Aspects of Protein Structure

Edited by G. N. RAMACHANDRAN

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

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

Library of Congress Catalog Card Number: 58–4033

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

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

SECTION IV—Electron Microscopic Studies

SECTION V—Chemical Studies

SECTION VI—General Discussion

PRESIDENTIAL ADDRESS

X-Ray Analysis of Biological Molecules

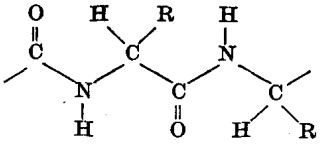
W. L. BRAGG

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

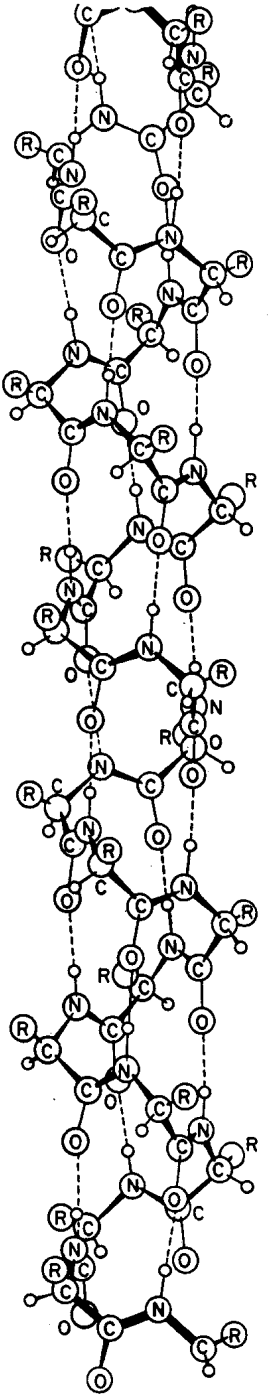


FIG. 1. The α-helix of Pauling.

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

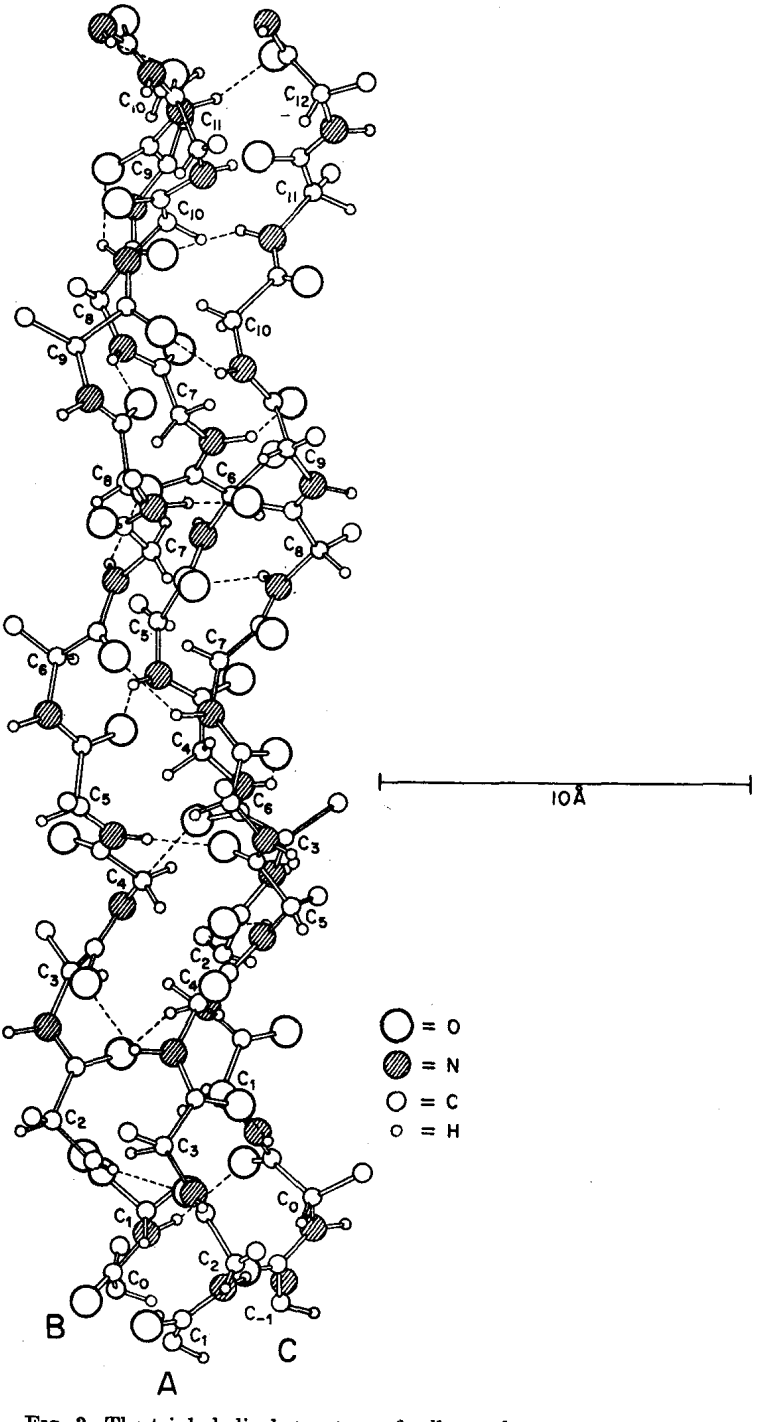


FIG. 3. The triple helical structure of collagen due to Ramachandran.

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

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.

[redacted] P17-13-B2

SECTION I
X-Ray Diffraction Studies