the diffraction rings of form I, encouraged us to reinvestigate the problem. We found that pastes of forms IA and I with acetic acid gave identical crystalline X-ray patterns. Furthermore, a sample of IA prepared with ethyl alcohol, which unlike acetic acid does not cause mutarotation, was found to have the same specific rotation as poly-L-proline I. Thus it seems clear that form IA is not an intermediate between forms I and II, but in fact a crystalline form of poly-L-proline I. Indeed we have observed appreciable variations in the degree of crystallinity of the IA patterns we have obtained with the use of different solvents. The intermediate specific rotation value obtained for IA by Sasisekharan may possibly have been caused by poly-L-proline II contamination, which is suggested by the presence of form II lines in his photographs of IA. A similar contamination might account for the observation by Blout and Fasman (1958) of poly-L-proline II features in the infra-red spectrum of poly-L-proline I which had been treated with acetic acid.

Well defined powder photographs were obtained from poly-L-proline I recrystallized from glacial acetic acid or propionic acid (Fig. 1 (c)). Nearly

twenty lines were observed and the spacings measured, using silver chloride as a standard (Table I).

It proved difficult to prepare well oriented specimens, and a number of solvents and techniques were tried in an effort to obtain suitable films or fibres. The best oriented photographs were obtained with films grown from ethyl alcohol solution and from fibres pulled from concentrated solutions of form I in propionic acid. One of these photographs is shown in Fig. 2. Although the orientation is insufficient for the assignment of the various reflections to different layer lines, it is possible to determine the approximate orientation of most of them as indicated in Table I.

TABLE I. Observed and calculated spacings and intensities of poly-L-proline I reflections

d_0(Å)	I_0	Orientation	hkl	d_c (Å)	I_c
7·84	150	equatorial	100	7·83	148
Not resolved			101	7·25	31
—	—	—	102	6.04	1
4·89	110	near meridional	103	4·93	127
4·49	9	equatorial	110	4·53	7
4·38		equatorial	111	4·40	10
4·06	18	diagonal	112	4·08	14
			104	4·07	4
3·92	3	equatorial	200	3·91	10
—	—	—	201	3·83	2
3·64	13	diagonal	113	3·68	13
			202	3·62	13
—	—	—	105	3·43	—
3·32	2	(?)	203	3·33	4
			114	3·28	3
—	—	—	204	3·02	2
2·94	22	broad equatorial	210	2·98	4
			106	2·94	—
			211	2·92	7
			115	2·91	1

Very weak reflections were also observed at 2·70 Å, 2·59 Å, 2·34 Å, 2·14 Å (diagonal), 2·03 Å (meridional), 1·90 Å (meridional) and 1·85 Å (meridional). Calculated and observed intensities are approximately on the same relative scale.

The density of crystalline preparations of poly-L-proline I were determined by the flotation method in two different ways. In the first a chloroform–benzene mixture was used; in the second a mixture of aqueous solutions of potassium carbonate (K_2CO_3) of different concentrations. The results from the two methods agreed well, the observed density being 1·24 g/cm^3.

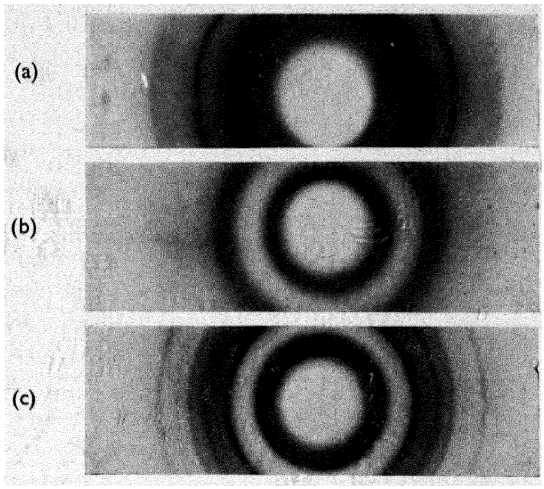


FIG. 1. X-Ray powder photographs. (a) Poly-L-proline II. (b) Poly-L-proline I, amorphous form. (c) Poly-L-proline I, recrystallized from propionic acid.

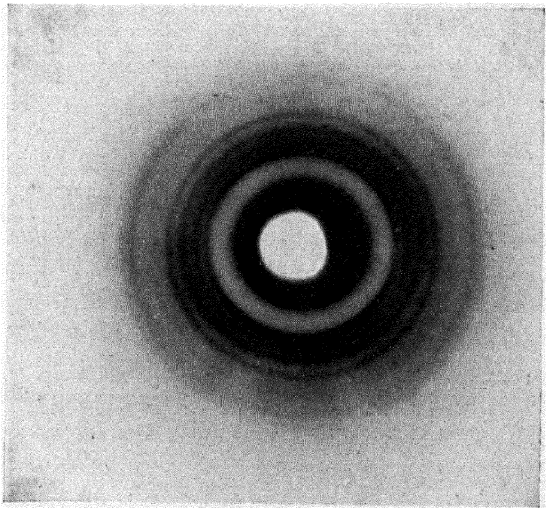


FIG. 2. X-Ray diffraction photograph of a fibre of poly-L-proline I drawn from propionic acid solution.

3. Materials and Methods

Samples of poly-L-proline of molecular weight 10,000–15,000 were obtained from the Biophysics Department of the Weizmann Institute, through the courtesy of Professor E. Katchalski.

Oriented films were grown at first on glass and later on Teflon from which it proved easier to remove the specimens.

Powder photographs were taken with 114·6 mm and 57·3 mm diameter cylindrical powder cameras and standard X-ray units. Oriented films or fibres were photographed on a Norelco microcamera used with a Hilger microfocus X-ray tube. Some thicker oriented specimens were photographed with longer specimen-to-film distances on a Unicam flat-plate camera with a Philips fine-focus tube. All photographs were taken with Cu $K\alpha$ radiation. The intensities of the reflections were measured with a Joyce Loebl recording microdensitometer.

Optical rotation measurements were kindly performed by Dr. J. Kurtz. For these a Rudolph model 70 polarimeter was used.

Two kinds of molecular models were used to aid the structural investigations. Model components built of brass rods to a scale of 5 cm = 1 Å, produced by Cambridge Repetition Engineers, proved particularly useful for the measurement of molecular dimensions and intramolecular distances. In addition, Courtauld's close-packing models, built to a scale of 0·8 in. = 1 Å, were used to study possible angular distortions and van der Waals contacts.

4. Determination of Helical Constants

The spacings and equatorial orientation of the reflections at 7·84 Å, 4·49 Å, 3·92 Å and 2·94 Å indicate a hexagonal or pseudo-hexagonal unit cell with $a = b = 9{\cdot}05$ Å. If furthermore the c-axis is chosen as 19·0 Å, all the reflections can be satisfactorily indexed in accordance with their observed orientations. Observed and calculated spacings are listed in Table I.

The observed density, 1·24 g/cm^3 indicates that there are ten proline residues in this unit cell. This would give a calculated density of 1·20 g/cm^3, whereas nine or eleven residues would imply densities of 1·08 g/cm^3, or 1·32 g/cm^3 respectively. As 19·0 Å is a true crystallographic repeat, ten residues must correspond to an integral number of helical turns. This number and ten cannot have a common factor, as this would imply a shorter c-axis making it impossible to index all the reflections. In fact molecular-model studies (see below) indicate that this number can only be 3. There are thus ten residues in three turns and 19·0 Å, implying a helical screw axis with a translation of 1·90 Å and a rotation of 108°.

This result is in accord with helical diffraction theory (Cochran *et al.*, 1952), the strong near-meridional 4·89 Å reflection corresponding to a first-order Bessel function.

5. Molecular-model Studies and Determination of Atomic Co-ordinates

Molecular models were now used to study the possible configurations of poly-L-proline chains and to see which of these might be consistent with the helical dimension derived above.

The configuration of a chain molecule can be defined in terms of the rotation about the various bonds along the chain. Of the three kinds of bonds along a poly-L-proline chain, rotation about the N—αC bond is severely limited because of its position in the pyrrolidine ring, as is rotation about the peptide bond C′—N because of its partial double-bond character. It was found that rotation about the αC—C′ bond was also considerably restricted by steric hindrance. In fact, even allowing the full 24° angular distortions possible with the Courtauld's models, only three limited regions of rotational freedom were found possible. One region has about 60° freedom of rotation about the αC—C′ bond, and consists of left-handed helices with *trans* peptide configurations, including the poly-L-proline II structure. However, all these helices have a translation per residue far greater than 1·90 Å. The other two possibilities have very little freedom of rotation. Both are non-integral right-handed helices with between three and four residues per turn and translations per residue of the order of 2 Å. One of these has a *cis* peptide configuration with rotation about the αC—C′ bond such that the oxygen atom is nearly *trans* to the αC hydrogen; the other has a *trans* peptide configuration with the oxygen nearly *cis* to the αC hydrogen. It was clear from these studies of packing models that no configurations were possible that did not have between two and four residues per turn of the helix.

Rod models of these two right-handed helices were now carefully constructed to test whether either was in accord with the helical dimensions which had been determined. It was found that a *cis* structure with 1·90 Å translation and 108° rotation per residue with satisfactory van der Waals contacts could indeed be constructed, but that a *trans* peptide structure built to these dimensions implied an oxygen 1 to δ-carbon 4 separation of about 2·0 Å. Thus it appeared that a right-handed helix of proline residues with *cis* peptide bonds was the only possible structure consistent with the X-ray data.

This result was confirmed by accurate geometrical constructions based on cyclographic projections (Sasisekharan, 1961), which were used also to determine the cylindrical co-ordinates of the atoms in the peptide group.

The constructions were based on the Pauling-Corey dimensions for the peptide unit (Corey and Pauling, 1953; Pauling and Corey, 1951). As the distance between α-carbon atoms of successive residues is 2·83 Å for *cis* peptides and 3·80 Å for *trans* peptides, the 1·90 Å translation and 108° rotation per residue fix the positions of the α-carbon atoms which lie on helices with radii of 1·28 Å and 2·03 Å for the *cis* and *trans* cases respectively. The positions of the other atoms of the peptide group depend upon its orientation with respect to the helical axis. The value of the angle N_1—αC_1—C_1' was determined for all possible orientations of the peptide group about the line αC_1—αC_2. At 5° intervals over the range of peptide orientation for which this angle is between 100° and 120°, the co-ordinates of the C′-, N-, O- and δC-atoms were determined, and all the van der Waals separations calculated. From these calculations it was found that of the four possibilities investigated (left-handed and right-handed helices with either *cis* or *trans* peptide bonds) only a *cis* right-handed helical structure has satisfactory van der Waals contacts between unbonded atoms, and that indeed even this configuration is only possible over a very small range of peptide orientation.

Given the co-ordinates of the atoms in the peptide group, positions were determined for the β- and γ-carbon atoms with the aid of a model of the pyrrolidine ring. When the co-ordinates of these atoms were used to calculate additional van der Waals separations, a 3·1 Å βC_1—αC_2 distance, as well as several other short contacts involving the β-carbon atom, were found. It became clear, both from models and calculation, that these short contacts could not be substantially improved by merely shifting the β- and γ-carbon atoms to conform with any possible alternative shape of the pyrrolidine ring. This led to a reconsideration of the dimensions used for the peptide group and in particular the α-carbon to α-carbon separation of 2·83 Å suggested by Pauling and Corey (1951), mainly on theoretical grounds. In fact, since this suggestion was made, leucyl–prolyl–glycine, which has a peptide group identical with that of poly-L-proline, has been found to have bond angles differing appreciably from Pauling and Corey's values and a carbon–carbon separation of 2·92 Å (Leung and Marsh, 1958).

The peptide dimensions were therefore revised in accordance with the values found for leucyl-prolyl-glycine and other structures having a pyrrolidine ring (Donohue and Trueblood, 1952; Mathieson and Welsh, 1952), the cyclograms redrawn, and the peptide co-ordinates recalculated. Again it was found that configurations other than a right-handed helix with *cis* peptide bonds could be excluded, but now β- and γ-carbon atoms could be incorporated in the latter structure, without any unacceptably short contacts between unbonded atoms.

The final atomic co-ordinates after a further slight adjustment of the

γ-carbon position (see Section 6) are given in Table II. Bond distances and angles are shown in Fig. 3, and a portion of the structure projected along the fibre axis in Fig. 4. Because of the increased separation between successive α-carbon atoms, these atoms now lie on a helix of radius 1·39 Å. The β- and γ-carbon atoms are 0·6 Å and 0·3 Å respectively out of the plane of the peptide group and both lie above the plane as shown in Fig. 3. The bond angle N_1—αC_1—C_1', is 114°, and short intramolecular

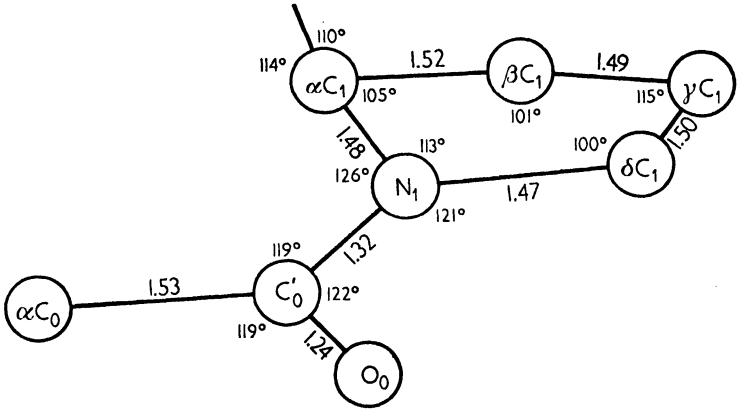


FIG. 3. The bond angles and distances of poly-L-proline I.

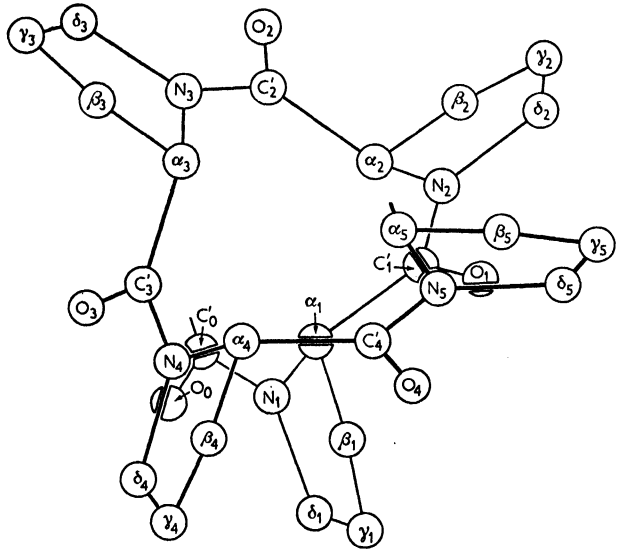


FIG. 4. Part of the structure of poly-L-proline I projected down the fibre axis.

contacts include O_1—N_1 = 2·80 Å, αC_1—C_2' = 3·50 Å, βC_1—αC_2 = 3·45 Å and βC_1—O_1 = 3·22 Å. The value of the angle at the α-carbon atom is larger than normal, though not unacceptably so. By choosing a different orientation of the peptide group about the line αC_1—αC_2, this value could

TABLE II. Cylindrical co-ordinates of poly-L-proline I.
There are ten residues in three turns in a height of 19·0 Å

Atom	r (Å)	ϕ (°)	z (Å)
αC_0	1·39	0·0	0·00
C_0	1·69	57·5	−0·21
O_0	2·41	58·0	−1·20
N_1	1·97	87·5	0·67
βC_1	2·59	107·5	2·83
γC_1	3·70	105·5	1·85
δC_1	3·37	97·0	0·49

be reduced, but this would also imply a reduction of the O_1—N_1 separation below the somewhat short van der Waals distance of 2·80 Å. On the other hand, a larger O_1—N_1 separation would imply a still larger N_1—αC_1—C_1' angle; hence the choice of peptide orientation from which the atomic co-ordinates of Table II were derived.

6. INTERMOLECULAR PACKING

From the molecular structure and the observed density it follows that there are ten proline residues comprising three turns of one helical chain in each unit cell. The only crystallographic symmetry which this group possesses is a two-fold screw axis along the axis of the helix. There is thus no true hexagonal symmetry in spite of the pseudo-hexagonal unit-cell dimensions and, in fact, the only possible space group is $P2_1$ of the monoclinic system.

The helix axis of the molecule must lie along the unique c-axis of the monoclinic cell, but the orientation of the molecule with respect to the other two crystallographic axes remains to be determined. Two projections of the molecule along the helix axis were drawn to a scale of 5 cm = 1 Å. These were kept 9·05 Å apart and parallel, and rotated together through intervals of $4\frac{1}{2}°$. For each position all short contacts between atoms of neighbouring molecules (calculated from the measured projected separation and known vertical separation) were noted. Because of the ten-fold helical symmetry of the molecule a rotation through 36° showed all the possible different modes of packing between neighbouring molecules. However, in each position the modes of contact along the a- and b-axes, which are 120° apart, differ from each other.

The molecules are not far from cylindrically symmetrical, and it was found that in all orientations the smallest separations between atoms of neighbouring molecules were near to the normal van der Waals contacts. This provides further confirmation of the essential correctness of the structure, especially when it is borne in mind that atoms in a helix of *trans* peptides would lie on very different radii. Slightly short contacts were found between γC_1 and O_2 of neighbouring molecules and γC_1 and

TABLE III. Unit-cell co-ordinates of poly-L-proline I

Atom	x	y	z
αC_0	0·064	0·175	0·000
βC_0	0·121	0·327	0·049
γC_0	0·188	0·469	−0·003
δC_0	0·228	0·430	−0·074
N_0	−0·167	0·246	−0·065
C_0'	−0·128	0·086	−0·011
O_0	−0·185	0·120	−0·063

δC_3, which varied from 2·9 Å to 3·7 Å and from 3·2 Å to 3·9 Å respectively as the molecules were rotated about their axes. An examination of the short contacts indicated a most favourable orientation, and by changing the γ-carbon position by 0·2 Å it was found possible to arrive at atomic co-ordinates which satisfy inter- as well as intra-molecular packing criteria. These are listed in Tables II and III.

7. COMPARISON OF CALCULATED AND OBSERVED INTENSITY DATA

Structure factors were computed from the co-ordinates of Table III on the Weizmann Institute electronic computer WEIZAC with a programme written by Dr. F. L. Hirshfeld. The value of the temperature factor used was $B = 4$ Å^2, a value commonly found in crystals of organic compounds. Calculated intensity values, derived from these structure factors, are listed in Table I.

The observed intensity values were estimated from microdensitometer traces of powder photographs. It was clear that there was a high background of scattering from amorphous material. An approximate correction for this was applied on the basis of microdensitometer traces of photographs of oriented specimens and amorphous poly-L-proline I. Intensity estimations were further complicated by the poor resolution of several spectra, particularly 200 and 101, the latter being almost com-

pletely hidden by the very strong and broad 100 peak. For these reasons the observed intensity values must be regarded as only semi-quantitative.

As can be seen from Table I, the agreement between calculated and observed intensities is nevertheless very good, and provides further strong confirmation of the correctness of the structure.

8. Conclusions

While it is important to show that a satisfactory solution can be formulated in terms of precise co-ordinates, it should not be thought that the structure analysis, based on so few reflections and perforce on many assumptions regarding bond distances and angles, allows of such accurate estimation of atomic positions.

However, there can be little doubt about the essential correctness of the structure. Not only is a right-handed helix with *cis* peptide groups the only configuration consistent with the spacings and orientations of the X-ray reflections, but this result is strongly supported by the satisfactory intermolecular packing and agreement between calculated and observed intensities, as well as a considerable amount of physico-chemical evidence.

Acknowledgements

We wish to thank Professors Ephraim Katchalski and Gerhard Schmidt for their encouragement and interest in this work and Dr. Joseph Kurtz for preparing samples of poly-L-proline and performing the optical rotation measurements. This investigation was supported by a P.H.S. research grant G-8608 from the Division of General Medical Sciences, United States Public Health Service.

References

Bamford, C. H., Elliot, A. and Hanby, W. E. (1956). In *Synthetic Polypeptides*. Academic Press, New York.

Berger, A., Kurtz, J. and Katchalski, E. (1954). *J. Amer. chem. Soc.* **76**, 5552.

Blout, E. R. and Fasman, G. D. (1958). In *Recent Advances in Gelatin and Glue Research* (G. Stainsby, ed.), p. 122. Pergamon Press, London.

Cochran, W., Crick, F. H. C. and Vand, V. (1952). *Acta cryst.* **5**, 581.

Corey, R. B. and Pauling, L. (1953). *Proc. roy. Soc.* **B141**, 10.

Cowan, P. and McGavin, S. (1955). *Nature, Lond.* **176**, 501.

Donohue, J. and Trueblood, K. N. (1952). *Acta cryst.* **5**, 414.

Edsall, J. T. (1954). *J. Polym. Sci.* **12**, 256.

Fitts, D. D. and Kirkwood, J. G. (1956*a*). *J. Amer. chem. Soc.* **78**, 2650.

Fitts, D. D. and Kirkwood, J. G. (1956*b*). *Proc. nat. Acad. Sci., Wash.* **42**, 33.

Harrington, W. F. and Sela, M. (1958). *Biochim. biophys. Acta* **27**, 24.

Kendrew, J. C., Watson, H. C., Dickerson, R. E., Phillips, D. C. and Shore, V. C. (1961). *Nature, Lond.* **190**, 666.

Kurtz, J., Berger, A. and Katchalski, E. (1956). *Nature, Lond.* **178**, 1066.

Kurtz, J., Berger, A. and Katchalski, E. (1958). In *Recent Advances in Gelatin and Glue Research* (G. Stainsby, ed.), p. 131. Pergamon Press, London.

Kurtz, J., Berger, A. and Katchalski, E. (1962). *Bull. Research Council Israel*, **11A**, 84.
Leung, Y. C. and Marsh, R. E. (1958). *Acta cryst.* **11**, 17.
Lindley, H. (1955). *Biochim. biophys. Acta* **18**, 197.
Mathieson, A. McL. and Welsh, H. K. (1952). *Acta cryst.* **5**, 599.
Pauling, L. and Corey, R. B. (1951). *Proc. nat. Acad. Sci., Wash.* **37**, 273.
Pauling, L. and Corey, R. B. (1953). *Proc. roy. Soc.* **B141**, 21.
Sasisekharan, V. (1959). *Acta cryst.* **12**, 897.
Sasisekharan, V. (1960). *J. Polym. Sci.* **47**, 373.
Sasisekharan, V. (1961). *Proc. Indian Acad. Sci.* **A53**, 296.
Steinberg, I. Z. (1960). *Bull. Research Council Israel*, **9A**, 118.
Steinberg, I. Z., Berger, A. and Katchalski, E. (1958). *Biochim. biophys. Acta* **28**, 647.
Steinberg, I. Z., Harrington, W. F., Berger, A., Sela, M. and Katchalski, E. (1960). *J. Amer. chem. Soc.* **82**, 5263.
Watson, H. C. and Kendrew, J. C. (1961). *Nature, Lond.* **190**, 670.

Discussion

E. R. BLOUT: As you probably remember, several years ago we reported the infrared dichroism of oriented solid films of poly-L-proline I and poly-L-proline II. The results we obtained indicated that the C=O orientation in both forms was predominantly perpendicular to the orientation direction. These data fit well with the diffraction data of Cowan and McGavin for poly-L-proline II. However, when I was in Israel a few days ago, I inspected the model of poly-L-proline I, which indicates that the carbonyl groups lie predominantly parallel to the fibre direction. If your model for poly-L-proline I is the correct one, it is then necessary to assume that the fibre axis of the preparations we examined lie perpendicular to the orientation direction. Have you any experience which would indicate that poly-L-proline I orients with its fibre axis perpendicular to the orientation direction?

W. TRAUB: In the case of poly-L-proline with propionic acid, we did observe the orientation of the chain perpendicular to the fibre axis. This has also been observed in polyethylene and also in poly-L-proline II by Cowan and McGavin.

A. ELLIOTT (King's College, London): There is a possibility not apparently considered by the authors, namely that there could be molecules of opposite chain sense in the same crystallite (anti-parallel chains). A regular crystallographic arrangement would give rise to additional reflections, but these would probably not be observed with the low orientation achieved. Alternatively, a random chain sense on a hexagonal arrangement of sites is possible. (This occurs in poly-L-alanine and almost certainly occurs in the ω-form of poly-β-benzyl-L-aspartate, and this would likewise not be recognized.) In either case, the intensities would be different from those calculated for the one-molecule unit cell. It is rather unlikely that there would be only one chain sense in a crystallite, for the interaction forces which might make this a state of lower potential energy than an arrangement of anti-parallel chains becomes appreciable only when the chains are close together. With long-chain molecules, it seems unlikely that at this stage a rearrangement of chain sense would be possible.

W. TRAUB: Yes. A reversal would be more difficult.

The Molecular Structure of α-Keratin

G. SWANBECK

Department of Medical Physics, Karolinska Institutet, Stockholm, Sweden

ABSTRACT

The diffraction pattern of α-keratin has been described. It has been suggested that α-keratin crystallizes in a cylindro-helical lattice as well on the molecular as on the microfibrillar level. The wide-angle diffraction pattern has been interpreted on this basis and geometrically optimal conditions for side-chain interaction in a multi-stranded cable of α-helices has been derived. A model for the structure and arrangements of the protein chains in α-keratin has been proposed. According to this model the protein chains are arranged in three concentric layers. Further data about the model and its stereochemical consequences are discussed. The possibility of the occurrence of one or two more concentric layers of protein chains with a high sulphur content is considered. An explanation for the particular diameter of the keratin microfibril is suggested on a stereochemical basis.

The diffraction pattern of a slightly disoriented fibre does not give enough information for a complete structural analysis. Additional information has to be added in the form of stereochemical, ultrastructural or other data. This approach to structural analysis has successfully been applied to DNA (Watson and Crick, 1953) and collagen (Ramachandran and Kartha, 1955).

In the present investigation an attempt is made to interpret the diffraction pattern of α-keratin on the basis of stereochemical and ultrastructural information.

1. Description of the Diffraction Pattern of α-Keratin

The diffraction pattern of α-keratin has already been described in detail by MacArthur (1943), Bear (1943) and Lang (1956*b*). Outside the continuous scatter at low-angles at the equator, three maxima follow at about 80 Å, 41 Å and 27 Å respectively. The first of these is considerably stronger than the other two, and is also rather sharp, while the last one is rather blurred. The next strong maximum on the equator has its centre at 9·8 Å. In some patterns two components might be distinguished at 10·5 Å and 9·2 Å respectively. Outside and inside of this strong maximum extremely weak maxima can be seen. There is also a rather strong and diffuse maximum around 4·3 Å. The meridional and near-meridional pattern can be indexed as higher orders of a 197 Å repeat. On the 3rd layer-line there is a strong meridional maximum. The 7th layer-line has a

strong near-meridional maximum and the 8th a strong meridional maximum. The 10th layer-line has near-meridional and the 11th strong meridional maxima. The 16th and 19th layer-lines have strong meridional maxima. There are weak maxima on the 28th, 30th, 32nd and 35th layer-lines. On the 38th layer-line there is a very strong meridional maximum at 5·18 Å. There is also a maximum at 1·48 Å, which corresponds to the 133rd layer-line. Figure 1 shows the low-angle part of the diffraction pattern of an African porcupine quill tip.

2. Crystal Lattice of α-Keratin

As was mentioned above, the low-angle equatorial diffraction pattern has three maxima, which can be regarded as the first, second and third order of a 80 Å repeat. There is no maximum at a spacing of $1/\sqrt{3}$ or $1/\sqrt{7}$ times the repeat unit as is expected for a hexagonal lattice. Ramachandran and Sasisekharan (1956) have reported similar findings for collagen, feather keratin and α-keratin and interpret them as an indication of the presence of a cylindrical lattice.

Electron microscopic studies by Birbeck and Mercer (1957) have shown, that within small restricted regions a hexagonal packing seems to exist, but that a whorl-like packing is more common. The microfibrils, which seem to have a circular cross-section with a diameter of 60–80 Å, are arranged in twisted cables with a diameter of a few thousand Ångström units.

The wide-angle equatorial diffraction pattern does not give any clear-cut information about the packing of the protein chains. It is not consistent with a perfect hexagonal lattice but may be consistent with an imperfect cylindrical lattice.

If a number of straight circular cylinders are brought closely together with their axes parallel, they will prefer a hexagonal packing rather than packing in a cylindrical lattice. It is therefore not probable that the microfibrils of α-keratin or the three-strand cables of collagen are straight circular cylinders, as they pack in a cylindrical lattice. If there is some surface structure or some other factor, such that the fibrillar units tend to coil around each other, there might be conditions for packing in cylindrical concentric layers with the fibrillar units having a helical path within each cylindrical layer. Such a lattice can be called a cylindro-helical lattice. The pitch of the helical paths of the different cylindrical layers might be the same or different. If the pitch is different, the cylindrical layers will be more perfect than if it is the same. In the latter case the units will have a tendency to lie in the groves between the units of the neighbouring layers.

It should also be pointed out that the fact that the micro-fibrils and

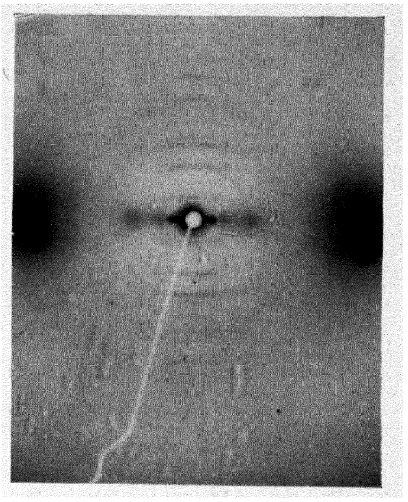
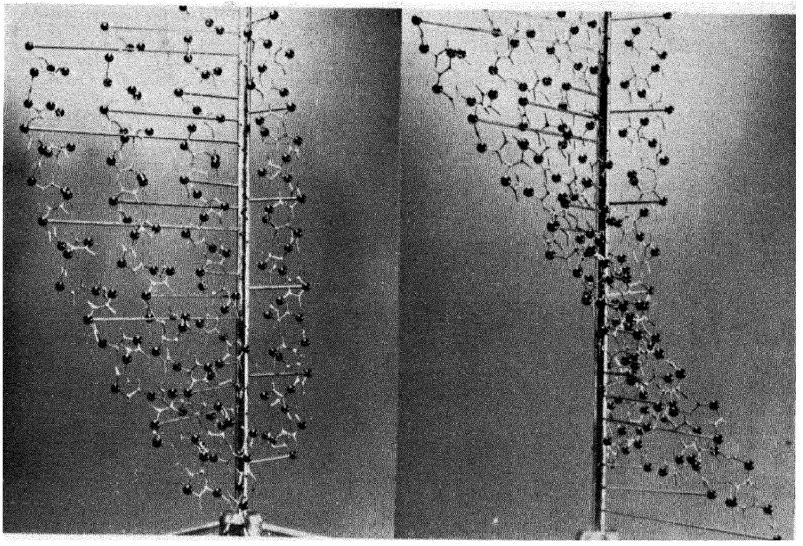


FIG. 1. Low-angle diffraction pattern of a porcupine quill tip.



(a) (b)

FIG. 3. Photograph of a three-dimensional model taken from two directions.

fibres of α-keratin, and collagen fibrils all have circular cross-sections, indicates that in these structures the units are packed in some sort of a cylindrical lattice.

3. Interpretation of the Wide-angle Diffraction Pattern

If the crystalline lattice of α-keratin is of the cylindro-helical type, the theory for the diffraction by helices (Cochran *et al.*, 1952; Klug *et al.*, 1958), coiled-coils (Crick 1953*a*; Lang 1956*a*), or multiple coiled structures shall be the basis for the interpretation of the diffraction pattern. For the more detailed discussion of these theories the reader is referred to the original papers. Only a short description of the diffraction theory of coiled-coils will be given here.

In the derivation of the diffraction theory of coiled-coils (Crick, 1953*a*) one uses two co-ordinate systems. In one system (x,y,z) the axis of the smaller helix follows a helical path, which is called the major helix. In the other co-ordinate system (x',y',z') the smaller helix is a perfect helix, and this co-ordinate system moves along the major helix in (x,y,z), so that the z'-axis is tangential to the major helix and the x'-axis always points away from the z-axis. The y'-axis is perpendicular to the z'-axis and the x'-axis.

A type of coiled-coil can be characterized by the parameters N_0, N_1, M, r_0, r_1 and c, where N_0 is the number of turns of the major helix in the repeat distance, c, along the z-axis; N_1 is the number of turns of the minor helix in the (x',y',z') co-ordinate system in the same distance, c; M is the number of amino acid residues or asymmetric units in c; r_0 and r_1 are the radii of the major and minor helices respectively. If the major helix is right-handed, N_0 is positive, and if it is left-handed, N_0 is negative, and correspondingly for the minor helix. This is partly contrary to Crick's (1953*a*) and Lang's (1956*a*) notation. When the minor helix makes N_1 turns in (x',y',z') it thus makes (N_1+N_0) turns in (x,y,z).

The Fourier transform of a coiled-coil is given by

$$
\begin{aligned}
F\left(R,\psi,\frac{l}{c}\right) = C \sum_{p=-\infty}^{+\infty} \sum_{q=-\infty}^{+\infty} \sum_{s=-\infty}^{+\infty} \sum_{d=-\infty}^{+\infty} & J_p(2\pi R r_0) \times \\
& \times J_q\left[2\pi R r_1 \frac{(1+\cos\alpha)}{2}\right] \times J_s\left[2\pi\left(\frac{l}{c}\right) r_1 \sin\alpha\right] \times \\
& \times J_d\left[2\pi R r_1 \frac{(1-\cos\alpha)}{2}\right] \times \exp\left[ip\left(\psi+\phi_0+\frac{\pi}{2}\right)+\right. \\
& +iq\left(-\psi+\phi_0+\phi_1+\frac{\pi}{2}\right)+is(\pi+\phi_1)+ \\
& \left.+id\left(\psi+\phi_1-\phi_0+\frac{\pi}{2}\right)-im\phi_M+2\pi i l z_0/c\right].
\end{aligned}
$$

subject to the conditions $N_0 p-(N_1+N_0)q-N_1 s-(N_1-N_0)d=l+Mm$ where p, q, s, d and m are integers.

Strong meridional maxima can be expected on layer-lines $l=N_1$ and $l=M$. For α-keratin these fall on spacings equal to 5·18 Å and 1·48 Å respectively (Lang, 1956*b*). The quotient M/N_1 gives the number of units per turn of the minor helix in the (x',y',z') co-ordinate system. For α-keratin $M/N_1 = 5{\cdot}18/1{\cdot}48 = 3{\cdot}50$. As the maxima at the above-mentioned spacings are the dominating meridional maxima in the wide-angle diffraction pattern, we can assume that the ratio 3·5 is true for the majority of protein chains in α-keratin. We can also state that the pitch of the minor helix in the (x',y',z') co-ordinate system projected on the z-axis is the same for most protein chains in α-keratin.

4. Effect of Side-chain Interaction

The α-helix is the most stable helical configuration of a protein chain, when no external forces are operating, or when the side-chain interaction is neglected. The relative instability of some other helical configurations related to the α-helix has been calculated by Donohue (1953).

At an early stage it was shown that the α-helix itself did not satisfy the diffraction pattern of α-keratin. Crick (1952) and Pauling and Corey (1953) therefore suggested a coiled-coil structure, very closely related to the α-helix structure, for the protein chains in α-keratin.

Crick (1953*b*) has discussed the geometrical aspects of the side-chain interaction and described two structures, the two-strand cable and the three-strand cable, with close side-chain interaction.

The present discussion of the side-chain interaction will be based on that of Crick (1953*b*), but extended to more complicated structures than the three-strand cable. As the majority of the side-chains reach further out from the axis of the protein chain than half the distance between two protein chains (see Waugh, 1958), the side-chains of one protein chain have to fit into the spaces between the side-chains of a neighbouring protein chain. Crick calls this packing knob-into-hole packing.

We assume that all α-helices are of the same sense, say right-handed. Crick (1953*b*) states that to get a perfect knob-into-hole packing of helices of the same sense, there must be $(n+1/2)$ knobs per turn. By tilting two α-helices slightly with respect to each other there will be 3·5 units per turn as referred to the line of contact, instead of 3·6 units, which occur in parallel α-helices. By deforming the chains slightly they can coil around each other forming a cable and having 3·5 units per turn as referred to the line of contact.

If we now assume that the protein chains in α-keratin are packed in a cylindro-helical lattice with the pitch of the major helices the same in all

cylindrical layers, the line of contact between two protein chains, in the same or in different layers, will be a helix with the same pitch as the major helices. To get a perfect knob-into-hole packing of side-chains there must thus be 3·5 units per turn of all minor helices in their (x', y', z') co-ordinate system. This is exactly what was predicted above from the wide-angle diffraction pattern of α-keratin. Another criterion for a perfect knob-into-hole packing under the above conditions for the cylindro-helical lattice is that N_1 is the same for all chains. This could also be predicted from the wide-angle diffraction pattern.

The next thing to consider is the intra-helical consequences of the side-chain interaction. However, it is first necessary to fix the values for the parameters N_0, N_1, M, r_0 and c. The value of r_1 is more or less fixed by the assumption that the protein chains have basically α-helical structure.

5. Description of the Model

There is now general agreement on the fact that the axial repeat distance c for α-keratin is about 197 Å (see Lang, 1956*b*). The most probable value for N_0 is -1. If the minor helix is right-handed the major helix has to be left-handed and *vice versa*.

$$N_1 = 197/5{\cdot}18 = 38 \text{ and } M = 197/1{\cdot}48 = 133.$$

There might be several different ways of starting the crystallization of protein chains in a cylindro-helical lattice under the above-mentioned conditions. Here only one way will be discussed. We start with a three-strand cable ($r_0 = 5$ Å) as the nucleus of crystallization. For the next

Table I

Cylindrical layer	r_0	Tilt angle α	Pitch of minor helix	Number of chains	Shortest distance between chains within the layer
1	5 Å	9°	5·25 Å	3	8·5 Å
2	14 Å	24°	5·65 Å	8	9·8 Å
3	23 Å	36°	6·40 Å	11	10·5 Å

cylindrical layer we put $r_0 = 14$ Å, and for the third $r_0 = 23$ Å. All minor helices are right-handed and all major helices left-handed. Table I gives the data for the protein chains in the three cylindrical layers. The values of r_0 are chosen so that there is 9 Å between the cylindrical layers, which

is compatible with the equatorial diffraction pattern. From the value of, c, and, r_0, the tilt angle, α, is calculated, and the pitch of the minor helix equals $5{\cdot}18$ Å/cos α. The number of chains per cylindrical layer is estimated to give a reasonable value for the shortest distance between two

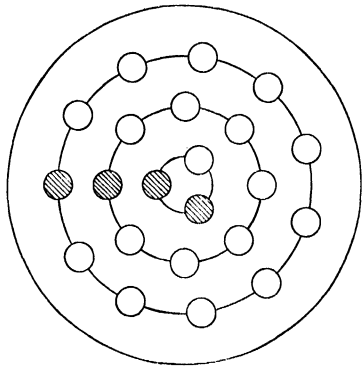


FIG. 2. Schematic drawing of the cross-section of the proposed model for the structure of α-keratin. One small circle represents one protein chain. The four protein chains shown in the three-dimensional model in Fig. 3 (a) and (b) are specially indicated.

protein chains in the same cylindrical layer. Figure 2 shows a cross-section of the model. The four chains in the three-dimensional model shown in Fig. 3 (a) and (b) are specially indicated in Fig. 2.

6. Stereochemical Consequences

The estimations of the stereochemical consequences of the deformation of the α-helical structure due to the side-chain interaction in the proposed model are as yet of a preliminary nature. Two methods have been used to estimate hydrogen-bond lengths and angles. The generalized mathematical relations for polypeptide chain helices by Low and Grenville-Wells (1953) and the building of a three-dimensional model are both based on the bond lengths and bond angles defined by Corey and Pauling (1953).

It is hereby found that for the first and second cylindrical layer the deformation of the α-helix structure is negligibly small. In the third layer, however, the hydrogen-bond length gets too long (about $3{\cdot}5$ Å), if we simply stretch the α-helix. Reasonable values are obtained, however, if residue 1 is hydrogen bonded to residue 3 instead of residue 4 as in the α-helix. In the third layer the protein chains will thus be of the $3{\cdot}0_{10}$ type of Donohue (1953) but slightly deformed. The small amount of energy needed for this deformation is probably easily supplied by the side-chain interaction. A study is in progress to investigate this further.

If a fourth layer is put on, then the pitch of the minor helix would be about 7 Å, which would require a β-structure, and then the 3·5 units per turn condition would be violated. The presence of such a layer, however, is very probable for reasons to be mentioned later.

According to the author's view the formation of the micro-fibrils of α-keratin is a crystallization in a cylindro-helical lattice of protein chains basically in α-helical configuration, and the crystallite size is limited by stereochemical factors. The diameter of the proposed model is about the same as that observed with electron microscopy.

7. Interpretation of the Meridional Low-angle Diffraction Pattern

In an earlier paper the author proposed a model for the structure of α-keratin (Swanbeck, 1961), which is similar to the one proposed here, although less detailed. The number of protein chains in the different cylindrical layers was there determined from the distribution of strong layer-lines in the low-angle diffraction pattern. Because of the rotational symmetry, the cylindrical layer with three chains will contribute to the third layer-line, and the layer with N chains to the Nth layer-line. These contributions to the different layer-lines should be off-meridional, if the microfibrils were parall packed. If they are packed in a cylindro-helical lattice, however, these contributions could very well be meridional. In that case the numbers of the layer-lines with meridional maxima indicate the rotational symmetry of the different cylindrical layers, which thus would be 3-, 8-, 11-, 16- and 19-fold. The three first will thus agree with the model derived here from the wide-angle diffraction pattern and steric considerations. As yet the author can see no explanation for the high intensity on some layer-lines in the low-angle pattern with off-meridional maxima.

8. Correlation with Other Data about Keratinization

The proposed model with three cylindrical layers of protein chains has a good stability without cross-linking with –S–S– bridges. To add one or two layers in the form of extended chains, would probably require –S–S–bridges to achieve necessary stability. It has also been shown (Ryder, 1958) that the sulphur incorporation into keratin occurs rather late in the hair or wool follicle. By degrading hair or wool two protein fractions are obtained (Alexander and Hudson, 1954). One (α-keratose) has a low sulphur content, high-molecular weight and gives an α-diffraction pattern when precipitated. The other (γ-keratose) has a high sulphur content, low molecular weight and gives a disoriented β-diffraction pattern. It is probable that the α-keratose occurs in the inner

three layers of the microfibril and that γ-keratose is situated in one or two layers outside. The inner three layers might also constitute what is usually characterized as the crystalline regions of keratin.

Fraser *et al.* (1962), have suggested a model for the structure of α-keratin consisting of two central units surrounded by nine outer units. Each unit is a three-strand segmented rope with 70 Å long, straight segments. The segments are inclined 10° with respect to the axis. No support for this model can be found in the equatorial low-angle diffraction pattern of α-keratin. This model would give rise to a strong maximum at about 20 Å corresponding to the distance between the three-strand ropes. No such maximum is found in the experimental pattern. There are also difficulties in packing three-strand ropes tightly.

9. Conclusion

The present model is derived to give the geometrically most favourable conditions for side-chain interaction of the knob-into-hole type. Information from the wide-angle diffraction pattern has been used partly for the derivation of the model, and partly as redundant information to confirm the geometrical considerations. As the distribution of strong low-order layer-lines has not been used in the derivation of the present model, the fact that the model satisfies this part of the diffraction pattern rather well can be considered as a confirmation of the model.

The present model shows a high degree of symmetry. Apart from the rotational symmetry of the individual cylindrical layer, the whole structure has a helical symmetry such that a translation of 10·36 Å along the z-axis combined with a rotation of $-19°$ brings back the same conformation.

References

Alexander, P. and Hudson, R. F. (1954). *Wool: its Chemistry and Physics.* Chapman and Hall, London.

Bear, R. S. (1943). *J. Amer. chem. Soc.* **65**, 1784.

Birbeck, M. S. C. and Mercer, E. H. (1957). *J. biophys. biochem. Cytol.* **3**, 203.

Cochran, W., Crick, F. H. C. and Vand, V. (1952). *Acta cryst.* **5**, 581.

Corey, R. B. and Pauling, L. (1963). *Proc. roy. Soc.* **B141**, 10.

Crick, F. H. C. (1952). *Nature, Lond.* **170**, 882.

Crick, F. H. C. (1953*a*). *Acta cryst.* **6**, 685.

Crick, F. H. C. (1953*b*). *Acta cryst.* **6**, 689.

Donohue, J. (1953). *Proc. nat. Acad. Sci., Wash.* **39**, 470.

Fraser, R. D. B., MacRae, T. P. and Rogers, G. E. (1962). *Nature, Lond.* **193**, 1052.

Klug, A., Crick, F. H. C. and Wyckoff, H. W. (1958). *Acta cryst.* **11**, 199.

Lang, A. R. (1956*a*). *Acta cryst.* **9**, 436.

Lang, A. R. (1956*b*), *Acta cryst.* **9**, 446.

Low, B. and Grenville-Wells, H. J. (1953). *Proc. nat. Acad. Sci., Wash.* **39**, 785.
MacArthur, I. (1943). *Nature, Lond.* **152**, 38.
Pauling, L. and Corey, R. B. (1953). *Nature, Lond.* **171**, 59.
Ramachandran, G. N. and Kartha, G. (1955). *Nature, Lond.* **176**, 593.
Ramachandran, G. N. and Sasisekharan, V. (1956). *Arch. Biochem. Biophs.* **63**, 255.
Ryder, M. L. (1958). *The Biology of Hair Growth*, p. 305. Academic Press, New York.
Swanbeck, G. (1961). *Exp. Cell Res.* **23**, 420.
Watson, J. D. and Crick, F. H. C. (1953). *Nature, Lond.*, **171**, 737.
Waugh, D. F. (1958). *Biophysical Science—A Study Program*, p. 84. Wiley, New York.

Discussion

G. N. RAMACHANDRAN: I agree with you that the cylindro-helical lattice of keratin can lead to the occurrence of fibrils with a finite diameter. In fact, we have been thinking on similar lines in connection with collagen. Dr. Sasisekharan and myself explained the equatorial pattern of collagen in terms of a cylindrical lattice. But I am not quite convinced that a side group arrangement leading to a helical packing can by itself produce a cylindro-helical lattice, since there seems to be no reason why such a phenomenon does not occur in the case of poly-L-alanine or poly-γ-methyl-L-glutamate, where also we have fairly bulky side groups.

G. SWANBECK: Preparation of synthetic polypeptides has been made under conditions which are quite different from those existing in the cells producing keratin. I think much more work should be done on fibre formation of synthetic polypeptides in different chemical environments.

Structure and Stability of Proteins and Allied Materials

P. R. Blakey and F. Happey

Department of Textile Industries, Institute of Technology, Bradford, Yorkshire, England

ABSTRACT

A study has been made of the structure and stability of synthetic polypeptides and of the α-, β- and collagenous proteins. Possible mechanisms of stabilization in these materials have been outlined; in particular the importance of the cross-linking of molecular chains, high crystallinity of the molecular fibrils and the possible influence of intercrystalline matrices on stability have been discussed. The significance of this last mechanism involving the interaction of polysaccharides in collagen and chitin is described, and its possible relationship to the appearance of native cellulose in mammalian skin tissues is reviewed.

1. Introduction

In the fibrous proteins of animal origin three major types of organized stable structure can be demonstrated by X-ray methods. They are the alpha, the beta and the collagenous forms consisting of amino acids arranged as long-chain polypeptide molecules. Each has its own characteristic in respect of its stability, but may have structural modifications varying with the physiological function which the particular molecule is required to fulfil. The α-form is found in the keratin–myosin–epidermin–fibrinogen group of tissues (Astbury, 1945) and the beta form in structures as widely differing as fibroin (Meyer and Mark, 1928; Mark and Meyer, 1929) and the degenerate fibrils of the aged nucleus pulposus (Naylor *et al.*, 1954). In collagen the apparently soluble protein is stabilized in association with polysaccharides, such as chondrotin sulphuric acid in mammalian connective tissues (Meyer and Rapport, 1951), and may form a minor protein constituent of the skeletal chitin in arthropods (Fraenkel and Rudall, 1940, 1947).

It is clear, therefore, that in such a wide field of protein specificity one of the main characteristics of the protein structure is to be found in the arrangement of side-chains, i.e. about nineteen amino acid residues are used of varying composition with side groups ranging from the hydrogen in glycine to such complexes as those in tyrosine and cystine, etc. Even in the present stage of knowledge it is clear that the organization of structure at a molecular level is only one part of a complex macromolecular arrangement and any results obtained from wide-angle X-ray

patterns may be complicated by the influence of the organized pattern of the apparent intermolecular arrangement of the protein. This is probably the case in the higher equatorial spacings on the X-ray diagram of keratin which indicates the possibility of a structural arrangement different from the classical unit cell of the α-proteins of Astbury (Astbury and Street, 1931).

Some years ago, it was clear to one of us (and his colleagues) that a new approach to the structure of the polypeptide chain might be made from a study of the synthetic polypeptides (Bamford *et al.*, 1951). This work led to a new series of structural models of polypeptides in which, in particular, the whole range of molecular transformation $\alpha \rightleftarrows \beta$ was demonstrated. From work on highly crystalline synthetic specimens, the originally postulated models for the α-polypeptide structures were replaced by those based on the later α-helix of Pauling and Corey (Pauling and Corey, 1951). The experimental findings did, however, give the general impression that the basic structures of the natural and synthetic polypeptides followed the same general pattern.

In the following, therefore, a comparative study will be made of the structural arrangements of the synthetic polypeptides, the natural polypeptides and the regenerated modifications of the latter group.

2. The α-Diagram in Proteins and Synthetic Polypeptides

In general, this diagram occurs in synthetic polypeptides mainly with large side-chains although it has been obtained in poly-L-alanine (Bamford *et al.*, 1953), in normal α-proteins of the keratin–myosin–epidermin–fibrinogen group of Astbury (1945), and in certain regenerated proteins from wool and other alpha proteins after they have been rendered soluble (Wormell and Happey, 1949). In the latter case, great care was taken to ensure that the protein was dispersed in solution in single chains and that no large aggregates of α-keratin were retained in the spinning solution.

The natural and synthetic polypeptides in the α-form give two basic types of X-ray diagram. First, those from simple synthetic polypeptides such as poly-L-alanine and poly-methyl-L-glutamate show a comparatively highly crystalline structure (Fig. 1) which can, with some reservations, be indexed on a hexagonal lattice involving only one chain per unit cell. Figure 5 shows a possible arrangement of the Pauling-Corey helix as the basis of the screw along the fibre axis. Secondly, there is the more diffuse α-diagram found amongst the synthetic co-polymers (Fig. 4), the fibrous α-protein molecules (Figs. 2 and 7), and, in certain cases, the synthetic polymers made from identical monomers

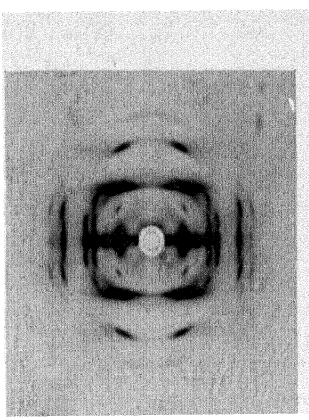


FIG. 1. X-Ray diagram of fibres of α-poly-L-alanine. Cu $K\alpha$ radiation. Camera radius ≈ 3 cm.

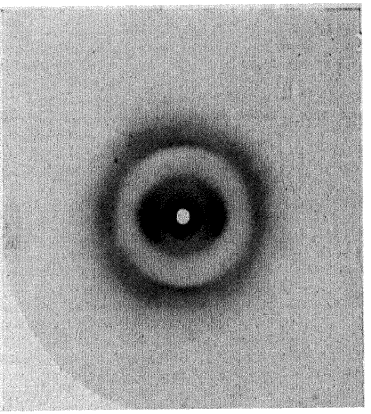


FIG. 2. X-Ray diffraction photograph of regenerated α-keratose. Cu $K\alpha$ radiation. $d = 3$ cm.

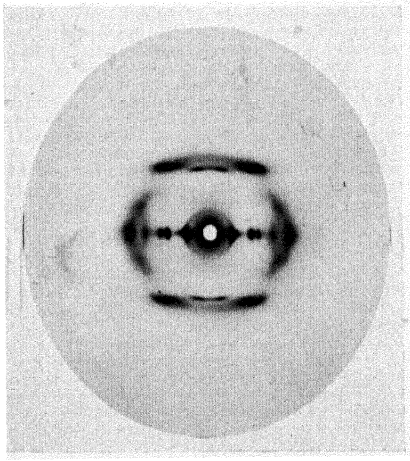


FIG. 3. X-Ray diagram of poly-L-glutamic benzyl ester.

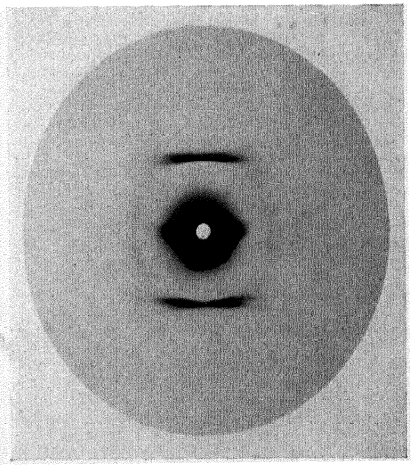


FIG. 4. X-Ray diagram of α-co-polymer of 1:1 DL-β-phenylalanine and DL-leucine. $d = 3$ cm.

The figures reproduced here are reduced from the original photographs. The dimensions given are consequently approximate.

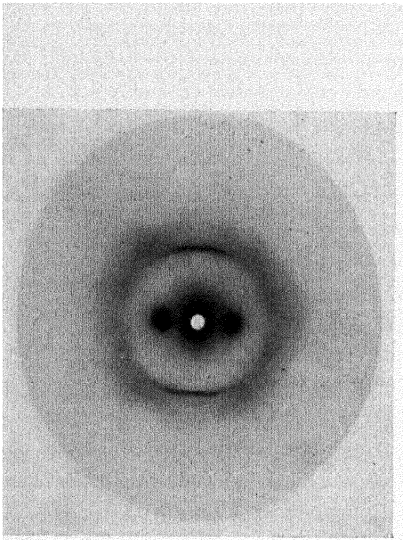


FIG. 7. X-Ray diagram of porcupine quill tip showing partially resolved polar arcs of 5·3 Å layer line. Cu $K\alpha$ radiation. $d = 3$ cm.

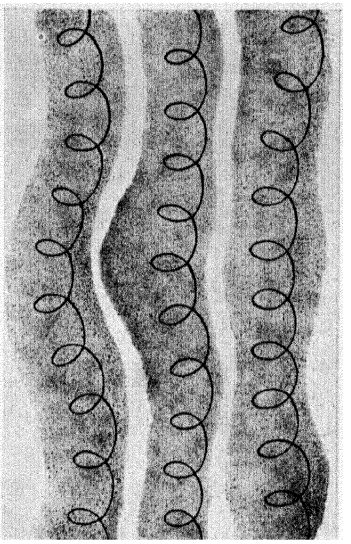


FIG. 8. Diagrammatic representation of α-chains showing secondary folding due to the packing of side chains of variable size and chemical constitution.

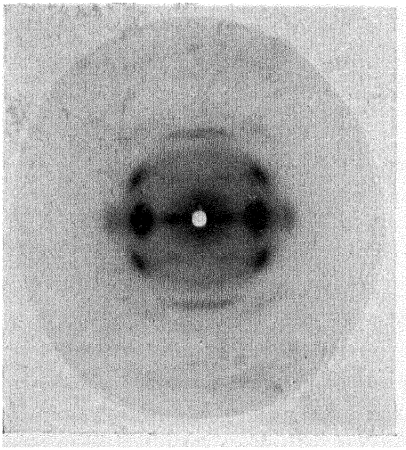


FIG. 9. X-Ray diffraction photograph of β-fibroin from silk. Cu $K\alpha$ radiation. $d = 3$ cm.

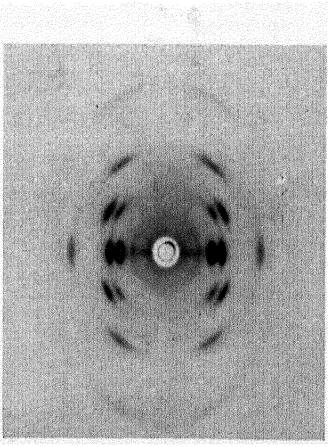


FIG. 10. X-Ray diagram of β-poly-L-alanine. Cu $K\alpha$ radiation. Camera radius ≈ 3 cm.

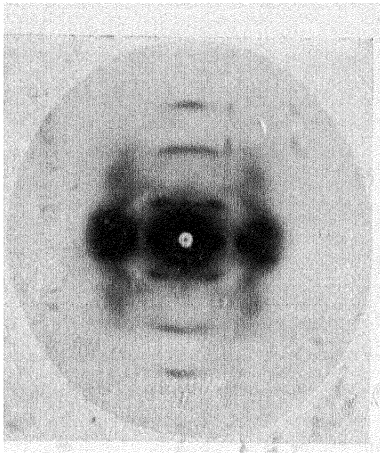


FIG. 12. X-Ray diffraction photograph of collagen from rat-tail tendon. Cu $K\alpha$ radiation. $d = 3$ cm.

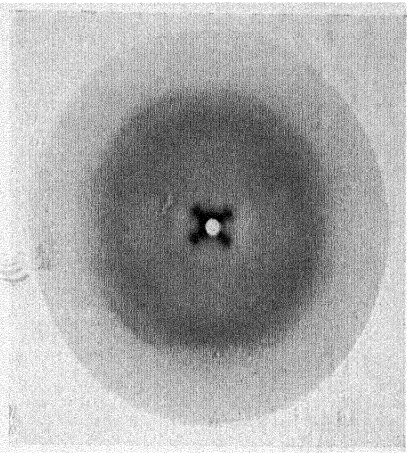


FIG. 13. X-Ray diffraction photograph of structure of annulus fibrosis of human intervertebral disc showing cross-structure of fibrils of the disc wall. Cu $K\alpha$ radiation. $d = 3$ cm.

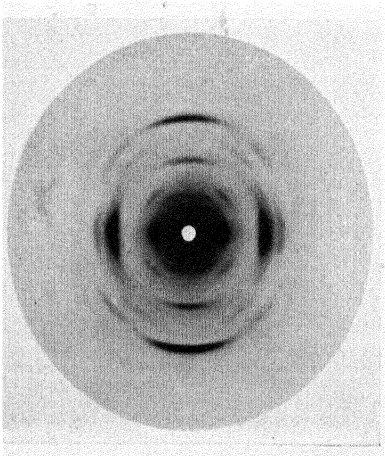


FIG. 14. X-Ray diffraction photograph of chitin of crab tendon. Cu $K\alpha$ radiation. $d = 3$ cm.

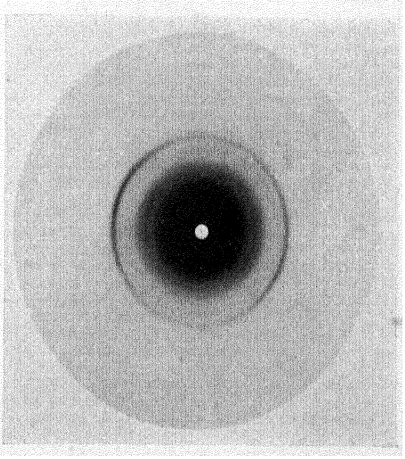


FIG. 15. X-Ray diffraction photograph of cellulose fibrils extracted from epidermal tissue and showing an angular dispersion of 75°. Direction of fibre axis marked with an arrow. Cu $K\alpha$ radiation. $d = 3$ cm.

of large molecular weight, e.g. poly-L-glutamic benzyl ester (Fig. 3). The fibre diagrams from these specimens are more amorphous in character and the polar reflections consist mainly of strong reflections ~ 5·2 Å to ~ 5·3 Å and ~ 1·5 Å; and excluding the latter benzyl ester a band of diffuse equatorial scatter varying in mean spacing from 9 Å to 12 Å.

3. Chain Displacement and the Pauling and Corey Helix

It has been shown that the intense 5·4 Å layer line on the X-ray diagram of the polyhomopeptides (i.e. those containing only one isomorphic type of amino acid residue) is associated with a single turn of the helix. This structure forms what might be considered as an idealized crystalline polypeptide. Such forms have been shown to occur in poly-L-alanine and poly-L-glutamic methyl ester. Chemically the variant applicable in other polypeptides is a change in the side-chains of the amino acid residues.

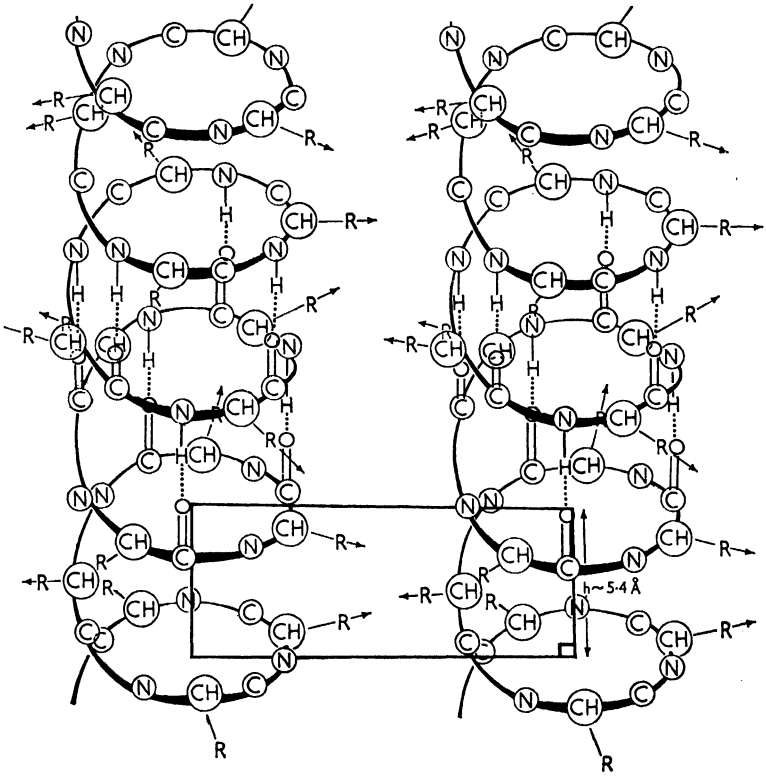


Fig. 5. The orthogonal arrangement of the Pauling-Corey helix which would represent the hexagonal lattice postulated for the polyhomopeptides of poly-L-alanine and poly-L-glutamic methyl ester.

This can modify the regular arrangement of the side-chains and may cause a displacement of the individual molecular chains from a mutually orthogonal packing (Fig. 5) to a staggered arrangement along the fibre axis (Blakey and Happey, 1961) (Fig. 6).

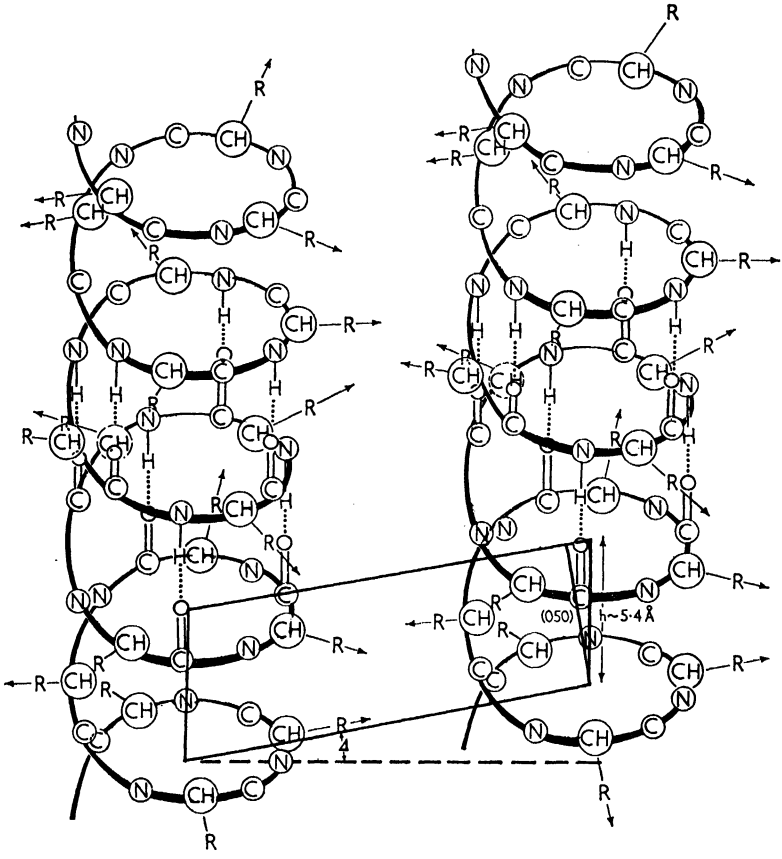


FIG. 6. The staggered arrangement of the Pauling-Corey helix possibly giving rise to the α-diagram of the synthetic co-polymers, etc. The hexagonal symmetry of the structure of the poly-L-alanine type is exhibited by the chain displacement and the (0*k*0) planes are no longer orthogonal.

The main organized X-ray scatter on the X-ray diagram of a polypeptide has been associated with the peptide backbone, and in particular an intense scatter is present on the (050) layer line associated with the molecular repeat of one turn of the helix. This is likely to obtain even when

(a) the individual chains are mutually displaced along the fibre axis, and

(b) the side groups of the amino acid residues of differing molecular weight and constitution inhibit the true hexagonal form of the unit cell in the *ac*-plane. (The fibre axis being the *b*-axis (Blakey and Happey, 1961).)

If the latter transformation is applied to the highly crystalline lattice structure, then the reflections on the equator of the X-ray diagram showing hexagonal symmetry will be replaced by a diffuse band giving the mean distance of approach of the chains. Thus, the more variable the side-chains the more diffuse will this reflection become. This variation can be seen clearly from a comparison of X-ray diagrams of polyhomopeptides (Fig. 1) mentioned above and those of the synthetic D- and L-polypeptides and the co-polymers (Fig. 4).

If displacement of molecules occurs in adjacent sheets parallel to the fibre axis (*b*) a further smearing out of the highly crystalline diagram may occur. The hexagonal symmetry will be distorted, concomitant with this a new series of crystallographic planes would be formed and the (0*k*0) planes would give reflections if $(\Delta + \phi) > \theta$ and be capable of resolution when $\Delta > \theta$ and $\phi < (\Delta - \theta)$ about the meridian.

Thus, Δ' (max.) the angle between (0*k*0) planes and that normal to the fibre axis would be approximately determined as shown on Fig. 6, where

$$\tan \Delta' = \frac{h}{2S} \tag{1}$$

that is, with a displacement of half the pitch of the helix (*h*) along the fibre axis, *S* being the mean separation of the chains.

For the purpose of this discussion Eq. (1) is an adequate approximation to the equation (Happey, 1954):

$$\cos^2 \Delta = \frac{\sin^2 \beta - \cos^2 \alpha - \cos^2 \gamma + 2 \cos \alpha \cos \beta \cos \gamma}{\sin^2 \beta} \tag{2}$$

Thus, the polar reflection on the fifth layer line, assuming a fibre repeat of 27 Å would be from the (050) planes and may in highly oriented samples be resolved into a doublet about the meridian (see Fig. 4). This spacing could vary between $d\ (0k0) = h \cos(\tan^{-1} h/2S)$ as a minimum value up to a value less than *h* which would be obtained when $\Delta \rightarrow 0$, i.e. in the systems of hexagonal symmetry found in the polyhomopeptides of L-alanine and L-methyl esters.

The X-ray diagrams of the α-synthetic polypeptides whether highly crystalline or otherwise show a polar reflection of $\sim 1{\cdot}5$ Å from the (0 18 0) planes; this is determined by the repeat of one amino acid residue along the axis of the helical chain. In the hexagonal polyhomopeptides this arc would be a single reflection 0 18 0. In the polypeptides with an inhibited hexagonal structure the arc should be capable of resolu-

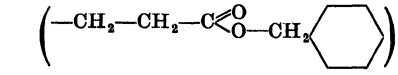
tion but modified slightly in value due to the non-orthogonal arrangement of the ($0k0$) planes. The $\sim 1{\cdot}5$ Å polar arc can be readily identified in porcupine quill, but, owing to the poor crystallinity of keratin and related fibrous α-proteins, this polar arc is difficult to identify even when the fibres have been inclined at the Bragg angle to the X-ray beam corresponding to the 1·5 Å spacing.

4. Application of the Displacement Theory

There are four particular cases in decreasing order of fibrous crystallization to which the displacement theory can be applied:

(a) the poly-L-glutamic benzyl ester type of structure;
(b) the synthetic co-polymers;
(c) the highly crystalline fibrous proteins (porcupine quill);
(d) the normal α-proteins (k.m.e.f. group of proteins).

First, it is assumed that the polyhomopeptides (e.g. poly-L-alanine) have a value of $\Delta = 0$; hence, if $0k0$ appears as a polar arc it is of value of the fifth order of the fibre repeat of 27 Å, i.e. $\sim 5{\cdot}4$ Å. In both cases where this structure is present the 050 arc is weak or absent, and the strong reflection is in general 150, etc. In (a), therefore, it might have been expected that the 050 reflection, if present, would be a true polar arc at $\sim 5{\cdot}4$ Å, from such a highly oriented polyhomopeptide. As can be seen from Fig. 3, this is not the case, and in the X-ray diagrams of these fibres the 5·25 Å polar arc can be resolved into a doublet about the meridian. It would appear, therefore, that the large side-chains



may influence the mutual displacement of the spiral molecules parallel to the fibre axis. A more detailed analysis of other possible aspects of this structure will be given later in the discussion.

The co-polymers of (b) (Fig. 4) are generally not very crystalline and show an intense and diffuse reflection at 10–12 Å on the equator of their X-ray diagrams and, as in (a), polar reflections $\sim 5{\cdot}25$ Å and $\sim 1{\cdot}5$ Å. In these polypeptides the side-chains cannot be symmetrical and perpendicular to the helix for obvious reasons. Parallel to it, the side-chain displacement, and also Δ, may vary from polymer to polymer, and this may be reflected in the small variations in spacing of the non-orthogonal (050) spacings found, particularly in the synthetic polypeptides.

The spacing variation in the polar reflections has been one difficulty in the comparison between the synthetic polypeptides, in which d (050) $\approx 5{\cdot}25$ Å, and the proteins where d (050) is 5·15 Å. From this

point of view it is of interest in (c) to consider the polar arcs in the X-ray diagram of porcupine quill (Fig. 7). The reflections can be identified as a polar arc of 5·15 Å (unresolved since $\Delta + \phi > \theta$) spreading to a doublet of 4·9 Å on the same layer line. On stretching the specimen of quill to its Hooke's Law limit the 5·15 Å spacing increases to 5·25 Å (approximately that found in the synthetic polypeptides), with a consequent increase in the 4·9 Å spacing. In X-ray diagrams of unstretched porcupine quill, where high resolution had not been obtained, the (050) spacing was approximately 5·05 Å (Happey, 1932). This diffuse reflection, therefore, may on these grounds comprise a superposition of the unresolved 050 reflections where d (050) may vary from 4·9 Å–5·15 Å (mean value 5·05 Å), and, on stretching, increase to 5·0 Å–5·25 Å (mean value 5·15 Å).

It is, therefore, possible that the spacing variations between 5·15 Å and 4·9 Å may be related to variations in the angle due to chain displacement, and possibly to variations in the distance of side-chain separation causing variations in the parameter S of Eq. (1) in different components of the structure. On the equator of the X-ray diagram of porcupine quill the medium angle pattern shows reflections at 10·5 Å (v.s.), 9·22 Å (v.s.) and 7·5 Å (weak). It is therefore possible that the polypeptide structure is multiphase in the sense that discrete crystalline aggregates of differing average residue weight are present, which would give rise to a further complication in the interpretation of the composite X-ray diagram. In such a case, the minimum value of the polar (050) reflection could be $5{\cdot}25 \cos [\tan^{-1}(2{\cdot}6/7{\cdot}6)] = 4{\cdot}95$ Å. In the proteins the helical pitch is $\sim 5{\cdot}25$ Å, but Hooke's Law stretching can increase this to $\sim 5{\cdot}35$ Å of the synthetic polypeptides. It must be realized that the foregoing discussion assumes that the polar reflections on the fifth layer line are from ($0k0$) planes and do not directly involve reflections from (hkl) planes, with d_h and d_l forming part of a larger unit of repeat perpendicular to the fibre axis. This aspect of the work will be discussed later, as the spacings reported originally as ~ 27 Å, etc., on the equator of the wool diagram have been identified with a large-scale packing of bundles of polypeptide chains in keratin (Fraser and MacRae, 1958).

In the k.m.e.f. group of proteins of lower crystallinity than porcupine quill, the previous arguments may equally well apply. On the X-ray diagram of keratin under tension, the diffuse polar arc is 5·15 Å in spacing, and this may be as low as 5·05 Å in unstretched protein fibres. In consequence, the diagram may be considered as almost amorphous in character, arising from the close packing of cylinders of variable and probably non-circular cross-section. These cylinders have a repeat of 5·25 Å along the axis, but may be staggered with maximum displacement of 5·25/2 Å to give a non-orthogonal arrangement of the ($0k0$) planes.

5. Comparison of Side-chain Configuration in Proteins and Synthetic Polypeptides

It is clear, therefore, that a reasonable comparison can be made between the natural and synthetic polypeptides when the former have been extended to their Hooke's Law limit, but a discrepancy occurs in the unstretched proteins which needs consideration. If the chemical constitution of the natural and synthetic materials is considered, then some points become fairly clear. The polyhomopeptides can form a true α-helical structure and adjacent chains can pack in true crystalline form. Where variations are made in this side-chain arrangement then accurate packing sideways cannot occur and some lattice distortions must exist in the crystalline pattern. Thus, a falling away in crystalline definition occurs. In proteins the change is most dramatic since the chains contain, in the main, unknown arrangements of side groups of widely varying size and chemical constitution. Thus in the interstices between the backbones of the polypeptide chains, widely differing electronic environments are present (Fig. 8). It is obvious that the spatial arrangements must be determined mainly by the larger side-chains, particularly where they are restricted in orientation. In consequence, it is possible that the backbone of the helix may not run perfectly straight along the fibre axis, but oscillate about it in a manner determined primarily by the packing of the side-chains. It is, therefore, likely that one factor which determines the decrease in fibre repeat along the axis of the molecule may be the secondary folding required to accommodate spatial variations of the side-chains from a truly hexagonal structure.

From the point of view of packing side-chains, it is of interest to consider the poly-benzyl ester, which is a polyhomopeptide, but which does not give a simple hexagonal lattice structure. The X-ray pattern obtained from the polymer can be varied by the use of differing solvents in its preparation—particularly as a film. Here a planar orientation can be obtained suggesting that under differing forces the side-chains of the polymer can orientate in different ways, even though the backbone helix remains unmodified. This may be of considerable biological interest in the sense that a synthetic polymer based on a uniaxial skeletal molecule can develop biaxial properties from the differing behaviour of similar side-chains under differing forces or in a modified environment. This could explain the biaxial orientation of an α-protein described by Rudall (1954).

6. The β-Proteins

In this series of structures the X-ray diagrams of β-proteins are generally characterized by a fibre repeat of 6·7 Å comprising two amino acid residues in a fully extended molecule. Perpendicular to this the

lattice repeats are 4·65 Å (hydrogen bonding distance) and a repeat mutually perpendicular to those determined by the lengths of the side-chains of the amino acid residues. In the fibre direction, this repeat is less than that calculated from normal bond lengths and angles.

In the beta synthetic polypeptides (Fig. 10), and in silk (Fig. 9), the fibre repeat compares with that calculated as about 7·2 Å. This suggests that in the crystalline parts of these fibres the molecules are fully extended and exclude a secondary molecular folding. Further, in polypeptides with small side-chains the planes associated with hydrogen bonding have a spacing of about 4·3 Å. Thus, in the case of silk (with side-chains H, CH_3, CH_2OH mainly in the crystalline phase) and poly-L-alanine (side-chains CH_3) good and uniform packing of side-chains is obtained giving an enhanced stability of these structures over the β-synthetic polypeptides with large side-chains and the normal β-proteins.

This does, therefore, in a general sense, tend to confirm that where good crystalline packing of the side-chains can occur in proteins, either α or β, the cell dimensions correspond reasonably well with the calculated lattice dimensions. Where this is not the case, the chain has to distort from a straight polymer to allow for side-chain interaction causing a shortening of the fibre repeat.

7. The Cross-β Structure

A further and less usual form of protein has been identified in the "cross-β" structure, which may occur in the folding of certain labile proteins, usually from the α-configuration. This was first described by Rudall (1946, 1952), later by Happey and Wormell (1946) in casein, and Mercer and workers with other labile α-proteins (Mercer, 1949). This protein form was defined mainly from its X-ray diagram, which showed a strong polar arc of $\sim$ 4·65 Å, an equatorial spacing of $\sim$ 10 Å and, in certain cases, a polar arc of about 4·2 Å. In most cases, however, there was also a weak 5·15 Å arc showing the presence also of an oriented normal α-form of the polypeptide molecule.

Rudall suggested that the cross-β was formed from fully extended β-protein molecules which folded on shrinking with intrachain hydrogen bonding between the folds, perpendicular to the fibre axis (Fig. 11). This would require the polar arc at 4·65 Å to represent a repeat of 4·65 Å along the fibre axis. It would also be consistent with a repeat of $\sim$ 10 Å perpendicular to the fibre axis and also one of $\sim$ 3·5 Å mutually perpendicular to these two directions representing the length of one amino acid residue.

From a comprehensive study of the dispersion of the 4·65 Å arc and that of the equatorial spacing of $\sim$ 10 Å, it appears that even with

approximate estimates of the limits of the arcs concerned, the semi-angle of dispersion χ_P of the polar arc is found to be considerably greater than that of the equatorial arc χ_E.

The values found were (Blakey, 1958): $\chi_P = 38°$, $\chi_E = 27°$. From a comparison of dispersions using Happey's equations for non-orthogonal planes it can be shown that $(\Delta + \phi) \approx 39°$, $\Delta \approx 12°$.

Although these measurements were approximate, from the usual estimation of the arc extremities, the differences were sufficiently great to suggest that the 4·65 Å reflections were not from orthogonal planes perpendicular to the fibre axis. This would need to be the case if the 4·65 Å spacing were given by the hydrogen bonding between β-folds perpendicular to the folds in the protein molecule.

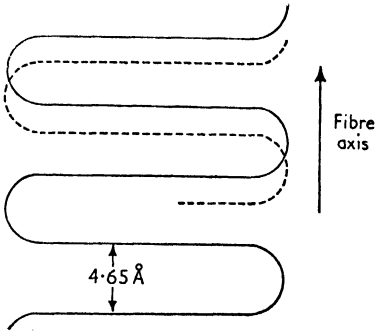


FIG. 11. The cross-β configuration postulated by Rudall showing the folding of the β-chains perpendicular to the fibre axis with an interfold separation of 4·65 Å.

Other difficulties arise in such a molecular fold if the present values of the mean molecular weights of the keratin molecules of ~ 30,000 are acceptable. To give X-ray reflections on a fibre diagram it is generally accepted that the crystallite dimensions need to be between 50 Å and 100 Å. Hence, such molecules in this type of configuration, with a mean residue weight of ~ 110 in a distance of 3·5 Å along the molecule, would no longer be fibrous in character. In point of fact the molecule could well be greater perpendicular to its axis than along it. Further, to draw out such a molecule from its folded form to a fully extended β-configuration the extension could well be $100/4{\cdot}5 \approx 20$ times (i.e. 2000%).

The cross-β pattern has been obtained from labile proteins of the k.m.e.f. group obtained from stable keratins in a variety of ways, and also in normally occurring labile proteins. In most cases, however, the final necessity to obtain this form from the normal α-form was to heat the specimen in water at a temperature just below boiling point (e.g. 90°C). Where no readjustment of the individual helices can occur,

as in cross-linked polypeptides, then no transformation to cross-β has been found to occur. If higher temperatures (100°C) are used then denaturation of the labile protein into the normally disoriented insoluble β-form usually takes place. In some of our experiments the transformation has been carried out from labile α-structure to a cross-β form with the fibres held at constant length (Blakey, 1958). Added to this transformation some degree of disorientation was also found to occur.

8. The Small Angle and Large Angle X-Ray Patterns from Polypeptides

In this simple representation of the protein helix, it is necessary to consider the possibility that the unit cell of the structure contains more than one chain. A detailed analysis of such a possibility was given by Astbury and Street (1931), Pauling and Corey (1951) and Crick (1953) in the case of keratin. More recently, Fraser and MacRae (1958) have shown, using small-angle reflections, that the larger patterns found in keratin may possibly be due to the packing of cells of approximately 50 Å–100 Å in diameter, perpendicular to the fibre axis. Hence, it is possible, that the pseudo-crystalline pattern of the medium-angle scatter perpendicular to the fibre axis is due to individual chains and the smaller angle scatter; e.g. the strong equatorial spacing at 27 Å (used by Astbury as the *b*-axis of his unit cell "*a*" dimension) may be due to the pseudo-hexagonal packing of the molecular bundles of fibrils perpendicular to this fibre axis.

This later conception has been developed further by Fraser (1961), who has postulated that the X-ray diagrams of porcupine quill can be possibly associated with microfibrils containing eleven subfibrils, each subfibril being a rope of three α-helices. Critical studies of this ingenious model have been made by Sikorski and others (c.f. Johnson and Sikorski, 1962).

9. Regenerated Proteins

In these laboratories, studies on the regeneration of protein fibres from keratin and other related fibres have been made. From the work it is clear that oriented α-protein can be regenerated showing an X-ray diagram not substantially different from that of the parent fibre (Happey, 1955) (Fig. 8).

(a) *The secondary spirals in keratin*

It has been shown clearly that the fibre repeat in proteins is lower than that expected from a calculation using known atomic distances and bond angles. To explain this apparent anomaly, in addition to the contraction due to side-chain bulkiness, secondary helical windings of the parent

chains have been postulated in keratin, and a series of ingenious models of coiled-coil formations have been adduced (Pauling and Corey, 1953). Since the introduction of these concepts the coiled forms are now a well-established basis for the elucidation of many problems in protein synthesis, particularly with reference to protein specificity. In highly crystalline proteins these spirals have been clearly demonstrated by Fourier techniques, both in individual α-protein chains and in specific aggregates of such molecules. In many cases the secondary folds have been associated solely with this secondary mutual spiralling of chains (i.e. generally referred to as "coiled-coils"). From the work on regenerated proteins, however, such a specificity can hardly be acceptable from a random collection of aligned protein chains; yet in this case, a very similar X-ray diagram to that of the parent keratin is obtained. In particular the two specimens have the same fibre repeat. If, therefore, the two forms of packing, specific and random, produce the same lattice contraction parallel to the fibre axis then the side-chain interactions have probably the same ultimate effect in the general formation of the structure. It is, therefore, possible that a random spiral may be formed due to the asymmetric packing of the side-chains of the molecules during fibre regenerations from solution. Thus specific interaction during growth may produce the same effect of apparent lattice contraction as does random interaction on regeneration.

Thus, in considering the apparent decrease in fibre repeat in polypeptide molecules three parameters may be involved as follows.

(a) The chains may be fully extended but non-orthogonally packed.
(b) The variable side-chain spacings may cause a secondary folding without any specific spiralling occurring.
(c) A secondary spiralling may be imposed on the molecular alignment which itself may be specific or may be of a random character.

10. The Collagenous Fibres

In the collagen of connective tissues, a series of comparatively crystalline structures is found, which are capable of high orientation, e.g. rat-tail tendon (Fig. 12). Collagen may, however, be rendered soluble very readily and the regenerated structures obtained give X-ray diagrams very similar to those from the parent material. It is, therefore, necessary to look for an additional mechanism to explain the stability of these proteins in aqueous media. This appears to arise from an interaction between the constituent mucopolysaccharides and the proteins of collagen. In certain cases, this interaction destroys the normal function of the material (the degeneration of the nucleus pulposus of the intervertebral disc). In other cases, without it the connective tissue would not retain its fibrous character in its aqueous environment (tendons, etc.). Any X-ray diagram

of collagen can apparently be interpreted as arising from a helical structure with a 3-fold screw axis, with an amino acid residue repeat of 2·86 Å (Ramachandran and Kartha, 1954). Such a similar spiral structure has been identified in one form of the synthetic polypeptide, poly-glycine (Crick and Rich, 1955). This latter structure, as might be expected, is very unstable and readily transforms into the β-form to give a highly crystalline and intractable hydrogen-bonded structure. This change of structure has not been obtained with collagen and it can be considered as an extended and stable helical polypeptide chain. It is, however, possible to show that the collagen molecule is capable of a reversible extension of about 7–10%, but no molecular transformation has been made in this molecule by stretching comparable with the transformation in the k.m.e.f. group of proteins.

It is significant, therefore, that in the normal collagen and in its regenerated protein component, gelatin, the X-ray diagrams are very similar. If, therefore, the polysaccharides are associated with the main complex they do not appear to form a part of the structure responsible for the X-ray diagram, and hence are probably extra-crystalline in character. It is likely, therefore, that the mucopolysaccharides form part of the stabilizing pellicle around the aligned crystalline protein, which itself determines the fibrous character of the collagenous mucoprotein.

(a) *Formation of "stabilizing" fibrils in the human intervertebral disc*

In the human intervertebral disc the stabilization of the protein molecules forms a degenerative process in ageing. The annulus fibrosus is at all times a stable collagenous structure formed of interlayering sheets crossed at an angle and giving an X-ray diagram as shown in Fig. 13 (Naylor *et al.*, 1954). These comparatively inelastic sheets are rendered elastic by allowing the sheets to shear and cause a change in the angle between the sheets themselves (Hall *et al.*, 1957). It is, however, in the nucleus pulposus that the greatest changes of structure occur on ageing.

In early life the nucleus pulposus is a viscous aqueous medium of soluble proteins and polysaccharides, and with age a precipitation of

(a) oriented collagenous fibres takes place which tends to replace the viscous medium, and
(b) there is a clear formation of a disoriented β-protein component in the nucleus (Blakey *et al.*, 1962).

Thus, the disc which ideally acts as a shock absorber in early life forms in later decades a structure which is tending to stabilize into an intractable mass of collagen fibres, and is further stabilized by a disoriented

β-component which replaces the viscous fluid of the nucleus pulposus of the early decades.

(b) *Crystalline polysaccharides in tissues*

In at least two cases in the animal, as distinct from the vegetable, world, crystalline polysaccharides have been identified. One, well-known, is in the arthropods, where the skeleton is composed mainly of the hardened chitin whose main crystalline component is poly-*N*-acetyl-glucosamine (Fig. 14).

The second case is rare, but cellulose has been found in certain tunicia and in some mammalian epidermal tissues (Hall *et al.*, 1958). This polysaccharide is in the native form and in the latter case has a crossed helical structure which can be identified (Fig. 15) by X-rays, and its spiral arrangement further demonstrated by the electron microscope (Hall *et al.*, 1958).

