FEBS

Enzymes:
Structure and Function

Volume 29

Organized by:
J. DRENTH
R. A. OOSTERBAAN
C. VEEGER

FEBPBY

FEDERATION OF EUROPEAN BIOCHEMICAL SOCIETIES
EIGHTH MEETING, AMSTERDAM, 1972

ENZYMES: structure and function

Volume 29

Organized by

J. DRENTH, *Groningen*
R. A. OOSTERBAAN, *Rijswijk*
C. VEEGER, *Wageningen*

NORTH-HOLLAND / AMERICAN ELSEVIER

Library of Congress Catalog Card Number: 72-9460 1
ISBN North-Holland
Series: 0 7204 4300 8
Volume: 0 7204 4329 6
ISBN American Elsevier: 0 444 10424 0

Publishers:

North-Holland Publishing Company - Amsterdam - London

Sole distributors for the USA and Canada:

American Elsevier Publishing Company, Inc.
52 Vanderbilt Avenue
New York, N.Y. 10017

Printed in The Netherlands

INTRODUCTION

The aim of this symposium was to provide a review of our present knowledge of structure and function of enzymes to the participants of the Eighth FEBS Meeting.

It was the intention of the organizers to invite mainly speakers from countries participating in FEBS, but to have, in addition, scientists from other countries for special topics. In our preliminary discussions the question arose as to whether the invited speakers had to deal with specific enzymes or with specific developments. We chose the latter and asked the speakers to emphasize new technical developments and ideas rather than present mainly new data. The speakers were allowed to include new data, of course.

Thus this edition contains, in addition to contributions concerning the three-dimensional structure of proteins and the interaction of proteins with water, papers dealing with the basic problems of enzyme catalysis, mobility of bound substrates, cooperative interactions, mechanism of enzyme catalysis and multi-enzyme complexes. Furthermore we considered it very useful to discuss the problem of transfer of information obtained in X-ray crystallographic studies to general enzymologists.

The organizers were handicapped to some extent by the simultaneously organized symposium on Analysis and simulation of biochemical systems; a certain overlap of interests could not be avoided.

The final edition of the book contains fewer contributions than originally intended, because some of the invited speakers were unable to submit a manuscript of their contribution on time. The organizers feel that this handicap is less important than the advantage of rapid publication of these proceedings which, in our opinion, give a good review of the present state of knowledge in enzymology.

Drenth
Oosterbaan
Veeger

LIST OF CONTENTS

THE COMPARISON OF PROTEIN STRUCTURES IN THE CRYSTALLINE STATE AND IN SOLUTION

J. Drenth
Laboratorium voor Structuurchemie, Rijksuniversiteit Groningen,
The Netherlands

When in 1912 Laue suggested the study of diffraction of X-rays by a crystal, the application to biological macromolecules was not of immediate concern. Biochemistry was still in its infancy. X-ray work on proteins with fibrous material started around 1938 with Astbury's pioneering studies. Bernal took the first X-ray diffraction picture of a globular protein, pepsin, in 1934. The processing of the data provided by the pictures had to pass the high barrier called phase problem before crystal structures of proteins could be produced. I shall not trouble you with an explanation of this phase problem. It has been solved by the application of isomorphous heavy atom derivatives of the protein, where isomorphous means that attachment of the heavy atoms does not change the protein structure itself. So X-ray diffraction is a tool to obtain the 3-dimensional structure of a protein molecule, provided it can be obtained in a crystalline form, because only for molecules in a regular crystalline array is the pattern of the diffracted beams both strong enough and easily interpretable; the fuzzy diffraction picture of amorphous material is uninterpretable.

If the two prerequisites: first of all good quality crystals -not all crystals are suitable- and secondly isomorphous heavy atom derivatives are fulfilled, the way is open to determine the crystal structure of a protein provided its size is not too large, let us say with a molecular weight of not more than 300,000. Modern electronic equipment for collecting and processing the enormous amount of data is indispensable. So far, the detailed crystal structures of at least 30 proteins have been reported. Having determined such a crystal structure, one is always faced with the question: Is the crystal structure relevant for the protein in its physiological medium, normally an aqueous solution? It is for this reason that some biochemists look with suspicion on a protein crystal structure. Of course they had reasons to be suspicious when only a few protein structures had been determined, but now a wealth of information supports the view that a protein molecule has essentially the same structure in the crystal as in solution.