(c) *Protein stabilization*

In general, it would appear that a single protein chain can readily be taken into solution provided its side-chains have sufficient hydrophilic groups to retain the molecule in solution. In the α-proteins this solubility is not inhibited by the peptide CO···HN hydrogen bonds, as they are satisfied internally in the chain. Using these chains as the parent materials for the protein content of animal tissue it is necessary to see how in an aqueous environment they may be rendered stable. In all fibrous structures, in the first instance, it is essential to have aligned molecules. When this has been achieved there are then possibly three ways in which a stable structure can be achieved.

(i) The chains may be cross-linked.

(ii) The stabilization may be achieved by high crystallinity of the resulting structure and a transformation from a soluble to an insoluble protein. This is probably achieved spontaneously in the regeneration of silk (Happey and Hyde, 1952).

(iii) The formation of an inter-crystalline component of high stability.

Keratin is probably the main example of stabilization by the first method in which cystine forms the main cross-linking agent. Here, however, it does appear that there is also an amorphous inter-crystalline matrix richer probably in cystine than is the oriented crystalline component; thus enhancing the general stability of the proteins. In the membranes of the wool cells a β-component of protein is also present, and hence an increased crystallinity is produced which gives rise to additional stability of the complex by method (ii).

Where α-protein is required to act as a labile molecule, then the cystine content may be reduced, e.g. in muscles, etc., or sulphur may be

present as the cysteine form SH, as in the unkeratinized wool cells in the follicle. Work is at present being carried out in Bradford on the problem and will be described at a later date. Such a structure of stabilized keratin is borne out by the work of Fraser and his co-workers in their study of the particular distribution of crystallites across the wool fibre from low-angle X-ray scatter measurements. The high-angle X-ray diffraction pattern of keratin gives only a lead on the inter-molecular packing of the chains which are probably staggered to accommodate the side-chain interactions. Where a cystine link is formed between chains, the chains are mutually fixed at that point. Where an intra-chain cystine link is formed, the helix is fixed over one turn and normal extension is inhibited.

High crystallinity generally applies in the stabilization of silk which has been described previously (Happey and Hyde, 1952). It also plays some part in the stabilization of porcupine quill and the stiffer keratins such as lions' whiskers (Blakey, 1958). Here the crystallinity of the protein is much more evident and the X-ray diagrams of such materials may probably be interpreted as composed of multiphase crystalline components.

The importance of (iii) in the stabilization of keratin has been mentioned, but its importance is probably more manifest in the case of collagen. The highly oriented phase of this material may be readily soluble as in gelatin and it is the inter-crystalline matrix which is mainly responsible for its stability. This is a complex of mucopolysaccharides and mucoproteins, and it is the interaction of these two materials which, in the main, produces stable collagen.

Finally, it has been shown that intercellular polysaccharides in the skin may develop as cellulose, and this material in its native form has been identified both chemically and by X-rays as a component forming a helical structure around proteins of a collagenous type in mammalian skin tissues. Thus, it would appear that polysaccharides appear as stabilizing agents by method (iii)—non-crystalline in the mucopolysaccharides of chondroitin sulphate, etc.–crystalline poly-glucosamine as chitin in the arthropods—and as native cellulose in some tunicia and in certain mammalian epidermal tissues.

Acknowledgements

The authors wish to thank their present and past colleagues of the Textile Department for their collaboration; and Mr. A. Naylor, Orthopaedic Surgeon to the Bradford "A" group hospitals, for his co-operation in the work on the intervertebral disc. Figures 1 and 10 are by kind permission of my former colleagues (F.H.) at Courtaulds Ltd. The work forms part of a programme on fibre structure financed mainly by the Wool Textile Research Council, the Bradford Education Authority and the Bradford "A" and "B" Group Hospital Executive Council.

References

Astbury, W. T. (1945). *Nature, Lond.* **155**, 667.
Astbury, W. T. and Street, A. (1931). *Phil. Trans.* **A 230**, 75.
Bamford, C. H., Hanby, W. E. and Happey, F. (1951). *Proc. roy. Soc.* **A 205**, 30.
Bamford, C. H., Brown, L., Elliott, A., Hanby, W. E. and Trotter, I. F. (1953). *Nature, Lond.* **171**, 1149.
Blakey, P. R. (1958). M.Sc. Thesis, University of London.
Blakey, P. R. and Happey, F. (1961). In *Structure de la Laine.* Institute Textile de France, July, 1961.
Blakey, P. R., Happey, F., Naylor, A. and Turner, R. L. (1962). *Nature, Lond.* **195**, 73.
Crick, F. H. C. (1953). *Acta cryst.* **6**, 685, 689.
Crick, F. H. C. and Rich, A. (1955). *Nature, Lond.* **176**, 780.
Fraenkel, G. and Rudall, K. M. (1940). *Proc. roy. Soc.* **B 129**, 1.
Fraenkel, G. and Rudall, K. M. (1947). *Proc. roy. Soc.* **B 134**, 111.
Fraser, R. D. B. (1961). In *Structure de la Laine.* Institut Textile de France, July, 1961).
Fraser, R. D. B. and MacRae, T. P. (1958). *Biochim. biophys. Acta* **29**, 229.
Hall, D. A., Lloyd, P. F., Happey, F., Horton, W. G. and Naylor, A. (1957). *Nature, Lond.* **179**, 1078.
Hall, D. A., Lloyd, P. F., Saxl, H. and Happey, F. (1958). *Nature, Lond.* **181**, 470.
Happey, F. (1932). Ph.D. Thesis, University of Leeds.
Happey, F. (1954). *X-ray Diffraction by Polycrystalline Materials*, p. 486. Institute of Physics.
Happey, F. (1955). *Proceedings of International Wool Textile Research Conference, Australia*, Vol. E, p. 431.
Happey, F. and Hyde, A. J. (1952). *Nature, Lond.* **169**, 921.
Happey, F. and Wormell, R. L. (1946). Symposium: "Fibrous Proteins", p. 160. Society of Dyers and Colourists, Bradford.
Johnson, D. J. and Sikorski. J. (1962). *Nature, Lond.* **194**, 31.
Mark, H. and Meyer, K. H. (1929). *Z. phys. Chem.* **B 4**, 115.
Mercer, E. H. (1949). *J. Text. Inst.* **40**, 1640; *Biochim. biophys. Acta* **3**, 101.
Meyer, K. H. and Mark, H. (1928). *Ber. dtsch. chem. Ges.* **61**, 1932.
Meyer, K. and Rapport, M. M. (1951). *Science* **113**, 596.
Naylor, A., Happey, F. and MacRae, T. P. (1954). *Brit. med. J.* **2**, 570.
Pauling, L. and Corey, R. B. (1951). *Proc. nat. Acad. Sci., Wash.* **37**, 241, 729.
Pauling, L. and Corey, R. B. (1953). *Proc. nat. Acad. Sci., Wash.* **39**, 253.
Ramachandran, G. N. and Kartha, G. (1954). *Nature, Lond.* **174**, 269.
Rudall, K. M. (1946). Symposium: "Fibrous Proteins", p. 15. Society of Dyers and Colourists, Bradford.
Rudall, K. M. (1952). *Advanc. Protein Chem.* **7**, 255.
Rudall, K. M. (1954). *Trans. Faraday Soc.* **50**, 1014.
Wormell, R. L. and Happey, F. (1949). *Nature, Lond.* **163**, 18.

Discussion

G. SWANBECK: I want to point out that the ratio between the 5 Å layer line spacing and the 1·5 Å layer line spacing should be 3·6 if the alpha helices are packed in parallel and 3·5 if the helices are twisted like the 3-strand cable or in the form of a cylindro-helical lattice. In the case of α-keratin, accurate measurements with a diffractometer show this ratio to be exactly 3·5, which is in agreement with the latter structure.

F. HAPPEY: It is quite interesting. At some stage it has got to be explained whether this is due to the secondary helix or not.

G. SWANBECK: It would be interesting to measure the 1·5 Å spacing with the fibre stretched to give a 5·4 Å spacing, instead of 5·15 Å.

F. HAPPEY: It would appear very difficult to determine accurately the 3rd and 18th layer lines from the unresolved reflections of about 5·1 and about 1·5 Å in wool keratin. The suggestion, however, is interesting and at some stage it has got to be explained whether the contraction is due to random packing or a secondary helix.

From a staggered general structure, even with fully aligned and extended chains, one would not get a 5·4 Å meridional reflection, but rather an unresolved doublet at about 5·25 Å. It is well known (I. MacArthur, (1943). *Nature, Lond.* **152,** 38; W. T. Astbury, E. Beighton and K. D. Parker (1959). *Biophys. biochim Acta* **35,** 17) that the 5·15 → 5·25 transformation is accompanied by a similar increase in the ~ 1·5 Å polar arc in porcupine quill tip. In our model the maximum value of the polar arc is approximately 5·25 Å, and this polar arc would appear from its spread to be an unresolved doublet. If such is the case, then this would give rise to a true helical repeat of approximately 5·4 Å. The main comparison has been drawn between the 5·25–5·3 Å spacing of the synthetic copolymer (fibre repeat here 5·4 Å) and the ~ 5·25 Å spacing of the unresolved polar arc in the stretched porcupine quill.

Stereochemistry of Polypeptide and Polysaccharide Chain Conformations

G. N. RAMACHANDRAN, C. RAMAKRISHNAN AND V. SASISEKHARAN

Department of Physics, University of Madras, Madras, India

ABSTRACT

A notation is developed for representing the configuration of a polypeptide or polysaccharide chain. It is shown that the sugar residue in the latter has a rigid configuration and the parameters describing its standard configuration have been worked out, analogous to the Pauling-Corey parameters for the peptide group. The relative configuration of two peptide groups joined at an α-carbon atom may be represented by two angles ϕ, ϕ', which are the angles by which the two groups are rotated about the single bonds C_α—N and C_α—C′ from a defined standard configuration (0°, 0°). A similar notation is also possible for two sugar residues linked by two single bonds at a bridge oxygen atom. The allowed and forbidden ranges of (ϕ,ϕ') are worked out by examining the short contacts between the various atoms in the two linked units. In a helical structure, the configuration (ϕ, ϕ') is the same at every linkage, and the helical parameters n (= number of residues per turn) and h (= resolved height of a residue along the helical axis) have been worked out over the complete range of ϕ and ϕ' for polypeptide chains and over a range close to the observed structures in the case of polysaccharide chains. These stereochemical studies have revealed various interesting results, such as (a) the chain conformation of the individual chains in the triple helical structure of collagen is a very "natural" one for a polypeptide chain, (b) the "ribbon structure" is a structure likely to be observed in simple peptides and polypeptides, (c) the γ-helix is very unlikely to occur and (d) the configuration at each bridge oxygen in cellulose and chitin is very nearly the same as in cellobiose.

1. INTRODUCTION

Both polypeptides and polysaccharides are long-chain polymers, whose monomeric units are the peptide residues and sugar residues respectively. The polypeptide chains which occur in most proteins are linear and unbranched, although, occasionally, there may be cross-links between regions of a chain, mainly through –S—S– bonds. This is particularly so in fibrous proteins. In the same way, the polysaccharide chains in biological fibres, such as cellulose and chitin, are also believed to be linear and unbranched. The problem of theoretically working out the possible configurations of such a chain therefore consists of two steps:

(a) finding out the configuration of an individual unit, i.e. the bond distances and bond angles between the atoms in it; and

(b) finding out the relative configurations of two units linked to each other.

In the case of polypeptides, the step (a) was carried out by Pauling and Corey about ten years ago (Pauling *et al.*, 1951; Pauling and Corey, 1951; Corey and Pauling, 1953). These Pauling-Corey parameters (PC parameters) of the peptide group have been used in various investigations on protein and polypeptide structures. However, the nature of the linkage between two such groups was not studied systematically until very recently (Sasisekharan, 1962, see also Ramachandran, 1962). The planar peptide residues are linked together at an α-carbon atom and there is a possibility of free rotation about the single bonds by which these groups are attached at the α-carbon atom. Starting from a standard relative configuration, as shown in Fig. 1, a general one may be obtained

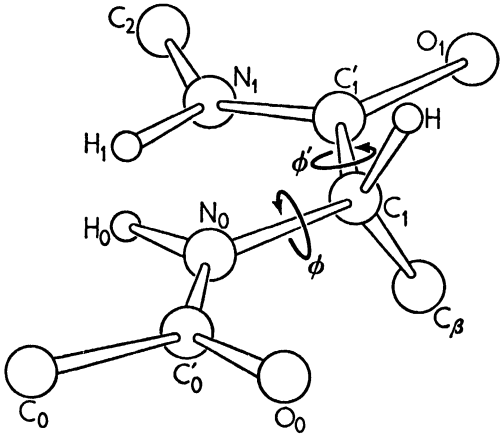


FIG. 1. Two peptide groups linked at an α-carbon atom. The angles ϕ, ϕ' which define a general configuration are marked in the figure. The location of the β-carbon atom for an L-type residue is also shown.

by rotating the two residues linked at this atom through angles ϕ, ϕ' as marked in this figure. If now we have a chain of residues denoted by indices 0, 1, 2,... linked successively at α-carbon atoms C_1, C_2,..., then the configuration of the whole chain may be described by means of the parameters (ϕ_1, ϕ_1'), (ϕ_2, ϕ_2'), (ϕ_3, ϕ_3')....

The problem is essentially the same with the polysaccharide chain also. The first step, namely that of finding the configuration of the backbone of the glucopyranose ring (by backbone is meant the six-membered ring composed of five carbon atoms and one oxygen atom), has now been carried out and is discussed in Section 4 below. The extension to step (b) is completely analogous to the case of the polypeptides and this is also

discussed in Section 5. The two monomeric units are linked in this case by single bonds at the bridge oxygen atom.

Thus, both for a polypeptide chain and a polysaccharide chain, the relative configurations of the units in the chain can be given by the parameters (ϕ_1, ϕ_1'), (ϕ_2, ϕ_2'), (ϕ_3, ϕ_3').... Now, if (ϕ, ϕ') is the same at every linkage, then the chain will take up a regular helical form, in which every monomeric unit is symmetrically equivalent to the others. Such a regular helix can be typified by two parameters n and h, where n is the number of monomeric units per turn of the helix and h is the height of the unit along the axis of the helix. We shall restrict ourselves to such regular helical forms in this paper.

A question that arises in this connection is whether all values of ϕ and ϕ' are permissible; or whether there are any restrictions. Such restrictions do exist because of the occurrence of short contacts between atoms of neighbouring units for certain values of ϕ and ϕ'. This leads to allowed and forbidden regions of (ϕ, ϕ') and, in consequence, it is only necessary to consider the possible configurations in the allowed region in discussing a structure. In addition, there may be certain configurations which are preferred, e.g. the α-helix for polypeptides, or the polysaccharide chain in cellulose, because they lead to the formation of hydrogen bonds. These are also discussed in the following sections.

2. Helical Polypeptide Chains

The convention adopted in defining ϕ and ϕ' in the case of two linked peptide residues is as shown in Fig. 1. The angle τ ($N_0C_1C_1'$) at the α-carbon atom C_1 normally lies between 110° and 115°. The location of the β-carbon atom for an L-type residue is also shown in Fig. 1. All the calculations reported in this paper are for the naturally occurring L-configuration for the peptide residues.

(a) *Calculation of the helical parameters* n *and* h

As mentioned earlier, when (ϕ, ϕ') is the same for every point of linkage, symmetry demands that the chain must have a helical symmetry, i.e. the succeeding residue is obtained from the preceding one by the operation of a rotation about some axis followed by a translation parallel to it. If the rotation is about an axis having direction cosines l, m, n with reference to a chosen co-ordinate system and it is through an angle ω, then it can readily be shown that the matrix of this rotational operation is given by

$$R = \begin{bmatrix} (a^2+b^2-c^2-d^2) & 2(bc+ad) & 2(bd-ac) \\ 2(bc-ad) & (a^2-b^2+c^2-d^2) & 2(cd+ab) \\ 2(bd+ac) & 2(cd-ab) & (a^2-b^2-c^2+d^2) \end{bmatrix} \quad (1)$$

where

$$a = \cos\frac{\omega}{2}, \quad b = l\sin\frac{\omega}{2}, \quad c = m\sin\frac{\omega}{2}, \quad d = n\sin\frac{\omega}{2} \qquad (2)$$

Taking the Y-axis parallel to C_0C_1 in the peptide group of index 0, the X-axis perpendicular to it in the plane of this peptide group and the Z-axis to be the third perpendicular direction normal to this plane (XYZ forming a right-handed co-ordinate system), it is possible to work out the matrix R corresponding to a given pair of parameters (ϕ, ϕ') (Ramakrishnan, 1963). Then a, b, c, d can be solved for from the elements of the matrix. Thus $\cos(\omega/2)$, and hence ω, the unit rotation of the helix, and from it n, the number of residues per turn, can be evaluated. Since the length L ($= C_0C_1$) is known, the unit height h ($= mL$) can also be obtained.

In this way, n and h have been obtained for the complete range of ϕ and ϕ' from 0° to 360°, corresponding to values of the angle τ at the α-carbon atom equal to 105°, 110° and 115°. The results for $\tau = 110°$ are shown in Fig. 2, along with the allowed ranges of ϕ and ϕ' to be discussed below. The detailed numerical results are being published elsewhere (Ramakrishnan, 1963).

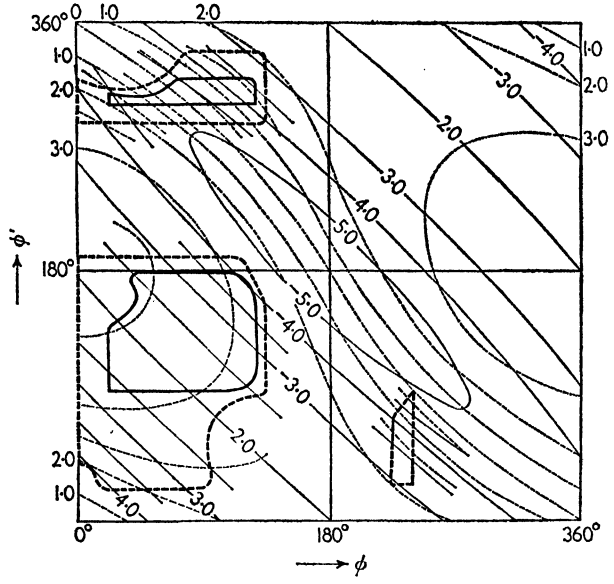


FIG. 2. The fully allowed region (———) and the outer limit (— — — —) for $\tau = 110°$. Also shown are the contours of constant n (———) and constant h (-----). The values of n are marked on the contours and the values of h are marked along the boundary.

(b) *Allowed ranges of ϕ and ϕ'*

The allowed range was worked out by calculating the interatomic distances between the various atoms in the two units (Sasisekharan, 1962). For this purpose, the β-carbon atom was also taken into account in addition to the atoms C'_0, O_0, N_0, H_0 of residue of index 0 and C'_1, O_1, N_1, H_1 of residue 1. The details are being published elsewhere (Ramakrishnan, 1963), but the results are shown in Fig. 2, corresponding to a value of $\tau = 110°$. The region within the thick lines is fully allowed and the broken lines are the outer limits. The contact distances assumed for the fully allowed region and the outer limit were as follows, in which the latter are shown within brackets:

C...C > 3·20 Å (3·00 Å)
C...O > 2·80 Å (2·70 Å)
C...N > 2·90 Å (2·80 Å)
C...H > 2·40 Å (2·20 Å)
O...O > 2·80 Å (2·70 Å)
O...N > 2·70 Å (2·60 Å)
O...H > 2·40 Å (2·20 Å)
N...N > 2·70 Å (2·60 Å)
N...H > 2·40 Å (2·20 Å)
H...H > 2·00 Å (1·90 Å)

3. Possible Helical Configurations

In order to compare the above theoretical predictions with observation, the parameters ϕ and ϕ' were evaluated for the configurations observed in various di-, tri-, and polypeptides (Sasisekharan, 1962). We shall call the pair (ϕ, ϕ') as the "configuration" of the two linked peptide residues. These are plotted in Fig. 3, and it will be seen that a large majority of them occur in the fully allowed region, although a few occur only within the outer limits (for a comparison with the earlier study of Sasisekharan, 1962, see Ramachandran, 1962). The main conclusions are the following:

Considering first di- and tripeptides, we find:

(i) At a glycyl α-carbon atom (i.e. when the β-carbon atom is absent), the usual configuration is close to a planar fully extended one, i.e. (0°, 180°).

(ii) When a β-carbon atom is present, ϕ is close to 120° and ϕ' lies between 120° and 180°. In fact, a value of ϕ' between 150° and 180° is observed in most of the amino acids (Sasisekharan, 1962).

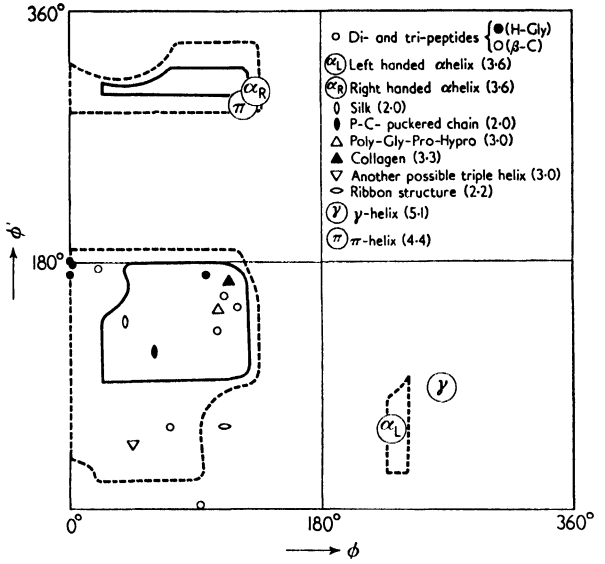


FIG. 3. The fully allowed region (———) and the outer limit (— — — — —) for $\tau = 110°$ and the configurations observed in the various known di- tri- and polypeptides and in proteins.

In the case of the known polypeptide configurations, we have:

(iii) The silk structure is close to the extended chain, but is not fully extended. Although the glycine content is large, the observed configuration is (40°, 140°), which is slightly shifted from the pure glycine configuration (0°, 180°) because of the presence of alanine and other residues. It now comes into the fully allowed region.

(iv) The right-handed α-helix (α_R of Fig. 3) is just outside the fully allowed region, while both the right-handed as well as the left-handed α-helices are within the outer limits. The helix α_R has $\phi \simeq 120°$, which therefore allows a proline residue to occur, but only with considerable distortion because of additional bad contacts due to the pyrrolidine ring. However, α_L has $\phi \simeq 240°$ and cannot at all have a pyrrolidine ring. Apart from this, Sasisekharan (1962) has found that the contact improves in α_R if the planes of the residues are slightly tilted from the vertical. Thus there are definite stereochemical reasons why the right-handed α-helix should occur with the naturally occurring L residues.

The right-handed π-helix (π in Fig. 3) has a configuration close to that of the α-helix, and it also occurs only within the outer limits. Its counterpart, the left-handed π-helix (not marked in Fig. 3) will also occur close to α_L and within the outer limits. Thus, the π-helix is not much inferior to the α-helix from the point of view of short contacts, and since both

have all the NH's hydrogen-bonded to CO's, the π-helix is quite likely to occur in some polypeptides.

On the other hand, the γ-helix occurs outside even the outer limits. The right-handed γ-helix considered by Pauling and Corey (Pauling and Corey, 1951) has the configuration (270°, 90°) and is marked γ in Fig. 3. It has two very short contacts from C_β, namely $C_\beta \ldots C_0 \simeq 2{\cdot}75$ Å and $C_\beta \ldots O_0 \simeq 2{\cdot}45$ Å. The left-handed γ-helix (not marked) will therefore have the configuration (90°, 270°). This also has one very short contact, namely $C_\beta \ldots N_2 \simeq 2{\cdot}67$ Å. In view of this both are unlikely to occur.

(v) The collagen triple helix (Ramachandran and Sasisekharan, 1961) and the related triple helices in polyglycine II (Rich and Crick, 1955), poly-L-proline II (Sasisekharan, 1959*a*), and poly-L-hydroxyproline A (Sasisekharan, 1959*b*) all have the parameters (120°, 150°). It is also interesting to note that the linkages in a number of di- and tri-peptides have a configuration close to this. Thus, apart from the fact that the Madras triple helix has $\phi \simeq 120°$, and hence can accommodate the pyrrolidine ring readily, the linkage between the peptide residues is also a very natural one in it. As already mentioned, the value of ϕ' is close to 150° in most amino acids. Consequently, even without internal hydrogen bonding, the collagen chain configuration is likely to be stable.

(vi) Another possible structure that was suggested earlier is the so-called ribbon structure (see Donohue, 1953), also known as the $2{\cdot}2_7$ helix. Its configuration is (112°, 60°) and it is outside the outer limits. However, the short contacts which occur in it are $N_0 \ldots O_1 = 2{\cdot}75$ Å and $H_0 \ldots O_1 = 1{\cdot}90$ Å. Since these contacts are involved in hydrogen-bond formation, and the angle NH∧NO is reasonable (24°), this is not an undesirable bad contact, but a stabilizing bond. It is therefore quite likely that the ribbon structure will be found in some polypeptides.

(vii) Finally, there is another possible triple helix, marked $\triangledown$ in Fig. 3, which is only in the outer limit region. It cannot accommodate proline or hydroxyproline, but can form interchain hydrogen bonds like the collagen helix. It is a right-handed helix, unlike the collagen helix, which is left-handed. Its properties would require further study.

Thus, apart from the analogues of the α-helix (which includes the π-helix), the extended β-structure (e.g. silk), and the Madras triple helix, the only configuration that is highly likely to occur is the ribbon structure. This also has a two-fold screw axis as in silk, but unlike the latter it is internally hydrogen-bonded. The right-handed triple helix is also a possibility.

4. Standard Parameters for a Sugar Residue

The basic unit in the case of polysaccharides is the monosaccharide unit or the sugar residue, e.g. the glucopyranose residue in cellulose. As

TABLE I. The co-ordinates of the atoms in different structures containing the glucopyranose ring

The origin is at atom C_3, the X-axis is along C_3C_5 and $C_3C_1C_5$ is the XY-plane. The mean co-ordinates of the atoms with the probable errors are given in the last column.

Atom	Structure 1	2	3	4	5	6	7	8	9	10	Mean
						x					
C_1	1·45	1·30	1·36	1·39	1·31	1·32	1·27	1·34	1·28	1·37	1·34 ± 0·04
C_2	0·03	0·07	0·05	0·09	0·10	0·01	−0·01	0·07	0·00	0·06	0·05 ± 0·03
C_3	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00
C_4	1·24	1·22	1·32	1·23	1·24	1·23	1·24	1·25	1·22	1·23	1·24 ± 0·02
C_5	2·47	2·54	2·56	2·56	2·53	2·48	2·51	2·49	2·41	2·50	2·50 ± 0·03
O_5	2·48	2·47	2·41	2·46	2·47	2·46	2·43	2·47	2·37	2·46	2·45 ± 0·02
O_1	1·52	1·37	1·40	1·52	1·39	1·38	1·33	1·48	1·43	1·46	1·43 ± 0·05
O_4	1·24	1·23	1·39	1·28	1·26	1·23	1·19	1·24	1·22	1·25	1·25 ± 0·03
O_2	−1·03	−1·09	−1·05	−1·05	−1·06	−1·04	−1·16	−1·04	−1·07	−1·06	−1·06 ± 0·02
O_3	−1·17	−1·20	−1·13	−1·19	−1·16	−1·02	−1·15	−1·19	−1·18	−1·02	−1·15 ± 0·03
C_6	3·76	3·85	3·86	3·77	3·78	3·80	3·83	—	3·70	—	3·80 ± 0·04
						y					
C_1	2·07	2·08	2·08	2·08	2·06	2·05	2·15	2·10	2·07	2·04	2·08 ± 0·02
C_2	1·47	1·41	1·40	1·49	1·37	1·46	1·47	1·44	1·42	1·44	1·44 ± 0·03
C_3	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00
C_4	−0·73	−0·84	−0·74	−0·74	−0·79	−0·77	−0·73	−0·78	−0·73	−0·78	−0·76 ± 0·03
C_5	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00
O_5	1·27	1·33	1·38	1·31	1·31	1·32	1·41	1·37	1·31	1·31	1·33 ± 0·03

O_1	3·33	3·38	3·42	3·37	3·34	3·24	3·38	3·39	3·33	3·35	3·35 ± 0·03
O_4	−2·15	−2·04	−2·02	−1·99	−2·03	−2·09	−2·05	−2·11	−2·07	−2·03	−2·06 ± 0·03
O_2	2·25	2·21	2·06	2·15	2·17	2·19	2·19	2·22	2·11	2·13	2·17 ± 0·04
O_3	−0·61	−0·61	−0·59	−0·54	−0·73	−0·72	−0·51	−0·62	−0·65	−0·81	−0·62 ± 0·05
C_6	−0·73	−0·71	−0·52	−0·61	−0·74	−0·66	−0·57	—	−0·59	—	−0·64 ± 0·06
						z					
C_1	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00
C_2	0·37	0·55	0·36	0·53	0·55	0·53	0·44	0·51	0·39	0·47	0·47 ± 0·06
C_3	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00
C_4	0·44	0·27	0·37	0·38	0·36	0·48	0·50	0·49	0·52	0·38	0·42 ± 0·06
C_5	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00
O_5	0·52	0·41	0·50	0·57	0·50	0·46	0·35	0·46	0·53	0·55	0·48 ± 0·04
O_1	0·65	0·47	0·35	0·52	0·52	0·52	0·60	0·52	0·63	0·49	0·53 ± 0·05
O_4	0·04	−0·25	−0·32	−0·37	−0·32	0·02	0·01	−0·06	0·07	−0·30	−0·13 ± 0·14
O_2	−0·03	0·24	−0·23	−0·13	0·19	0·00	0·14	0·00	−0·15	−0·05	0·00 ± 0·10
O_3	0·53	0·50	0·57	0·52	0·39	0·72	0·42	0·49	0·38	0·56	0·50 ± 0·06
C_6	0·40	0·39	0·65	0·61	0·64	0·38	0·48	—	0·37	—	0·49 ± 0·09

1. β-D-Glucose (Ferrier, 1960).
2, 3. Cellobiose (Jacobson *et al.*, 1961).
4, 5. Cellobiose (Brown, 1962*a*).
6. α-D-Glucose (McDonald and Beevers, 1950).
7. α-D-Rhamnose monohydrate (McGeachin and Beevers, 1957).
8. β-Arabinose (Hordvik, 1961).
9. Barium glucose orthophosphate (Kartha, 1962).
10. β-Methyl-xyloside (Brown, 1962*b*).

mentioned in Section 1, the first step in the investigation of the polysaccharide chain configuration is to work out the standard configuration for the backbone of the sugar residue, analogous to the single peptide residue in the case of polypeptides. The backbone forming the pyranose ring is the same in the various sugars and only the side groups attached to the ring atoms differ.

The β-glucopyranose unit is shown in Fig. 4, wherein the designation of the atoms in the usual way is also indicated. Unlike the peptide residue,

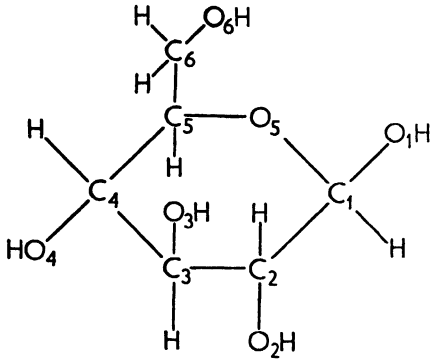


FIG. 4. The schematic representation of β-D-glucose. The designation of the various atoms is also indicated.

which is planar, the configuration of the glucopyranose unit is not unique. It can take up either the so-called boat form or the chair form. This aspect has been examined by Reeves (1950) and he has shown that the chair form is more likely to occur than the boat form. Even for this form, it can be seen from a model that there is a certain amount of flexibility for the ring. This observation is however misleading, for, a careful analysis of the published data on various crystal structures containing the pyranose ring shows that the ring skeleton is remarkably uniform in its configuration.

The sugars whose structures are known to a good degree of accuracy are listed in Table I. In all these substances, the ring (backbone) $C_1C_2C_3C_4C_5O_5$ is the same and only the positions of the substituents (or side-groups) are different. The data on their structure can therefore be used to arrive at a standard configuration for the glucopyranose ring. The following system of axes was used to co-ordinate the different data. The atom C_3 was taken to be the origin of co-ordinates and C_3C_5 the X-axis, with $C_3^{\prime}C_1C_5$ forming the XY-plane. The Z-axis was taken perpendicular to this plane, so that XYZ formed a right-handed system. Special techniques involving the stereographic projection were developed

for the purpose of transforming the co-ordinates from the structural reports to this system.

The co-ordinates of the atoms for the different structures are given in Table I, along with the mean co-ordinates of the various atoms and their probable errors. The atom O_6 is not included in this Table, since the position of O_6 in a crystal need not be specific owing to the free rotation about the C_5C_6 bond. It will be seen from Table I that the co-ordinates of the atoms forming the backbone (ring) are highly specific and that the maximum probable error is only $\pm 0{\cdot}06$ Å in the z-co-ordinate, i.e. normal to the plane of the ring. It can be seen further that the three atoms C_2, C_4 and O_5 also lie in a plane very nearly parallel to the XY plane containing the atoms $C_1C_3C_5$. The probable errors in the z-co-ordinates of O_4 and O_2 are fairly large, $\pm 0{\cdot}14$ Å and $\pm 0{\cdot}10$ Å respectively, while the probable error in the other two co-ordinates x and y is only about $\pm 0{\cdot}04$ Å. Thus, these two atoms can have a slight wagging motion of the order of $\pm 0{\cdot}1$ Å in a direction perpendicular to the plane of the ring. In Table II are given the bond lengths and bond angles calculated from the above mean co-ordinates.

The polysaccharide chains are built up from their monomer units, by means of linkages through the bridge oxygen atoms. In the case of cellulose and chitin, it is a 1,4-linkage and so the angle at the bridge

TABLE II. Bond lengths and bond angles in the glucopyranose ring. Calculated from the mean co-ordinates of the atoms given in Table I.

Bond	Length (Å)	Bond angle	Value (°)
C_1—C_2	1·52	O_5—C_1—C_2	110
C_2—C_3	1·51	C_1—C_2—C_3	109
C_3—C_4	1·52	C_2—C_3—C_4	111
C_4—C_5	1·53	C_3—C_4—C_5	110
C_5—O_5	1·42	C_4—C_5—O_5	110
C_1—O_5	1·42	C_5—O_5—C_1	114
C_1—O_1	1·38	O_5—C_1—O_1	108
C_4—O_4	1·41	O_1—C_1—C_2	109
C_2—O_2	1·41	C_1—C_2—O_2	110
C_3—O_3	1·40	O_2—C_2—C_3	112
C_5—C_6	1·52	C_2—C_3—O_3	109
		O_3—C_3—C_4	111
		C_3—C_4—O_4	111
		O_4—C_4—C_5	110
		C_4—C_5—C_6	114
		C_6—C_5—O_5	109

oxygen O plays the role of the angle at C_α in the configuration of polysaccharides. Among the various substances studied so far, only in the two disaccharides, namely cellobiose and sucrose, are there two units linked through an oxygen atom. The values of the angle at the bridge oxygen in these structures are 117·5°, 116·8° from the two determinations on cellobiose and 118·3° in sucrose (Beevers *et al.*, 1952), so that a mean value of 117·5° was taken to be the standard value for this angle.

5. Configuration of Polysaccharide Chains

The case of the polysaccharides is very similar to that of the polypeptides; here also the two sugar residues linked at the oxygen atom (O) are capable of free rotation about the single bonds C_1O and $C_4'O$, the angles of rotation being again denoted by ϕ and ϕ' respectively. The initial configuration $(\phi, \phi') = (0°, 0°)$ is taken to be the one in which the other glycosidic oxygens, i.e. O_4 and O_1' lie in the plane of C_1OC_4', as shown in Fig. 5.

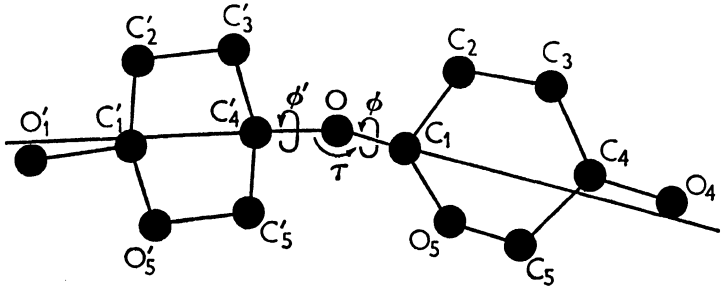


Fig. 5. Two sugar residues linked through the bridge oxygen atom O. The initial configuration (0°,0°) is shown along with the angles ϕ, ϕ' which define a general configuration.

Analogous to the intramolecular N—H...O=C bond of the polypeptide structures, in this case also, a hydrogen bond can be formed between the atom O_3 of one residue and the ring oxygen atom O_5 of the adjacent residue. The importance of the formation of such a bond has been dealt with by Carlström (1962) in discussing the bent chain structure of α-chitin and by Jones (1960) in connection with the structure of cellulose.

In our notation, this hydrogen bond occurs around the configuration $(-30°, 210°)$. Making use of techniques involving stereographic projection, the bond distance around this region was calculated. Taking 2·5 Å and 2·8 Å as the lower and upper limits for the hydrogen-bond length, the region of (ϕ, ϕ') within which this bond is formed was evaluated

and is shown in Fig. 6 (the region between the two curves marked 2·5 and 2·8).

However, all the configurations within this range of (ϕ, ϕ') may not be permitted due to some possible bad contacts. Using the fully allowed and outer limit contact distances mentioned in Section 2B, the region around was examined for the short contacts between atoms C_1, C_3'; C_1, C_5'; C_4', O_5; and C_4', C_2. The boundaries of the fully allowed and outer limit regions thus obtained are also shown in Fig. 6.

Using the same method as described in Section 2A, the helical parameters n and h were worked out over this range. The contours of n and h are also shown in Fig. 6.

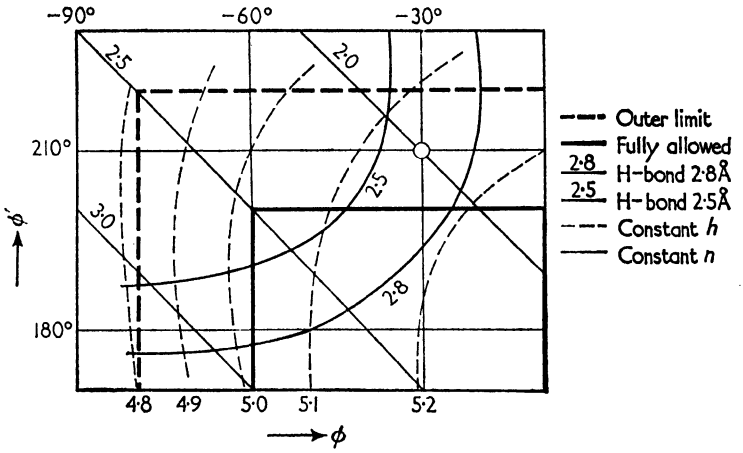


FIG. 6. The fully allowed region and the outer limit around the configuration (−30°, 210°) (marked by a circle) for a pair of sugar residues. The region where a reasonably good hydrogen bond can be formed is the one between the contours marked 2·5 and 2·8. The contours of n and h are also shown in the figure.

Thus, the configurations likely for polysaccharide chains are those which occur within the strip outlined by the hydrogen-bond lengths 2·5 Å and 2·8 Å and the outer limit boundary. It will be noticed that the number of residues per turn (n) in the helix can vary between 2 and 3 for these configurations. If the chains have to form a regular lattice, then n has to be either 2 or 3. On the other hand, it is quite likely that n may be equal to 2·5 and the structure may have five residues in two turns. The repeat spacing in this case will be five times the unit height h. It will be seen from Fig. 6 that the residue height h varies only a very little over the whole range, and is around 5 Å. The above results are in good agreement with observation. Thus, Meyer (1950) has divided the various

crystalline modifications of cellulose into three families, having an observed repeat spacing along the fibre axis of about 10 Å (2 × 5 Å), 15 Å (3 × 5 Å) and 25 Å (5 × 5 Å). Further, the actual repeat spacings found for cellulose (Meyer, 1950) and chitin (Carlström, 1962) are close to 10·3 Å, i.e. $n = 2$, $h = 5{\cdot}15$ Å. This agrees perfectly with $\phi = -30°$, $\phi' = 210°$ (marked by a circle in Fig. 6). Although the contact $C_1 \ldots C_3$ is only in the outer limit, the occurrence of a strong hydrogen bond offsets this. In fact, the natural configuration in cellobiose itself is also close to this, which again shows that the cellulose and chitin structures should be based on the cellobiose configuration. Attempts are being made to refine these two polysaccharide structures on this basis.

References

Beevers, C. A., McDonald, T. R. R., Robertson, J. H. and Stern, F. (1952). *Acta cryst.* **5**, 689.

Brown, C. J. (1962*a*). Personal communication.

Brown, C. J. (1962*b*). Personal communication.

Carlström, D. (1962). *Biochim. biophys. Acta* **59**, 361.

Corey, R. B. and Pauling, L. (1953). *Proc. roy. Soc.* **B141**, 10.

Donohue, J. (1953). *Proc. nat. Acad. Sci., Wash.* **39**, 470.

Ferrier, W. G. (1960). *Acta cryst.* **13**, 678.

Hordvik, A. (1961). *Acta chem. scand.* **15**, 16.

Jacobson, R. A., Wunderlich, J. A. and Lipscomb, W. N. (1961). *Acta cryst.* **14**, 598.

Jones, D. W. (1960). *J. Polym. Sci.* **42**, 173.

Kartha, G. (1962). Personal communication.

McDonald, T. R. R. and Beevers, C. A. (1950). *Acta cryst.* **3**, 394.

McGeachin, H. Mc.D. and Beevers C. A. (1957). *Acta cryst.* **10**, 227.

Meyer, K. H. (1950). In *Natural and Synthetic High Polymers*, Vol. 4, pp. 300, 340.

Pauling, L. and Corey, R. B. (1951). *Proc. nat. Acad. Sci., Wash.* **37**, 235.

Pauling, L., Corey, R. B. and Branson, H. R. (1951). *Proc. nat. Acad. Sci., Wash.* **37**, 205.

Ramachandran, G. N. (1962). In *Collagen* (N. Ramanathan, ed.), p. 3. Wiley, New York.

Ramachandran, G. N. and Sasisekharan, V. (1961). *Current Sci. (India)* **30**, 127.

Ramakrishnan, C. (1963). To be published.

Reeves, R. E. (1950). *J. Amer. chem. Soc.* **72**, 1499.

Rich, A. and Crick, F. H. C. (1955). *Nature, Lond.* **176**, 915.

Sasisekharan, V. (1959*a*). *Acta cryst.* **12**, 897.

Sasisekharan, V. (1959*b*). *Acta cryst.* **12**, 903.

Sasisekharan, V. (1962). In *Collagen* (N. Ramanathan ed.), p. 39. Wiley, New York.

Discussion

N. S. ANDREEVA: I would like to say in this connection that the polymer (-Gly-Pro-Hypro-)$_n$ forms a major helix of the collagen type with ten diffractional units in three turns. The most prominent layer lines of its X-ray pattern are zero, third, seventh and tenth.

E. KATCHALSKI: Will your calculations concerning free rotation be affected by the nature of the amino acid side chain? How do five- or six-membered rings affect the rotation?

G. N. RAMACHANDRAN: In evaluating the restrictions on the orientations shown in Figs. 2 and 3, the interaction of the β-carbon atom with the rest of the atoms has also been taken into account. If a five membered ring comes, it joins the β-carbon atom to the nitrogen with $\phi = 120°$. A similar locking will occur also with a rigid six-membered ring, as in poly-L-pipecolic acid. If it is any other side chain, then only the β-carbon atom matters, and the full range shown in Figs. 2 and 3 is effective.

Structural Investigations of Polymers Related to Collagen

N. S. Andreeva, M. I. Millionova and
Yu. N. Chirgadze

Institute of Biological Physics, Academy of Sciences of the U.S.S.R., Moscow, U.S.S.R.

ABSTRACT

The relation of secondary and tertiary structure to the primary structure of collagen is dealt with. The possibility of obtaining polypeptides isomorphous with a fibrous protein such as collagen is also considered. The structural features of the polypeptide (-Gly-Pro-Hypro-)$_n$ are presented with details of X-ray, optical and infra-red data and also the determination of molecular weights of the two forms of the polytripeptide. The relations between the polytripeptide and collagen structures are discussed.

The regular configuration of a polymeric chain is the consequence of stereochemical requirements both for its backbone and side groups. In many cases, the regular secondary structure of proteins is caused only by the polypeptide backbone, because the structural differences of various amino acids and the irregularity of their arrangement in chains prevent effective packing of side groups. Thus polypeptide chains of many proteins have α-helical configuration—the most stable structural form of polypeptide backbone (Pauling *et al.*, 1951; Donohue, 1953).

However, in some cases a requirement of effective packing of side groups might not be compatible with the most stable configuration of backbone. Hence the general form of a chain will depend on the energy of side group packing (stabilization). When it is prevalent, the α-helix breaks down and new configurations may appear. Only regularity in arrangement of some specific side groups along the chain can provide enough energy for formation of regular configurations of this type.

There are several examples of such structures for monotonic synthetic polypeptides (Bradbury *et al.*, 1962; Fraser *et al.*, 1962; Cowan and McGavin, 1955).

There are two examples of such structures for fibrous proteins—collagen and silk fibroin. Thus the amino acid sequence in these proteins must be more or less regular, at least in some parts of the chains.

At present certain chemical data support this conclusion. The sequence of amino acids in some parts of *Bombyx mori* silk fibroin chains can

be described as the sequence (-Gly-Ala-)$_n$ with substitution of Ala by Ser in some places (Ioffe, 1954). The other type of fibroin has considerable amounts of alanine. Its structure is isomorphous with poly-L-alanine (Marsh *et al.*, 1955). There is no doubt that the structure of this fibroin is due to a monotonic arrangement of alanine residues into chains.

The primary structure of collagen is less well known than that of fibroin. But there is strong evidence to suggest a regular arrangement of glycine residues along a collagen chain (Grassmann *et al.*, 1960). Many chemical data show the main part of the collagen molecule to be formed from triplets containing glycine, imino acid residues and alanine (Kroner *et al.*, 1953; Schroeder and Kay, 1954; Schrohenloher *et al.*, 1959). The most plausible models of the collagen structure, proposed by Ramachandran and Kartha (1955), Rich and Crick (1955) and Cowan *et al.* (1955), are based on some type of regular amino acid sequence.

As the structure of these two proteins must be due to the regular arrangement of amino acid pairs or triplets, it can be investigated more effectively by examination of model polypeptides each with a regular sequence of residues. Such polymers exhibit better crystallization and thus more effectively reveal structural properties.

We chose this method for collagen structure investigations. Various polypeptides need to be synthesized from peptides containing glycine, imino acids and alanine. The first problem was to obtain a polypeptide isomorphous with collagen.

A polypeptide that showed collagen-like properties (poly-(-Gly-L-Pro-L-Hypro-) or, for convenience, poly-(-Gly-Pro-Hypro-)) was synthesized by Debabov and Shibnev about two years ago (Debabov and Shibnev, 1961; Andreeva *et al.*, 1961). Non-oriented preparations of this polymer show X-ray patterns very similar to those of collagen with a prominent 2·82 Å reflection. They had also the significant negative optical rotation and in the infra-red spectrum they gave the collagen-specific displacement of the NH stretching vibration band (Andreeva *et al.*, 1961a, b).

Katchalski and his collaborators also began to work on the synthesis of polypeptides from peptides containing glycine, alanine and imino acids. Berger and Wolman (1961) reported preliminary data on the optical rotation and mutarotation of these compounds. Seifter *et al.* (1961) investigated the splitting of such polymers with collagenase. The work on the synthesis of regular polymers from tripeptides containing glycine, proline and leucine was performed also by Japanese investigators (Kitaoka, 1958). In all these cases, however, the polymers obtained differed to some extent from our polymer. Only our polymer is reported to be isomorphous with collagen.

But the complete proof of isomorphism of this polymer and collagen can be obtained only by more detailed investigations, including exami-

nation of oriented specimens. The present studies are designed to obtain information about the physico-chemical properties of the polymer. Improved methods for its synthesis have been found.

1. General Structural Properties of Poly-(-Gly--Pro-Hypro-) and its Optical Rotation

The synthesis of poly-(-Gly-Pro-Hypro-) is to be the subject of a special paper. While searching for the most effective methods of synthesis, every preparation was checked by X-ray and optical methods. Some of them were also analysed by chemical and ultracentrifuge methods. Preparations were fractionated on DEAE-Sephadex.

The main results of this part of the work are the following.

(1) Fractions with low molecular weights (up to about 4000 determined by the ultracentrifuge) do not possess collagen-like features.

(2) Polymers with average molecular weights greater than 4000 (ultracentrifuge data) can have the two different structural modifications, A (collagen-like) and B (unknown). As usual, after synthesis we obtained mixtures of the two forms containing varying amounts of each of them.

The mixtures have an optical rotation of about $-200°$. Their non-oriented X-ray patterns (Fig. 1) contain the same total number of reflections as in collagen (see below) with additional reflections of the second (B) form. The most prominent among them is the reflection corresponding to 3·25 Å. The relative intensity of reflections for the A and B forms changes from specimen to specimen. The whole background in the inner part of the X-ray patterns is also variable. According to this, optical rotation also changes.

The first task was to separate these forms and investigate their properties in as pure a state as possible. To some extent they could be separated on DEAE-Sephadex. The collagen-like form (A) was predominant in the highest molecular weight fraction, but the second (B) form can also exist if the molecular weight is large enough. From some solutions crystallization gave a collagen-like form. The best were solutions in *m*-cresol and water. In all cases, we were not quite sure that complete separation had been achieved.

X-Ray and optical data obtained for the A and B forms are presented in Table I and Fig. 1.

The reflections listed in Table I correspond to Bessel functions of zero order and first order on the X-ray pattern of collagen.

$(\text{-Gly-Pro-Hypro-})_n$ in concentrated solution in water and *m*-cresol solutions forms liquid crystals and spherulites. The distance between molecules changes with concentration, but at some concentrations in

TABLE I. X-Ray and optical data for A and B forms

Preparation	Spacings (Å) for the strong reflections				Optical rotation
	a†	b	c	d	$[\alpha]_D$
Collagen	$11 \pm 0{\cdot}5$	$7{\cdot}5 \pm 0{\cdot}2$	$2{\cdot}86 \pm 0{\cdot}02$	$4{\cdot}1 \pm 0{\cdot}1$	$-350°$
Fraction A of (-Gly-Pro-Hypro-)$_n$	$11 \pm 0{\cdot}5$	$7{\cdot}4 \pm 0{\cdot}2$	$2{\cdot}82 \pm 0{\cdot}02$	$3{\cdot}87 \pm 0{\cdot}05$	about $-280°$
Fraction B of (-Gly-Pro-Hypro-)$_n$	$11 \pm 0{\cdot}5$	$6{\cdot}8 \pm 0{\cdot}3$	$3{\cdot}25 \pm 0{\cdot}04$	diffuse background	about $-150°$

† This spacing increases with increasing water or solvent content for all preparations.

m-cresol solutions it is stable. Such crystallization was observed for preparations of molecular weight ~ 10,000–15,000 and an asymmetry of about 8·5. The common type of spherulite growing in water solutions is shown in Fig. 2. After drying, these spherulites became fragile and disintegrated.

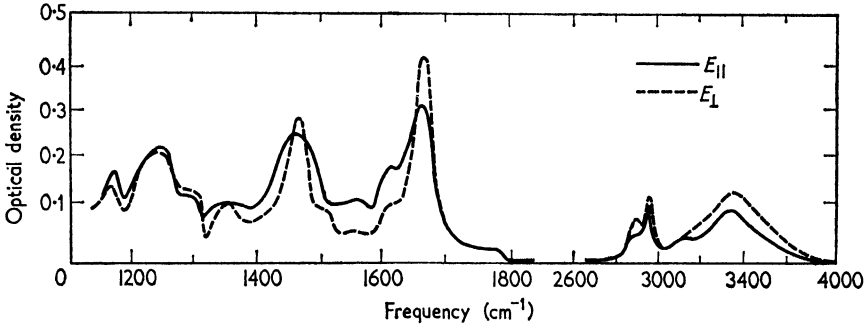


FIG. 3. Infra-red spectrum of poly-(-Gly-Pro-Hypro-), fraction A.

The tendency for the polymer to crystallize in this manner shows the presence of a considerable amount of elongated aggregates of the same thickness. These specimens were used in experiments on the orientation of the polymer.

The best orientation was obtained for films from *m*-cresol solutions. However, this orientation is very unstable and rapidly disappears completely. The X-ray investigation of oriented polymer was very difficult because orientation in films was destroyed during the exposure time. We could obtain only X-ray patterns with poor orientation. The

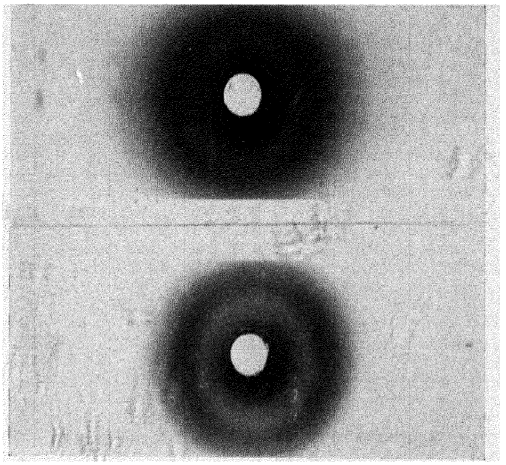


FIG. 1. X-Ray diffraction pattern of the polymer showing form A (above) and form B (below).

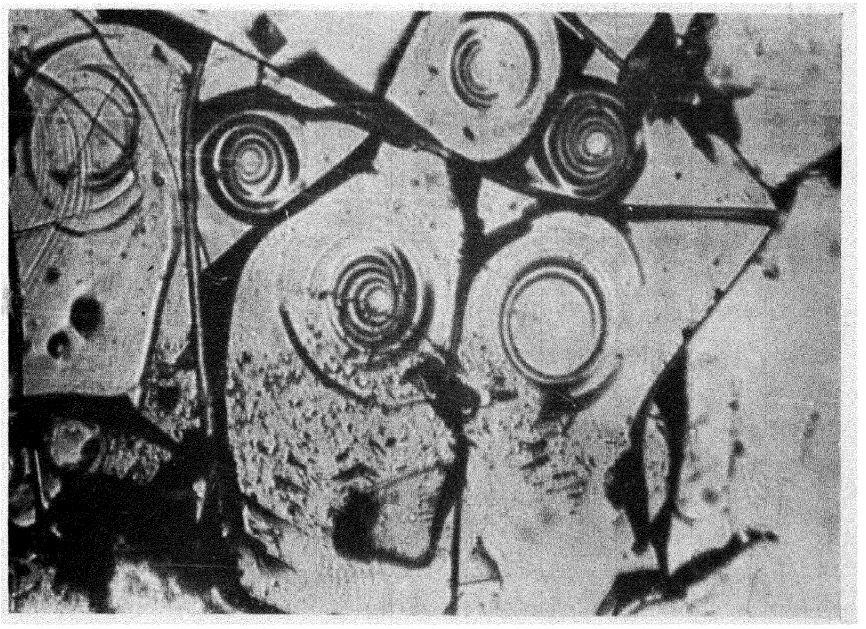


FIG. 2. Spherulites of polymer (x 140).

11·5 Å reflection shows equatorial orientation; it is opposite to the orientation of the 2·82 Å reflection.

Because of this we could obtain good polarized infra-red spectra. They show prominent perpendicular dichroism for CO and NH stretching vibrations which is one of the features of collagen (Fig. 3). The frequency of the last vibration is shifted to 3350–3360 cm^{-1}, in accordance with the specific shift for collagen structures. The polarization and frequencies of other vibrations are in good agreement with theoretical calculations, explaining the main features of the spectra of polymers containing imino acid residues, such as poly-(-Gly-Pro-Hypro-) and polyproline (Chirgadze, 1962).

2. Molecular Weight Determinations

Debabov and Shibnev have determined molecular weights of synthesized preparations by chemical methods. Moroskin (1963) investigated sedimentation properties of the polymer and its diffusion. He determined also the molecular weights of various preparations by the Archibald equilibrium method. The results obtained are presented in Table II.

Table II. Molecular weights of poly-(-Gly-Pro-Hypro-)

Preparation	Fraction	Chemical data	Ultracentrifuge data			
			$S_{20,w} \times 10^{13}$	Molecular weight	Asymmetry	m/L (Avogr.)
I	A	1700–2000	1·9	25,000	8·5	170
	B	1200–1500	1·3	13,100	7·5	120
	C	about 800	0·54	3070	5·0	85
II	A	1700–2000	1·3	14,800	8·5	120
	B	1200–1500	0·72	4540	5·0	85
	C	about 800	0·47	1900	2·5	85

The most interesting fact is the following. Chemical methods showed molecular weights that were always much lower (average about 1500–2000) than those obtained with the ultracentrifuge. This fact can be explained if the polymer forms aggregates. Further experiments showed that such aggregates dissociate in solutions having the property of breaking down hydrogen bonds.

The highest molecular weight was about 36,000. In accordance with X-ray investigations of $(\text{-Gly-Pro-Hypro-})_n$, these data support the main idea about the tendency of the polymer to form multichain aggregates.

But the heterogeneity of the relatively low molecular weight individual chains seems to make all aggregates not very regular.

No evident relations between average molecular weights and average asymmetries for various fractions were obtained. The average asymmetry of aggregates of weight $\sim$ 15,000 is about 8. Simple calculations show such aggregates to contain more than three laterally bound chains. Based on the data obtained, we propose the scheme shown in Fig. 4 for the structure of the aggregates that poly-(-Gly-Pro-Hypro-) forms in water solutions. We assume these aggregates to have regions with triple chain structures to provide a collagen-like feature. At the same time, in other regions, the second structural form (B) can exist. The data

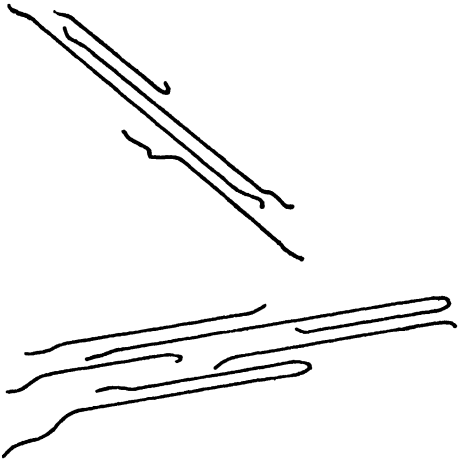


FIG. 4. The scheme of aggregation of polymer chains.

obtained show that lateral aggregation increases with increasing molecular weight. For example, preparations of molecular weight $\sim$ 36,000 had an asymmetry of about 9. This makes experiments on the orientation of the polymer very difficult.

3. CONCLUSION

At present we have no doubt that form A of the polymer has a structure very similar to that of collagen. The basis of this structure is formed with left-handed helices. Increasing the content of form A always leads to an increase in the negative optical rotation. Left-handed helices are bound in multichain aggregates by hydrogen bonds. As infra-red data showed, these bonds are nearly perpendicular to the chain axis.

The strongest of the polymer X-ray pattern reflections correspond to those of collagen, the structure factors depending mainly on the values

of Bessel functions of zero order and first order. The comparison of the relative intensities of these reflections using atomic co-ordinates proposed by Rich and Crick (1961), Burge *et al.* (1958) and Ramachandran *et al.* (1960), is not quite convincing, though there is general agreement. Until the positions of the water molecules, which play an important role in scattering, are found, all such calculations will not be quite correct.

We believe that the polymer (Gly-Pro-Hypro-)$_n$ is one of the best models for further investigations of the atomic parameters in collagen-type structures.

The specimens we investigated were not convenient for obtaining ordered structures. The molecular weights of individual chains were rather low; the asymmetry of multichain aggregates was also low. The polymer contains one phosphate group per chain bound to its amino end. Although such a small group cannot change the main structural properties of the polymer, the possibility of branching is not completely excluded. This causes additional difficulties in obtaining the ordered structures.

We hope that results of these investigations will stimulate the development of more effective methods for the synthesis of the regular polytripeptides related to collagen.

Acknowledgement

We are grateful to the Director of the Institute of Biological Physics, Academy of Sciences of the U.S.S.R., Professor G. M. Frank, for his constant interest in the development of protein structure investigations.

References

Andreeva, N. S., Debabov, V. A., Millionova, M. I., Shibnev, V. A. and Chirgadze, Yu. N. (1961*a*). *Biofizika* **6**, 244.

Andreeva, N. S., Millionova, M. I. and Chirgadze, Yu. N. (1961*b*). *Report on the Vth International Biochemical Congress* (Symposium N I).

Berger, A. and Wolman, Y. (1961). *Abstracts for Sessional Communications of the Vth International Biochemical Congress*, p. 24.

Bradbury, E. M., Brown, L., Downie, A. R., Elliott, A., Fraser, R. D. B. and Hanby, W. E. (1962). *J. mol. Biol.* **5**, 230.

Burge, R. E., Cowan, P. M. and McGavin, S. (1958). In *Recent Advances in Gelatin and Glue Research* (G. Stainsby, ed.), p. 25. Pergamon Press, London.

Chirgadze, Yu. N. (1962). *Biofizika* **7**, 523, 657.

Cowan, P. M. and McGavin, S. (1955). *Nature, Lond.* **176**, 501.

Debabov, V. A. and Shibnev, V. A. (1961). *Report on the Vth International Biochemical Congress* (Session I).

Donohue, J. (1953). *Proc. nat. Acad. Sci., Wash.* **39**, 470.

Fraser, R. D. B., MacRae, T. P. and Stapleton, J. W. (1962). *Nature, Lond.* **193**, 573.

Grassmann, W., Hannig, K. and Schleyer, M. (1960). *Hoppe-Seyl. Z.* **322**, 71.

Ioffe, K. G. (1954). *Biokhimia* **19**, 495.
Kitaoka, H. (1958). *Bull. chem. Soc. Japan* **31**, 802.
Kroner, D., Tabroff, W. and McGarr, J. (1953). *J. Amer. chem. Soc.*, **75**, 4084.
Marsh, R. E., Corey, R. B. and Pauling, L. (1955). *Acta cryst.* **8**, 710.
Moroskin, A. (1963). *Biofizika* **8**, 509.
Pauling, L., Corey, R. B. and Branson, H. R. (1951). *Proc. nat. Acad. Sci., Wash.* **37**, 205.
Ramachandran, G. N. and Kartha, G. (1955). *Proc. Indian Acad. Sci.* **A42**, 215.
Ramachandran, G. N., Sasisekharan, V. and Thathachari, Y. T. (1960). In *Collagen* (N. Ramanathan, ed.), p. 81. Wiley, New York.
Rich, A. and Crick, F. H. C. (1955). *Nature, Lond.* **176**, 915.
Rich, A. and Crick, F. H. C. (1961). *J. mol. Biol.* **3**, 483.
Schroeder, W. A. and Kay, L. M. (1954). *J. Amer. chem. Soc.* **76**, 3356.
Schrohenloher, R. E., Ogle, J. D. and Logan, M. A. (1959). *J. biol. Chem.* **234**, 58.
Seifter, S., Gallop, P. M. and Franzblau, C. (1961). *Trans. N.Y. Acad. Sci.* **23**, 540.

Note added in proof: As further experiments showed, form B can also exist in low molecular weight peptides, from hexapeptides up to peptides containing 15–18 residues. It has a structure close to that of polyproline II, with straight left-hand helices, but forms triple-chain aggregates with hydrogen bonding arranged like that in collagen II. A detailed description of this form will soon be published. Sedimentation data presented here are concerned with more complex molecular aggregates.

Discussion

E. R. BLOUT: In previous publications from your group it has been stated that the polymeric tripeptide (-Gly-Pro-Hypro-)$_n$ has a molecular weight of approximately 25,000. In this paper you now report that the molecular weight of this polymeric tripeptide is less than 2000. Can you tell us the methods used for the molecular weight determinations, which you indicate were performed by chemical means?

N. S. ANDREEVA: The molecular weight reported earlier was determined by the ultracentrifuge method. The data listed in Table 2 were obtained for low and high molecular weight fractions of polymer both by chemical methods (end groups analysis) and ultracentrifuge methods (sedimentation properties and Archibald equilibrium technique). Ultracentrifuge data support the previous result, but the chemical molecular weight is always much lower. It proves that the polymer has a strong tendency to form multichain aggregates.

W. TRAUB: I would like to report in this connection that we have been investigating the structure of the ordered copolymer poly-(L-prolyl-glycyl-glycine). It has a hexagonal unit cell and an axial repeat of 9·2 Å per tripeptide, and thus appears to have a structure similar to that of polyproline II and polyglycine II rather than the coiled structure of collagen.

N. S. ANDREEVA: I suppose it is necessary to have hydroxyproline residues to form or induce the forming of a collagen-like structure.

A. J. HODGE: In view of the possible complications introduced by the hydrogen bonding capacity of hydroxyproline and since hydroxylation occurs only after synthesis, have you studied the behaviour of poly-(-Gly-Pro-Pro-)?

N. S. ANDREEVA: We investigated the other polymers related to collagen, but the whole data are preliminary because effective methods of their synthesis and crystallization have not yet been developed. The polymer (-Gly-Pro-Pro-)$_n$ did not give a collagen-like X-ray pattern.

SECTION II
Genetic Code

Synthetic Polynucleotides and the Genetic Code

SEVERO OCHOA

Department of Biochemistry, New York University School of Medicine, New York, U.S.A.

ABSTRACT

Further study of the effect of synthetic polyribonucleotides, prepared with polynucleotide phosphorylase, has shown that, contrary to previous belief, the homopolynucleotides, poly A and poly C are meaningful and stimulate the incorporation into protein-like products of the amino acids lysine and proline, respectively. These polynucleotides are without effect on the incorporation of any of the remaining nineteen amino acids. The code triplets AAA and CCC can therefore definitely be assigned to lysine and proline. There are indications that polylysine is the product of lysine incorporation in the presence of poly A by the *E. coli* system. With use of copolynucleotides rich in either adenylic or cytidylic acid residues, but containing no uridylic acid residues, non-U containing code triplets have been found for a number of amino acids bringing the number of code triplets to date to forty-one out of sixty-four triplets in RNA. The amino acid code thus appears to be extensively degenerate and there are no indications that the code triplet list is as yet complete. To date the base composition of the code is A, 28·5%; U, 26·8%; C, 23·6%; and G, 21·1%; A + U = 55·3%, G + U = 44·7%. Roughly A = U, and C = G. The high U content of the code had been a matter for speculation until non-U containing triplets were found. It had always appeared likely that the existence of a number of non-U code triplets would explain the apparent anomaly.

A beginning has been made in the determination of base sequence of some code triplets and, from the results of experiments with polymers containing triplets of known sequence at one end of a poly U chain, the sequences GUU, and AUU have tentatively been assigned to the amino acids cysteine and tyrosine, respectively. It is to be noted that, to date, only one code triplet has been found for each of these two amino acids.

Results of experiments with synthetic polyribonucleotides containing uridylic acid, which led to the assignment of code triplets to each of nineteen amino acids, have been previously reviewed (Ochoa, 1963; Speyer *et al.*, 1962*b*). This work shed light on the base composition, but not the base sequence, of nucleotide triplets all of which contained one or more uridylic acid residues. More recent experiments, to be reported in this paper, shed some light on the base sequence of the code triplets for the amino acids tyrosine and cysteine. Recent experiments have also shown that synthetic polynucleotides containing no uridylic acid can

stimulate the incorporation of amino acids into protein-like products. This has led to the assignment of additional code triplets to a number of amino acids bringing the total number of code triplets to date to forty-one, out of sixty-four triplets in RNA. Of these triplets twenty-three contain uridylic acid while the remaining eighteen contain no uridylic acid. The amino acid code would thus appear, in agreement with suggestions from genetic experiments (Crick, 1963) to be extensively degenerate. It is of interest to note that in addition to poly U, which promotes the incorporation of phenylalanine and leads to the synthesis of polyphenylalanine (Lengyel *et al.*, 1961; Nirenberg and Matthaei, 1961), two other homopolymers have now definitely been shown to be active in the cell-free *E. coli* system of amino acid incorporation. Poly A stimulates the incorporation of lysine, and of no other amino acid, leading to the synthesis of polylysine, and poly C stimulates the incorporation of proline, and of no other amino acid, leading presumably to the synthesis of polyproline (Gardner *et al.*, 1962; Wahba *et al.*, 1963).

1. Base Sequence of Code Triplets

An important task in the studies of the genetic code is the determination of the base sequence of the code triplets. This is no easy task and it may take a long time. The simplest approach to the problem is to place triplets of known sequence at one end of a homopolymer chain, e.g. at one end of the poly U chain. This is made possible by the finding of Heppel and collaborators that short oligonucleotides prime polynucleotide phosphorylase by acting as nuclei for growth of the polynucleotide chains (Singer *et al.*, 1960). By hydrolysis of poly UA (5 : 1) and poly UG (5 : 1) with pancreatic ribonuclease, we have isolated mixtures of tri- and dinucleotides which in the main consist of ApApUp, ApUp, and GpGpUp, GpUp units, respectively. After removing the terminal phosphate residue with phosphomonoesterase, the oligonucleotides (ApApU together with ApU in one case and GpGpU together with GpU in the other) were used as primers for the synthesis of poly U with *Azotobacter* polynucleotide phosphorylase. The AU primers yielded mixtures of ApApUpUpU.....pU and ApUpUpU.....pU polymers; they will be referred to as AUU...U. The GU primers yield mixtures of GpGpUpUpU.....pU and GpUpUpU.....pU polymers; they will be referred as GUU...U.

In the *E. coli* amino acid incorporation system (Wahba *et al.*, 1962) AUU...U promoted the incorporation of phenylalanine and of small amounts of tyrosine but little or no isoleucine. No incorporation of asparagine or lysine was detected with this polymer (Table I). In preliminary experiments GUU...U promoted the incorporation of

phenylalanine and of some amounts of cysteine but not that of valine, lysine or tryptophan. Since the code triplet for tyrosine contains 2U1A and that for cysteine 2U1G, it would appear from this experiment that the tyrosine triplet is AUU and the cysteine triplet GUU. Determination of the position of the incorporated amino acid in the polypeptide chain is made difficult by the high insolubility of the phenylalanine-rich polypeptides made by this system under the direction of polynucleotides rich in uridylic acid residues. However, end group assays have indicated (Wahba *et al.*, 1962) that while some of the phenylalanine incorporated in

TABLE I. Effect of poly U, poly AUU...U, and poly UA on the incorporation of various amino acids in *E. coli* system†

	Amino acid§				
Polynucleotide‡	Phenylalanine	Isoleucine	Tyrosine	Asparagine	Lysine
None	600	22	25	58	35
U	22,000	29	35	62	26
AUU...U	20,000	29	63	63	27
UA (4:1)	9400	1400	1540	500	200

† Values are expressed in $\mu\mu$moles/mg ribosomal protein. They are averages of at least three (phenylalanine) or two (isoleucine, tyrosine) duplicate experiments.

‡ 20 μg of poly U and 40 μg of each poly AUU...U and poly UA per sample.

§ Per sample, isoleucine, tyrosine, and lysine, 12·5 mμmoles; phenylalanine and asparagine, 50 mμmoles.

the presence of either poly U or poly AUU...U is present in an N-terminal position, as detected by hydrazynolysis, no tyrosine is present in the N-terminal position in the polypeptide synthesized in the presence of the latter polymer while some was found, by use of the dinitrofluorobenzene method, in a C-terminal position. These results suggest that the nucleotide code is read from right to left, if the convention used above for writing polynucleotide chains is followed, or in other words that the code is read beginning at the end of the polynucleotide template chain bearing two unesterified hydroxyl residues at positions 2′ and 3′ of the ribose moiety. This follows from the finding (Bishop *et al.*, 1960; Dintzin, 1961; Goldstein, 1961) that synthesis of the polypeptide chain starts with the N-terminal amino acid.

Studies of the base sequence of code triplets are still in a preliminary phase and are being actively pursued. On the assumption that AUU is the code triplet for tyrosine and GUU that for cysteine and with use of the now extensive amino acid replacement data occurring as a result of spontaneous or induced mutations in haemoglobin, TMV protein, and other proteins, Jukes (1962) has proposed tentative sequences for all of the U-containing code triplets deduced from previous work. However, with the possibility of extensive degeneracy of the code suggested by the more recent work, sequence assignments based on amino acid replacement data cannot be made with any degree of assurance.

2. Stimulation of Amino Acid Incorporation by Polynucleotides containing no Uridylic Acid

While studying the effect of synthetic polynucleotides on amino acid incorporation by a cell-free rat liver system, it was observed that poly A consistently produced a small stimulation of the incorporation of lysine into products insoluble in trichloroacetic acid. As no such effect had previously been noted with *E. coli* preparations, this matter was re-investigated. Polylysine is soluble in trichloroacetic acid (Sela and Katchalski, 1959) and it was possible that poly A-promoted synthesis of this polypeptide might have escaped detection in earlier experiments with use of trichloroacetic acid as the protein precipitating agent. Since polylysine is insoluble in tungstic acid (Sela and Katchalski, 1959), the effect of poly A on the incorporation of lysine in the *E. coli* system was therefore studied with the use of a mixture of trichloroacetic and tungstic acids as the precipitating reagent (Gardner *et al.*, 1962). Under these conditions poly A consistently promoted a marked incorporation of [^{14}C]lysine into trichloroacetic-tungstic acid insoluble material. As previously found for the poly U-mediated incorporation of phenylalanine, the poly A-dependent lysine incorporation was dependent on the presence of transfer RNA and an ATP-generating system but did not require the presence of other amino acids. This suggested formation of a lysine homopolypeptide under poly A direction. The incorporation of lysine was inhibited by puromycin and by chloramphenicol. Since the product formed from lysine in the presence of poly A is soluble in water and aqueous solvents, it was readily identified as a polypeptide with use of proteolytic enzymes (Gardner, *et al.*, 1962). Incubation with trypsin resulted in the almost complete disappearance of the tungstic acid-insoluble radioactive material formed by the *E. coli* system on incubation with [^{14}C]lysine in the presence of poly A. This was not the case following incubation with chymotrypsin. Since polylysine is susceptible to trypsin but very resistant to chymotrypsin (Katchalski, 1951), these results

suggest that poly A directs the synthesis of poly-L-lysine in cell-free systems of protein synthesis.

As mentioned in the introduction it was also recently found that poly C stimulates the incorporation of proline into protein-like products (Wahba *et al.*, 1963). Following the finding of stimulation of lysine incorporation by poly A it appeared desirable to reinvestigate whether or not poly C stimulates the incorporation of proline. It may be remembered that poly C had been reported to be active (Nirenberg and Matthaei, 1961) for proline incorporation although later this activity appeared to be explainable by the presence of some uridylic acid residues in the poly C preparations used (Matthaei *et al.*, 1962). In our experience (Speyer *et al.*, 1962a), stimulation of proline incorporation by poly C was very small while copolymers containing both U and C were much more effective. Reinvestigation of the problem was made even more desirable by indications that poly C, in contrast to poly U, had low affinity for the ribosomes. Thus, while poly U effectively competed with poly UC and markedly decreased stimulation of serine incorporation by the latter polymer, poly C decreased this incorporation to a very slight extent. Poly C also decreased very slightly, if at all, the poly U promoted incorporation of phenylalanine. In these experiments all polymers were used at the concentration 160 μg/ml.

In view of the above it was possible that, in previous experiments, poly C might have been essentially inactive not because of intrinsic meaninglessness of CCC triplets but because of the low affinity of the polymer for the ribosomes. This view was substantiated by experiments showing a marked polymer concentration dependence for proline incorporation in the presence of poly C. High concentrations of poly C (800 μg/ml), effectively promoted the incorporation of proline. It was further found that when 5% trichloroacetic acid was used as the protein precipitating agent, as in previous work, only about half as many counts were recovered as acid-insoluble material when [^{14}C]proline was incorporated in the presence of poly C. To ensure complete precipitation of polyproline, synthetic polyproline was added as a carrier. These results show that previous negative results with poly C were due mainly to low affinity for the ribosomes and, to a lesser extent, to partial solubility of polyproline in 5% trichloroacetic acid. The above experiments definitely allow the assignment of an AAA code letter to lysine, and a CCC code letter to proline. With techniques available to detect the formation of lysine- and proline-rich polypeptides by cell-free systems of protein synthesis it was possible to search for non-U code triplets by the use of synthetic polyribonucleotides containing no uridylic acid but rich in either adenylic or cytidylic acid residues. It will be remembered that U-containing code triplets were determined with use of polynucleotides rich in uridylic acid

residues so that formation of even relatively small polypeptides could be detected due to the high insolubility of polyphenylalanine since polypeptides rich in uridylic acid residues would yield polypeptides with a number of unbroken phenylalanine sequences.

TABLE II. Amino acid incorporation in *E. coli* system with various polynucleotides†

	Polynucleotide							
Amino acid	A	AU (5:1)	AC (5:1)	AG (5:1)	ACG (4:1:1)	ACG (6:1:1)	C	CI (5:1)
Ala	0	0	0	0	0·12	0·06	0	0·11
Arg	0	0	0	0·05	0·55	0·35	0	0·09
AspN	0	0·13	0·30	0	0·35	0·39	0	0
Asp	0	0	0	0	0·08	0·06	0	0
Cys	0	0	0	0	0	0	0	0
Glu	0	0	0	0·11	0·43	0·47	0	0
GluN	0	0	0·44	0·02	0·62	0·64	0	0
Gly	0	0	0	0·02	0·07	0·03	0	0·02
His	0	0	0·09	0	0·26	0·32	0	0
Ile	0	0·10	0	0	0	0	0	0
Leu	0	0	0	0	0	0	0	0
Lys	1·2	0·47	0·99	0·36	2·06	2·92	0	0
Met	0	0	0	0	0	0	0	0
Phe	0	0	0	0	0	0	0	0
Pro	0	0	0·05	0	0·18	0·08	0·72	0·44
Ser	0	0	0	0	0·16	0·10	0	0
Thr	0	0	0·23	0	0·49	0·57	0	0·07
Try	0	0	0	0	0	0	0	0
Tyr	0	0·02	0	0	0	0	0	0
Val	0	0	0	0	0	0	0	0

† mμmoles/mg ribosomal protein. Blanks without polynucleotide subtracted. Data with poly A, AU, AC, and AG, and with poly ACG, C, and CI, are from separate experiments. Experimental details as previously described (Gardner *et al.*, 1962; Wahba *et al.*, 1963).

Each of twenty amino acids was tested with and without the addition of poly A, poly AU, poly AC, poly AG, poly ACG, poly C, and poly CI (Gardner *et al.*, 1962; Wahba *et al.*, 1963). Poly CI was used in place of poly CG for, as previously shown (Basilio *et al.*, 1962), inosinic acid can replace guanylic acid in coding. The results of these experiments are summarized in Table II. They led to the assignment of additional (non-U) code triplets as follows: alanine, 1C1A1G, 2C1G; arginine, 1G2A, 1G2C;

asparagine, 1C2A; aspartic acid, 1G1C1A; glutamic acid, 2A1G; glutamine, 1A2G, 2A1C; glycine, 1A2G, 1C2G; histidine, 1A2C; isoleucine, 2A1U; lysine, AAA; proline, CCC, 1A2C; serine, 1A1C1G; threonine, 2A1C, 2C1G. Bretscher and Grunberg-Manago (1962) had reported stimulation of the incorporation of glutamine, histidine, proline, and threonine, by AC-containing copolymers. Stimulation of the incorporation of several amino acids by non-U copolymers, with use of conventional techniques, was also recently reported by Nirenberg (1962).

3. Code Triplet Assignments

Table III presents a summary of the effects of all the polynucleotides tested to date in our laboratory on the incorporation of each of twenty amino acids. The polynucleotides are divided in three groups, uridylic acid-rich, adenylic acid-rich, and cytidylic acid-rich. The figures in parentheses under each copolynucleotide give the proportion of the corresponding nucleoside-5′ diphosphates used in the preparation of the polymers. As shown in Table IV these proportions are closely reflected in the actual base composition of the polymers. The base composition of poly ACG and poly CI has not yet been determined but from the results with the other polymers it would be expected to mirror closely the ratios of nucleoside diphosphates used in the preparation of these polymers. The values in Table III are relative values giving the net stimulation of incorporation of amino acids by the various polymers relative to the incorporation, taken as 100, of phenylalanine, lysine, or proline, by the U-rich, A-rich, and C-rich polymers, respectively. In each of these three groups of polymers, the incorporation of phenylalanine, lysine, or proline, is stimulated maximally because of the higher frequency of UUU, AAA, and CCC triplets relative to other triplets (cf. Table II for lysine and proline, and Tables II and III of a previous publication (Ochoa, 1962) for phenylalanine). In making code triplet assignments the stimulation by a polymer of the incorporation of a given amino acid relative to that of phenylalanine (U-rich polymers), lysine (A-rich polymers), and proline (C-rich polymers), was matched to the calculated frequency of a given triplet relative to that of the UUU, the AAA, or the CCC triplet in this polymer for the U-rich, A-rich, and C-rich group of polymers, respectively. The calculation of triplet frequencies is based on two assumptions, (a) that the proportions of bases in the polymers are the same as the proportions of the corresponding nucleoside diphosphates used in their preparation, and (b) that the nucleotides are randomly distributed in the polymers. Evidence supporting the first assumption has just been presented. The second assumption namely random distribution of the nucleotides in the copolymers prepared with polynucleotide

TABLE III. Effect of synthetic polynucleotides on the incorporation of amino acids into protein in cell-free *E. coli* system†

Amino acid	Polynucleotide													
	U-rich							A-rich					C-rich	
	U	UA (5:1)	UC (5:1)	UG (5:1)	UAC (6:1:1)	UCG (6:1:1)	UAG (6:1:1)	A	AU (5:1)	AC (5:1)	AG (5:1)	ACG (4:1:1)	C	CI (5:1)
Ala	0	0	0	0	0	3	0	0	0	0	0	6	0	22
Arg	0	0	0	0	0	3	0	0	0	0	13	27	0	19
AspN	0	7	0	0	7	0	5	0	28	30	0	17	0	0
Asp	0	0	0	0	0	0	3	0	0	0	0	4	0	0
Cys	0	0	0	20	0	25	32	0	0	0	0	0	0	0
Glu	0	0	0	0	0	0	2	0	0	0	30	21	0	0
GluN	0	0	0	0	—	0	—	0	0	44	6	30	0	0
Gly	0	0	0	4	0	3	0	0	0	0	5	3	0	5
His	0	0	0	0	3	0	0	0	0	9	0	13	0	0
Ile	0	20	0	0	16	0	32	0	20	0	0	0	0	0
Leu	0	14	20	13	25	25	27	0	3	0	0	0	0	0
Lys	0	3	0	0	2	0	0	100	100	100	100	100	0	0
Met	0	0	0	0	0	0	4	0	0	0	0	0	0	0
Phe	100	100	100	100	100	100	100	0	0	0	0	0	0	0
Pro	0	0	8	0	3	3	0	0	0	5	0	9	100	100
Ser	0	0	25	0	25	26	0	0	0	0	0	8	0	0
Thr	0	0	0	0	9	0	0	0	0	23	0	24	0	16
Try	0	0	0	5	0	4	—	0	0	0	0	0	0	0
Tyr	0	25	0	0	25	0	20	0	3	0	0	0	0	0
Val	0	0	0	20	0	20	25	0	0	0	0	0	0	0

† In each of the three groups the experimental results are expressed as per cent of the incorporation of the one amino acid, e.g. phenylalanine, lysine, proline, which is incorporated maximally. Blank spaces, no effect (zero).

TABLE IV. Base ratios of synthetic polynucleotides

Polynucleotide	Ratio of nucleoside diphosphates in reaction mixture	Base ratios of isolated polynucleotides
Poly UA	5:1	4·8:1
Poly UC	5:1	4·7:1
Poly UG	5:1	5·2:1
Poly AC	5:1	4·9:1
Poly AG	5:1	4·9:1
Poly AU	5:1	4·7:1
Poly UAC	6:1:1	6:1·25:0·9
Poly UAG	6:1:1	6:1·3:1
Poly UCG	6:1:1	6:0·6:1

TABLE V. Examples of code triplet assignments

Polynucleotide	Triplets	Frequency of each triplet (%)	Amino acid incorporation (%)	Code triplet composition
UG (5:1)	UUU	100	Phe, 100	UUU
	UUG, UGU, GUU	20	Cys, 20; Val, 20	2U1G
	UGG, GUG, GGU	4	Gly, 4; Try, 4	1U2G
	GGG	0·8		
AC (5:1)	AAA	100	Lys, 100	AAA
	AAC, ACA, CAA	20	AspN, 30; Thr, 23	2A1C
	ACC, CAC, CCA	4	Pro, 5	1A2C
	CCC	0·8		
CG (5:1)†	CCC	100	Pro, 100	CCC
	CCG, CGC, GCC	20	Ala, 22; Arg, 19	2C1G
	CGG, GCG, GGC	4	Gly, 5	1C2G
	GGG	0·8		

† Poly CI used in place of poly CG.

phosphorylase is based on the results of nearest neighbour frequency studies with poly AU (1:1) (Heppel *et al.*, 1957) and polyAGUC (1:1:1:1) (Ortiz and Ochoa, 1959).

The matching of triplet frequency and amino acid incorporation in making code triplet assignments is shown by the examples given in Table V for one copolymer of each group U-rich (poly UG (5:1)), A-rich

TABLE VI. Amino acid code triplets

Amino acid	U-triplets†	Non-U triplets	Shared doublets
Ala	CUG	CAG, CCG	C °G
Arg	GUC	GAA, GCC	G °C
AspN	UAA, CUA	CAA	°AA, C °A
Asp	GUA	GCA	G °A
Cys	GUU	...	...
Glu	AUG	AAG	A °G
GluN	...	AGG, AAC	...
Gly	GUG	GAG, GCG	G °G
His	AUC	ACC	A °C
Ile	UUA, AAU‡	...	...
Leu	UAU, UUC, UGU	...	U °U
Lys	AUA	AAA	A °A
Met	UGA	...	...
Phe	UUU	...	...
Pro	CUC	CCC, CAC	C °C
Ser	CUU	ACG	...
Thr	UCA	ACA, CGC	°CA
Try	UGG	...	...
Tyr	AUU	...	...
Val	UUG	...	...

† Sequence from Jukes (1962).
‡ Not in Jukes' list.

(AC (5:1)), and C-rich (CG (5:1)). The calculation of frequencies has been explained previously (Ochoa, 1962); the percent amino acid incorporation values have been taken from Table III.

A summary of the code triplet assignments to date is shown in Table VI. The U-containing triplets have been arranged in the sequences proposed by Jukes (1962). When possible, the sequence of non-U triplets was fitted to that of U-triplets as if triplets for a given amino acid were derived from each other through a single base replacement, e.g. GAG and GCG for an assumed GUG sequence of the U-containing lysine triplet. When this was not possible, non-U sequences for an amino acid were

written to avoid duplication with sequences of the same base composition for another amino acid. Since, as already discussed, sequences derived from amino acid replacement data are not reliable if the code is extensively degenerate, all sequences in Table VI are arbitrary, except for GUU and AUU for cysteine and tyrosine, respectively. These are based as already mentioned on direct experimental determination. Table VI includes forty-one out of sixty-four triplets in RNA, and the list is probably still incomplete. It may be mentioned that recent experiments with poly UCG (6 : 1 : 1) failed to substantiate the prediction (Speyer *et al.*, 1962a) that 1U1C1G is the code triplet for glutamine. Hence, 1U1C1G has been excluded from the list as a glutamine triplet. Also, previous results suggesting 1U2C as a code triplet for threonine (Speyer *et al.*, 1962a) could not be confirmed, and this has been removed from the list as a threonine triplet.

It may be noted that, in many cases (Table VI, column 4) a doublet is shared, with the same relative position of its bases, by two or three triplets of the same amino acid. It might be that in these cases the chemical nature of only two bases in the triplet is meaningful, that of the third base being immaterial except as a position sign as indicated by the circle in column 4. This possibility bears some resemblance to Roberts' doublets (Roberts, 1962). In this case the code for alanine, arginine, aspartic acid, glutamic acid, lysine, histidine, proline, and threonine, might be more degenerate and the extent of degeneracy of the amino acid code as a whole would be correspondingly decreased. It remains to be seen whether, in these cases, one or several transfer RNA's are involved in the read out of the code. Weissblum, Benzer and Holley (1962) have reported on two leucine transfer RNA's corresponding to the 2U1G and 2U1C code triplets to this amino acid.

4. Universality of Code

Arguments supporting the view that the genetic code is universal, that is, one for all living things, have been previously presented (Ochoa, 1962; Speyer *et al.*, 1962b). Additional evidence in favour of this view has been obtained in experiments in several laboratories on the effect of synthetic polyribonucleotides on the incorporation of amino acids into acid-insoluble products by cell-free preparations from sources other than bacterial cells. The results of our laboratory (Gardner *et al.*, 1962), shown in Table VII, with a cell-free system from rat liver are in good agreement with those reported by Weinstein and Schechter (1962) and Maxwell (1962) and show that *E. coli* and rat liver systems share code triplets for those amino acids so far investigated with the latter system. Similar

results have been obtained with plasma tumor cells and *Chlamydomonas* preparations (Weinstein and Sager, 1962).

TABLE VII. Amino acid incorporation in rat liver system with various polynucleotides

Amino acid	Polynucleotide†								
	None	A	U	UA (5:1)	UC (5:1)	UG (5:1)	Found‡	Calculated§	Code triplet
Phenylalanine	27	33	778	245	370	346	100	—	UUU
Isoleucine	26	—	—	56	—	—	14	20	2U1A
Lysine	15	55	—	15	—	—	—	—	AAA
Serine	13	—	—	—	59	17	14	20	2U1C
Tyrosine	55	—	—	101	—	—	21	20	2U1A
Valine	26	—	—	—	28	134	34	20	2U1G

† Amino acid incorporation given in $\mu\mu$moles/mg ribosomal protein.

‡ Amino acid incorporation given as percentage of the phenylalanine incorporation (blanks without polynucleotide subtracted).

§ Per cent frequency relative to UUU of triplet most closely matching percentage amino acid incorporation relative to phenylalanine.

REFERENCES

Basilio, C., Wahba, A. J., Lengyel, P., Speyer, J. F. and Ochoa, S. (1962). *Proc. nat. Acad. Sci., Wash.* **48**, 613.

Bishop, J., Leahy, J. and Schweet, R. (1960). *Proc. nat. Acad. Sci., Wash.* **46**, 1030.

Bretscher, M. S. and Grunberg-Manago, M. (1962). *Nature, Lond.* **195**, 283.

Crick, F. H. C. (1963). *Progress in Nucleic Acid Research.* Academic Press, New York, in press.

Dintzin, H. M. (1961). *Proc. nat. Acad. Sci., Wash.* **47**, 247.

Gardner, R. S., Wahba, A. J., Basilio, C., Miller, R. S., Lengyel, P. and Speyer, J. F. (1962). *Proc. nat. Acad. Sci., Wash.* **48**, 2087.

Goldstein, A. (1961). *Biochim. biophys. Acta* **53**, 468.

Heppel, C. A., Ortiz, P. J. and Ochoa, S. (1957). *J. biol. Chem.* **229**, 695.

Jukes, T. H. (1962). *Proc. nat. Acad. Sci., Wash.* **48**, 1809.

Katchalski, E. (1951). In *Advances in Protein Chemistry*, Vol. 6, p. 123. Academic Press, New York.

Lengyel, P., Speyer, J. F. and Ochoa, S. (1961). *Proc. nat. Acad. Sci., Wash.* **47**, 1936.

Matthaei, J. H., Jones, O. W., Martin, R. G. and Nirenberg, M. W. (1962). *Proc. nat. Acad. Sci., Wash.* **48**, 666.

Maxwell, E. S. (1962). *Proc. nat. Acad. Sci., Wash.* **48**, 1639.

Nirenberg, M. W. (1962). Reported at Symposium on Informational Macromolecules. Rutgers—The State University, New Brunswick, New Jersey, September, 1962.

Nirenberg, M. W. and Matthaei, J. H. (1961). *Proc. nat. Acad. Sci., Wash.* **47**, 1588.

Ochoa, S. (1963). Symposium on Genetic Mechanisms, *Fed. Proc.* **22**, Part I, 62.

Ortiz, P. J. and Ochoa, S. (1959). *J. biol. Chem.* **234**, 1208.

Roberts, R. B. (1962). *Proc. nat. Acad. Sci., Wash.* **48**, 897.

Sela, M. and Katchalski, E. (1959). In *Advances in Protein Chemistry* M. L. Anson and J. T. Edsall, eds.), Vol. 14, p. 391. Academic Press, New York.

Singer, M. F., Heppel, C. A. and Hilmoe, R. J. (1960). *J. biol. Chem.* **235**, 738.

Speyer, J. F., Lengyel, P., Basilio, C. and Ochoa, S. (1962*a*). *Proc. nat. Acad. Sci., Wash.* **48**, 63.

Speyer, J. F., Lengyel, P., Basilio, C. and Ochoa, S. (1962*b*). *Proc. nat. Acad. Sci., Wash.* **48**, 441.

Wahba, A. J., Basilio, C., Speyer, J. F., Lengyel, P., Miller, R. S. and Ochoa, S. (1962). *Proc. nat. Acad. Sci., Wash.* **48**, 1683.

Wahba, A. J., Gardner, R. S., Basilio, C., Miller, R. S., Speyer, J. F. and Lengyel, P. (1963). *Proc. nat. Acad. Sci., Wash.* **49**, 116.

Weinstein, I. B. and Sager, R. (1962). Presented at Symposium on Informational Macromolecules. Rutgers—The State University, New Brunswick, New Jersey, September, 1962.

Weinstein, I. B. and Schechter, A. N. (1962). *Proc. nat. Acad. Sci., Wash.* **48**,1686.

Weissblum, B., Benzer, S. and Holley, R. W. (1962). *Proc. nat. Acad. Sci., Wash.* **48**, 1449.

Discussion

E. KATCHALSKI: I am very glad that the polylysine and polyproline, which were synthesized by us several years ago, find use in the elucidation of the genetic code.

It is worth mentioning in this connection that one of the reasons for my synthesis of polylysine and polyarginine was the suggestion of Kossel many years ago that, from the evolutionary point of view, polyarginine was the simplest of proteins formed. The elegant experiments presented by Professor Ochoa perhaps indicate that Kossel's theory might have some truth in it.

Recently we have succeeded in the preparation of a water-insoluble ribonuclease preparation. We are thus able to digest RNA and remove the insoluble ribonuclease by centrifugation or filtration. I wonder whether such an enzyme preparation would be of use to you?

S. OCHOA: Such an enzyme will be very valuable.

E. KATCHALSKI: Have you determined whether the polyproline formed in the presence of poly C is in the form I or form II?

S. OCHOA: We do not have any idea. The only thing we know is that it was insoluble in tungstic acid.

G. N. RAMACHANDRAN: I want to ask about the "adaptor" mechanism. Professor Wilkins believes that it is due to the three residues present in the bend of the RNA molecule. Since there should be two places of attachment, one to the messenger-RNA and the other to the amino acid, and since the transfer RNA must be attached to the amino acid, it looks that there should be a specific pairing occurring in the system.

S. OCHOA: The amino acid is attached to the adenosine residue, at the end of the transfer-RNA chain, through a covalent linkage. An aminoacyl-RNA is thereby formed. The "adaptor" would appear to be exclusively concerned with attachment of this molecule to a complementary base sequence in messenger-RNA through a base-pairing mechanism.

Proteins, Coding and Medicine

L. D. HAMILTON

Sloan-Kettering Institute for Cancer Research, New York, New York, U.S.A.

ABSTRACT

Three subjects in medicine are concerned with proteins and coding. The first is hereditary disease, where abnormality of one gene may cause defective synthesis of a single protein. The classic example of enzymic deficiency as primary biochemical defect is absence of phenylalanine hydroxylase in phenylketonuria. The abnormal haemoglobins provide additional examples.

The second subject concerns the recognition of alien material and includes immune-state problems of homotransplantation and diseases of auto-immunity. The weight of evidence suggests that the recognition by the body of alien proteins is a general phenomenon initiated probably by the foreign cell protein, the main recognition being a property of lymphocytes. Auto-immune diseases may represent erratic associations between exogenous chemicals or other agents and self-cell membranes that are then treated as foreign and so induce antibodies against their own tissues.

Finally, the third subject is cancer. The cancer cell may be considered as a mistake—presumably a somatic mutation—but it does not produce cells recognizably foreign; if anything, there appears to be a diminished specific-protein synthesis in such malignant transformations, reflecting de-differentiation. Cancer cells do not provoke the immune and connective tissue response that normally walls off foreign bodies. The cancer cell represents a cell released from repression mechanisms in which replication becomes potentially infinite. Besides having lost specific proteins concerned with recognition, it presumably has also lost growth-repressive proteins, in contrast, say, to a cell of normally regenerating liver, which becomes increasingly inhibited by the liver-specific regulatory principle found in liver itself.

1. INTRODUCTION

It might be of interest to the specialists here to review the relationship of recent developments in molecular biology to problems in medicine, and cover especially the significance of these developments for hereditary disease, immunogenetic mechanisms, and cancer.

The central dogma in contemporary biology can briefly be summarized by the following familiar:

$$\text{DNA} \rightarrow \text{RNA} \rightarrow \text{Protein}$$

The arrows indicate the flow of information. The presumption is that the

genetic information coded in the base-sequence of DNA is ultimately transcribed into the amino acid sequence of protein by a polyribonucleotide intermediary. Those of us who have ready access to New York newspapers have no difficulty in keeping abreast of, and indeed in some instances (Osmundsen, 1962) keeping ahead of, this field, as Fig. 1, taken from a Sunday newspaper seven years ago, clearly shows. For those not having any access to this fount of learning, I recapitulate the main points. It is believed that most protein synthesis in the cell takes place in the cytoplasm outside the nucleus. The actual sites are probably the

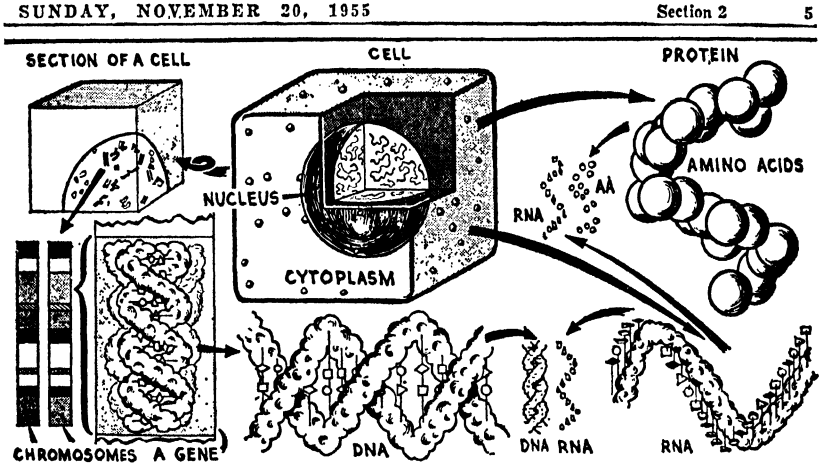


FIG. 1. DNA→RNA→Protein (Reproduced by kind permission of the *New York Herald Tribune.*)

small particles—ribosomes—consisting of RNA protein. Recent work suggests that the special RNA called messenger RNA is synthesized in the nucleus as a copy of DNA in a way very similar to the way DNA itself is replicated; as the two chains of DNA separate, a copy of the information on the DNA is made in the messenger RNA by the proper base-pairing. This RNA goes into the ribosomes where it acts as the actual template for the protein synthesis.

How does the messenger RNA arrange the amino acids in the correct order? The amino acids cannot by themselves recognize the correct base-sequence of RNA. This is done by each amino acid being provided with an adapter. This adapter is another kind of RNA, known as transfer RNA. A special activating enzyme which can recognize the amino acid is used to join the amino acid on to the special transfer RNA. There is at least one special transfer RNA and a special activating enzyme for each

of the 20 amino acids. Transfer RNA, with the amino acid, goes to the ribosomes, recognizes the proper triplet of bases on the messenger RNA by forming base-pairs between its bases and the bases of the messenger, and this gets the amino acid into the right place.

This dogma, so palatable to contemporary biologists, has medical implications, and indeed, I believe, much of medicine can be reclassified as to diseases associated with one or other of these three classes of macromolecules, DNA, RNA and protein.

2. Hereditary Diseases

The hereditary diseases, of which there are many examples, reflect DNA defects. I shall speak here of only two.

(a) *Phenylketonuria*

Phenylketonuria was the first hereditary disease in which the classical relationship between gene, enzyme and clinical abnormality postulated by Garrod was unequivocally demonstrated 25 years later. The error is

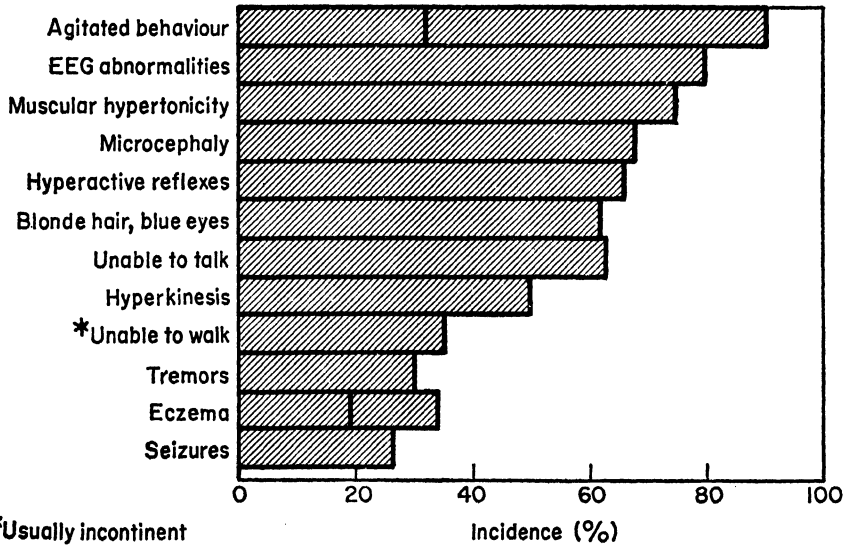


Fig. 2. Major clinical findings in phenylketonuria.

inability to oxidize phenylalanine to tyrosine, reflecting the absence of phenylalanine hydroxylase in the liver of patients with this disease. To develop this disease, both of a patient's parents must have had a defect in one of the two genes controlling phenylalanine hydroxylase. A reduction in phenylalanine hydroxylase activity can be detected in the parents,

but there is no disease as a result of this inadequacy. On an average, one out of four offspring from two heterozygous parents has both genes defective; in these offspring there is no active phenylalanine hydroxylase and phenylketonuria results.

Disability first becomes apparent several weeks after birth; it begins with an elevation of plasma phenylalanine to 30 times normal level and by excretion of phenylpyruvic acid. After 6 months, mental development is obviously retarded, often accompanied by seizures and other neurological abnormalities, deficient pigmentation of hair and skin, and eczema. In older children and in adults the process remains stationary, but life-expectancy is reduced. Most phenylketonurics are imbeciles; only a few have passable intellectual development. Strikingly, then, lack of a single enzyme causes a disease affecting many systems (Fig. 2), notably the central nervous system. The immediate incitant of the symptoms is the accumulation of by-products of the alternate metabolic pathways that phenylalanine follows in the absence of phenylalanine hydroxylase.

Present treatment with some success consists of limiting phenylalanine intake; the major biochemical abnormalities of phenylketonuria are reversed.

(b) *The human haemoglobins*

The second example is provided by the haemoglobinopathies. The inheritance of haemoglobins follows strict genetic patterns and certain diseases of the blood may reflect quite precisely the genetic pattern of the constituent haemoglobins. A familiar example is sickle-cell haemoglobin. This can exist associated with a profound disease, sickle-cell anaemia, and in other individuals as a harmless trait. There are differences in the tendency to sickle on deoxygenation between the cells of patients with sickle-cell anaemia and those with the trait. The cells of those with the disease have a shorter life-span than normal cells, whereas the trait cells have a normal survival time. The sickling phenomenon is familial, and the mode of the inheritance is that of a single Mendelian dominant.

At first it was generally assumed that the same gene produced in some persons the sickling trait and in others the severe form of the disease with anaemia and other changes. It now appears that the sickle-cell anaemia is present in patients who are homozygous for the abnormal gene; the red cells contain only sickle-cell haemoglobin. In contrast, the carriers of the sickle-cell trait are heterozygous, and in their blood both sickle-cell and adult haemolgobin are found. Occasionally foetal haemoglobin is found in patients homozygous for sickling.

The findings in sickle-cell disease can be explained by the physical properties of reduced sickle haemoglobin, and especially by its low

solubility. Many of the clinical findings are due to the clogging of small blood vessels with the gelling of haemoglobin inside the sickle cells.

The abnormality in sickle-cell haemoglobin, as is well known, reflects a change in a single amino acid in the polypeptide chain forming the globin molecule. A glutamic acid is replaced by valine in one peptide. After this discovery, several other single amino acid changes accompanying changes in haemoglobin were reported. It is of interest to compare, as others have done (Jukes, 1962, 1963), the relationship between these single amino acid changes and their relation to the formation of polypeptides in cell-free systems by synthetic polyribonucleotides. If we take into account the relationship between the code triplets and amino acids provided by the experimental work of Nirenberg and associates (Jones and Nirenberg, 1962) and Ochoa and associates (Wahba *et al.*, 1963), for both triplets containing the uracil and the more recent synthetic polyribonucleotides containing adenine, cytosine and guanine, but no uracil, it is possible to relate the single amino acid change to a corresponding base change in "messenger RNA". There is no direct evidence identifying the synthetic polyribonucleotides with messenger RNA. Moreover, the exact sequence of bases in the synthetic polyribonucleotide triplet is unknown, and there is considerable degeneracy, that is, duplication in the code. For these reasons, Table I relating the amino acid change in the different haemoglobins to the corresponding base change in "messenger RNA" is speculative. Nevertheless, for most haemoglobins the single amino acid change does correspond to a single change in base in the messenger RNA, that presumably must also correspond to a single complementary change in the coding DNA. Thus in haemoglobin I, in which aspartic acid replaces lysine, the corresponding base change in messenger RNA is guanine in place of adenine: in haemoglobin Norfolk, the change in the amino acid is aspartic acid replacing glycine; this could be accomplished by a change in the messenger RNA of guanine to either adenine or cytosine. There are examples where the single amino acid change could be accomplished by changes in two bases of messenger RNA; I have restricted consideration to changes in single bases only. As one may infer, the change in sickle-cell haemoglobin from glutamic acid to valine is probably mediated by the change in adenine to uracil in messenger RNA. Three possible single base changes in messenger RNA could be responsible for the mutation glutamic acid to glutamine that occurs in haemoglobin G Honolulu and haemoglobin D β-Punjab, i.e. guanine to cytosine or adenine to guanine, or uracil to guanine. The exact correspondence will undoubtedly be revealed when the sequences of the triplets are known more accurately. Meanwhile, the haemoglobinopathies—for some of these abnormal haemoglobins are in fact accompanied by serious clinical disease—provide a neat illustration of how a single change in an amino

acid corresponds to a single change in one base of messenger RNA and must correspond therefore to a complementary single change in the DNA of the individual with the disease. One can only marvel at the way in which this event is prevented from happening more often in a molecule which, after all, contains at least 10^9 base-pairs in each human cell, and we each contain approximately 10^{14} cells.

TABLE 1. Single amino acid changes in haemoglobins

Haemoglobin	Amino acid change	Corresponding base change in messenger RNA
I (Murayama, 1960)	Lys to Asp	A to G
Norfolk (Gordon *et al.*, 1961)	Gly to Asp	G to A or G to C
Zurich (Muller and Kingma, 1961)	His to Arg	A to G
M Boston (Gerald and Efron, 1961)	His to Tyr	C to U
M Saskatoon (Ingram, 1962)	His to Tyr	C to U
G Philadelphia (Baglioni and Ingram, 1961)	$AspNH_2$to Lys	C to A
C, E (Hunt and Ingram, 1959)	Glu to Lys	G to A
G San Jose (Hill and Schwartz, 1959)	Glu to Gly	A to G
A_2 (Ingram and Stretton, 1961)	Glu to Ala	A to C
S (Ingram, 1957)	Glu to Val	A to U
M Milwaukee (Gerald and Efron, 1961)	Val to Glu	U to A
A_2 (Ingram and Stretton, 1961)	Ser to Thr	U to A
A_2 (Ingram and Stretton, 1961)	Thr to Ser	A to U
A_2 (Ingram and Stretton, 1961)	Thr to $AspNH_2$	C to U
G Honolulu (Swenson *et al.*, 1962)	Glu to $GluNH_2$	G to C or A to G
D β-Punjab (Baglioni, 1962)	Glu to $GluNH_2$	U to G

Parenthetically, as we are near the monazite area in Kerala and Madras which, I understand is approximately 200 km long and several hundred metres wide with an approximate population of 100,000, it would be of interest to ask whether the population here—one exposed to approximately 10 times the normal amount of natural radiation from the high external radiation of radioactivity from soil and rock—shows any evidence of an increased incidence of somatic gene mutations, such as one might expect from increased radiation. Similarly, one would like to know whether there are observable increases in hereditary changes in this population, since undoubtedly the DNA molecule is the target molecule for the mutagenic action of ionizing radiation for, as we have

seen, a single base change is all that is needed to give rise to recognizable disease.

3. Immunogenetic Problems

Let us turn from the hereditary diseases—diseases of defects in DNA—to immunogenetic problems in disease in man. From the central dogma DNA→RNA→Protein if antibody production is an example of induced protein synthesis, then theories of antibody production centre on what controls initiation of antibody.

(a) *Theories of antibody production*

According to the "instructor theory", the antigen determines the conformation and construction of the specific antibody by direct template action. Haurowitz (1952) conceives of the template as a peptide layer producing a complementary replica by specific adsorption of amino acids; the template reproduction takes place only when the protein is in an expanded state. Then the peptide layer is released from the template and folds into a globular molecule. Pauling (1940) proposed a mechanism based on the assumption that antigen does not affect the order of amino acids, but that the antigen determines the conformation of the specific antibody by coming into contact with the globulin molecule while it is being folded into its third-order structure. The antigen determining groups in contact with antibody globulin impress the complementary conformation on the end parts of the chain to be stabilized by normal folding of the centre of the chain. Thus antibody specificity is visualized by Pauling as being determined primarily at the protein end of the flow of information and by Haurowitz as taking place at the interaction of RNA and protein. The current idea is that the linear sequence of amino acid units in each protein determines the way in which the helical polypeptide chains fold on themselves to produce the third-order structure. The fact that the amino acid sequence is coded in the base sequence of DNA and that the amino acid sequence probably determines the spatial conformation of the protein argues against the "instructor theory" of antibody production. Nevertheless, one of the primary interests of immunologists in antibody production is that this may represent the one situation in which protein synthesis is specifically affected by other proteins. The interrelationship of messenger RNA and ribosomes is not sufficiently clear to exclude this possibility. Moreover, many haemoglobin molecules have similar conformations, although their amino acid sequences differ considerably.

An alternate theory is that antigens somehow stimulate selectively the production of different groups of cells having inherent ability to produce the specific antibody to react with that particular antigen. This "clonal

selection theory" (Burnet, 1959) postulates that the antigen simply selects for proliferation the particular clone of cells that can react with it. Although this theory has been much mooted (Burnet, 1962), it merely embroiders the fact that antibody production is inherent in DNA, that the ability to encode all antibodies must reside in the primary genetic code, and that the response to antigen stimulation is essentially the stimulation of a particular DNA (Hamilton, 1956; 1957). Whether the DNA is in one cell or another or in many is not vital. What is vital is the deployment of a portion of the DNA molecule that can specifically encode a particular antibody. The information is then passed to the corresponding messenger and the corresponding antibody is produced. The "clonal selection theory" says nothing about mechanism: it adds another *ad hoc* assumption of selection of antibody producing cells and returns us to the original mystery of cellular differentiation in ontogeny.

(b) *Plasma cells*

Antibodies are made by plasma cells in lymphatic tissue; lymphatic tissue is all over the body but concentrated in lymph nodes and spleen. Plasma cells are normally present in nodes and other lymphatic tissues only in small numbers, and there is reason to believe that these are cells already committed to the production of specific antibodies. When antigenic material is introduced, plasma cells appear and multiply rapidly and at the same time more antibody is produced. In studies of isolated single cells (Nossal, 1958; Attardi *et al.*, 1959), plasma cells predominate among single cells capable of antibody synthesis. However, in the studies of Attardi *et al.*, cells indistinguishable from medium-size or small lymphocytes also produced antibody. It appears that the principle source of both antibody and γ-globulin is the plasma-cell series; recent evidence suggests that lymphocytes may also contribute.

(c) *Lymphocytes and cell transfer*

Lymphocytes, small white cells in the peripheral blood, carry antibodies and, more importantly, information controlling the synthesis of additional antibody, as shown by experiments in animals where it has been possible to transfer the *ability* to produce specific antibodies from one animal to another by the transfer of lymphoid cells, including small lymphocytes. Moreover, new information leads to the belief that small lymphocytes under suitable antigenic stimulation may be transformed into antibody-producing plasma cells. Cell-transfer studies (Roberts and Dixon, 1955; Neil and Dixon, 1959) seem to provide a simple explanation for antibody production by transfer of lymphocytes. These experiments indicate that populations of mainly small lymphocytes may be the cell stimulated with antigen, but after injection into an irradiated recipient,

and appropriate residence within the host, these cells are transformed into antibody-synthesizing plasma cells. Morphologic evidence for the transformation of lymphocytes to plasma cells seems good. There is no question that the new population of cells produces antibody. Possibly but now unlikely the accumulation of plasma cells at the site of injection of stimulated lymphoid cells is not the consequence of transformation of the population of lymphocytes into a population of plasma cells, but rather the rapid proliferation of a few plasmablasts or proplasmacytes present in the original cell suspension stimulated with antigen.

The evidence that one may transfer the *ability* to produce specific antibodies from one animal to another by the transfer of lymphoid cells was pioneered by Chase's success (1945) in transferring delayed hypersensitivity to tuberculin by means of cells from lymph nodes, spleen, and peritoneal exudates of tuberculous guinea-pigs. Using similar cells, Chase (1951, 1953) also transferred delayed hypersensitivity to picryl chloride from previously sensitized guinea-pigs to normal guinea-pigs; the recipient animals developed skin-sensitivity and anaphylactogen antibodies to picryl chloride; the increase in antibodies in time was measured.

Another example of the transfer of the ability to produce antibodies by cellular transfer is the homograft studies of Billingham (1953), Mitcheson (1955) and Medawar. A skin or homograft from a mouse of strain X is transplanted to a mouse of strain Y. In about 11 days, depending on the genetic disparity between X and Y, the graft is destroyed and the effect of this first graft is to increase greatly the rapidity with which subsequent grafts are rejected. If, at the time of this immune reaction, cells from the regional lymph nodes of mouse Y are injected into another mouse of strain Y, the recipient of the lymph node cells will reject a homograft from a strain X mouse at an earlier time 4–5 days (accelerated reaction) as if it has been previously immunized. The hypersensitivity and homograft reponse achieved after cell transfer is too immediate to be due to transfer of the original antigen with the cells. Since antibodies cannot self-replicate and it is clear that not enough passive antibody could be transferred to account for the long continued immunity conferred on the recipients— and indeed, for the striking increase of antibody with time in the recipient—it is likely that some actively metabolic mechanism has been carried over in this way. Cytological observations by several investigators on the cell-types present in the transfer of lymph node cells have shown that 95% or more of the viable cells are in fact small lymphocytes.

(d) *Delayed type hypersensitivity*

"Delayed type hypersensitivity" is applied to responses in which there

is a relatively slow development of specific inflammatory processes after introduction of antigen into the skin. In contrast with wheal and erythema sensitivity, in which inflammation begins in seconds, ends in minutes, and the Arthus type sensitivity, in which activity appears within minutes after injection of the antigen into the skin, delayed hypersensitivity generally does not appear until 4–8 hours after injection of the antigen, achieves a gradual evolution to maximal reaction 24–48 hours later, and only gradually disappears. Such delayed skin reactions follow not only association with infections and skin contact with chemicals, but also injection of soluble protein antigens.

Delayed type hypersensitivity responses are associated with certain bacterial infections, notably infectious diseases characterized clinically and pathologically as granulomatous, e.g. tuberculosis, where the tuberculin reaction represents the specific immunologic expression of delayed hypersensitivity. Delayed reactions also characterize many other infectious diseases. Contact reactions to many organic chemicals, e.g. plant oils found in poison ivy and poison oak, and even allergic reactions to extremely simple chemical compounds also appear to belong to the class of delayed hypersensitivity reactions. There are strong arguments for delayed hypersensitivity being responsible for rejection of homotransplants. These delayed hypersensitivity responses depend not on circulating antibodies, but on specifically committed lymphocytes.

One concludes that the small lymphocyte that circulates in enormous numbers in the peripheral blood is *the* recogition cell for the immune mechanism. Thus when foreign tissues are transplanted and are vascularized, the cells calling them foreign are the small lymphocytes—these are also the cells that reject the strangers. Gowans *et al.* (1962) has demonstrated that small lymphocytes after antigenic stimulation transform into large pyroninophilic cells that position themselves in lymphatic tissue. Small lymphocyte descendants of these large pyroninophilic cells in the nodes then accumulate in the skin at the site of homograft rejection. Animals injected with adult lymphoid cells which lack antigens present in the tissues of the host may develop a fatal wasting disease. There is convincing evidence that in this situation the small lymphocyte induces this disease by reacting with the tissue antigens of the host and initiating an immune response against them. In delayed hypersentivity reactions it is still not proven whether it is the committed small lymphocytes themselves that recognize the stimulating antigen and give rise to hypersensitivity; evidence is accumulating that this is so.

(e) *Auto-immune diseases*

In an increasing number of disease states it is recognized that the

disease ensues from production of antibodies against specific tissues of the body by the individual himself; these are the "auto-immune" diseases. Considerable evidence relates delayed hypersensitivity to the destructive process in these diseases. For example, experimental allergic encephalomyelitis (Waksman and Morrison, 1951) induced by injection of the brain with adjuvants and auto-immune thyroid disease (McMaster *et al.*, 1961) correlate better with delayed hypersensitivity than with circulating antibody. Furthermore, extensive morphologic data (Waksman, 1960) links a whole group of auto-immune pathologic processes to the delayed response.

One theory of these diseases is that they represent somatic gene mutations of the antibody-producing cells that now produce antibodies against their own tissues (Burnet, 1959). An explanation more consonant with the evidence is that auto-immune diseases are a variation of the delayed hypersensitivity phenomenon, i.e. they may well represent erratic associations between exogenous chemicals or other agents and self-cell proteins. The resulting complexes are then treated by the body's recognition system as foreign, and hence antibodies are produced against the body's own tissues, giving rise to various diseases.

An excellent example of such an association between exogenous chemical and cell proteins is afforded by the thrombocytopenic purpura induced by sedormid (allyl isopropyl acetyl urea) in sensitive subjects (Ackroyd, 1955). In this condition, sedormid complexes with the platelet. The complex is now treated as foreign, antibodies are produced against the platelets and thrombocytopenia ensues. Although the primary initiating condition that gives rise to the immune reaction was the chance association of exogenous chemical and cell-membrane, it is clear that variation in susceptibility to such complex formation must be genetically determined; this too probably holds for all delayed hypersensitivity phenomena. There is something unique in the cell-membrane that facilitates complex-formation and hence its acting as a foreign antigen. Alternately the genetic specficity could be in the particular type of antibody produced.

The immune diseases may be regarded as diseases arising primarily from alterations in proteins. Such alterations can result from complexes with exogenous agents or possibly from alterations in protein molecules associated with aging. In both situations there are probably also genetic factors determining susceptibility to such protein-complexing and aging. Thus the immunogenetic diseases, although reflecting primary defects in protein, are also DNA-determined. Certainly the ensuing production of antibody is DNA-determined; if one accepts Haurowitz's instructor hypothesis, one could regard immunogenetic diseases as being primary RNA diseases.

4. Cancer

The cancer cell may be considered as a mistake, presumably a somatic gene mutation resulting in a defect in a growth-controlling system. Induced and repressible enzyme systems in micro-organisms provide possible models for the regulation of protein synthesis. Such effects have been observed frequently in higher organisms; it has not been possible to analyse any in detail. The problem of chemical embryology is to understand why cells do not express all the time all potentialities inherent in their genomes. Cellular differentiation requires that many, and in some tissues most of these potentialities must be repressed.

During differentiation in higher organisms, rates of enzyme synthesis are set. They do not vary greatly under different nutritional or metabolic conditions, since animal cells are not exposed to the same environmental extremes as are bacteria, except in animals subjected experimentally to nutritional extremes, e.g. changes in gluconeogenesis (Krebs *et al.*, 1963). In animal cells regulation depends on reversible or irreversible activation and inactivation of enzymes; these activities are controlled by specific regulatory sites on enzymes that can in combination or dissociation with small molecules modify the stability and activity of the enzyme (Pardee and Wilson, 1963).

The problem of the origin of malignant transformation is thus bound up with differentiation in multicellular organisms. Possibly relevant is the fact that higher organisms have histones applied to their nucleic acids; in contrast, in *E. coli* quantitative amino acid analysis shows that the protein is not histone (Zubay and Watson, 1959), and X-ray diffraction of nucleoprotein preparations shows that the DNA is largely protein free (Wilkins and Zubay, 1959). Recent work has shown that thymus histones added to isolated thymus nuclei or to nuclear ribosomes inhibit many biosynthetic reactions, including RNA synthesis (Allfrey *et al.*, 1963; Allfrey, 1963). The arginine-rich histones are strong inhibitors of RNA synthesis, while the lysine-rich histones are relatively inert. One feature of the X-ray diffraction pattern given by nucleohistone could be explained if histone is distributed throughout the space between DNA molecules and is not in contact with DNA, except for the basic ends of the side-chains of lysine and arginine (Wilkins, 1959). Thus the proximity of the lysine and arginine side-chains could affect local regions of the DNA molecule. Moreover, the removal of 60–70% of the total acid-extractable thymus nuclear histones from the nucleus by selective digestion with trypsin apparently greatly stimulates RNA synthesis as judged by isotope uptake; the RNA synthesized in the histone-depleted nuclear suspension has a different base composition from RNA synthesized in untreated nuclei (Allfrey, 1963).

Inhibition by histone may involve several steps; added histones can inhibit nuclear ATP synthesis and thus prevent ATP-dependent activation of amino acids for protein synthesis and diminish kinase activity and the ATP "pool" for RNA synthesis (Allfrey *et al.*, 1963). It is too early to say how specific are these histone effects: histones inhibit many reactions through their formation of complexes with many enzymes, their substrates, and cofactors.

One wonders therefore whether histones are the repressors that modulate the flow of DNA-carried information, since most of the DNA in the thymus nucleus and in other highly differentiated cell-types is repressed, not participating in messenger-RNA synthesis. Were differentiation histone dependent, one might find differences in histones from different organs of the same species. But chromatographic analysis of histones isolated to date from many tissues have failed to show tissue-specific differences, although they have revealed differences between histones of different species (Crampton *et al.*, 1957). Present technique may not yet be discriminating enough. There have been a few suggestive observations on tumours: thus nuclear histone in papilloma nuclei is greatly reduced relative to the histone content of normal skin nuclei (Rogers, 1962), and it has been claimed that some tumours have a high rate of lysine-rich histone synthesis (Davis and Busch, 1960); such histones might be ineffective inhibitors of nuclear activity, as suggested by Allfrey (1963).

Another point is the observation that most specialized cells, when taken from the body and grown in tissue culture, lose their special abilities, but instead during dedifferentiation acquire at least one property of malignant cells: capacity for infinite replication.

Further hints on the role of proteins in regulating cell-replication are the inhibition of regenerative changes after partial hepactectomy by albumin (Glinos, 1958). It has been suggested that plasma-albumin concentration in the immediate environment of the hepatic cells controls regeneration. These experiments have their counterpart in work on embryonic tissue, where extracts of organs inhibit the development of homologous organ grown in the extracts (Rose, 1957; Braverman, 1961).

There has been increased emphasis on the virus-induction of cancer, especially in experimental animals. This has led to theories of carcinogenesis postulating that the viral nucleic acid takes over control of the replicative mechanism of the host cells, i.e. gives rise to tumours. Nevertheless, the evidence indicates that most physical and chemical carcinogens, the latter including a wide range of structures, exert their effect directly on the DNA molecule, changing the normal base sequence in the polynucleotide chains, changes ultimately expressing themselves as the

anarchic cancer cell. Stereochemical evidence for this mechanism of carcinogenesis may be forthcoming before long.

References

Ackroyd, J. F. (1955). *Brit. Med. Bull.* **2**, 28.

Allfrey, V. G. (1963). In *Exp. Res. Suppl., Nucleus of the Cancer Cell* (in press).

Allfrey, V. G., Littau, V. C. and Mirsky, A. E. (1963). *Proc. nat. Acad. Sci., Wash.* **49**, 414.

Attardi, G., Cohn, M., Houbata, K. and Lennox, E. S. (1959). *Bact. Rev.* **23**, 213.

Baglioni, C. (1962). *Biochim. biophys. Acta* **59**, 437.

Baglioni, C. and Ingram, V. M. (1961). *Biochim. biophys. Acta* **48**, 253.

Billingham R. E., Brent, L. and Medawar, P. B. (1953). *Nature, Lond.* **172**, 603.

Braverman, M. H. (1961). *J. Morph.* **108**, 263.

Burnet, F. M. (1959). *The Clonal Selection Theory of Acquired Immunity.* Vanderbilt University Press, Nashville, Tenn.

Burnet, F. M. (1962). *The Integrity of the Body.* Harvard University Press, Cambridge, Mass.

Chase, M. W. (1945). *Proc. Soc. exp. Biol., N.Y.* **59**, 134.

Chase, M. W. (1951). *Fed. Proc.* **10**, 404.

Chase, M. W. (1953). In *The Nature and Significance of the Antibody Response.* Columbia University Press, New York.

Crampton, C. F., Stein, W. H. and Moore, S. (1957). *J. biol. Chem.* **225**, 363.

Davis, J. R. and Busch, H. (1960). *Cancer Res.* **20**, 1208.

Gerald, P. A. and Efron, M. L. (1961). *Proc. nat. Acad. Sci., Wash.* **47**, 1758.

Glinos, A. D. (1958). In *The Chemical Basis of Development.* Johns Hopkins University Press, Baltimore.

Gordon, W. G., Basch, J. J. and Kalan, E. B. (1961). *J. biol. Chem.* **236**, 2908.

Gowans, J. L., McGregor, D. D., Cowen, D. M. and Ford, C. E. (1962). *Nature, Lond.* **196**, 651.

Hamilton, L. D. (1956). *Nature, Lond.* **178**, 597.

Hamilton, L. D. (1957). In *Brookhaven Symposia in Biology, X. Homeostatic Mechanisms*, p. 52.

Haurowitz, F. (1952). *Biol. Rev.* **27**, 247.

Hill, R. L. and Schwartz, H. C. (1959). *Nature, Lond.* **184**, 641.

Hunt, J. A. and Ingram, V. M. (1959). *Nature, Lond.* **184**, 640, 870.

Ingram, V. M. (1957). *Nature, Lond.* **180**, 326.

Ingram, V. M. (1962). *The Molecular Control of Cellular Activity.* McGraw-Hill, New York.

Ingram, V. M. and Stretton, A. O. W. (1961). *Nature, Lond.* **190**, 1079.

Jones, O. W. and Nirenberg, M. W. (1962). *Proc. nat. Acad. Sci., Wash.* **48**, 2115.

Jukes, T. H. (1962). *Proc. nat. Acad. Sci., Wash.* **48**, 1809.

Jukes, T. H. (1963). *Biochem. biophys. Res. Comm.* **10**, 155.

Krebs, H. A., Bennett, D. A. H., De Gasquet, P., Gascoyne, T. and Yoshida, T. (1963). *Biochem. J.* **86**, 22.

McMaster, P. R. B., Lerner, E. M. and Exum, E. D. (1961). *J. exp. Med.* **113**, 611.

Mitcheson, N. A. (1955). *J. exp. Med.* **102**, 157.

Muller, C. J. and Kingma, S. (1961). *Biochim. biophys. Acta* **50**, 595.

Murayama, M. (1960). *Fed. Proc.* **19**, 78.

Neil, A. L. and Dixon, F. J. (1959). *Arch. Path. (Lab. Med.)* **67**, 643.

Nossal, G. J. V. (1958). *Brit. J. exp. Path.* **39**, 544.
Osmundsen, J. (1962). *New York Times*, February 2, pp. 1, 14.
Pardee, A. B. and Wilson, A. C. (1963). In: American Cancer Society Conference. Problems Basic to Cancer Chemotherapy. *Cancer Res.* (in press).
Pauling, L. (1940). *J. Amer. chem. Soc.* **62**, 2643.
Roberts, J. C. and Dixon, F. J. (1955). *J. exp. Med.* **102**, 379.
Rogers, S. (1962). In *The Molecular Basis of Neoplasia.* University of Texas Press, Austin.
Rose, S. M. (1957). *Biol. Rev.* **32**, 357.
Swenson, R. T., Hill, R. L., Lehmann, H. and Jim, R. T. S. (1962). *J. biol. Chem.* **237**, 1517.
Wahba, A. J., Gardner, R. S., Basilio, C., Miller, R. S., Speyer, J. F. and Lengyel, P. (1963). *Proc. nat. Acad. Sci., Wash.* **49**, 116.
Waksman, B. H. (1960). *Cellular Aspects of Immunity.* Little, Brown, and Co., Boston.
Waksman, B. H. and Morrison, L. R. (1951). *J. Immunol.* **66**, 421.
Wilkins, M. H. F. (1959). In *Solvay Institute 11th Chemistry Conference. Nucleoproteins*, p. 45. Interscience, New York.
Wilkins, M. H. F. and Zubay, G. (1959). *J. biophys. biochem. Cytol.* **5**, 55.
Zubay, G. and Watson, M. (1959). *J. biophys. biochem. Cytol.* **5**, 51.

General References

Hereditary Diseases

Harris, H. (1959). *Human Biochemical Genetics.* Cambridge University Press.
Stanbury, J. B., Wyngaarden, J. B. and Frederickson, D. S. (1960). *The Metabolic Basis of Inherited Disease.* McGraw-Hill, New York.

Immunology

Kallos, P. and Waksman, B. H. (eds.) (1962). *Progress in Allergy* VI. Karger, Basel.

Differentiation

Symposium on Specificity of Cell Differentiation and Interaction (1962). *J. cell. comp. Physiol.* **60**, Suppl. 1.

Discussion

D. C. PHILLIPS: It is now possible to understand the physiological effects of some abnormal haemoglobins in terms of their molecular structures. Thus, following Perutz (1962), three kinds of haemoglobin-M have been observed so far, in which histidine residues 58*d* or 63β in the normal chains are replaced by tyrosine or in which valine 67β is replaced by glutamic acid. These residues all lie in helices close to the oxygen-combining sites of the haem groups, which suggests the possibility that direct links are formed between the iron atoms and the phenolic groups of the tyrosines or the γ-carboxyl groups of the glutamic acid residues. Such linkages would block the oxygen-combining sites and protect the ferric iron atoms from the action of methaemoglobin reductase, the enzyme which helps to keep the iron atoms in the ferrous state. (M. F. Perutz, *Proteins and Nucleic Acids: Structure and Function.* Elsevier, Amsterdam (1962).)

SECTION III
Optical Studies

Ultra-violet Spectra and Structure of Proteins

JOHN T. EDSALL

Biological Laboratories, Harvard University, Cambridge, Massachusetts, U.S.A.

ABSTRACT

The use of absorption spectra in studying protein conformation will be discussed. The absorption bands above 260 mμ, due to the aromatic amino acid side chains, can be displaced by various changes of solvent. Addition of substances such as ethylene glycol and glycerol increases the polarizability of the solvent without much change in the conformation of the protein. The absorption of aromatic side chains that are exposed to the solvent at the surface of the protein should be shifted to longer wavelengths under these circumstances; buried groups in the interior of the native protein should undergo little or no change of absorption (Herskovits and Laskowski, 1960, 1962). Reagents such as dioxane and urea, at high concentrations, alter the native protein conformation drastically; thus urea produces the well-known "denaturation blue shift" near 287 mμ. Observation of difference spectra as a function of time permits the study of the kinetics of the shift when ribonuclease is dissolved in urea (Nelson and Hummel, 1962).

It is well known that tyrosyl hydroxyl groups are often unavailable for titration in native proteins, becoming available only after denaturation. A review of the data for numerous proteins will be given, with a discussion of recent studies by L. M. Riddiford in our laboratory on titration of human erythrocyte carbonic anhydrase.

Glazer and Smith (1960, 1961) reported a very large peak in the difference spectrum near 235 mμ after acid denaturation of many proteins. Eisenberg (1961) in our laboratory has shown by titration of human serum mercaptalbumin, from pH 7 to pH 2, that the 235 mμ peak in the difference spectrum runs exactly parallel to the smaller peak at 287 mμ, which is known to be due to tyrosyl residues. However, this correlation cannot be taken as a general rule for proteins.

Changes in absorption near 190 mμ can give information about helix coil transitions in proteins (Imahori and Tanaka, 1959; Doty and Gratzer, 1962). Accurate knowledge of contributions made by amino acid side chains to absorption is essential for interpretation of the results. New and highly accurate data for absorption spectra of amino acids between 180 and 230 mμ are now being obtained in Professor Doty's laboratory. Tentative calculations of helical content of proteins, based on ultra-violet absorption data, will be discussed, and illustrated by studies of S. A. S. Ghazanfar in our laboratory on human carbonic anhydrase. Comparison with optical rotary dispersion and deuterium exchange studies will be presented, with critical evaluation of the results.

The use of ultra-violet spectra for the study of the conformation of protein molecules has undergone considerable refinement in the last few years. Alterations of conformation with changing temperature or pH, or

with the addition of organic molecules to an aqueous solvent medium, are accompanied almost invariably by shifts in position, and often by changes in intensity, of ultra-violet absorption bands. Interpretation of these changes has depended partly on a deeper understanding of the nature of the energy level transitions involved, but even more on empirical correlations with information derived by other methods, such as optical rotatory dispersion, deuterium exchange, acid–base titrations, and the classical techniques for the study of macromolecules, such as sedimentation, viscosity and light scattering. The detailed information concerning the conformation of myoglobin and haemoglobin, provided by the crystallographic studies of Kendrew, Perutz and their associates, is now beginning to provide correlations between specific features of protein fine structure and the ultra-violet spectra; we shall return later to one or two of these.

Most earlier studies of ultra-violet spectra of proteins concerned the wavelength region between 250 and 300 mμ, in which the absorption is due almost entirely to the aromatic side chains of phenylalanine, tyrosine and tryptophan. The far stronger absorption bands of these side chains at shorter wavelengths, between 190 and 240 mμ, received comparatively little study until lately, partly because of technical difficulties, and still more because of the overlapping absorption in this region from histidine, methionine, cystine and ionized cysteine, and the long wavelength end of the absorption due to the peptide bond itself.

In the last few years there has been a rapid increase in interest in the absorption bands located below 200 mμ, with particular emphasis on the absorption band of the peptide linkage centred near 190 mμ. The work of Imahori and Tanaka (1959) on polyglutamic acid gave the first decisive demonstration that the transition from a helix to a random coil was accompanied by a large increase of ultra-violet absorption in this region. The more recent work of Rosenheck and Doty (1961) gave a great impetus to the study of absorption in this region, particularly because of the close relation between the ultra-violet absorption bands and the changes in optical rotatory dispersion that accompany the helix-coil transition.

The intermediate region, between 200 and 250 mμ, is also the subject of much current study. The ionization of tyrosine residues in alkaline solutions can be conveniently followed by observation of the absorption band with its peak near 245 mμ. This is as characteristic of the ionized phenolic group, and is several times as intense, as the more familiar absorption band centred at 295 mμ. Also, the studies of Glazer and Smith (1960, 1961*a*) on difference spectra in neutral and acid solutions of many proteins have revealed a large peak near 235 mμ. This peak in the difference spectrum is far higher than the peaks due to tyrosine residues

near 278 and 287 mμ, or those due to tryptophan at slightly longer wavelengths. These studies are further discussed below.

Before proceeding to further detailed discussion I would call attention to two important recent reviews, one by Beaven (1961) and one by Wetlaufer (1962). In this much briefer discussion I shall concentrate on some recent developments, including some of the new studies in our own laboratory on carbonic anhydrase and those in Professor Doty's laboratory on a variety of proteins at very short wavelengths.

1. Ionization of Tyrosine Groups in Relation to Protein Structure

In the compact structure of native globular proteins, the amino acid side chains may project outward towards the solvent from the main peptide chain, or may be folded into the interior of the molecule and enclosed by surrounding groups of the protein itself. Intermediate situations may of course exist. An amino acid side chain may be partly buried near the surface of the molecule, but flexible enough to emerge at times and interact with the surrounding solvent. Some proteins may enclose water molecules trapped in the interior, so that internal groups may still be in contact with water. There may be crevices at the surface of the protein molecule, through which small solvent molecules, but not large ones, can penetrate. Until very recently all such conceptions were largely speculative, although supported by much indirect evidence. For at least one protein, sperm-whale myoglobin, however, the investigations of Kendrew (1962) provide a detailed picture of the folding of the peptide chain in space, and of the distribution and orientation of the side chains as they project from the peptide chain. In some respects myoglobin may not be a typical protein—more than three-quarters of the peptide chain is folded into right-handed α-helical segments, and many (perhaps most) proteins are almost certainly far less helical than this. The distribution of the side chains in myoglobin, however, may be more characteristic of native globular proteins in general. The charged side chains, and the aliphatic hydroxyl groups of serine and threonine, are generally at the surface of the molecule, in contact with the solvent; they tend to be flexible, and to move freely around their points of attachment to the main chain. The interior of the molecule is predominantly made up of non-polar residues, in van der Waals contact with one another; a few such residues, especially small ones like alanine, are at the surface.

In the case of the two tryptophan residues in myoglobin, the ring N—H is at the surface of the molecule, although the large hydrophobic part of

the ring is buried inside. In one tryptophan residue the NH group appears free; in the other it is bonded to a glutamine residue.

However, we are here concerned particularly with the distribution of tyrosine side chains which, like those of tryptophan, have both polar and non-polar characteristics. The acidic phenolic group, with its capacity for hydrogen bonding, interacts readily with water. On the other hand the large benzene ring gives the tyrosine side chain a strong tendency to imbed itself in non-polar regions of the molecule. Thus the tyrosine groups in protein molecules may be found either in close contact with the solvent, or deeply buried in the interior of the molecule, or half buried, with the hydroxyl group near the surface, like the tryptophan groups of myoglobin.

In myoglobin Kendrew (1962) finds that one of the three tyrosine groups of myoglobin (H22), in the H-helix not far from the -COOH terminal end of the peptide chain, is buried within the molecule, with its hydroxyl group bonded to a carbonyl residue of a quite different portion of the main peptide chain in position FG5. This tyrosine residue serves to cross-link the helices designated as G and H by Kendrew. The other two tyrosine residues of myoglobin are close to the surface and would appear to be free to ionize.

The spectrophotometric titrations of Hermans (1962) have now made a significant beginning in establishing a correlation between the detailed structural data from X-ray measurements and the spectrophotometric pH titration of the tyrosine groups. Hermans, in his titration, observed the ionization of the tyrosine groups in sperm-whale myoglobin, and also in human and horse haemoglobin, by measuring the absorption at 245 mμ, near the peak at 240 mμ, which is highly characteristic of the ionized phenolic group and is several times as intense as the peak at 295 mμ. In the study of haem proteins measurement at 245 mμ also has the advantage of involving less overlap with the characteristic absorption bands of the haem group. Hermans found that, of the three tyrosine groups in myoglobin, two ionize reversibly, one with a pK value of 10·3, and the other with the abnormally high pK value of 11·5. A third group was apparently quite unavailable for ionization in the native molecule. It appears natural to identify the latter group with the tyrosine residue at H22, which Kendrew finds to be hydrogen bonded to a CO group of the peptide chain in the interior of the molecule. It is premature as yet to attempt an explanation of the difference in the ionization behaviour of the other two groups.

The effect of removing the haem group from the myoglobin has been studied by Breslow (1962). She has shown that in the resulting globin molecule all three tyrosine residues are titratable, and moreover the entire spectrophotometric curve is shifted to lower pH values, by

approximately 1 pH unit. It is also interesting that at least two imidazole groups of histidine, which are unreactive in native myoglobin, are readily titratable in the globin derived from it.

These studies on myoglobin offer the first definitive structural clue to the unusual behaviour of many tyrosine groups in proteins, which has been apparent ever since the classical work of Crammer and Neuberger (1943) revealed a profound difference between the tyrosine groups of insulin and those of ovalbumin. Crammer and Neuberger, like nearly all later investigators, followed the ionization of tyrosine groups by measurement of the intensity of absorption at 295 $m\mu$. They found that insulin resembles a simple tyrosine peptide, in that all four of the tyrosine groups of the monomer molecule are readily ionizable, whereas the nine or ten tyrosine groups of ovalbumin are virtually unavailable for reaction with the solvent medium until the molecule is denatured in strong alkaline solution above pH 12. A large number of other proteins have now been studied in similar fashion, and a list of most of them is given in Table I.

The data of Table I would indicate that the great majority of proteins contain at least some tyrosine groups which, in the native molecule, are somehow shielded from the solvent so that they are unavailable for ionization until some sort of molecular unfolding has occurred. Most of the proteins in Table I contain several tyrosine groups that are apparently free to ionize; the number of these, in each case, is listed in Table I. The remainder are apparently unavailable to the solvent until very high pH values are attained and the native structure is lost. However, no summary in such a table can give full justice to the complexity of the situation. In a sense, for example, the tyrosine groups of serum albumin are apparently free, since all, or nearly all, ionize reversibly. Nevertheless they are clearly different from tyrosine groups in simple peptides (Tanford *et al.*, 1955 *a*, *b*). The intrinsic ionization constant of the albumin tyrosine groups, corrected for electrostatic interactions, is characterized by the unusually high pK value of 10·4. The heat of ionization is 11·5 kcal/mole, in contrast to the normal value of about 6, and ΔS^0 is -9 e.u. as against the normal value near -25 for tyrosine groups. Clearly there are impediments to ionization, possibly, but not certainly, due to hydrogen bonds in which these groups may be involved. In any case the tyrosine groups of serum albumin, as they exist in the native molecule, are held in some kind of internal structure in the native molecule.

Likewise, the study by Smillie and Kay (1961) on trypsinogen reveals complexities in the ionization of the tyrosine groups which are not revealed in the summary given in Table I. In this case the effects of temperature on ionization are profound. The molecule contains nine or ten tyrosine residues, four of which ionize freely and reversibly at all

TABLE I. Free and total tyrosine groups and other aromatic amino acid residues in certain proteins†

Protein	Mol. wt.	Phe	Tryp	Tyr (total)	Tyr (free)	Reference
α-Corticotropin (sheep)	4500	3	1	2	2	1
Insulin (beef)	5733	3	0	4	4	2
Ribonuclease (beef pancreas)	13,383	3	0	6	3	3, 4
Lysozyme (hen's egg)	14,800	3	6	3	3	5
Myoglobin (sperm whale)	16,700	6	2	3	2	6
Trypsinogen	24,500	3	(4?)	9–10	4	7
α-Chymotrypsin	25,100	7	7	4	2	8
α-Chymotrypsinogen	25,200	7	7	4	2	9
Papain	20,700	4	5	17	11	10
Human carbonic anhydrase (fraction I)	28,000	10	6	7–8	2	11
Δ^5-3-Ketosteroid isomerase	40,800	25	0	10	4	12
Ovalbumin	44,000	21	3	9–10	(2?)	2
Bovine serum albumin	66,000	27	2	20	‡	13
Human Haemoglobin	64,500	30	6	12	8	6
Conalbumin	76,600	25	11	18	11	14
Fe-Conalbumin	76,700	25	11	18	5	14

(1) Léonis and Li (1959); (2) Crammer and Neuberger (1943); (3) Shugar (1952); (4) Tanford *et al.* (1955*a*); (5) Tanford and Wagner (1954); (6) Hermans (1962); (7) Smillie and Kay (1961); (8) Havsteen and Hess (1962); (9) Chervenka (1959); (10) Glazer and Smith (1961*b*); (11) Riddiford (1962); (12) Kawahara *et al.* (1962); (13) Tanford *et al.* (1955*b*); (14) Wishnia *et al.* (1961). For references to work on other proteins see Wetlaufer (1962), p. 341.

† The molecular weights and numbers of aromatic amino acid residues as given in the table can usually be found in the data given by the above authors themselves. For the tabulation of the amino acid composition data for haemoglobin and myoglobin see Braunitzer *et al.* (1961); also, with respect to human haemoglobin see Hill *et al.* (1962).

‡ For a discussion of the tyrosine residues in serum albumin see the text.

temperatures, with fairly normal pK and ΔH values. At 10° another set of four residues are reversibly unmasked at about pH 11·5; at 25 and 37°, on the other hand, this unmasking is associated with irreversible changes in the trypsinogen molecule. Independent studies of sedimentation and viscosity revealed a change in molecular conformation beginning about pH 10, well below the pH value at which the tyrosine groups begin to be unmasked.

Recent studies in our own laboratory by L. M. Riddiford, on fraction I of human carbonic anhydrase (prepared by the method of Rickli and Edsall (1962)), have shown a behaviour of the tyrosine residues which differs somewhat from that of any other reported protein. This fraction of human carbonic anhydrase has a molecular weight near 28,000; the

molecule contains one zinc atom, seven or eight tyrosine and six tryptophan residues. Two of the tyrosine groups ionize readily and reversibly below pH 11. All of them are ionized above pH 13. In the intermediate pH zone, however, the character of the tyrosine ionization curve, as measured by the absorption at 295 mμ, is profoundly dependent on both time and temperature. At 1°, the ionization of most of the tyrosine groups occurs very slowly; even at pH 13 only about half the tyrosine groups are ionized at first. The final value of the absorption at 295 mμ is not attained until a period of time of the order of a week has elapsed. At 10° the process, although more rapid, is still extremely slow and requires 2 or 3 days to reach a steady state. At 25° the final value is reached in 3 to 4 hours. Important changes occur in the conformation of the protein during these slow transitions, but these changes appear to be largely reversible when the solution is brought back to pH 11 or below. Correspondingly, Riddiford has found that the enzyme activity is retained to a remarkable degree, even after long exposure to alkaline pH. After 3 h at pH 12·2 and 10°, for example, virtually 100% of the original enzyme activity of this carbonic anhydrase preparation is recovered on readjusting the pH to 7. Even after 3 h at pH 12·9 and 10°, 80% of the enzyme activity is still found.

The very slow, and apparently largely reversible, unfolding of this carbonic anhydrase molecule in alkaline solutions sets it apart from nearly all the other proteins that have been studied with respect to ionization of the tyrosine groups. It should be added that the native molecule is highly compact, as indicated by a very low intrinsic viscosity. As judged by the criteria of optical rotatory dispersion, short wavelength ultra-violet absorption, and deuterium exchange studies—carried out by S. A. S. Ghazanfar in our laboratory—one may tentatively conclude that the helix content of the native molecule is very low, and might even be zero. We return later to the changes in its ultra-violet absorption during acid denaturation, to which it is very susceptible.

There have been two general hypotheses regarding the nature of the forces that prevent ionization of tyrosine groups in native protein molecules. On the one hand, as was first suggested by Crammer and Neuberger (1943) the tyrosine hydroxyl groups may be hydrogen bonded to suitable acceptor groups, for instance to the ionized carboxyl groups of aspartate or glutamate residues, or, as in the one actually established case in myoglobin, to a C=O group in the main peptide chain. On the other hand the tyrosyl groups may be shielded from the solvent by enclosure in a surrounding sheath of non-polar side chains, in the native protein molecule. The latter view has been favoured by a number of authors, for instance by Williams and Foster (1959); it is perhaps most explicitly stated in the discussion by Yanari and Bovey (1960). These two proposed

mechanisms, however, should not be regarded as mutually exclusive. Hydrogen bonds between tyrosine groups are easily broken by competing water molecules in an aqueous medium, as was shown, for instance, by Wetlaufer (1956). To be strong such hydrogen bonds must be shielded from water by adjacent non-polar groups. Conversely a tyrosine hydroxyl group, if enclosed by other non-polar groups, will form hydrogen bonds with any suitable acceptor group that is sterically available, since hydrogen bonding of this type increases stability. It is highly likely, therefore, that the "unavailable" tyrosine groups in proteins are hydrogen bonded and *also* buried in the predominantly non-polar interior of the molecule.

2. Changes of Absorption Bands of Aromatic Side Chains, on Change of Solvent, in Neutral and Acid Solutions

In neutral and acidic solutions tyrosine remains un-ionized. Hence, the observed spectral perturbations in such solutions, resulting either from change of pH or from the addition of salts or organic molecules, are small in comparison with those resulting from ionization of tyrosine. Nevertheless they have yielded information of comparable importance concerning protein structure. Before considering these effects on protein molecules in solution we must first consider briefly the effects on small molecules containing similar aromatic side chains. Many of the underlying considerations have been conveniently summarized by Yanari and Bovey (1960) who give references to earlier fundamental theoretical papers on the effect of solvent media on absorption spectra. No discussion of the underlying theoretical considerations will be attempted here. It is an experimental fact, however, that increase of the refractive index of the solvent medium generally causes a red shift of the absorption bands; that is, increasing polarizability of the solvent commonly decreases the *difference* in energy level between the ground state and the first excited state, for $\pi \rightarrow \pi^*$ transitions, as in the aromatic amino acids. This might in general be expected, as a result of the interaction of the transition dipole of the chromophore with the polarizable molecules of the surrounding medium; the higher the polarizability of the solvent molecules the greater the decrease in the energy required for the transition to the excited state. For $n \rightarrow \pi^*$ transitions the effect of changing refractive index of the medium is almost invariably in the opposite direction (see Kasha, 1961) but this class of transitions does not concern us here.

The effects of hydrogen bonding require separate consideration. A relatively simple system, for observation of the effects of hydrogen bonding on the spectra of substances containing aromatic hydroxyl

groups, would consist of phenol, or a similar molecule, in a hydrocarbon solvent, with a basic molecule added, at varying concentrations, to serve as a hydrogen bond acceptor. Such systems have been studied by several investigators, notably by Nagakura and Gouterman (1957) and by Baba and Suzuki (1961). We may consider particularly the latter, very careful and detailed study, which involved phenol, α-naphthol, and β-naphthol, dissolved in iso-octane, with dioxane as the hydrogen bond acceptor. Amines and aliphatic ethers react similarly as acceptors, so far as their effects on the spectrum are concerned. The data show that, as dioxane is added, there is a pronounced red shift of the absorption bands near 275 mμ (36,000–38,000 cm^{-1}). The shift is of similar magnitude for all the vibrational components of the absorption band in this region; for phenol this red shift in the H-bonded complex amounts to about -340 cm^{-1}. The stronger absorption bands in the 200–230 mμ region (45,000–47,500 cm^{-1}) also undergo a red shift, of the order of -700 cm^{-1}. (One of the band components of α-naphthol in this region appears, on the contrary, to undergo a displacement to higher frequency, but all the other bands of both phenol and the naphthols undergo the characteristic red shift. The reason for this one anomalous shift is not clear.)

The absorption bands of simple phenolic compounds such as tyrosine and *O*-methyltyrosine generally undergo a red shift when the solvent is changed from water to media of higher refractive index, such as concentrated urea and sucrose solution (Wetlaufer *et al.*, 1958; Bigelow and Geschwind, 1960). Changes in the ionization of the adjoining amino and carboxyl groups in the free amino acids also cause similar spectral shifts in tyrosine, *O*-methyltyrosine and tryptophan (Wetlaufer *et al.*, 1958; Hermans *et al.*, 1960). These shifts involve primarily a displacement of the long wavelength side of the absorption band, above 270 mμ, to still longer wavelengths, with little change in the height of the maximum; the resulting difference spectra, when for instance the spectrum of a solution of the amino acid in concentrated urea is measured against a reference solution of the amino acid in water, show two peaks, one near 278 and another higher peak near 287 mμ (Fig. 1). Similar difference spectra for tryptophan and other indole derivatives give peaks near 284 and 293 mμ in the difference spectra (Bigelow and Geschwind, 1960; Donovan *et al.*, 1961). These peaks are considerably higher than those in the tyrosine difference spectra, corresponding to the much higher molar absorbancy values for tryptophan, and the correspondingly higher values for $d\epsilon/d\lambda$ in the neighbourhood of the peak absorption.

In the ionization of either tyrosine or tryptophan the *removal* of a proton, whether from the carboxyl or the amino group, produces a red shift in the spectrum of the chromophore group. Except in rare cases, however, the difference spectra of proteins resulting from change of pH

in the acid and neutral region, cannot be explained by perturbations due to the ionization of neighbouring groups. These effects of proton removal fall off very rapidly with increase in the distance between the ionizing group and the chromophore: compare, for instance, the virtually negligible effect on the spectrum resulting from the ionization of the amino group in glycyl-*O*-methyltyrosine, with the strong effect produced by the same ionization in *O*-methyltyrosine (Wetlaufer *et al.*, 1958). Likewise the ionization of the amino group in tryptophan produces a large peak in the difference spectrum ($\Delta\epsilon \cong 700$ at 293 mμ) whereas the corresponding ionization in glycyltryptophan produces a shift which is

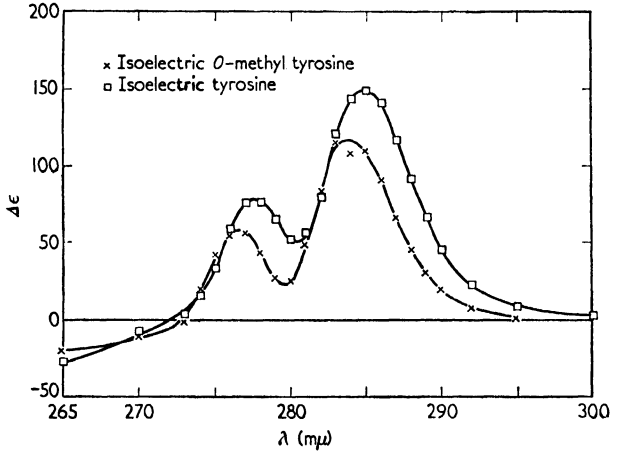


FIG. 1. Difference spectra produced by 6 M urea (with reference to aqueous solutions). From Wetlaufer *et al.* (1958).

smaller by nearly a factor of 10 (Donovan *et al.*, 1961). In general therefore the effects of acidification on the absorption spectra of proteins must be sought in conformational changes in the protein, with resulting spectral perturbations as the environment of the aromatic side chains is altered, and (rarely) in the ionization of groups immediately adjacent to the chromophore.

The exposure of proteins to solvents of higher refractive index than water gives more complicated results than with simple amino acids or peptides. If the conformation of the protein remains essentially unchanged the increase of refractive index will cause the usual red shift in the spectrum. The magnitude of this shift, however, will depend upon the topography of the native protein molecule. Aromatic amino side chains that are located close to the surface of the protein, so that they can readily interact with the molecules of the solvent, will undergo the

characteristic red shift as molecules of higher polarizability are added to the solvent. Aromatic amino acid residues deeply buried in the interior of the molecule, however, are shielded from such interactions with the solvent and would not be expected to undergo such displacements. These considerations have led Herskovits and Laskowski (1960, 1962) to the development of a solvent perturbation method of difference spectroscopy which they have applied to the study of the location of tyrosyl residues in ribonuclease and in serum albumin. Ribonuclease contains no tryptophan residues; bovine serum albumin contains only two tryptophan

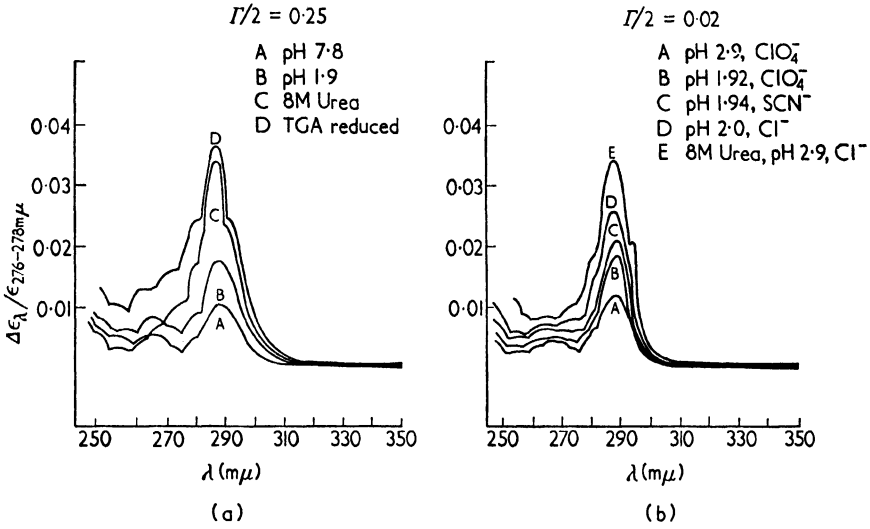


FIG. 2. (a) Typical solvent perturbation difference spectra of bovine serum albumin (BSA) due to 20% sucrose. (b) Effect of strong anion binding on the difference spectra of BSA, obtained with 20% sucrose. Protein concentration, 0·6%; temperature, 25 ± 1°. TGA = thioglycolic acid. From Herskovits and Laskowski (1962*a*).

residues, and human albumin only one, as against eighteen tyrosine residues. Hence in these proteins the observed difference spectra induced by alteration of the medium are primarily, and in the case of ribonuclease exclusively, due to the perturbations of the tyrosine group. Herskovits and Laskowski employed such solvents as ethylene glycol, glycerol and sucrose which in moderate concentrations (up to 20%) cause no significant changes in the conformation of the protein as judged by such criteria as optical rotatory dispersion (see for instance Tanford *et al.*, 1962). An example of their method is given in Fig. 2 which shows the effect of pH on the difference spectra of native and denatured human and bovine serum albumin with 20% glycerol as perturbant.

The maximum effect of change of medium on the difference spectra is obtained when all the seventeen disulphide bonds in the albumin molecule are broken by reduction with thioglycolic acid, and the resulting –SH groups are carboxymethylated to prevent reoxidation. This reduced albumin is studied in 8 M urea to keep it in solution. The "perturbation difference spectrum" produced by addition of 20% sucrose to the reduced protein in this medium is shown in curve D of Fig. 2 (a).

It is seen that the perturbations produced by the added sucrose are far greater than in the native protein at pH 7·8 (Fig. 2 (a), curve A) and still approximately twice as large as in the native protein when it is in the "expanded" form in acid solution (Fig. 2 (a), curve B). In Fig. 3 the form of the transition is shown as a function of pH, with 20% glycerol added as the perturbing solvent. Above pH 5, in the native protein, the value of $\Delta\epsilon$ at 287 mμ is only about 0·4 of the maximum $\Delta\epsilon$ value attained for the reduced carboxymethylated protein in urea. Comparison with simple model compounds indicated that the tyrosine residues, even in the completely reduced albumin in urea, are not quite fully exposed to the solvent. Allowing for this consideration, the data of Fig. 3 might mean that about six of the twenty tyrosine residues in serum albumin are exposed to the solvent in the native molecule. The acid transition in the difference spectrum occurs near pH 4, and at low pH it would appear that an additional three to five tyrosine groups become exposed to the solvent. Such interpretations must be made with caution. There is probably no sharp distinction between "exposed" and "unexposed" groups; we may expect that some, perhaps many groups, are partially but not completely exposed, as Herskovits and Laskowski are careful to point out. Moreover, the choice of the perturbing solvent makes a difference. If the molecules of the perturbant are small, they may have a more pronounced effect on the difference spectrum than similar but slightly larger molecules (see the studies by Herskovits and Laskowski (1962*b*) on ovomucoid).

The change in conformation of the serum albumin molecule in acid solution near pH 4 is known to involve two distinct steps. Both steps involve changes in optical rotation (Leonard and Foster, 1961). The step that occurs at higher pH is also clearly revealed by electrophoretic measurements but does not involve a change in the difference spectrum; the second step, at lower pH, is manifested clearly in the difference spectrum, but not in electrophoresis. No detailed interpretation of these phenomena is yet available, but they are obviously of major importance for the elucidation of the structure of serum albumin.

It should be noted that tyrosine residues can apparently be shielded from the solvent, not only by being buried in the interior of a globular protein molecule, but to a considerable extent simply by forming part

of an α-helix. This is shown by the work of Doty and Gratzer (1962) who have measured the difference spectrum of copoly-L-tyrosine-L-glutamic acid (5 mole-% tyrosine) in the helical form (pH 3·88) against the random coil form (pH 6·8). The helical molecule absorbed more strongly in the region from 270 to 300 mμ, with the usual peaks near 278 and 287 mμ in the difference spectrum. The absorbancy at 288 mμ, plotted as a function

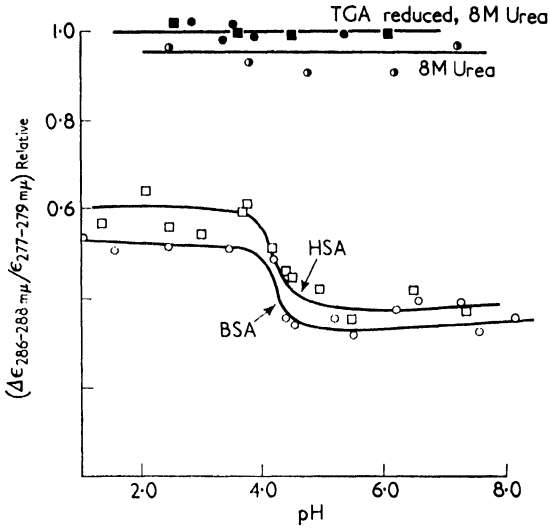


FIG. 3. Comparison of the difference spectral data for native and denatured bovine and human serum albumins (BSA and HSA) obtained with 20% glycerol as perturbant. BSA: ○, 0·25 M Cl^-; ◑, 8 M urea, 0·25 M Cl^-; ●, thioglycolic acid (TGA)-reduced BSA in 8 M urea, $\Gamma/2 = 0{\cdot}2$. HSA: □, 0·25 M Cl^-; ■, thioglycolic acid-reduced HSA in 8 M urea, $\Gamma/2 = 0{\cdot}2$. Protein concentration, 0·5–0·6%; temperature, 25 ± 1°. From Herskovits and Laskowski (1962*a*).

of pH, fell abruptly between pH 4·0 and 4·5, when only a few of the γ-carboxyl groups of the glutamic acid residues had ionized. Presumably the helical regions involving the tyrosine residues are weaker, and break more easily, than those involving only glutamic acid residues. In contrast to this large copolymer, the dipeptide L-tyrosyl-L-glutamic acid showed the usual *increase* in absorbancy characteristic of simple tyrosine peptides as the carboxyl group became ionized on raising the pH from 1·7 to 8.

The change in molar absorbancy at 287 mμ, per tyrosine residue, accompanying the helix-coil transition in the tyrosine-glutamic acid copolymer, was −140. Much higher values have been reported for the blue shift accompanying the denaturation of such proteins as ribonu-

clease and serum albumin. Therefore, as Doty and Gratzer point out, the apparent shielding of the tyrosine residues that is achieved by enfolding them in an α-helix is presumably less complete than that provided by enclosing them in the interior of a globular protein. Even so, it is highly significant.

3. Alterations of Protein Conformation in Denaturing Solvents: the Denaturation Blue Shift

Solvents such as urea, which at high concentrations cause profound alterations of protein conformation, naturally produce alterations in the spectra more complex than those we have just considered. The effects obtained on the addition of urea to a protein are particularly well illustrated by the work of Bigelow (1960) on ribonuclease (Fig. 4). At low

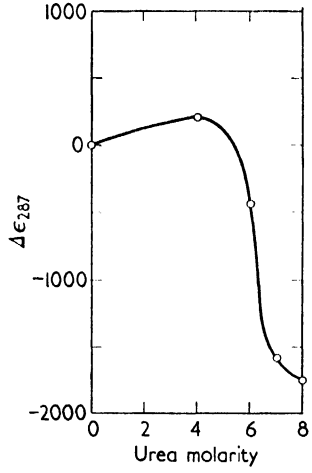


Fig. 4. The effect of concentration of urea on the change in the molar extinction coefficient of ribonuclease at 287 mμ. From Bigelow (1960).

concentrations of urea, below about 4 M, the usual red shift occurs, with a maximum in the difference spectrum near 287 mμ, $\Delta\epsilon_{287}$ being nearly a linear function of concentration. As the concentration of urea increases further, the value of $\Delta\epsilon_{287}$ plunges abruptly downwards, reaching a minimum when the urea concentration is near 8 M. Studies on viscosity, sedimentation and other properties of ribonuclease show that it is in this region of urea concentration that the molecule undergoes a general unfolding. As the concentration of urea increases still further, above 8 M, there is a small rise in $\Delta\epsilon_{287}$ once more, which again corresponds to the

usual red shift due to the increase in the refractive index of the solvent. The rapid decrease of $\Delta\epsilon_{287}$ over a critical range of urea concentration is the "denaturation blue shift", first explicitly described as such by Bigelow. Clearly it serves as an index of the re-arrangement of molecular structure as the molecule unfolds in the denaturing solution, and the three "unreactive" tyrosine residues of ribonuclease are removed from the interior of the native molecule with its high polarizability (i.e. high effective refractive index) and are brought into contact with the solvent, which is of lower refractive index. Although the matter was first explicitly

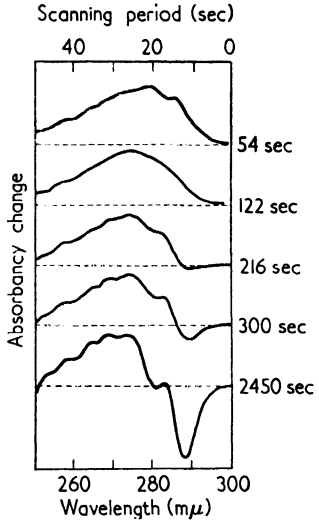


FIG. 5. The change in the difference spectrum of $1{\cdot}8 \times 10^{-4}$ M RNase with time in 7 M urea, 0·04 ionic strength imidazole, pH 7·0, at 23–24°. Each spectrum was scanned starting at the time shown at 300 mμ and was completed at 250 mμ 50 sec later. Each division on the vertical axis represents 0·1 absorbancy unit. Values above each base line represent increased absorbancy of RNase in urea over that in its absence. From Nelson and Hummel (1962).

discussed in essentially these terms by Bigelow, it is interesting to note that the denaturation blue shift was clearly observed by Chervenka (1959) in his studies of chymotrypsinogen. Since this protein contains seven tryptophan and only four tyrosine residues it is not surprising to find that the difference spectrum accompanying the blue shift on denaturation in urea, shows peaks at 284 and 293 mμ, closely corresponding to the usual difference spectra of tryptophan derivatives. As Chervenka pointed out, it is to be expected that the tryptophan residues, in a protein of this composition, will have a dominating influence upon the spectrum.

Recently Nelson and Hummel (1962) studied the kinetics of the unfolding of ribonuclease in urea by observing the progress of the denaturation blue shift as a function of time. Their results (Fig. 5) show that, immediately after dissolving ribonuclease in 7 M urea, the difference spectrum for the urea solution, measured against the aqueous solution, shows a large positive value of $\Delta\epsilon$ in the region around 287 mμ. This corresponds to the usual red shift that is produced by increasing the refractive index of the solvent. In the course of several minutes this positive $\Delta\epsilon$ value decreases to zero and then continues downward to the large negative value characteristic of the denaturation blue shift. Nelson and Hummel have made use of the phenomenon in an elegant study of the kinetics of molecular unfolding in urea solutions, and of the refolding which occurs when the urea concentration is drastically reduced. Ultra-violet spectra should furnish a powerful tool for similar kinetic observations in other cases.

4. Formation of Ribonuclease-S: Difference Spectra and Kinetics

An interesting application of difference spectra to the study of ribonuclease structure has been reported by Richards and Logue (1962). The earlier studies of Richards and Vithayathil (1960) had shown in detail the effects of cleaving a single peptide bond in ribonuclease with subtilisin, to give the S-peptide (residues 1–20 of the enzyme) and the S-protein (residues 21–124 inclusive). Neither fragment is active alone, but the two combine with high affinity, even though no covalent bond unites them, to form the active enzyme ribonuclease-S. This, like native ribonuclease, contains three readily titratable tyrosine groups, and three which are not titratable in the native molecule. In the S-protein, however, which contains all six tyrosine groups, five are now readily titratable and only one is unavailable for titration. Richards and Logue (1962) have followed the changes in absorption spectrum that accompany the combination of S-peptide and S-protein in acid and neutral solution (Fig. 6). The striking increase in molar absorbancy in the 280–290 mμ region in formation of ribonuclease-S, undoubtedly reflects an increased "tightening" of the internal structure of the molecule which involves moving two tyrosine side chains into the interior, with an accompanying red shift of their absorption bands. (Some of the other tyrosine groups, of course, may be involved in the observed effects also.) The difference spectrum shown in part (*c*) of Fig. 6 indicates the perturbation of a phenylalanine group also, when S-peptide and S-protein combine; this may be the single phenylalanyl residue of S-peptide.

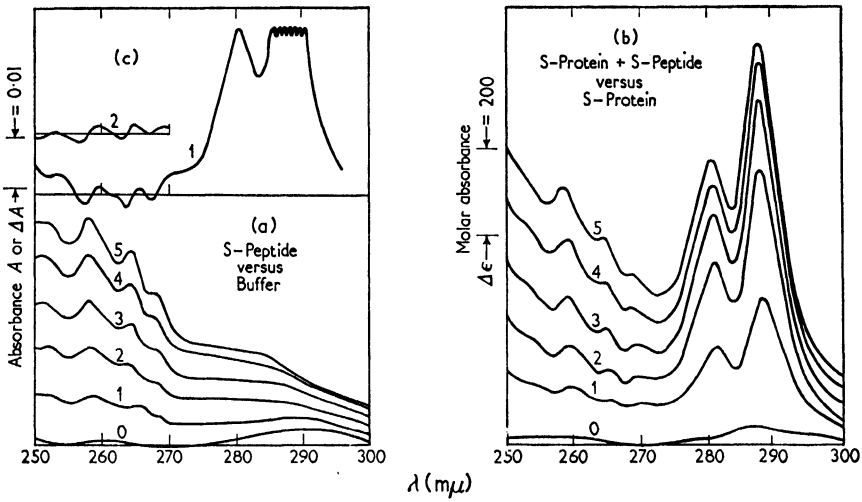


FIG. 6. Some spectra and difference spectra involving S-peptide and S-protein from ribonuclease. In all cases the solvent was 0·1 M acetate buffer pH 4·50, temperature 22°C, 1 cm cuvettes. (*a*) reference cell, acetate buffer; sample cell, S-peptide at $A \times 2{\cdot}9 \times 10^{-5}$ M, where A is the number of the curve as shown in the figure. (*b*) Reference cell, S-protein at $8{\cdot}7 \times 10^{-5}$ M (1·0 mg/ml); sample cell, same S-protein solution plus $A \times 2{\cdot}9 \times 10^{-5}$ M S-peptide, A as defined above. (*c*) Curve 1 represents the difference between curve 3 in part (b) and curve 3 in part (a); curve 2 represents a difference spectrum of *N*-acetyl-L-phenylalanine methyl ester scaled to represent a concentration of $8{\cdot}7 \times 10^{-5}$ M. From Richards and Logue (1962).

Richards and Logue studied the time dependence of the ultra-violet absorption to determine the kinetics of the combination of the two fragments. They observed a very rapid initial reaction, followed by an intramolecular re-arrangement with a half-time of the order of 1 min at 25°.

5. Spectral Changes accompanying the Acid Denaturation of Human Carbonic Anhydrase

Earlier I have discussed the work of L. M. Riddiford on the ionization of the tyrosine groups of fraction I of human carbonic anhydrase. The enzyme, as obtained from human erythrocytes, is readily separated into two major fractions (Rickli and Edsall, 1962) on hydroxylapatite columns. Our physico-chemical studies have hitherto been concentrated on fraction I, which is present in larger amount than fraction II, although its specific enzyme activity is lower. The activity of the enzyme is remarkably stable in alkaline solutions up to pH 12 or above, as

Riddiford's work indicates; but it is rapidly and apparently irreversibly lost in acid solution at pH 4. Accompanying this transition there is a sharp rise in sedimentation constant and in specific viscosity, in a manner indicating aggregation. The work of Dr. E. E. Rickli also indicates that the single zinc atom is released under these conditions. The preparation can be re-dissolved, either in acid solution below pH 2, or in neutral solution, but we have as yet failed to find any conditions for restoring enzyme activity after exposure to pH below 4.

When the native enzyme at pH near 7 is compared with the acid-denatured enzyme at pH below 2, the difference spectrum shows three peaks, as shown by the work of Dr. Ghazanfar (see Fig. 7). The two peaks

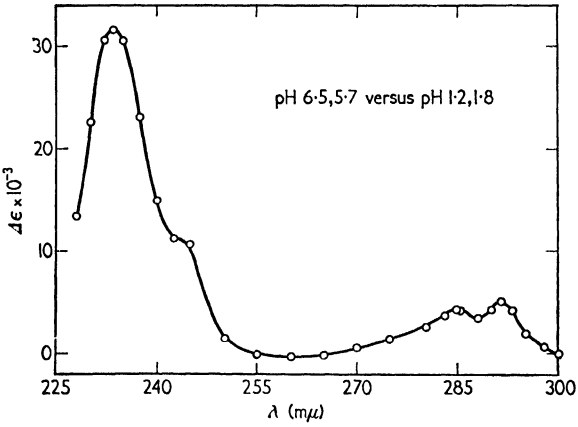


FIG. 7. Difference spectrum of human carbonic anhydrase: neutral solution *v.* acid denatured solution. From work of S. A. S. Ghazanfar (unpublished).

in the longer wavelength region at 285 and 292 mμ are characteristic of tryptophan difference spectra; the latter is the stronger; the molar value of $\Delta\epsilon_{292}$ being about 5000 per mole of protein. The molecule contains seven to eight tyrosine and six tryptophan residues; clearly the latter, although less numerous, dominate the observed difference spectrum. For comparison we note that $\Delta\epsilon_{293}$ for the ionization of the carboxyl group in tryptophan is near 300 (Hermans *et al.*, 1960). I would interpret the difference spectrum of carbonic anhydrase in Fig. 7 as due to molecular unfolding, with resulting exposure of tryptophan and tyrosine residues to the solvent. Ionization of acidic groups adjoining the tryptophan residues may possibly contribute to the difference spectrum, but its contribution is probably minor.