Let us first of all review very briefly the solution methods.
It is impossible to determine the detailed structure of a complicated molecule in solution. From low angle X-ray scattering just the dimensions and the shape of the molecule can be derived. The method should be applied to well-defined molecular species. Unexpected differences in the low angle scattering of α-lactalbumin and lysozyme could be explained by the presence of an α-lactalbumin dimer (Achter and Swan, 1971). Low angle X-ray scattering results for myoglobin are in close agreement with the crystallographic structure (Beeman, 1967).
A wide variety of methods is available for pin-pointing structural details in a large molecule. Chemical experiments are often directed to the accessibility of functional groups in proteins. An enormous amount of titration data on internal and external tyrosines is available which generally agree with the positions of these residues in the crystal structure. However, one has to be careful with the chemical approach; e.g. in the nitration of tyrosine residues with tetranitromethane, residues in a hydrophobic environment may be preferentially modified (Myers and Glazer, 1971). Horse heart cytochrome c contains four tyrosine residues, two of

which, 67 and 48, are buried and the other two, 74 and 97, have their hydroxyl group exposed; but it is the buried Tyr 67 that is highly reactive towards tetranitromethane. The same applies to iodination and acetylation. Therefore availability of a side chain for a reaction does not immediately indicate its actual location: whether at the surface or buried in the protein molecule. The reaction of papain with 1,3-dibromoacetone is an example of a chemical reaction in which the distance between a cysteine and a histidine residue could be determined as 5 Å (Husain and Lowe, 1968). The topography of papain in solution was further studied by means of photo-oxydation (Jori et al., 1971). The essential cysteine was first caused to react with a photo-sensitizer group. Subsequently in a photo-oxydation process the protein molecule was attacked in the neighbourhood of the sensitizing group. The first residues to be attacked are Trp 177 and His 159. Their oxydation does not lead to a disruption of the papain molecule. When however on further oxydation Trp 26 was also modified, the papain structure broke down completely. The conclusion must be that Trp 177 and His 159 are at the surface and do not contribute to the stability of the papain molecular structure, but that Trp 26 does. This is also very obvious from the structure of this protein as determined by X-ray diffraction (Drenth et al., 1968). In a great many experiments optical and magnetic spectroscopy have been applied to detect details of the structure, structural changes and the extent of flexibility. The estimation of secondary structure from CD and ORD measurements is never in great conflict with the X-ray data. The U.V. spectral shift by solvent perturbation is dependent on the degree of exposure of absorbing chromophores in the protein (Kronman and Robbins, 1970).

The fluorescence of tryptophan and tyrosine has been studied extensively. Fluorescent coenzymes have given important information on structural features in a number of proteins. It is also possible to attach a fluorescent probe to the protein, e.g. a fluorescent substrate to the active site of an enzyme, and to follow the fluorescence as catalysis proceeds. This was applied to carboxypeptidase A (Lasser and Feitelson, 1971) and specific interactions of the substrate with the active site could be indicated in addition to, but not in conflict with the X-ray data on substrate binding to carboxypeptidase (Quiocho and Lipscomb, 1971). Similar satisfying results are obtained from NMR data on the binding of N-formyl-L-tryptophan as inhibitor to chymotrypsin (Gerig and Rimerman, 1970).

Besides chemical and physical techniques, measurements of enzymatic activity of crystalline enzymes and immunological experiments have been used for checking whether there are differences between the crystal structure as determined by X-ray diffraction and the structure in solution.

If we attempted to review all available information on the structure of proteins in solution, we would drown in the abundance of data. Let us instead turn to crystallography because interestingly enough the strongest support for the assumption that the crystal structure is a true representation of the native protein structure comes from crystallography itself. The information is so abundant that even the most suspicious biochemist can trust the crystallographic results.

The polypeptide chain in the hemoglobin family of proteins is folded around the heme group in such a way that its iron atom can react reversibly with oxygen. The folding is very similar, not only for the α and β chain in the various hemoglobins and myoglobins but also for evolutionarily more remote heme proteins. Representatives of this last category are lamprey hemoglobin (Hendrickson and Love, 1971), the hemoglobin from an annelid worm (Love et al., 1972) and erythrocruorin from an insect (Huber et al., 1971). Lamprey hemoglobin is monomeric and has 26% degree of homology with sperm whale myoglobin and an extra tail of 10 residues at the N-terminus. The

heme group is oriented and wrapped in the surrounding polypeptide chain in a similar way to the other members of this family of heme proteins. Rather major configurational changes, induced by deletions and insertions are only found at points not affecting the heme. The same is true for the worm hemoglobin. The monomeric insect hemoglobin erythrocruorin has only 20% identical residues at equivalent positions when compared with sperm whale myoglobin. Superimposing this molecule with myoglobin and comparing equivalent atomic positions showed an average separation of 1.25 Å. For the atoms in helices the mean distance was only 0.5 to 0.6 Å. All deletions occur in parts of the molecule away from the heme group. The observed differences in the atomic positions can be easily explained by small shifts in the molecule as a result of the deletions and residue substitutions. The heme group is slightly rotated but the general pattern of apolar heme contacts is remarkably constant.

Thus the whale and the worm, the human and the horse, the insect and the lamprey have very similar crystal structures of their hemoglobins with differences easily explained by the amino acid sequence. If this structure does not basically represent the native structure then all these hemoglobin crystal structures must differ from it, which is extremely unlikely.