Figure 8 shows the difference spectrum at 285 and 292 mμ as a function of pH. The acid transition is centred near pH 3·5 and it is obviously

extremely sharp, indicating some kind of co-operative transition in the molecular conformation, which is triggered by the change in acidity.

By far the largest peak shown in the difference spectrum of Fig. 7 is centred near 235 mμ with a value of $\Delta\epsilon$ near 31,000. Both in relative height and in position it is closely similar to the 235 mμ peaks already reported by Glazer and Smith (1960, 1961*a*) in the difference spectra of a large number of proteins, when neutral solutions are compared with acid solutions near pH 2 or below. In all cases the neutral solutions show stronger absorption than the acid solution in the entire range of wavelength from 230 to 300 mμ. The origin of the peak near 235 mμ is still by

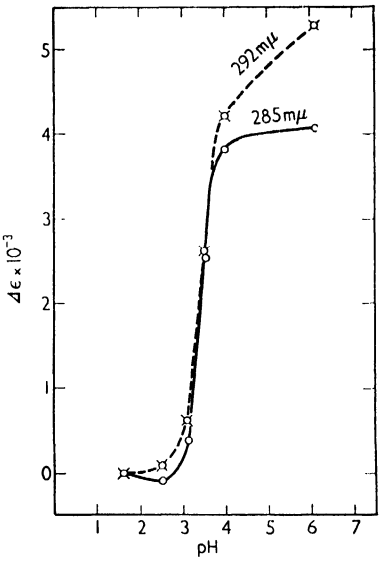


FIG. 8. pH Dependence of the difference spectrum of human carbonic anhydrase at 285 and 292 mμ. From work of S. A. S. Ghazanfar (unpublished).

no means completely clarified. Glazer and Smith infer that, in some cases at least, it is due primarily to the absorption of the peptide linkage and to changes in conformation of the peptide chain that accompany acidification. In some cases this may be true. However, David Eisenberg (1961) in our laboratory has made a careful study of the pH dependence of the difference spectra of human serum albumin. His results show that $\Delta\epsilon_{236}$ is exactly proportional to $\Delta\epsilon_{287}$ over the entire pH range from 1 to 7. The molar value of $\Delta\epsilon$ (pH 7–pH 1) for serum albumin ($\Delta\epsilon_{\text{maximum}}$) is 3800 at 287 m$\mu$ and 12 500 at 236 mμ, and the curves for $\Delta\epsilon$ as a function of pH, between 1 and 7, at these two wavelengths superimpose exactly if plotted as $\Delta\epsilon/\Delta\epsilon_{\text{maximum}}$. Thus it would appear that both these peaks, for

serum albumin, arise from changes in the environment of the tyrosine residues. This conclusion, of course, does not necessarily apply to the other proteins studied by Glazer and Smith. The situation is complex and requires further study.

Optical rotary dispersion studies, carried out by Ghazanfar in our laboratory, in association with Dr. Peter Urnes in the laboratory of Professor Doty, yield for native carbonic anhydrase a value of $b_0 = -4{\cdot}8$ when the data are formulated in terms of the Moffitt-Yang equation; a value that would commonly be interpreted as corresponding to a nearly complete absence of α-helix in the structure, in spite of the fact that it is a highly compact, globular molecule. In the acid denatured protein, the value of $b_0 = -106$ is considerably more negative than in the native protein, as if the content of helix had increased somewhat. Conclusions of this sort, derived from a single technique, must be regarded with great caution. In the following section we consider other evidence bearing on the same question.

6. Short Wave Ultra-violet Spectra: the Peptide Absorption Bands and the Helix–Coil Transition

Imahori and Tanaka (1959) were, as we have noted, the first to report a marked change in the ultra-violet absorption of a polypeptide (poly-L-glutamic acid) near 190 mμ, accompanying the helix–coil transition. The absorption of the random coil form, in the region near 190 mμ, was found to be markedly greater than that of the helix, even after correcting for the increased absorption due to the ionization of the carboxyl groups that accompanied the transition from helix (at low pH) to random coil (at high pH). Recent work, notably that of Rosenheck and Doty (1961) has extended these observations to other polypeptides and to certain proteins. If the attempt is made to draw inferences concerning the percentage of helix in a protein, from ultra-violet absorption at short wavelengths, a number of calculations are necessary. Some amino acid side chains show strong absorption in the region near 190 mμ; these include methionine, cystine, histidine, arginine, tryptophan, tyrosine and phenylalanine. The amide groups of asparagine and glutamine, and ionized carboxyl groups of aspartic and glutamic acid residues, also contribute to the absorption. To interpret the absorption spectrum of a protein in the neighbourhood of 190 mμ therefore requires knowledge of its amino acid composition, with appropriate corrections for the contribution of the various side chains, amide groups and ionized carboxyl groups to the absorption at the given wavelength. The necessary data have been tabulated by Rosenheck and Doty (1961). After correcting

for side-chain and other contributions they calculated the molar extinction coefficient for the peptide group at 190 mμ as 6900 per residue for the random coil and 4100 for the α-helix; at 197 mμ their estimated values were 6350 (coil) and 3200 (helix); at 205 mμ, they report 3200 (coil) and 2000 (helix). On this basis they estimated percentage helical content of paramyosin, myosin, insulin, ribonuclease and β-lactoglobulin from absorption measurements at these three wavelengths, corrected for the non-peptide absorption. In general there was a fairly close correlation between the estimated helical contents of the proteins, as derived from ultra-violet absorption and from rotatory dispersion.

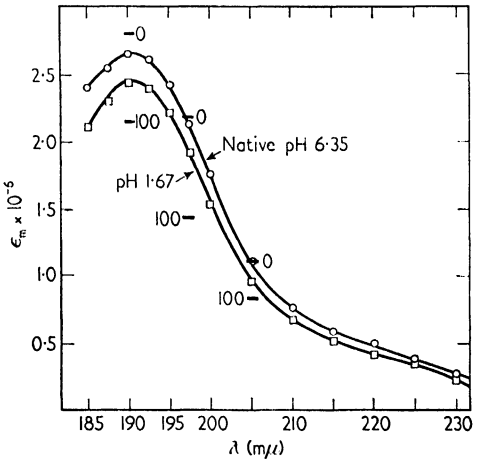


FIG. 9. Short-wave ultra-violet absorption of human carbonic anhydrase fraction I. Native protein at pH 6·35; acid denatured protein at pH 1·67. The short horizontal lines marked 0 and 100 in the figure indicate calculated absorption values corresponding to 0 and 100% helix at 197 and 205 mμ. See text. From work of S. A. S. Ghazanfar (unpublished).

Figure 9 presents similar data obtained by Dr. Ghazanfar on one of our preparations of human carbonic anhydrase fraction I. The measurements were made in Professor Doty's laboratory with the aid and advice of Dr. W. B. Gratzer. The figure shows the absorption spectrum of the carbonic anhydrase preparation, both for the native protein at pH 6·35 and for the acid denatured material at pH 1·67. The absorption curve of the acid denatured material, at all wavelengths shown in the figure, is distinctly lower than that of the native protein. This is concordant with Ghazanfar's data from optical rotatory dispersion, already mentioned, which indicate very low content of helix in the native protein and a somewhat greater helical content in the acid denatured protein. Following

the procedure employed by Rosenheck and Doty (1961) we have introduced scale marks corresponding to the calculated values for 0 and 100% helix, for the three wavelengths 190, 197 and 205 mμ. These are calculated values, based on the assumptions outlined above, for a protein with the amino acid composition of carbonic anhydrase. At 197 and at 205 mμ the experimental value of ϵ is very close to the value calculated for zero helix content. The value at 190 mμ would suggest the possible presence of some helix, perhaps 20–25%. The curve for the acid denatured material might correspond to a helix content of 30–50%. Only rough estimates are justified; the data however do suggest that the native protein contains very little helix, the acid denatured protein somewhat more.

Dr. Ghazanfar has also observed, by means of infra-red absorption, the rate of exchange of hydrogen for deuterium in carbonic anhydrase when it is dissolved in D_2O. From this he has calculated the percentage of "hard-to-exchange amide hydrogen" by the method of Blout *et al.* (1961). The resulting data are compatible with the assumption that the helix content of the native protein is very low and that it increases somewhat upon denaturation in acid; qualitatively, therefore, they are in accord with the studies of rotary dispersion and short wave ultra-violet absorption.

Personally I would emphasize the uncertainty of the inferences that can be drawn concerning the helix content of proteins by any or all of these methods. A simple polypeptide such as poly-L-glutamic acid generally exists only in two forms, the helix and the random coil; a long peptide chain may, of course, contain some helical and some randomly coiled regions. A globular protein in its native state, however, can be made up of some helical regions, and of others that are very far from helical in character, but which nevertheless are not random but are restricted to quite definite conformations, due to the rigidity of the framework of the whole protein molecule. There appears to be no adequate theory at present to predict the contribution of such regions to the optical rotation or to the ultra-violet absorption of the peptide linkages. Moreover, the amide hydrogens of peptide chains may be unable to exchange readily with deuterium, due to any one of a variety of structural features that prevent them from coming into contact with the solvent. There is no necessity to assume, and in general it is perhaps rather unlikely, that such shielded peptide hydrogens are always part of a helical structure. At present therefore I regard estimates of helical content in proteins as a kind of shorthand summary of certain numerical values derived from the experimental data. I would not take them as clear-cut evidence of the actual content of helix in the protein until we know much more about the detailed three-dimensional structure of a

considerable number of proteins, and can correlate these structural data with other kinds of physical measurement.

Recent studies have begun to establish fundamental relations between short wave ultra-violet absorption and the optical rotary dispersion of helical structures. Moffitt (1956) concluded that the $N \rightarrow V_1$ transition of the peptide chromophore would give rise to two bands, due to resonance exciton interaction of the electronic transition moments of the oriented peptide groups in the helix. One component should be polarized parallel to the helix axis, and there should be a degenerate pair with perpendicular polarization. The recent work of Gratzer *et al.* (1961), with oriented films of polypeptides, affords striking confirmation of Moffitt's conclusions.

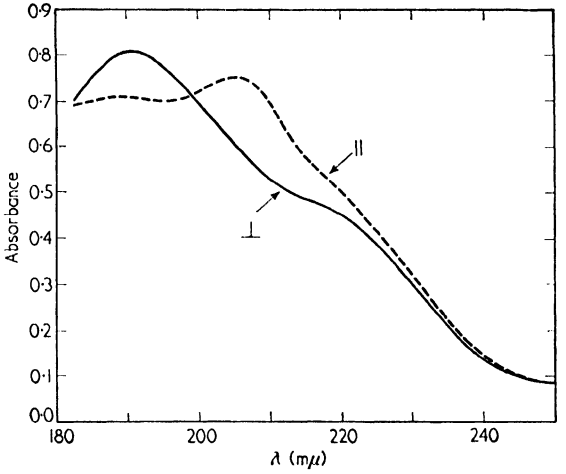


FIG. 10. Short-wave ultra-violet spectra of an oriented film of poly-γ-methyl-L-glutamic acid in helical form. The curves are shown when the incident beam is polarized parallel and perpendicular to the axis of the helix. From work of Gratzer *et al.* (1961).

Figure 10 shows the polarization spectrum of a film of poly-γ-methyl-L-glutamate; the band maxima lie at 191 and 206 mμ, and they are clearly polarized perpendicular and parallel, respectively to the helix axis, with a separation of 2700 cm^{-1}, close to Moffitt's prediction. There is also a strong indication in Fig. 10 of a shoulder on the long wavelength side, corresponding to a much weaker absorption band centred near 222 mμ. There are strong reasons, as Gratzer *et al.* have indicated, for believing that this corresponds to an $n \rightarrow \pi^*$ transition. Recent studies by rotatory dispersion (Blout *et al.*, 1962) and circular dichroism (Holzwarth *et al.*, 1962) have revealed a positive Cotton effect at 190 mμ for helical peptides, and for several native proteins, in addition to the previously

recognized Cotton effect near 225 mμ. These effects appear to be related to the absorption bands in the same regions; the n→ π* transition near 225 mμ, though very weak in absorption, may make a major contribution to the optical rotation. Knowledge in this field is developing rapidly; these new developments are briefly mentioned here simply to call attention to their interest and importance.

Acknowledgements

I am indebted to Professor Paul Doty and Dr. W. B. Gratzer for helpful comments. The researches in our laboratory discussed in this paper were supported by grants from the National Science Foundation (G9116) and the United States Public Health Service (H-3169).

References

Baba, H. and Suzuki, S. (1961). *J. chem. Phys.* **35**, 1118.

Beaven, G. H. (1961). In *Advances in Spectroscopy*, Vol. II, p. 331. Interscience, New York.

Bigelow, C. C. (1960). *C. R. trav. lab. Carlsberg* **31**, 305.

Bigelow, C. C. and Geschwind, I. I. (1960). *C. R. trav. lab. Carlsberg* **31**, 1.

Blout, E. R., de Lozé, C. and Asadourian, A. (1961). *J. Amer. chem. Soc.* **83**, 1895.

Blout, E. R., Schmier, I. and Simmons, N. W. (1962). *J. Amer. chem. Soc.* **84**, 3193.

Braunitzer, G., Gehring-Muller, R., Hilschmann, N., Hilse, K., Hobom, G., Rudloff, V. and Wittmann-Liebold, B. (1961). *Hoppe-Seyl. Z.* **325**, 283.

Breslow, E. (1962). *J. biol. chem.* **237**, PC3308.

Chervenka, C. H. (1959). *Biochim. biophys. Acta* **31**, 85.

Crammer, J. L. and Neuberger, A. (1943). *Biochem. J.* **37**, 302.

Donovan, J. W., Laskowski, M., Jr. and Scheraga, H. A. (1961). *J. Amer. chem. Soc.* **83**, 2686.

Doty, P. and Gratzer, W. B. (1962). In *Polyamino Acids, Polypeptides and Proteins* (M. A. Stahmann, ed.), p. 111. University of Wisconsin Press, Madison.

Eisenberg, D. (1961). Honours Thesis in Biochemical Sciences, Harvard University.

Glazer, A. N. and Smith E. L. (1960). *J. biol. Chem.* **235**, PC43.

Glazer, A. N. and Smith E. L. (1961*a*). *J. biol. Chem.* **236**, 2942.

Glazer, A. N. and Smith, E. L. (1961*b*). *J. biol. Chem.* **236**, 2948.

Gratzer, W. B., Holzwarth, G. M. and Doty, P. (1961). *Proc. nat. Acad. Sci., Wash.* **47**, 1785.

Havsteen, B. H. and Hess, G. P. (1962). *J. Amer. chem. Soc.* **84**, 448.

Hermans, J., Jr. (1962). *Biochemistry* **1**, 193.

Hermans, J., Jr., Donovan, J. W. and Scheraga, H. A. (1960). *J. biol. Chem.* **235**, 91.

Herskovits, T. T. and Laskowski, M., Jr. (1960). *J. biol. Chem.* **235**, PC56.

Herskovits, T. T. and Laskowski, M., Jr. (1962*a*). *J. biol. Chem.* **237**, 3418.

Herskovits, T. T. and Laskowski, M., Jr. (1962*b*). *J. biol. Chem.* **237**, 1481.

Hill, Robert J., Konigsberg, W., Guidotti, G. and Craig, L. C. (1962). *J. biol. Chem.* **237**, 1549.

Holzwarth, G., Gratzer, W. B. and Doty, P. (1962). *J. Amer. chem. Soc.* **84**, 3194.
Imahori, K. and Tanaka, J. (1959). *J. mol. Biol.* **1**, 359.
Kasha, M. (1961). In *Light and Life* (W. D. McElroy and B. Glass, eds.), p. 31. Johns Hopkins Press, Baltimore.
Kawahara, F. S., Wang, S.-R. and Talalay, P. (1962). *J. biol. Chem.* **237**, 1500.
Kendrew, J. C. (1962). *Brookhaven Symposia in Biology* (in press).
Leonard, W. J., Jr. and Foster, J. F. (1961). *J. biol. Chem.* **236**, 2662.
Lóonis, J. and Li, C. H. (1959). *J. Amer. chem. Soc.* **81**, 415.
Moffitt, W. (1956). *Proc. nat. Acad. Sci., Wash.* **42**, 736; *J. chem. Phys.* **25**, 467.
Nagakura, S. and Gouterman, M. (1957). *J. chem. Phys.* **26**, 881.
Nelson, C. A. and Hummel, J. P. (1962). *J. biol. Chem.* **237**, 1567.
Richards, F. M. and Logue, A. D. (1962). *J. biol. Chem.* **237**, 3693.
Richards, F. M. and Vithayathil, P. J. (1960). *Brookhaven Symposia in Biology*, **13**, *Protein Structure and Function*, p. 115.
Rickli, E. E. and Edsall, J. T. (1962). *J. biol. Chem.* **237**, PC58.
Rosenheck, K. and Doty, P. (1961). *Proc. nat. Acad. Sci., Wash.* **47**, 1775.
Shugar, D. (1952). *Biochem. J.* **52**, 142.
Smillie, L. B. and Kay, C. M. (1961). *J. biol. Chem.* **236**, 112.
Tanford, C. and Wagner, M. L. (1954). *J. Amer. chem. Soc.* **76**, 3331.
Tanford, C., Hauenstein, J. D. and Rands, D. G. (1955*a*). *J. Amer. chem. Soc.* **77**, 6409.
Tanford, C., Swanson, S. A. and Shore, W. S. (1955*b*). *J. Amer. chem. Soc.* **77**, 6414.
Tanford, C., Buckley, C. E., De, P. K., and Lively, E. P. (1962). *J. biol. Chem.* **237**, 1168.
Wetlaufer, D. B. (1956). *C. R. trav. lab. Carlsberg, Ser. Chem.* **30**, 135.
Wetlaufer, D. B. (1962). In *Advances in Protein Chemistry* (M. L. Anson and J. T. Edsall, eds.), Vol. 17, p. 303. Academic Press, New York.
Wetlaufer, D. B., Edsall, J. T. and Hollingworth, B. R. (1958). *J. biol. Chem.* **233**, 1421.
Williams, E. J. and Foster, J. F. (1959). *J. Amer. chem. Soc.* **81**, 865.
Wishnia, A., Weber, I. and Warner, R. C. (1961). *J. Amer. chem. Soc.* **83**, 2071.
Yanari, S. and Bovey, F. A. (1960). *J. biol. Chem.* **237**, 2818.

Discussion

K. S. V. SAMPATHKUMAR: To what extent may the absorption changes at acid pH be due to the removal of metal from carbonic anhydrase? Can you get similar changes if the metal is removed at neutral pH with chelating agents?

J. T. EDSALL: Apparently the changes are not due to the removal of the metal. S. Lindskog and Bo. G. Malmström (*J. biol. Chem.* **237**, 1129 (1962)) have made optical rotatory dispersion measurements on bovine carbonic anhydrase, both on the native and on the undenatured metal-free enzyme, and have found no difference between them in this respect. If their results apply to the enzyme from human blood also, as they probably do, we cannot explain the alterations occurring in acid solution simply in terms of removal of the zinc.

M. S. NARASINGA RAO (Regional Research Laboratory, Hyderabad): You mention that the helical content of a protein may be evaluated by measuring the absorption in the 190 mμ region. Is this only an empirical method or does it have a theoretical basis?

J. T. EDSALL: I would regard estimates of helix content from absorption near 190 mμ as empirical at present. The corrections for absorption due to the amino

acid side chains are important; their accuracy depends on the assumption that the side chains in the protein make the same contribution to the absorption as in the corresponding free amino acid; this is not likely to be exactly true. Also the accuracy of the amino acid analysis of the protein is obviously crucial in making this correction. On the other hand, it is true, as indicated in my paper, that Gratzer *et al.* (1961) have confirmed the major general conclusions of Moffitt's theory for the absorption due to a helical array of peptide linkages; but the theory is not precise enough to give exact numerical values of the absorption. Inevitably therefore the calculation involves the use of numerical data obtained from measurements on simple substances (amino acids and peptides) which are taken as reference materials. Of course I should point out that the calculation of helix content from optical rotatory dispersion is also to be regarded as empirical in the same sense.

M. S. NARASINGA RAO: One of the usual methods of determining the availability of a group in a protein for titration has been to determine the pK and heat of dissociation of the group and compare them with those of the group in a small molecule like a peptide. In your opinion, to what extent is this valid?

J. T. EDSALL: The pK value, also the ΔH^0 and ΔS^0 values characteristic of a particular group, may be greatly modified when it is incorporated into a protein. For instance, Tanford showed that ΔH^0 and ΔS^0 for the tyrosyl groups in serum albumin were quite different from the values characteristic of simple tyrosyl peptides. The whole situation is well reviewed by Tanford in the forthcoming Volume 17 of "Advances in Protein Chemistry", Academic Press, New York.

Behaviour in Solution of Polypeptides Related to Collagen

EPHRAIM KATCHALSKI, ARIEH BERGER AND JOSEPH KURTZ

Department of Biophysics, The Weizmann Institute of Science, Rehovoth, Israel

ABSTRACT

Hydrodynamic and optical rotatory studies have shown that the two conformational forms of poly-L-proline, the structures of which in the solid state have been elucidated, can persist also in solution. Form I ($[\alpha]_D \simeq +50°$), characterized by a right-handed helix with all peptide bonds in the *cis* configuration, is stable in poor solvents such as pyridine and aliphatic alcohols. Form II ($[\alpha]_D \simeq -540°$), a left-handed helix with all peptide bonds in the *trans* configuration, is stable in good solvents such as formic acid and water. The transformation of the poly-L-proline I helix into the poly-L-proline II helix and vice versa, in the forward and reverse mutarotations respectively, was shown to be the result of a series of acid catalysed *cis–trans* isomerizations of peptide bonds. The macromolecular asymmetry of poly-L-proline II could be destroyed by neutral salts in high concentrations ($[\alpha]_D = -240°$ in 12 M aqueous LiBr).

Copolymers of L-proline with *O*-acetylhydroxy-L-proline, glycine or sarcosine were synthesized by simultaneous polymerization of the corresponding *N*-carboxy-anhydrides. With the two former amino acids random copolymers were obtained, whereas the latter yielded block copolymers.

Poly-3,4-dehydro-L-proline was shown to exist in solution in two forms (I and II) similar in their conformations to the corresponding forms of poly-L-proline.

Two forms of poly-*O*-acetylhydroxy-L-proline were also detected. Form II ($[\alpha]_D \simeq -175°$) is stable in formic acid; Form I ($[\alpha]_D \simeq +25°$) is stable in pyridine, dimethylformamide and acetic anhydride. Forward and reverse mutarotations were shown to occur in the appropriate solvents. In acetic anhydride–$HClO_4$, stable intermediate forms were observed.

Poly-L-pipecolic acid and poly-D-pipecolic acid were obtained by the polymerization of the corresponding *N*-carboxyanhydrides. Poly-L-pipecolic acid as isolated from the polymerization mixture in dioxane (Form I) has a specific optical rotation of $[\alpha]_D = -325°$ in acetic acid. Form I mutarotates in formic acid to yield Form II with $[\alpha]_D = -50°$ in acetic acid. Both forms give different X-ray powder diagrams, as well as different infra-red absorption spectra in chloroform.

1. INTRODUCTION

During the past few years theories on the structure of collagen have in general agreed that the proline and hydroxyproline residues, which amount to between 15 to 30% of the total weight of the protein, are of crucial importance in the molecular pattern of this substance. The

elegant proposal of Ramachandran and Kartha (1954), that collagen consists of three helical polypeptide chains coiled about each other in a three-stranded rope, has been given strong support by the X-ray studies of Rich and Crick (1955, 1958, 1961) and Cowan, McGavin and North (1955). At present the most promising proposed structure for collagen appears to be one in which each of the polypeptide chains making up the collagen molecule is coiled into a left-handed helix of the type described for poly-L-proline II (Cowan and McGavin, 1955; Sasisekharan, 1959) and for polyglycine II (Crick and Rich, 1955). The striking similarity in the X-ray diffraction spacings of poly-L-proline II and stretched collagen strongly suggest similar chain configurations, and it is, therefore, reasonable to suppose that a study of the solution properties of poly-L-proline and its copolymers should help substantially in understanding the chemistry of collagen.

In the following a description is given of the configurational changes of poly-L-proline in solution. The mutarotation of poly-3,4-dehydro-L-proline, is discussed, and the behaviour in solution of polyhydroxy-L-proline and some of its derivatives is described. A short discussion of the structure and conformation of some proline containing copolymers is also included. Finally, the synthesis of poly-L- and poly-D-pipecolic acid is briefly described, and the conformational changes in solution of these optically active polypeptides are compared with those of poly-L-proline.

2. Poly-L-proline

It has been demonstrated (Kurtz *et al.*, 1956) that the synthetic polymer, poly-L-proline, obtained by the polymerization of *N*-carboxy-L-proline anhydride, can exist in two forms which exhibit markedly different optical rotations. If the polymer is precipitated from the polymerization medium (pyridine) with ether and redissolved in acetic acid, its initial specific rotation, $[\alpha]_D^{25}$, is $+50°$. This material has been denoted Form I (Blout and Fasman, 1958; Harrington and Sela, 1958). On standing, however, the acetic acid solution slowly changes in rotation over a period of several days, reaching a final value of $[\alpha]_D^{25} = -540°$ (Kurtz *et al.*, 1956; 1958*a*). The polyproline with this specific rotation has been denoted Form II. The mutarotation may be greatly accelerated by heating, but once the state characterized by an $[\alpha]_D = -540°$ has been reached, no further optical rotatory changes are observed. A number of other solvents including formic acid, benzyl alcohol and chloroethanol, also favour the mutarotation of Form I into Form II. Thus formic acid brings about the mutarotation at 30° in less than 5 min.

Later studies have demonstrated that these optical rotatory changes may be reversed by means of certain solvent systems (Steinberg *et al.*,

1958). Thus, a solution of Form II in a solvent consisting of 10% acetic acid and 90% 1-propanol gradually decreases in laevorotation, the final value of $[\alpha]_D$ approaching that of Form I. Since this cyclic process of mutarotation in water (or in formic or acetic acid) followed by reverse mutarotation in the acetic (or formic) acid–1-propanol (or 1-butanol) system may be repeated indefinitely, and since both Form I and Form II yield L-proline on hydrolysis, it may be concluded that the variations in specific optical rotation observed are not induced by chemical changes but are rather a reflection of configurational transitions in the individual polymer molecules. This view has been supported by infra-red studies (Blout and Fasman, 1958; Steinberg *et al.*, 1958) showing that Form I and Form II exhibit distinctly different infra-red spectra in the solid state. Moreover, the X-ray diffraction powder diagrams of Form I and Form II (Cowan and McGavin, 1955; Sasisekharan, 1959) suggest substantial differences between the molecular structure of the two forms.

In early studies it has been suggested (Kurtz *et al.*, 1956) that the polyproline I–polyproline II interconversion results from a series of *cis-trans* isomerizations at the peptide bonds of the polymer. This possibility has been explored (Harrington and Sela, 1958) and shown to be consistent with the observed viscosity, sedimentation and optical rotatory properties of the two forms. It was, therefore, suggested that Form I in solution is a right-handed helix with all its peptide bonds in the *cis* configuration, whereas Form II in solution has the configuration of the left-handed helix described by Cowan and McGavin (1955) with all the peptide bonds in the *trans* configuration.

The structure of polyproline I in the solid state, has been determined recently by Traub and Shmueli (1963). Form I forms, in accord with the proposal described above, a right-handed helix with a translation of 1·90 Å and a rotation around the helix axis of 108° per proline residue. The peptide groups are in the *cis* configuration. A detailed re-examination of the X-ray diffraction photographs of Form II (Sasisekharan, 1959) has confirmed that this form of poly-L-proline possesses, in the solid state, a left-handed helical conformation with peptide bonds in the *trans* configuration. The structure has three residues per turn of the helix, and a repeat distance of 3·12 Å per residue along the longitudinal axis.

In the following a summary of the optical rotatory properties of poly-L-proline in various solvent systems is given, and the kinetics of interconversion of the two forms of polyproline is described. It will be shown that the right- and left-handed helical conformations suggested for Forms I and II respectively in the solid state, prevail under suitable conditions, also in solution.

Two accounts on the behaviour of poly-L-proline in solution have recently been published (Downie and Randall, 1959; Steinberg *et al.*, 1960).

(a) *Hydrodynamic properties*

It seems clear that any proposal concerning the spatial arrangement of the L-proline residues of Form I and Form II of poly-L-proline in solution should be consistent with the hydrodynamic properties of the two forms. If Form I and Form II retain in solution the conformations they have in the solid state, both should exhibit the hydrodynamic

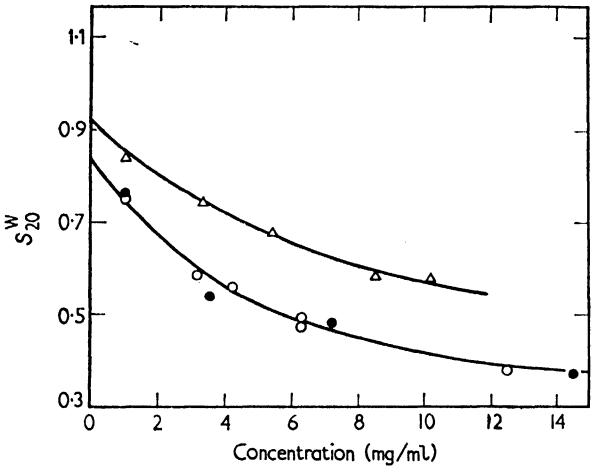


FIG. 1. Sedimentation coefficients of poly-L-proline (sample B4) as a function of concentration: △——△, Form I in propionic acid; ●——●, Form II in propionic acid; ○——○, Form II in acetic acid. From Steinberg *et al.* (1960).

characteristics of rigid, highly asymmetric macromolecules. Furthermore, the hydrodynamic friction coefficient of a given sample of poly-L-proline in Form II is expected to exceed that of the same sample in Form I.

The mutarotation of Form I in propionic acid was found to be extremely slow at room temperature. It was thus possible to determine the intrinsic viscosities, and sedimentation and diffusion coefficients of both forms in this solvent system. The curves given in Fig. 1 for a poly-L-proline sample (B4) with a number average molecular weight of 19,000 show, as expected, that within the whole range of concentrations measured, the sedimentation coefficients of Form II are lower than those of Form I. Significant differences were also observed between the reduced viscosity η_{sp}/c, of Form I and II in propionic acid (Fig. 2).

Attempts were made to calculate the axial ratios from viscosity data alone, using the equation of Simha (1940) (Eq. 1), where ν is the well-known shape factor.

$$[\eta] = N\nu V_e/100M \tag{1}$$

Assuming tentatively that the volume of the effective hydrodynamic ellipsoid, V_e, is equal to the molecular volume $M\bar{v}/N$, where $\bar{v}$ is the partial specific volume (a value of $\bar{v} = 0{\cdot}758$ cm^3/g was found for poly-L-proline, Cowan and McGavin, 1955) an axial ratio of 40 for Form II of sample B4 was found. Since this ratio is close to that derived from the

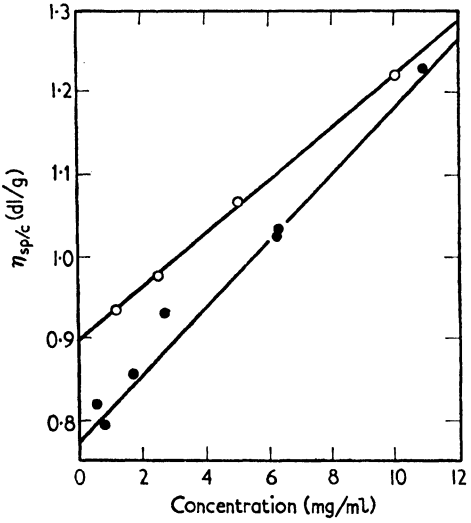


FIG. 2. The reduced viscosity of poly-L-proline (sample B4) in propionic acid at 30° as a function of concentration: ○——○, Form II; ●——●, Form I. From Steinberg *et al.* (1960).

rigorous Scheraga and Mandelkern equation (1953), it was felt that, for purposes of comparison, a rough estimate of the axial ratios of the prolate effective hydrodynamic ellipsoids of other polyproline samples may be obtained, from viscosity data alone, making the above assumption ($N\bar{v}/M \simeq V_e$). The axial ratios thus calculated for a number of poly-L-proline samples, as Form II, with average molecular weights of 12,000, 14,300, 19,000, and 52,000 are 33 (49), 36 (59), 40 (78), and 52 (214) respectively. However, the axial ratios of the equivalent ellipsoids derived from the intrinsic viscosities were markedly smaller than those calculated for rod-like particles possessing the Cowan-McGavin (1955) configuration (figures given in brackets). This discrepancy becomes larger with increasing molecular weight, indicating that if poly-L-proline

(Form II) helices persist in solution, they possess a certain degree of flexibility.

Viscosity measurements of polyproline samples in Form I seem to show that the polypeptide possesses marked molecular asymmetry also in this configuration. From the intrinsic viscosity of sample B4, Form I, ($[\eta] = 0{\cdot}78$ dl/g in propionic acid) an axial ratio of 37 was derived following the reasoning discussed above. A similar calculation for sample P8 ($[\eta] = 0{\cdot}56$ dl/g in propionic acid) gave an axial ratio of 29. If the molecular axial ratios are calculated from the molecular dimensions proposed by Traub and Shmueli (1963) for poly-L-proline I, values of 39 and 27 are obtained for samples B4 and P8 respectively. This agreement obtained with low molecular weight materials does not hold, however, for high molecular weight preparations. Thus, a sample of molecular weight 52,000 (Blout and Fasman, 1958) showed, in Form I an intrinsic viscosity of 0·99 dl/g in acetic acid corresponding to an axial ratio of 44, whereas the calculated value for this degree of polymerization is 116.

(b) *Kinetics of forward mutarotation in acetic acid*

The change in the specific optical rotation of poly-L-proline I in glacial acetic acid at 25° at three different concentrations is given in Fig. 3. The course of the reaction is seen to be practically independent of concentration in the range studied (0·25 to 2·0 g/100 ml). This independence on concentration was also observed in other solvent systems and with other polymer samples. Similar results were obtained in the studies of the reverse mutarotation. These findings justify the use of $[\alpha]$ as the variable in the presentation of the kinetic experiments discussed below.

In order to evaluate the apparent order of the forward mutarotation in acetic acid, values of $\log \mathrm{d}[\alpha]_t/\mathrm{d}t$ taken from an experiment performed at 44°, were plotted against $\log([\alpha]_t - [\alpha]_\infty)$. A straight line with a slope of 4/3 was obtained. The course of the reaction may thus be represented by

$$-\frac{\mathrm{d}[\alpha]_t}{\mathrm{d}t} = k([\alpha]_t - [\alpha]_\infty)^{4/3}$$

where k is a constant.

In the experiments described above, in which the mutarotation was studied at different polymer concentrations, c_0, it was found that $\mathrm{d}\alpha/\mathrm{d}t$, i.e. the measured rate of change in optical rotation is, at any given $[\alpha]$, proportional to c_0. This would indicate that the mutarotation reaction obeys first-order kinetics. On the other hand, we demonstrated that in any single experiment, i.e. when c_0 is constant, the change in $[\alpha]$ with time seems to proceed according to an order of 4/3. This seeming contradiction stems from the fact that the observed mutarotation is the result of a

series of intramolecular configurational changes in every single macromolecule. A theoretical analysis of the kinetics of mutarotation as a function of total concentration of L-proline residues, as well as of residues associated with a *cis* or *trans* peptide bond has been published elsewhere (Steinberg *et al.*, 1960).

In order to evaluate the enthalpy of activation, the mutarotation in acetic acid was carried out at several temperatures between 30 and 45°. An enthalpy of activation $\Delta H^* = 20{\cdot}6$ kcal per mole peptide bond was obtained from a plot of $\log k$ *versus* the reciprocal of the absolute temperature.

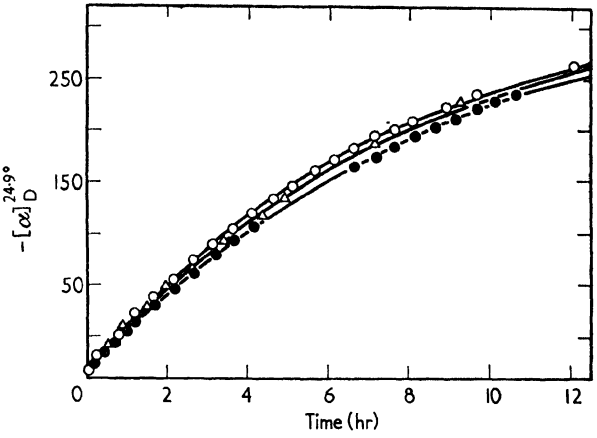


FIG. 3. The forward mutarotation of poly-L-proline (sample B1, Form I) in glacial acetic acid at 24·9° at three different concentrations: ●——●, 2 g per 100 ml; △——△, 0·5 g per 100 ml; ○——○, 0·25 g per 100 ml. From Steinberg *et al.* (1960).

(c) *Kinetics of reverse mutarotation*

When a solution of poly-L-proline II in formic, acetic or propionic acid is diluted tenfold with 1-propanol, the specific rotation $[\alpha]_D = -370°$ changes over several days, at room temperature, to a final value of about $[\alpha]_D = -20°$. The reverse mutarotation at different temperatures is given in Fig. 4. The solvent was acetic acid–1-propanol (1 : 9 v/v), and the polymer concentration (sample B4) was 0·25%. Similar data were obtained with 0·5% solutions. First-order kinetics was obeyed throughout to a good approximation.

The Arrhenius plot of $\log k$ against $1/T$ gives $\Delta H^* = 20{\cdot}2$ kcal/mole peptide bond demonstrating that the forward and reverse mutarotation reactions have essentially identical activation energies.

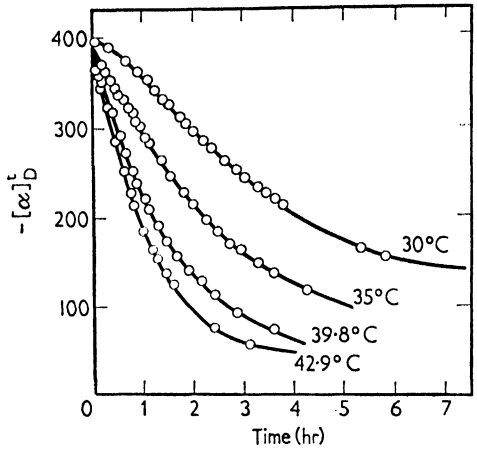


FIG. 4. Reverse mutarotation of poly-L-proline (sample B4) in acetic acid–propanol (1:9 v/v) at different temperatures. From Steinberg *et al.* (1960).

(d) *Acid catalysis of mutarotation*

The rate of forward mutarotation in aliphatic acids is greatly increased by minute quantities of strong mineral acids such as perchloric acid. In Fig. 5 the rate of forward mutarotation in propionic acid at 20% conversion ($[\alpha]_D = -80°$) is plotted against the molar ratio of perchloric acid to peptide bonds. It can be seen that amounts of perchloric acid from 0·027 to 0·047 mole per mole peptide bond caused a linear increase in the velocity of mutarotation. Larger amounts of acid caused precipitation. The addition of $HClO_4$ up to an amount of 0·027 mole per mole

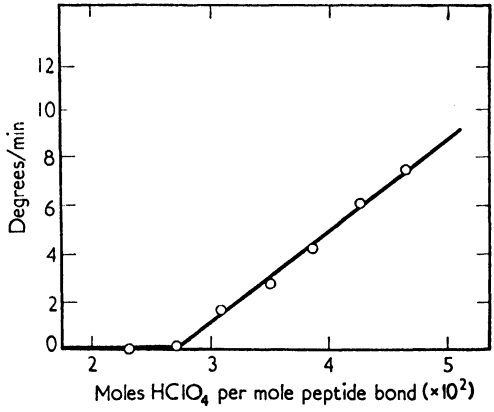
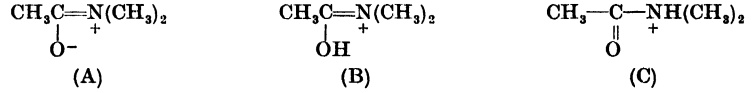


FIG. 5. The rate of forward mutarotation of poly-L-proline (sample B4) at $[\alpha]^{30}_D - 80°$ (corresponding to 20% conversion) as a function of moles $HClO_4$ per mole peptide bond; temperature 30°. From Steinberg *et al.* (1960).

peptide bond had no catalytic effect. This was found to be due to the presence of titrable basic groups. The strong catalytic effect of $HClO_4$ on the forward mutarotation was also observed in glacial acetic acid as well as in aqueous acetic acid.

The reverse mutarotation was found to be similarly catalysed by $HClO_4$. This effect was investigated in detail in the case of poly-*O*-acetylhydroxy-L-proline in acetic anhydride (see section 5). In this case it was also found that the state of equilibrium between Forms I and II depends on the concentration of perchloric acid.

The experiments on acid catalysis provide important evidence for the occurrence of *cis-trans* isomerization of the peptide bonds during mutarotation. The absence of free rotation about the C—N bond in amides, due to its partial double-bond character is well established. The mechanism of protonation and its effect on the rotation of the C—N bond in *N*,*N*-dimethylacetamide has been investigated by Berger *et al.* (1959) using the nuclear magnetic resonance technique. It was demonstrated that under strongly acid conditions free rotation takes place according to a mechanism involving an equilibrium between the three species (A), (B) and (C), of which the last one is capable of free rotation.



(A) (B) (C)

It may be inferred that in the polymers of proline, protonation results in a "loosening" of the peptide bonds allowing *cis-trans* isomerization to occur.

Support for the proposed mechanism is obtained from several experiments demonstrating the strong binding of acid by polyproline and related polymers. Polyproline precipitates from its solution in acetic acid on the addition of anhydrous solutions of perchloric acid. The precipitates obtained with either polyproline I or polyproline II contained from 0·26 to 0·31 mole $HClO_4$ per mole peptide bond. When the precipitates derived from either form were dissolved in water, the specific rotation was that of Form II. Proton binding in solution by peptides containing cyclic secondary amino acids could be demonstrated by potentiometric titration with $HClO_4$ in acetic anhydride. Since poly-L-proline is insoluble in acetic anhydride experiments were carried out with poly-*O*-acetylhydroxy-L-proline, and the results are reported in section 5.

(e) *Viscosity changes during mutarotation*

The viscosity of poly-L-proline in Form I is always less than that of Form II, in a given solvent at the same concentration. Forward and

reverse mutarotations are, therefore, necessarily accompanied by changes in viscosity. Figure 6 shows these changes during mutarotation under three different conditions. The three mutarotation experiments described graphically, demonstrate that a poly-L-proline molecule may attain different shapes, as reflected by different hydrodynamic properties, but still possess the same optical rotation.

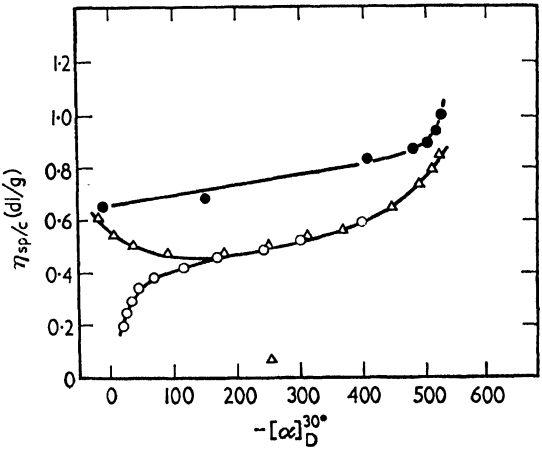


FIG. 6. Reduced viscosity ($c = 1{\cdot}0$ g per 100 ml) at 30° as a function of $[\alpha]_D^{30}$: ●——●, data taken during forward mutarotation in glacial acetic acid; △——△, data taken during forward mutarotation in acetic acid–water (7:3 v/v); ○——○, data taken during reverse mutarotation in acetic acid–(propanol (2:8 v/v); △, value in aqueous 12 M LiBr. From Steinberg *et al.* (1960).

(f) *Optical rotatory properties of low molecular weight proline derivatives*

Optical rotation measurements were carried out with *p*-toluensulphonyl-L-prolyl-L-proline and L-proline anhydride in acetic acid, acetic acid–propanol (1:9 v/v), water and saturated aqueous LiBr. The toluenesulphonyl derivative had an $[\alpha]_D^{25} = -147°$ in the first two solvents, and the anhydride gave $[\alpha]_D^{25}$ values of $+135°$–$147°$ in the four solvents investigated. None of the optical rotations changed with time. These findings show that solvents which produce large optical rotatory changes in poly-L-proline have little effect on the above simple proline containing substances. The observations made confirm the view that the optical rotatory changes of poly-L-proline in solution result from configurational alterations along the polypeptide chain.

(g) *The effect of neutral salts on the configuration of poly-L-proline*

Configurational changes occur in poly-L-proline under the influence of neutral salts such as LiBr, $CaCl_2$ and KCNS at high concentrations

(Harrington and Sela, 1958). These changes seem to be of an entirely different nature from the ones taking place because of *cis–trans* isomerizations. Poly-L-proline II dissolved in concentrated aqueous LiBr shows an optical rotation of $[\alpha]_D^{25} = -240°$, which does not change with time, and a low intrinsic viscosity of $[\eta] = 0{\cdot}06$. It seems that the poly-L-proline molecules in this environment have no macromolecular asymmetric structure, a view supported by the optical rotatory data of L-proline-glycine copolymers (see section 3). Dilution of a solution of poly-L-proline II in LiBr, gives rise to a mutarotation and a value of $[\alpha]_D^{25} = -540°$ is finally reached (Steinberg *et al.*, 1960). This mutarotation is not catalysed by either acid or LiBr, and is much more rapid than the acid catalysed mutarotation described above. Measurements of the rates of mutarotation in the temperature range 2° to 10° led to an enthalpy of activation of $\Delta H^* = 20{\cdot}6$ kcal per mole prolyl residue. Since this value is very similar to the one found for the *cis–trans* isomerization, the high reaction rate must be due to a low entropy of activation. A value of $\Delta S^* = -0{\cdot}84$ e.u. was calculated for the mutarotation in LiBr, as compared with $\Delta S^* = -12{\cdot}5$ e.u. obtained for the acid catalysed mutarotation. This shows that fundamentally different mechanisms must be involved in the two processes. It is plausible to attribute the new type of mutarotation to isomerization of the C(α)—CO bond which, in the poly-L-proline molecule, is spatially stabilized by steric hindrance to free rotation. Additional stabilization may be due to solvation effects.

(h) *Discussion*

The structure of both forms of poly-L-proline in the solid state are now well established. Poly-L-proline I forms a right-handed helix in which all peptide bonds possess a *cis* configuration, whereas poly-L-proline II forms a left-handed helix with all peptide bonds in the *trans* configuration. These findings, in conjunction with the results described here, make it possible to determine the macromolecular conformations of poly-L-proline in solution and to elucidate the mechanism by which configurational changes take place.

Poly-L-proline I and poly-L-proline II, when dissolved in propionic acid, yield solutions with $[\alpha]_D = +50°$ and $[\alpha]_D = -540°$ respectively. Both forms can be recovered unchanged from these solutions by precipitation or by evaporation of solvent. It is therefore concluded that poly-L-proline molecules showing in solution a specific optical rotation of $[\alpha]_D = +50°$ have all their peptide bonds in the *cis* configuration, whereas solutions with $[\alpha]_D = -540°$ contain poly-L-proline molecules with all peptide bonds in the *trans* form. Moreover, the large difference in specific optical rotation between the two forms, and the finding that the specific rotation of an isolated L-prolyl residue ($[\alpha]_D = -240°$, as

derived from the optical rotations of glycine–proline copolymers, see section 3) is of an intermediate value, strongly suggests that the highly asymmetric helical conformations of Forms I and II encountered in the solid state, persist also in solution. This conclusion is corroborated by the hydrodynamic properties of poly-L-proline which indicate that the molecules of both forms are highly extended. In aqueous concentrated aqueous LiBr in which a specific rotation close to that of an isolated L-prolyl residue was recorded, poly-L-proline exhibits an extremely low intrinsic viscosity, indicating the collapse of the asymmetric helical structure. Relatively low viscosities were also measured with poly-DL-proline.

For poly-L-proline samples with average molecular weights of up to 10,000, good agreement was found between the axial ratios of the prolate effective hydrodynamic ellipsoids evaluated according to Simha (1940), and the corresponding axial ratios calculated for rod-like molecules having the dimensions of the right-handed helix of poly-L-proline I and the left-handed helix of poly-L-proline II. For samples with higher average molecular weight, however, it was found that the axial ratios, calculated from the model helices, were always larger than those estimated from the intrinsic viscosities. Since this difference increased with molecular weight it appears that helices of Form I and Form II, which behave in solution as stiff rods at low molecular weight, exhibit increasing flexibility as the length of the contour of the molecules increases. Assuming that the molecules in Form II approach a flexible chain at high molecular weights (above 50,000), it is possible to calculate the length of a statistical element, l, with the aid of the well-known equation of Kirkwood and Riseman (1948). A value of $l = 43$ Å, corresponding to 14 prolyl residues, i.e. approximately four turns of the helix, was obtained.

In view of the finding (Yaron and Berger, 1961) that poly-L-proline chains composed on the average of only six prolyl residues already exhibit the full optical rotation of poly-L-proline II ($[\alpha]_D = -520°$), it can be concluded that while the viscosity is determined primarily by the end to end distance of the polymer, the optical rotation is a reflection of the helical structure of relatively short segments. These considerations explain why a given molecular species can take up different shapes, as measured by viscosity, which, however, possess essentially identical rotatory properties. This situation is clearly seen in Fig. 6.

The experimental data reported above suggest that there is little difference between the enthalpies of poly-L-proline I and poly-L-proline II. It is, therefore, to be expected that solvation may be a decisive factor in determining which of the two forms is the stable one in a given solution. It may be recalled that aliphatic acids, *m*-cresol and water, in which polyproline II is readily soluble, stabilize Form II, while the

addition of alcohols, in which polyproline is insoluble, favours the existence of Form I. Similarly, polymerization of *N*-carboxy-L-proline anhydride in pyridine, in which poly-L-proline cannot be dissolved in either form, yields Form I. Also in poly-*O*-acetylhydroxy-L-proline one observes forward mutarotation in good solvents (aliphatic acids and *m*-cresol) and reverse mutarotation in poor solvents (acetic anyhdride and dimethylformamide). A plausible explanation of this situation is that Form II, because of the more open character of its helix, is more easily subject to interaction with the solvent, thus gaining stability due to the energy of solvation. The stability of Form II of poly-*O*-acetylhydroxy-L-proline in acetic anhydride in the presence of perchloric acid was explained to be due to electrostatic repulsion between the charged peptide groups of the protonated peptide chain.

The transformation of the poly-L-proline I helix into the poly-L-proline II helix, *and vice versa*, in the forward and reverse mutarotations respectively, is obviously the result of a series of *cis-trans* isomerizations of amide bonds within the polypeptide chain. The finding that the forward and reverse mutarotations are acid catalysed thus finds its explanation in the observation that free rotation in amide bonds is enhanced by strong acids. The mechanism of this catalysis, which has been discussed above, involves the protonation of the amide nitrogen with the consequent loss of the partial double-bond character of the –CON– linkage. No satisfactory mechanism is as yet available to explain the conformational changes brought about in poly-L-proline by salts such as LiBr, $CaCl_2$ and KCNS.

3. Copolymers of L-Proline with Other Amino Acids

Amino-acid copolymers containing proline seem of interest since their study may lead to a better understanding of the properties of prolyl residue in peptides and in proteins. Copolymers obtained by the polymerization of *N*-carboxy-L-proline anhydride with *N*-carboxyanhydrides of other amino acids have been chosen as the subject of the work described below because of the particular ease with which these compounds can be prepared.

The preparation of proline–glycine copolymers was undertaken because these two amino acids occur to a high percentage in collagen. Furthermore, since glycine is devoid of an asymmetric carbon atom, optical rotatory characteristics of a prolyl residue in a Pro–Gly copolymer can be measured directly. Because both *N*-carboxy-L-proline anhydride and *N*-carboxyglycine anhydride have comparable, exceptionally high, polymerization rates, the formation of random copolymers could be expected. Finally, it is pertinent to note that polyglycine II was shown

to have, in the solid state, a helical structure similar to that of poly-L-proline II (Crick and Rich, 1955).

Proline–sarcosine copolymers were chosen for study because of their good solubility in water and various organic solvents. Furthermore, sarcosyl residues are devoid of optical activity and lack the ability to form hydrogen-bonded structures. They do not interfere, therefore, with the measurement of the optical rotatory properties of prolyl residues, and are not expected to disturb the formation of the characteristic poly-L-proline conformations.

Preliminary studies were carried out on the possibility of copolymerizing *N*-carboxy-L-proline anhydride with other *N*-carboxyanhydrides of optically active amino acids.

(a) *With glycine*

Copolymers of glycine and L-proline, having molar residue ratios of 1:1; 2:1 and 3:1, were prepared by the copolymerization in pyridine of *N*-carboxyglycine anhydride and *N*-carboxy-L-proline anhydride in the appropriate molar ratios. The copolymers obtained after precipitation with ether were soluble in formic acid and trifluoroacetic acid, and insoluble in water. They separated out quantitatively from their solution in formic acid by adding water. As polyglycine is insoluble in formic acid and poly-L-proline II is soluble in water, it can be concluded from the above solubility characteristics that the three preparations obtained do not contain either polyglycine or polyproline, but represent true glycine–proline copolymers.

The specific optical rotations per proline residue of the three copolymers in formic acid are given in Fig. 7. The data presented show that the dilution of the prolyl residues with glycine along the peptide chain leads to a decrease in laevorotation. This is due to the gradual disruption of the poly-L-proline conformation by the introduction of the optically inactive glycine residues. Extrapolating the curve given in Fig. 7 to high glycine to proline ratios, a value of $[\alpha]_D^{25} = -250°$ to $-300°$ is reached for the specific rotation of a proline residue. The latter value seems, therefore, to represent the intrinsic residue rotation of L-proline. This figure is in accord with the corresponding value ($[\alpha]_D^{25} = -240°$) derived from the specific rotation of poly-L-proline in concentrated aqueous LiBr.

Experiments with branched polyamino acids containing poly-L-proline side chains of varying length (Yaron and Berger, 1961) have demonstrated that for the formation of the characteristic polyproline II helix a minimum of five to six prolyl residues in sequence is required. The finding that the specific optical rotation per prolyl residues changes as in Fig. 7 shows that the fraction of prolyl residues in a sequence of six or more decreases with increasing glycine content of the copolymers,

approaching zero at glycine to proline ratios larger than seven. An elementary statistical consideration indicates, however, that a true random distribution of prolyl and glycyl residues along the copolymer chains should cause the rotation per proline residue to reach its intrinsic value at a much lower glycine content than that found experimentally. It might be thus concluded that during the copolymerization reaction an N-terminal prolyl residue has a greater tendency to add an *N*-carboxy-L-proline anhydride monomer than an *N*-carboxyglycine anhydride.

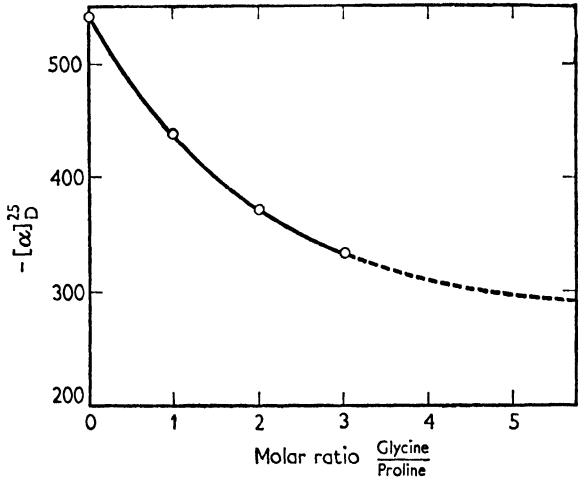


FIG. 7. Specific optical rotation, $[\alpha]_D^{25}$, per proline residue of glycine: L-proline copolymers in formic acid. From Harrington and Sela (1958).

(b) *With sarcosine*†

L-Proline-sarcosine copolymers, in approximate molar residue ratios of 1 : 1 and 1 : 8, were prepared by copolymerization of the corresponding *N*-carboxyanhydrides in anhydrous nitrobenzene using sodium methoxide as initiator. A solution of *N*-carboxy-L-proline anhydride was added to the polymerization mixture at the rate necessary to maintain the required molar ratio between the two anhydrides. This procedure was adopted to compensate for the higher polymerization rate of *N*-carboxyproline anhydride. The copolymers obtained had average molecular weights in the range of 5000 to 9000 as estimated from sedimentation and diffusion measurements.

The formation of copolymers as a result of the simultaneous polymerization of *N*-carboxy-L-proline anhydride and *N*-carboxy-L-sarcosine anhydride was ascertained by (a) the solubility of the products formed

† This section is a preliminary account of a study carried out by A. Morawiecki and E. Katchalski.

in propanol, a solvent in which poly-L-proline is insoluble, (b) solubility in hot water, in which poly-L-proline is insoluble, and (c) by the observation that fractionation procedures, such as column chromatography and paper electrophoresis failed to yield fractions with different amino acid compositions.

Both copolymers, Pro:Sarc (1:1) and Pro:Sarc (1:8), gave specific optical rotations per proline residue of $[\alpha]_D^{25} = -500$ to $-520°$ in formic acid (after 30 min at room temperature), and of $[\alpha]_D^{25} = -230$ to $-250°$ in 10 M aqueous LiBr. These values are similar to those recorded for poly-L-proline in these solvents.

The copolymers synthesized undergo forward mutarotation, similarly to poly-L-proline I, in formic acid, acetic acid, *m*-cresol and water. The course of mutarotation of the copolymer Pro:Sarc (1:8) in glacial acetic acid is given in Fig. 8. The forward mutarotation in acetic acid as well

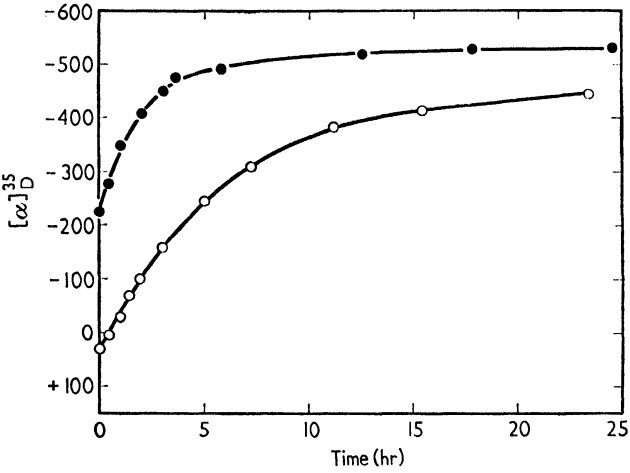


FIG. 8. Mutarotation of poly-L-proline (—○—○—) and of a sarcosine: L-proline (8:1) copolymer (—●—●—) in acetic acid at 35°. Specific rotations given $[\alpha]_D^{25}$, are calculated per proline residue. From Morawiecki and Katchalski (1962).

as in water was found to be considerably faster than that of poly-L-proline.

As in the case of poly-L-proline, the mutarotation of the proline–sarcosine copolymers is acid catalysed. It is faster in formic acid than in acetic acid and is greatly retarded in formic acid in the presence of sodium formate.

Specific rate constants and reaction orders (defined as for the case of poly-L-proline) of the forward mutarotation of copolymer Pro:Sarc (1:8) in water and glacial acetic acid were determined in the temperature

range of 30 to 45°. The forward mutarotation was found to be of order 1·0 in water, and of order 1·6 in glacial acetic acid. From the temperature dependence of the rate constants, enthalpies of activation per mole prolyl residue of $\Delta H^* = 13{\cdot}3$ kcal in water and of $\Delta H^* = 27{\cdot}8$ kcal in acetic acid were calculated.

The reverse mutarotation of the Pro:Sarc (1:8) copolymer (Form II) in propanol–formic acid (9:1 v/v) is given in Fig. 9. The final specific rotation reached as a result of the reverse mutarotation was also attained when the copolymer, as isolated from the polymerization mixture, was dissolved in propanol–formic acid (9:1 v/v) and left to undergo forward

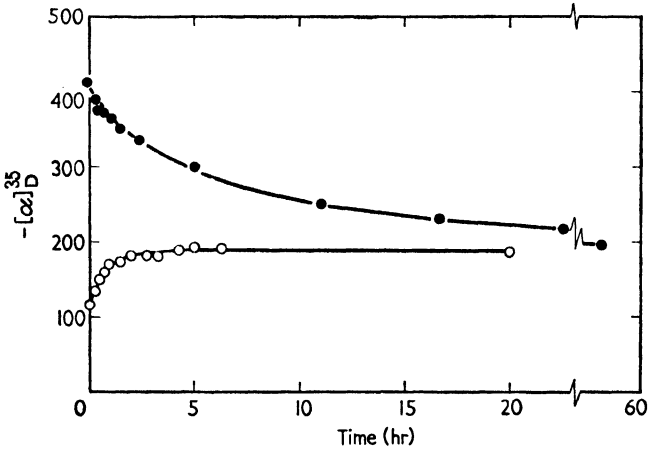


FIG. 9. Forward and reverse mutarotation of a sarcosine: L-proline (8:1) copolymer in propanol: formic acid (9:1 v/v) at 35°. Specific rotations given, $[\alpha]_D^{30}$, are calculated per proline residue. From Morawiecki and Katchalski (1962).

mutarotation to a constant specific optical rotation. This means that the specific optical rotation reached in both cases represents a true equilibrium value. The specific optical rotation at equilibrium, as a function of solvent composition in the propanol–formic acid system at 35°, is given in Fig. 10.

Copolymer Pro:Sarc (1:8) has a specific optical rotation per proline residue of $[\alpha]_D^{35} = -84°$ in propanol and $[\alpha]_D^{35} = -470°$ in formic acid. Assuming that the prolyl peptide bonds of the copolymer in propanol are of the *cis* configuration and in formic acid of the *trans* configuration, the equilibrium constant K = (*trans* prolyl peptide bonds)/(*cis* prolyl peptide bonds) for any propanol–formic acid mixture, can be calculated according to the procedure described for the case of poly-*O*-acetylhydroxy-L-proline (Steinberg *et al.*, 1960). Values of $K = 0{\cdot}205$ at 35° and $K = 0{\cdot}225$ at 45° were obtained for propanol–formic acid (9:1 v/v). From these

values an enthalpy of $\Delta H = -2{\cdot}6$ kcal per mole was calculated for the *cis-trans* isomerization of the prolyl peptide bonds.

Our findings that the specific optical rotation per proline residue in formic acid of both proline–sarcosine copolymers synthesized (1:1 and 1:8), ($[\alpha]_D^{25} = -500°$ to $-520°$), is very close to that of poly-L-proline II ($[\alpha]_D^{25} = -540°$), strongly suggest that these copolymers are block copolymers, in which most of the prolyl residues are arranged in poly-prolyl sequences of more than five to six residues. This view is corroborated by the finding that the X-ray powder diagrams of the copolymers, after treatment with formic acid, resemble that of poly-L-proline II. Furthermore, proline–sarcosine copolymers show, under the suitable

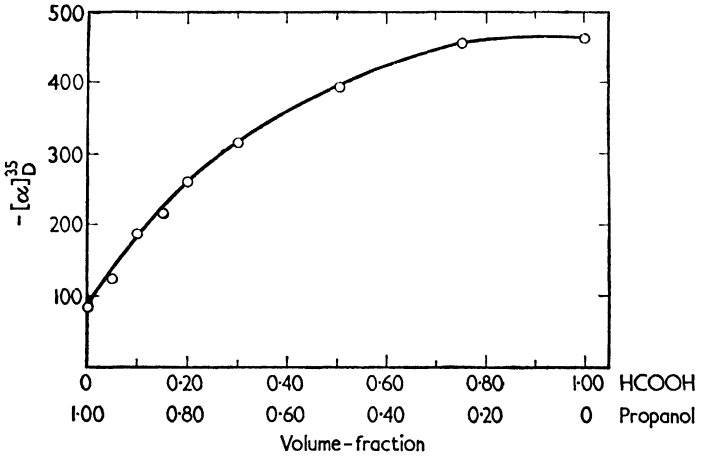


FIG. 10. Specific optical rotations per proline residue, of a sarcosine:L-proline (8:1) copolymer at equilibrium, in propanol–formic acid mixtures of different compositions, at 35°. From Morawiecki and Katchalski (1962).

conditions, a forward and a reverse mutarotation which closely resembles that of poly-L-proline. It should be remarked, however, that the copolymer mutarotates in glacial acetic acid and in water considerably faster than poly-L-proline. Also, the specific optical rotation per proline residue of the copolymer, Form I, in acetic acid ($[\alpha]_D^{25} = -150°$), differs markedly from that of poly-L-proline I, ($[\alpha]_D^{25} = +50°$), in the same solvent. Finally, it might be noted that unlike poly-L-proline, equilibria between Forms I and II have been established for the Pro:Sarc (1:8) copolymer in propanol formic-acid mixtures.

(c) *With other amino acids*

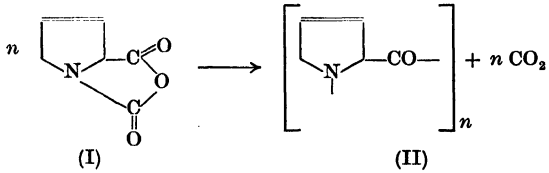
Preliminary experiments were carried out to copolymerize *N*-carboxy-L-proline anhydride with *N*-carboxy-*O*-acetylhydroxy-L-proline an-

hydride or with *N*-carboxy-γ-benzyl-L-glutamate. A copolymer containing proline and *O*-acetylhydroxy-L-proline in a molar residue ratio of 1:1 gave in formic acid $[\alpha]_D^{25} = -320°$. On deacetylation a Pro:Hypro (1:1) copolymer was obtained with a specific optical rotation of $[\alpha]_D^{25} = -400°$ in water. In view of the structural similarity between the L-prolyl residue and the hydro-L-prolyl residue, and the conformational similarity between poly-L-proline II and poly-*O*-acetylhydroxy-L-proline II, it is plausible to assume that the Pro:*O*-acetyl-Hypro copolymer synthesized is a random copolymer and might attain the conformations of poly-L-proline I or II under suitable conditions. It will be of particular interest to study the conformations of Pro-Hypro copolymers, since polyhydroxy-L-proline has been observed only in Form II.

When *N*-carboxy-L-proline anhydride and *N*-carboxy-γ-benzyl-L-glutamate anhydride at molar ratios of 1:1 or 1:5, were simultaneously polymerized in dioxane using sodium methoxide as initiator, materials were obtained which could be separated into two fractions by extraction with dimethylformamide. The insoluble fraction proved to be poly-L-proline, whereas the soluble one was poly-γ-benzyl-L-glutamate containing traces of proline. Polymerization of the above *N*-carboxy-anhydrides in a molar ratio Pro:Glu of 1:10, on the other hand, gave a product completely soluble in dimethylformamide and insoluble in water, which seems to represent a true copolymer. These findings illustrate the marked tendency of the terminal prolyl residue of a growing peptide chain to react with *N*-carboxy-L-prolyl anhydride in preference to *N*-carboxy-γ-benzyl-L-glutamate anhydride.

4. Poly-3,4-dehydro-L-proline‡

Poly-3,4-dehydro-L-proline (II) was prepared, similarly to poly-L-proline, by the polymerization of *N*-carboxy-3,4-dehydro-L-proline anhydride (I) in pyridine.



(I) (II)

The *N*-carboxyanhydride (I) was obtained by treatment of 3,4-dehydro-L-proline (Robertson and Witkop, 1962) with phosgene followed by cyclization of the *N*-carbonyl chloride formed with silver oxide.

‡ This section is a preliminary account of a study carried out by J. Kurtz, A. V. Robertson, B. Witkop and A. Berger.

The polymer, isolated from the polymerization mixture by ether precipitation, gives immediately after dissolution in 99% formic acid a specific optical rotation of $[\alpha]_D^{25} = -500°$ ($c = 1{\cdot}0$). The laevorotation gradually increases to a value of $[\alpha]_D^{25} \simeq -1200°$ within 2 hr. Mutarotation in trifluoroacetic acid seems to be practically instantaneous, the specific rotation reaching a value of $[\alpha]_D^{25} = -1250°$ within a few minutes.

Because of the structural similarity between the L-prolyl and dehydro-L-prolyl residues, and the resemblance in the general course of forward mutarotation of the two corresponding polymers, it seems plausible that the conformation of polydehydro-L-proline with $[\alpha]_D^{25} = -500°$ corresponds to that of poly-L-proline I ($[\alpha]_D^{25} = +50°$), whereas the conformation of polydehydro-L-proline with $[\alpha]_D^{25} \simeq -1200°$ corresponds to that of poly-L-proline II ($[\alpha]_D^{25} = -540°$).

It is of interest to note that the rate of mutarotation of polydehydro-L-proline in 99% formic acid, at 25°, is considerably slower (half-life time about 30 min) than that of poly-L-proline (half-life time less than 1 min). Furthermore, the two polymers differ markedly in their solubility properties. Thus poly-L-proline I is soluble in acetic acid and propionic acid, whereas polydehydro-L-proline is insoluble in these solvents; poly-L-proline II dissolves readily in water, while poly-L-dehydroproline II is insoluble in water. The reverse mutarotation of polydehydro-L-proline II could not be carried out under the conditions employed in the case of poly-L-proline because of the insolubility of the unsaturated polymer in formic acid–alcohol mixtures.

In order to determine the intrinsic specific optical rotation of the dehydro-L-prolyl residue, a number of dehydro-L-prolyl-glycine copolymers (molar residue ratios Gly : dehydroPro from 1 to 30) were prepared and their optical rotations in trifluoroacetic acid and in 12 M aqueous LiBr measured. The data presented in Fig. 11 shows that the absolute value of the specific rotation per dehydroprolyl residue in trifluoroacetic acid decreases with increasing glycine content of the copolymer. This trend seems to indicate that the fraction of dehydroprolyl residues located in sequences capable of forming the characteristic dehydro-L-proline II helical conformation decreases as the dehydroproline is diluted by glycine along the copolymer chain. The specific rotation per dehydroprolyl residue in copolymers with high glycine to dehydroproline ratios, $[\alpha]_D^{25} = -550°$, obviously gives the rotation of the dehydroprolyl residue devoid of any contribution from an asymmetric macromolecular structure. This interpretation is corroborated by the observation that in aqueous LiBr, a solvent known to destroy the helical conformation of poly-L-proline, the specific rotation per dehydro-L-prolyl residue, $[\alpha]_D^{25} = -550°$, is independent of the glycine content of the copolymer and equal to that found in trifluoroacetic acid for a copolymer containing

glycine and dehydro-L-proline in a molar residue ratio of 30:1. In view of the high specific residue rotation obtained it is of interest to note that the specific laevorotation of the parent amino acid, dehydro-L-proline, $[\alpha]_D^{20} = -335°$ (water), is considerably larger than that of L-proline, $[\alpha]_D^{20} = -86°$ (water).

A comparison of Fig. 11 with Fig. 7 shows that the relative amounts of glycine required to approach the residue rotation of L-proline in glycine–proline copolymers is considerably smaller than that necessary to reach the dehydro-L-prolyl residue rotation in glycine–dehydroproline copolymers. This seems to indicate that in the copolymers investigated the dehydroprolyl residues have a greater tendency to form block sequences than the prolyl residues.

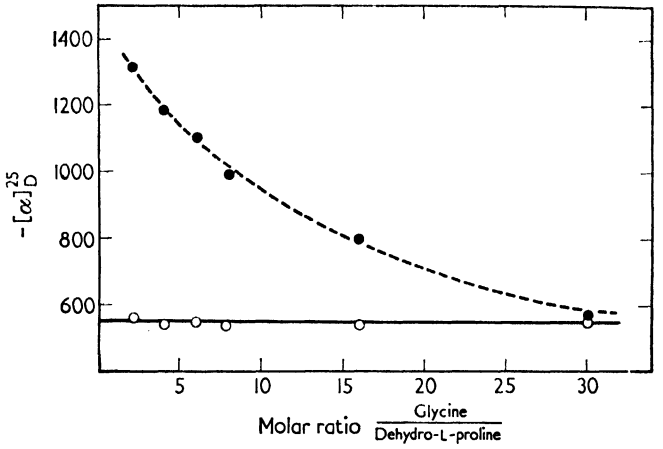


FIG. 11. Specific optical rotation, $[\alpha]_D^{25}$, per dehydroproline residue of glycine:dehydro-L-proline copolymers in trifluoroacetic acid, (—●—●—) and in 12 M LiBr (—○—○—). From Kurtz *et al.* (1962).

5. POLYHYDROXY-L-PROLINE, POLY-*O*-ACETYL- AND POLY-*O*-TOLYLSULPHONYL-HYDROXY-L-PROLINE

Poly-*O*-acetylhydroxy-L-proline was derived from *O*-acetyl-*N*-carboxyhydroxy-L-proline anhydride on polymerization in pyridine (Kurtz *et al.*, 1958*b*). Deacetylation with aqueous ammonia gave rise to polyhydroxy-L-proline. Osmotic and sedimentation measurements in aqueous solution gave average molecular weights of 10,600 and 10,700 respectively, for the sample of polyhydroxy-L-proline synthesized. Poly-*O*-*p*-tolylsulphonylhydroxy-L-proline was obtained by polymerization of *N*-carboxy-*O*-*p*-tolylsulphonylhydroxy-L-proline anhydride in pyridine.

As with poly-L-proline mutarotation was observed with poly-*O*-

acetylhydroxy-L-proline and with poly-*O*-*p*-tolylsulphonylhydroxy-L-proline. Poly-*O*-acetylhydroxy-L-proline, precipitated from the pyridine polymerization mixture by ether (Form I), showed a specific optical rotation of $[\alpha]_D^{25} = +25°$, immediately after dissolution in 90% formic acid. The optical rotation of the solution decreased with time to a final value of $[\alpha]_D^{25} = -175°$ (Form II) after 6 hr at room temperature. Precipitation of the polymer from the highly laevorotatory solution with ether yielded a preparation with $[\alpha]_D^{25} = -175°$, immediately after dissolution in 90% formic acid.

Boiling of poly-*O*-acetylhydroxy-L-proline II ($[\alpha]_D^{25} = -175°$) in dimethylformamide for several seconds, cooling and precipitation by

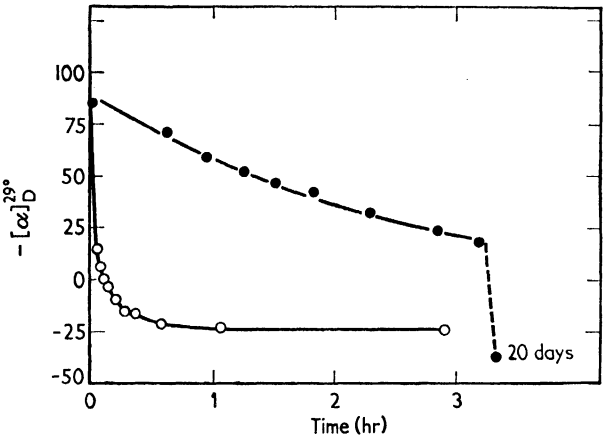


FIG. 12. The effect of perchloric acid on the reverse mutarotation of poly-O-acetylhydroxy-L-proline ($c = 0{\cdot}5$ g per 100 ml) in acetic anhydride at 29°: ●——●, in acetic anhydride; ○——○, $7{\cdot}1 \times 10^{-4}$ M $HClO_4$ in acetic anhydride (0·022 mole $HClO_4$ per mole peptide bond). From Steinberg *et al.* (1960).

ether yielded Form I ($[\alpha]_D^{25} = +25°$). The material obtained when dissolved in formic acid undergoes normal forward mutarotation. A material with $[\alpha]_D^{25} = +10°$ (Form I) was obtained when a concentrated solution of Form II ($[\alpha]_D^{25} = -175°$) in formic acid was diluted with pyridine (40 volumes), left for 24 hr at room temperature, and precipitated with ether.

The effect of $HClO_4$ on the reverse mutarotation of poly-*O*-acetylhydroxy-L-proline in acetic anhydride is given in Fig. 12. This polymer undergoes reverse mutarotation in this solvent (initial $[\alpha]_D^{29} = 84°$, final $[\alpha]_D^{29} = +38°$) with a half-life time of 175 min. On the addition of a small amount of $HClO_4$ (0·022 mole $HClO_4$ per mole peptide bond) the reverse mutarotation is accelerated to a half-life time of less than 2·5 min.

Proton binding by poly-*O*-acetylhydroxy-L-proline in acetic anhydride

could be demonstrated potentiometrically by anhydrous titration with $HClO_4$ in dichloroethane–acetic acid. The end point of the titration curve corresponded to protonation of about 37% of the number of peptide nitrogen atoms present in the polymer. It is of interest that this value is close to the amount of $HClO_4$ bound by poly-L-proline when precipitated by means of $HClO_4$ from acetic-acid solutions.

As mentioned above, poly-*O*-acetylhydroxy-L-proline is stable in its dextrorotatory form (Form I, $[\alpha]_D = +40°$) in acetic anhydride. However, Form II ($[\alpha]_D = -130°$) becomes the stable one in the presence of perchloric acid at concentrations exceeding 0·3 mole $HClO_4$ per mole peptide bond. With lesser amounts of acid, intermediate $[\alpha]_D$ values are obtained as illustrated in Fig. 13.

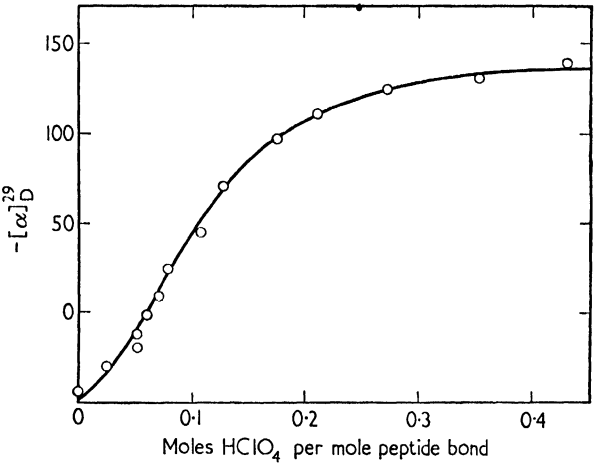


FIG. 13. Equilibrium values of $[\alpha]_D$ of poly-*O*-acetylhydroxy-L-proline in acetic anhydride ($c = 0·5$ g per 100 ml) as a function of moles $HClO_4$ per mole peptide bond; temperature 29°. From Steinberg *et al.* (1960).

The stabilization of Form II at relatively high $HClO_4$ concentrations is most likely connected with the protonation of the peptide chain. Electrostatic repulsion forces in the charged chain favour the extension of the molecule. It will be remembered that Form II, as represented by the all-*trans* Cowan-McGavin helix, is the most extended form which the polypeptide under discussion can attain. It should be further noticed that maximum extension coincides with maximum protonation.

An attempt was made to evaluate the enthalpy of the *cis–trans* isomerization reaction. The optical rotation of a solution containing poly-*O*-acetylhydroxy-L-proline and 0·106 mole $HClO_4$ per mole peptide bond in acetic anhydride was measured at several temperatures in the range from 19 to 46°. The value of $[\alpha]_D$ ($-42°$) did not change within the limits

of experimental error. It is thus concluded that the ΔH of isomerization in the system investigated does not exceed 0·5 kcal per mole peptide bond. In this connection it is pertinent to note that practically equal values were obtained for the enthalpies of activation of the forward mutarotation of poly-L-proline in acetic acid and its reverse mutarotation in propanol–formic acid (9:1 v/v).

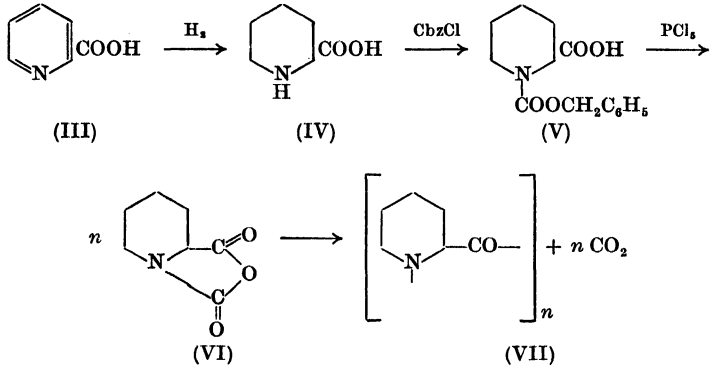
Poly-*O*-tolylsulphonylhydroxy-L-proline precipitated from pyridine by ether, $[\alpha]_D^{25} \simeq 0°$ (Form I), mutarotated in glacial acetic acid to give a final specific rotation of $[\alpha]_D^{25} = -120°$ within several hours. The polymer precipitated by ether from its solution in acetic acid ($[\alpha]_D^{25} = -120°$, Form II) could be transformed into Form I ($[\alpha]_D^{25} \simeq 0°$) by reverse mutarotation in pyridine.

The specific rotation of polyhydroxy-L-proline, $[\alpha]_D^{25} = -400°$ (in water), obtained from poly-*O*-acetylhydroxy-L-proline by treatment with aqueous ammonia, did not change within 2 weeks at room temperature. The treatment with aqueous ammonia seems, therefore, to produce the form of polyhydroxy-L-proline with the highest laevorotation in water. No conditions could be found under which polyhydroxy-L-proline undergoes mutarotation.

6. Poly-L- and Poly-D-pipecolic Acid

(a) *Synthesis*

Poly-L-pipecolic acid (VII) was synthesized according to the following scheme:



Picolic acid (III) was hydrogenated in the presence of platinum oxide as catalyst (Harris and Pollock, 1953) and the DL-pipecolic acid (IV) obtained was resolved to its optically active enantiomorphs by means of L- and D-tartaric acids (Mende, 1896). The specific optical rotations measured for L-pipecolic acid and D-pipecolic acid in water were

$[\alpha]_D^{25} = -27°$ and $[\alpha]_D^{25} = +27°$ respectively. Similar values were recorded by Harris and Pollock (1953). The *N*-benzyloxycarbonyl derivatives of both enantiomorphs (V), obtained in the usual way, yielded on treatment with phosphorous pentachloride *N*-carboxy-L- (or D)-pipecolic acid anhydride (VI). The anhydride (VI) could also be obtained, though in poor yield, from pipecolic acid and phosgene by a procedure analogous to that used in the preparation of *N*-carboxy-L-proline anhydride (Kurtz *et al.*, 1958*b*). The required optically active polypipecolic acids (VII) were finally obtained by the polymerization of *N*-carboxy-L-(or D)-pipecolic acid anhydride in benzene or dioxane using diethylamine as initiator.

Poly-L-pipecolic acid, precipitated by ether from the polymerization mixture (Form I), is soluble in glacial acetic acid and chloroform and insoluble in water, dioxane and dimethylformamide. It has a specific optical rotation of $[\alpha]_D^{25} = -325°$ in glacial acetic acid and $[\alpha]_D^{25} = -260°$ in chloroform.

The sample investigated gave an intrinsic viscosity $[\eta] = 0{\cdot}35$ dl/g in glacial acetic acid. From its sedimentation and diffusion coefficients, $s_{20} = 0{\cdot}36S$ and $D_{20} = 8{\cdot}1 \times 10^{-7}$ cm^2 sec^{-1}, and its partial specific volume $\bar{v} = 0{\cdot}75$, all measured in glacial acetic acid, an average molecular weight of 4350, corresponding to an average degree of polymerization of $\overline{DP} = 40$, was calculated.

When poly-L-pipecolic acid ($[\alpha]_D^{25} = -325°$ in glacial acetic acid) was dissolved in 98% formic acid, its specific optical rotation changed gradually to reach a final value of $[\alpha]_D^{24} = -50°$ within 2 hr. Isolation of the polypeptide from its solution in formic acid yielded a product (Form II) with a specific rotation of $[\alpha]_D^{24} = -50°$ in glacial acetic acid. Form II is soluble in acetic acid, propionic acid, chloroform and water.

Both forms of poly-L-pipecolic acid (I and II) gave a nearly quantitative yield of L-pipecolic acid on acid hydrolysis. They seem, therefore, to differ in their macromolecular conformation but not in their primary structure.

Poly-D-pipecolic acid (Form I), prepared analogously to the L-polymer, showed a specific optical rotation of $[\alpha]_D^{24} = +325°$ in glacial acetic acid. It mutarotated in formic acid to give Form II with $[\alpha]_D^{24} = +50°$.

The optically inactive poly-DL-pipecolic acid, obtained by the polymerization of *N*-carboxy-DL-pipecolic acid anhydride in benzene, is water-soluble, and separates out from its aqueous solution on heating to 37°. The precipitate thus formed redissolves on cooling.

(b) *Properties of Forms I and II*

The optical rotatory dispersion curves of Forms I and II of poly-L-pipecolic acid in glacial acetic acid and in chloroform, in the range 350 mμ

to 600 mμ, are given in Fig. 14. For comparison the optical rotatory dispersion of L-pipecolic acid in glacial acetic acid is included. The optical rotatory dispersion values of Forms I and II of poly-D-pipecolic acid in glacial acetic acid and in chloroform were equal with, but opposite

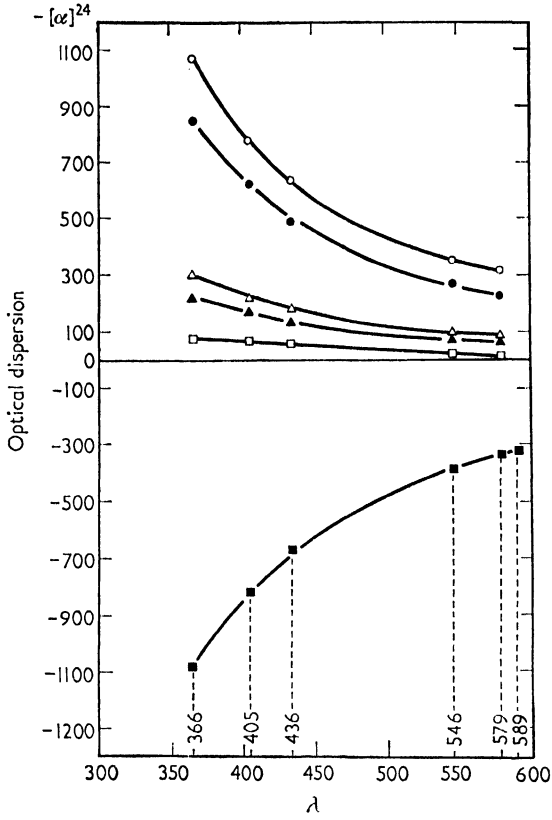


FIG. 14. Optical rotatory dispersion of the two forms of poly-L-pipecolic acid. ○——○, Form I in glacial acetic acid; ●——●, Form I in chloroform; △——△, Form II in chloroform; ▲——▲, Form II in glacial acetic acid; □——□ L-pipecolic acid in glacial acetic acid; ■——■ Poly-D-pipecolic acid, Form I in glacial acetic acid.

in sign to, the corresponding values measured for the L-polymer. All rotatory dispersion curves of the polymer could be described by a one-term Drude equation with dispersion constants, λ_c, between 210 and 220 mμ.

The infra-red absorption spectra of poly-D-pipecolic acid I and II in chloroform are given in Fig. 15. Both absorption curves show the characteristic C=O absorption at 1650 cm^{-1} and, as expected, lack the NH absorption at 1540–1550 cm^{-1}. Form II shows a weak but distinct

band at 1720 cm^{-1} which is absent in Form I. Further differences are also found at longer wavelengths. The characteristic absorption peaks at 980–990 cm^{-1} found in Form I but absent in Form II, indicate that the latter is practically free of the former.

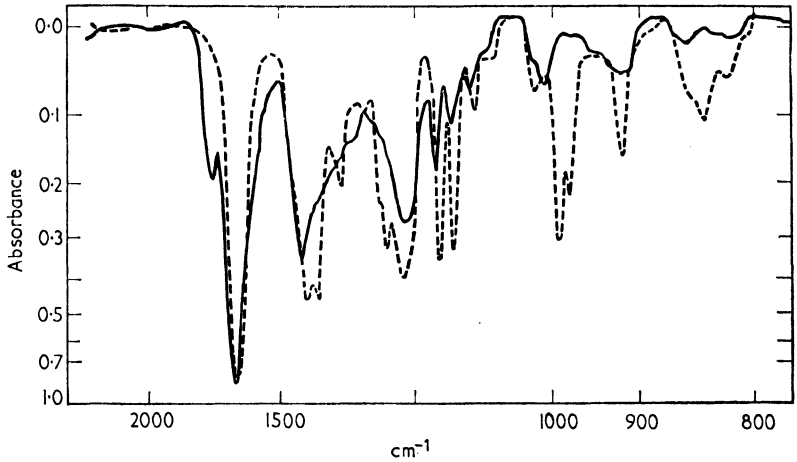


FIG. 15. Infra-red absorption spectra of poly-D-pipecolic acid Form I (---) and Form II (——) in chloroform (10% w/v).

X-Ray powder diagrams of both forms of poly-L-pipecolic acid are given in Fig. 16. X-Ray diagrams of partially oriented fibres of poly-L-pipecolic acid I indicate that this form is characterized by a left-handed helix with peptide bonds in the *trans* configuration (Traub, 1962).

(c) *Mutarotation*

The course of forward mutarotation of poly-L-pipecolic acid I in several formic acid–acetic acid mixtures, at 24°, is given in Fig. 17. Unlike poly-L-proline, poly-L-pipecolic acid does not mutarotate in glacial acetic acid. In formic acid the mutarotation of poly-pipecolic acid I is quite fast, though slower than that of poly-L-proline. The rate of mutarotation in acetic acid–formic acid increases with increasing content of formic acid in the mixture. Thus, whereas the half-life time of mutarotation in a solvent mixture containing 20% by volume of formic acid, was about 24 hr, it was only 1 hr in a mixture containing 50% by volume of formic acid. The results obtained with poly-D-pipecolic acid in acetic acid–formic acid mixtures were similar to those obtained with poly-L-pipecolic acid, the optical rotation changing from $[\alpha]_D^{24} = +325°$ to $[\alpha]_D^{24} = +50°$.

Since poly-L-pipecolic acid, Form I, retains its specific optical rotation in acetic acid and in chloroform, it was possible to test for initiation of mutarotation by mineral acids in these solvents. Indeed, mutarotation of poly-L-pipecolic acid I was observed in acetic acid 2×10^{-3} M in $HClO_4$, and in chloroform $3{\cdot}8 \times 10^{-2}$ M in HCl. In the former case $[\alpha]_D^{24}$ reached $-220°$ in 2 hr, whereas in the latter case $[\alpha]_D^{24}$ reached $-100°$ in 2 hr. Additional proof for the catalytic effect of hydrogen ions on the mutarotation was obtained through the finding that increasing concentrations of sodium formate in formic acid, markedly decreased the rate of mutarotation.

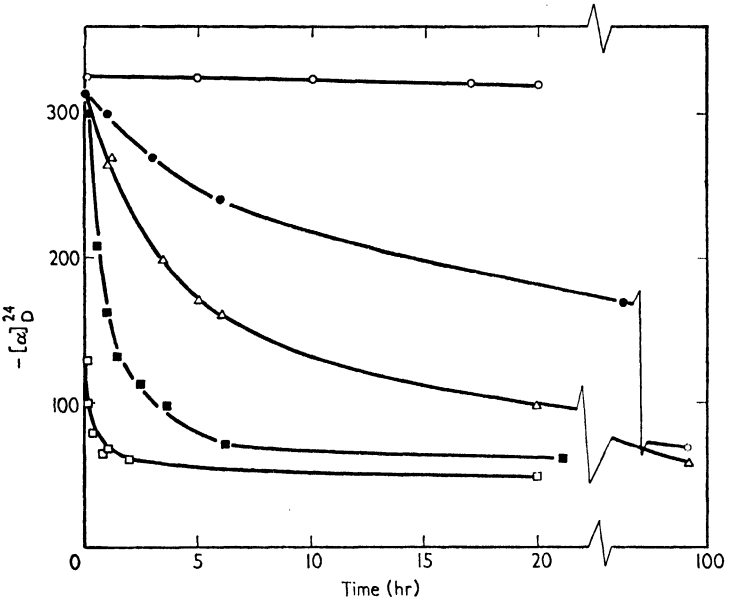


FIG. 17. Mutarotation of poly-L-pipecolic acid in acetic acid (—○—○—); 98% formic acid (—□—□—); and in acetic acid–formic acid mixtures of the following compositions: —●—●— 8:2 v/v; —△—△— 7:3 v/v; —■—■— 5:5 v/v.

No suitable solvent system has as yet been found in which reverse mutarotation of Form II to Form I occurs. Poly-L-pipecolic acid is insoluble in formic acid–propanol or acetic acid–propanol mixtures, in which the reverse mutarotation of poly-L-proline has been demonstrated.

(d) *Degradation*

Preliminary results indicate that poly-L (or D)-pipecolic acid is readily degraded in formic acid (98%) or in formic acid–acetic acid mixtures. Mutarotation is thus accompanied by the appearance of new carboxyl

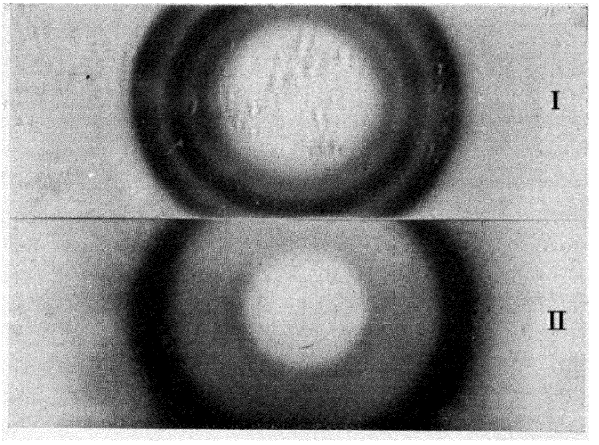


FIG. 16. X-Ray powder diagrams of poly-D-pipecolic acid form I (above) and form II (below).

and imino end groups. Hydrolysis is most likely enhanced by the interaction of imide bonds with the protons of the medium; a neutralization reaction that has been shown to catalyse mutarotation.

(e) *Discussion*

The data presented show that poly-L-pipecolic acid can be obtained in two forms, identical in primary structure, but differing in macromolecular conformation. Forms I and II are distinguished by their specific optical rotations, $[\alpha]_D^{25} = -325°$ and $[\alpha]_D^{25} = -50°$ (in glacial acetic acid) respectively, by their characteristic optical rotatory dispersions, infra-red absorption spectra, and X-ray diffraction patterns.

The observation that the mutarotation of Form I to Form II is acid catalysed indicates that the macromolecular conformational changes taking place during the mutarotation reaction of poly-L-pipecolic acid are caused by a series of *trans-cis* isomerizations of peptide bonds, similarly to the events occurring during the mutarotation of poly-L-proline. The rate of mutarotation of poly-L-pipecolic acid is as a rule slower than that of poly-L-proline, under similar conditions. This may be due to increased steric hindrance by the bulkier six-membered ring of pipecolic acid as compared with the five-membered ring of proline.

The experimental data available so far does not permit us to attribute unequivocally a definite conformation to each of the two forms of poly-L-pipecolic acid. In accord with the findings with poly-L-proline, however, it is plausible to assume that poly-L-pipecolic acid I and II possess helical conformations with opposite senses of twist. Since the laevorotation of Form I (at the D-line) is 275° larger than that of Form II, it might be suggested that Form I represents the left-handed helix and Form II the right-handed one. In analogy with poly-L-proline, the peptide bonds of the left-handed helix would have to be in the *trans*-configuration, and those of the right-handed helix would have to be in the *cis*-configuration. The suggestion made as to the conformation of poly-L-pipecolic acid I is in accord with the preliminary X-ray data mentioned previously. Poly-L-pipecolic acid II gave, in aqueous 10 M LiBr, a solvent system shown to cause breakdown of the helical asymmetry of poly-L-proline II, a specific optical rotation of $[\alpha]_D^{25} = -100°$. The finding that the polypeptide investigated has in its random form a specific rotation intermediate between those of the two extreme forms, further supports the above suggestion as to the sense of twist of the proposed helices.

Accepting the above proposal as to the conformations of the two forms of poly-L-pipecolic acid, it is of interest to note that the polymerization of *N*-carboxy-L-pipecolic acid anhydride leads to the formation of the left-handed *trans*-helix of poly-L-pipecolic acid I, whereas the poly-

merization of *N*-carboxy-L-proline anhydride leads to the formation of the right-handed *cis*-helix of poly-L-proline I.

Furthermore, under strongly acid conditions the stable configurations are the right-handed *cis*-helix in the case of poly-L-pipecolic acid and the left-handed *trans*-helix in the case of poly-L-proline. A similar relationship has been observed between poly-γ-benzyl-L-glutamate and poly-β-benzyl-L-aspartate. These two homologous polypeptides may form in the same solvent system helices with opposite sense of twist (Karlson *et al.*, 1960).

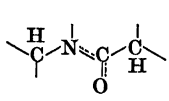
The availability of *N*-carboxy-D-pipecolic acid anhydride enabled us to prepare poly-D-pipecolic acid and to compare its properties with those of the L-polymer. The optical rotation of poly-D-pipecolic acid under various conditions was found to be equal in magnitude with, but of opposite sign to that of poly-L-pipecolic acid, when measured under the same conditions. In other properties, such as solubility and infra-red absorption, both polymers were identical. Thus, poly-L-pipecolic acid and poly-D-pipecolic acid, each built up from enantiomorphic units, are shown to be enantiomorphs also with respect to macromolecular conformation.

7. Concluding Remarks

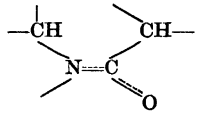
The study of the hydrodynamic and optical rotatory properties of poly-L-proline in solution revealed that the two forms of this polypeptide, shown to exist in the solid state, prevail also in solution. Form I, characterized by a right-handed helical conformation, is stable in poor solvents, such as ethanol, methanol and pyridine, whereas Form II, characterized by a left-handed helical conformation is stable in good solvents, such as formic acid, acetic acid and water. The presence of the two conformational forms encountered in the case of poly-L-proline has now been shown to exist also in the other analogous polyamino acids, polydehydro-L-proline, poly-*O*-acetyl hydroxy-L-proline and poly-L-pipecolic acid, described in this paper.

The question arises what are the factors which stabilize the polyproline helices and prevent their disruption in solution. In the absence of hydrogen bonding, which is the main stabilizing factor of the well-known α-helix, it seems that other causes restricting free rotation must be operative. Such rotational restrictions must obviously apply to all the bonds along the polypeptide backbone. The three bonds to be considered are the N—C(α) bond, the N—CO bond, and the C—CO bond. In polyproline, rotation about the N—C(α) bond is obviously impossible as a result of its position in the pyrrolidine ring. Rotation about the peptide bond N—CO is restricted by its partial double-bond character. It has been estimated that rotations about similar bonds require an energy of

activation of more than 20 kcal per mole (Pauling, 1948). As a result of this restriction there exist two possible orientations of the peptide bond: the *trans* configuration in which the two α-carbon atoms of adjacent prolyl residues are *trans* to each other, and the *cis* configuration in which the two α-carbon atoms attain a *cis* position.



trans



cis

Inspection of molecular models of poly-L-proline shows that rotation about the third type of bond along the peptide chain, the C(α)—CO bond, is restricted by steric hindrance. This bond can assume two rotational positions, each having considerable freedom of oscillation. The X-ray configurational analysis of the two forms of poly-L-proline revealed that in both cases the C(α)—CO bond assumes the same position resulting in a left-handed helix for the all-*trans* poly-L-proline II and in a right-handed helix for the all-*cis* poly-L-proline I. The restrictions to free rotation imposed on all the bonds of the peptide backbone should suffice to stabilize the helices described. On the other hand, any freedom of oscillation about these bonds will lend flexibility to the macromolecules resulting in hydrodynamic behaviour deviating from that expected for rigid rods.

In all the polymers of the cyclic amino acids investigated, strong acids were found to catalyse mutarotation, i.e. the transformation of the *cis* helices into *trans* helices and *vice versa*. This catalysis was explained in terms of the temporary abolishment of the double-bond character of the peptide bond, due to protonation of the NCO nitrogen. Neutral salts as LiBr, $CaCl_2$ and KCNS, were also found to effect markedly the conformation in solution of poly-L-proline, as well as of some of the other polyamino acids investigated. These salts, at high concentrations, destroy the regular asymmetric macromolecular conformation, and lead to the formation of randomly coiled chains. The mechanism of this effect is still obscure; however, since no *cis-trans* isomerization seems to be involved, it is plausible to assume that the macromolecular randomization is caused by the induction of free rotation about the C(α)—CO bond.

During the forward and reverse mutarotations described intermediate conformations appear. Assuming that the fraction of peptide bonds in the *cis* or *trans* configuration, at any given specific rotation [α] is the same for every polypeptide molecule in solution, the conformation of

each macromolecule will be determined by the distribution of the *cis* and *trans* CON bonds along the backbone. No information is as yet available on the nature and distribution of these macromolecular conformations. It is obvious, however, that while the hydrodynamic properties are determined by the overall shape of the various polypeptide conformations in solution, the optical rotatory properties are determined by the macromolecular asymmetry of relatively short helical segments. In this connection it is of interest to recall that in the case of poly-L-proline it was possible to derive, from one sample, a number of solutions with the same specific optical rotation but with markedly different relative viscosities (see Fig. 6).

The structures in the solid state of polydehydro-L-proline, poly-*O*-acetyl and poly-*O*-tosylhydroxy-L-proline have not as yet been worked out. The optical rotatory characteristics, as well as the course of their forward and reverse mutarotations, strongly suggest, however, that they also may attain conformations similar to those of poly-L-proline. The form with the largest laevorotation was denoted Form I and the other extreme form was denoted Form II. In all polymers of this group the product of polymerization in pyridine or dioxane proved to be Form I, which could be converted into Form II by treatment with formic acid. In view of the above it may be assumed that Forms I and II of the three polymers discussed correspond with respect to conformation to Form I and II of poly-L-proline respectively. In the case of polyhydroxy-L-proline only one form is known. Because of its high laevorotation we are inclined to assign to it a *trans* left-handed helical conformation analogous to that of poly-L-proline II.

Poly-L-pipecolic acid could also be obtained in two conformational forms. In this case, however, the product of polymerization of the corresponding *N*-carboxyanhydride in dioxane, denoted as Form I, shows the largest laevorotation, whereas the product of complete mutarotation in formic acid (Form II) has the lowest laevorotation. It is thus difficult to assign definite conformations to the two forms. On the basis of optical rotatory properties alone, one would assign a *trans* left-handed conformation to Form I and a *cis* right-handed conformation to Form II.

An investigation of the simultaneous polymerization of *N*-carboxy-L-proline anhydride with *N*-carboxyanhydrides of other amino acids revealed that, whereas proline copolymerizes readily with glycine, sarcosine and *O*-acetylhydroxy-L-proline, it does not interact to an appreciable extent with γ-benzyl-L-glutamate and ϵ-carbobenzoxy-L-lysine. From the optical rotatory behaviour of the various proline copolymers obtained it could be concluded that they differ widely in the manner in which the prolyl residues are distributed along the chains of the copolymer molecules. In the sarcosine–proline copolymers blocks of

prolyl residues of considerable size alternate with blocks of sarcosyl residues. This was concluded from their optical rotatory properties, such as specific optical rotation per proline residue as well as forward and reverse mutarotation, found to be very close to those of poly-L-proline. The above properties persist even at relatively high molar ratios of sarcosine to proline (8:1). In the glycine–proline copolymers, on the other hand, a random distribution of both residues could be demonstrated at least in copolymers with a high glycine to proline molar ratio (6:1). In these copolymers the optical rotatory characteristics of poly-L-proline disappear completely and the proline residues behave like those of poly-L-proline when in a random conformation. At lower glycine to proline ratios features such as the existence of two forms, as well as mutarotation, can be observed. In this connection it is pertinent to note that copolymers of dehydro-L-proline and glycine resemble in their behaviour in solution the copolymers of proline and glycine. However, a much larger glycine to dehydro-L-proline ratio (30:1) is necessary in order to eliminate the tendency of the dehydro-L-prolyl residues to form dehydroprolyl blocks during copolymerization.

The findings presented in this article show that the characteristic conformation of the peptide chains of the collagen molecule can be attained in solution by poly-L-proline as well as by the other analogous polymers of the cyclic secondary amino acids investigated. Furthermore, the conformational changes studied in considerable detail in the case of poly-L-proline can be reproduced in poly-*O*-acetylhydroxy-L-proline, polydehydro-L-proline, as well as in poly-L-pipecolic acid. Finally, it may be significant that those amino acids which readily copolymerize with L-proline, i.e. glycine and hydroxy-L-proline, also occur together with L-proline in the sequence Gly.Pro.Hypro which most likely determines the main structural characteristics of the collagen molecule.

Acknowledgement

Most of the work described in this paper has been supported by grant No. A-3083 of the National Institutes of Health, United States Public Health Service.

References

Berger, A., Loewenstein, A. and Meiboom, S. (1959). *J. Amer. chem. Soc.* **81**, 62.
Blout, E. R. and Fasman, G. D. (1958). In *Recent Advances in Gelatin and Glue Research* (G. Stainsby, ed.), p. 122. Pergamon Press, London.
Cowan, P. M. and McGavin, S. (1955). *Nature, Lond.* **176**, 501.
Cowan, P. M., McGavin, S. and North, A. C. T. (1955). *Nature, Lond.* **176**, 1062.
Crick, F. H. C. and Rich, A. (1955). *Nature, Lond.* **176**, 780.
Downie, A. R. and Randall, A. A. (1959). *Trans. Faraday Soc.* **55**, 2132.
Harrington, W. F. and Sela, M. (1958). *Biochim. biophys. Acta* **27**, 24.
Harris, G. and Pollock, J. R. A. (1953). *Chem. & Ind. (Rev.)* 462.

Karlson, R. H., Norland, K. S., Fasman, G. D. and Blout, E. R. (1960). *J. Amer. chem. Soc.* **82**, 2268.
Kirkwood, J. G. and Riseman, J. (1948). *J. chem. Phys.* **16**, 565.
Kurtz, J., Berger, A. and Katchalski, E. (1956). *Nature, Lond.* **178**, 1066.
Kurtz, J., Berger, A. and Katchalski, E. (1958*a*). In *Recent Advances in Gelatin and Glue Research* (G. Stainsby, ed.), p. 131. Pergamon Press, London.
Kurtz, J., Fasman, G. D., Berger, A. and Katchalski, E. (1958*b*). *J. Amer. chem. Soc.* **80**, 393.
Kurtz, J., Robertson, A. V., Witkop, B. and Berger, A. (1962). Unpublished.
Mende, F. (1896). *Ber. dtsch. chem. Ges.* **29**, 2887.
Morawiecki, A. and Katchalski, E. (1962). Unpublished.
Pauling, L. (1948). *The Nature of the Chemical Bond*, p. 207. Cornell University Press, New York.
Ramachandran, G. N. and Kartha, G. (1954). *Nature, Lond.* **174**, 269.
Rich, A. and Crick, F. H. C. (1955). *Nature, Lond.* **176**, 915.
Rich, A. and Crick, F. H. C. (1958). In *Recent Advances in Gelatin and Glue Research* (G. Stainsby, ed.), p. 20. Pergamon Press, London.
Rich, A. and Crick, F. H. C. (1961). *J. mol. Biol.* **3**, 483.
Robertson, A. V. and Witkop, B. (1962). *J. Amer. chem. Soc.* **84**, 1697.
Sasisekharan, V. (1959). *Acta cryst.* **12**, 897.
Scheraga, H. A. and Mandelkern, L. (1953). *J. Amer. chem. Soc.* **75**, 179.
Simha, R. (1940). *J. phys. Chem.* **44**, 25.
Steinberg, I. Z., Berger, A. and Katchalski, E. (1958). *Biochim. biophys. Acta* **28**, 647.
Steinberg, I. Z., Harrington, W. F., Berger, A., Sela, M. and Katchalski, E. (1960). *J. Amer. chem. Soc.* **82**, 5263.
Traub, W. (1962). Personal communication.
Traub, W. and Shmueli, U. (1963). This Symposium.
Yaron, A. and Berger, A. (1961). *Bull. Res. Council Israel* **10A**, 46.

Discussion

G. N. RAMACHANDRAN: The reason why polyhydroxyproline behaves in quite a different way from polyproline and poly-*O*-acetyl-hydroxyproline seems to be that, in solution, poly-L-hydroxyproline would occur as a triple chain, while the other two would occur as single chains. This is so because in polyhydroxyproline there are strong OH···O bonds, leading to the triple chain structure, and it will be comparatively difficult for this triple chain to change over to the *cis* form, unlike the case of single chains.

As regards poly-L-pipecolic acid, it looks rather inappropriate to draw the analogy between benzyl aspartate and benzyl glutamate, because when one tries to build up the collagen type helix with the side groups in the L-configuration, one obtains only the left-handed helix unlike the alpha helix (which could be of either hand). Hence, the *trans* isomer should have only a negative optical rotation, whether it is form I or form II.

E. KATCHALSKI: Yes. For the sodium line, it is about $-320°$ and it drops to about $-50°$ when put in formic acid.

M. S. NARASINGA RAO (Regional Research Laboratory, Hyderabad): You seem to suggest that the differences of sedimentation coefficient (as a function of concentration) of the two forms can be explained on the basis of differences in viscosity of the two forms. If this is so, then the product of $S_{20,w}$ and viscosity for the two forms should fall on the same curve.

E. KATCHALSKI: They do. I would like to add that for low molecular weights the polyproline molecule can be taken to be something like an ellipsoid of revolution and the axial ratios calculated fit well with the model. But for high molecular weights, this does not agree, since then there can be a number of kinks in the molecule.

M.S. NARASINGA RAO: I want to mention that we have found that lithium thiocyanate degrades protein. Recently it has been found that lithium salts in general denature protein.

Optical Rotatory Dispersion and the Structure of Polypeptides and Proteins in Solution

ELKAN R. BLOUT

Children's Cancer Research Foundation

and

Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts, U.S.A.

ABSTRACT

Recent experimental work on the optical rotatory dispersion of synthetic polypeptides and proteins is reviewed. It is shown that measurements of rotatory dispersion allow the determination of both helix contents and the sense of helix of α-helical polypeptides and proteins. Rotation measurements in the region 185–250 mμ have revealed the presence of two characteristic Cotton effects in α-helical polypeptides and proteins. Recent optical rotatory dispersion measurements on poly-L-proline II and native collagen solutions in the region 190–240 mμ disclosed an intrinsic negative Cotton effect characteristic of helical forms of these materials.

Although X-ray diffraction is a powerful tool for the determination of the structure of proteins in the solid state, it is not now possible to determine the structure of such biological macromolecules in solution using this technique. In fact, there are no methods for structural determination of organic compounds in solution which compare with the precision attainable with X-ray diffraction on crystalline compounds. However, since many of the important biological functions of proteins and nucleic acids occur in water solution or in hydrated gels, much work has been done recently to develop methods for molecular conformational determinations in solution. One of the most powerful techniques is that of optical rotatory dispersion and in this paper we will briefly review some of the recent optical rotatory results on proteins and polypeptides and present some new data on the far ultra-violet rotatory dispersion of such compounds.

Although the optical rotatory power of organic compounds has been known for about 100 years, it is only comparatively recently that equipment for the measurement of rotatory dispersion, that is the measurement of optical rotation as a function of wavelength, has been developed. In fact, optical rotatory dispersion investigations, both in the

sense of improved experimental approaches and the fundamental theory, are under active development at this time.

Rotatory dispersion measurements on proteins were recorded in 1927 (Hewitt, 1927; Jessen-Hansen, 1927) but little further work was done until about 1955 perhaps because of the lack of ability to interpret the data in terms of structure. It had been known for some time that native proteins showed much lower negative values of rotation (usually at the sodium D line, 589 mμ) than in their denatured states. Because of such phenomenological observations, measurements of optical rotation were often used as an indication of the native states of proteins. Following the discovery of the α-helix as an important structure in proteins (Pauling and Corey, 1950, 1951), it was but a step to the recognition that such structures play a significant part in the optical rotatory properties of proteins. Thus, Cohen's suggestion (Cohen, 1955) that the rotation of native proteins was due in large part to their α-helical content was soon followed by theoretical work by Moffitt (Moffitt, 1956*a*, *b*; Moffitt and Yang, 1956) on the rotatory dispersion of synthetic polypeptides.

From recent investigations on synthetic polypeptides and proteins, it is clear now that the structureless "random chain" conformation of these substances show simple rotatory characteristics in spectral regions far removed from their absorption bands—that is in the readily accessible ultra-violet and visible, 350–600 mμ. In this spectral region the rotatory dispersions of random chain conformations of polypeptides and denatured proteins follow the one-term Drude equation

$$[\alpha]_\lambda = \frac{k}{(\lambda^2 - \lambda_c^2)} \tag{1}$$

However, when a synthetic polypeptide is in a helical conformation, the rotatory dispersion data do not follow the one-term Drude relation but a new result is found, namely that the dispersion is anomalous and a two-term equation is required. The equation proposed by Moffitt

$$[R']_\lambda = \frac{a_0 \lambda_0^2}{\lambda^2 - \lambda_0^2} + \frac{b_0 \lambda_0^4}{(\lambda^2 - \lambda_0^2)^2} \tag{2}$$

has been shown to be the most successful for fitting the experimental data both for synthetic polypeptides and proteins. It has been found experimentally that the coefficient of the second term, b_0, has a value of approximately $-630°$ in almost all of the fully helical synthetic polypeptides examined thus far (Blout, 1960; Urnes and Doty, 1961). Further, it has been shown that with synthetic polypeptides which are not fully helical, such as copolymers of L-glutamic acid and L-lysine, the extent of the helix formation can be estimated from the magnitude of b_0 (Blout and Idelson, 1958). As can be seen from Eq. (2), when $b_0 = 0$, Eq. (2) reduces

in form to the Drude equation (Eq. 1). If the assumption is made that proteins consist only of helical and random segments (and this may not be a valid assumption) then the helix contents of many proteins can be obtained by rotatory dispersion measurements in the region 350–600 mμ.

In addition to showing the presence of helix, rotatory dispersion can also reveal the sense of helix and possibly the presence of other ordered structures. Before describing some recent work on the ultra-violet rotatory dispersion of polypeptides and proteins, I would like to review briefly rotatory work on such compounds involving (1) the determination of helix contents; (2) the determination of the sense of helix, and (3) attempts to characterize β conformations.

Helix Contents of Polypeptides and Proteins

As noted above, the helix contents of polypeptides and proteins can be estimated from rotatory dispersion data using the coefficient of the second term of the Moffitt equation, b_0, and this has been done for some fifteen well-characterized proteins and the data reviewed (Blout, 1960; Urnes and Doty, 1961). However, two other techniques for helix measurements employing rotatory dispersion have been described recently. The first method involves the binding of dyes to polypeptides and proteins (Blout and Stryer, 1959; Stryer and Blout, 1961), and the second involves rotation measurements in the far ultra-violet (Simmons and Blout, 1960; Simmons *et al.*, 1961; Beychok and Blout, 1961; Beychok *et al.*, 1962). It has been found that certain symmetric dye molecules, such as acriflavine and acridine orange, when bound to helical forms of synthetic polypeptides, show anomalous rotatory dispersion, that is Cotton effects, in the absorption bands of the dyes. Since the dyes alone possess no optical rotatory power, and the dyes bound to random conformations of synthetic polypeptides show only the rotatory characteristics of the synthetic polypeptides, it is clear that the observed anomalous rotatory dispersion of the dye:polypeptide complex is indicative of the fact that in these complexes the chromophoric group of the dye has acquired asymmetry. An example of this effect is shown in Fig. 1, where the rotatory dispersion of acridine orange bound to both random and helical forms of poly-α,L-glutamic acid is shown. The exact nature of this dye–helix phenomenon and its usefulness as a probe for determining specific protein structures has not yet been investigated in detail. However, recent experimental observations with insulin–acridine orange complexes have been made and similar types of anomalous rotatory dispersion have been recorded (Beychok and Blout, to be published).

The observed rotatory dispersion of synthetic polypeptides and

proteins in the visible and near ultra-violet is clearly a consequence of much larger rotatory effects associated with absorption bands in the far ultra-violet region and therefore it has been of some interest to determine

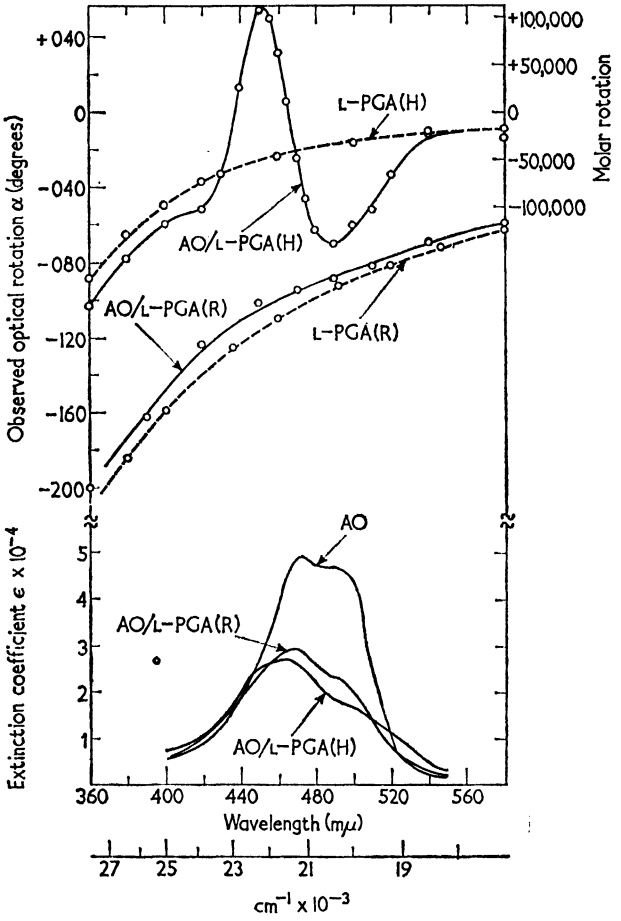


FIG. 1. Optical rotatory dispersions and absorption spectra of complexes of acridine orange (AO) and poly-L-glutamic acid (L-PGA). The absorption spectrum of unbound acridine orange markedly changes upon binding to either the random conformation of poly-L-glutamic acid (AO/L-PGA(R)) or the helical conformation (AO/L-PGA(H)); only the complex of acridine orange and the helical conformation of L-PGA (AO/L-PGA(H)) exhibits a Cotton effect. (Stryer and Blout, 1961.)

whether it is possible to measure the rotatory behaviour of such materials near their intrinsic ultra-violet absorption bands. Secondary amides, peptides and polypeptides show fundamental absorption bands due to the

peptide group at about 145 and 190 mμ (Peterson and Simpson, 1957). The absorption band at 145 mμ can only be measured using vacuum spectroscopic methods and, therefore, is of little interest in the determination of macromolecular structures in water solution. But, in addition to the 190 mμ $\pi \rightarrow \pi^*$ absorption of peptides, it has been shown recently that the $n \rightarrow \pi^*$ transition around 225 mμ can be observed in the ultra-violet spectra of polypeptides and proteins (Glazer and Smith, 1960; Rosenheck

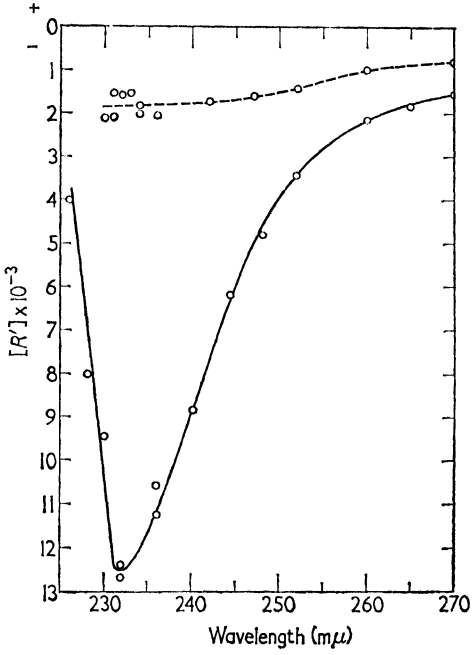


FIG. 2. The ultra-violet rotatory dispersion of (full line) poly-L-glutamic acid at pH 4·5 in water; (broken line) poly-L-glutamic acid (sodium salt) at pH 7·4 in water. (Simmons *et al.*, 1961.)

and Doty, 1961; Gratzer *et al.*, 1961; Doty and Gratzer, 1962; Tinoco *et al.*, 1962).

In an attempt to extend the range and usefulness of rotatory dispersion methods with polypeptides and proteins, Dr. Norman Simmons and other colleagues in our laboratory have recently successfully measured the rotation of such substances in the wavelength region 225–350 mμ. The initial measurements with tobacco mosaic virus protein (Simmons and Blout, 1960) suggested the presence of a negative Cotton effect with a trough at 233 mμ. Further measurements on α-helical polypeptides, muscle proteins, and some globular proteins have confirmed those

experiments and the more recent observations (Simmons *et al.*, 1961; Beychok and Blout, 1961; Beychok *et al.*, 1962) show the presence of a negative Cotton effect with an inflection point at 225 mμ and a trough around 233 mμ (Figs. 2 and 3). This 225 mμ Cotton effect is characteristic of the helical conformation since conversion of polyglutamic acid to its random form eliminates the 225 mμ Cotton effect (Fig. 2). The 225 mμ Cotton effect is decreased in size or removed if proteins are treated with known denaturing agents such as urea (Fig. 3). It has been shown also that the magnitude of the rotation at 233 mμ in these model substances

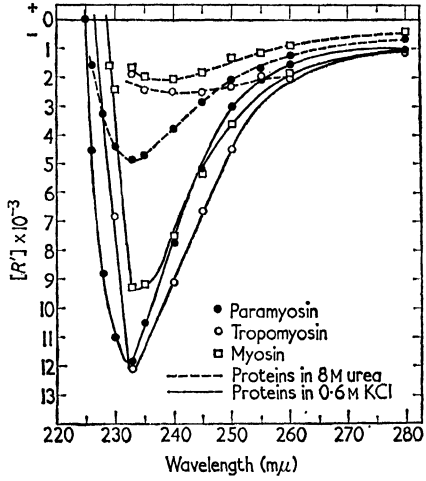


FIG. 3. The ultra-violet rotatory dispersions of some fibrous α-proteins. (Simmons *et al.*, 1961.)

correlates with their known helix contents. Some data are shown in Table I.

Ultra-violet optical rotatory dispersion measurements in addition to providing a new parameter for estimates of helix contents in simple proteins and polypeptides, may provide structural information in more complex systems, for example proteins containing visible absorbing chromophores (*vide infra*) which are difficult to analyse by rotatory dispersion measurements using visible and near ultra-violet wavelengths.

SENSE OF HELIX

Early results using the Moffitt equation to analyse optical rotatory dispersion data of simple helical homopolypeptides showed that all such substances gave b_0 values which were not only relatively constant but

TABLE I. Some rotatory properties of certain synthetic polypeptides and proteins

Material	Helical				Random	
	b_0	% Helix[a] calc. from b_0	$[R']_{233}$	% Helix[b]	b_0	$[R']$
Polypeptides						
Poly-L-glutamic acid[c]	−600	95	−12,600	99	0	−1780
Poly-L-methionine[c]	−630	100	−10,400	79	0	—
Poly-γ-benzyl-L-glutamate[c]	−670	106	−12,800	101	0	—
Poly-β-benzyl-L-aspartate[d]	+630	100	+9500	102 (left-handed)		
	Native				Denatured	
Proteins						
Paramyosin[e]	~ −600	95	−11,600	90	−220	−4850
Tropomyosin[e]	~ −600	95	−12,100	95	0	−1920
Myosin[e]	−380	60	−8700	63	0	−1670
Ferrimyoglobin[e]	−509	78	−10,000	75		
Ferrihaemoglobin[e]	−520	80	−10,000	75		
Ribonuclease[f]	−100[g]	16[g]	−4000	13(?)–20		
Bovine serum albumin[g]	−290[g]	46[g]	−7000[h]	48[h]		
Tobacco mosaic virus protein[i]	−160	25	−5600	35		

[a] Assuming mean value of $b_0 = -630$ for 100% helix, $b_0 = 0$ for 0% helix, and a linear interpolation.
[b] Assuming mean value of $[R']_{233} = -12{,}700$ for 100% helix, $[R']_{233} = 1800$ for 0% helix, and a linear interpolation.
[c] Simmons *et al.* (1961).
[d] Blout and Karlson (1958); Karlson *et al.* (1960).
[e] Beychok and Blout (1961); Beychok *et al.* (1962).
[f] Zimmerman and Schellman (1962).
[g] Doty (1960).
[h] Blout *et al.* (1962).
[i] Simmons and Blout (1960).

had one sign—negative. However, when measurements were made with poly-β-benzyl-L-aspartate (Blout and Karlson, 1958; Karlson *et al.*, 1960; Bradbury *et al.*, 1960) it became clear that this substance, while completely helical, showed a positive b_0 value of about +630. Proof that this positive b_0 value indicated the opposite sense of helix was obtained by preparing copolymers of benzyl-L and benzyl-D-aspartate with γ-benzyl-L-glutamate. Recently we have observed the rotatory dispersion of poly-β-benzyl-L-aspartate in the region 225–250 mμ (Blout, unpublished)

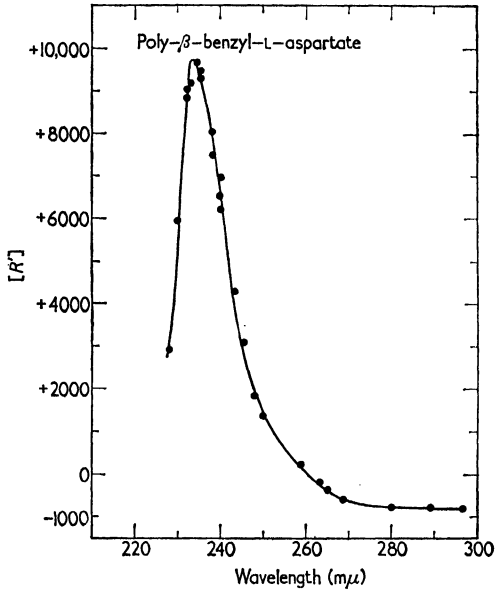


FIG. 4. Ultra-violet rotatory dispersion of poly-β-benzyl-L-aspartate in methylene dichloride solution. These data show the beginning of a *positive* Cotton effect.

(Fig. 4) and find that a positive Cotton effect is observed with a maximum at ~230 mμ and an inflection point around 225 mμ. The conclusion thus seems inescapable that poly-β-benzyl-L-aspartate has the opposite sense of helix to poly-γ-benzyl-L-glutamate and many other L-polypeptides.

Another method of obtaining the information regarding the sense of helix is to make use of the dye-binding technique mentioned above and some experiments have been performed. In Fig. 5 there is shown the rotatory dispersion of acridine orange bound to (a) poly-α,L-glutamic acid and to (b) poly-α,D-glutamic acid. It should be noted that whereas a negative Cotton effect is observed for the L-polypeptides, a positive Cotton effect in the same spectral region but opposite in magnitude, is

observed for the D-polypeptide. Since the D-polypeptide must have the helical sense opposite to that of the L-polypeptide, it is possible that such dye-binding rotatory dispersion studies can prove useful in determining the helical sense in more complex systems such as proteins.

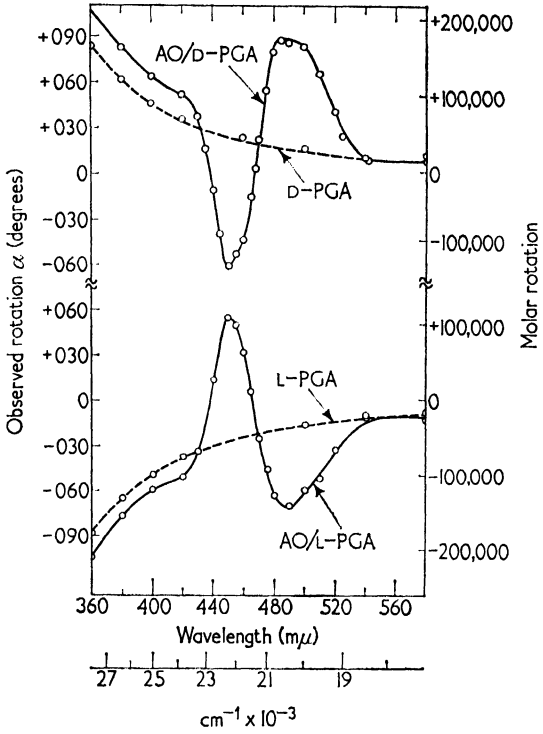


FIG. 5. Optical rotatory dispersions of complexes of acridine orange and helical poly-L-glutamic acid (AO/L-PGA), and of acridine orange and the polypeptide of opposite screw-sense of helix, poly-D-glutamic acid (AO/D-PGA). (Stryer and Blout, 1961.)

OTHER STRUCTURES

In addition to α-helical and random chain conformations, extended or β conformations have been observed in fibrous proteins by X-ray diffraction techniques. However, pure β conformations, by their very nature, being intermolecularly hydrogen bonded, generally result in much decreased solubility of the polypeptide. Thus, little work has been done to ascertain the rotatory dispersive properties of such structures in solution. At this point it can be said that the characterization of such structures by rotatory dispersive techniques does not rest on as nearly a firm foundation

as does the characterization of helical conformations. What is interesting, however, is that two investigations (Fasman and Blout, 1960; Wada *et al.*, 1961) have yielded significantly different results. However, the dispersion data of both groups of workers show that the β structures investigated have rotatory properties which are qualitatively different from helix and random chains. Thus it appears likely that if such conformations form a large component of a protein's structure, they will significantly affect the rotatory properties of the protein.

Recent Investigations

Now let us consider three recent investigations related to the rotatory dispersive properties of proteins.

(a) *Ultra-violet rotatory dispersion of myoglobin and haemoglobin*

The X-ray diffraction studies of Kendrew *et al.* (1961) on ferrimyoglobin have demonstrated directly that there are eight segments of right-handed α-helix in this compound accounting for some 75% of the molecule. Since these studies were performed on wet protein crystals, it was of some interest to determine whether the helix contents of this molecule in aqueous solution corresponded to that observed in the solid state. Because, as is well known, both myoglobin and haemoglobin show strong absorption in the visible and near ultra-violet region, due to the presence of the heme moiety, measurements of optical rotatory dispersion were made in the ultra-violet (Beychok and Blout, 1961). As can be seen from Fig. 6, the rotatory dispersions in the region 400–225 mμ for both ferrimyoglobin and ferrihaemoglobin are very similar, and both show the 233 mμ trough of the 225 mμ Cotton effect. Analyses of these data in terms of the Moffitt equation over the wavelength region 236–334 mμ (outside the Cotton effect) gave values of b_0 for myoglobin of -509 and for haemoglobin of -520. Using the scale of b_0 values, random $\rightarrow$ 100% helix $= 0 \rightarrow -630$, we find helix contents of about 75%. Independent measurements by other workers have confirmed the value for ferrimyoglobin (Urnes *et al.*, 1961). Also, estimates of helix content have been made using the magnitude of the rotation trough at 233 mμ (Table I). Calculations using these data agree well with those obtained from ultra-violet dispersion measurements and thus provide additional examples of satisfactory agreement between the two methods. Similar types of measurements have been performed on solutions of denatured ferrihaemoglobin and ferrimyoglobin (Beychok *et al.*, 1962). The denatured proteins show helix contents ranging from 10 to about 30% depending on the conditions of denaturation. It has also been shown that a visible Cotton effect around 405 mμ associated with the heme moiety disappears

upon denaturation, thus providing additional evidence of the sensitivity of rotatory dispersion measurements to molecular conformation. Another relevant observation is the finding that haemin can be complexed with helical poly-L-lysine to show a visible Cotton effect in the haem absorption band whereas when haemin is complexed with the random form of poly-L-lysine no such Cotton effect is observed (Stryer, 1961).

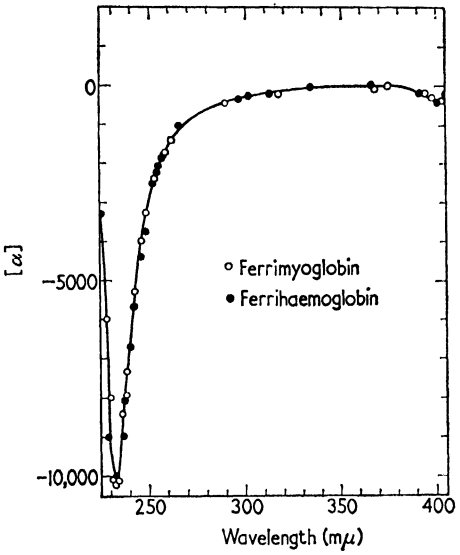


FIG. 6. Ultra-violet optical rotatory dispersion of water solutions of ferrihaemoglobin and ferrimyoglobin at pH 6·6 with no added electrolyte. (Beychok and Blout, 1961.)

(b) $\pi \rightarrow \pi^*$ *Transitions and the rotatory dispersion of polypeptides and proteins*

Recent improvements in experimental apparatus have allowed the determinations of optical rotatory dispersions in the wavelength region 185 to 225 mμ (Blout *et al.*, 1962). As might have been expected, anomalous optical rotations are observed in this region with polypeptides and proteins. Using poly-α,L-glutamic acid as a model compound, we have measured the rotatory dispersion of both the helical and random conformations (Fig. 7). The helical conformation shows a large positive Cotton effect with a maximum or peak at approximately 198 mμ and an inflection point at about 190 mμ. Like the 225 mμ Cotton effect, the 190 mμ Cotton effect is conformation dependent. The random coil form of this polypeptide shows much lower rotations in this spectral region and the presence of a weaker negative Cotton effect with a trough around 204 mμ

and an inflection point at 197 mμ. Some measurements on proteins as well as on other synthetic polypeptides have confirmed the original findings. For example, bovine serum albumin shows a minimum at 233 mμ and a maximum at 198 mμ in its optical rotatory dispersion curve; the

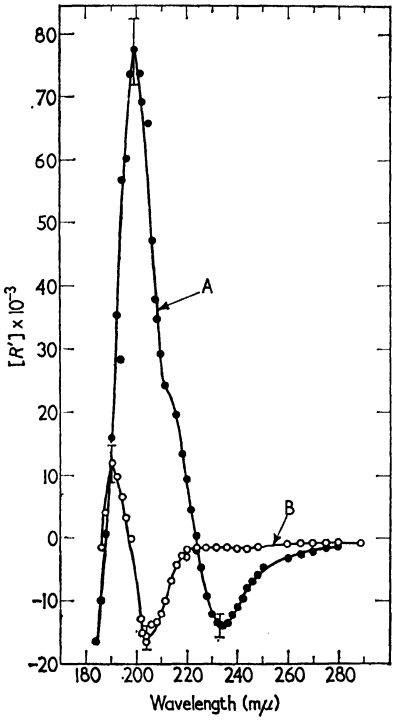


FIG. 7. Curve A: the ultra-violet optical rotatory dispersion of the helical form of poly-α,L-glutamic acid in water solution, pH 4·3. A 1-mm cell was used. Concentrations ranged from 0·0137 to 0·0456%. Curve B: the ultra-violet optical rotatory dispersion of the random coil form of poly-α,L-glutamic acid (sodium salt), pH 7·1 in water solution. A 1-mm cell was used. Concentrations ranged from 0·0176 to 0·400%. The vertical lines at the peaks and troughs indicate the range of experimental uncertainty. (Blout *et al.*, 1962.)

magnitude of the minimum and maximum correspond with helical constants in the range 55–60%. It appears from the position of the 190 mμ Cotton effect that this effect is directly related to the fundamental $\pi \rightarrow \pi^*$ absorption observed for helical polypeptides. It now seems that there are three rotatory phenomena which allow at least semi-quantitative measurement of helix constants of polypeptides and proteins, namely, (1) the rotatory dispersion data over a long wavelength range in

either visible or ultra-violet, (2) the magnitude of the rotation trough at 233 mμ, and (3) the magnitude of the rotation peak at 198 mμ.

(c) *Ultra-violet rotatory dispersion of collagen*

Because of the research interests and contributions of the convener of this symposium, Professor Ramachandran, I am very pleased that we are able to report here some new data on collagen. The structure of collagen in the solid state is now generally accepted to be that of three left-handed

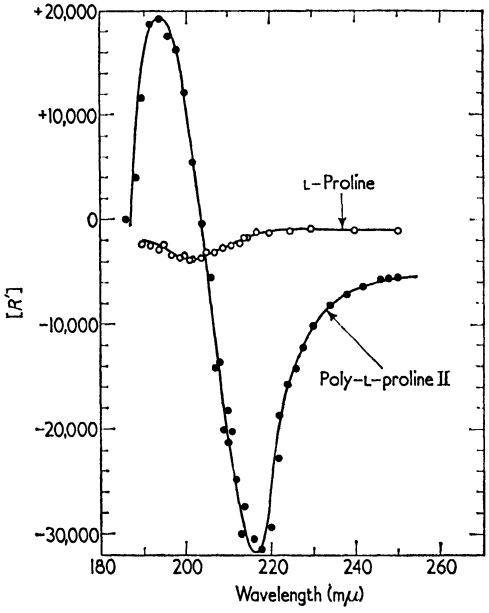


FIG. 8. The ultra-violet rotatory dispersions of poly-L-proline II and L-proline in water solution. (Blout *et al.*, 1963.)

helices coiled into a right-handed super-helix (Ramachandran and Kartha, 1954, 1955; Ramachandran and Sasisekharan, 1961; Rich and Crick, 1955, 1961). Although much work on collagen solutions has been reported, few definitive structural conclusions have emerged. Recently, however, stimulated by new rotatory data on poly-L-proline, we have begun investigation of the far ultra-violet rotatory dispersion of collagen. Work with poly-L-proline revealed the presence of a fairly strong absorption band between 202 and 208 mμ in both poly-L-proline I and poly-L-proline II (Fasman and Blout, 1963). This absorption band probably has its origin in a $\pi \rightarrow \pi^*$ transition as does the band observed around 190 mμ in α-helical polypeptides.

Further investigations with poly-L-proline II, the water soluble form of this polypeptide, have now shown the presence of a strong negative Cotton effect with a minimum at 216 mμ, a maximum at 194 mμ and an inflection point at 203 mμ (Blout *et al.*, 1963). This Cotton effect appears to be related directly to the conformation of poly-L-proline, since L-proline shows no such effect (Fig. 8). Following this observation, the far ultra-violet spectra of solutions of native and denatured calfskin collagen

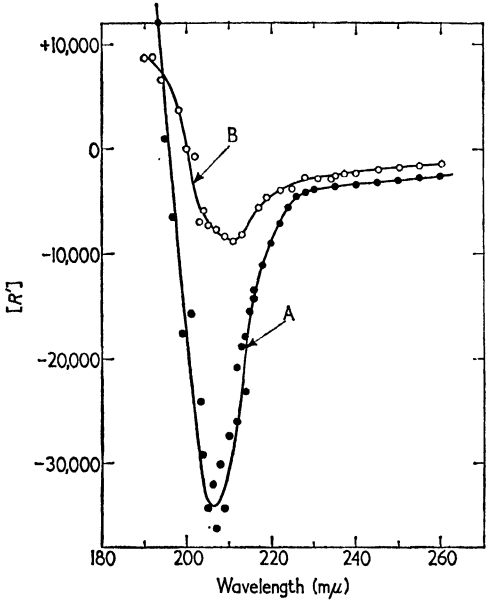


FIG. 9. Curve A: the ultra-violet rotatory dispersion of native calfskin collagen in 0·01 M acetic acid solution. Curve B: the ultra-violet rotatory dispersion of the same preparation of calfskin collagen denatured by heating at 50° C for 30 min and measured immediately on cooling to 25° C. (Blout *et al.*, 1963.)

were measured and absorption maxima in the region 185–220 mμ observed. Rotatory dispersion measurements in this region revealed that native collagen solutions also show a strong negative Cotton effect (Fig. 9) similar in magnitude to that observed with poly-L-proline but with an inflection point at 195 mμ (Blout *et al.*, 1963). Thus, it appears that the helical structure of poly-L-proline and collagen are similar *in solution.*

Acknowledgement

I wish to acknowledge my indebtedness to my colleagues Drs. S. Beychock, C. Cohen, C. de Lozé, G. D. Fasman, J. Gross, I. Schmier, N. S. Simmons, L. Stryer,

A. G. Szent-Gyorgyi and D. B. Wetlaufer, and Messrs. J. P. Carver and K. S. Norland, without whose tireless efforts much of the work reported here would not have been carried out. I am also pleased to acknowledge financial support from the U.S. Public Health Service and the Office of the Surgeon-General, Department of the Army.

References

Beychok, S. and Blout, E. R. (1961). *J. mol. Biol.* **3**, 769.
Beychok, S., de Lozé, C. and Blout, E. R. (1962). *J. mol. Biol.* **4**, 421.
Blout, E. R. (1960). In *Optical Rotatory Dispersion* (C. Djerassi, ed.), p. 238. McGraw-Hill, New York.
Blout, E. R. and Idelson, M. (1958). *J. Amer. chem. Soc.* **80**, 4909.
Blout, E. R. and Karlson, R. H. (1958). *J. Amer. chem. Soc.* **80**, 1259.
Blout, E. R. and Stryer, L. (1959). *Proc. nat. Acad. Sci., Wash.* **45**, 1591.
Blout, E. R., Schmier, I. and Simmons, N. S. (1962). *J. Amer. chem. Soc.* **84**, 3193.
Blout, E. R., Carver, J. P. and Gross, J. (1963). *J. Amer. chem. Soc.* **85**, 644.
Bradbury, E. M., Downie, A. R. Elliott, A. and Hanby, W. E. (1960). *Proc. roy. Soc.* **A259**, 110.
Cohen, C. (1955). *Nature, Lond.* **175**, 129.
Doty, P. (1960). *Proc. 4th Int. Cong. Biochem.*, Vol. 8, p. 8. Vienna.
Doty, P. and Gratzer, W. B. (1962). In *Polyamino Acids, Polypeptides, and Proteins*, (M. A. Stahmann, ed.), p. 111. University of Wisconsin Press.
Fasman, G. D. and Blout, E. R. (1960). *J. Amer. chem. Soc.* **82**, 2262.
Fasman, G. D. and Blout, E. R. (1963). *Biopolymers*, **1**, 3.
Glazer, A. N. and Smith, E. L. (1960). *J. biol. Chem.* **235**, PC43.
Gratzer, W. B., Holzwarth, G. M. and Doty, P. (1961). *Proc. nat. Acad. Sci., Wash.* **47**, 1785.
Hewitt, L. F. (1927). *Biochem. J.* **21**, 216.
Jessen-Hansen, H. (1927). *C. R. trav. Lab. Carlsberg, ser. chim.* **16**, No. 10.
Karlson, R. H., Norland, K. S., Fasman, G. D. and Blout, E. R. (1960). *J. Amer. chem. Soc.* **82**, 2268.
Kendrew, J. C., Watson, H. W., Strandberg, B. E., Dickerson, R. E., Phillips, D. C. and Shore, V. C. (1961). *Nature, Lond.* **190**, 666.
Moffitt, W. (1956*a*). *J. chem. Phys.* **25**, 467.
Moffitt, W. (1956*b*). *Proc. nat. Acad. Sci., Wash.* **42**, 736.
Moffitt, W. and Yang, J. T. (1956). *Proc. nat. Acad. Sci., Wash.* **42**, 596.
Pauling, L. and Corey, R. B. (1950). *J. Amer. chem. Soc.* **72**, 5349.
Pauling, L. and Corey, R. B. (1951). *Proc. nat. Acad. Sci., Wash.* **37**, 235.
Peterson, D. L. and Simpson, W. T. (1957). *J. Amer. chem. Soc.* **79**, 2375.
Ramachandran, G. N. and Kartha, G. (1954). *Nature, Lond.* **174**, 269.
Ramachandran, G. N. and Kartha, G. (1955). *Nature, Lond.* **176**, 593.
Ramachandran, G. N. and Sasisekharan, V. (1961). *Nature, Lond.* **190**, 1004.
Rich, A. and Crick, F. H. C. (1955). *Nature, Lond.* **176**, 915.
Rich, A. and Crick, F. H. C. (1961). *J. mol. Biol.* **3**, 483.
Rosenheck, K. and Doty, P. (1961). *Proc. nat. Acad. Sci., Wash.* **47**, 1775.
Simmons, N. S. and Blout, E. R. (1960). *Biophys. J.* **I**, No. 1, 55.
Simmons, N. S., Cohen, C., Szent-Gyorgi, A. G., Wetlaufer, D. B. and Blout, E. R. (1961). *J. Amer. chem. Soc.* **83**, 4766.
Stryer, L. (1961). *Biochim. biophys. Acta* **54**, 395.
Stryer, L. and Blout, E. R. (1961). *J. Amer. chem. Soc.* **83**, 1411.

Tinoco, I. Jr., Halpern, A. and Simpson, W. T. (1962). In *Polyamino Acids, Polypeptides and Proteins* (M. A. Stahmann, ed.), p. 147. University of Wisconsin Press.

Urnes, P. and Doty, P. (1961). In *Advances in Protein Chemistry* (C. B. Anfinsen, N. L. Anson, K. Bailey and J. T. Edsall, eds.), Vol. 16, p. 401. Academic Press, New York.

Urnes, P. J., Imahori, K. and Doty, P. (1961). *Proc. nat. Acad. Sci., Wash.* **47**, 1635.

Wada, A., Tsuboi, M. and Konishi, E. (1961). *J. phys. Chem.* **65**, 1119.

Zimmerman, S. B. and Schellman, J. A. (1962). *J. Amer. chem. Soc.* **84**, 2259.

Discussion

R. NATH (All India Institute of Medical Sciences, New Delhi): Have you performed the studies on the renaturation of protein and how does it affect the Cotton effect?

E. R. BLOUT: Yes, we have done such experiments with several proteins, including denaturation studies with various muscle proteins and both denaturation and renaturation studies with myoglobin and haemoglobin. The data obtained from the latter two proteins show that the negative trough of the 225 mμ Cotton effect is markedly diminished in magnitude upon denaturation with acid. When the pH is returned to near neutrality, this trough increases in magnitude and attains values close to those of the native proteins.

G. N. RAMACHANDRAN: The helix content in paramyosin and tropomysin seems to be much more than in myosin itself. I would like to know the proline content of these in comparison with myosin, since we know that the α-helix cannot continue when a proline residue occurs.

E. R. BLOUT: As far as I remember both paramyosin and tropomyosin have very low or no proline content.

A. ELLIOTT (King's College, London): The work described by Dr. Blout enables the contribution of chromophoric groups in side-chains to be separated from that of the polypeptide chains and so removes a source of uncertainty in the interpretation of optical rotatory dispersion. It will be very interesting to know how the Cotton effect changes from one conformation to another, and in principle this could give a valuable method of assessing the proportions of alpha helix and other conformations present in solutions. I might mention that we have recently found that the value of Moffit's constant b_0 for the β form of poly-*O*-benzyl-L-serine (in solution) is about $+190°$, hence the β form should have a characteristic Cotton effect, if it could be observed.

E. R. BLOUT: I agree with Dr. Elliott that the finding of these new Cotton effects opens up an interesting field for the exploration of the relationships between optical rotatory dispersion and many of the postulated protein and polypeptide conformations.

Infra-red Studies of the Conformations of Polypeptides and Proteins

T. MIYAZAWA

Institute for Protein Research, Osaka University, Kitaku, Osaka, Japan

ABSTRACT

Polypeptides or proteins exhibit strong infra-red absorption bands at about 3300 cm^{-1} (amide A), 3100 cm^{-1} (amide B), 1650 cm^{-1} (amide I), 1550 cm^{-1} (amide II), and 700–600 cm^{-1} (amide V), which arise from the CONH groups. The amide A and B bands are due to the Fermi resonance between the fundamental N–H stretching vibration and the first overtone of the amide II vibration. The frequencies of these bands are not very sensitive to the change in the chain conformations, but the dichroic behaviour depends upon the conformations. The amide I band is due to the C = O stretching mode and the amide II band is due to the hybridization of the N—H bending and C—N stretching modes. The frequencies as well as dichroism of the amide I and II bands depend upon the conformations, namely the α-helical form, the parallel-chain β-form, the antiparallel-chain β-form, and the random coil form. Those correlations have been analysed in terms of the vibrational interactions among peptide groups in the chain as well as across interchain hydrogen bonds. The amide V bands change markedly with the chain conformations; the β-form: 700 cm^{-1}, the random coil form: *ca.* 650 cm^{-1}, and the α-form: 620 cm^{-1}. For the co-polymers of methyl-L-glutamate and methyl-D-glutamate, the fractions of the right-handed helix, the left-handed helix and the random coil form have been estimated by measuring the intensities of the amide V bands of the α-form as well as the optical rotatory dispersion.

Infra-red spectroscopy has been a useful method for studying the chain conformations as well as for studying the hydrogen bonds of polypeptides and proteins. Polypeptides or proteins exhibit strong infra-red absorption bands at about 3300 cm^{-1} (amide A), 3100 cm^{-1} (amide B), 1650 cm^{-1} (amide I) and 1550 cm^{-1} (amide II) which are characteristic of the CONH group (Sutherland, 1952; Beer *et al.*, 1959). The nature of these bands have been discussed in detail (Miyazawa, 1962).

The amide A and B bands are undoubtedly due to the N—H group, since they disappear on *N*-deuteration. The A and B bands of *N*-methylacetamide and *N*-methylformamide have been measured in the gaseous, liquid and solid states, and it has been found that, as the N—H $\cdots$O=C hydrogen bonds become stronger, the A frequencies decrease whereas the B frequencies increase. A quantitative treatment of the intensities as well

as the frequencies of the amide A and B bands (Miyazawa, 1960*a*) has confirmed the explanation that these bands arise from the Fermi resonance between the fundamental N—H stretching vibration and the first overtone of the amide II vibration (Badger and Pullin, 1954; Cannon, 1960). The amide A frequency has been used for investigating the strengths of the N—H···O=C hydrogen bonds. However, since this band arises from the Fermi resonance, the shift due to the resonance should be corrected before the frequency value is treated in connection with the strength of the hydrogen bonds.

The amide A frequencies are not very sensitive to the chain conformations. However, the dichroic behaviour of the amide A and B bands are useful for distinguishing between the α-helical conformation and the extended β-conformation. It may be recalled that the N—H bonds in the α-conformation are nearly parallel to the helix axis, whereas the N—H bonds of the β-conformation are nearly perpendicular to the axis. In accord with the direction of the N—H bonds, the amide A band (at 3293 cm^{-1}) of poly-L-alanine in the α-form shows parallel dichroism, whereas the amide A band (at 3283 cm^{-1}) of poly-L-alanine in the β-form exhibits perpendicular dichroism (Elliott, 1954).

In the 6 μ region two strong bands, namely the amide I and II bands, are observed for polypeptides and proteins. These arise from the in-plane vibrations of the CONH group. These bands were considered to arise from the vibrational interactions among the C=O stretching, C—N stretching and N—H bending modes (Fraser and Price, 1952). The quantitative nature of these bands has been elucidated by a normal co-ordinate analysis (Miyazawa *et al.*, 1958) as well as the infra-red and Raman studies of a series of monosubstituted amides (Miyazawa *et al.*, 1956). The amide I band is essentially due to the C=O stretching mode. The amide II band arises from the hybridization of the C—N stretching and the N—H bending modes. The atomic displacements for the amide I and II vibrations have been found to be localized in the CONH group.

The amide I and II bands of monosubstituted amides have been measured in various phases (Richards and Thompson, 1947; Miyazawa *et al.*, 1956). As the C=O···H—N hydrogen bonds become stronger, the amide I frequency decreases, whereas the amide II frequency increases. These frequency shifts may be understood with the above-mentioned vibrational assignments. It is expected that an increase in the hydrogen-bond strength will cause a decrease in the double-bond character of the C=O bond and an increase in the double-bond character of the C—N bond, together with an increase in the potential energy associated with the N—H bending mode.

Certain polypeptides, such as sodium poly-L-glutamate or poly-L-lysine hydrochloride, are considered to exist in the random coil form in

aqueous solution (Doty *et al.*, 1957). The strength of the hydrogen bonds has been studied for these polypeptides in deuterium oxide solutions (Miyazawa, 1962). In deuterium oxide solution the amide hydrogen atom is replaced by the deuterium atom and the amide I′ band (due to the C=O stretching mode) and the amide II′ band (primarily due to the C—N stretching mode) are observed. Thus the amide I′ and II′ bands are observed at 1645 cm^{-1} and 1460 cm^{-1}, respectively, for the COND groups which form intergroup deuterium bonds, whereas the corresponding bands are observed at 1655 cm^{-1} and 1440 cm^{-1} for the COND groups which are solvated by the D_2O molecules. Accordingly the intergroup deuterium bonds are considered to be stronger than the deuterium bonds between the peptide groups and the D_2O molecules. This conclusion has also been supported by the temperature dependence of the relative intensities (Miyazawa, 1962).

An important result emerging from the infra-red studies of polypeptides and proteins is that various chain conformations give absorption spectra which differ from each other. The correlations between the amide I and II bands and the chain conformations were first found by Ambrose and Elliott (1951), and were subsequently applied for elucidating the chain conformations of a number of polypeptides or proteins from infra-red measurements. The usefulness of these correlations may be somewhat limited if a number of different conformations co-exist as is probably the case for some proteins (Beer *et al.*, 1959).

The vibrational displacements of the amide I and II vibrations are highly localized in the CONH group as mentioned before. Accordingly the vibrational interactions among peptide groups may be treated by the first-order perturbation theory. Thus the vibrational interactions among adjacent peptide groups in the chain or through intergroup hydrogen bonds have been taken into account and the correlations between the frequencies and the chain conformations have been analysed theoretically (Miyazawa, 1960*b*). Furthermore, certain new correlations have been found, allowing one to distinguish between the α-helical conformation, the parallel-chain extended conformation, the antiparallel-chain extended conformation, and the random coil form (Miyazawa and Blout, 1961). More recently the perturbation treatment has been refined and applied to polyamides (Elliott, private communication) and to polyglycine II and feather keratin (Krimm, private communication). The frequencies and relative intensities of the amide I and II bands of polypeptides in various conformations are summarized in Table I.

For the α-helical conformation, where there are 3·6 peptide groups per helical turn, two vibrations are active in the infra-red absorption. The $\nu(0)$ vibrations, for which the phase difference (δ) between the adjacent groups in the chain is equal to zero, give rise to parallel bands, whereas

the $\nu(\theta)$ vibrations for which the phase difference is equal to $\theta = 2\pi/3{\cdot}6$ give rise to perpendicular bands. For the amide I vibration, the parallel and perpendicular components overlap each other while for the amide II vibration the two components are split by about 30 cm^{-1}. The α-helical conformation may be characterized by the strong parallel amide I band at about 1650 cm^{-1} and the strong perpendicular amide II band at about 1550 cm^{-1}. These frequencies, however, are not much different from the corresponding frequencies of the random coil form. For some cases, where

TABLE I. The frequencies (cm^{-1}) and relative intensities† of the amide I and II bands of polypeptides in various conformations (Miyazawa and Blout, 1961)

Conformation	Designation	Amide I	Amide II
Random coil‡	—	1655 (s)	1535 (s)
α-Helix§	$\nu(0)$	1650 (s)	1516 (w)
	$\nu(\theta)$	1652 (m)	1546 (s)
Parallel-chain β‖	$\nu(0,0)$	1645 (w)	1530 (s)
	$\nu(\pi,0)$	1630 (s)	1550 (m)
Antiparallel-chain β¶	$\nu(0,\pi)$	1685 (w)	1530 (s)
	$\nu(\pi,0)$	1632 (s)	1540 ††
	$\nu(\pi,\pi)$	1668 ††	1550 (w)

† Relative intensities are given in parentheses: (s) = strong; (m) = medium; (w) = weak.
‡ Poly-serine.
§ Poly-γ-benzyl-L-glutamate.
‖ β-Keratin.
¶ Polyglycine I.
†† Calculated value.

the sample of polypeptide in question may be oriented, the α-helical form may be identified by the weak parallel amide II band at about 1520 cm^{-1}. The inclination of the amide I or II transition moment from the helix axis has been estimated from the intensity ratios of the parallel and perpendicular components (Miyazawa and Blout, 1961).

For the parallel-chain β-conformation as well as for the antiparallel-chain β-conformation, the infra-red active vibrations are designated as $\nu(\delta,\delta')$, where δ is the phase difference between the adjacent group in the chain and δ' is the phase difference between the adjacent group through intergroup hydrogen bonds. For the parallel-chain β-conformation, the $\nu(0,0)$ vibrations give rise to parallel bands, whereas the $\nu(\pi,0)$ vibrations

give rise to perpendicular bands. For the antiparallel-chain β-conformation, the $\nu(0,\pi)$ vibrations give rise to parallel bands, whereas both the $\nu(\pi,0)$ and $\nu(\pi,\pi)$ vibrations give rise to perpendicular bands. The β-conformation may be characterized by the strong perpendicular amide I band at about 1630 cm^{-1}, and the strong parallel amide II band at about 1530 cm^{-1}. The antiparallel-chain β-conformation may be identified by the presence of a weak but well-defined amide I band at about 1690 cm^{-1} (parallel dichroism).

The correlations between the chain conformations and the frequencies and intensities of the amide I or II bands apply satisfactorily to polypeptides consisting only of single amino-acid residues. If, however, several chain conformations co-exist as in the case of co-polymers of various amino acids or in the case of proteins, the fractions of various conformations may not necessarily be estimated by the measurements only of the amide I and II bands. In view of these situations, characteristic amide bands in other frequency regions have been studied. For a series of monosubstituted amides three characteristic amide bands (IV, V, and VI) have been observed in the region 800–500 cm^{-1} (Miyazawa *et al.*, 1956). These bands arise from the C=O in-plane bending, N—H out-of-plane bending, and C=O out-of-plane bending modes, and accordingly are considered to be more sensitive to the change in the chain conformation than are the amide I or II bands. Therefore the infra-red spectra of a variety of polypeptides and their *N*-deuterated species have been measured in the region 800–400 cm^{-1} (Miyazawa, 1962; Miyazawa *et al.*, 1962). For example, a strong perpendicular band has been observed at 620 cm^{-1} for poly-γ-methyl-L-glutamate in the α-conformation and this band is shifted to 465 cm^{-1} (perpendicular dichroism) on *N*-deuteration. Accordingly this band is considered to arise from the N—H out-of-plane bending mode (the amide V). The frequencies of the amide V band of the α-conformation have been located in a narrow region of 610–620 cm^{-1} for a variety of polypeptides. For elephant hair the perpendicular amide V band has been observed at about 600 cm^{-1} (Beer *et al.*, 1959), indicating that this protein is made up primarily of the α-helical conformation.

The amide V band of the antiparallel-chain β-conformation is observed at about 700 cm^{-1}, which is higher than the corresponding frequency of the α-form by almost 100 cm^{-1}. For example, a strong band is observed at 700 cm^{-1} for poly-γ-methyl-L-glutamate in the β-form, and this band is shifted to 530 cm^{-1} on *N*-deuteration. The amide V band of the β-form is also located in a narrow frequency region of 695–705 cm^{-1} for polypeptides so far studied. For the random coil form the amide V band is appreciably broad and the band centre is located at about 650 cm^{-1} (for example, sodium poly-L-glutamate and poly-L-lysine hydrochloride).

As remarked before, the amide I and II frequencies of the α-conformation are not much different from the corresponding frequencies of the random coil form. Therefore if the α-helical form and the random coil form co-exist, the fractions of these two forms may not necessarily be determined from the intensity measurements of the amide I or II bands. In such cases, the measurements of the amide V bands appear to be more useful, since the amide V band of the α-helical form is sufficiently lower than the corresponding frequencies of the random coil form. For example two amide V peaks have been observed at 665 cm^{-1} and 630 cm^{-1} for a solid film of poly-DL-alanine; these peaks may be ascribed to the random coil and the α-helical form, respectively. The presence of a certain amount of the α-helical form has also been indicated by the ultra-violet hypochromic effect (Imahori and Tanaka, 1959; Doty and Gratzer, 1962).

For certain proteins, the right-handed α-helix and the left-handed α-helix are considered to co-exist (e.g. insulin, Narita *et al.*, 1961). The difference between the fractions of the right-handed helix (x_{H+}) and the left-handed helix (x_{H-}) may be determined by the measurements of the optical rotatory dispersion and the sum of these fractions has been estimated from the decay rate of the amide II bands in deuterium oxide solutions (Lenormant and Blout, 1953, 1954; Blout *et al.*, 1961). The sum of the fractions of the right-handed and left-handed helices may also be estimated even for the solid samples of polypeptides if the intensity of the amide V band at 620 cm^{-1} is properly measured. In our previous study a series of the co-polymers of methyl-L-glutamate and methyl-D-glutamate have been treated (Masuda and Miyazawa, unpublished). Since the thicknesses of the solid films of these co-polymers were not easily set, the relative intensities of the amide V band at 620 cm^{-1} and the C—H stretching band at about 2950 cm^{-1} were measured. If the assumption may be made that the absorption coefficient of the C—H stretching band does not change appreciably with the L:D ratios of the co-polymers, the fraction of the α-helical form for the co-polymers may be estimated with reference to the corresponding ratio as measured for polymethyl-L-glutamate, which exists only in the α-helical form. The fractions of the α-helix thus estimated for various co-polymers are shown in Table II. It may be seen that the fraction of the α-helix ($x_{H+}+x_{H-}$) decreases as the L:D ratio changes from 1:0 to 0·5:0·5. It should be remarked, however, that the fraction of the α-helix does not reduce to zero even for the DL co-polymer. This indicates the profound tendency towards the formation of sequences of like residues (Wada, 1961). In Table II, are also listed the difference ($x_{H+}-x_{H-}$) between the fractions of the right-handed and left-handed helices as determined from the measurements of the optical rotatory dispersion in chloroform solution. If an additional assumption is made that the fractions of the

left-handed helix, right-handed helix and random coil form do not change on passing from the solid state to the solution, each of these fractions may also be estimated. Although the assumption made here

TABLE II. The fractions $(x_{H^+} + x_{H^-})$ and $(x_{H^+} - x_{H^-})$ of the co-polymers of methyl-L-glutamate and methyl-D-glutamate as estimated from the intensities of the 620 cm^{-1} band (solid) and from the optical rotatory dispersion (solution), respectively (Masuda and Miyazawa, unpublished)

Ratio L:D	$x_{H^+} + x_{H^-}$	$x_{H^+} - x_{H^-}$
1·00:0·00	1·00	1·00
0·95:0·05	0·95	0·93
0·90:0·10	0·89	0·82
0·85:0·15	0·88	0·79
0·80:0·20	0·92	0·75
0·75:0·25	0·88	0·68
0·70:0·30	0·79	0·58
0·65:0·35	0·77	0·51
0·60:0·40	0·66	0·38
0·55:0·45	0·73	0·18
0·50:0·50	0·78	—0·01

should be examined and refined if necessary, the treatment may well turn out to be a useful new method for more detailed structure analyses of polypeptides and proteins by means of infra-red spectroscopy.

REFERENCES

Ambrose, E. J. and Elliott, A. (1951). *Proc. roy. Soc.* A **205**, 47.

Badger, R. M. and Pullin, A. D. E. (1954). *J. chem. Phys.* **22**, 1142.

Beer, M., Sutherland, G. B. B. M., Tanner, K. N. and Wood, D. L. (1959). *Proc. roy. Soc.* A **249**, 147.

Blout, E. R., deLozé, C. and Asadourian, A. (1961). *J. Amer. chem. Soc.* **83**, 1895.

Cannon, C. G. (1960). *Spectrochim. Acta* **16**, 302.

Doty, P. and Gratzer, W. B. (1962). In *Polyamino Acids, Polypeptides and Proteins* (M. Stahmann, ed.), p. 117. The University of Wisconsin Press, Madison.

Doty, P., Wada, A., Yang, J. T. and Blout, E. R. (1957). *J. Polym. Sci.* **23**, 851.

Elliott, A. (1954). *Proc. roy. Soc.* A **226**, 408.

Fraser, R. D. B. and Price, W. C. (1952). *Nature, Lond.* **170**, 490.

Imahori, K. and Tanaka, J. (1959). *J. mol. Biol.* **1**, 359.

Lenormant, H. and Blout, E. R. (1953). *Nature, Lond.* **172**, 770.

Lenormant, H. and Blout, E. R. (1954). *Bull. Soc. chim. Fr.* 859.

Miyazawa, T. (1960*a*). *J. mol. Spect.* **4**, 168.
Miyazawa, T. (1960*b*). *J. chem. Phys.* **32**, 1647.
Miyazawa, T. (1962). In *Polyamino Acids, Polypeptides and Proteins* (M. Stahmann, ed.), p. 201. The University of Wisconsin Press, Madison.
Miyazawa, T. and Blout, E. R. (1961). *J. Amer. chem. Soc.* **83**, 712.
Miyazawa, T., Shimanouchi, T. and Mizushima, S. (1956). *J. chem. Phys.* **24**, 408.
Miyazawa, T., Shimanouchi, T. and Mizushima, S. (1958). *J. chem. Phys.* **29**, 611.
Miyazawa, T., Masuda, Y. and Fukushima, K. (1962). *J. Polym. Sci.* (in press).
Narita, K., Kakutani, Y. and Imahori, K. (1961). The Fifth International Congress of Biochemistry, Moscow.
Richards, R. E. and Thompson, H. W. (1947). *J. chem. Soc.* 1248.
Sutherland, G. B. B. M. (1952). In *Advances in Protein Chemistry* (M. L. Anson and J. T. Edsall, eds.), Vol. 7, p. 304. Academic Press, New York.
Wada, A. (1961). *J. mol. Biol.* **3**, 507.

Discussion

A. ELLIOTT (King's College, London): Dr. Miyazawa has referred to the greater proportional changes in frequency with conformation which are seen in the amide V band (700–600 cm^{-1}) compared with the amide I and II bands of polypeptides. It seems likely, however, that the amide V band may be proportionally more sensitive to other changes also (such as changes in the chemical nature of the side chains). It is these other changes which sometimes obscure the effects of chain conformation on frequency. The amide II band is, in our experience, rather sensitive. We have found that ν_0 (the frequency when interaction effects are eliminated) is not independent of conformation in this band.

T. MIYAZAWA: Actually, for a variety of polypeptides, the amide V frequencies, have been located in the narrow regions of 610–620 cm^{-1} and 695–705 cm^{-1} for the α-helical form and the β-form, respectively. Also, I would like to add that the two crystal forms, α and γ, of polyamides may be distinguished from each other by the amide VII bands lying in the region 350–250 cm^{-1}.

G. N. RAMACHANDRAN: In trying to explain the variation in the infra-red frequency in an actual protein or polypeptide, the effect cannot be taken to be due only to the configuration alone. The hydrogen bonds and their strengths would also play an important part.

N. S. ANDREEVA: The normal vibrations of peptide groups containing tertiary nitrogen (connected with imino acid residues) have to be different from the normal vibrations of peptide groups in *trans* configuration. Dr. Yu. N. Chirgadze has done special theoretical work on the normal vibrations of peptide groups in various configurations. He has shown that amide II bands for a peptide group with tertiary nitrogen have to appear near 1450 cm^{-1}. Investigations of various special model compounds are in very close agreement with this result. These data can be employed for investigations of collagen and related compounds. (Chirgadze, Yu. N. (1962). *Biofizika* **7**, 382, 523.)

T. MIYAZAWA: The 1450 cm^{-1} bands of disubstituted amides lie exactly in the same region as the CH_2 bending bands.

SECTION IV
Electron Microscopic Studies

The Electron Microscope in Research on Proteins

RALPH W. G. WYCKOFF

Department of Physics, University of Arizona, Tucson, Arizona, U.S.A.

ABSTRACT

The fact that it is no longer difficult to see with the electron microscope particles the size of the molecules of proteins, nucleic acids and other biologically important substances gives this instrument a necessary place in their study. It enables us to see how these large molecules are arranged in the crystalline and paracrystalline solids they form in nature and in the laboratory and it often shows something of their internal structure. In this paper it is proposed to indicate what the electron microscope can now do and how the information it yields supplements that derived from other sources.

During the last generation research on proteins and the other essential macromolecular components of living matter has developed in a truly fantastic way. We easily forget that forty years ago scarcely a dozen were known in the pure state and these were not very typical of the innumerable molecular species which have now been isolated. In those days protein was almost synonymous with protoplasm, the jelly-like substrate of all life. As biochemistry developed, the new methods it evolved led to the isolation and crystallization of more and more of these extraordinarily complex and varied substances at the same time that it showed them all to be built up of a few simple amino acids. With the ultracentrifuge came the then-surprising demonstration that each of these innumerable proteins was composed of definite molecular entities, hundreds or thousands of times more massive than the familiar molecules of synthetic organic chemistry but just as uniform from one molecule to its neighbours as are the molecules of simple compounds.

When this point had been reached it became feasible to begin to move from the vague question of what a protein is to more precise questions about its molecular architecture. At the outset such of these questions as were answerable dealt with the sizes and shapes of the molecules of various proteins and with their contents of water. Proteins could not then, and still cannot, be synthesized and their molecular structures established through the conventional methods of organic chemistry. Nevertheless as enzyme chemistry has grown over the last generation, it has provided tools for breaking up the molecules of a protein in a wide

variety of controlled ways, and the invention of chromatographic analysis has given a means of identifying the molecular fragments thus produced. In this way much has already been learned about how various amino acids are associated together within the molecule of a protein. What this type of analysis cannot tell is the spatial distribution of these molecular components, and hence how they are knit together to form the comparatively rigid atomic networks that distinguish a protein molecule from its polypeptide fragments. X-Ray diffraction has been so vital to the further development of our knowledge of proteins because it supplies the means to answer this question of spatial distribution, to tell where atoms are within the space occupied by a protein molecule. You are hearing from Sir Lawrence Bragg and others about the truly marvellous insight already gained into the molecules of haemoglobin and myoglobin and about our increasing understanding of the atomic architecture of the molecules of other proteins.

Though X-ray methods thus hold the possibility of showing in previously unimaginable detail the structure of many proteins essential to life, they can never provide all the information needed for a deeper understanding of vital processes. Life proceeds through an elaborate series of chemical reactions involving protein and other macromolecular substances, and the orderly control of these essential reactions depends both on the substances themselves and on the way they are distributed within the cells that are the functioning units of vital activity. To understand the mechanism of life we must know both what the molecules are and where they are within the semi-solid framework of the cells and tissues of which they form the active ingredients. X-Ray diffraction can tell us how the molecules are distributed in a crystal as well as how the atoms are arranged in such crystallizable molecules; but a cell contains many different kinds of molecules, and they are very rarely arranged in an order that is crystalline. Furthermore, many of the most important of these do not crystallize, even after they have been extracted and purified. The electron microscope has established itself as an indispensable tool in protein research because it can render these molecules visible as individual entities and hence can show where they are within the cell, irrespective of whether or not there is order in their distribution. In the present paper it is proposed to discuss the application of this instrument to the study of protein and other macromolecular substances, to indicate the sort of results already obtained and the kind of knowledge to be expected from its continuing application.

The experimental knowledge we now have of the fine structure of matter has been gained in two fundamentally different ways. In one we measure some property which theory indicates is sensitive to its molecular or atomic fine-structure and then imagine or deduce a structural

design that will explain the measured property. This is what we do when determining with X-rays the molecular and atomic arrangement in a protein crystal. The success of this indirect investigation of fine structure depends on many factors, including the richness and sensitivity of the data, the adequacy of the theory and sometimes that of our imagination. In the other approach to fine structure we seek to "see" directly what is thereby causing the matter to interact with a form of radiation which can then be imaged to show the details of the interaction and thus, as our sensory experience has led us to expect, of the structure itself. This is the way of microscopy.

The theory underlying optical microscopy is very adequately known. It shows that an inescapable relation exists between the light we use and the smallest size of detail that can be imaged. We cannot hope to see things smaller than half the wavelength of the light employed and this lower limit of vision can only be reached with lenses that are substantially perfect. Even the biggest macromolecules are less than 1000 Å in diameter, and hence forever beyond the reach of optical microscopes. Electron microscopes and the ability to see below this optical limit have resulted from the realization that definite wavelengths are associated with particulate as well as electromagnetic radiation. The electrons in these microscopes have wavelengths a hundred thousand times shorter than those of light and the limit to their resolving power, if electron lenses were perfect, would be but a small fraction of an atomic diameter. Existing microscopes are very far from attaining this goal but they are able to image particles with diameters of about 5 Å—and even the smallest protein molecules are bigger than this. It is evident that whatever difficulties may be experienced in visualizing protein molecules they do not arise from trying to see objects below the resolving power of the instruments at our command.

We sometimes fail to realize how essential microscopy has been to the growth, first of biology and then of biochemistry. The morphological information which only the optical microscope can give is the framework necessary for understanding the significance of reactions with which biochemistry deals. Without this instrument nothing would be known of the cells that are the units of life and of their organization into tissues composing the organs of all but the simplest of living forms; inside the cells themselves it has revealed their nuclei and other organized structures that are the sites of so many of the chemical reactions essential to life. But these structures, such as chromosomes and mitochondria, are so small that few details of their organization are visible with light; it has become a primary task of the electron microscope to penetrate the world of these sub-cellular structures. With the demonstration that the macromolecules of living matter lie within the range of this new instrument, the

ultimate objective of its application to living matter becomes the extended description of the fine-structure of cells and tissues in terms of these molecules. This is a goal towards which we are already aiming though it can scarcely be approached short of a generation of work. It is interesting to see in what directions and with what success we are proceeding.

The first successful application of the electron microscope was in the visualization of the macromolecules in pure preparations of virus and other proteins. We now know by direct observation the sizes and shapes of the macromolecular particles of many viruses, and of the molecules of numerous proteins. This information has been gained in two ways. It can be obtained by examining isolated molecules deposited from dilute solution on to a very thin support, or by studying the solids they form. If the molecule is more or less spherical, the solid is likely to be crystalline and we can in favourable instances see not merely the molecules themselves but also how they are arranged to produce a single crystal. If the molecule is filamentous, as in collagen, we can ascertain how these filaments are aligned in the distinctive semi-crystalline solids they form. Since these filamentous solids constitute the frameworks that support the more fluid cells of all the higher forms of life and provide in muscle the device by which animals move from place to place, such a knowledge of their molecular composition is essential to a thorough understanding of the functioning of the higher forms of life.

The principal difficulties that arise in portraying protein molecules are the inevitable consequence of the small amounts of matter they contain. We see matter in an electron microscope by reason of the electrons it scatters and this scattering is minimal for single biological molecules composed of light elements. To examine them these molecules must rest on a substrate that even when very thin will scatter more electrons than the molecules it supports. As a result they are hard to distinguish, not because they are too small to be resolved by the microscope, but because they do not contribute enough to the total scattering to create the necessary contrast in the image that is formed. This is the situation that prevailed before shadowing was introduced and it continues to set a limit to the smallness of the molecules that can be seen when for one reason or another shadowing cannot be employed. Nearly twenty years ago we discovered that the contrast required to bring out molecular outlines and surface details could be created by covering a preparation with an obliquely deposited layer of metal only a few atoms thick. The use of metal shadowing is largely responsible for the extensive knowledge we now have of the sizes and shapes of the elementary particles of viruses and many proteins. There are, however, certain limitations to the method which must be borne in mind in interpreting

the pictures the electron microscope gives. It was soon noticed that the visible detail was particularly coarse when a relatively low-melting metal such as gold was used and that it became even coarser under the electron bombardment of the microscope. This was due to crystallization and recrystallization of the evaporated metal. The false detail thus created can be much reduced by shadowing with high-melting metals like platinum, or with compounds like tungsten oxide that evaporate but do not crystallize readily on deposition or in the microscope. No shadowing material seems structureless below about 20 Å, however, and hence the technique must be used with great caution for particles and details smaller than this. In spite of such shortcomings shadowing remains the best way to examine isolated molecular particles and it faithfully reveals the molecules of proteins and nucleic acids if the preparations are sufficiently pure and the substrate thin and smooth.

Information that the electron microscope can give is not restricted to the external form of individual macromolecular particles. Many years ago it was shown through studies with the ultracentrifuge that as the pH is altered the weight of haemocyanin molecules in solution varies by integral multiples of a sub-unit. The electron microscope revealed these units and the way they associate to form larger molecules. We are now seeing that many virus particles as well as protein molecules are similar aggregates of ordered sub-units.

Obviously shadowing can give information about internal structure only if, as with haemocyanin molecules, this is reflected in the configuration of the surface of the particle. The carefully dehydrated elementary particles of many virus proteins have a polyhedral shape in shadowed preparations; but the negative staining that involves treatment with strong phosphotungstic acid often brings out better this shape and the definite sub-units whose ordered distribution is responsible for it. In this way, for example, one can see the icosahedral distribution of the sub-units in herpes and adenovirus particles and what may be the spiralling nucleoprotein threads packed within the elementary particles of influenza-like viruses. Sometimes when, as is the case with the liver protein ferratin, the molecule contains heavy atoms no such treatment is necessary to bring out its ordered sub-structure. The increasing ability of chemists to incorporate metallic atoms into protein molecules should allow us to add considerably to our knowledge of internal molecular structure by using this visibility of heavy metal sites. The staining with uranium and lead salts that has already given much information about ordered structures within virus particles is a technique of this general sort.

A further insight into the sub-units of macromolecular particles can sometimes be gained by disrupting them on a membrane and looking

at the liberated fragments. Thus the fine threads lying around the debris of a bacteriophage particle collapsed by osmotic shock are undoubtedly its liberated nucleic acid molecules. In the case of the smaller bacteriophages such as X-174, these threads may be many phage-particle diameters long. We can measure their dimensions but how they were intertwined and coiled within the intact particle is something the electron microscope has not yet told us.

It is important to realize that in spite of its successes the electron microscope gives only fragmentary data about the internal structure of protein macromolecules, and these must be evaluated in view of the ease with which proteins and their molecules can be altered during preparation. Certainly they are profoundly modified by the vigorous chemicals used in staining: we probably see so many details of their internal structure precisely because of these changes and of those that occur when the large amounts of water they contain are abstracted.

What the electron microscope can reveal about how the protein and other biological macromolecules are arranged in the solids they form is at least as valuable as the knowledge it gives about them individually. Because of the great power of X-ray methods as applied to crystals it is probable that to many the aesthetic pleasure in seeing the ordered arrangement of the molecules in a protein crystal outweighs the scientific merits of this visualization; nevertheless the microscope can often supply information beyond the reach of diffraction. The crystals we see packed with molecules commonly have the same characteristic outlines as before desiccation. This demonstration that our methods of specimen preparation have not drastically altered either the dimensions or the spatial distribution of these molecules is an added assurance that electron microscopy is a valid way to establish molecular shapes.

Several techniques have now been devised for displaying with the electron microscope the way the molecules are arranged in a protein crystal. The one first used more than 15 years ago involves making a shadowed replica of the crystalline surface. If the crystals employed are small enough so that their individual outlines and recognizable faces and edges are reproduced, the molecular arrangement within them can be deduced from the molecular patterns on their several faces. This has been done in several instances but sometimes the method fails because the molecular framework collapses when the water is removed; more often it has been unsuccessful because the crystalline faces have been covered with adhering foreign molecules that could be dislodged only by dissolving the crystal itself. Crystals with their faces clean enough to reveal their molecular order have been especially difficult to obtain for proteins such as haemoglobin and the albumins whose ready solubility in water makes it necessary to crystallize them from high concentrations

of salt. Crystals of various plant virus proteins, of the poliomyelitis virus, and of proteins such as haemoglobin and a number of enzymes have been portrayed in this fashion.

The molecular arrangement has been seen within crystals that have been thinly sectioned after embedding according to procedures employed for the electron microscopy of tissues. In most cases these crystals have been inclusions of virus particles within cells that had produced them in extraordinary profusion, but extracellular crystals can be similarly prepared for microscopy. In such slices through crystals, the molecules themselves are sectioned, as is immediately apparent from their different diameters, and in general they will overlie one another in the section. This, combined with the fact that it is rarely possible to control the orientation of the crystals with respect to the cut, makes it hard to deduce with certainty how the particles are arranged. Nevertheless if enough crystalline sections are observed, this technique can supply useful information about molecular configuration and order.

A third way of visualizing molecular arrangement is through the direct examination of sufficiently thin crystals. This can be successful only when the molecules themselves are small enough so that the minimal stacking required for crystallinity is not too thick for electron microscopy. Few, if any, proteins meet this requirement but numerous synthetic organic compounds do so, and it is these that have thus far been studied by this technique. The first to be examined were the phthalocyanins which have large plate-like molecules less than 10 Å thick. When suitably oriented for electron microscopy, the image of one of their crystals is covered with parallel stripes whose constant distance apart is determined by the molecular separations. Though we may talk naïvely about "seeing" the flat molecules of these compounds, the image is in fact something between a direct visualization and an electron diffraction pattern. The molecular, and even the atomic, order in very thin crystals is also reflected in another kind of striated image that arises when the overlying layers are not quite in register. The distances between the striae in these moiré patterns are much greater than the crystalline spacings and are not dependent on them alone. These striae have been observed from very diverse substances—metals, graphite, metallic oxides and silicates as well as organic compounds of large molecular weight. They may be encountered and prove useful in the examination of protein crystals.

Striated patterns that resemble those described above frequently cover the images of polyhedral bodies included in sections through virus-diseased insects. These bodies are crystals of proteins of rather low molecular weight produced during the course of the virus infection. Too little work has yet been done to give a complete interpretation of these

striae and to show if they can, under favourable conditions, be resolved into molecular images. Similar striae may be expected from other thinly sectioned protein crystals and they may supply useful information about molecular distribution.

Most of the proteins that crystallize well have molecules that are more or less spherical or polyhedral in shape. Proteins with filamentous molecules are equally numerous and the solids they form have roles in life processes which are far more important than those played by protein crystals. These fibrous solids constitute the basic frameworks of the higher organisms. Though their elongated molecules are not usually in crystalline form, they are in ordered arrays and one of the great contributions of the electron microscope has been to make this order visible. The comparative values of the microscope and of X-ray diffraction are very different when dealing with these fibrous solids and with crystals. Most of that which the electron microscope can show about molecular order in a protein crystal would be learned early in a complete X-ray determination of its structure. The different situation that prevails with fibrous solids is apparent if we compare, for instance, the wealth of new information about collagen already supplied by the electron microscope with the limited rewards from years spent in careful investigation of its X-ray diffractions. Collagen is the basis of the connective tissue that ties our cells together to form organs and binds these to one another and to our bones to create a functioning whole. We obviously need to know all we can about this essential protein and the solids it produces. The electron microscope has revealed the ordered macromolecular arrangement in connective tissue from many animals and the different kinds of molecular order that exist in collagenous solids formed under various physico-chemical conditions. Its direct visualization of how the molecular units are arranged in these several forms of collagen is at once a way to study the submolecular structure of these units, to understand the conditions under which connective tissues are laid down in the animal body and perhaps to learn more about the diseases involving these tissues. Other fibrous solids essential to life have filamentous molecules not protein in composition. Conspicuous amongst them is, of course, the cellulose that is the framework of plants. The electron microscope has already shown how it is uniquely able to manifest the various ordered and disordered ways in which the elementary fibrils of cellulose are formed and arranged through the activity of the living plant.

Not all filamentous molecules occur as fibrous solids in living organisms. Some, like the nucleic acids, function singly or in combination with other macromolecules. They can be seen individually in purified preparations by the same techniques used for visualizing the molecules of globular proteins. These exceedingly thin and indefinitely long molecular

threads have been repeatedly photographed and it is clear that what can be learned about them depends above all else on the ability to make preparations in which they appear unaltered from their native condition.

The most highly ordered, non-crystalline, protein structure with which we are familiar is striated muscle. Since it was first shown more than a dozen years ago to consist of closely-packed orderly stackings of filamentous protein molecules, the microscope has given an extraordinary insight into its molecular structure and what happens when it contracts. This instrument has revealed how more than one kind of filamentous molecule participates in producing the order we observe and has made an essential contribution to the molecular understanding we now have of how muscle functions.

The greatest problem that now presents itself to the person wishing to use our ability to see macromolecules for the interpretation of living processes is one of recognizing these entities when he sees them in cells and tissues. This is not too difficult when there is a characteristic inner order in the structures they form, as with muscle and most forms of connective tissue. Collagen is easily recognized wherever it occurs, by its striae repeated every 220 Å or 650 Å and, to take another example, there is no difficulty in identifying the myelin enveloping a transversely cut nerve by the constant repetition of its layers about 80 Å apart. The chloroplasts of plants provide another repetitive structure it would be hard to miss. Ordered molecular structures have already been observed in many specialized tissues such as the eye and the kidney; this order is unquestionably related to their characteristic functions and its elucidation may be expected to furnish the same kind of understanding which is arising through the electron microscopy of muscle.

The various organelles common to most cells—nucleoli, mitochondria, microsomes of one sort or another, chromosomes, etc.—are examples of structures which have a molecular order, not necessarily repetitive, that becomes more accessible to observation as methods of tissue preparation are improved. We may hope with time to recognize more and more of the nucleic acids, carbohydrates and lipoids, as well as proteins, which constitute these structures. Somewhat different problems of identification are presented by the many macromolecular substances that occur in tissues without forming fixed parts of their structures. Some of their molecules are large and of a distinctive appearance that makes them easy to recognize wherever they may be found. This is true for many of the products, infectious and otherwise, of viral diseases and of such a protein as ferratin containing its identifying iron. Most, however, are distributed widely within a protoplasmic net from which we would not be able to distinguish them even after having studied them in purified solution. The obvious way to approach this very fundamental problem of molecular

identification in tissues is through some kind of "staining" which will result in characteristic reaction products identifiable under the microscope. Such a "staining" using heavy atoms has been referred to earlier in this paper but it is to be hoped that additional ways of molecular labelling will be discovered. The amount of meaning we can give to the mass of detail now visible in electron micrographs is clearly dependent on the success of this search.

Whatever the future may hold in this respect it is by now clear that the visualization the electron microscope gives to objects of macromolecular size has made this instrument an essential tool in research on proteins. As the preceding discussion has indicated, it is the most direct way imaginable to measure the dimensions of their molecules and it often shows us something of their internal structure. At the same time it is supplying unique information about how the molecules are arranged in the solids that proteins and other macromolecular substances form, both in nature and in the laboratory.

Bibliographical Note

In a subject that is developing as rapidly as is the electron microscopy of macromolecular particles, an adequate bibliography is far too voluminous to be attached to such a general article as the present one. Those who are interested will find a record of much of the recent work in published proceedings of the international congresses on electron microscopy:

The Proceedings of the Third International Conference on Electron Microscopy, London, 1954 (Royal Microscopical Society, London, 1956).

Electron Microscopy, Proceedings of the Stockholm Conference, September 1956 (Academic Press, New York).

Proceedings of the Fourth International Conference on Electron Microscopy, Berlin, 1958 (Springer-Verlag, Berlin, 1960).

Proceedings of the European Regional Conference on Electron Microscopy, Delft, 1960, 2 volumes (Netherlands Society for Electron Microscopy, Delft).

Electron Microscopy, Proceedings of the Fifth International Congress for Electron Microscopy, Philadelphia, 1962, 2 volumes, (Academic Press, New York).

The rest of the recent literature is covered in the extensive bibliography for the years 1956–61 published this year in:

The International Bibliography of Electron Microscopy, Vol. II, 1956–61 (New York Society of Electron Microscopists, New York, 1962).

The beginnings of the subject can be found by reference to the earlier issues of this bibliography or to such books as:

Electron Optics and the Electron Microscope, by V. K. Zworykin *et al.* (Wiley, New York, 1945).

Electron Microscopy, Technique and Applications, by Ralph W. G. Wyckoff (Interscience, New York, 1949).

Introduction to Electron Microscopy, by C. E. Hall (McGraw-Hill, New York, 1953).

The World of the Electron Microscope, by Ralph W. G. Wyckoff (Yale University Press, New Haven, 1958).

Investigations of the Renaturation of Tropocollagen

K. KÜHN

Anorganisch-Chemisches Institut der Universität, Heidelberg, Germany

ABSTRACT

Long-spacing segments obtained from thermally denatured and renatured collagen solutions can be separated by reprecipitation from a non-striated fraction. Tropocollagen molecules, which are incompletely renatured and which do not occur as rigid rods in solution, participate in this formation of completely striated segments. On attachment to preformed segments, these molecules are subjected to an auxiliary orientation under the influence of electrostatic forces.

Treatment of renatured collagen solutions with trypsin effects hydrolysis of incompletely renatured molecules. The trypsin-resistant fraction behaves just like native tropocollagen molecules.

Rod-like trypsin-resistant tropocollagen molecules can be reformed not only from the γ-component, but also from completely separated α- and β-components.

1. INTRODUCTION

The tropocollagen (TC) molecule is a comparatively rigid rod with a length of about 3000 Å, a cross-section of about 14 Å, and a molecular weight of 360,000 (Boedtker and Doty, 1956). It consists of three peptide helices, which are intertwined to form the so-called triple helix, being held together by hydrogen bridges (Ramachandran and Kartha, 1954, 1955*a*, *b*; Rich and Crick, 1955; Ramachandran and Sasisekharan, 1961).

In part, the three individual peptide chains in the triple helix are cross-linked together intramolecularly. On denaturation by heating to 40°C in acid solution or by treatment with 6 M urea or potassium thiocyanate, neutral salt-soluble collagen (the primary precursor of mature insoluble collagen) is degraded into three peptide chains of equal molecular weights, the α-components (Orekhovich *et al.*, 1960). If acid-soluble collagen (the second precursor) is subjected to similar treatment, it dissociates into an α- and a β-component. The latter has a molecular weight which is twice that of the α-component (Orekhovich and Shpikiter, 1958) and appears to be comprised of two peptide chains which are held together by an alkali-labile bond (Doty and Nishihara, 1958). In addition, minute amounts of a third entity, the γ-component, occur; this appears to consist of three interlinked α-molecules (Grassmann *et al.*, 1961).

Thermal denaturation of TC in citrate buffer at pH 3·7 proceeds in two stages (Engel, 1962*a*). First, in a rapid process, the ordered asymmetric structure of the molecule disappears. At 38°C, physical properties such as optical rotation, viscosity, and the initial gradient of the light scattering curve attain limiting values after 11 min. The intermediate formed, in which the individual chains are loosely connected together, dissociates in a second, slower step. Hence, the average molecular weight reaches its limiting value only after 90 min at 38°C.

The alterations in optical activity, viscosity, and molecular weight can be partially reversed (renaturation) by cooling the collagen solutions to low temperatures (von Hippel and Harrington, 1960; Harrington and von Hippel, 1961; Flory and Weaver, 1960; Engel, 1962*a*). Altgelt *et al.* (1961) and Veis *et al.* (1961) discovered that, in the cold, the pure denatured γ-component, in which the three individual peptide chains are held together by thermostable cross-linkages, completely regains its native rod-like molecular shape. On addition of ATP, the molecule reforms regular long-spacing segments (SLS). However, this component forms only 6–12% of the total mixture along with the α- and β-components (Grassmann *et al.*, 1961).

Using electron microscopic techniques, we were able to show (Engel *et al.*, 1962) that collagen which has been denatured to the end of the first stage can be regenerated almost quantitatively by isothermal renaturation at 4°C to give native fibrils and segments. After the first stage of denaturation, the TC molecules are still not broken down into their subunits and hence are just as capable as the γ-component of reforming the collagen molecule. After completion of the second step of denaturation, when the TC is entirely broken down into its subunits, considerably fewer (40%) striated segments and fibrils are produced on renaturation. Nevertheless, the proportion is greater than that of the γ-component in acid-soluble collagen.

This discovery is in contrast to physico-chemical investigations (Engel, 1962*a*), according to which over 90% of the TC molecules in collagen solutions that have been denatured to the second stage do not reassume rod-like structures but remain at a more primitive renaturation stage. This difference raises the question whether native fibrils and segments can be rebuilt from molecules which have an extended conformation but which do not yet possess sufficient rigidity to occur in solution as rods.

In the following discussion, two problems will be examined. First, can incompletely renatured collagen molecules give typical native-collagen patterns in the electron microscope? Second, to what extent can rigid, rod-like collagen molecules be regenerated from totally separated α- and β-components?

2. Materials and Methods

Materials

The collagen (calf skin) was obtained following Gallop's modification (Gallop, 1955) of the method of Orekhovich *et al.* (1948), and purified by a single reprecipitation.

Methods

Denaturation was effected isothermally by maintaining at 38°C for 11 or 90 min, respectively ($c = 0{\cdot}1$ g coll./100 ml).

Renaturation was accomplished either by isothermal renaturation at 4°C for 7 days, or by cooling slowly from 38°C to 4°C over a period of 7 days (the temperature of the solutions was reduced stepwise by steps of 5°C per day) or by temperature fluctuation according to Engel (1962*b*) (the temperature was altered four times from 4°C to 22°C and back in 24-hr periods).

Stepwise thermal denaturation (Figs. 2 and 5). The temperature was increased in steps every 15 min. Analytical measurements were carried out at the middle of each 15-min period for each temperature (Engel, 1962*b*).

Optical rotation. This was measured using a Zeiss photoelectric polarimeter at 405 mμ (0·05° accuracy).

Viscosity measurements were carried out in a capillary viscometer of the Ubbelohde type (temperate constancy $\pm 0{\cdot}1$°C).

Trypsin treatment was effected in 0·5 M calcium chloride solution at pH 8 (adjusted with 0·05 M tris(hydroxymethyl) aminomethane (tris) buffer) and 20°C for 15 hr. Source of trypsin preparation: Trypure Novo from Novo Industry, Mainz, Germany.

The method of preparing the long-spacing segments and fibrils and the electron-microscopic procedures are described in the paper by Engel *et al.* (1962).

3. Results

(a) *The participation of incompletely renatured collagen molecules in the formation of striated long-spacing segments*

Collagen solutions which had been denatured by heating at 38°C for 11 and 90 min respectively, and renatured by cooling slowly to 4°C over a period of 7 days in citrate buffer at pH 3·7, were treated with ATP to precipitate the collagen. The material was reprecipitated several times by dissolving in 0·05% acetic acid, centrifuging off insolubles, and reprecipitating with ATP. This fractionation was continued until exclusively striated segments were deposited from the solutions (Fig. 1).

The purified solutions obtained in this way were then dialysed against citrate buffer and their thermal stability then compared with that of renatured initial solutions and that of native collagen solutions (Engel, 1962*b*). The temperature was increased stepwise every 15 min, and the optical activity measured. The denaturation curves obtained are shown in Fig. 2. Undenatured collagen molecules are stable up to 30°C. Above this temperature, a sharp decrease in the optical activity occurs. All the other solutions exhibit decreases in the optical rotation at even lower

temperatures. This implies that they contain in part incompletely renatured collagen molecules and that these lose their optically active configurations at lower temperatures than native collagen molecules do. The shape of the curves indicates that the fractionated solutions, which deposit striated segments quantitatively, are more like native collagen than the initial unfractionated solutions but that they are by no means made up entirely of perfectly renatured rod-like molecules.

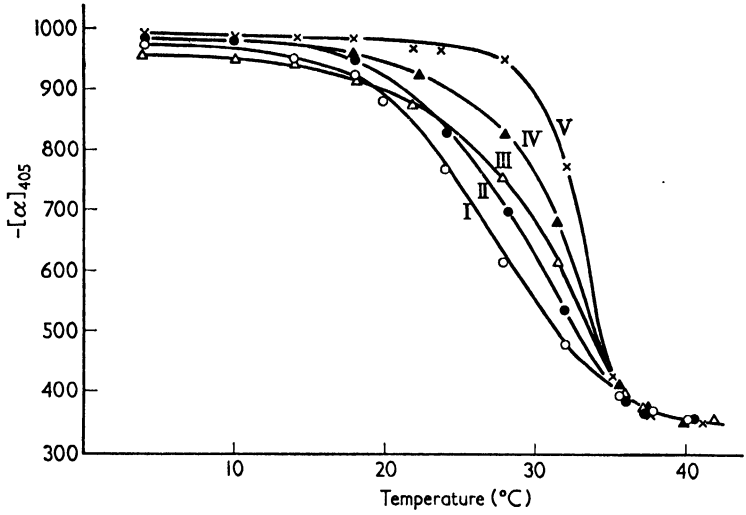


FIG. 2. Stepwise thermal denaturation of various denatured and renatured collagen solutions.

I—Denatured at 38°C for 90 min, renatured by cooling slowly to 4°C over 7 days.

II—Tropocollagen molecules obtained from I by precipitation as long-spacing segments.

III—Denatured at 38°C for 11 min, renatured by cooling slowly to 4°C over 7 days.

IV—Tropocollagen molecules obtained from III by precipitation as long-spacing segments.

V—Native collagen.

Values of the trypsin-resistant tropocollagen content in mg/g total collagen are I = 430, II = 520, III = 670, IV = 790.

Figure 2 also shows the values obtained for the content of collagen molecules renatured to form rigid rod-like molecules as determined by treatment with trypsin (see below). These values constitute a quantitative confirmation of the denaturation curves. It can be seen that in every instance tropocollagen molecules are involved which are incompletely renatured and which probably do not occur in solution as rods. It is conceivable that these entities become attached to preformed

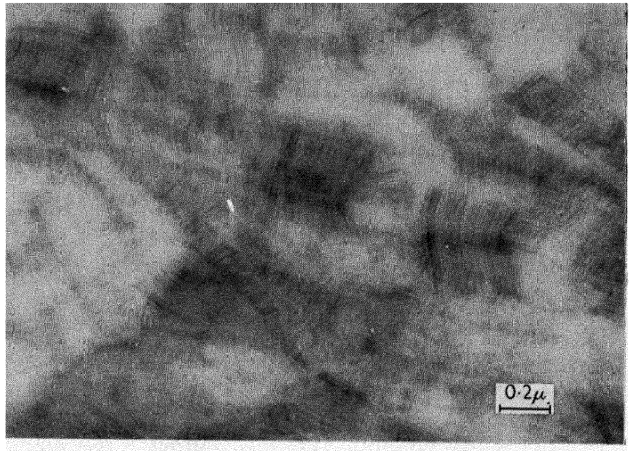


FIG. 1. Long-spacing segments from a collagen solution after denaturation for 90 min at 38°C, renaturation by cooling slowly to 4°C over 7 days, and reprecipitating twice. Stained with phosphotungstic acid and uranyl acetate.

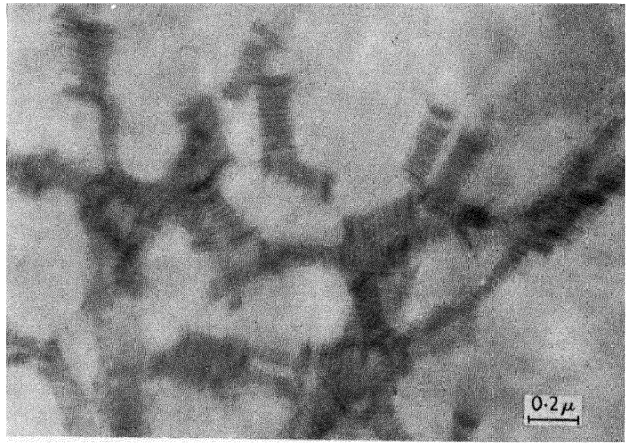


FIG. 3. Long-spacing segment derived from a collagen solution after denaturation for 90 min at 38°C, renaturation by temperature fluctuation, and trypsin treatment (collagen: enzyme = 100:1). Stained with phosphotungstic acid and uranyl acetate.

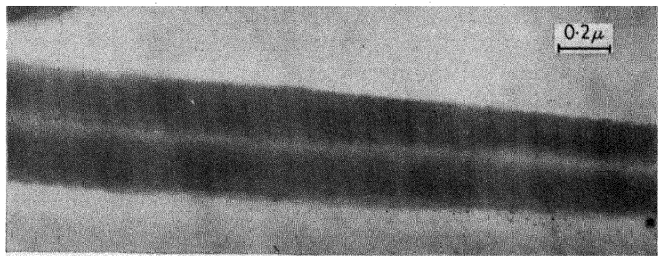


FIG. 4. Native-type fibrils from the same collagen solution as in Fig. 3.

segment sections by means of electrostatic forces and that they become stretched in this way.

(b) *Treatment of renatured collagen with trypsin*

Native collagen molecules are effectively resistant to attack by trypsin at 20°C (Kühn *et al.*, 1961), whereas denatured collagen is degraded rapidly by this proteolytic enzyme. If a collagen solution, which has been denatured to the second stage by keeping at 38°C for 90 min and renatured by altering the temperature according to Engel (1962*b*), is treated with trypsin at 20°C in 0·5 M calcium chloride at pH 8, its optical rotation decreases by a certain amount. This can be taken as a measure of the amount of degraded, i.e. incompletely renatured, material. The variations in optical rotation and the amounts of trypsin resistant tropocollagen molecules calculated therefrom for various enzyme to collagen concentrations are given in Table I.

TABLE I. Variation of optical rotation during trypsin treatment of renatured collagen solutions†

Collagen : enzyme	$-[\alpha]_{405}$ after trypsin treatment	TC trypsin resistant (mg/g total collagen)‡
25:1	612	390
50:1	640	430
100:1	660	470

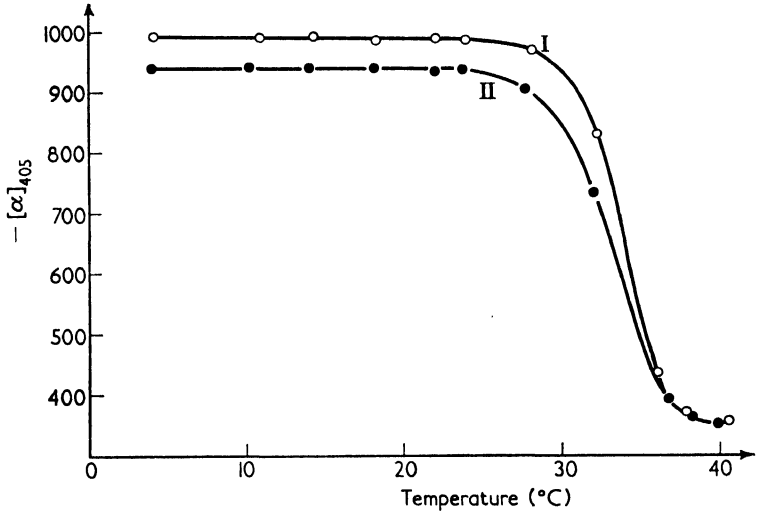
$[\alpha]_{405}$ native $= -1000$, $[\alpha]_{405}$ denatured $= -360$, $[\alpha]_{405}$ renatured $= -894$.

† Denatured isothermally at 38°C for 90 min. Renatured by temperature fluctuation.

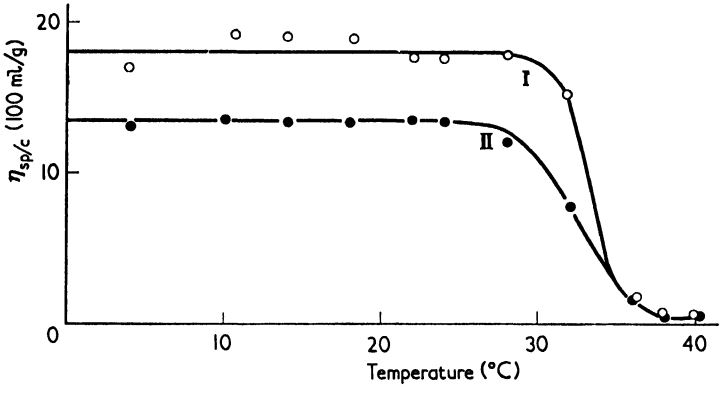
‡ mg TC/g total collagen $= \dfrac{[\alpha]_{405}\text{ trypsin treated} - [\alpha]_{405}\text{ denatured}}{[\alpha]_{405}\text{ native} - [\alpha]_{405}\text{ denatured}} \times 1000.$

Trypsin-resistant tropocollagen molecules behave just like native collagen; they form well-striated segments quantitatively (Fig. 3) and give fibrils typical of the native type (Fig. 4). Their denaturation curves are absolutely identical with those of native collagen molecules (Fig. 5 (a) and (b)).

Since even native collagen molecules are slightly attacked by higher enzyme concentrations, a trypsin : collagen ratio of 1 : 100 is most suitable for determining the degree of renaturation. This concentration effects complete hydrolysis of incompletely renatured collagen without any significant degradation of native collagen.



(a)



(b)

FIG. 5. Stepwise thermal denaturation of trypsin-resistant collagen molecules obtained after denaturation at 38°C for 90 min and renaturation by temperature fluctuation (collagen : enzyme = 100 : 1).

I—native collagen ($c = 0{\cdot}089$ g collagen/100 ml).

II—renatured collagen ($c = 0{\cdot}081$ g collagen/100 ml).

(a) According to optical rotation determinations.

(b) According to viscosity determinations.

(c) *The degree of renaturation of various renatured collagen molecules*

The amounts of trypsin-resistant tropocollagen molecules in various denatured and renatured solutions are given in Table II. As expected, renaturation of solutions denatured only to the first stage is greater than that of completely denatured solutions. The degree of renaturation is largely dependent upon the experimental conditions. Solutions renatured isothermally at 4°C show the lowest yields of trypsin-resistant

TABLE II. Content of trypsin-resistant tropocollagen molecules of various denatured and renatured solutions

Denaturation	Renaturation	TC trypsin-resistant (mg/g total collagen)†
11 min, 38°C	7 days, isothermally at 4°C	610
90 min, 38°C	7 days, isothermally at 4°C	230
11 min, 38°C	Cooling slowly at 4°C	670
90 min, 38°C	Cooling slowly at 4°C	430
90 min, 38°C	Temperature fluctuation	470

† See Table I(‡).

molecules. At lower temperatures, the peptide chains are too immobile to reorient themselves at a reasonable rate. On the other hand, on lowering the temperature gradually or on fluctuating the temperature, the molecules can arrange themselves more easily into a triple helical pattern. Engel draws a comparison between the effect of varying the temperature and the crystallization of quenched melts by local fusing of regions with energetically unfavourable configuration (Engel, 1962*b*).

It is remarkable that in every case more than 10% of the denatured collagen is resistant to trypsin. This means that not only the γ-component, but also completely dissociated α- and β-components can revert into the solid rod-like form of the native collagen molecule.

4. DISCUSSION

These investigations have shown that, apart from the γ-component, the α- and β-components can also be renatured to give intact TC molecules. Whereas the peptide chains in the γ-component are maintained in register by strong bonds, and thus are in the proper orientation for forming the triple helix directly on regeneration, the tangled α- and β-components have to achieve the proper mutual orientation first.

It is easily comprehensible how the ordered array of rigid, rod-like native TC molecules in the fibrils originates under the control of electrostatic forces. The characteristic distribution of positive and negative charges induces the molecules to aggregate laterally in a staggered sequence with quarter molecule length displacements. In this manner, the molecules always come together automatically in the same way (Kuhn and Zimmer, 1961*b*). A similar mechanism might also apply in principle with regard to the recombination of the peptide chain subunits to form tropocollagen molecules. However, it is difficult to conceive how more or less tangled peptide chains 3000 Å long can come together to fit so exactly that after being wound together and stretched, the ends terminate flushly.

The problem becomes even more complicated when it comes to the intertwining of the three α-components. At present, we are investigating the extent to which α-components can recombine *in vitro* to give rigid tropocollagen molecules.†

The problem of the formation of tertiary structure is of basic importance for protein chemistry. It is known that ribonuclease molecules can reform rapidly even after complete denaturation of its tertiary structure. Even after cleavage of the molecule into two fractions using subtilisin, the enzyme activity in the mixture, and hence the tertiary structure, remains unaffected. Evidently with this molecule, the tertiary structure is predetermined by the primary structure, i.e. by the arrangement of the amino acids along the peptide chain. Thus, the tertiary structure originates automatically under suitable conditions (Anfinsen and White, 1961).

It is an open question whether an analogous process is in operation with the significantly larger polypeptide chain of the collagen molecule. Collagen synthesis can in principle proceed in two ways. Firstly, the three peptide chains could be synthesized independently of each other and then intertwine in a process monitored by the primary structure. On the other hand, it is possible that the three chains are synthesized synchronously, and that the triple helix is constructed right from the start of the peptide chain growth. Our investigations do not exclude the former premise. It appears, too, that a supplementary orientation and extension of the molecules by electrostatic forces during the attachment of more or less tangled molecules on to a preformed fibril structure matrix in connective tissues might be possible.

† *Note added in proof*: Recently we investigated neutral salt-soluble collagen from rat skin (α-component only). After denaturation by heating at 38° C for 90 min and renaturation by cooling slowly to 4° C over a period of 7 days in citrate buffer at pH 3·7, 480 mg of trypsin-treated TC per g total collagen were obtained.

Acknowledgements

Thanks are due to Prof. U. Hofmann and Prof. W. Grassmann for their particularly valuable discussions and for the use of the facilities of their institutes, to Miss M. Kaestner for her expert technical assistance, and to the Deutsche Forschungsgemeinschaft for the provision of funds and apparatus.

References

Altgelt, K., Hodge, A. J. and Schmitt, F. O. (1961). *Proc. nat. Acad. Sci., Wash.* **47**, 1914.

Anfinsen, C. B. and White F. H., Jr. (1961). In *The Enzymes* (P. D. Boyer, H. Lardy, K. Myrbäck, eds.), Vol. 5, p. 95.

Boedtker, H. and Doty, P. (1956). *J. Amer. chem. Soc.* **78**, 4267.

Doty, P. and Nishihara, T. (1958). In *Recent Advances in Gelatin and Glue Research* (G. Stainsby, ed.), p. 92. Pergamon Press, London.

Engel, J. (1962*a*). *Arch. Biochem. Biophys.* **97**, 150.

Engel, J. (1962*b*). *Hoppe-Seyl. Z.* **328**, 95.

Engel, J., Grassmann, W., Hannig, K. and Kühn, K. (1962). *Hoppe-Seyl. Z.* **329**, 69.

Flory, P. J. and Weaver, E. S. (1960). *J. Amer. chem. Soc.* **82**, 4518.

Gallop, P. M. (1955). *Arch. Biochem. Biophys.* **54**, 486.

Grassmann, W., Hannig, K. and Engel, J. (1961). *Hoppe-Seyl. Z.* **324**, 284.

Harrington, W. F. and von Hippel, P. H. (1961). *Arch. Biochem. Biophys.* **92**, 109.

von Hippel, P. H. and Harrington, W. F. (1960). *Proc. Symposium Protein Structure and Function, No. 13*, p. 213. Brookhaven National Laboratories, June 6–8, 1960.

Kühn, K. and Zimmer, E. (1961). *Z. Naturforsch.* **16b**, 648.

Kühn, K., Kühn, J. and Hannig, K. (1961). *Hoppe-Seyl. Z.* **326**, 50.

Orekhovich, V. N. and Shpikiter, V. O. (1958). In *Recent Advances in Gelatin and Glue Research* (G. Stainsby, ed.), p. 87. Pergamon Press, London.

Orekhovich, V. N., Tustanovsky, A. A., Orekhovich, K. D. and Plotnikova, N. J. (1948). *Biokhimiia* **13**, 55.

Orekhovich, V. N., Shpikiter, V. O., Mazurow, V. J. and Kounina, D. V. (1960). *Bull. soc. chim. biol.* **42**, 505.

Ramachandran, G. N. and Kartha, G. (1954). *Nature, Lond.* **174**, 269.

Ramachandran, G. N. and Kartha, G. (1955*a*). *Nature, Lond.* **176**, 593.

Ramachandran, G. N. and Kartha, G. (1955*b*). *Proc. Indian Acad. Sci.* **A42**, 215.

Ramachandran, G. N. and Sasisekharan, V. (1961). *Nature, Lond.* **190**, 1004.

Rich, A. and Crick, F. H. C. (1955). *Nature, Lond.* **176**, 915.

Veis, A., Anesey, J. and Cohen, J. (1961). *Arch. Biochem. Biophys.* **94**, 20.

Recent Studies with the Electron Microscope on Ordered Aggregates of the Tropocollagen Macromolecule†

A. J. Hodge and J. A. Petruska

California Institute of Technology, Division of Biology
Gordon A. Alles Laboratory for Molecular Biology, Pasadena, California, U.S.A.

ABSTRACT

Examination of SLS-type aggregates in the electron microscope by the negative contrast method allows precise measurement of the length of the tropocollagen macromolecule in terms of the separation of δ bands in the SLS band pattern (this interval is equal to the axial period in native-type collagen fibrils). The normalized length found is $4{\cdot}40 \pm 0{\cdot}02$, indicating that there must be an end-to-end overlap of 0·40 in the "quarter-stagger" packing arrangement of native-type fibrils. Measurement of the $\delta_1 - \delta_4$ interval across the A–B junctions of a fibrous form of SLS (F–SLS) shows that the same overlap occurs in protofibrils formed by water dialysis.

Two possible models for the tropocollagen macromolecule have been examined in the light of present-day evidence. In the first (Model I), the three constituent polypeptide chains have a normalized length of 4·0, but longitudinal displacement of the chains relative to one another gives an over-all molecular length of 4·40. The second model (Model II) comprises three polypeptide chains each 4·40 in length with no longitudinal displacements. The latter model best explains the corrugated appearance of shadow-cast native-type fibrils, the apparent lengths of SLS from both enzyme-treated and control solutions of soluble collagen, and the regions of high mass thickness (of length 0·4) found in negative contrast images of both native-type and F–SLS fibrils.

An important consequence of Model II is that it predicts the presence of "holes" in the non-overlap regions of the fibril. In embryonic bone the sites of initial appearance of hydroxyapatite crystals appear to coincide with the ends of these "holes" in the structure.

The calculated molecular weight for Model II is 305,000, assuming a residue repeat of 2·86 Å, a mean residue weight of 93 for calf-skin collagen, and an axial period of 700 Å for native-type fibrils. Model I would have a molecular weight about 9% less.

The discovery of long-spacing ordered aggregates of the macromolecules of soluble collagen, especially the paracrystalline aggregation state known as "segment long-spacing" (SLS) by Schmitt *et al.* (1953), has had important consequences in the whole field of investigation of collagen

† Supported by a grant (RG-6965) from the National Institutes of Health, U.S. Public Health Service.

structures. Firstly, it led to the concept of the tropocollagen (TC) macromolecule as the monomeric unit in solutions of soluble collagen (Gross *et al.*, 1954). Furthermore, the lengths of the SLS forms indicated that the constituent TC monomers must be about four times the dimension of the well-known axial periodicity observed by low-angle X-ray diffraction and by electron microscopy in native-type collagen fibrils (see Bear, 1952). This result was later confirmed by physical chemical methods (Boedtker and Doty, 1956) and by direct electron microscopic visualization of individual macromolecules (Hall and Doty, 1958). Secondly, it became clear from the lengths of SLS and from the polarized character of SLS that the native-type band pattern must arise by longitudinal, unidirectional displacement of neighbouring protofibrils by multiples of approximately 1/4 of the molecular length, the protofibril being defined as a linear polymer of TC macromolecules. That this "quarter-stagger" packing arrangement is indeed the case for native-type fibrils has been demonstrated by direct "optical synthesis" of the native-type band pattern using the SLS band pattern as a starting point (Hodge and Schmitt, 1960). These authors were also able to establish the precise localization of the TC macromolecules in native-type fibrils by electron microscopic examination of dimorphic forms in which native-type fibrils were utilized as nucleation sites for the subsequent growth of SLS. The exact correspondence of SLS bands with those of the native-type fibrils (Fig. 1) with respect to axial location (but not with respect to intensity of staining) in these dimorphic forms clearly shows that all bands in the native-type pattern arise by summation of sets of four "equivalent bands", which contribute to the staining intensity by virtue of their lateral apposition in the staggered array (Fig. 2). Thus, the "molecular fingerprint" or linear map of the TC macromolecule as revealed by appropriately stained SLS has the property that all bands (representing clusters of both basic and acidic side-chains) may be classified into sets of four "equivalent bands" separated by intervals equal to the axial period in native-type fibrils, a feature which allows a rational assignation of SLS band nomenclature (Fig. 1). Given these restrictions on the band positions, it will be apparent that the SLS pattern must contain an intrinsic pseudo-period, although this is not immediately obvious from visual inspection of the pattern. The ordered aggregation states of the TC macromolecule have been recently reviewed by Hodge and Schmitt (1961).

The "molecular fingerprints" obtained by staining SLS with phosphotungstic acid (PTA) and cationic uranium to show the distribution of basic and acidic side-chains, respectively, (Fig. 3) show two main features of interest. Firstly, they match each other precisely in terms of band position, thus confirming the localization of the polar side-chains

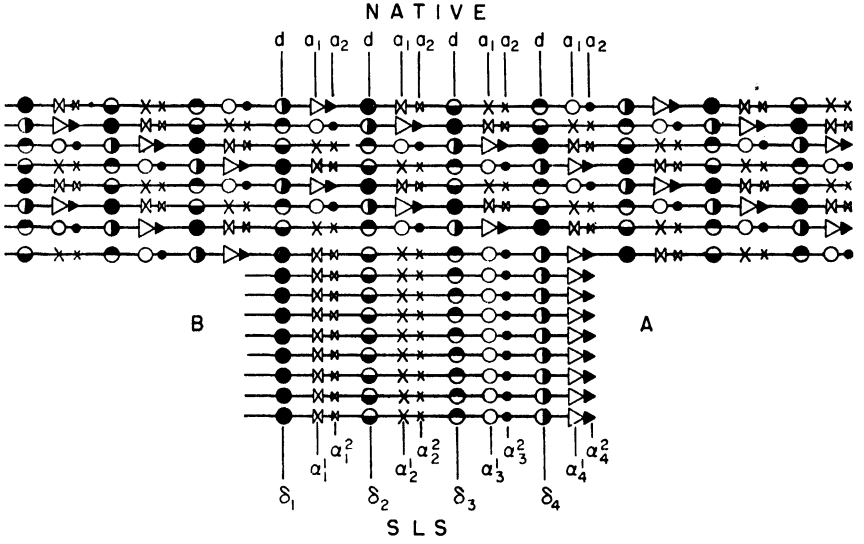


FIG. 2. Diagrammatic representation of the packing arrangement of TC macromolecules in the dimorphic aggregate shown in Fig. 1. Only three of the twelve or thirteen bands usually observable in the native-type structure are shown in order to minimize complexity. Note that in the SLS-type packing only like features are in register (i.e. homo-register), while in the native-type each band arises by the alignment of the four corresponding "equivalent loci" of the TC macromolecules (i.e. hetero-register), as a result of the "quarter-stagger" arrangement of the protofibrils. The TC are here depicted as having a normalized length of 4·0, i.e. four times the spacing between δ bands. See Figs. 7 and 9 for a more recent interpretation of protofibril structure. (From Hodge and Schmitt, 1960.)

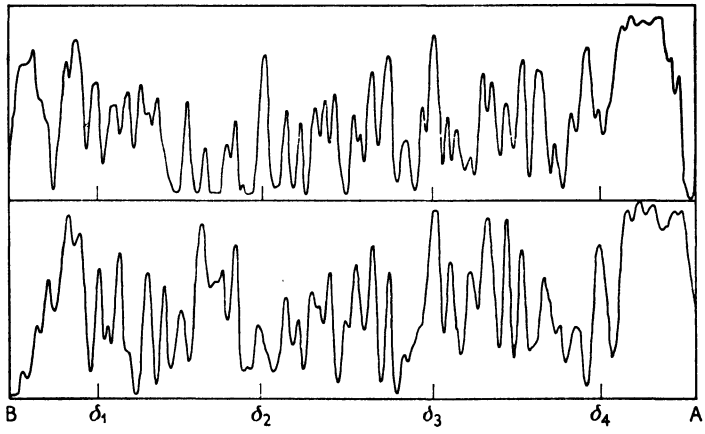


FIG. 4. Densitometric tracings of the SLS band patterns shown in Fig. 3. At top is the uranyl staining pattern, at bottom the PTA pattern. Note particularly the differences in staining intensity of the δ bands in the two patterns.

in clusters separated by relatively long non-polar regions, a concept originally derived from low-angle X-ray diffraction observations (see Bear, 1952). Secondly, there are obvious differences in the relative intensities of various bands in the two patterns, particularly when densitometric traces are compared (Fig. 4). These observations indicate that, while basic groups predominate in some of the polar loci, relative parity or an excess of acidic groups exists in others. The most likely explanation of these results is that the "charge profile" of the TC macromolecules is such that there is maximal complementation of charge when they are arranged in the "quarter stagger" array characteristic of the native-type fibril, under physiological conditions of pH and ionic strength.

1. The Tropocollagen Macromolecule

The physical and chemical evidence indicates that the monomeric unit of soluble collagen is a stiff, three-stranded rod about four times the axial period in native-type fibrils in length (see Harrington and von Hippel (1961) for detailed references), and having a molecular weight in the range 300,000 to 350,000. While earlier observations of SLS showed considerable variation in length of the TC macromolecule, recent measurements using improved specimen preparation techniques (Hodge and Schmitt, 1960) indicated a very narrow spread of molecular length. Indeed, the observed variance was within the degree of reproducibility to be expected for electron microscopic observations, in which factors such as dimensional changes occurring during drying of the specimen and also during electron irradiation, errors resulting from image distortion, and magnification calibration are to be contended with. Thus, the macromolecules of soluble collagen must be regarded as highly homogeneous, both with respect to molecular length and constancy of charge profile.

In addition to uncertainties resulting from errors in magnification calibration, the determination of molecular length by observation of positively stained SLS is beset by a more fundamental difficulty, namely, the possibility that the macromolecules extend beyond the last observable bands at one or both ends of the SLS. However, as we shall see, electron microscopic observations of SLS under negative contrast conditions (so-called negative staining) has allowed very accurate measurement of the molecular length in terms of the characteristic native-type macroperiod.

In a "quarter-stagger" array of TC macromolecules giving rise to a highly ordered band pattern, it is a geometrical consequence that if individual protofibrils arise by simple end-to-end abutment, the con-

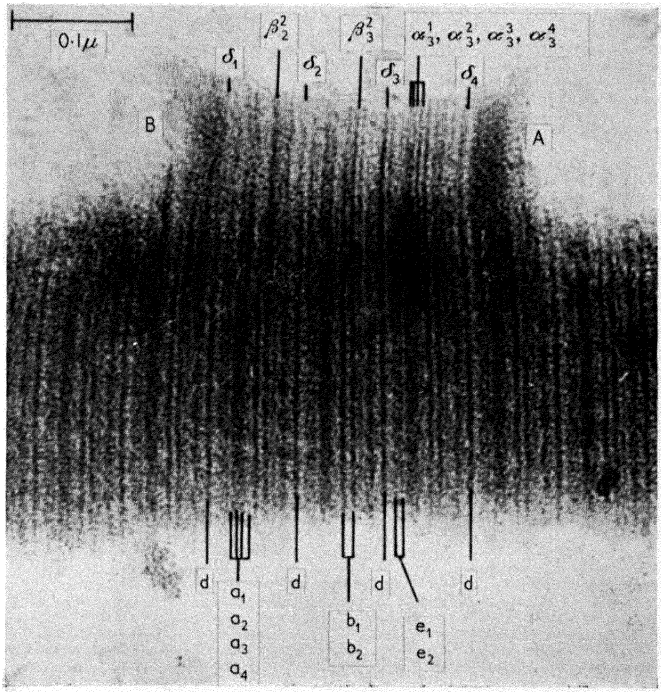


FIG. 1. Dimorphic ordered aggregate of TC produced by exposing native-type fibrils to a solution containing TC macromolecules and ATP at a pH favouring the formation of SLS. Note the continuity of the bands across both structures, especially that of the prominent *d* bands with the set of four δ bands in SLS each differing in staining intensity. (From Hodge and Schmitt, 1960.)

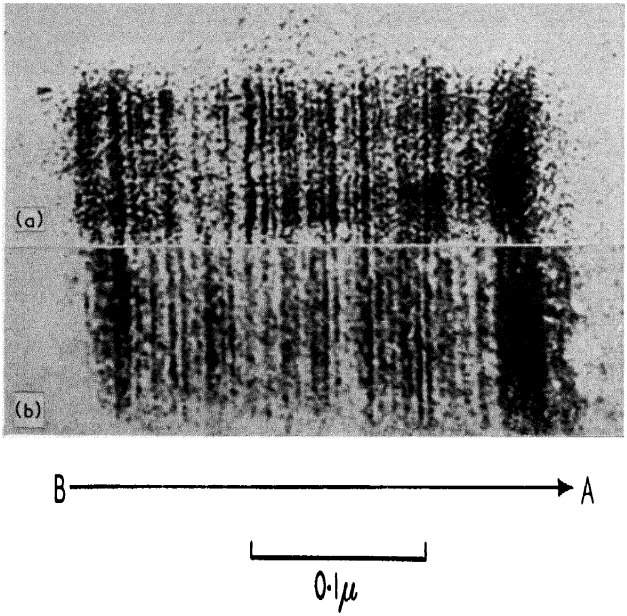


FIG. 3. Positively stained SLS-type aggregates, showing the close correspondence between the two "fingerprints" with respect to the location of basic and acidic staining loci: (a) stained with cationic uranium, (b) with phosphotungstic acid, pH 4·2. (From Hodge and Schmit, 1960.)

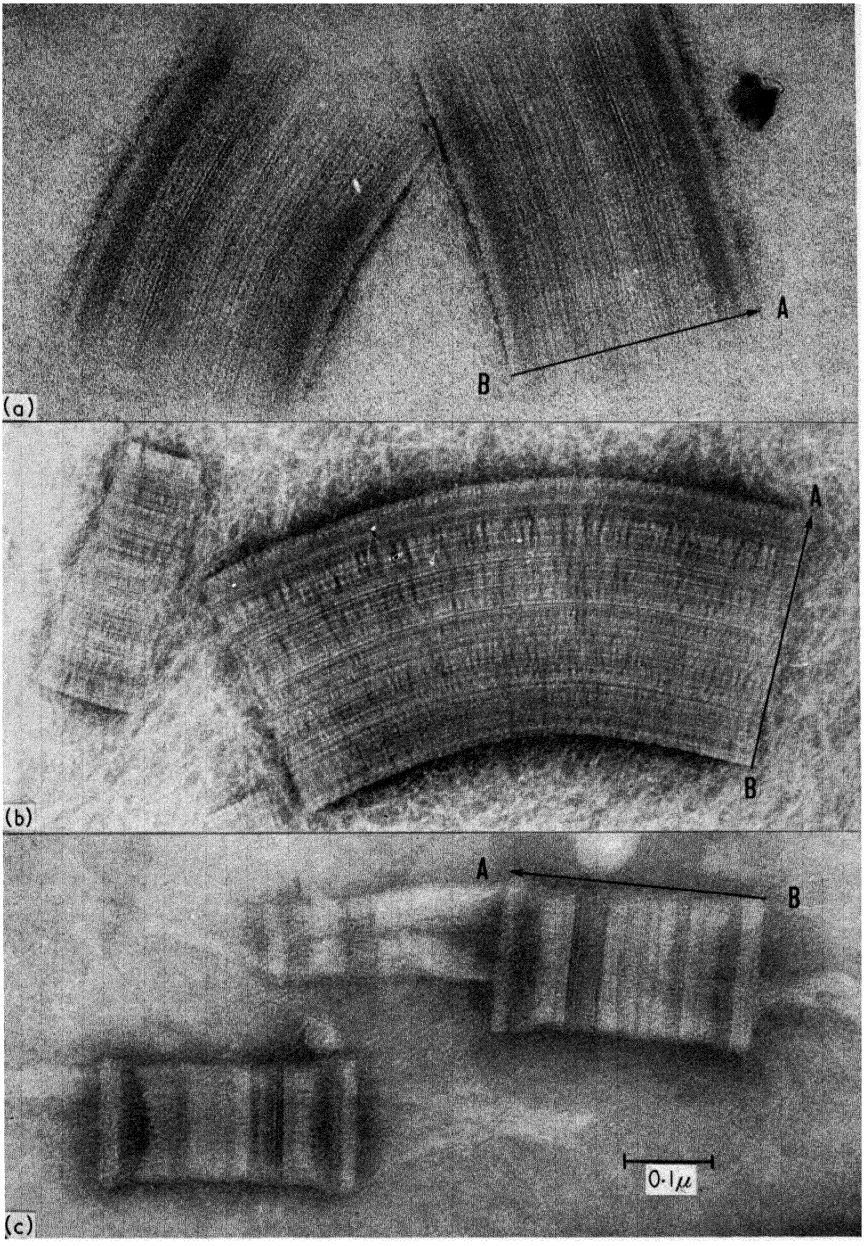
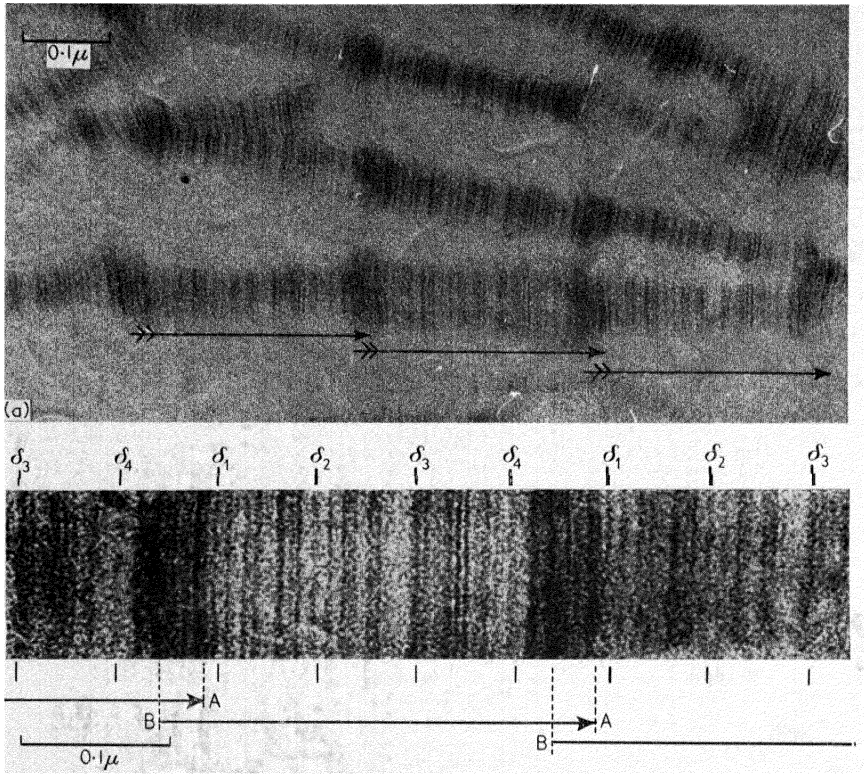


FIG. 5. SLS-type aggregates observed by the negative contrast method, showing the ends in sharp relief; (a) and (b) are two representative areas from the same specimen grid using sodium phosphotungstate, pH 7·0, as the contrast medium, illustrating the variation in appearance of the band pattern with local concentration of PTA; (c) shows a similar SLS preparation using PTA at pH 4·2 as the contrast medium.



(b)

FIG. 6. F-SLS fibrils positively stained with PTA; (a) shows the strikingly high density and "bulging" in the junctional regions, attributed to overlap; (b) the equidistant spacing of all δ bands. (From Hodge and Petruska, 1962.)

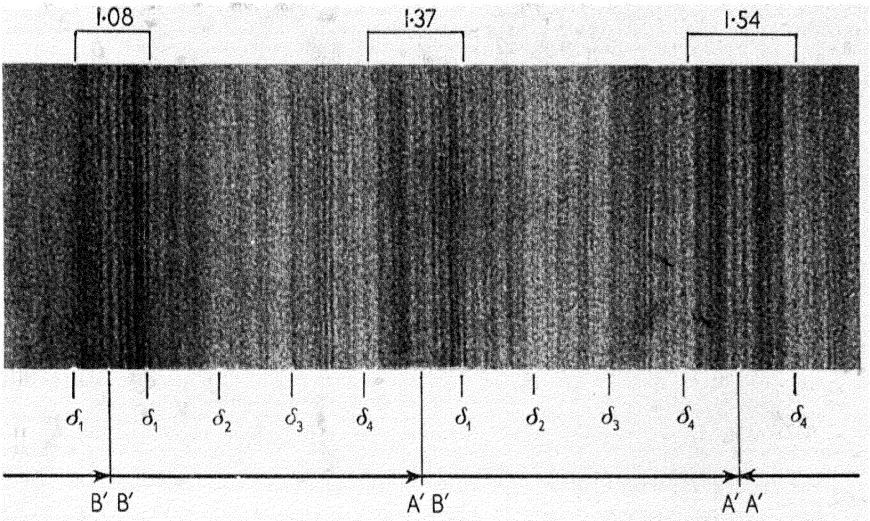


FIG. 8. Polymeric SLS-type aggregate obtained from sonicated solution of calf-skin collagen, positively stained with PTA, showing abnormal end-to-end interactions with little or no overlap. Note that the $\delta_1 - \delta_4$ interval across the A′B′ junction measures 1·37 (close to the expected value of 1·40 for no overlap) instead of 1·0 as in F–SLS (Fig. 6).

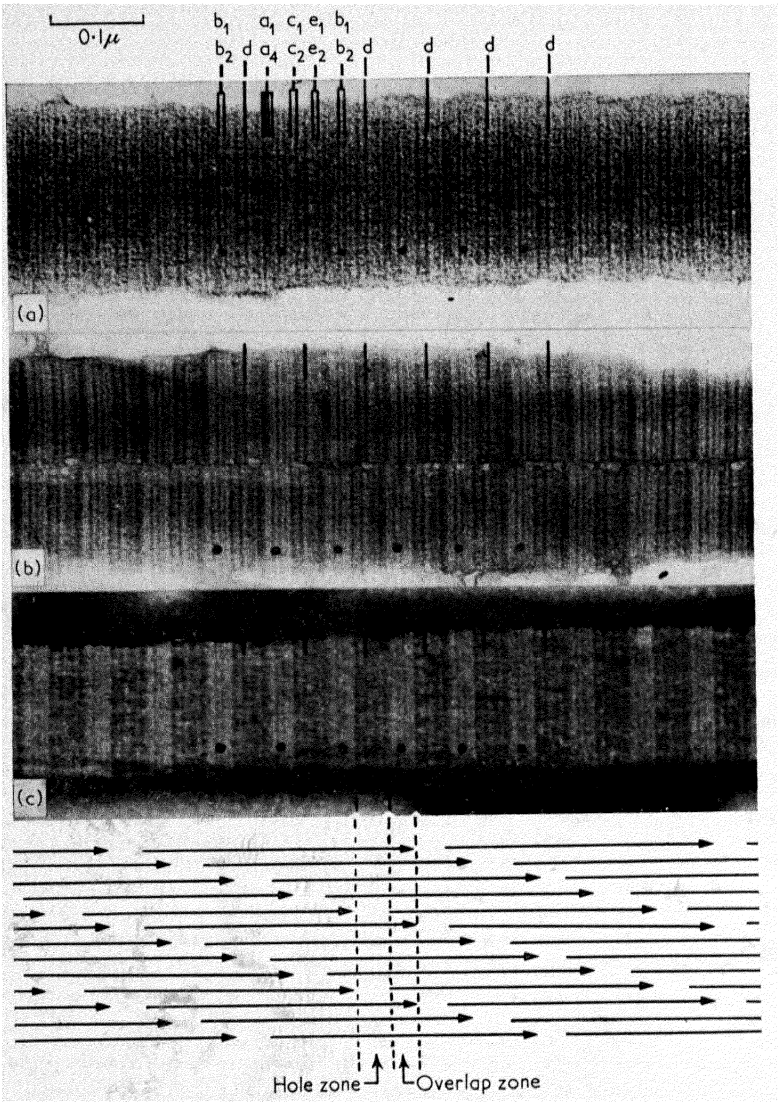


FIG. 9. Native-type fibrils arranged to show the relationship between the band patterns observed in positive and negative staining: (a) positively stained, PTA, pH 4·2, (b) positively stained as in (a) but with a local excess of PTA giving rise to a partial negative contrast effect, and (c) negatively stained, PTA, pH 7·0. The location of the *b* doublet is indicated by dots, the sharp *d* band by lines. In the negative contrast image (c) the regions of low density, bounded by the a_3 band on one side and the c_2 band on the other, correspond to the overlap regions in the protofibril. At bottom is a schematic representation of the deduced packing arrangement of the TC macromolecules, showing the overlap and hole zones of the macroperiod.

stituent monomeric units must be exactly four times the native-type macroperiod of *ca.* 700 Å (the exact value depends on the collagen being examined and on its degree of hydration (Bear, 1952)). For the purposes of subsequent discussion, let us assign this period (equal to the spacing between consecutive "equivalent bands" in the SLS band pattern, e.g. $\delta_1 - \delta_2$) a value of 1·00 and let all other length parameters be normalized with respect to this interval, thus minimizing the various causes of error arising in the absolute determination of dimension by electron microscopy.

When SLS derived from acid-soluble solutions of calf-skin collagen are examined by the negative contrast technique using neutralized phosphotungstic acid (PTA) as the contrast medium it is found that the ends of individual segments are sharply delineated, thus allowing accurate measurements of molecular lengths (Fig. 5). Although the details of the SLS band pattern vary considerably depending on the amount of contrast medium present in the vicinity of any particular segment, it is possible to determine the apparent molecular length (normalized to the interval between consecutive "equivalent bands", e.g. the $\delta - \delta$ separation) with a precision approaching 0·5% (Hodge and Petruska, 1962). The actual value obtained for the apparent length of the TC macromolecule by this technique is $4{\cdot}40 \pm 0{\cdot}02$. It should be noted that this represents a minimum value, since any "end-chains" such as those proposed by Boedtker and Doty (1956), and by Hodge and Schmitt (1958) on the basis of electron microscopic examination of polymeric SLS-type structures obtained from sonically treated collagen solutions, might be in a random coil configuration in any non-polymeric ordered aggregate. If, on the other hand, we assume that the TC macromolecule is three-stranded throughout its apparent length (as determined by negative contrast observations of SLS), it becomes possible to compute a molecular weight based: (1) on the average residue weight (93 for calf-skin collagen); (2) on the assumption that the residue repeat is 2·86 Å; and (3) on the assumption that the axial period derived from low-angle X-ray diffraction is directly related to the $\delta - \delta$ interval observed in electron micrographs of SLS. Assuming a low-angle period of *ca.* 700 Å for hydrated vertebrate collagens, we arrive at a molecular weight for the TC macromolecule of *ca.* 300,000.

2. Protofibril Formation

When an acid solution of collagen is dialysed against water, its viscosity rises and it can be shown that this is due to formation of protofibrils (Hodge *et al.*, 1960). Addition of ATP gives rise to a fibrous form of SLS (F–SLS) shown in Fig. 6. Close examination of the band pattern

of these F–SLS fibrils shows that all of the $\delta - \delta$ intervals, including the $\delta_1 - \delta_4$ distances across the A–B junctions, are equal to 1·0. It is clear, therefore, both from this result and the value of 4·40 for single segments, that the formation of protofibrils must involve a specific end-to-end overlap of 0·4, i.e. about 10% of the molecular length. Similarly, since it has already been shown both by direct optical synthesis and by band correspondence in dimorphic forms that all bands in the native-type period arise by lateral juxtaposition of the four corresponding "equivalent bands" in the SLS pattern (e.g. the *d* band arises by summation of the δ_1, δ_2, δ_3 and δ_4 bands), it follows directly that the same degree of overlap must be present in native-type fibrils.

3. Nature of the Overlap Regions

As we have seen, there is unequivocal evidence for an overlap of about 10% of the molecular length in collagen protofibrils, and also in F–SLS and native-type fibrils. B. R. Olsen (unpublished data)† of the University of Oslo has also concluded from negative contrast examination of various fibrous long-spacing (FLS) forms, SLS and native-type fibrils that there must be an overlap of *ca.* 300 Å in the latter type of fibril.‡ However, these findings still leave open such questions as the type of interaction involved in protofibril formation and the number of polypeptide chains present in the overlap regions (Fig. 7). Let us examine some of the evidence relevant to these problems.

On the basis of a TC model with dangling terminal peptide chains proposed by Boedtker and Doty (1956) and after comparison of monomeric SLS with certain abnormal polymeric SLS forms obtained by the addition of ATP to sonicated TC solutions, Hodge and Schmitt (1958) proposed that formation of protofibrils might involve a specific coiling of such terminal chains to form a highly ordered junctional region. In any such model, the length of the monomer must necessarily exceed 4·0 (see Hodge and Schmitt, 1961, for a summary). Hodge *et al.* (1960) also tested the effects of various proteolytic enzymes such as trypsin and pepsin on TC macromolecules in solution in order to determine whether such terminal peptides could be isolated. It was found that such enzymatic treatment, while not producing any appreciable change in the size and shape of the TC monomers, nevertheless resulted in: (1) depolymerization of any linear polymers present; and (2) complete abolition or inhibition of protofibril formation in the water dialysis system and

† We are greatly indebted to Professor T. W. Blackstad of the Anatomy Department, University of Oslo, for bringing this work to our attention.

‡ This value is in good agreement with our own estimate of an overlap region with a normalized length of 0·4, but differs radically from that of K. Kühn (*Leder*, **13**, 86 (1962)), who claimed an end-to-end overlap of 30–40 Å in native-type fibrils.

reduction of the rate of formation of native-type fibrils in the "thermal gelation" system. Since it had been reported that iodination of tyrosyl residues markedly increased the rate of fibril formation (Bensusan and Scanu, 1960), and because an apparently terminal tyrosine-containing peptide had been isolated from tryptic digests of thermally denatured collagen (Grassman *et al.*, 1956), Hodge *et al.* (1960) labelled the digests with ^{131}I after trypsin treatment of both native and denatured TC. Paper curtain electrophoresis yielded an acidic peptide fraction containing hydroxyproline and having a high ^{131}I specific activity associated with the tyrosine content. However, insufficient material was recovered to allow isolation of a chromatographically and electrophoretically pure

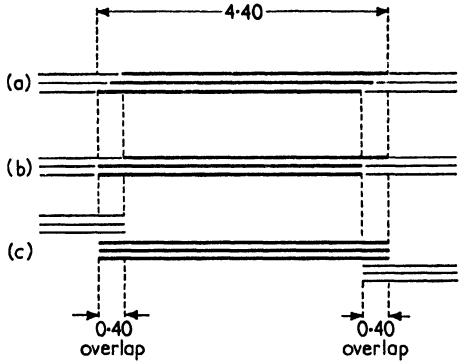


FIG. 7. Three possible models for the TC macromolecule and their corresponding modes of overlap in the protofibril; (a) and (b) are examples of a general class of structure (Model I) in which the three constituent polypeptide chains each have a normalized length of 4·0 and are staggered relative to one another; (c) is the type of structure (Model II) suggested by the negative contrast observations. In the latter model each chain has a length of 4·40 and there is no longitudinal displacement.

peptide containing hydroxyproline and with the appropriate lysine or arginine C-terminal group, a result which would have been necessary to clearly establish peptide linkage of any such enzymatic product with the "body" of the TC macromolecule. Thus, although the results of these experiments are compatible with the concept of dangling terminal chains, they could also be explained by the presence of a small amount of gelatin in the TC solutions together with small amounts of a tyrosine-rich peptide or protein contaminant non-covalently linked to the TC macromolecules. Such a protein might be a specific co-factor involved in fibril formation *in vivo*. At the present time, it can fairly be stated that while enzymes such as trypsin, chymotrypsin, and pepsin undoubtedly affect the capacity of the TC macromolecule for end-to-end polymerization, the mechanisms which give rise to such effects remain obscure.

However, recent results of ours suggest that the effects of these enzymes may be physical rather than enzymatic.

It could be argued in favour of the end-chain model (Fig. 7 (a), (b)) that in view of the extensive evidence in support of water being involved in the stabilization of collagen, such terminal chains need not necessarily be in a random coil configuration, but might be sufficiently ordered to contribute bands to the ends of the SLS band pattern. In support of such a concept are data showing that reformation of helicity in single gelatin strands probably involves stabilization by water molecules (see review by Harrington and von Hippel, 1961). The observed buoyant density changes in passing from TC in solution to the hot form of gelatin, and from there to the cold form in density gradient ultracentrifugal studies (Fessler and Hodge, 1962) would also be consistent with such a view. On the other hand, it has been shown (Altgelt *et al.*, 1961) that in a cooled thermally denatured acid solution of TC, only those TC macromolecules with thermostable (presumably covalent) cross-links between all three strands (the γ-component) are capable of complete renaturation as judged by the following criteria: (1) complete thermal reversibility of viscosity and optical rotatory properties; and (2) restoration of the original charge profile as manifested by the capacity to form SLS. It must be concluded, therefore, that the relatively slow mutarotation and viscosity recovery of the α- and β-components (single and two-stranded, respectively) does not seem to result in a charge profile sufficiently ordered to allow formation of SLS-type structures.

While such a conclusion does not necessarily apply to specialized terminal peptide structures (it could for example be argued that a composition different from that of the bulk of the TC macromolecule allows single or double-stranded peptides to maintain a high degree of helicity, and hence presumably rigidity), recent electron microscopic analyses would seem to favour the view (Fig. 7 (c)) that the TC macromolecule is three-stranded throughout its normalized length of 4·40. The main lines of argument are as follows:

(1) Close examination of the abnormal polymeric forms of SLS produced after sonication of acid solutions of TC (Fig. 8) reveal that the $\delta - \delta$ intervals across the A′ – A′, B′ – B′, and especially across the A′ – B′ junctions, are not equal to 1·0 (Hodge and Petruska, 1962). In fact, the separation between the centres of A′ – A′ and B′ – B′ junctions is almost exactly that of the molecular length determined from negative contrast observation of single SLS, indicating that there is little or no overlap in these particular polymeric SLS forms. In particular, the $\delta_1 - \delta_4$ interval across the A′ – B′ junction measures 1·37 instead of the value of 1·0 found for the same interval across the A–B junction in F–SLS fibrils (Fig. 6 (b)).

(2) In positively stained F–SLS fibrils one gains a definite qualitative impression of increased bulk in the junctional regions (Fig. 6 (a)), as well as increased band density to be expected from the overlap.

(3) To date, SLS made from TC solutions after treatment with trypsin or pepsin under appropriate conditions do not differ significantly from control SLS with respect to length when examined by negative contrast technique (Petruska and Hodge, unpublished data).

(4) Perhaps the most telling argument is that native-type collagen fibrils, whether examined by negative contrast technique or after shadow-casting, exhibit a periodic distribution of mass thickness which cannot be attributed solely to the pattern of polar side-chains seen in appropriately stained SLS. In fibrils observed by negative contrast technique (Fig. 9 (c)), the regions of high mass thickness, likely corresponding to the raised regions observed in shadowed electron micrographs, have a normalized length of *ca.* 0·4, in agreement with the overlap already described. Furthermore, the regions intermediate between the predicted zones of overlap are more penetrable by the negative contrast agent and show a longitudinal structure indicative of disorder resulting from dehydration. In this connection it is of interest that Tomlin (1955) suggested a model with four-fold units and overlapping ends to explain certain features of the changes in the X-ray diffraction pattern on drying, and attributed such alterations to disorder in the overlap regions. However, the bulk of evidence to date would suggest that most of the disorder resulting from dehydration occurs in the non-overlap region of the native-type period, i.e. the depressed regions in shadow-cast fibrils. Our data also support the interpretation of Bear (1952) that the characteristic "fanning" of the low-angle diffraction pattern resulting from dehydration is due to disorder arising in the polar regions of the fibrils. Thus, in positively stained SLS we have consistently observed that the $\delta_1 - \delta_2$ and $\delta_3 - \delta_4$ intervals may be as much as 3–4% less than the $\delta_2 - \delta_3$ separation, as would be expected from the higher concentration of polar groups observed in the distal regions of the SLS band pattern. Such differences appear to be minimized in SLS observed by the negative contrast method, in agreement with the results of others on a wide variety of particulates of biological origin, namely, that structural order is better preserved with this technique than in specimens prepared by conventional drying and staining methods.

4. "Holes" in Native-type Fibrils

If we accept the concept supported by the bulk of the evidence already reviewed, namely, that the TC macromolecule is three-stranded throughout its normalized length of 4·40 and that protofibril formation must

normally involve an end-to-end overlap of 0·4, it follows that there is no possible structural arrangement in which all TC macromolecules can be completely close-packed throughout their length when neighbouring protofibrils are longitudinally displaced in the "quarter-stagger" array required for the formation of the native-type periodicity. It also follows that in the hydrated structure there must be "holes" or "pores" about 400 Å in length (normalized length = 0·6) and having an effective cross-sectional diameter about that of the TC macromolecules, even if we assume perfect close-packing for the overlap region. The distribution of such "holes" within a native-type collagen fibril clearly depends on the type of packing assumed (hexagonal or other array) and on the precise manner in which a protofibril is made up from TC monomers.

It is of considerable interest that pores have been demonstrated in vacuum-dried tendon fibrils immersed under Hg at pressures around 10,000 p.s.i. (Swerdlow and Stromberg, 1955). Under these conditions, the fibrils took up mercury which could be seen in the electron microscope to be distributed in bands or spots spaced about 700 Å apart in the axial direction. It seems very likely that the mercury was penetrating the same regions entered by PTA in the negatively stained fibrils (Fig. 9 (c)). The presence of pores in native-type fibrils thus seems well-established, and we are currently attempting to devise means to study their three-dimensional distribution within the fibril and in relation to the band pattern.

The presence of discrete holes in collagen fibrils has some important implications in relation to their mineralization both *in vivo* and *in vitro*. It has already been shown (Glimcher *et al.*, 1957) that of the various possible reconstituted aggregation states of the TC macromolecule, only the native-type fibrils are able to initiate nucleation of hydroxyapatite from metastable solutions containing calcium and phosphate ions in concentrations comparable with those present in physiological fluids. These authors postulated that only the "quarter-stagger" type of packing allowed suitable juxtaposition of the side-chain groups involved and thus the formation of stereochemically-specific nucleation sites within each axial period (see Glimcher, 1959, for details), as required by the electron microscopic observation that the crystallites form within the fibrils and only within a certain region of the axial period. Close examination of the electron micrographs of embryonic bone published by Fitton-Jackson (1957) reveals that the crystallites form initially in those regions of the axial period corresponding to the ends of the pores, i.e. in close proximity to the ends of the TC macromolecules. It seems likely, therefore, that the nucleation sites comprise groups at or near the ends of the TC macromolecules together with others brought into proximity with the former as a result of the "quarter-stagger" packing arrangement of

protofibrils. Further growth of the crystallites which become characteristically oriented with respect to the fibre axis, as shown by selected area electron diffraction of sections of embryonic bone, could presumably occur within the gaps in the non-overlap zones. Recent electron microscopic examination of early stages of calcification in embryonic bone (Glimcher and Hodge, unpublished data) show clearly that the hydroxyapatite crystals are confined to a definite region of the axial period probably corresponding to the non-overlap zones described. However, their precise position with respect to the period is difficult to establish since the fibrils in such osmium-fixed preparations show relatively few bands per period.

5. On the Negative Staining of Tropocollagen Structures

The examination of TC aggregation states, especially SLS and native-type fibrils, has yielded some interesting and unexpected results. Initially it was reasoned that since the bands contain the bulky polar side-chains, these should show up in negative contrast as white lines against a darker background. However, as shown in Figs. 5 and 9, the band pattern obtained is highly variable and depends on the pH and local concentration of the negative contrast agent (PTA in the examples indicated). Specifically, the ideal negative contrast image is approached where there is a large excess of PTA at neutral pH, but is modified by penetration of the agent into the structure (high ionic strength tends to disperse the relatively unstable SLS forms). In regions of relatively low PTA concentration both the SLS and native-type band patterns bear a strong resemblance to the corresponding positively stained band patterns, only the ends of the SLS showing up in negative contrast. This suggests that positive binding of PTA occurs even at neutral pH and that the over-all image is a composite of both positive and negative staining characteristics. The discovery of a dense non-ionizable inert compound would greatly facilitate negative contrast observations of structures involving charged groups.

References

Altgelt, K., Hodge, A. J. and Schmitt, F. O. (1961). *Proc. nat. Acad. Sci., Wash.* **47**, 1914.

Bear, R. S. (1952). In *Advances in Protein Chemistry* (M. L. Anson and J. T. Edsall, eds.), Vo. 7, p. 69. Academic Press, New York.

Bensusan, H. B. and Scanu, A. (1960). *J. Amer. chem. Soc.* **82**, 4990.

Boedtker, H. and Doty, P. (1956). *J. Amer. chem. Soc.* **78**, 4267.

Fessler, J. H. and Hodge, A. J. (1962), *J. mol. Biol.* **5**, 446.

Fitton-Jackson, S. (1957). *Proc. roy. Soc.* **B 146**, 270.

Glimcher, M. J. (1959). *Rev. Mod. Phys.* **31**, 359.

Glimcher, M. J., Hodge, A. J. and Schmitt, F. O. (1957). *Proc. nat. Acad. Sci., Wash.* **43**, 860.

Grassmann, W., Hannig, K., Endres, H. and Riedel, A. (1956). *Hoppe-Seyl. Z.* **306**, 123.

Gross, J., Highberger, J. H. and Schmitt, F. O. (1954). *Proc. nat. Acad. Sci., Wash.* **40**, 679.

Hall, C. E. and Doty, P. (1958). *J. Amer. chem. Soc.* **80**, 1269.

Harrington, W. F. and von Hippel, P. H. (1961). In *Advances in Protein Chemistry* (M. L. Anson and J. T. Edsall, eds.), Vol. 16, p. 1. Academic Press, New York.

Hodge, A. J. and Petruska, J. A. (1962). *Proc. of the Vth International Congress for Electron Microscopy, Philadelphia*, Aug. 29–Sept. 5, 1962 (S. S. Breese, Jr., ed.). Academic Press, Paper QQ–1.

Hodge, A. J. and Schmitt, F. O. (1958). *Proc. nat. Acad. Sci., Wash.* **44**, 418.

Hodge, A. J. and Schmitt, F. O. (1960). *Proc. nat. Acad. Sci., Wash.* **46**, 186.

Hodge, A. J. and Schmitt, F. O. (1961). In *Macromolecular Complexes* (M. V. Edds, Jr., ed.). Ronald Press Co.

Hodge, A. J., Highberger, J, H., Deffner, G. G. J. and Schmitt, F. O. (1960). *Proc. nat. Acad. Sci., Wash.* **46**, 197.

Schmitt, F. O., Gross, J. and Highberger, J. H. (1953). *Proc. nat. Acad. Sci., Wash.* **39**, 459.

Swerdlow, M. and Stromberg, R. R. (1955). *J. Res. nat. Bur. Stand.* **54**, 83.

Tomlin, S. G. (1955). *Proc. Internat. Wool Textile Res. Conf., Australia*, B, 187.

Discussion

G. N. RAMACHANDRAN: I have been trying to understand how the one-fourth shift between the protofibrils which you have found, could be pictured so as to give a cylindrical packing of these. Do you think the electron microscope can give any indication about this?

A. J. HODGE: The problem is worse now when we have overlap like this. We can find the arrangement of best close packing. We were expecting to see if the distribution of holes would have anything to do with the X-ray pattern further in than the usual 10 or 11 Å, for which there seems to be already some indication according to Dr. Corey.

G. N. RAMACHANDRAN: There were some photographs taken by Cowan and co-workers in King's College some years ago, where they found a definite pattern on the equator inside the 12 Å spots. It has a number of maxima in the low angle region. Dr. Sasisekharan and I have been able to explain them in terms of a cylindrical lattice with just seven layers (G. N. Ramachandran and V. Sasisekharan (1956). *Arch. Biochem. Biophys.* **63**, 255; V. Sasisekharan and G. N. Ramachandran (1957). *Proc. Indian Acad. Sci.* **A45**, 363).

Electron Microscopic and Sedimentation Studies on Human Haemoglobins A, F and E

N. N. Das Gupta, D. N. Misra, P. Ganguly, A. B. Sanyal and J. B. Chatterjea*

Biophysics Division, Saha Institute of Nuclear Physics, Calcutta, India

ABSTRACT

The sedimentation analysis has been performed on human haemoglobins A and F and the results compared with those of previous workers. It is found that the HbF has higher S values compared to HbE, and that the sedimentation coefficients of HbE and HbA are identical. At very low concentrations, sedimentation coefficients S tend to increase with c in both cases. Electron micrographs of negatively stained HbA and HbF indicate a number of particles with small holes at the centre. The dimensions of individual particles seem to be about 62 Å × 76 Å in both cases.

Recent experiments have brought to light the existence of a large number of human haemoglobin (Hb) variants which differ from one another in their primary structure. These differences may be the substitution of one amino acid for another in the same chain as in haemoglobin A, S, C and E, or the substitution of one chain as a whole by another as between HbA (α_2^A, β_2^A), HbF (α_2^A, γ_2^F), HbH (β_4^A) and Hb "Bart's" (γ_4^F) (Wintrobe, 1961). Such primary structural changes may bring about secondary and tertiary changes in the coiling of the polypeptide chains and in the folding of these coils. Changes in the physical characters, viz. in the surface charge, surface structure, size, shape, intermolecular association, etc., may be the ultimate result of the differences in the primary structure. The present investigation was undertaken with a view to explore whether any variation in the biophysical character of three common haemoglobin variants, viz. HbA, F and E could be recognized with the techniques of electron microscopy and ultracentrifugation.

The normal foetal haemoglobin of man, designated HbF, is present in infants, but it may also be found in children and adults with thalassaemia major and other abnormal haemoglobin diseases. Haemoglobin E is found either in the homozygous or in the heterozygous state in a significant proportion of the population in South-Eastern Asia. In West Bengal,

* Haematology Department, School of Tropical Medicine, Calcutta, India.

approximately 4% of the people have this abnormal haemoglobin. The human haemoglobins A and F have been shown to differ in solubility, in the spread on monomolecular films, in affinity for oxygen, stability towards alkalies, in ultraviolet absorption, in crystallographic behaviour, etc.; however, electrophoretic separation of Hb A and F is not obtained (White and Beaven, 1959). Electrophoretically, Hb E is completely separable from Hb A, but with respect to other physical properties Hb E does not seem to differ significantly from Hb A (Beaven and Gratzer, 1959).

The previous sedimentation analysis of haemoglobins was mostly concerned with Hb A. Kegeles and Gutter (1951) made a systematic study of the sedimentation of human carboxyhaemoglobin within the concentration range of 0·1–2%. Field and O'Brien (1955) reported the value of 4·18S for human carboxyhaemoglobin at 0·71% protein concentration. Kekwick and Lehmann (1960) recently obtained a value of 4·13S for a mixture of 89% Hb A and 11% Hb Bart's at 1% protein concentration. No sedimentation study has been reported on Hb F and Hb E and also no comparative study has been done on the different haemoglobins under identical conditions. The only electron microscopic studies of the haemoglobins reported earlier were with the shadowed preparations (Chatterji *et al.*, 1961; Sadhukhan *et al.*, 1962). It is well known that accurate determination of the size and shape of a macromolecule from shadowed electron micrographs is difficult, due to the distortions introduced by the shadowing metal. The present studies were therefore carried out with negative staining technique (Horne, 1961), which enabled an examination of the molecules at a much higher magnification than was possible before.

Materials and Methods

(a) *Preparation of samples*

Samples of Hb A were obtained by drawing fresh blood from the veins of normal adults, and those of Hb F from the cord blood of newborn babies. The oxalated red blood cells were first freed from serum by repeated washing with normal saline. The serum-free cells were then haemolysed with ten times their volume of distilled water. The haemolysed cells were spun down at 3500 rev/min for 30 min and the clear haemoglobin solution pipetted off and stored with a little toluene. The haemoglobin F samples analysed by the alkaline denaturation method of Singer *et al.* (1951) showed about 80% haemoglobin F, the rest being haemoglobin A. Hb E was obtained from a homozygous E subject, whose blood contained about 98% Hb E and only 2% Hb F.

The concentrations of the protein were determined from the absorption

at 540 $m\mu$ with a Zeiss spectrophotometer. A solution with a concentration of the order of 10^{-3} g/ml was used in the electron microscopic work; for ultracentrifugal studies the concentrations varied from 3×10^{-3} to 4×10^{-4} g/ml.

(b) *Sedimentation analysis*

The sedimentation analysis was done with a Spinco Model E analytical ultracentrifuge with Schlieren optical system. The sedimentation runs were given in double distilled water as well as in 0·1 M NaCl, pH approximately 7·0, at a speed of 56,100 rev/min at varying protein concentrations. $S^0_{20,w}$ was obtained by extrapolation to zero concentration.

(c) *Electron microscopy*

Electron microscopy was done with a Siemens Elmiskop I at 60 kV, at an electronic magnification of × 40,000. Drops of Hb solutions at concentrations of about 10^{-3} g/ml were settled on holed carbon-coated collodion films and after subsequent rinsing, stained with 2% PTA for 5 min at pH 6–7. It was found that films of PTA with haemoglobin molecules were sometimes suspended across the holes when the best contrasts were obtained in the electron micrographs. Preliminary experiments with PTA of different pH, showed that the Hb F was best stained at pH 6, and Hb A at pH 6·5.

Results and Discussions

Figure 1 shows a plot of the sedimentation coefficient $S_{20,w}$ against concentration of Hb in g/100 ml. It is found that there is substantial agreement between the results of Kegeles and Gutter (1951), Field and O'Brien (1955), Kekwick and Lehmann (1960) on Hb A as well as those obtained in the present work on Hb E. It is found that the S–c curve deviates appreciably from linearity for concentrations below 0·3%, showing an increase of S with dilution. The results indicate that there is no dissociation of Hb molecules with dilution up to 0·04% concentration. From the present measurements, made within the concentration range 0·30–0·04%, the extrapolated value of sedimentation coefficient for Hb A corresponding to zero concentration ($S^0_{20,w}$) is found to be 4·53S.

The sedimentation coefficients for Hb F (80%) obtained under identical conditions are also indicated Fig. 1. The data for Hb F are few and do not extend over as wide a range of concentrations as in the case of Hb A. However, the present measurements for concentrations less than 0·3% show consistently higher S-values in the case of Hb F. If approximately the same shape and hydration are assumed, the present results indicate that Hb F has a slightly larger molecular volume than Hb A.

Representative electron micrographs of negatively stained preparations of Hb F and A are shown in Fig. 2 (a) and (b). The most striking feature of these micrographs is the presence of a large number of particles with small holes at the centre. Figure 3 (a), (b), where portions of the previous figure have been enlarged to higher magnification (× 400,000), show the holes very clearly. Sometimes 2, 3 or 4 particles are seen to be lying close together, when the holes of individual particles can also be distinguished (Fig. 3 (c), (d) and (e)). When a number of particles are thus joined together, they appear to be somewhat flattened towards the common side. The smallest particles, presumably representing the individual molecules, are definitely elliptical with an average ratio of 1:2 between the two dimensions at right angles. The average dimensions of the

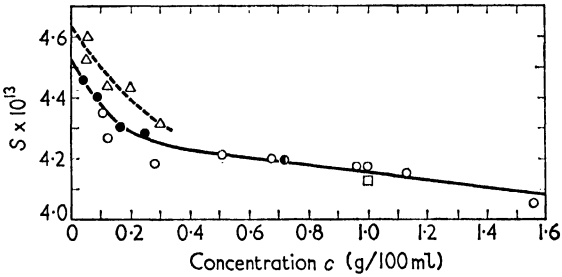


FIG. 1. Variation of sedimentation coefficient S with the concentration c of haemoglobin (g/100 ml). Data of the present investigation have been plotted along with those of previous workers.

○, Kegeles and Gutter; ◐, Field and O'Brien;

□, Kekwick and Lehmann; Present work: Hb E, ●; Hb F, △.

individual small particles (both Hb A and F) measured from their electron micrographs seemed to be about 62 Å × 76 Å. A large number of particles will have to be measured more thoroughly before difference in size, if any, between Hb A and F molecules could be established with precision. The hole diameter, which is almost at the limit of resolution of the microscope, appears to be about 18 Å.

The presence of a hole in the haemoglobin molecule from horse blood was deduced by Perutz *et al.* (1960) from X-ray analysis. The present electron micrographs made with human haemoglobin confirm their prediction in a striking manner.

The authors are indebted to Messrs. M. L. De, S. B. Bhattacherjee and P. K. Ganguli for helpful advice and discussions. The work was supported by grants from the Ministry of Scientific Research and Cultural Affairs, Government of India.

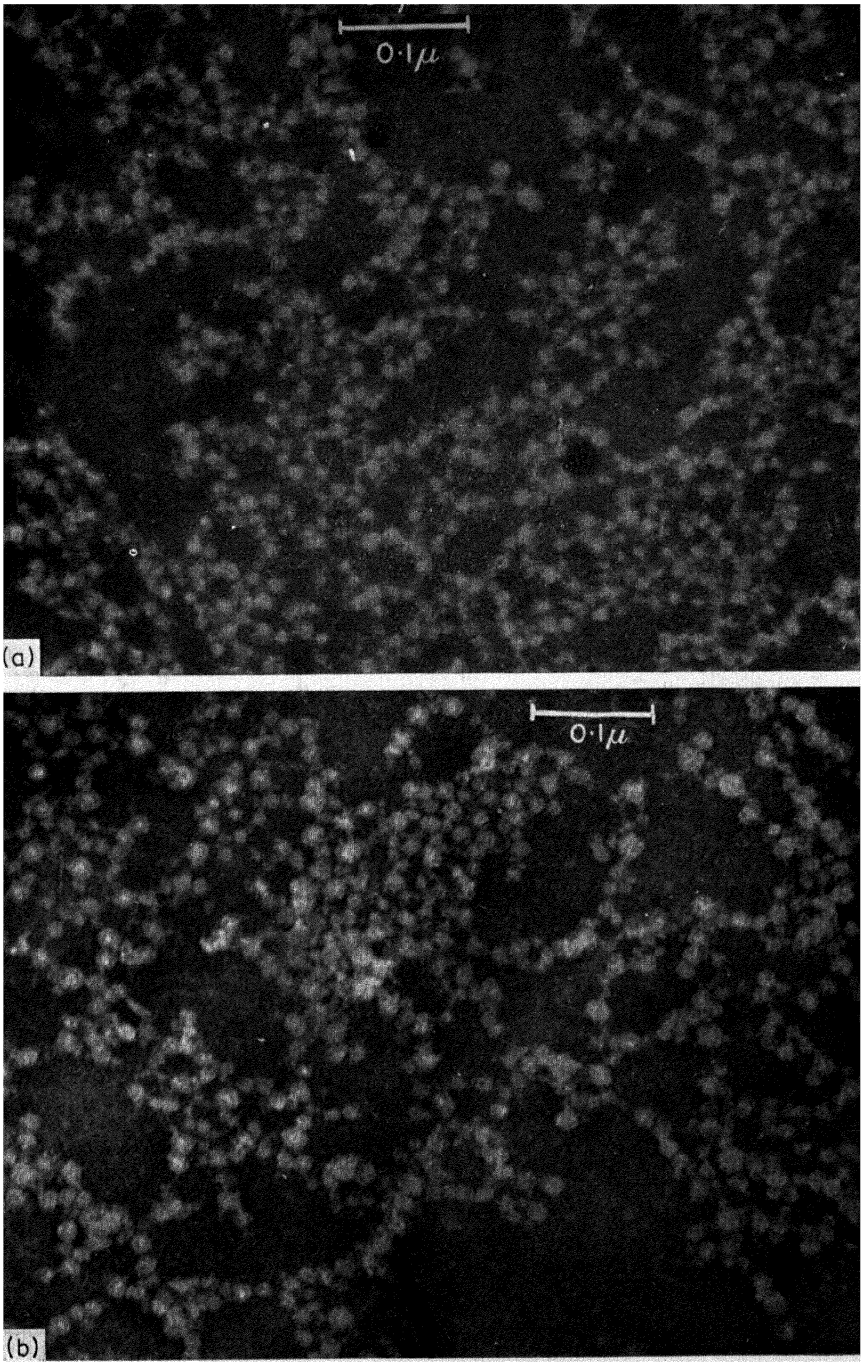


FIG. 2. Electron micrographs of normal human haemoglobin molecules negatively stained with PTA, × 165,000: (a) HbF, (b) HbA.

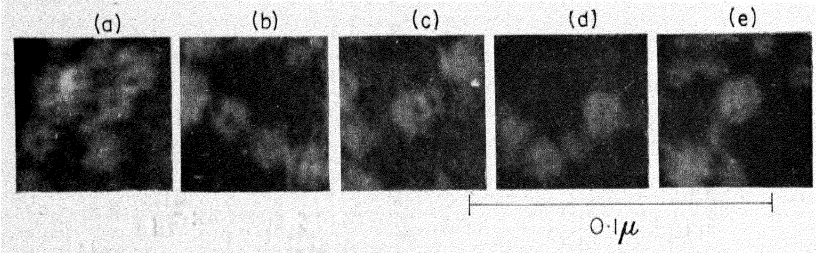


FIG. 3. Electron micrographs of individual haemoglobin molecules, × 400,000. (a), (b) Single molecules with central holes, (c) a pair of molecules, (d) and (e) two or more molecules lying side by side.

References

Beaven, G. H. and Gratzer, W. B. (1959). *J. clin. Path.* **12**, 101.
Chatterji, S. N., Sadhukhan, P. and Chatterjea, J. B. (1961). *J. biophys. biochem. Cytol.* **10**, 113.
Field, E. O. and O'Brien, J. R. P. (1955). *Biochem. J.* **60**, 656.
Horne, R. W. (1961). In *Techniques for Electron Microscopy* (D. Kay, ed.), p. 150. Blackwell, Oxford.
Kegeles, G. and Gutter, F. J. (1951). *J. Amer. chem. Soc.* **73**, 3770.
Kekwick, R. A. and Lehmann, H. (1960). *Nature, Lond.* **187**, 158.
Perutz, M. F., Rossman, M. G., Cullis, A. F., Muirhead, H. and Will, G. (1960). *Nature, Lond.* **185**, 416.
Sadhukhan, P., Das Gupta, N. N., Misra, D. N. and Chatterjea, J. B. (1962). In *Electron Microscopy* (S. S. Breese, Jr., ed.), Vol. 1, p. T2. Academic Press, New York.
Singer, K., Chernoff, A. F. and Singer, L. (1951). *Blood* **6**, 429.
White, J. C. and Beaven, G. H. (1959). *Brit. med. Bull.* **15**, 33.
Wintrobe, M. W. (1961). *Clinical Hematology*, p. 150. Lea & Febiger, Philadelphia (for references to earlier work).

Discussion

J. SEGAL (Humboldt University, Berlin): Two years ago we published a theoretical picture of haemoglobin in which the monomeric molecule appeared as an oval structure with an extracentral haem group. In the global molecule, the monomers were shown to slide one upon another, to an angle of about 13°. Viewed by an electron microscope, such a structure ought to appear as an oval with a slightly blurred boundary at one end and with sharp contours and a central hole at the other. The investigation of Prof. Das Gupta having been conducted without any knowledge of our paper, the conformity of the two results seems very satisfactory.

M. S. NARASINGA RAO (Regional Research Laboratory, Hyderabad): At what pH were your sedimentation studies made?

N. N. DAS GUPTA: It was at about 6·8–7.

D. C. PHILLIPS: The 6 Å resolution model of horse oxyhaemoglobin described by Perutz and his colleagues is believed to show, for the most part, the course of the polypeptide chain, much of which is in the α-helical configuration. Many of the apparent holes and crevices in this model in fact must represent regions filled by side chains in close interaction with one another. It is therefore dangerous to interpret the appearance of the electron micrographs too closely in terms of this model.

SECTION V
Chemical Studies

A Discussion of Methods that have Proved Useful in Research on Ribonuclease

STANFORD MOORE

The Rockefeller Institute, New York, New York, U.S.A.

ABSTRACT

The methods used in the determination of the chemical structure of bovine pancreatic ribonuclease are reviewed briefly under the following headings: 1. Isolation of a homogeneous protein. 2. Amino-acid composition. 3. Amino-terminal residues. 4. Cleavage of disulphide bonds by oxidation or reduction. 5. Separation of peptides yielded by enzymatic hydrolysis. 6. Step-wise degradation of peptides. 7. Pairing of half-cystine residues.

A thorough understanding of the reactions in which proteins participate in living organisms will require, ultimately, knowledge of the detailed structures of many proteins of different origins and different functions. It is only within the last decade or two that chemical techniques have become available which make it possible to consider elucidation of the chemistry of protein molecules. The accumulation of information is coming from laboratories in many parts of the world where methods are being shaped and improved to make the research more expeditious. In reviewing some of the experimental procedures that have proved useful in the determination of the structure of ribonuclease, we are summarizing only a part of a rapidly expanding subject. The results are helping the experimental pace of protein chemistry to keep up with the fascinating new possibilities that the glimpse of the structures of a few proteins is eliciting.

Two of the series of investigations that set the scene for the subsequent work on this subject were the studies that led to determination of the structure of oxytocin by du Vigneaud and his associates (1953), and the elucidation of the structure of insulin by Sanger and his colleagues (1955). The goal of the work to be discussed in this review was the determination of the chemical structure of bovine pancreatic ribonuclease (Fig. 1). This protein is an enzyme, and the structural formula, in two dimensions, forms a starting point for the study of the specific chemical

and physical properties that endow the three-dimensional structure with catalytic activity.

The experiments from this laboratory summarized in the following review are those of C. H. W. Hirs, Darrel H. Spackman, Arthur M. Crestfield, George R. Stark and Derek G. Smyth, in collaboration with William H. Stein and the present reporter.

Within the limits of a brief report, we will discuss the methods with which we have had particular experience; it should be emphasized that the results obtained concurrently by Christian B. Anfinsen and his associates (1961) have formed an integral part of the development of the

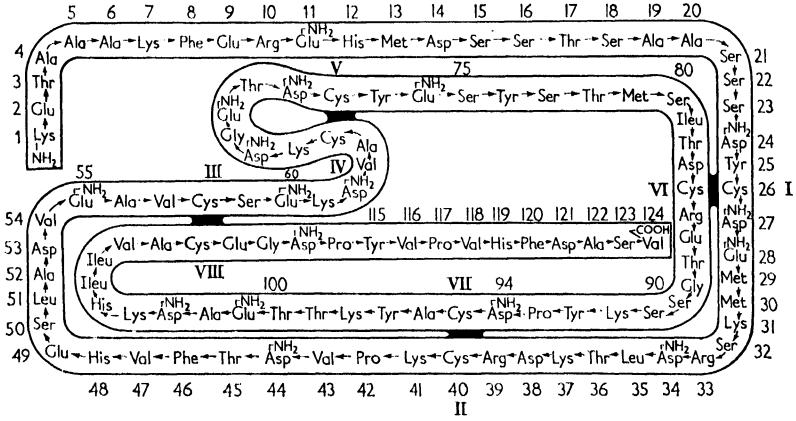


FIG. 1. The sequence of amino-acid residues in bovine pancreatic ribonuclease-A, based upon the experiments of Hirs *et al.* (1960), Spackman *et al.* (1960) and Smyth *et al.* (1962, 1963).

knowledge of the structure of ribonuclease. He and his group have often used different methods, and the diversification thus introduced has been valuable.

1. Isolation of a Homogeneous Protein

A number of the techniques introduced in recent years for the purification of proteins have grown directly or indirectly from the renewal of interest in chromatography stimulated by Martin and Synge and from Craig's studies on counter-current distribution. Their fundamental experiments focused attention on multiplate systems of high resolving power for the separation of a wide variety of organic compounds, including proteins.

(a) *Gel filtration*

Separations on a size basis by filtration through columns of particulate gels have recently been introduced and are proving very useful in protein

chemistry. Porath and Flodin (1959) have employed granular preparations of cross-linked dextran gels in a separation process based on the phenomenon of molecular sieving. Limiting the present discussion to results obtained with ribonuclease, an example of the use of a column of Sephadex G-75 is given in Fig. 2. Crestfield *et al.* (1962) have been able to separate ribonuclease (mol. wt. 14,000) from a dimer (mol. wt. 28,000), which in turn is separated from a tetramer. The aggregates of ribonuclease arose during lyophilization of solutions of the enzyme in 50% acetic

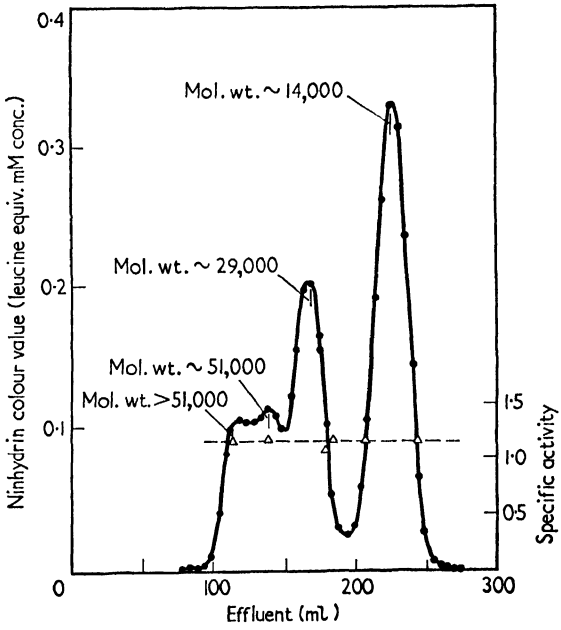


FIG. 2. Separation of ribonuclease-A and its aggregates by gel filtration on a 2 × 143 cm column of Sephadex G-75 with 0·2 M sodium phosphate buffer at pH 6·47 as eluant. (From Crestfield *et al.*, 1962.)

acid; their occurrence emphasizes the way that physical processes sometimes complicate an isolation problem.

A more-tightly cross-linked preparation (Sephadex G-25; water-regain 2·5) was utilized by Porath and Flodin for separating proteins from inorganic salts. Crestfield *et al.* (1963*a*, *b*) have used the gel columns for removing salts, urea and reducing and alkylating agents from ribonuclease and its derivatives. The process has generally been more rapid and more quantitative than dialysis. When a protein or a protein derivative is insoluble in dilute aqueous salt solution, 50% acetic acid has been a useful solvent for gel filtration on Sephadex G-75. Also, protein–protein

and protein–peptide associations are minimized in the strong acetic acid solution; the solvent can be removed by rotary evaporation and dilution prior to lyophilization or by gel filtration against dilute salt followed by ultrafiltration.

A protein preparation that has been successfully freed of material of higher or lower molecular weight should give a single peak by gel filtra-

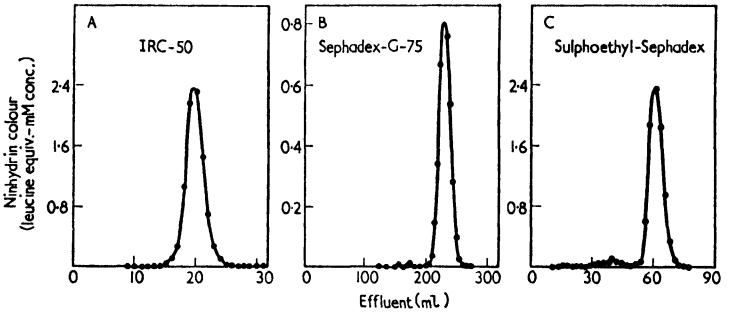


FIG. 3. Behaviour of chromatographically purified ribonuclease on three types of columns. Recovery was complete in each of the three systems. A. Amberlite IRC-50, 0·9 cm × 30 cm column, 0·2 M phosphate, pH 6·47, 30 ml/h, 3 mg load. B. Sephadex G-75, 2 cm × 157 cm column, same buffer as for A, 5 ml/hr, 25 mg load. C. Sulphoethyl–Sephadex (C-25, minus 200 mesh), 0·9 cm × 60 cm column, 1:1 dilution of the 0·2 M phosphate buffer, 5 mg load. (From Crestfield *et al.*, 1963*a*.)

tion, as does purified ribonuclease (Fig. 3 B). (A bibliography on the uses of the Sephadexes of various porosities is available from the manufacturer, AB Pharmacia, Uppsala, Sweden.)

(b) *Chromatography*

The preparation of chromatographically purified ribonuclease-A has recently been re-examined by Crestfield *et al.* (1963*a*). The procedure of Hirs *et al.* (1953) for the chromatography on columns of the carboxylic acid exchanger, Amberlite IRC-50, is applied on a gramme scale to preparations of the enzyme isolated from beef pancreas by salt and solvent precipitation methods patterned after those of Kunitz (1940) and McDonald (1948). The final preparations, if stored with prescribed precautions, are chromatographically homogeneous when examined with an IRC-50 column on an analytical scale (Fig. 3 A). The preparation shows a few per cent of impurity when chromatographed on a column of sulphoethyl–Sephadex C-25 (Fig. 3 C). The use of the dextran gel with a sulphonic acid group as the functional ion introduces an additional adsorbent for protein chromatography. Peterson and Sober (1962)

opened up new possibilities for the chromatography of a wide variety of proteins when they introduced ion exchangers with cellulose as the matrix instead of a synthetic polymethacrylate or polystyrene backbone for the ionic groups. Carboxymethyl (CM)-cellulose and diethylamino (DEAE)-cellulose have proved applicable to many problems in protein chemistry. The availability of cross-linked dextran which has been treated with chloroethylsulphonic acid provides for the first time an exchanger with a carbohydrate backbone and a relatively high density of sulphonic acid groups (capacity, 2·0 meq/g). The results obtained when ribonuclease is chromatographed on this exchanger have been encouraging (Fig. 3 C); the sulphoethyl–Sephadex has proved easier to use than IRC-50 for the chromatography of a streptococcal proteinase (Liu *et al.*, 1963).

2. Amino-acid Composition

Quantitative determination of amino acids bears a relationship to the chemistry of proteins similar to that which elementary analysis bears to the chemistry of simpler organic molecules. An empirical formula for ribonuclease can be expressed in terms of the numbers of each type of amino-acid residue contributing to the 124 residues in the molecule. The degree to which such a calculation gives integral molar ratios for the constituent residues is a function of the purity of the protein and the accuracy of the amino acid analysis. The most accurate results with hydrolysates of ribonuclease have been obtained using ion exchange chromatography on columns of sulphonated polystyrene (Moore *et al.*, 1958) and automatic recording equipment (Spackman *et al.*, 1958) for measurement of the ninhydrin colour (cf. Moore and Stein, 1962). Improvements in the conditions for hydrolysis have recently been introduced by Crestfield *et al.* (1963*b*). Thorough removal of oxygen before sealing the hydrolysis tubes under vacuum is important in minimizing decomposition of amino acids in 6 N HCl at 110°C; rapid removal of HCl by rotary evaporation avoids artifacts that may arise if the acid is removed slowly in a desiccator. The analysis of peptides can be made more rapidly (in 7 hr) on 0·1 μmole scale using a 55 cm column for the neutral and acidic amino acids and an 8 cm column for the basic amino acids (cf Smyth *et al.*, 1963; Stark and Smyth, 1963).

The determination of the number of residues of half-cystine plus cysteine can be made by chromatographic measurement of cysteic acid in a hydrolysate of the performic acid oxidized protein by a recent modification (Moore, 1963) of the procedure of Schram *et al.* (1954) or by determination of *S*-carboxymethylcysteine in a hydrolysate of the

reduced and alkylated protein (Crestfield *et al.*, 1963*b*). The amount of cysteine, as distinct from half-cystine, can be ascertained if the protein is alkylated in 8M urea without prior reduction.

3. Amino-terminal Residues

The research of Sanger on insulin began with his introduction of fluorodinitrobenzene as an end-group reagent (1949), and this reagent was employed throughout most of the work on ribonuclease. The phenylisocyanate method of Edman (1950, 1953) and Eriksson and Sjöquist (1960) has provided an additional procedure for this purpose. Recent observations of Stark *et al.* (1960) on the ease of carbamylation of NH_2-groups in proteins have led Stark and Smyth (1963) to use cyanate for the determination of NH_2-terminal residues. The hydantoins formed can be hydrolysed back to the amino acids and the final measurement can be made with the equipment used for the accurate determination of amino acids. With ribonuclease the method gives 0·94 residue of NH_2-terminal lysine per molecule, and the technique has been applied to a number of proteins (Stark and Smyth, 1963; Liu *et al.*, 1963). Although the yields of serine and threonine are low, the cyanate method provides a useful and convenient addition to the list of techniques for determining NH_2-terminal residues in both proteins and peptides.

There is still no fully satisfactory chemical method for the determination of COOH-terminal residues. Smyth *et al.* (1963) have noted that the results obtained with carboxypeptidase were consistently dependable in the studies on peptides from ribonuclease.

4. Cleavage of Disulphide Bonds by Oxidation or Reduction

An investigation into the structure of a protein molecule frequently requires as an early step the cleavage of the disulphide bonds of cystine residues. When tryptophan is absent, as in insulin (Sanger, 1955) and ribonuclease (Hirs, 1956), oxidation by performic acid has proved satisfactory. If more typical proteins which contain tryptophan are oxidized, however, the sensitivity of the indole ring to oxidation yields a variety of products. Since tryptophan is normally stable towards reducing agents, a number of investigators have studied the cleavage of –S—S– to –SH and subsequent coverage of the sulphydryl groups. In the course of their extremely significant experiments on the regeneration of active ribonuclease from the reduced protein, White (1960) and Anfinsen and Haber (1961) have described conditions for the reduction with mercaptoethanol in 8M urea. The resulting –SH groups can be alkylated by iodoacetic acid; Crestfield *et al.* (1963*b*) have introduced small modifications in the

alkylation step, to avoid the possible side reactions with methionine, lysine, histidine and tyrosine residues, and have successfully applied this modified procedure to a number of proteins.

The use of a reduced and carboxymethylated protein as a starting product for a structural study presents some problems on which there is little experience at the present time. Attention has to be given to the protection of the thio-ether sulphur of methionine from oxidative changes during chromatographic separation of the peptides, and probably also to the indole group of tryptophan.

5. Separation of Peptides Yielded by Enzymatic Hydrolysis

In the studies on oxidized ribonuclease, the peptides obtained by cleavage of the oxidized protein by trypsin, chymotrypsin, and pepsin were separated mainly by chromatography on Dowex 50-X2 by manual procedures (cf. Hirs, 1960). Smyth *et al.* (1963) have used higher loads (300 mg of peptide mixture on columns 0·9 cm in diameter) and have also found it useful to employ conditions described by Eaker† for the initial separation of the peptides into groups by gel filtration on Sephadex G-25, followed by ion-exchange chromatography. Sephadex G-25 (150 × 0·9 cm column) equilibrated with 50% acetic acid provides a convenient means for the de-salting and further purification of peptides obtained from buffered ion-exchange columns. Paper electrophoresis furnishes a further test for homogeneity.

Columns of the anion exchanger Dowex 1-X2 were employed by Rudloff and Braunitzer (1961) for the separation of peptides obtained in studies on the structure of haemoglobin. Konigsberg and Hill (1962) have found such columns very useful in their studies of the same protein. Volatile pyridinium acetate or formate and *N*-ethylmorpholinium acetate buffers have been used. These systems proved advantageous for the recent isolations of the four large peptides from tryptic hydrolysates of ribonuclease by Eaker† and Smyth *et al.* (1963).

Only a beginning has been made in the use of automatic recording techniques for peptide chromatography. The 15 cm column of Amberlite IR-120 on the amino-acid analyser has proved effective in some instances for peptide separations. An example is shown in Fig. 4 in which analyses of tryptic hydrolysates of performic acid oxidized and reduced-alkylated ribonuclease are compared (Crestfield *et al.*, 1963*b*). The precise comparison of peptide patterns from protein samples of different origins is an aim of the research on recording techniques; further study is required on both ion exchangers and eluant systems.

† Doctoral dissertation by David L. Eaker on "Structural and Enzymatic Studies with des-Lysyl Forms of Bovine Pancreatic Ribonuclease", The Rockefeller Institute, June 1962.

6. Step-wise Degradation of Peptides

The phenylthiocarbamyl method introduced by Edman has been the key technique employed in the elucidation of the sequence of the amino-acid residues in purified peptides obtained from ribonuclease. Recent experience of Smyth *et al.* (1963) in the use of the reaction when the process is followed by amino acid analysis of the residual peptide (the subtractive method, as we have employed it) has explained the reasons why the first experiments from this laboratory (Hirs *et al.*, 1960) contained errors which have required revision (Smyth *et al.*, 1962, 1963). When the cyclization is carried out in glacial acetic acid/HCl at 100°C there are complicating side reactions which can lead to misleading conclusions. Fortunately, these difficulties are completely avoided under the modified conditions recently recommended by Konigsberg and Hill (1962) in which the cyclization is carried out at 25°C in anhydrous trifluoroacetic acid and the residual peptide is purified by passage through a short column of Dowex 50-X2. Smyth *et al.* (1963) occasionally found it advantageous to purify the residual peptide by precise chromatography on a long ion-exchange column before continuing the step-wise degradation.

The allocations of the amide groups of glutamine and asparagine residues were made from the behaviour of the peptides upon paper electrophoresis (Smyth *et al.*, 1962, 1963; apparatus of Crestfield and Allen, 1955) and the results of chromatographic analyses for glutamine and asparagine (Hirs *et al.*, 1960) performed after the peptides had been hydrolysed by leucine aminopeptidase.

7. Pairing of Half-cystine Residues

A complete knowledge of the sequence of the residues in performic acid oxidized ribonuclease still leaves open the question of the pairings of the eight half-cystine residues in the native enzyme. The important first step employed by Spackman *et al.* (1960) in solving this problem was the hydrolysis of the protein (with the –S—S– bonds intact) by pepsin. The enzyme acts optimally at pH 2, the pH value at which the rate of the disulphide interchange reaction is at a minimum. Brief subsequent hydrolysis by trypsin and chymotrypsin led to a mixture of peptides from which those containing cystine could be separated and oxidized with performic acid; the peptides comprising each pair of cysteic acid-containing peptides were separated, analysed for amino acid content, and identified by reference to the sequence of residues in the oxidized protein.

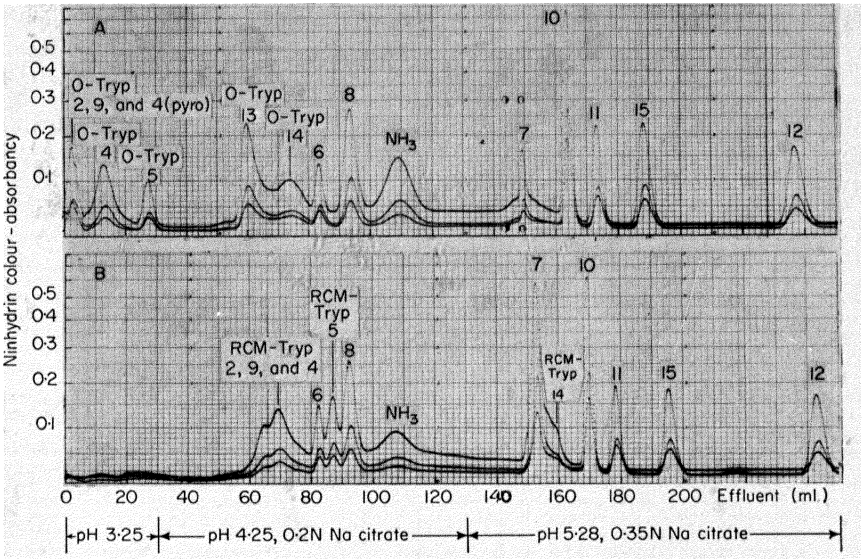


Fig. 4. Comparison of the chromatographic separations of peptides obtained by 2-hr tryptic hydrolysis (cf. Hirs *et al.*, 1960) of performic-acid oxidized and reduced-carboxymethylated ribonuclease. The 15 cm column of Amberlite IR-120 on the amino-acid analyser was used. Load: peptides from 5 mg of protein. Flow rate: 30 ml/hr. Temperature: 50°. The peptides containing cysteic acid residues move more rapidly on the acidic resin than their counterparts containing carboxymethylcysteine residues. (From Crestfield *et al.*, 1963*b*.)

References

Anfinsen, C. B. and Haber, E. (1961). *J. biol. Chem.* **236**, 1361.
Anfinsen, C. B. and White, F. H., Jr. (1961). In *The Enzymes* (P. D. Boyer, H. Lardy, and K. Myrbäck, eds.), Vol. 5, p. 95. Academic Press, New York.
Crestfield, A. M. and Allen, F. W. (1955). *Analyt. Chem.* **27**, 422.
Crestfield, A. M. Stein, W. H. and Moore, S. (1962). *Arch. Biochem. Biophys.* Suppl. 1, 217.
Crestfield, A. M., Stein, W. H. and Moore, S. (1963*a*). *J. biol. Chem.* **238** (in press).
Crestfield, A. M., Moore, S. and Stein, W. H. (1963*b*). *J. biol. Chem.* **238** (in press).
du Vigneaud, V., Ressler, C. and Trippett, S. (1953). *J. biol. Chem.* **205**, 949.
Edman, P. (1950). *Acta chem. scand.* **4**, 283.
Edman, P. (1953). *Acta chem. scand.* **7**, 700.
Eriksson, S. and Sjöquist, J. (1960). *Biochim. biophys. Acta* **45**, 290.
Hirs, C. H. W. (1956). *J. biol. Chem.* **219**, 611.
Hirs, C. H. W. (1960). *J. biol. Chem.* **235**, 625.
Hirs, C. H. W., Moore, S. and Stein, W. H. (1953). *J. biol. Chem.* **200**, 493.
Hirs, C. H. W., Moore, S. and Stein, W. H. (1960). *J. biol. Chem.* **235**, 633.
Konigsberg, W. and Hill, R. J. (1962). *J. biol. Chem.* **237**, 2547.
Kunitz, M. (1940). *J. gen. Physiol.* **24**, 15.
Liu, T.-Y., Neumann, N. P., Elliott, S. D., Moore, S. and Stein, W. H. (1963). *J. biol. Chem.* **238**, (in press).
McDonald, M. R. (1948). *J. gen. physiol.* **32**, 39.
Moore, S. (1963). *J. biol. Chem.* **238** (in press).
Moore, S. and Stein, W. H. (1962). In *Methods in Enzymology* (S. P. Colowick and N. O. Kaplan, eds.), Vol. VI. Academic Press, New York.
Moore, S., Spackman, D. H. and Stein, W. H. (1958). *Analyt. Chem.* **30**, 1185.
Peterson, E. A. and Sober, H. A. (1962). In *Methods in Enzymology* (S. P. Colowick and N. O. Kaplan, eds.), Vol. V, p. 3. Academic Press, New York.
Porath, J. and Flodin, P. (1959). *Nature, Lond.* **183**, 1657.
Rudloff, V. and Braunitzer, G. (1961). *Hoppe-Seyl. Z.* **323**, 129.
Sanger, F. (1949). *Biochem. J.* **45**, 563.
Sanger, F. (1955). *Bull. Soc. Chim. biol., Paris* **37**, 23.
Schram, E., Moore, S. and Bigwood, E. J. (1954). *Biochem. J.* **57**, 33.
Smyth, D. G., Stein, W. H. and Moore, S. (1962). *J. biol. Chem.* **237**, 1845.
Smyth, D. G., Stein, W. H. and Moore, S. (1963). *J. biol. Chem.* **238** (in press).
Spackman, D. H., Stein, W. H. and Moore, S. (1958). *Analyt. Chem.* **30**, 1190.
Spackman, D. H., Stein, W. H. and Moore, S. (1960). *J. biol. Chem.* **235**, 648.
Stark, G. R. and Smyth, D. G. (1963). *J. biol. Chem.* **238** (in press).
Stark, G. R., Stein, W. H. and Moore, S. (1960). *J. biol. Chem.* **235**, 3177.
White, F. H., Jr. (1960). *J. biol. Chem.* **235**, 383.

Discussion

K. S. V. SAMPATHKUMAR: You have mentioned that there is no reaction of iodoacetate with histidine residues in ribonuclease when the protein is dissolved in 8 M urea at pH 5·5. Is there any reaction at other pH values?

S. MOORE: The same result was obtained at pH 8.

S. OCHOA: When pancreatic ribonuclease of a different species is studied, you pointed out that certain amino acid residues are replaced and that those residues are non-essential. Do these ribonucleases have the same specific activity?

S. MOORE: The pancreatic ribonuclease isolated from sheep by Aquist and Anfinsen had the same specific activity as the bovine enzyme.

G. J. S. RAO (Indian Institute of Science, Bangalore): You said that two histidines are important for the activity of ribonuclease. Weil and Seibles had shown that photo-oxidation of histidine in ribonuclease reduces activity; which of the two histidines that you mentioned corresponds to the one altered by Weil and Seibles?

S. MOORE: So far as I know, the 1955 studies of Weil and Seibles, which provided the first evidence that histidine was probably crucial, have not been repeated on intact ribonuclease in order to determine which residue is altered, now that the covalent structure is known.

G. J. S. RAO: You also mentioned that the sequence of amino acid residues determines the pairing of half-cystine residues in the active protein. But in the case of lysozyme, the yield of active material after reduction and reoxidation, as observed by White and by Isemura, was considerably less.

S. MOORE: What was the yield?

G. J. S. RAO: I think that it was about 47%.

S. MOORE: White obtained high yields (over 90%) in the experiments with ribonuclease by working with very dilute solutions. Perhaps the intermolecular condensations are more difficult to avoid with lysozyme.

M. S. NARASINGA RAO (Regional Research Laboratory, Hyderabad): In addition to ribonuclease, there are other enzymes known which, on degradation, also exhibit enzymatic activity. Does the rest of the protein molecule have any specific role in enzymatic activity?

S. MOORE: In the case of papain, Smith and associates were able to degrade the native enzyme, with about 200 residues, to an active fragment that is about the same size as ribonuclease. The smallest enzyme that I know of is thrombin, which Scheraga and associates have shown to have a molecular weight of 8000. It looks, at the moment, as if 80–100 residues may be needed to give the necessary three-dimensional structure for a protein to possess a highly specific catalytic action; the so-called "excess baggage" in the natural forms of some enzymes probably does play a role in increasing the stability of the molecules.

Identification and Sequence of Amino-acid Residues around the Thiol of the Active Site of Carboxypeptidase A

K. S. V. Sampath Kumar, Kenneth A. Walsh, Jean-Pierre Bargetzi and Hans Neurath

Department of Biochemistry, University of Washington, Seattle, Washington, U.S.A.

ABSTRACT

Bovine pancreatic carboxypeptidase A is a zinc metalloenzyme, zinc being bound in the native enzyme to the single sulphydryl group and to a nitrogen atom which has been identified with the α-amino group at the N-terminus of the protein (Vallee *et al.*, 1960; Coombs and Omote, 1962). After removal of zinc, the free thiol group reacts with certain sulphydryl reagents such as silver, *p*-mercuribenzoate or ferricyanide, but reaction with typical alkylating agents such as iodoacetate requires disruption of the tertiary structure of the apoenzyme. Prior treatment with reducing agents such as β-mercaptoethanol or sodium borohydride exposes two thiol groups, the cysteinyl of the active site and a second cysteinyl which does not exist as such in the native enzyme and which can be selectively alkylated without impairment of catalytic activity of the enzyme. Selective reaction can be accomplished by the addition of specific inhibitors, such as β-phenylpropionate which protects the thiol of the active site. These observations were utilized to label specifically the zinc binding site of the enzyme and to determine the amino acid sequence around the thiol of the active site and around the second artifactitious thiol. These findings are documented in detail.

Carboxypeptidase A is an exopeptidase of the pancreas and it specifically catalyses the cleavage of certain amino acids from the carboxyl terminal portion of a variety of peptides and proteins (Neurath, 1960). The purified crystalline enzyme has a molecular weight of 34,300, whereas its precursor, procarboxypeptidase A, has a molecular weight of 90,000. Some of the characteristic properties of this zymogen are summarized in Table I (Keller *et al.*, 1956, 1958). This zymogen, like others of the bovine pancreas, is activated by trypsin; however, procarboxypeptidase A is unique since, on activation, it gives rise not to one but to two enzymes—carboxypeptidase A and an endopeptidase which hydrolyses the synthetic substrate, acetyl-L-tyrosine ethyl ester. The time course of these activation reactions is indicated in Fig. 1, which shows that the endopeptidase activity appears rapidly, through the action of low concentrations of trypsin (1:1000) at 0°, but the carboxypeptidase activity appears only after prolonged activation at 37° in the presence of much

higher concentrations of trypsin (1:10). More detailed studies of the activation process have revealed that the endopeptidase inherent in procarboxypeptidase is itself necessary for the conversion of the zymogen to carboxypeptidase.

TABLE I. Properties of procarboxypeptidase A

Molecular weight	90,000
$S^0_{20,w}$	5·9 S
$D_{20,w}$	$6{\cdot}23 \times 10^{-7}$ cm^2/sec
Isoelectric point (0·2 ionic strength)	$< 4{\cdot}5$
Nitrogen	15·9%
$E^{1\%}_{280}$	19
N-Terminal groups	½ Cystine Lysine Aspartic acid (or the amide)

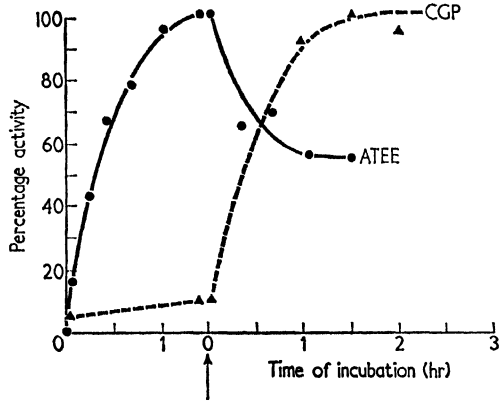


FIG. 1. The activation of procarboxypeptidase (from Keller *et al.*, 1958). The appearance of enzymic activity towards acetyltyrosine ethyl ester (the solid line) and towards carbobenzoxyglycylphenylalanine (the dotted line) upon incubation of procarboxypeptidase with a low concentration of trypsin. At the time indicated by the arrow, the trypsin concentration was raised by a factor of 100 and the temperature raised from 0° to 37°.

The role of the endopeptidase in this process is illustrated in Fig. 2. Recent studies in our laboratory have provided evidence that procarboxypeptidase is in fact an aggregate of three protein sub-units, one of which is seemingly the immediate precursor of carboxypeptidase A; the second one gives rise to the endopeptidase, while the identity and the biological function of the third sub-unit are not known (Brown *et al.*, 1961).

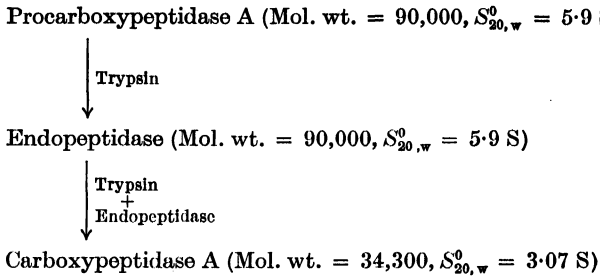


FIG. 2. Activation of procarboxypeptidase A.

Carboxypeptidase A may be isolated as a homogeneous protein by any of several procedures either after activation of crude extracts of whole pancreatic tissues or after activation of the purified zymogen. In either case, the protein can be crystallized and is found to be homogeneous by the usual criteria of protein chemistry. The molecular characteristics of the enzyme are summarized in Table II (Neurath, 1960). The protein is a single polypeptide chain with a molecular weight of 34,300, containing only one identifiable N-terminal group and one C-terminal group. It has a highly helical structure as demonstrated by the low specific laevorotation, $[\alpha]_D$, and is a characteristic euglobulin—being soluble in dilute salt solutions but completely insoluble in water. An important feature of the enzyme is its content of 1 g-atom of zinc per mole (Vallee and Neurath, 1955)—which, as will be outlined presently—plays an essential role in the function of carboxypeptidase as an enzyme.

TABLE II. Some properties of carboxypeptidase A

Molecular weight	34,300
Number of amino-acid residues	301
Metal content	1 g-atom of Zn per mole
N-Terminal	Asparagine
C-Terminal	Asparagine
Isoelectric point	
Ionic strength 0·2	6·0
Ionic strength 0·3	5·6
$[\alpha]_D^{0°}$	−18°

The amino acid composition of purified carboxypeptidase A is summarized in Table III (Bargetzi *et al.*, 1963). The protein consists of 301 amino-acid residues based on a molecular weight of 34,300; and it is strikingly deficient in sulphur-containing amino acids. In view of the involvement of a thiol, viz. a cysteinyl residue, in the binding of zinc (*vide infra*), special emphasis was placed on the study of the distribution

of sulphur atoms in the proteins presented in Table IV (Walsh *et al.*, 1962). Clearly, there are five sulphur atoms: three of these are found in methionine residues, whereas the other two appear after performic acid oxidation as two residues of cysteic acid. Also, reduction of the protein by

TABLE III. Amino acid composition of carboxypeptidase A

	Residues per molecule		Residues per molecule
Aspartic acid	27	Methionine	3
Threonine	26	Isoleucine	20
Serine	32	Leucine	23
Glutamic acid	25	Tyrosine	19
Proline	10	Phenylalanine	16
Glycine	22·5	Lysine	15
Alanine	19	Histidine	8
"Half cystine"	2	Arginine	10
Valine	16	Tryptophan	8

TABLE IV. Sulphur distribution per molecule of carboxypeptidase A

Methods	"Half-cystine" (residues)	Methionine (residues)	Total sulphur (atoms)
Schöniger ignition			5 (4·94, 4·82)
Performic oxidation	2·0 ± 0·2	3·0 ± 0·2	
ME† + Iodoacetate	1·9	2·9	
ME† + Iodoacetamide	2·0	2·6	
ME† + NEM	(1·6)‡	3·1	
ME† + DDPM	(2)‡	3·0	
BAL + Iodoacetamide	1·6	2·9	
Acid hydrolysis	0–2·0	3·0	

† β-Mercaptoethanol.

‡ Using an approximate correction factor for conversion to *S*-2-succinylcysteine on hydrolysis of protein alkylated with *N*-ethylmaleimide (NEM) or its coloured derivative, DDPM.

β-mercaptoethanol (ME) or BAL, followed by alkylation with iodoacetate, iodoacetamide, *N*-ethylmaleimide or DDPM (diethylaminodinitrophenylmaleimide) yielded in all cases two residues of *S*-alkylcysteine. In all cases, the residues indicated were determined after acid hydrolysis by quantitation with a Spinco amino acid analyser by the technique of Spackman, Stein and Moore (1958). While it may thus appear that in carboxypeptidase A the sulphur is distributed between

three methionine and either two cysteine or one cystine residues, this is, in fact, not the case since the native protein contains one and only one cysteinyl group. Before considering this apparent anomaly, the enzymic specificity of the enzyme and certain characteristics of the active centre will be discussed.

The narrow specificity of the highly purified crystalline enzyme is briefly summarized in Table V. The structural features absolutely necessary for catalysis are as follows (Neurath and Schwert, 1950). (1) There must be a free carboxyl group. This absolute specificity has made possible the use of carboxypeptidase as a specific reagent for the elucidation of carboxyl terminal groups of peptides and proteins. (2) The

TABLE V. Substrate specificity of carboxypeptidase A

Type	Example
N-Acyl peptide R″CONHCHCONHCHCOOH R′ R	Carbobenzoxyglycyl-L-phenylalanine
N-Acyl amino acid R′CONHCHCOOH R	*N*-Chloroacetyl-L-phenylalanine
O-Acyl hydroxy acid R′COOCHCOOH R	Hippuryl-β-L-phenyllactic acid

carboxyl group must be alpha to an amino or hydroxyl group, and this residue must be of the L-configuration. (3) The amino group of the adjacent residue should not be free. In addition, the catalysis is more rapid if the C-terminal residue is aromatic.

Competitive inhibitors of the enzyme are structural analogues of the substrates in two respects: they contain both a free carboxyl group and an aromatic ring. Examples of these inhibitors are given in Table VI, which also summarizes values for the enzyme-inhibitor dissociation constant, K_i. As is evident, β-phenylpropionate is the most effective agent, and the inhibitory effect of this compound towards carboxypeptidase A is approximately 100 times greater than on chymotrypsin. As will be discussed later, the inhibition by β-phenylpropionate has been used to advantage in the elucidation of the structural features of the active site of carboxypeptidase.

TABLE VI. Competitive inhibitors for carboxypeptidase A

Inhibitor	K_i 10^{-4} M
Phenylacetic acid	3·9
β-Phenylpropionic acid	0·62
γ-Phenylbutyric acid	11·3
Indoleacetic acid	0·78
β-Indolepropionic acid	5·5
α-Indolebutyric acid	33
β-Cyclohexylpropionic acid	20

The first step toward the localization of the active centre of the enzyme came in 1955, when Vallee and Neurath demonstrated that this protein contains 1 g-atom of zinc per mole of the protein; and provided proof that the enzyme conforms to the operational criteria of a metallo-enzyme. The metal can be removed by dialysis against buffers of pH below 5·5, or at neutral pH by dialysis against 1,10-phenanthroline. The loss of enzyme activity is directly proportional to the loss of the metal, as shown in Fig. 3 (Vallee *et al.*, 1958). Clearly, the zinc is an essential constituent of the native enzyme and its presence is necessary for the operation of its catalytic mechanism.

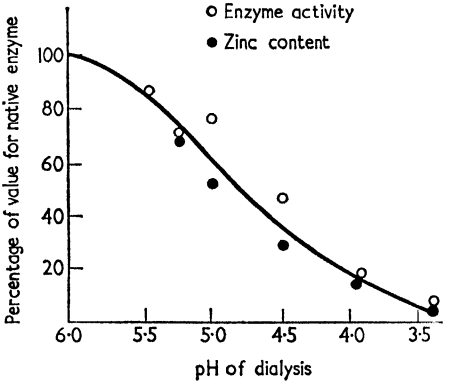


FIG. 3. The loss of zinc and peptidase activity from carboxypeptidase in acid solution (from Vallee *et al.*, 1958).

The physical properties and chemical reactivities of the metal ions, firmly incorporated into the structure of the native enzymes, afford different avenues of approach for the delineation of the active catalytic sites of metalloenzymes. Such approaches towards the definition of the

active catalytic site of carboxypeptidase have been and are being strenuously pursued by Dr. Vallee and his collaborators in Boston. To bring the present paper in proper perspective, some of their work along these lines will be briefly summarized.

The removal of zinc from carboxypeptidase A with concomitant loss of activity is fully reversible. Furthermore, the resultant apoenzyme can be made to recombine with various other metals, particularly those of the first transition series and group II B elements (Coleman and Vallee, 1961). The various metals are shown in Table VII to differ strikingly in the degree to which they restore enzymic activity. Even qualitative differences in specificity appear. Thus, the cobalt, nickel and

TABLE VII. Peptidase and esterase activities of metallo-carboxypeptidases (Coleman and Vallee, 1961)

[(CPD)Me]	Peptidase activity (C)	Esterase activity ($k \times 10^3$)
	Substrate = CGP	Substrate = HPLA
Zn	7·5	1·15
Co	12·0	1·10
Ni	8·0	1·00
Mn	0·6	0·40
Cu	0	0
Hg	0	1·34
Cd	0	1·75
Pb	0	0·60

manganese enzymes are active towards both peptides and ester substrates, while the mercury and cadmium enzymes are ineffective as peptidases, but effective as esterases. The copper protein is inactive towards both tryps of substrates. In all the cases, only one mole of the metal is bound to each mole of the apoenzyme.

The lack of reactivity of the cadmium and mercury carboxypeptidases towards peptides, and the ease of incorporating different metals into the apoenzyme, have aided in developing approaches towards a study of the binding sites on the protein irrespective of the catalytic process. Thus, it was shown by Coleman and Vallee (1962*a*) that peptides are bound to the apoenzyme at the active centre since they prevent the access of zinc to the protein, whereas esters are ineffective and do not bind to the apoenzyme. The binding sites apparently exist even in the inactive copper protein, since the substrates for carboxypeptidase prevent the removal of the metal from the copper protein. These, and other related studies involving the use of specific synthetic substrates, are preparing the way

for the fuller understanding required to describe in definite terms the mechanism for the binding of substrates and their cleavage by the enzyme.

Zinc and cobalt enzymes are interconvertible. Dialysis of cobalt enzyme against solutions of zinc results in replacement of the metal and *vice versa* (Coleman and Vallee, 1961). Manganese, cobalt and nickel enzymes undergo measurable dissociation under specific conditions of dialysis. From equilibrium dialysis measurements and isotope exchange studies, the stability constants of the various metallocarboxypeptidases have been calculated as summarized in Table VIII. There is good correlation between the stability constants derived for the various metalloenzymes and those for nitrogen–metal–sulphur ligands. Thus, Vallee

TABLE VIII. Stabilities of metallocarboxypeptidases and of metal chelates of known structure (Coleman and Vallee, 1961)

	Log K for [(CPD)Me]		Log K_1 for Me chelates		
Metal	Apparent	Corr. for Cl^-, buffer	Nitrogen–Sulphur	Nitrogen–Oxygen	Nitrogen–Nitrogen
Mn^{++}	5·6	5·6	4·1	3·4	2·7
Co^{++}	5·8	7·0	7·7	5·2	5·9
Ni^{++}	5·7	8·2	9·9	6·2	7·7
Cu^{++}	5·1	10·6	—	8·6	10·7
Zn^{++}	8·3	10·5	10·2	5·2	5·7
Cd^{++}	7·9	10·8	11·0	4·8	5·5
Hg^{++}	6·7	21	22·0	10·3	12·0

has pointed out that the order and magnitude of these constants strongly suggest that the metal in carboxypeptidase is bound to a thiol—obviously a cysteine residue—and to a nitrogen. Titration data with the apoenzyme show that two hydrogen ions are displaced by zinc from the metal-binding site, with pK values of 7·7 and 9·1, respectively, compatible with an involvement of an α-amino group and a thiol. Furthermore, the formation of the cobalt enzyme is accompanied by the appearance of a red colour with an absorption maximum at 530 mμ, consistent again with a metal to sulphur ligand (Coleman and Vallee, 1960).

Additional evidence for the involvement of a thiol in the binding of zinc comes from the differential reactivity of site-specific sulphydryl reagents with the metal-containing and metal-free enzyme. In Table IX, the data of Vallee, Coombs and Hoch are summarized, and these show that one mole of apoenzyme reacts with one mole of silver, *p*-mercuribenzoate or ferricyanide, indicating one reactive sulphydryl group; in

contrast, the native, metal-containing enzyme does not react with these reagents. Furthermore, additions of mole fractions of zinc to the zinc-free enzyme result in commensurate decrease of the mole fraction of —SH groups titratable with silver, as shown in Table X. The sum of the number of g-atoms of zinc and the number of sulphydryl groups titrated per mole is equal to one. *p*-Mercuribenzoate also blocks the titrability with silver.

TABLE IX. Reaction of native and zinc-free carboxypeptidase with sulphydryl reagents (Vallee, Coombs and Hoch, 1960)

Reagent	Native enzyme	Metal-free enzyme
	(Moles reagent per mole protein)	
Silver	0	0·91
p-Mercuribenzoate	0	0·95
Ferricyanide	0	1·22

TABLE X. Complementarity of Ag^+ titrable –SH groups and zinc content of carboxypeptidase A (Vallee, Coombs and Hoch, 1960)

Zn^{++} Content g-atoms/mole	–SH Titrated (moles/mole)	Σ (SH + Zn^{++})
0	0·91	0·91
0·07	0·84	0·91
0·27	0·63	0·90
0·57	0·42	0·99
0·77	0·35	1·12
1·00	0	1·00

The well-known reactivity of cysteine or cysteinyl peptides with alkylating agents has resulted in their wide application to the detection of cysteinyl residues in proteins. Yet, neither iodoacetate, iodoacetamide, *N*-ethylmaleimide, nor DDPM reacts readily with the thiol group of the metal-free apoenzyme, even after treatment with denaturing agents, such as urea, sodium dodecyl sulphate, or the non-ionic detergent BRIJ-35 (Walsh *et al.*, 1962). Thus, the obvious approach to locate the active site thiol by direct alkylation of the metal-free enzyme cannot be applied in this case. However, the data in Table XI indicate that if carboxypeptidase was first heat-denatured to render it susceptible to proteolytic degradation, and then digested with chymotrypsin, the thiol group could be readily alkylated, provided that the metal was first

removed from the digest by the addition of *o*-phenanthroline. Thus, only under conditions of extensive disruption of the structure of the apoenzyme can the thiol group of the active site of carboxypeptidase be made to react with alkylating agents in a manner characteristic of -SH groups of model compounds of simpler structure.

Although an alkyl label could be introduced onto the zinc-binding thiol in this degraded enzyme, the advantage of correlating the extent of introduction of the label with the disappearance of enzyme activity was lost. To permit this correlation, an approach was sought whereby the label could be introduced into the intact protein.

TABLE XI. Reaction of iodoacetamide with carboxypeptidase A (Walsh *et al.*, 1962)

Enzyme	Treatment prior to alkylation	CM-Cys (residues/molecule)
Native	—	0
Metal-free	—	0
Heat-denatured	Chymotryptic digestion	0
Heat-denatured	Chymotryptic digestion + *o*-phenanthroline	0·93
Heat-denatured	Chymotryptic digestion + β-mercaptoethanol	1·93

In the determination of the distribution of sulphur atoms in carboxypeptidase (Table IV) it became clear that treatment of the enzyme with β-mercaptoethanol or BAL followed by alkylating agents resulted in the formation of two *S*-alkylcysteines. The reducing agents render the thiol of the active site available for alkylation; but, in this process, a second sulphur is converted to a form which is analytically characterized as a cysteine residue. Presumably, β-mercaptoethanol, besides removing the zinc from the nitrogen–sulphur ligand and exposing the cysteine, modified the protein in such a way as to make a second thiol available for alkylation. The problem, how to put a label at the active site thiol, resolves operationally into two aspects—one, introduction of a second thiol group into the protein molecule by treating the enzyme with a reducing agent and two—subsequent differentiation of this newly formed thiol from the one that binds the zinc in the native enzyme. To this end, advantage was taken of the enzymatic specificity of carboxypeptidase A by exposing it to a competitive inhibitor during reduction. This prevents the removal of the metal, and yet permits the reduction of the second sulphur to a cysteine residue. In Fig. 4, reduction with mercaptoethanol in the absence of a competitive inhibitor,

β-phenylpropionate (βPP) results in a rapid loss of the activity of the enzyme, and the production of two carboxymethylcysteines upon alkylation, whereas in the presence of βPP, only 6% of the activity was lost, yet 0·4 residues of carboxymethylcysteine were formed. This protection of the enzyme by βPP from inactivation by mercaptoethanol increases with increasing concentrations of βPP. When this experiment was repeated on a metal-free enzyme, it was found that βPP did not

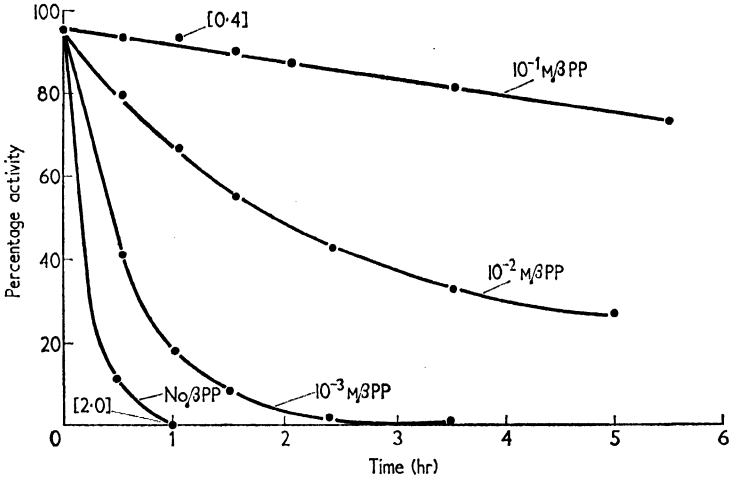


FIG. 4. Inactivation of carboxypeptidase A in 4 M urea by mercaptoethanol. Protection by β-phenylpropionate (βPP). The reduction was carried out at pH 8 at 0° with 0·06 molar β-mercaptoethanol. The protein was then alkylated with an excess of iodoacetamide. The figures in square brackets refer to the number of equivalents of carboxymethylcysteine formed per mole of protein. Relative activities were measured using hippurylphenyllactate as substrate. These data are partly presented in Walsh *et al.* (1962).

protect the enzyme from mercaptoethanol reduction, in keeping with the observations of Coleman and Vallee (1962*b*) that βPP is not bound to the metal-free enzyme.

The protective effect of BPP takes place in the concentration range where the inhibitor is bound tightly. This may be inferred from the position of the inhibition constant in Fig. 5, since more than 90% of the enzyme must be in the form of the enzyme–inhibitor complex before it protects carboxypeptidase significantly from mercaptoethanol.

In Figs. 4 and 5, the effectiveness of βPP was described in the presence of 4 M urea. The same general picture is seen in Fig. 6 in the absence of urea. Again, βPP protects carboxypeptidase against mercaptoethanol reduction. It offers this protection most effectively at pH 7·2, but the

reduction takes place most quickly at pH 9·0. pH 8·0, then, seems a reasonable compromise where selective alkylation can be obtained. In this way, it is possible to obtain 0·5 residues of *S*-alkylcysteine without

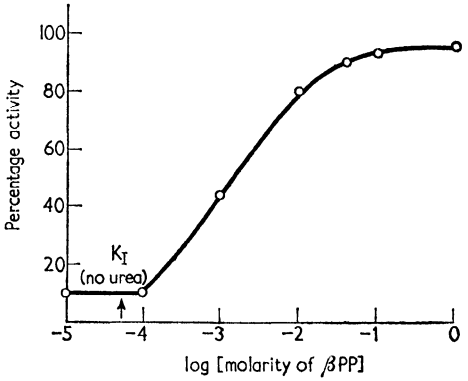


FIG. 5. Protection of carboxypeptidase A by β-phenylpropionate (βPP) against reduction by mercaptoethanol in urea. The experimental conditions were the same as in Fig. 4.

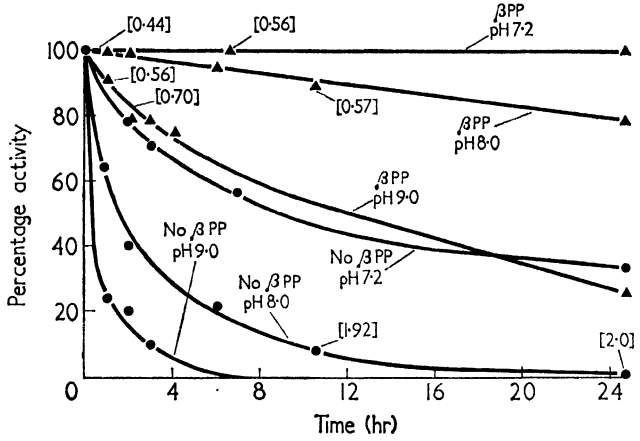


FIG. 6. Inactivation of carboxypeptidase by mercaptoethanol in the absence of urea. The enzyme was treated with 0·33 M β-mercaptoethanol in the presence or absence of 0·1 M β-phenylpropionate (βPP), at the indicated pH values, at 0° C. Inserts indicate the number of equivalents of carboxymethylcysteine formed per mole of carboxypeptidase upon addition of an excess of iodoacetamide to stop the reduction.

a proportional loss of enzymic activity. These data are summarized in Table XII which shows how, in a control reduction with mercaptoethanol in the absence of βPP, two *S*-alkylcysteine residues are formed with

complete loss of enzymatic activity. In the presence of βPP, about 0·5 residues of *S*-alkylcysteine can be obtained without significant loss of enzyme activity. When, in these experiments, β-mercaptoethanol is replaced by sodium borohydride, the formation and subsequent alkylation of the second thiol becomes quantitative and fully selective, as is indicated in the last line by the fact that the mono-alkylprotein retains full enzymic activity. Thus, the second, artifactitious, thiol is clearly non-essential for the enzymic function of carboxypeptidase. Under more vigorous conditions of treatment with sodium borohydride, in the absence of βPP, enzymic activity is progressively lost and this loss can be correlated with the additional yields of carboxymethylcysteine over and above the one equivalent contributed by the newly formed thiol.

TABLE XII. Alkylation of thiols of carboxypeptidase A (Walsh *et al.*, 1962)

Enzyme	Urea	β-Phenyl propionate	Reducing agent	pH	Enzyme inactivation (%)	Carboxymethylcysteine (residues/molecule)
+	+	−	ME† 0·06 M	8·0	100	2·00
+	+	+	ME 0·06 M	8·0	5	0·40
+	−	+	ME 0·33 M	7·2	0	0·56
+	−	+	ME 0·33 M	8·0	2	0·44
+	−	−	$NaBH_4$	9·0	26	1·10
+	−	−	$NaBH_4$	9·0	48	1·44
+	−	+	$NaBH_4$	9·0	0	0·95

† ME, β-mercaptoethanol.

On the basis of the experiments just described, it has been possible to label selectively the thiol at the active site, and to isolate the labelled peptides from enzymic digests of the protein. Two reagents were found to be most suitable for this purpose—one is radioactive [^{14}C] iodoacetate, and the other the yellow-coloured maleimide derivative of Witter and Tuppy (1960), DDPM. The radioactive peptides were purified by chromatography on Dowex 50 × 2, whereas the yellow peptides were selectively adsorbed on talc. In either case, final purification was achieved by high voltage electrophoresis and chromatography.

Selective labelling of the thiol at the active site with DDPM was achieved as follows. The protein was first treated with sodium borohydride in the presence of βPP, and the newly formed thiol blocked with iodoacetamide. As indicated in Table XII, the monoalkylcarboxypeptidase that was obtained was fully active. This protein was then

treated with mercaptoethanol and subsequently alkylated with a different alkylating agent, namely DDPM, which could couple only with the thiol group that bound zinc in the native enzyme. The compositions of the cysteinyl peptides obtained from Nagarse digests of the alkylated enzyme are summarized in Table XIII. The top row indicates the composition of the cysteinyl peptide containing the active site thiol which was labelled with DDPM after the second sulphur was blocked with iodoacetamide. The bottom row describes the composition of the cysteinyl peptides obtained by labeling both the thiols after reduction with mercaptoethanol. As is to be expected, two families of peptides are found—one similar to the peptide at the active site, the other obviously representing the region of the non-essential sulphur which became reduced and alkylated in the intact protein. Thus, the cysteine peptide containing

TABLE XIII. S-DDPS-Cysteinyl peptides from Nagarse digests of carboxypeptidase A (Walsh *et al.*, 1962)

Carboxypeptidase A	Isolated peptides	
Active centre thiol alkylated after second thiol was blocked with iodoacetamide	(DDPS–Cys,Ser_2) (DDPS–Cys,Ser_2,Glu)	Trace of (DDPS–Cys,Val,Gly)
Both thiols alkylated	(DDPS–Cys,Ser_2) Ser(DDPS–Cys,Ser,Pro) (DDPS–Cys,Ser_2,Glu)	(DDPS–Cys,Val)Gly (DDPS–Cys,Val,Gly,Asp)

serine represents the thiol at the active site which binds zinc in the intact protein.

Having identified the thiol of the active site by its adjacent serine residues, in contrast to the non-essential sulphur peptide with its valine and glycine, the amino-acid sequences around these two residues were examined. In these studies, DDPM was not used, since the yields of the labelled peptides were low, and their separation and purification were complicated by the lability of the imide ring structure of the alkyl group. Instead, the protein was oxidized with performic acid and the cysteic acid peptides were isolated from a chymotryptic digest. These peptides were purified by high voltage electrophoresis at pH 2·1, followed by electrophoresis at pH 6·5, and finally by descending chromatography in butanol–acetic acid–water. The sequences around the two cysteic acid residues in oxidized carboxypeptidase are summarized in Fig. 7. The serines adjacent to the cysteic acid in peptide A identifies this cysteic acid with the sequence around the active centre. Peptide B is the sequence

around the non-essential sulphur, and as expected, contains valine and glycine.

The N-terminal groups were determined by FDNB, which gave DNP-glycine with peptide A, and DNP-cysteic acid with peptide B. Carboxypeptidase A released only tyrosine from peptide A. Trypsin split the A peptide into a dipeptide (Gly,Lys) (residues 1–2), and a dodecapeptide with alanine as the N-terminal group. Nagarse digest of this large peptide yielded a tetrapeptide with the composition

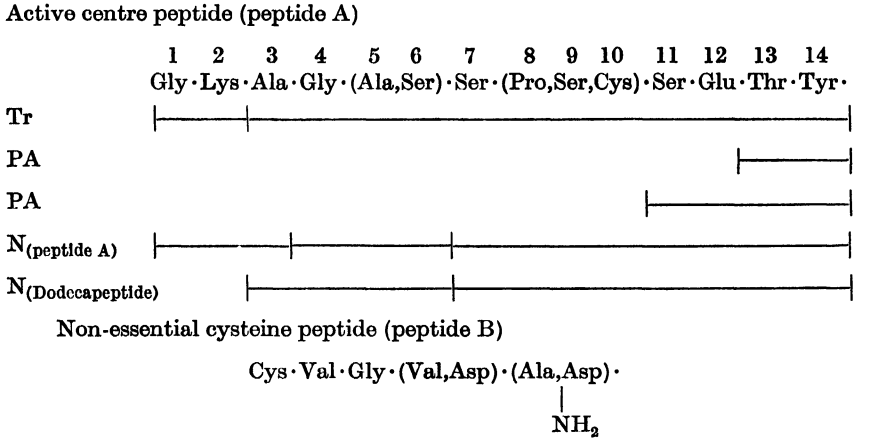


FIG. 7. Sequences of the cysteic acid peptides from chymotryptic digests of oxidized carboxypeptidase A. The abbreviations on the left indicate the type of partial degradation employed, namely Tr, tryptic digestion; PA, partial acid hydrolysis; N, digestion with Nagarse.

(Ala_2,Ser,Gly) (residues 3–6), and an octapeptide with the composition of residues 7–14 and with serine as the N-terminal group. Partial acid hydrolysis of the octapeptide yielded a dipeptide (Thr,Tyr) (residues 13–14), and a tetrapeptide Ser.(Glu,Thr,Tyr) (residues 11–14). Nagarse digests of the A peptide yielded peptides 1–3, 4–6 and 7–14. By an integration of the compositions of these various peptides with the compositions in Table XIII, the partial sequence depicted in Fig. 7 has been worked out. The sequences of residues 5–6 and 8–10 are yet to be determined.

In the case of peptide B, Nagarse digest yielded a pentapeptide with N-terminal cysteic acid, and a dipeptide of asparagine and alanine. From the neutral character of the dipeptide at pH 6·5, it was possible to establish that it is asparagine and not aspartic acid. Carboxypeptidase digestion of the pentapeptide yielded both valine and aspartic acid at the same rate. From Nagarse digests of the original dialkyl protein, a

tripeptide with the composition of Cys,Val,Gly with a C-terminal glycine was isolated. From these results, the partial sequence around the second sulphur was deduced.

In summary, carboxypeptidase A, a zinc metalloenzyme, has a sulphydryl group that binds the metal at the active site. However, this group in the apoenzyme is strikingly unreactive towards alkylating agents unless the protein is extensively and irreversibly disrupted. This lack of reactivity may well relate to details of tertiary configuration of the apoenzyme and to interactions of thiols with neighbouring groups such as have been encountered with other proteins; or, it may be an inherent property of certain protein thiol groups.

A second thiol group is formed by reduction of the protein, but it plays no functional role in the catalytic mechanism. Its chemical nature in the native protein is still not known, although it gives rise to cysteine on reduction and cysteic acid on oxidation. It does not seem to be a thiolester either since treatment with hydroxylamine under conditions usually employed for the hydrolysis of thiolesters does not yield a reactive thiol.

The thiol at the active site could be labelled with an alkyl group only after prior reduction of the protein and incidental formation of the second thiol. However, by protecting the functional thiol with a competitive inhibitor during reduction, the second thiol could be selectively alkylated. Next, by removing the competitive inhibitor, the thiol at the active site could be specifically substituted with an identifiable label.

In concluding, one is tempted to infer some features of significance concerning the structure of the peptide from the active centre. The abundance of amino acids with short side chains like serine, alanine, and glycine is striking and could well be of functional significance in providing space to accommodate the approach of large substrate molecules. In this respect, the sequences around the active centre serine in trypsin and chymotrypsin bear some resemblance to carboxypeptidase. But if one proposes on the one hand that substrates—large and small—can approach the active centre, one must account for the fact that the alkylating agents will not react with the thiol of the apoenzyme at the active site unless the structure of the protein is modified by reducing agents or partial degradation. Obviously, it would be premature to infer structural features of the whole active centre from the structure of a single peptide fragment representing a region of the active catalytic site in the protein. Such an understanding would require the elucidation of the complete structural features of the protein—primary, secondary and tertiary—by chemical, physical and crystallographic means. Also it would require the complete description of the active catalytic site by the identification of all of the functional groups involved in the

catalytic mechanism, as well as those that are involved in the binding of substrates. Work along these lines is currently in progress in our laboratory and in Boston.

Acknowledgements

The research described in this report has been supported in part by the National Institutes of Health (RG-4617), by the Department of the Navy, Office of Naval Research (NONR 477-04) and by the American Cancer Society (P-79). The authors wish to acknowledge also many valuable discussions and advice by Dr. Bert L. Vallee.

References

Bargetzi, J.-P., Cox, D. J., Kumar, K. S. V. S., Walsh, K. A. and Neurath, H. (1963). In preparation.

Brown, J. R., Cox, D. J., Greenshields, R. N., Walsh, K. A., Yamasaki, M. and Neurath, H. (1961). *Proc. nat. Acad. Sci., Wash.* **47**, 1554.

Coleman, J. E. and Vallee, B. L. (1960). *J. biol. Chem.* **235**, 390.

Coleman, J. E. and Vallee, B. L. (1961). *J. biol. Chem.* **236**, 2244.

Coleman, J. E. and Vallee, B. L. (1962*a*). *J. biol. Chem.* **237**, 3430.

Coleman, J. E. and Vallee, B. L. (1962*b*). *Fed. Proc.* **21**, 247.

Coombs, T. L. and Omote, Y. (1962). *Fed. Proc.* **21**, 234.

Keller, P. J., Cohen, E. and Neurath, H. (1956). *J. biol. Chem.* **223**, 457.

Keller, P. J., Cohen, E. and Neurath, H. (1958). *J. biol. Chem.* **230**, 905; **233**, 344.

Neurath, H. (1960). In *The Enzymes* (P. D. Boyer, H. Lardy and K. Myrback, eds.), 2nd Ed., p. 11. Academic Press, New York.

Neurath, H. and Schwert, G. W. (1950). *Chem. Rev.* **46**, 69.

Spackman, D. H., Stein, W. H. and Moore, S. (1958). *Analyt. Chem.* **30**, 1190.

Vallee, B. L. and Neurath, H. (1955). *J. biol. Chem.* **217**, 253.

Vallee, B. L., Coombs, T. L. and Hoch, F. L. (1960). *J. biol. Chem.* **235**, PC45.

Vallee, B. L., Rupley, J. A., Coombs, T. L. and Neurath, H. (1958). *J. Amer. chem. Soc.* **80**, 4750.

Walsh, K. A., Kumar, K. S. V. S., Bargetzi, J.-P. and Neurath, H. (1962). *Proc. nat. Acad. Sci., Wash.* **48**, 1443.

Witter, A. and Tuppy, H. (1960). *Biochim. biophys. Acta* **45**, 429.

Discussion

M. S. NARASINGA RAO (Regional Research Laboratory, Hyderabad): I see from one of your slides that the isoelectric point of carboxypeptidase A is strongly dependent on the ionic strength of the buffer. One interpretation is that there is specific ion binding. Does this have any effect on enzymatic activity?

K. S. V. SAMPATHKUMAR: To some extent.

G. T. RAJAGOPALAN (Duke University, Durham, U.S.A.): If there is similarity in sequence between trypsin and chymotrypsin and carboxypeptidase, why is not carboxypeptidase A inhibited by DFP?

K. S.V. SAMPATHKUMAR: We just say that the sequences have some similarity as far as the abundance of short chain amino acids at the active site is concerned. As regards the functional group that is involved, the enzymes are different—one is cysteine in carboxypeptidase A and, in trypsin and chymotrypsin, it is a serine hydroxyl group.

Resilin, a Rubber-like Protein, and its Significance

TORKEL WEIS-FOGH

Zoophysiological Laboratory B,
University of Copenhagen, Denmark

ABSTRACT

Physical and chemical analyses of resilin show (1) that this structural protein behaves as an almost perfect rubber; (2) that the polypeptide chains are held together in a three-dimensional network by very stable groups involving co-valent bonds; (3) that these cross links consist of residues of either diamino-dicarboxylic or triamino-tricarboxylic α-amino acids built into the polypeptide chains, probably by ordinary peptide bonds and at quite regular distances; (4) that the cross-linking groups are derived from tyrosine and behave as fluorescent monophenols.

The protein has no tendency to form any regular secondary structure or to crystallize. It shows little contrast and no structure in the electron microscope. It is suggested that similar proteins may be present in many cells and extracellular structures.

Resilin is a structural protein which has been discovered recently as an essential part of some elastic ligaments in arthropod cuticle (Weis-Fogh, 1960; Bailey and Weis-Fogh, 1961). In certain places, it is secreted by the epidermal cells in a pure form so that, in spite of its complete insolubility, its physical and chemical properties can be studied in the native state. The only disadvantage is that the pieces are small and must be dissected free under the microscope.

It may seem strange, in this book, to mention a protein which has no tendency to take up any regular secondary structure or to crystallize, but since interest in proteins is due to the fact that they are the main building stones of living organisms it could, nevertheless, be useful to discuss an odd-man-out, particularly since structural proteins of similar nature may turn out to be quite common.

We are interested in this protein (a) because resilin swollen in water behaves as a perfect rubber and therefore consists of a three-dimensional network of very flexible polypeptide chains in vivid thermal agitation, held together by means of a few stable cross-links, (b) because these cross-links are of a novel type and (c) because similar networks may play an important role as supporting molecular structures both outside and inside living cells.

1. Physical Properties

A piece of resilin is hard and glass-like when dry but swells and becomes rubbery in water and other polar solvents like formamide, ethylene glycol and glycerol (Weis-Fogh, 1960). Even in 70% ethanol, where it swells to only 1·5 times its dry volume, it shows typical long-range elasticity with lack of flow and complete elastic recovery after prolonged deformation. It is characteristic that the elastic force is associated with entropy changes rather than with changes in internal energy (Weis-Fogh, 1961*a*) and that the mechanical and optical properties are consistent with an isotropic network of chains which do not interact and whose flexibility is unaffected by pH (from 1·8 to 12·3) and is so great that each residue is virtually free to rotate relative to its neighbours. Thus, there can be no significant amount of intrachain hydrogen bonding at any pH or between 20 and 90°C (Weis-Fogh, 1961*b*). Although few in number, the lack of flow indicates that every chain is linked by stable cross-links to its neighbours. Moreover, the very high elastic efficiency (about 97%) points towards these bonds being regularly spaced, in contrast to those of ordinary rubber. In fact, this is probably the reason why resilin is a better elastomeric material than any known natural or synthetic rubber (Jensen and Weis-Fogh, 1962).

In water the swelling of resilin changes with pH from 1·8 to 4·5 times the dry volume, i.e. with the dissociation of ionizable groups, but the fundamental properties are identical in water, pure glycol, glycerol and formamide and can therefore not be explained on the basis of an electrostatic model (unpublished experiments). It is also characteristic and consistent with our concepts of an ideal isotropic rubber that X-ray diffraction of highly stretched and slowly dried resilin "fibres" failed to show any regular ordering of the chains or any other sign of crystallinity in the now hard and glass-like preparations; similarly, no structure can be seen in the electron microscope (unpublished observations by G. F. Elliott, A. F. Huxley and Weis-Fogh). It is therefore justified to consider resilin as an isotropic network of highly flexible chains which show no tendency to interact when swollen and no possibility for clicking into any regular lattice when dried. Such networks would be difficult to detect in small quantities in living tissues.

2. Chemical Properties

Resilin is completely insoluble in all solvents that do not break peptide bonds and is heat stable in water up to 140°C, but it is readily digested by all proteases, acids and bases. A complete hydrolysate contains sixteen ordinary amino acids of which glycine is the most abundant while Hypro, Met, Cys and Try are absent (Bailey and Weis-Fogh, 1961).

We have no explanation of the great flexibility and lack of intrachain hydrogen bonding because we still do not know anything about the primary sequence, but the relatively high content of Gly (31%), Pro (8%) and residues with polar groups (34%) is noteworthy. The significant point is that nature seems able to design polypeptide chains with no tendency to form stable secondary structures and with very little interaction between neighbouring chains, irrespective of whether the quantity of polar residues is high, as in resilin, or very low, as in elastin. It is reasonable to assume that such properties are utilized in many other proteins.

In addition to the sixteen ordinary amino acids, Andersen (1963) has recently shown that there are two new α-amino acids present in small amounts. They both react as monophenols; one is a diamino-dicarboxylic acid and the other a triamino-tricarboxylic acid. In native resilin the α-amino groups are not free to react with DNFB and partial enzymic digestion shows that they are built into the polypeptide network. Since they are ideally suited for linking together two and three polypeptide chains, respectively, and since their number agrees well with the number of cross-links estimated from mechanical experiments (Weis-Fogh, 1961*b*), we have strong reasons for believing that they represent the actual stable cross-links that must be present in resilin. Other obvious possibilities could be ruled out (SH and S—S groups are absent, all ϵ-amino groups of lysine are free, phosphate and carbohydrate are absent). We also know from Neville's (1963) experiments that the amount of the unusual amino acids increases linearly with the total amount of resilin laid down during morphogenesis. This indicates that they constitute a fixed proportion of native resilin and that proresilin is cross-linked almost as soon as it is secreted, in other words not by a random bulk reaction similar to the vulcanization of rubber. It is likely that the superior mechanical properties of resilin is due to finely controlled cross-linking which may be achieved in refined co-polymers like proteins but not in ordinary rubbers. As to the precursors of the cross-linking amino acids, unpublished experiments by Andersen and Kristensen with radioactive amino acids show that one is made from three and the other from five tyrosine rings, but we do not know their exact constitution yet or any details about their formation.

3. Biological Significance

There is good reason to think that, as to fundamental properties, elastin has much in common with resilin although its amino-acid composition is rather different. Thus, elastin is cross-linked by stable groups of unknown nature and, if these groups are similar to those in resilin, they

have probably escaped notice partly because they can be present only in small quantities but mainly because such aromatic compounds tend to stick firmly to the usual ion exchange materials (Andersen, 1963).

In arthropods, resilin is used as an elastic extracellular material, i.e. for mechanical springs, in the same way as elastin is used for the construction of elastic blood vessels and tissues in vertebrates. However, highly deformable and stable molecular networks may be of fundamental importance as internal constituents of many cells and their organelles. Whereas there is no doubt that aggregation of polymers goes a long way to explain the "spontaneous" formation of membranes, envelopes, fibrils and several other constituents, many cells exhibit a feature of stability which is hard to reconcile with aggregations due only to alignment of molecules and formation of secondary bonds. Thus, some cells may swell and shrink quickly and reversibly by a factor of 5–10 without any permanent damage. Other cells may change their shape enormously due to externally applied strains or due to internal deposits or vacuoles and still revert to their former dimensions and, finally, many unicellular organisms have an elastic cortex just beneath the plasma membrane. I do not claim that these and similar well-known observations make it necessary to postulate the existence of molecular networks in cells but they suggest their presence, not as a universal cyto-skeleton, but in specific areas. In this context, it should be remembered that even the concentrated network of resilin (40% dry matter at pH 7) tends to escape notice due to its stability, flexibility, lack of colour, lack of specific reactive groups, lack of crystallinity and, finally, due to its low contrast and lack of structure in the electron microscope.

Regarding cross-links, I should not be surprised if, in due course, we find several types of cross-link in proteins in addition to the few known ones and that they will be found in the insoluble structural compounds of cells and tissues which biochemists tend to throw away precisely because they are insoluble.

References

Andersen, S. O. (1963). *Biochim. biophys. Acta* **69**, 249.
Bailey, K. and Weis-Fogh, T. (1961). *Biochim. biophys. Acta* **48**, 452.
Jensen, M. and Weis-Fogh, T. (1962). *Phil trans.* **B245**, 137.
Neville, A. C. (1963). *J. insect Physiol.* (in press).
Weis-Fogh, T. (1960). *J. exp. Biol.* **37**, 889.
Weis-Fogh, T. (1961*a*). *J. mol. Biol.* **3**, 520.
Weis-Fogh, T. (1961*b*). *J. mol. Biol.* **3**, 648.

Discussion

S. MOORE: I recall having an opportunity to discuss the interesting experiments on resilin with Dr. Weis-Fogh about a year ago. Their results afford a good example

of the fact that protein chemists must be prepared for surprises. Not all chains in proteins will turn out to follow the simple polypeptide pattern. Some of those investigated so far have been built only from polypeptide chains with disulphide bonds as cross-links; the unusual cross-linkage that Dr. Weis-Fogh and his associates have discovered is one of the several additional types of linkage that may be encountered with increasing frequency as more proteins with special functions are studied in detail.

Amino Acid Composition of Ichthylepidin from Fish Scales

R. V. Seshaiya, P. Ambujabai and M. Kalyani

Marine Biological Station, Porto Novo, South India, India

ABSTRACT

The amino acid composition of ichthylepidin in the scales of six species of fish has been determined. The percentage of glycine, proline and hydroxyproline is almost constant in the ichtylepidins from different species. Cystine content is higher in the ichthylepidins in the species investigated now than in the pilchard, but more or less the same as in the herring scale. The total imino acid content is relatively higher in ichthylepidin than in gelatin of the scale. In its X-ray diffraction pattern ichthylepidin resembles collagen. In the presence of cystine and in hydrothermal stability ichthylepidin resembles elastoidin.

1. Introduction

The scales of teleosts, familiarly known as bony fishes, contain mineral matter up to 59% of their dry weight and organic matter varying from 41% to 84% in the different species. The organic matter is almost all protein, which, for a long time, was regarded as wholly composed of collagen. But the investigations of Morner (1898) showed that the scales of many species of fish contain besides mineral matter and collagen a peculiar albuminoid, to which he gave the name ichthylepidin. Green and Tower (1902) studied the distribution of ichthylepidin in some of the common American fish.

Ichthylepidin can be fractionated from the scales after demineralization and subsequent removal of guanine and chondroitin sulphuric acid, and the extraction of all collagen as gelatin. Pure ichthylepidin thus obtained is in the form of transparent, homogeneous and flexible plates more or less conforming to the shape of the scale. The two proteins, collagen, and ichthylepidin, are present in approximately equal proportions in the scales.

The most characteristic property of ichthylepidin, which is in contrast to collagen, is that it is not converted into gelatin on boiling with water. There have been very few investigations on the composition of ichthylepidin. Block *et al.* (1949) carried out a qualitative and quantitative analysis of some of the amino acids of ichthylepidin from the herring scales. Winter (1954) determined some of the amino acids in the ichthylepidin of the scales of the roach, and Burley and Solomons (1957) carried out a complete analysis of the ichthylepidin of the pilchard (*Sardina ocellata*).

But the teleosts are a very large and varied group of animals inhabiting different types of environment, and to have a satisfactory understanding of the nature and composition of ichthylepidin, it would be necessary to investigate it in different types of fish. It is also to be borne in mind that the composition of fibrous proteins like collagen is variable, and may, sometimes, differ even between organs of the same animal.

In the present study the amino acid composition of ichthylepidin of six species of fish has been investigated.

2. Methods

The extraction of ichthylepidin was effected by adopting Morner's method as described by Green and Tower (1902) and slightly modified by Winter (1954).

The amino acid composition was studied on acid hydrolysates of ichthylepidin, mostly with the aid of uni-dimensional and two-dimensional chromatography. Whatman chromatographic paper No. 2 was used and the solvent systems consisted of butanol/acetic acid/water and phenol/ammonia. Leucine, isoleucine, phenylalanine, valine and methionine were separated using *m*-cresol (pH 8.4). Quantitative determinations were made with a photovolt densitometer adopting the maximum density method. For estimation of hydroxyproline the method of Neuman and Logan (1950) modified by Leach (1960) was adopted. For proline determination the method described by Troll and Lindsley (1955) was followed. Hydroxylysine and tryptophan were not determined. In all cases not less than five determinations were made and the reproducibility of the results checked. X-ray diffraction patterns of ichthylepidin were obtained with Cu $K\alpha$ radiation of wavelength 1·542 Å, using a flat camera.

It may be noted that although the analytical procedures adopted by us differ to some extent from those of Burley and Solomons (1957), the results have on the whole a remarkably close agreement.

We have also made a preliminary investigation of the hydrothermal stability of ichthylepidin which is much higher than for gelatin. The shrinkage temperature for ichthylepidin is about 63°C.

3. Results

The amino acid composition of ichthylepidin in the six species of fish investigated by us is shown in Table I. Table I also shows the amino acid composition of the ichthylepidin of the scales of the pilchard, herring and roach, investigated by Burley and Solomons (1957), Block *et al.* (1949) and Winter (1954) respectively.

As already mentioned, there is a very close correspondence between the results obtained in the present investigation and in the investigations by Burley and Solomons (1957). Amino acids like glycine, proline and hydroxyproline, which are important in relation to the structure of the fibrous protein, have almost identical values in all the species investigated by us and in the pilchard. Histidine, glutamic acid, threonine,

TABLE I. Amino acid composition of ichthylepidin from different species of fish

Values in g amino acid/100 g protein†

Acid	*Mugil speigleri*	*Mugil dussumieri*	*Mugil cephalus*	*Polynemus tetradactylus*	*Hilsa toli*	*Opisthopterus tardoore*	Pilchard (Burley and Solomons, 1957)	Herring (Block *et al.*, 1949)	Roach (Winter, 1954)
Alanine	9·62	9·54	9·59	10·13	9·80	10·14	9·20	—	—
Aspartic acid	7·50	7·56	7·57	6·72	7·40	6·65	6·50	—	—
Arginine	7·06	7·13	7·26	8·31	9·00	8·88	7·33	5·4	5·5
Histidine	1·48	1·53	1·53	1·38	1·64	1·48	1·56	3·1	—
Lysine	3·43	3·38	3·45	3·58	4·10	3·46	3·69	3·7	2·9
Glutamic acid	9·67	9·64	9·73	9·92	10·30	10·20	9·56	—	9·2
Glycine	24·48	24·50	24·56	25·30	25·44	24·93	24·00	—	22·0
Serine	5·91	5·92	5·86	5·20	5·80	5·58	4·56	—	7·0
Threonine	3·06	3·04	3·12	3·22	3·80	3·40	3·14	—	3·7
Proline	13·10	13·20	13·25	13·04	13·20	13·30	13·34	—	—
Hydroxyproline	9·51	9·61	9·69	9·69	9·80	9·53	9·80	—	—
Leucine	2·20	2·26	2·30	2·50	2·68	2·65	2·84	—	—
Isoleucine	2·10	1·97	2·17	1·40	1·30	1·26	1·46	—	—
Phenylalanine	2·59	2·68	2·70	2·07	2·00	2·00	3·12	—	—
Valine	2·23	2·20	2·33	2·33	2·60	2·20	2·44	—	—
Methionine	1·69	1·58	1·48	1·91	1·33	1·79	2·06	3·0	—
Tyrosine	1·26	1·17	1·13	1·03	1·07	0·98	1·71	5·2	2·4
Cystine	1·14	1·17	1·19	1·01	1·07	1·18	0·50	1·1	—
Cysteine	—	—	—	—	—	—	0·03	—	—
Hydroxylysine	—	—	—	—	—	—	1·63	—	—
Tryphtophan	—	—	—	—	—	—	0·50	0·6	—

† Mean values for five determinations. The standard error is of the order of 0·1 g/100 g and in some cases even less. Hydroxylysine was not determined.

leucine, valine and serine have slightly higher values in the species investigated by us. Isoleucine shows slight variation in the different species. But these variations and differences are what may reasonably be expected when the materials are obtained from different species.

The content of sulphur-containing amino acids in the different ichthylepidins calls for some comment. The cystine content in the ichthylepidins of the species in the present investigation is higher than that in the pilchard, but comes close to that in the herring scale. It must also be pointed out that Burley and Solomons (1957) estimated both cysteine and cystine but only cystine was recovered from the hydrolysate in our investigation, cysteine having been probably converted into cystine. Further, the methionine content in the ichthylepidins investigated by us is slightly lower than in the pilchard ichthylepidin. According to Green and Tower (1902) the content of total sulphur varies within wide limits with each species.

As Block *et al.* (1949) and Winter (1954) carried out only a partial analysis of ichthylepidin, it is not possible to discuss their findings in detail.

4. Nature of Ichthylepidin

From the analyses discussed so far, we are justified in inferring that the over-all picture of the amino acid composition of ichthylepidin is more or less similar, though slightly variable in the different species. In the light of this we might attempt to define the relation of ichthylepidin to other scleroproteins.

So far as we know, no X-ray studies of ichthylepidin have been published. Astbury (1947) referred the protein of the scale of the fish to the collagen type, but ichthylepidin was not investigated separately from collagen. A general resemblance of the pattern of ichthylepidin to that of collagen is indicated in a preliminary X-ray study made by us.

Block *et al.* (1949) assumed that ichthylepidin might be regarded as an albuminoid of the collagen type, since it contains hydroxyproline. Winter (1954) compared ichthylepidin with other scleroproteins and defined it as a special kind of scleroprotein. Solomons (1955) suggested that ichthylepidin is related to keratin, although he mentioned that it is similar to elastoidin. Jacquot (1961) also considered ichthylepidin as a keratin on account of the presence of cystine.

A comparison between gelatin, ichthylepidin and elastoidin, in respect of their amino acid residues is shown in Table II. The gelatin and ichthylepidin were obtained from the scales of *Mugil cephalus* and the values for elastoidin are taken from the work of Damodaran *et al.* (1956).

It will be observed that the total imino acid (proline + hydroxyproline)

content is slightly higher in ichthylepidin than in gelatin. Piez (1960) and Piez and Gross (1960) have presented evidence to indicate that the varying stabilities exhibited by collagens are related to the pyrrolidine ring content rather than to the hydroxy group of hydroxyproline and that it is the total hydroxyproline + proline content that is important. Their suggestion is that the pyrrolidine ring of imino acids plays an important role in intramolecular stability of collagen. Thus a decrease of about

TABLE II. Comparison of the amino acid composition† of gelatin and ichthylepidin from the scales of *Mugil cephalus* and of elastoidin from shark fin-ray

Acid	Gelatin‡		Ichthylepidin‡		Elastoidin (Damodaran *et al.*, 1956)	
Alanine	117·50		107·75		128·00	
Aspartic acid	57·06		56·84		48·10	
Arginine	52·93		41·96		45·50	
Histidine	9·48		9·87		11·10	
Lysine	28·90		23·63		25·60	
Glutamic acid	70·30		66·18		74·90	
Glycine	343·06		326·13		338·00	
Serine	54·84		55·81		31·50	
Threonine	28·67		26·13		20·30	
Proline	111·20	176·9	115·20	189·1	115·30	182·1
Hydroxyproline	65·72		73·90		66·80	
Leucine	21·98		17·55		20·00	
Isoleucine	14·70		16·56		20·50	
Phenylalanine	9·45		16·36		12·70	
Valine	19·40		11·36		23·20	
Methionine	8·65		9·98		12·00	
Tyrosine	4·57		6·24		39·50	
Cystine	—		4·75		1·50	

† Residues per 1000 total residues.
‡ The values for gelatin and ichthylepidin are based on the present investigation.

three residues in the imino acid content is associated with a fall of 1°C in shrinkage. The greater stability of ichthylepidin is probably associated with its higher imino acid content. There is a difference of about thirteen residues between the gelatin and ichthylepidin of the scales of *Mugil cephalus*, which we believe, is statistically significant, and a preliminary study of the physical properties of ichthylepidin has shown that its shrinkage temperature is about 63°C.

The resemblance of ichthylepidin to elastoidin is seen not only in the presence of cystine but also in the slightly higher imino acid content

which is higher in both of them than in gelatin, and also in the high shrinkage temperature. Gustavson (1956, 1958) has pointed out that elastoidin behaves hydrothermally as a cross-linked (tanned) collagen, and Pickens (1960) considers that the properties of elastoidin indicate that the polypeptide chains are very firmly and completely cross-linked. Ichthylepidin may be expected to show a similar structure.

The physical properties of ichthylepidin like its insolubility, its stability and high shrinkage temperature are of advantage to the scales in functioning as a flexible protective exoskeleton.

Acknowledgement

Out thanks are due to Dr. K. Venkateswarlu, formerly Professor of Physics in the Annamalai University, for help in the X-ray diffraction study of ichthylepidin.

References

Astbury, W. T. (1947). *Proc. roy. Soc.* **A134**, 303.
Block, R. J., Horwitt, M. K. and Bolling, D. (1949). *J. dent. Res.* **28**, 518.
Burley, R. W. and Solomons, C. C. (1957). *S. Afr. industr. Chem.* **11**, 154.
Damodaran, M., Sivaraman, C. and Dhavalikar, R. S. (1956). *Biochem. J.* **62**, 621.
Green, E. H. and Tower, R. W. (1902). *Bull. U.S. Fish. Comm.* **21**, 97.
Gustavson, K. H. (1956). *The Chemistry and Reactivity of Collagen*, p. 217. Academic Press, New York.
Gustavson, K. H. (1958). In *Recent Advances in Gelatin and Glue Research* (G. Stainsby, ed.), p. 253. Pergamon Press, London.
Jacquot, R. (1961). In *Fish as Food* (G. Borgstrom, ed.), p. 157. Academic Press, New York.
Leach, A. A. (1960). *Biochem. J.* **74**, 70.
Morner, C. Th. (1898). *Z. phys. Chem.* **24**, 125.
Neuman, R. E. and Logan M. A. (1950). *J. biol. Chem.* **184**, 299.
Pickens, L. (1960). *Organization of Cells*, p. 399. Clarendon Press, Oxford.
Piez, K. A. (1960). *J. Amer. chem. Soc.* **82**, 247.
Piez, K. A. and Gross, J. (1960). *J. biol. Chem.* **235**, 995.
Solomons, C. C. (1955). *S. Afr. J. med. Sci.* **20**, 27.
Troll, W. and Lindsley, J. (1955). *J. biol. Chem.* **215**, 655.
Winter, H. (1954). *Arch. Fischereiwissenschaften* **5**, 170.

Discussion

S. VENKATARAMAN (Madras Institute of Technology): You mentioned the similarity between elastoidin and ichthylepidin. Elastoidin exhibits a reversal of birefringence on thermal shrinkage. Do you find a similar reversal of birefringence when ichthylepidin is thermally shrunk?

R. V. SESHAIYA: There are dissimilarities also. Ichthylepidin does not show a reversal of shrinkage (birefringence has not been investigated).

G. N. RAMACHANDRAN: Does the X-ray pattern actually show the 2·95 ring as for collagen?

R. V. SESHAIYA: It is only a general resemblance.

G. N. RAMACHANDRAN: Unless one gets the particular ring clearly, one cannot be sure of the collagen structure.

R. V. SESHAIYA: The collagen group was originally named after the collagen of histology, and the high content of glycine, proline and hydroxyproline has been considered as the characteristic feature of all collagens.

M. S. NARASINGA RAO (Regional Research Laboratory, Hyderabad): How do you prepare the protein for amino acid analysis from the scales of fish?

R. V. SESHAIYA: The general procedure adopted for the preparation of ichthylepedin was to remove the mineral matter, chondroitin sulphuric acid, etc. and then treat the residual scales with 0·1% HCl at 40°C for several days. This removes all collagen as gelatin, and ichthylepidin is left behind.

The Study of Early Phases in Collagen Biosynthesis

M. Chvapil
in collaboration with
J. Hurych, V. Kobrle and B. Čmuchalová

Institute of Industrial Hygiene and Occupational Diseases, Department of Experimental Biology, Prague, Czechoslovakia

ABSTRACT

Evidence is given that the content of free and bound ultrafilterable hydroxyproline in tissues depends on the rate of collagen metabolism. At the inhibition of collagen synthesis with cortisone the level of free hydroxyproline is parallelly decreased and, on the other hand, at the artificially increased fibrillogenesis, after application of carrageenin, the level of ultrafilterable hydroxyproline is also increasing in a certain phase of granulation tissue formation. During the incubation of skin slices of chicken embryos with [^{14}C]proline the specific activity of the fraction of [^{14}C]hydroxyproline of the low molecular substances is significantly higher than of soluble collagens.

In this paper we present a certain part of our experiments which aim to contribute to the understanding of the function of hydroxyproline (Hypro), either free or bound in low molecular weight substances, in the metabolism of collagen. Although it was commonly agreed that free Hypro is not utilized for collagen synthesis, our results, however, show its dependence on the rate of collagen metabolism.

The question of hydroxyproline metabolism arose for the first time in 1957 when we studied the mechanism of the effect of cortisone on the biosynthesis of collagen. We administered cortisone into the chorioallantoic membrane of 8-day-old-embryonated eggs and determined the content of free and total bound Hypro in whole embryos after different intervals. We proved that the action of cortisone results in a decrease in the absolute amount of collagen and in the absolute amount of free hydroxyproline in the embryo (Chvapil, 1959).

The same conclusion was reached when using another experimental model. We administered cortisone repeatedly to pregnant rats and followed free proline (Pro) and Hypro as well as total Hypro in new-born rats (Chvapil, 1958). From the results summarized in Fig. 1 it is clear that a significant inhibition of the formation of bound collagen Hypro as well as free Hypro with a simultaneous increase of Pro content occurred.

These experiments led us to the conclusion that cortisone as well as a deficiency of vitamin C interferes with the mechanisms leading to the hydroxylation of proline in hydroxyproline.

These results favour the possibility of hydroxylation of Pro already in its free state (Robertson *et al.*, 1959), whereas it is generally presumed that hydroxylation occurs of even peptide-bound proline (as well as lysine) (Stetten, 1949; Sinex *et al.*, 1959).

The results on the mode of action of cortisone on fibrillogenesis are equally in favour of a certain inter-relation between the content of free Hypro and the degree of collagen metabolism. This was, however, a

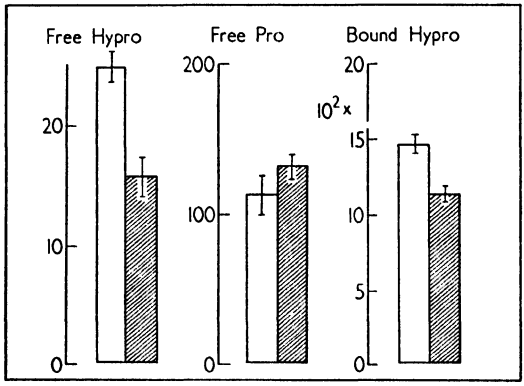


FIG. 1. Influence of cortisone on free hydroxyproline, free proline and total collagen hydroxyproline in newborn rats—unshaded without, and shaded with cortisone. The values are presented in μg/g wet tissue. No differences in water content in both groups of rats were found. The variability is given by the standard deviation shown.

contradiction of the findings of Wolf and Berger (1958) that "no metabolic turnover of Hypro in the organism occurs". We tried to solve these contradictions by means of the following experiments.

After giving guinea-pigs subcutaneous injections of carrageenin, we studied the content of DNA being formed in the granulation tissue, and also free and collagen Hypro. Further we followed the content of ultrafilterable bound Hypro whose presence in animal tissues, especially in embryonal tissues we have proved before (Kobrle and Chvapil, 1961). Owing to the properties of the applied filtration membrane the maximum molecular weight of ultrafilterable substances is about 30,000.

Two other figures illustrate the course of changes of the above-mentioned substances during the development of carrageenin granuloma. The changes are shown schematically in Fig. 2, where they are expressed as weight concentration, i.e. the quantity of studied substances in the granuloma. Figure 3 shows the absolute changes describing the actual formation of the studied substances. From this illustration

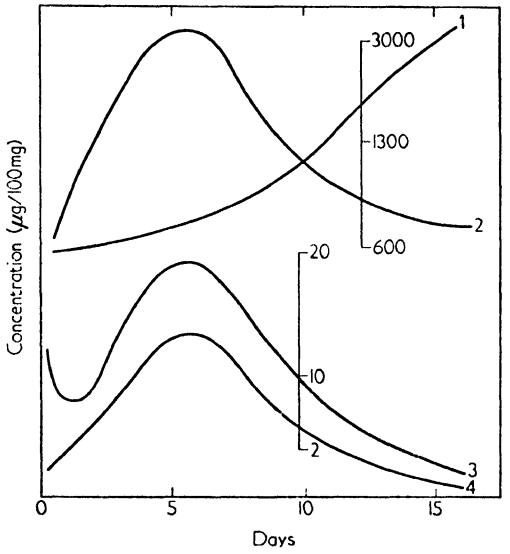


FIG. 2. Scheme of changes in the concentration of collagen (1), DNA (2), free (3) and peptide-bound (4) ultrafilterable hydroxyproline during the development of carrageenin granuloma. Data are given in mg dry substance as a reference basis.

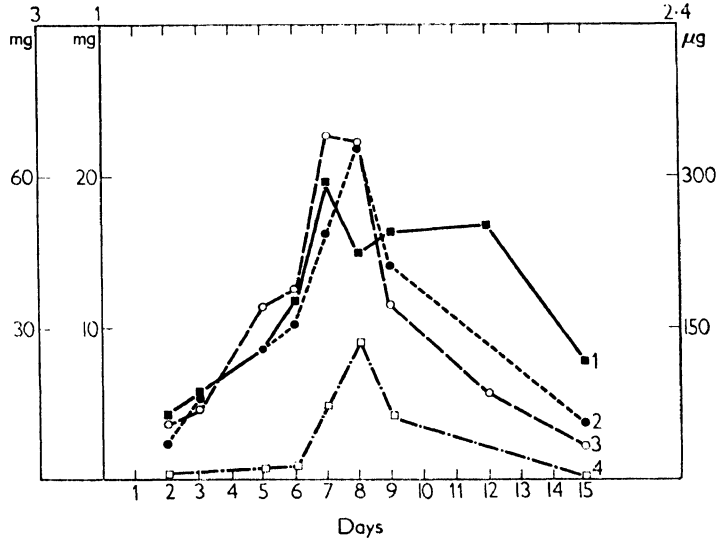


FIG. 3. Changes in the absolute amount of collagen hydroxyproline (1), free hydroxyproline (2), DNA (3) and ultrafilterable bound hydroxyproline (4) in granulation tissue from one guinea-pig.

follows the inter-relation of the collagen formation with the accumulation of cells and in addition there is a quite distinct inter-relation between the content of cells with the content of free Hypro, and in the period from the 6th to the 15th days also with the content of bound ultrafilterable Hypro. From both ways of illustration it is, however, evident that during the period of degradation of collagen structures no increase of either free or peptide Hypro occurs. Consequently, we find also in this

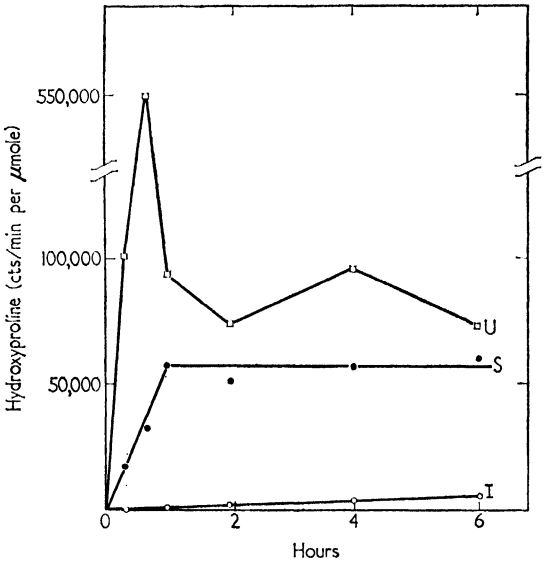


FIG. 4. Specific activities of [^{14}C]-hydroxyproline in the ultrafilterable fraction (U), soluble collagen (S), and insoluble collagen (I) isolated from skin slices of 17-day-old-chicken embryos during 6 hr incubation.

experimental arrangement the inter-relation of changes of low molecular substances containing Hypro with collagen formation (Chvapil and Čmuchalová, 1960, 1961).

A parallel histological examination shows, in agreement with other authors (Williams, 1957; Chapman, 1961), that the greatest accumulation of fibroblasts occurs only at the 7th to the 8th days of granuloma formation, that means at the time when both the free and the peptide Hypro reach the maximum. If afterwards the fractionation of granuloma cells from this period is carried out by differential centrifuging, it is of interest that Hypro found in microsomes is ultrafilterable from 80%, in mitochondria at an average from only 25%, in nuclei from 1·3% and in extracellular "heavy" fraction only from 0·1% (Chvapil *et al.*, 1962).

Intracytoplasmatic structures connected with protein formation contain predominantly low molecular bound Hypro.

Although the majority of the results mentioned hitherto would favour the fact that low molecular substances containing Hypro are interrelated with the degree of collagen synthesis, it remains to be seen whether these substances are actual precursors or possibly degradation products of collagen formation. To this effect we incubated skin slices from 17-day-old-chicken embryos and added [^{14}C]-L-proline to the incubation medium. At intervals of 20 min to 6 hr we then studied the specific activity of hydroxyproline both in the fraction obtained by ultrafiltration of neutral salt extract of homogenized skin, and in collagens soluble in 0·2N NaCl, and finally in the insoluble collagen (Hurych and Chvapil, 1962). The results are given in Fig. 4.

It has been shown in this experimental arrangement that the specific activity of the low molecular weight fraction was significantly higher during all periods. It can, therefore, be concluded that this fraction does not originate from the collagen degradation. Nevertheless, it is not possible to decide definitely whether the specific activity is due to free or to ultrafilterable peptide-bound Hypro, as the separation of these substances has not yet been made by us. In a recent paper Prockop *et al.* (1962) found, however, in a similar experimental arrangement a relatively small specific activity of free Hypro. It seems, therefore, that the high specific activity we have found is due to bound Hypro in substances of molecular weight lower than 30,000. These substances have a course of activity in agreement with the hypothesis that they are collagen precursors.

References

Chapman, J. A. (1961). *J. biophys. biochem. Cytol.* **9**, 639.

Chvapil, M. (1958). *Physiol. Bohemoslov.* **7**, 391.

Chvapil, M. (1959). *Physiol. Bohemoslov.* **8**, 186.

Chvapil, M. and Čmuchalová, B. (1960). *Nature, Lond.* **186**, 806.

Chvapil, M. and Čmuchalová, B. (1961). *Exp. Med. Surg.* **19**, 171.

Chvapil, M., Holečková, E., Čmuchalová, B., Kobrlo, V. and Hurych, J. (1962). *Exp. Cell Res.* **26**, 1.

Hurych, J. and Chvapil, M. (1962). *Biochim. biophys. Acta* **165**, 170.

Kobrle, V. and Chvapil, M. (1961). *Nature, Lond.* **190**, 909.

Prockop, D. J., Peterkofsky, B. and Udenfriend, S. (1962). *J. biol. Chem.* **237**, 1581.

Robertson, W. von B., Hiwett, J. and Herman, C. (1959). *J. biol. Chem.* **234**, 105.

Sinex, F. M., Van Slyke, D. D. and Christman, D. R. (1959). *J. biol. Chem.* **234**, 918.

Stetten, M. (1949). *J. biol. Chem.* **181**, 31.

Williams, G. (1957). *J. Path. Bact.* **73**, 557.

Wolf, G. and Berger, C. R. A. (1958). *J. biol. Chem.* **230**, 231.

Discussion

G. N. RAMACHANDRAN: I think that in non-adult tissues there is more proline than hydroxyproline and that as the tissues develop, there is a progressive hydroxylation of proline.

M. CHVAPIL: In fact, we have proved this several times (Chvapil and Kobrle (1961). *Experientia* **17**, 226). Nevertheless, I suppose this is mainly a problem in extraction and purification of protein. For instance the method of isolation of collagen from adult tissues is quite different from that used for embryonic tissues.

P. S. SHARMA (Indian Institute of Science, Bangalore): What is the influence of vitamin C?

M. CHVAPIL: There is no doubt about the influence of ascorbic acid on collagen synthesis, especially on hydroxylation of proline to hydroxyproline. In our recent studies using skin slices of chicken embryos we proved the dependence of the hydroxylation rate on ascorbic acid level in the incubation medium. It was found that there is a certain optimal concentration of ascorbic acid in the medium leading to the maximal hydroxylation. This is in agreement with findings of Robertson and Hewitt (*Biochim. biophys. Acta* **49**, 404, 1961) and Stone and Meister (*Nature, Lond.* **194**, 555, 1962).

Methods for Determining the Nature of Linkages of Certain Constituents of the Carbohydrate Moiety to the Protein Core in Skin Mucoid

S. M. Bose

Biochemistry Laboratory, Central Leather Research Institute, Madras, India

ABSTRACT

The mode of linkages of sialic acids and certain other constituents of the carbohydrate moiety to the protein core in skin mucoid was investigated. In order to determine the nature of ester linkages, mucoid was subjected to the $LiBH_4$ treatment and the loss of aspartic and glutamic acids and the extent of removal of sialic acids, hexosamines and component hexoses were determined. The reduced dicarboxylic acids were also identified in the acid hydrolysate of $LiBH_4$-treated mucoid. The sialyl compound was isolated from the dialysate of $LiBH_4$-treated mucoid and characterized by the action of neuraminidase and by periodate oxidation. A sialo-glycopeptide was isolated from mucoid by sequential digestion with pepsin and trypsin and structural studies on the peptide were carried out. The results obtained showed that the sialic acids of mucoid were bound terminally in *O*-glycosidic linkages to the galactose residues which were in turn involved in ester linkages with the β-carboxyl group of aspartyl and the γ-carboxyl group of glutamyl residues of the peptide chain. Only a small amount of the sialyl-galactose units were joined to the protein core in some other linkages. Studies on the nature of linkages in the sialo-glycopeptide confirmed the observations with respect to the original mucoid.

1. Introduction

Animal skin contains several globular proteins of which the mucoid is most important. The mucoid was found (Thabaraj *et al.*, 1962) to constitute about 5·7% of dry goatskin. Bose and Das (1956) reported that the goatskin mucoid contained about 80% protein, composed of nineteen amino acids, and about 20% carbohydrate forming an integral part of the protein molecule. Joseph and Bose (1959) reported that it contained 6·4% hexoses consisting of galactose, mannose, glucose and fucose and 1·8% hexosamines consisting of glucosamine and galactosamine. Recently, Bose (1963) found the presence of 3·3% sialic acids in skin mucoid and identified the component sialic acids as *N*-acetylneuraminic acid (NANA) and *N*-glycolylneuraminic acid, the latter constituting only 9·6% of the total sialic acid content of the mucoid. Meyer and Rapport (1952) and Meyer (1953) had reported that skin mucoid contained hyaluronic acid and chondroitin sulphate.

The sialic acid of several sialo proteins and sialyl compounds was found (Gottschalk, 1956; Heimer and Meyer, 1956) to be linked terminally in glycosidic linkages through its reducing group. Galactosamine in bovine submaxillary mucoid (Gottschalk, 1957); galactose in orosomucoid (Popenoe, 1959), ox-brain mucolipid (Rosenberg and Chargaff, 1960), human-plasma barium-alpha$_2$ glycoprotein (Kamiyama and Schmid, 1961) and trisaccharide (Kuhn and Brossmer, 1958) of human milk; and lactose in trisaccharide (Kuhn and Brossmer, 1956) of cow colostrum or rat mammary glands were shown to be the major partners of sialic acids in glycosidic linkages. A nucleotide (Comb *et al.*, 1959), cytidine-5′ monophospho-NANA, in which the carbonyl group of NANA was bound to the 5′-phosphate group of cytidine by a glycosidic bond, and nucleotide-linked polyneuraminic acid peptides (O'Brien and Zilliken, 1959) were isolated from *E. coli*. In a previous communication (Bose, 1963), it was reported that the sialic acid of skin mucoid was bound terminally through its reducing group to the adjacent galactose residues in *O*-glycosidic linkages. No information, however, is available as to how the sialic-acid-linked galactose units are joined to the amino-acid residues of the polypeptide moiety of the mucoid. In the present paper, the mode of linkages of such sialyl units to the protein core of the mucoid has been investigated.

Recently, Cook *et al.* (1960) isolated a sialo-mucopeptide from human erythrocyte. Several investigators (Cunningham *et al.*, 1957; Johansen *et al.*, 1958, 1961; Jevons, 1958) isolated glycopeptides from ovalbumin and showed that the carbohydrate was a single oligosaccharide unit linked through an aspartic acid residue. Glycopeptides of similar nature were also isolated from human γ-globulin (Rosevear and Smith, 1958). Glycopeptides were isolated from pleuromucoid (Bourrillon and Michon, 1960) by sequential action of a series of proteolytic enzymes. Weinfeld and Tunis (1961) prepared a hexose-rich fraction from orosomucoid by sequential action of pepsin and trypsin. In the present investigation, a sialo-glycopeptide has been isolated from skin mucoid and structural studies on this peptide have been carried out in order to examine the nature of linkages of sialic acid and other carbohydrate constituents with the amino acids.

2. Experimental

Mucoid was prepared from fresh goat skins and purified as described previously (Bose, 1963).

(a) *Nature of ester linkages in mucoid by lithium borohydride method*

In order to determine the nature of ester type linkages between certain constituents of the carbohydrate moiety and polypeptide chains

of mucoid, the lithium borohydride ($LiBH_4$) method (Gottschalk and Murphy, 1961; Murphy and Gottschalk, 1961) was adopted with suitable modifications. $LiBH_4$ is known to effect a reductive cleavage of ester linkages. Nystrom *et al.* (1949) showed that under suitable conditions it reduces the ester group to the corresponding alcohol group but does not react with the acid amide group. In order to examine the mode of ester linkages in ovine and bovine submaxillary gland mucoproteins, Gottschalk and Murphy (1961) and Murphy and Gottschalk (1961) submitted them after suitable pre-treatments to the action of $LiBH_4$ in anhydrous tetrahydrofuran. To increase the solubility in tetrahydrofuran, they fragmented the mucoprotein by digestion with trypsin and reduced the dipolar ion character of the fragments by converting the free amino groups into their phenylthiocarbamyl derivatives according to Edman (1950, 1953).

The mucoid was digested with crystalline trypsin (Nutritional Biochemical Corporation, U.S.A.) in presence of phosphate buffer of pH 7·8 and toluene at 37°C for 4 hr. The reaction mixture was dialysed at 4°C in presence of toluene to remove the salts and dried *in vacuo*. The dried material was dissolved in a mixture of water and *N*-allylpiperidine–pyridine buffer of pH 9·0 and after addition of phenyl*iso*thiocyanate was heated in a glass-stoppered flask at 40°C for 1 hr. The cooled mixture was extracted several times with benzene and the aqueous layer was dried *in vacuo*. The dried product was dissolved in 0·3 M $LiBH_4$ in anhydrous tetrahydrofuran and refluxed at boiling point for 12 hr. After cooling, the reaction mixture was treated with methanolic HCl and the solvent was removed by vacuum distillation. The residue was dissolved in water, neutralized and exhaustively dialysed against water (saturated with toluene) at 4°C. The dialysate and the non-dialysable material were separately concentrated *in vacuo*. The untreated mucoid and the dried residue recovered from the non-dialysable portion were analysed for sialic acids, hexoses, hexosamines, aspartic and glutamic acids in order to investigate whether and to what extent these constituents are involved in ester linkages.

Total sialic acids were estimated by the resorcinol method of Svennerholm (1958) as adopted previously (Bose, 1963).

Total hexoses and hexosamines and the individual hexose constituents were analysed by the resin hydrolysis method as described previously (Bose, 1963). The modified anthrone method of Scott and Melvin (1953) as adopted by Moss (1955) for total hexoses and the method of Elson and Morgan (1933) as adopted by Rimington (1940) for total hexosamines were followed. The descending paper-chromatographic method of Gebhardt (1960) and the method of Joseph and Bose (1959) for the identification and quantitative estimation of component hexoses were followed.

For the estimation of aspartic and glutamic acids, the material was hydrolyzed with 6N HCl in an evacuated sealed tube at 150°C for 24 hr and subjected to two-dimensional paper chromatography of Levy and Chung (1953) as described previously (Dhar and Bose, 1961). The results were corrected for the contribution from trypsin which was found (Dhar and Bose, 1961) to contain 17·2% aspartic acid and 9·7% glutamic acid. The results are presented in Table I.

TABLE I. Dicarboxylic acids and carbohydrate constituents. Balance of mucoid after $LiBH_4$ treatment

Constituent	Untreated mucoid (mg/g)	Non-dialysable portion of mucoid after treatment (mg/g*)	Percentage remaining bound to mucoid after treatment
Sialic acids (as NANA)	35·47	1·98	5·58
Aspartic acid	144·95	81·08	55·93
Glutamic acid	96·08	61·15	63·64
Total hexoses	69·86	31·20	44·66
Total hexosamines	17·48	10·55	60·35
Galactose	38·29	8·07	21·07
Mannose	17·32	13·20	76·20
Glucose	9·03	6·12	67·77
Fucose	3·81	2·86	75·06

* Refers to mucoid.

(b) *Identification of the reduced dicarboxylic acids in acid hydrolysate of $LiBH_4$-treated mucoid*

As appreciable loss of dicarboxylic acids after treatment of mucoid with $LiBH_4$ was observed, the reductive cleavage of ester linkages involving free carboxyl groups of aspartic and glutamic acids was evident. In order to find out the position of the reduced carboxyl groups (α or β in aspartyl, α or γ in glutamyl), the acid hydrolysate of the $LiBH_4$-treated mucoid was examined for the formation of homoserine and α-amino-δ-hydroxy-*n*-valeric acid, as suggested by Murphy and Gottschalk (1961). The mucoid after $LiBH_4$ treatment was hydrolysed with 6N HCl and subjected to two-dimensional paper chromatography of Levy and Chung (1953) as adopted by Dhar and Bose (1961). Using a guide chromatogram, the overlapped spots were cut out from the undeveloped duplicate chromatogram, eluted with slightly acidic water and re-chromatographed using different solvent systems as described

previously (Dhar and Bose, 1961; Bose and Das, 1956). Homoserine and α-amino-δ-hydroxy-*n*-valeric acid were identified with reference to the control chromatogram of authentic samples of the same acids which were prepared by the method of Murphy and Gottschalk (1961) by $LiBH_4$ treatment of β-methyl aspartate (Coleman, 1951) and γ-methyl glutamate (Coleman, 1951; Kovacs *et al.*, 1953) respectively. The results indicated that the β-carboxyl group of aspartyl and the γ-carboxyl group of glutamyl residues of mucoid were involved in ester linkages.

In order to find out if asparagine and glutamine residues were present in mucoid, its amide nitrogen was determined by the method of Gottschalk and Simmonds (1960) after removal of terminal sialic acids by the action of *Cl. perfringens* neuraminidase (Bose, 1963) and dialysis. They observed that the determination of amide nitrogen in sialic acid containing mucoproteins is possible only after enzymic removal of the terminal sialic acid residues. The amide nitrogen content of mucoid, however, was found to be very low.

(c) *Isolation and characterization of sialyl-compounds from the dialysate of* $LiBH_4$*-treated mucoid*

As major amounts of sialic acid and galactose residues were found to be removed from mucoid by $LiBH_4$ treatment and as it was observed previously (Bose, 1963) that in mucoid sialic acid was joined terminally to galactose residues by *O*-glycosidic linkage, it was of interest to isolate and characterize the sialic acid containing compounds from the dialysate of the $LiBH_4$-treated mucoid. The dialysate which was concentrated *in vacuo* was passed through an ion-exchange column of Dowex-50 (H^+ form). The effluent was subjected to descending paper chromatography (Gottschalk and Graham, 1959) in Whatman No. 1 paper using butanol–pyridine–water (6:4:3) solvent. Using a guide chromatogram, the spots for the sialyl compounds were cut out from a number of undeveloped chromatograms, eluted with water and the eluate was concentrated *in vacuo*.

An aliquot of the eluate was subjected to mild acid hydrolysis and descending paper chromatography for the identification of sialic acid and hexose component as described previously (Bose, 1963). Another aliquot was subjected to the action of *Cl. perfringens* neuraminidase (Bose, 1963) and similarly examined. In both cases, NANA and galactose were identified.

The eluate was subjected to periodate oxidation under the same conditions as adopted previously (Bose, 1963). After the reaction the excess periodate was reduced by ethylene glycol and the mixture was lyophilized. It was analysed for NANA by the resorcinol method of Svennerholm (1958) and for galactose by paper chromatographic method

as described previously (Bose, 1963). For comparison, the original mucoid was subjected to periodate oxidation and similarly analysed for sialic acid and galactose content. As control, the dried materials before periodate oxidation were analysed for the same constituents. The results are presented in Table II.

TABLE II. Effect of periodate oxidation of mucoid and the sialyl-compound (isolated from the dialysate of $LiBH_4$-treated mucoid) on the recovery of sialic acids and galactose

Material	Periodate oxidation (hr)	Sialic acids (mg NANA/g)	Galactose (mg/g)
Mucoid	Nil	35·47	38·29
Mucoid	5	12·06	37·82
Sialyl-compound	Nil	576	331
Sialyl-compound	5	170	198

(d) *Isolation of sialo-glycopeptide from mucoid*

The mucoid was subjected to sequential digestion with pepsin and trypsin by the method of Weinfeld and Tunis (1961). It was digested with crystalline pepsin (Worthington Biochemical Corporation, U.S.A.) in dilute HCl solution of pH 2·3 for 2 hr at 37°C. The digestion mixture was cooled, adjusted to pH 4·2 and the precipitate produced by the addition of 4 vol. acetone and 0·1 vol. 5N NaCl was recovered by centrifugation and freed of acetone *in vacuo*. The precipitate was digested with crystalline trypsin (Nutritional Biochemical Corporation, U.S.A.) in phosphate buffer of pH 7·8 at 37°C for 2 hr. The mixture was cooled, treated with acetone and NaCl solution and the resulting precipitate was recovered and freed of acetone as before. The precipitate was dialysed against water (saturated with toluene) at 4°C and the content of the dialysis bag was passed through a column of Dowex-50 (H^+ form). The effluent was concentrated *in vacuo* and subjected to paper chromatography (Cook *et al.*, 1960) on Whatman No. 1 paper with butanol–acetic acid–water (4:1:5). In addition to a number of ninhydrin-positive spots which migrated from the starting point, a slow moving spot containing sialic acid was also detected. Using a guide chromatogram, this sialic acid containing material was eluted with 10% (v/v) aqueous isopropanol (Cook *et al.*, 1960) from a number of air-dried untreated chromatograms and the eluate was concentrated *in vacuo*.

(e) *Structural studies on the sialo-glycopeptide isolated from mucoid*

An aliquot of the eluate was hydrolysed with 6 N HCl and examined for the component amino acids by paper chromatography of Levy and Chung (1953) as adopted by Dhar and Bose (1961). The DNP-method of Sanger (1945) and Sanger and Thompson (1953) and the hydrazinolysis method of Akabori *et al.* (1952) as modified by Bradbury (1956) were followed for the identification (Joseph and Bose, 1958) of N- and C-terminal amino acids of the peptide respectively.

The peptide was incubated with crystalline carboxypeptidase (L. Light and Co., England) in the presence of phosphate buffer of pH 7·8

TABLE III. Characterization of sialo-glycopeptide isolated from mucoid

Tests	Results
Component amino acids	Phenylalanine, aspartic acid, leucine, threonine and serine
N-terminal amino acid	Phenylalanine
C-terminal amino acid	Serine
Action of carboxypeptidase	Leucine, threonine and serine liberated
Hexoses	Galactose, mannose (trace)
Hexosamines	Traces
Sialic acids	NANA
Hydroxylamine test	Positive
Action of neuraminidase	NANA released

at 37°C for 12 hr. The liberated amino acids were identified by dinitrophenylation, extraction of DNP-amino acids with ether and one-dimensional paper chromatography of DNP-amino acids as reported previously (Joseph and Bose, 1958).

The component hexoses, hexosamines and sialic acid of the peptide were identified by paper-chromatographic methods as reported previously (Bose, 1963).

The presence of ester linkage in the peptide was examined by the hydroxylamine method of Hestrin (1949).

The peptide was subjected to the action of *Cl. perfringens* neuraminidase under the same conditions as adopted previously (Bose, 1963) and the liberated sialic acid was identified by paper chromatography as mentioned before. The results obtained are presented in Table III.

3. DISCUSSION

It may be seen from Table I that treatment of nucoid with $LiBH_4$ effected appreciable loss of aspartic and glutamic acids content along with the removal of major amounts of sialic acid and galactose residues

and also comparatively smaller amounts of hexosamines and other hexoses. Practically the same amount of dicarboxylic acids was found in original mucoid and also in its trypsin digest after dialysis, indicating no loss of aspartic and glutamic acids during dialysis. Treatment of mucoid with $LiBH_4$ for more than 12 hr did not result in further loss of the dicarboxylic acids.

An analysis of the dialysate for total sialic acid content showed that only about 58% of the sialic acids which were removed from mucoid by $LiBH_4$ treatment could be recovered from the dialysate, indicating loss of some sialic acids during $LiBH_4$ treatment. Gottschalk and Murphy (1961) also reported loss of sialic acids by $LiBH_4$ treatment of ovine submaxillary gland mucoprotein and sialyl-galactosamine compound and explained that a portion of the released prosthetic groups underwent further chemical changes on heating in the system $LiBH_4$-tetrahydrofuran.

The sialyl-compound which was isolated from the dialysate of $LiBH_4$-treated mucoid was found to consist of only NANA and galactose. It was observed that NANA was easily released from the sialyl-compound by the action of neuraminidase which is very specific in liberating sialic acids from sialyl-compounds by the hydrolytic cleavage of *O*-glycosidic links. It was observed previously (Bose, 1963) that essentially all the sialic acids could be released from the original mucoid by prolonged action of neuraminidase. The periodate oxidation (Table II) of mucoid led to the rapid destruction of sialic acids without measurable loss of the galactose content whereas both NANA and galactose components of the sialyl-compound were susceptible to destruction by periodate. All this evidence indicated that the sialic acids were bound terminally in *O*-glycosidic linkages to the galactose residues of the mucoid.

Again, the loss of dicarboxylic acids concomitantly with the release of sialyl-galactose units after $LiBH_4$ treatment of mucoid showed that the galactose residues were in turn involved in ester linkages with the free carboxyl groups of aspartyl and glutamyl residues. The loss of dicarboxylic acids, however, was found to be much greater than what could be accounted for by their engagement in ester linkages with the sialyl-galactose units. It thus appears that other constituents like hexosamines, hyaluronic acid, chondroitin sulphate and other hexoses or the free hydroxyl groups of polypeptide chains may also be involved to some extent in ester linkages with the free carboxyl groups of aspartyl and glutamyl residues. The identification of homoserine and α-amino-δ-hydroxy-n-valeric acid in the acid hydrolysate of $LiBH_4$-treated mucoid indicated the engagement of the β-carboxyl group of aspartyl and the γ-carboxyl group of glutamyl residues in ester linkages. Chibnall and Rees (1958) reported that if the β- and γ-carboxyl groups of α-

aspartyl and α-glutamyl residues (in peptide chain or N-terminal) of proteins are involved in ester linkages and are reduced by $LiBH_4$ treatment, they would be present in the acid hydrolysate of the treated protein as α-amino-γ-hydroxybutyric acid and α-amino-δ-hydroxyvaleric acid respectively. Gottschalk and Murphy (1961) and Murphy and Gottschalk (1961) reported that in ovine and bovine submaxillary gland mucoproteins approximately 80% of the prosthetic groups were involved in glycosidic-ester linkages to the β-carboxyl of aspartyl and the γ-carboxyl of glutamyl residues and the residual prosthetic groups were linked by *O*-glycosidic bonds to serine and/or threonine residues. In the present investigation, only 5·6% of the total sialic acids was found to remain bound to the mucoid after $LiBH_4$ treatment, suggesting that a small amount of sialyl-galactose units may be linked in some other way, e.g. *N*- or *O*-glycosidic linkage with free amino or hydroxyl group of the peptide chain. According to Chibnall and Rees (1958) and Crawhall and Elliott (1955), the reduction of the amide groups of proteins by $LiBH_4$ is negligible and the reductive cleavage of peptide bonds under standard conditions does not amount to more than about 2% of the total peptide bonds and α-glutamyl and α-aspartyl residues are very resistant. Murphy and Gottschalk (1961) reported that $LiBH_4$ treatment does not reduce the free carboxyl groups of aspartyl and glutamyl residues of proteins.

The isolation of a sialo-glycopeptide from mucoid showed that the carbohydrate moiety was firmly linked to the polypeptide chain in a covalent linkage. The main amino acid constituents of the peptide were found to be phenylalanine, aspartic acid, leucine, threonine and serine; phenylalanine and serine being the N- and C-terminal amino acids of the peptide respectively. It was observed that only three amino acids, viz. leucine, threonine and serine were liberated by the action of carboxypeptidase and the peptide contained galactose and NANA as the major carbohydrate constituents, mannose and hexosamines being present only in traces. NANA was found to be easily released from the peptide by the action of neuraminidase. The presence of ester linkage in the peptide was also shown. An interpretation consistent with all these observations (Table III) and also with other results obtained on the original mucoid would be that in this sialo-glycopeptide NANA was bound in *O*-glycosidic linkage to the galactose residue which again was mainly joined in ester linkage to the β-carboxyl group of aspartic acid, the α-carboxyl group of which was involved in usual peptide linkage with leucine or threonine, the α-amino group of aspartic acid being normal peptide bonded with phenylalanine. The isolation of a sialo-glycopeptide of such structure from the mucoid indicated that such sialic acid linked galactose units were attached in ester linkages to a number of aspartyl and glutamyl residues in the protein core.

Acknowledgements

My thanks are due to Dr. Y. Nayudamma, Director, Central Leather Research Institute, for his kind encouragement in this work and permission to publish these results. I am very grateful to Dr. S. Roseman, Rackham Arthritis Research Unit, University of Michigan, U.S.A., and to The Dow Chemical Company, Midland, U.S.A., for the gift of authentic samples of NANA and Dowex-1 respectively, which were used in this work.

References

Akabori, S., Ohno, K. and Narita, K. (1952). *Bull. Chem. Soc. Japan* **25**, 214.
Bose, S. M. (1963). *Biochim. biophys. Acta* (in press).
Bose, S. M. and Das, B. M. (1956). *J. Amer. Leath. Chem. Ass.* **51**, 647.
Bourrillon, R., and Michon, J. (1960). *Biochim. biophys. Acta* **44**, 608.
Bradbury, J. H. (1956). *Nature, Lond.* **178**, 912.
Chibnall, A. C. and Rees, M. W. (1958). *Biochem. J.* **68**, 105.
Coleman, D. (1951). *J. chem. Soc.* 2294.
Comb, D. G., Shimizu, F. and Roseman, S. (1959). *J. Amer. chem. Soc.* **81**, 5513.
Cook, G. M. W., Heard, D. H. and Seaman, G. V. F. (1960). *Nature, Lond.* **188**, 1011.
Crawhall, J. C. and Elliott, D. F. (1955). *Biochem. J.* **61**, 264.
Cunningham, L. W., Nuenke, B. G. and Nuenke, R. B. (1957). *Biochim. biophys. Acta* **26**, 660.
Dhar, S. C. and Bose, S. M. (1961). *Bull. Centr. Leath. Res. Inst.* **8**, 1.
Edman, P. (1950). *Acta chem. scand.* **4**, 277, 283.
Edman, P. (1953). *Acta chem. scand.* **7**, 700.
Elson, L. A. and Morgan, W. T. J. (1933). *Biochem. J.* **27**, 1824.
Gebhardt, D. O. E. (1960). "A Biochemical Study on the Development of Collagen." Doctorate Thesis, University of Amsterdam, p. 38.
Gottschalk, A. (1956). *Biochim. biophys. Acta* **20**, 560.
Gottschalk, A. (1957). *Biochim. biophys. Acta* **24**, 649.
Gottschalk, A. and Graham, E. R. B. (1959). *Biochim, biophys. Acta* **34**, 380.
Gottschalk, A. and Murphy, W. H. (1961). *Biochim. biophys. Acta*, **46**, 81.
Gottschalk, A. and Simmonds, D. H. (1960). *Biochim. biophys. Acta* **42**, 141.
Heimer, R. and Meyer, K. (1956). *Proc. nat. Acad. Sci., Wash.* **42**, 728.
Hestrin, S. (1949). *J. biol. Chem.* **180**, 249.
Jevons, F. R. (1958). *Nature, Lond.* **181**, 1346.
Johansen, P. G., Marshall, R. D. and Neuberger, A. (1958). *Nature Lond.* **181**, 1345.
Johansen, P. G., Marshall, R. D. and Neuberger, A. (1961). *Biochem. J.* **78**, 518.
Joseph, K. T. and Bose, S. M. (1958). *Bull. Centr. Leath. Res. Inst.* **4**, 343.
Joseph, K. T. and Bose, S. M. (1959). *Bull. Centr. Leath. Res. Inst.* **6**, 132.
Kamiyama, S. and Schmid, K. (1961). *Biochim. biophys. Acta* **49**, 250.
Kovacs, J., Bruckner, V. and Kovacs, K. (1953). *J. chem. Soc.* 145.
Kuhn, R. and Brossmer, R. (1956). *Angew. Chem.* **68**, 211; *Chem. Ber.* **89**, 2013.
Kuhn, R. and Brossmer, R. (1958). *Angew. Chem.* **70**, 25.
Levy, A. L. and Chung, D. (1953). *Analyt. Chem.* **25**, 396.
Meyer, K. (1953). *Disc. Faraday Soc.* **13**, 271.
Meyer, K. and Rapport, M. M. (1952). *Advanc. Enzymol.* **13**, 199.
Moss, J. A. (1955). *Biochem. J.* **61**, 151.
Murphy, W. H. and Gottschalk, A. (1961). *Biochim. biophys. Acta* **52**, 349.
Nystrom, R. F., Chaiken, S. W. and Brown, W. G. (1949). *J. Amer. chem. Soc.* **71**, 3245.

O'Brien, P. J. and Zilliken, F. (1959). *Biochim. biophys. Acta* **31**, 543.
Popenoe, E. A. (1959). *Biochim. biophys. Acta* **32**, 584.
Rimington, C. (1940), *Biochem. J.* **34**, 931.
Rosenberg, A. and Chargaff, E. (1960). *Biochim. biophys. Acta* **42**, 357.
Rosevear, J. W. and Smith, E. L. (1958). *J. Amer. chem. Soc.* **80**, 250.
Sanger, F. (1945). *Biochem. J.* **39**, 507.
Sanger, F. and Thompson, E. O. P. (1953). *Biochem. J.* **53**, 353.
Scott, T. A. and Melvin, E. H. (1953). *Analyt. Chem.* **25**, 1656.
Svennerholm, L. (1958). *Acta chem. scand.* **12**, 547.
Thabaraj, G. J., Bose, S. M. and Nayudamma, Y. (1962). *Bull. Centr. Leath. Res. Inst.* (In press).
Weinfeld, H. and Tunis, M. (1961). *Biochim. biophys. Acta* **50**, 590.

Discussion

M. CHVAPIL: I would like to mention that we are adopting another approach to the understanding of the significance and linkage of mucopolysaccharides in collagen. We studied the kinetics of degradation of collagen fibres on adult rat tail tendon, during chemical contraction and relaxation in 2·5 M $NaClO_4$. The results indicate that there are at least three types of hexosamine-containing compounds in collagen. The proportion of individual hexosamine compounds in individual periods of contraction and relaxation of the fibres is age dependent. Approximately 60% of the total hexosamine compounds present in collagen fibre is firmly bound.

S. M. BOSE: Probably the hexosamine compounds which are firmly bound in collagen fibre are involved in covalent linkages. There are at least two types of linkages, viz., strong linkages of covalent type and salt linkages of electrovalent type. The stability of covalent linkages is generally greater than that of salt linkages.

P. S. SARMA (Indian Institute of Science, Bangalore): To what extent is the sialo-glycopeptide present in the total mucoid? How much is the particular peptide that you have isolated present in the whole of the mucoid?

S. M. BOSE: The isolation of the sialo-glycopeptide from the mucoid and the structural studies on the particular peptide isolated have been done only on a qualitative basis. As is evident from the different steps involved in the isolation process, it is difficult to carry out such work quantitatively.

SECTION VI
General Discussion

Strategy of Protein Research

Edited by

J. T. EDSALL

On the final afternoon of the conference, on January 18, an informal session was held in order to discuss the present status of the problems of protein structure and protein biosynthesis and to consider the strategy of future advance. About twenty-five members of the conference attended the session, which lasted for nearly two and a half hours. The discussion was very lively. No attempt will be made here to give any general summary of what went on, since the meeting was held on the understanding that no formal record would be kept. The highly informal atmosphere of the proceedings encouraged the ready discussion of unconventional ideas that were not yet fully developed. However, we mention here a few of the points that were brought up.

The earlier part of the session concerned the future of X-ray diffraction studies on proteins. The difficulties of interpretation of X-ray work on fibres compared with single crystals were discussed first. With regard to crystals, Dr. Harker appealed to protein chemists in general to concern themselves more systematically with the techniques of protein crystallization, and to describe their procedures in sufficient detail so that others may readily repeat them. If such information is fully recorded in the published literature, the number of protein crystals potentially available for study by crystallographers will be considerably increased. Generally the crystallographer will need to grow larger crystals than the protein chemists have troubled themselves to prepare; however, with adequate information how the protein chemist obtained his crystals, it is generally possible for the crystallographer, with a little patience and by slight modifications of the technique, to grow crystals that will be large enough to serve well for X-ray analysis. Dr. Harker estimated that a crystal of side about 0·5 mm is a good working size for protein crystals in X-ray work. Smaller crystals can also be used.

Attention was then turned to the problem of attaching heavy atoms to protein molecules in crystals. The procedures required for different crystals, and the kinds of heavy atom derivatives that are most suitable, differ widely from one protein to another. The best isomorphous derivatives of ribonuclease have been prepared by quite

different methods than those that Kendrew and Perutz had found to be successful in the case of haemoglobin and myoglobin.

X-Ray diffraction work on a number of crystalline proteins is now under way in at least 8 or 10 different laboratories and, in view of the success now being achieved, it is certain that the number of such laboratories will increase markedly within the next few years. There will be an urgent need for improving the means for automatic collection and coding of data, and a good deal of large-scale equipment will be needed for speeding up computations and improving their accuracy. This will undoubtedly impose some heavy demands on the agencies that make grants in support of research.

Next, some of the problems of refining our structural information on DNA and RNA were considered by Dr. Williams and others.

Dr. Moore discussed the prospects for further advances in the techniques of amino acid analysis and of sequence determinations in polypeptide chains. It is obviously of great importance to make such methods simpler, more rapid, and more nearly automatic, without sacrificing reliability. Also the more the scale of operations can be reduced the larger the number of rare proteins that can be studied. There was considerable discussion of possible means of achieving these very important ends.

Dr. Ochoa was interested in the present status of the coding problem, while Dr. Katchalski discussed the various aspects of the chemistry of polyamino acids and the possibilities they offer for studies related to protein structure.

There was some discussion of the factors in the primary structure of proteins which serve to determine the tertiary structure. However, the members of the group in general concluded that as yet our knowledge is far too limited for adequate analysis of this problem.

Dr. Ramachandran raised the question whether all collagens are alike in detail, in somewhat the same sense that we can say that globular protein crystals consist of molecules that are all alike in detail. Dr. Hodge and others discussed this point further.

Author Index

Numbers in italics indicate the page on which the reference is listed.

A

Abelman, M. B., 13, *22*
Ackroyd, J. F., 171, *174*
Akabori, S., 69, 70, *80*, 363, *366*
Alexander, P., 99, *100*
Allen, F. W., 316, *317*
Allfrey, V. G., 172, 173, *174*
Altgelt, K., 280, *287*, 296, *299*
Ambady, G. K., 40, 42, *53*
Ambrose, E. J., 259, *263*
Andersen, S. O., 339, 340, *340*
Andreeva, N. S., 138, *143*
Anesey, J., 280, *287*
Anfinsen, C. B., 286, *287*, 310, 314, *317*
Arndt, U. W., 58, *66*
Asadourian, A., 200, *202*, 262, *263*
Astbury, W. T., *37*, 103, 104, 113, *118*, 346, *348*
Attardi, G., 168, *174*
Avey, H. P., 21, *21*

B

Baba, H., 187, *202*
Badger, R. M., 258, *263*
Baglioni, C., 166, *174*
Bailey, K., 337, 338, *340*
Bamford, C. H., 81, *91*, 104, *118*
Bargetzi, J. -P., 321, 322, 327, 328, 329, 331, 332, *335*
Basch, J. J., 166, *174*
Basilio, C., 147, 148, 149, 150, 151, 152, 157, *158*, *159*, 165, *175*
Bear, R. S., 40, *53*, 93, *100*, 290, 292, 293, 297, *299*
Beaven, G. H., 180, *202*, 302, *305*
Beer, M., 257, 259, 261, *263*
Beevers, C. A., 129, 132, *134*
Bennett, D. A. H., 172, *174*
Bensusan, H. B., 295, *299*
Benzer, S., 157, *159*
Berger, A., 81, 82, *91*, *92*, 138, *143*, 206, 207, 208, 209, 211, 212, 213, 214, 215, 216, 218, 221, 225, 226, 227, 229, *237*, *238*
Berger, C. R. A., 352, *355*
Beychok, S., 243, 246, 247, 250, 251, *255*
Bigelow, C. C., 187, 192, *202*
Bigwood, E. J., 313, *317*
Billingham, R. E., 169, *174*
Birbeck, M. S. C., 94, *100*
Bishop, J., 149, *158*
Blake, C. C. F., 59, 64, *66*, *67*
Blakey, P. R., 106, 107, 112, 113, 115, 117, *118*
Block, R. J., 343, 344, 345, 346, *348*
Blout, E. R., 82, 83, *91*, 200, 201, *202*, 206, 207, 210, 234, *237*, *238*, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, *255*, 259, 260, 262, *263*, *264*
Blow, D. M., 15, *21*, 66, *67*
Boedtker, H., 50, *53*, 279, *287*, 290, 293, 294, *299*
Bolling, D., 343, 344, 345, 346, *348*
Bose, S. M., 357, 358, 359, 360, 361, 362, 363, 364, *366*, *367*
Bourrillon, R., 358, *366*
Bovey, F. A., 185, 186, *203*
Bradbury, E. M., 137, *143*, 248, *255*
Bradbury, J. H., 363, *366*
Bragg, W. L., 79, *80*
Branson, H. R., 122, *134*, 137, *144*
Braunitzer, G., 184, *202*, 315, *317*
Braverman, M. H., 173, *174*
Brent, L., 169, *174*
Breslow, E., 182, *202*
Bretscher, M. S., 153, *158*
Brossmer, R., 358, 361, *366*
Brown, C. J., 129, *134*
Brown, G. L,, *37*
Brown, J. R., 320, *335*
Brown, L., 104, *118*, 137, *143*

Brown, W. G., 359, *366*
Bruckner, V., 361, *366*
Buckley, C. E., 189, *203*
Burge, R. E., 49, *53*, 143, *143*
Burley, R. W., 343, 344, 345, 346, *348*
Burnet, F. M., 168, 171, *174*
Busch, H., 173, *174*

C

Cannon, C. G., 258, *263*
Carlisle, H. C., 21, *21*
Carlström, D., 132, 134, *134*
Carver, J. P., 253, 254, *255*
Chaiken, S. W., 359, *366*
Chapman, J. A., 354, *355*
Chargaff, E., 358, *367*
Chase, M. W., 169, *174*
Chatterjea, J. B., 302, *305*
Chatterji, S. N., 302, *305*
Chernoff, A. F., 302, *305*
Chervenka, C. H., 184, 193, *202*
Chibnall, A. C., 364, 365, *366*
Chirgadze, Yu. N., 138, *143*, 264
Christman, D. R., 352, *355*
Chung, D., 360, 363, *366*
Chvapil, M., 351, 352, 354, 355, *355*, 356
Čmuchalová, B., 354, *355*
Cochran, W., 86, *91*, 95, *100*
Cohen, C., 40, *53*, 242, 243, 245, 246, 247, *255*
Cohen, E., 319, 320, *335*
Cohen, J., 280, *287*
Cohn, M., 168, *174*
Coleman, D., 361, *366*
Coleman, J. E., 325, 326, 329, *335*
Comb, D. G., 358, *366*
Cook, G. M. W., 358, 362, *366*
Coombs, T. L., 319, 324, 327, *335*
Corey, R. B., 65, *67*, 81, 87, *91*, *92*, 96, 98, *100*, *101*, 104, 113, 114, *118*, 122, 127, *134*, 137, 138, *144*, 242, *255*
Cowan, P. M., 40, 49, 51, *53*, 82, *91*, 137, 138, *143*, 206, 207, 209, *237*
Cowen, D. M., 170, *174*
Cox, D. J., 320, 321, *335*
Craig, L. C., 184, *202*
Crammer, J. L., 183, 184, 185, *202*
Crampton, C. F., 173, *174*
Crawhall, J. C., 365, *366*
Crestfield, A. M., 311, 312, 313, 314, 315, 316, *317*
Crick, F. H. C., *37*, 42, 49, 50, 51, *53*, *54*, 65, *67*, 86, *91*, 93, 95, 96, *100*, *101*, 113, 115, *118*, 127, *134*, 138, 143, *144*, 148, *158*, 206, 218, *237*, *238*, 253, *255*, 279, *287*
Cullis, A. F., 62, *67*, 304, *305*
Cunningham, L. W., 358, *366*

D

Damodaran, M., 346, 347, *348*
Das, B. M., 357, 361, *366*
Das Gupta, N. N., 302, *305*
Davidson, J. N., *37*
Davies, D. R., 58, *67*
Davis, J. R., 173, *174*
De, P. K., 189, *203*
Debabov, V. A., 138, *143*
Deffner, G. G. J., 293, 294, 295, *300*
De Gasquet, P., 172, *174*
de Lozé, C., 200, *202*, 243, 246, 247, 250, *255*, 262, *263*
Dhar, S. C., 360, 361, 363, *366*
Dhavalikar, R. S., 346, 347, *348*
Dickerson, R. E., 58, 65, *67*, 81, *91*, 250, *255*
Dintzin, H. M., 149, *158*
Dixon, F. J., 168, *174*, *175*
Donohue, J., 87, *91*, 96, 98, *100*, 127, *134*, 137, *143*
Donovan, J. W., 187, 188, 196, *202*
Doty, P., 50, 51, *53*, 179, 180, 191, 198, 200, 201, *202*, *203*, 242, 243, 245, 247, 250, *255*, *256*, 259, 262, *263*, 279, *287*, 290, 293, 294, *299*, *300*
Downie, A. R., 137, *143*, 208, *237*, 248, *255*
du Vigneaud, V., 309, *317*

E

Edman, P., 314, *317*, 359, *366*
Edmundson, A. B., 61, *67*
Edsall, J. T., 81, *91*, 184, 187, 188, 195, *203*
Efron, M. L., 166, *174*
Eisenberg, D., 179, 197, *202*
Elliott, A., 81, *91*, 104, *118*, 137, *143*, 248, *255*, 258, 259, *263*
Elliott, D. F., 365, *366*
Elliott, S. D., 313, 314, *317*
Elson, L. A., 359, *366*
Endres, H., 295, *300*

Engel, J., 51, *53*, 279, 280, 281, 283, *287*
Eriksson, S., 314, *317*
Exum, E. D., 171, *174*

F

Fasman, G. D., 82, 83, *91*, 206, 207, 210, 225, 229, 234, *237*, *238*, 247, 248, 250, 253, *255*
Fenn, R. H., 64, *67*
Ferrier, W. G., 129, *134*
Fessler, J. H., 296, *299*
Field, E. O., 302, 303, *305*
Fitton-Jackson, S., 298, *299*
Fitts, D. D., 82, *91*
Flodin, P., 311, *317*
Flory, P. J., 280, *287*
Ford, C. E., 170, *174*
Foster, J. F., 185, 190, *203*
Fraenkel, G., 103, *118*
Franklin, R. E., *37*
Franzblau, C., 138, *144*
Fraser, R. D. B., 100, *100*, 109, 113, *118*, 137, *143*, 258, *263*
Frederickson, D. S., *175*
Fridrichsons, J., 79, *80*
Fukushima, K., 261, *264*
Fuller, W., *37*, 42, *53*
Furnas, T. C., 13, *22*

G

Gallop, P. M., 138, *144*, 281, *287*
Gardner, R. S., 148, 150, 151, 152, 157, *158*, *159*, 165, *175*
Gascoyne, T., 172, *174*
Gebhardt, D. O. E., 359, *366*
Gehring-Muller, R., 184, *202*
Gerald, P. A., 166, *174*
Geschwind, I. I., 187, *202*
Glazer, A. N., 179, 180, 184, 197, *202*, 245, *255*
Glimcher, M. J., 298, *299*, *300*
Glinos, A. D., 173, *174*
Goldstein, A., 149, *158*
Gordon, W. G., 166, *174*
Gosling, R. G., *37*
Gottschalk, A., 358, 359, 360, 361, 364, 365, *366*
Gouterman, M., 187, *203*
Gowans, J. L., 170, *174*
Graham, E. R. B., 361, *366*
Grassmann, W., 51, *53*, 138, *143*, 279, 280, 281, *287*, 295, *300*
Gratzer, W. B., 179, 191, 201, *202*, *203*, 245, *255*, 262, *263*, 302, *305*
Green, E. H., 343, 344, 346, *348*
Greenshields, R. N., 320, *335*
Grenville-Wells, H. J., 98, *101*
Gross, J., 50, 51, *53*, *54*, 253, 254, *255*, 289, 290, *300*, 347, *348*
Grunberg-Manago, M., 153, *158*
Guidotti, G., 184, *202*
Gustavson, K. H., 348, *348*
Gutter, F. J., 302, 303, *305*

H

Haber, E., 310, 314, *317*
Hall, C. E., 51, *53*, *277*, 290, *300*
Hall, D. A., 115, 116, *118*
Halpern, A., 245, *256*
Hamilton, L. D., *37*, 168, *174*
Hanby, W. E., 81, *91*, 104, *118*, 137, *143*, 248, *255*
Hannig, K., 51, *53*, 138, *143*, 279, 280, 281, 286, *287*, 295, *300*
Happey, F., 103, 104, 106, 107, 109, 111, 113, 115, 116, 117, *118*
Harker, D., 13, 21, *22*
Harrington, W. F., 82, *91*, *92*, 206, 207, 208, 209, 211, 212, 214, 215, 219, 221, 226, 227, *237*, *238*, 280, *287*, 292, 296, *300*
Harris, G., 228, 229, *237*
Harris, H., *175*
Hart, R. G., 58, *67*
Hauenstein, J. D., 183, 184, *203*
Haurowitz, F., 167, *174*
Havsteen, B. H., 184, *202*
Heard, D. H., 358, 362, *366*
Heimer, R., 358, *366*
Heppel, C. A., 148, 156, *158*, *159*
Herman, C., 352, *355*
Hermans, J., Jr., 182, 184, 187, 196, *202*
Herskovits, T. T., 179, 189, 190, 191, *202*
Hess, G. P., 184, *202*
Hestrin, S., 363, *366*
Hewitt, J., 356
Hewitt, L. F., 242, *255*
Highberger, J. H., 50, *54*, 289, 290, 293, 294, 295, *300*
Hill, R. J., 184, *202*, 315, 316, *317*
Hill, R. L., 166, *174*, *175*
Hilmoe, R. J., 148, *159*
Hilschmann, N., 184, *202*

Hilse, K., 184, *202*
Hirs, C. H. W., 61, *67*, 310, 312, 314, 315, 316, *317*
Hiwett, J., 352, *355*
Hobom, G., 184, *202*
Hoch, F. L., 319, 327, *335*
Hodge, A. J., 280, *287*, 290, 291, 292, 293, 294, 295, 296, 297, 298, *299*, *300*
Holečková, E., 354, *355*
Holley, R. W., 157, *159*
Hollingworth, B. R., 187, 188, *203*
Holzwarth, G. M., 201, *202*, *203*, 245, *255*
Hooper, C. W., *37*
Hordvik, A., 129, *134*
Horne, R. W., 302, *305*
Horton, W. G., 115, *118*
Horwitt, M. K., 343, 344, 345, 346, *348*
Houbata, K., 168, *174*
Hudson, R. F., 99, *100*
Hummel, J. P., 179, 193, 194, *203*
Hunt, J. A., 166, *174*
Hurych, J., 354, 355, *355*
Hyde, A. J., 116, 117, *118*

I

Idelson, M., 242, *255*
Imahori, K., 179, 180, 198, *203*, 250, *256*, 262, *263*, *264*
Ingram, V. M., 166, *174*
Ioffe, K. G., 138, *144*

J

Jacobson, R. A., 129, *134*
Jacquot, R., 346, *348*
Jensen, M., 338, *340*
Jessen-Hansen, H., 242, *255*
Jevons, F. R., 358, *366*
Jim, R. T. S., 166, *175*
Johansen, P. G., 358, *366*
Johnson, D. J., 113, *118*
Jones, D. W., 132, *134*
Jones, F. B., 58, *66*
Jones, O. W., 151, *158*, 165, *174*
Joseph, K. T., 357, 359, 363, *366*
Jukes, T. H., 150, 156, *158*, 165, *174*

K

Kakudo, M., 70, 73, *80*
Kalan, E. B., 166, *174*
Kallos, P., *175*
Kamiyama, S., 358, *366*
Karlson, R. H., 234, *238*, 247, 248, *255*
Kartha, G., 20, 39, 40, 42, 51, *53*, 79, *80*, 93, *101*, 115, *118*, 129, *134*, 138, *144*, 206, *238*, 253, *255*, 279, *287*
Kasha, M., 186, *203*
Katchalski, E., 81, 82, *91*, *92*, 150, *158*, *159*, 206, 207, 208, 209, 211, 212, 214, 215, 220, 221, 222, 225, 226, 227, 229, *238*
Katutani, Y., 262, *264*
Kawahara, F. S., 184, *203*
Kay, C. M., 183, 184, *203*
Kay, L. M., 138, *144*
Kegeles, G., 302, 303, *305*
Kekwick, R. A., 302, 303, *305*
Keller, P. J., 319, 320, *335*
Kendrew, J. C., 58, 60, *67*, 79, *80*, 81, *91*, *92*, 181, 182, *203*, 250, *255*
King, M. V., 13, *22*
Kingma, S., 166, *174*
Kirkwood, J. G., 82, *91*, 216, *238*
Kitaoka, H., 138, *144*
Klug, A., 95, *100*
Kobrle, V., 352, 354, *355*, 356
Konigsberg, W., 184, *202*, 315, 316, *317*
Konishi, E., 250, *256*
Kounina, D. V., 279, *287*
Kovacs, J., 361, *366*
Kovacs, K., 361, *366*
Krebs, H. A., 172, *174*
Kroner, D., 138, *144*
Kühn, J., 286, *287*, 294
Kühn, K., 280, 281, 283, 286, *287*
Kuhn, R., 358, *366*
Kumar, K. S. V. S., 321, 322, 327, 328, 329, 331, 332, *335*
Kunitz, M., 312, *317*
Kurtz, J., 81, 82, *91*, *92*, 206, 207, 225, 229, *238*

L

Lakshmanan, B. R., 42, 43, 50, *53*
Lang, A. R., 93, 95, 96, 97, *100*
Langridge, R., *37*
Laskowski, M., Jr., 179, 187, 188, 189, 190, 191, *202*
Leach, A. A., 344, *348*
Leahy, J., 149, *158*
Lehmann, H., 166, *175*, 302, 303, *305*
Lengyel, P., 147, 148, 149, 150, 151, 152, 157, *158*, *159*, 165, *175*

Lennox, E. S., 168, *174*
Lenormant, H., 262, *263*
Leonard, W. J., Jr., 190, *203*
Léonis, J., 184, *203*
Lerner, E. M., 171, *174*
Leung, Y. C., 79, *80*, 87, *92*
Levy, A. L., 360, 363, *366*
Lewis, M. S., 51, *53*
Li, C. H., 184, *203*
Lindley, H., 91, *92*
Lindskog, S., *203*
Lindsley, J., 344, *348*
Lipscomb, W. N., 129, *134*
Littau, V. C., 172, 173, *174*
Liu, T.-Y., 313, 314, *317*
Lively, E. P., 189, *203*
Lloyd, P. F., 115, 116, *118*
Loewenstein, A., 213, *237*
Logan, M. A., 138, *144*, 344, *348*
Logue, A. D., 194, 195, *203*
Low, B., 98, *101*

M

MacArthur, I., 93, *101*
McDonald, M. R., 312, *317*
McDonald, T. R. R., 129, 132, *134*
McGandy, E. L., 58, *66*
McGarr, J., 138, *144*
McGavin, S., 49, 51, *53*, 82, *91*, 137, 138, *143*, 206, 207, 209, *237*
McGeachin, H. Mc.D., 129, *134*
McGregor, D. D., 170, *174*
McMaster, P. R. B., 171, *174*
MacRae, T. P., 100, *100*, 103, 109, 113, 115, *118*, 137, *143*
Magdoff, B. S., 13, *22*
Malmström, Bo., G. *203*
Mandelkern, L., 209, *238*
Marghoshes, M., 46, *53*
Mark, H., 103, *118*
Marsh, R. E., 65, *67*, 79, *80*, 87, *92*, 138, *144*
Marshall, R. D., 358, *366*
Martin, G. R., 51, *53*, 151, *158*
Masuda, Y., 261, 262, *264*
Mathieson, A. McL., 79, *80*, 87, *92*
Matthaei, J. H., 148, 151, *158*, *159*
Maxwell, E. S., 157, *158*
Mazurow, V. J., 279, *287*
Medawar, P. B., 169, *174*
Meiboom, S., 213, *237*
Meister, A., 356
Melvin, E. H., 359, *367*
Mende, F., 228, *238*
Mercer, E. H., 94, *100*, 111, *118*
Meyer, K., 103, *118*, 357, 358, *366*
Meyer, K. H., 103, *118*, 133, 134, *134*
Michon, J., 358, *366*
Miller, R. S., 148, 149, 150, 151, 152, 157, *158*, *159*, 165, *175*
Millionova, M. I., 138, *143*
Mirsky, A. E., 172, 173, *174*
Misra, D. N., 302, *305*
Mitcheson, N. A., 169, *174*
Miyazawa, T., 257, 258, 259, 260, 261, 262, *264*
Mizushima, S., 258, 261, *264*
Moffitt, W., 201, *203*, 242, *255*
Moore, S., 173, *174*, 310, 311, 312, 313, 314, 315, 316, *317*, 322, *335*
Morawiecki, A., 220, 221, 222, *238*
Morgan, W. T. J., 359, *366*
Morner, C. Th., 343, *348*
Moroskin, A., 141, *144*
Morrison, L. R., 171, *175*
Moss, J. A., 359, *366*
Muirhead, H., 62, *67*, 304, *305*
Muller, C. J. 166, *174*
Murayama, M., 166, *174*
Murphy, W. H., 359, 360, 361, 364, 365, *366*

N

Nagai, Y., 69, 70, *80*
Nagakura, S., 187, *203*
Nakamoto, K., 46, *53*
Narita, K., 262, *264*, 363, *366*
Naylor, A., 103, 115, *118*
Nayudamma, Y., 357, *367*
Neil, A. L., 168, *174*
Nelson, C. A., 179, 193, 194, *203*
Neuberger, A., 183, 184, 185, *202*, 358, *366*
Neuman, R. E., 344, *348*
Neumann, N. P., 313, 314, *317*
Neurath, H., 319, 320, 321, 322, 323, 324, 327, 328, 329, 331, 332, *335*
Neville, A. C., 339, *340*
Nirenberg, M. W., 148, 151, 153, *158*, *159*, 165, *174*
Nishihara, T., 279, *287*
Noda, H., 69, 70, *80*

Norland, K. S., 234, *238*, 247, 248, *255*
North, A. C. T., 40, 49, *53*, 62, 64, *67*, 206, *237*
Nossal, G. J. V., 168, *175*
Nuenke, B. G., 358, *366*
Nuenke, R. B., 358, *366*
Nystrom, R. F., 359, *366*

O

O'Brien, J. R. P., 302, 303, *305*
O'Brien, P. J., 358, *367*
Ochoa, S., 147, 148, 149, 151, 152, 156, 157, *158*, *159*
Ogawa, Y., 70, *80*
Ogle, J. D., 138, *144*
Ohno, K., 363, *366*
Omote, Y., 319, *335*
Orekhovich, K. D., 281, *287*
Orekhovich, V. N., 51, *53*, 279, 281, *287*
Ortiz, P. J., 156, *158*, *159*
Osmundsen, J., 162, *175*

P

Pardee, A. B., 172, *175*
Pauling, L., 81, 87, *91*, *92*, 96, 98, *100*, *101*, 104, 113, 114, *118*, 122, 127, *134*, 137, 138, *144*, 167, *175*, 235, *238*, 242, *255*
Perutz, M. F., 62, *67*, 79, *80*, 175, 304, *305*
Peterkofsky, B., 355, *355*
Peterson, D. L., 245, *255*
Peterson, E. A., 312, *317*
Petruska, J. A., 293, 297, *300*
Phillips, D. C., 58, 59, 64, *66*, *67*, 81, *91*, 250, *255*
Pickens, L., 348, *348*
Piez, K. A., 51, *53*, 347, *348*
Pinkerton, M., 65, *67*
Plotnikova, N. J., 281, *287*
Poljak, R. J., 64, *67*
Pollock, J. R. A., 228, 229, *237*
Popenoe, E. A., 358, *367*
Porath, J., 311, *317*
Price, W. C., 258, *263*
Prockop, D. J., 355, *355*
Pullin, A. D. E., 258, *263*

R

Ramachandran, G. N., 39, 40, 42, 43, 44, 46, 48, 50, 51, 53, *53*, *54*, 79, *80*, 93, 94, *101*, 115, *118*, 127, *134*, 138, 143, *144*, 206, *238*, 253, *255*, 279, *287*, *300*
Ramakrishnan, C., 42, 43, 50, *53*, 122, 124, 125, *134*
Randall, A. A., 208, *237*
Randall, J. T., 40, *53*
Rands, D. G., 183, 184, *203*
Rapport, M. M., 103, *118*, 357, *366*
Reddy, J. M., 65, *67*
Rees, M. W., 364 365, *366*
Reeves, R. E., 130, *134*
Ressler, C., 309, *317*
Rich, A., 42, 49, 50, 51, *53*, *54*, 115, *118*, 127, *134*, 138, 143, *144*, 206, 218, *237*, *238*, 253, *255*, 279, *287*
Richard, F. M., 194, 195, *203*
Richards, R. E., 258, *264*
Rickli, E. E., 184, 195, *203*
Riddiford, L. M., 184
Riedel, A., 295, *300*
Rimington, C., 359, *367*
Riseman, J., 216, *238*
Roberts, J. C., 168, *175*
Roberts, R. B., 157, *159*
Robertson, A. V., 223, 225, *238*
Robertson, J. H., 132, *134*
Robertson, W. von B., 352, *355*, 356
Robinson, C., *37*
Rogers, G. E., 100, *100*
Rogers, S., 173, *175*
Rollett, J. S., 60, *67*
Rose, S. M., 173, *175*
Roseman, S., 358, *366*
Rosenberg, A., 358, *367*
Rosenheck, K., 180, 198, 200, *203*, 245, *255*
Rosevear, J. W., 358, *367*
Rossmann, M. G., 15, 18, *22*, 62, 66, *67*, 304, *305*
Rudall, K. M., 103, 110, 111, *118*
Rudloff, V., 184, *202*, 315, *317*
Rundle, R. E., 46, *53*
Rupley, J. A., 324, *335*
Ryder, M. L., 99, *101*

S

Sadhukhan, P., 302, *305*
Sager, R., 158, *159*

Sakakibara, S., 69, 70, *80*
Sanger, F., 309, 314, *317*, 363, *367*
Santhanam, M. S., 53, *53*
Sasada, Y., 70, 73, *80*
Sasisekharan, V., 42, 43, 44, 46, 48, 50, 51, *53*, *54*, 82, 83, 86, *92*, 94, *101*, 122, 125, 126, 127, *134*, 143, *144*, 206, 207, *238*, 253, *255*, 279, *287*, *300*
Saxl, H., 116, *118*
Scanu, A., 295, *299*
Schechter, A. N., 157, *159*
Schellman, J. A., 247, *256*
Scheraga, H. A., 187, 188, 196, *202*, 209, *238*
Schleyer, M., 51, *53*, 138, *143*
Schmid, K., 358, *366*
Schmier, I., 201, *202*, 247, 251, 252, *255*
Schmitt, F. O., 50, *53*, 280, *287*, 289, 290, 291, 292, 293, 294, 295, 296, 298, *299*, *300*
Schram, E., 313, *317*
Schroeder, W. A., 138, *144*
Schrohenloyer, R. E., 138, *144*
Schwartz, H. C., 166, *174*
Schweet, R., 149, *158*
Schwert, G. W., 323, *335*
Scott, T. A., 359, *367*
Seaman, G. V. F., 358, 362, *366*
Seeds, W. E., *37*
Seifter, S., 138, *144*
Sela, M., 82, *91*, *92*, 150, *159*, 206, 207, 208, 209, 211, 212, 214, 215, 219, 221, 226, 227, *237*, *238*
Shmueli, U., 207, 210, *238*
Shibnev, V. A., 138, *143*
Shimanouchi, T., 258, 261, *264*
Shimizu, F., 358, *366*
Shore, V. C., 58, *67*, 81, *91*, 250, *255*
Shore, W. S., 183, 184, *203*
Shpikiter, V. O., 51, *53*, 279, *287*
Shugar, D., 184, *203*
Shukla, P. D., 21, *21*
Sikorski, J., 113, *118*
Simha, R., 209, 216, *238*
Simmonds, D. H., 361, *366*
Simmons, N. S., 243, 245, 246, 247, 251, 252, *255*
Simmons, N. W., 201, *202*
Simpson, W. T., 245, *255*, *256*
Sinex, F. M., 352, *355*
Singer, K., 302, *305*
Singer, L., 302, *305*
Singer, M. F., 148, *159*
Sivaraman, C., 346, 347, *348*
Sjöquist, J., 314, *317*
Smillie, L. B., 183, 184, *203*
Smith, E. L., 179, 180, 184, 197, *202*, 245, *255*, 358, *367*
Smyth, D. G., 310, 313, 314, 315, 316, *317*
Sober, H. A., 312, *317*
Solomons, C. C., 343, 344, 345, 346, *348*
Spackman, D. H., 310, 313, 316, *317*, 322, *335*
Sparks, R. A., 60, *67*
Spencer, M., *37*
Speyer, J. F., 147, 148, 149, 150, 151, 152, 157, *158*, *159*, 165, *175*
Stanbury, J. B., *175*
Stanford, R. H., 65, *67*
Stapleton, J. W., 137, *143*
Stark, G. R., 313, 314, *317*
Stein, W. H., 173, *174*, 310, 311, 312, 313, 314, 315, 316, *317*, 322, *335*
Steinberg, I. Z., 82, *92*, 206, 207, 208, 209, 211, 212, 214, 215, 221, 226, 227, *238*
Steinhauf, L. K., 65, *67*
Stern, F., 132, *134*
Stetten, M., 352, *355*
Stokes, A. R., *37*
Stone, N., 356
Strandberg, B. E., 58, *67*, 250, *255*
Street, A., 104, 113, *118*
Stretton, A. O. W., 166, *174*
Stromberg, R. R., 298, *300*
Stryer, L., 243, 244, 249, 251, *255*
Sutherland, G. B. B. M., 257, 259, 261, *263*, *264*
Suzuki, S., 187, *202*
Svennerholm, L., 359, 361, *367*
Swanbeck, G., 99, *101*
Swanson, S. A., 183, 184, *203*
Swenson, R. T., 166, *175*
Swerdlow, M., 298, *300*
Szent-Gyorgi, A. G., 243, 245, 246, 247, *255*

T

Tabroff, W., 138, *144*
Talalay, P., 184, *203*
Tanaka, J., 179, 180, 198, *203*, 262, *263*
Tanaka, K., 70, *80*
Tanford, C., 183, 184, 189, *203*
Tanner, K. N., 257, 259, 261, *263*

Thabaraj, G. J., 357, *367*
Thathachari, Y. T., 42, 43, 44, 46, 48, 50, *53*, *54*, 143, *144*
Thompson, E. O. P., 363, *367*
Thompson, H. W., 258, *264*
Tinoco, I., Jr., 245, *256*
Tomlin, S. G., 297, *300*
Tower, R. W., 343, 344, 346, *348*
Traub, W., 207, 210, 231, *238*
Trippett, S., 309, *317*
Troll, W., 344, *348*
Trotter, I. F., 104, *118*
Trueblood, K. N., 87, *91*
Tsuboi, M., 250, *256*
Tunis, M., 358, 362, *367*
Tuppy, H., 331, *335*
Turner, R. L., 115, *118*
Tustanovsky, A. A., 281, *287*

U

Udenfriend, S., 355, *355*
Urnes, P., 242, 243, 250, *256*

V

Vallee, B. L., 319, 321, 324, 325, 326, 327, 329, *335*
Vand, V., 86, *91*, 95, *100*
Van Slyke, D. D., 352, *355*
Veis, A., 280, *287*
Vithayathil, P. J., 194, *203*
von Hippel, P. H., 280, *287*, 292, 296, *300*

W

Wada, A., 250, *256*, 259, 262, *263*, *264*
Wagner, M. L., 184, *203*
Wahba, A. J., 148, 149, 150, 151, 152, 157, *158*, *159*, 165, *175*
Waksman, B. H., 171, *175*
Wallwork, S. C., 49, *54*
Walsh, K. A., 320, 321, 322, 327, 328, 329, 331, 332, *335*
Wang, S.-R., 184, *203*
Warner, R. C., 184, *203*
Watson, H. C., 58, *67*, 81, *91*, *92*
Watson, H. W., 250, *255*
Watson, J. D., *37*, 93, *101*
Watson, M., 172, *175*
Waugh, D. F., 96, *101*
Weaver, E. S., 280, *287*
Weber, I., 184, *203*
Weinfeld, H., 358, 362, *367*
Weinstein, I. B., 157, 158, *159*
Weis-Fogh, T., 337, 338, 339, *340*
Weissblum, B., 157, *159*
Welsh, H. K., 87, *92*
Wetlaufer, D. B., 181, 184, 186, 187, 188, *203*, 243, 245, 246, 247, *255*
White, F. H., Jr., 286, *287*, 310, 314, *317*
White, J. C., 302, *305*
Wilkins, M. H. F., *37*, 172, *175*
Will, G., 304, *305*
Williams, E. J., 185, *203*
Williams, G., 354, *355*
Wilson, A. C., 172, *175*
Wilson, H. R., *37*
Winter, H., 343, 344, 345, 346, *348*
Wintrobe, M. W., 301, *305*
Wishnia, A., 184, *203*
Witkop, B., 223, 225, *238*
Witter, A., 331, *335*
Wittmann-Liebold, B., 184, *202*
Wolf, G., 352, *355*
Wolman, Y., 138, *143*
Wood, D. L., 257, 259, 261, *263*
Wormell, R. L., 104, 111, *118*
Wunderlich, J. A., 129, *134*
Wyckoff, H. W., 95, *100*
Wyckoff, R. W. G., *277*
Wyngaarden, J. B., *175*

Y

Yamasaki, M., 320, *335*
Yanari, S., 185, 186, *203*
Yang, J. T., 242, *255*, 259, *263*
Yaron, A., 216, 218, *238*
Yoshida, T., 172, *174*

Z

Zilliken, F., 358, *367*
Zimmer, E., 283, *287*
Zimmerman, S. B., 247, *256*
Zubay, G., 172, *175*
Zworykin, V. K., *277*