Let us now consider the cytochrome-c family of heme proteins. Its members also have considerable differences in amino acid sequence but are nevertheless functionally identical, which does indicate the same folding of the polypeptide chain for all members. The horse heart cytochrome-c molecule has a great deal of extended chain, wrapped around the heme. A striking invariancy of residues has been noticed for three locations in the molecular structure: in a basic and in an acid region and for the buried hydrophobic residues, although individual residues may differ (Dickerson et al., 1971). Thus X-ray analysis explains the homology in the amino acid sequences: acidic and basic regions on the molecular surface are present in all members even though the individual residues change.

Structural similarity within families of proteins in not confined to the heme proteins. It is also true for enzymes. In the tryptic family of proteolytic serine enzymes, X-ray analysis showed for trypsin (Stroud et al., 1972), chymotrypsin (Blow, 1971) and elastase (Hartley and Shotton, 1971) not only the same conformation of their main chains but also a very similar construction for their active sites. The precursor chymotrypsinogen can also be added to this list (Kraut, 1971). Another member is the α-lytic protease of a bacteria. Judged by the amino acid sequence, this enzyme has hardly any relationship to its relatives. Nevertheless a molecular model of α-lytic protease could be built with a structural core which is almost the same as that in the other tryptic enzymes (McLachlan and Shotton, 1971). The differences in conformation are restricted to the surface where entire loops are missing or new ones added. This model building procedure will soon be checked by the result of X-ray crystallographic studies of α-lytic protease, which are in progress (James).

A proteolytic serine enzyme that has much in common with the tryptic family of enzymes is subtilisin. Its molecule consists of 275 residues in a single chain which contains some pieces of α-helix and β-structure but is otherwise folded rather randomly. The spherical molecule has a diameter of 42 Å with a distorted pleated sheet in its centre, consisting of six stretches of chain. This sheet has, on both sides, hydrophobic side chains in contact with hydrophobic groups of two helices on one side and two helices on the other side. The structure of this protein has been determined completely independently in two laboratories: by Kraut and his colleagues in California (Schubert Wright et al., 1969) and in our laboratory (Drenth et al., 1972). Kraut grew his crystals in concentrated ammonium sulfate at pH 5.9 and we used completely different crystals which were grown at pH 9.1 from

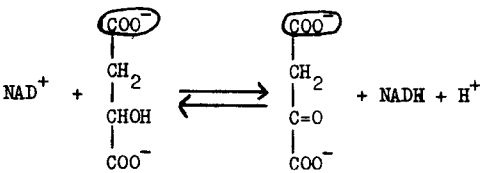
55% acetone. The important differences in the medium, in the pH and in the crystal form did not really affect the structure. The mean difference in position for equivalent atoms was 1.55 A. Even the high mobility of a stretch of five residues was present in both types of crystal. A very similar situation exists for ribonuclease which was crystallized from a 55% 2-methyl-2,4-pentanediol solution. Its crystal structure (Kartha et al., 1967) is identical in all respects to the structure of ribonuclease-S (Wyckoff et al., 1970) except for that part of the chain where it has been cut. And ribonuclease-S was crystallized from a completely different medium, a concentrated salt solution in water, in a quite different crystal form. The conclusion must be that the molecular structures of subtilisin and ribonuclease are sensitive neither to the packing forces in the crystal structure nor to the change in the medium.

It must of course be realized that not all proteins are as stable as subtilisin and ribonuclease. In principle it should be possible that a protein, when crystallized from a solution which is quite different from its natural medium, may assume a non-native structure. This, however, could only happen if the modified structure has one single conformation. Otherwise it would not crystallize or at least would not give a high quality X-ray diffraction pattern.

Returning to the tryptic enzymes we should mention that the structure of the pancreatic trypsin inhibitor has been determined crystallographically (Huber et al., 1972). It is a pear-shaped molecule with a length of 29 Å and a maximum diameter of 19 Å. We shall not go into the details of its structure. When a molecular model of this inhibitor is fitted to the trypsin model the two molecules match each other so beautifully in their contact area that any scepticism as to the reliability of the enzyme structure should vanish.

Actual structure determinations on the inhibitor complexes with proteinases are now being undertaken in Munich and in Cambridge. X-ray results on dehydrogenases are just as exciting as in other families of proteins.

Heart muscle malate dehydrogenase catalyses the following reaction:



The molecule has two identical subunits. The structure of the enzyme with one NAD^+ group bound per molecular dimer was determined, so only one of the monomers has the NAD^+ (Tsernoglou et al., 1972). When we cut off COO^- and replace it by H we have another important physiological reaction: the oxydation of lactate to pyruvate and this is catalyzed by LDH. LDH is a tetrameric enzyme. Its amino acid sequence is not yet completely determined but the X-ray structure of the apoenzyme as well as the structure of the binary complex with the coenzyme NAD^+ and of several ternary complexes have been worked out. As a result the conformation of the polypeptide chain and the binding of the coenzyme are known (Rossmann et al., 1972). When the LDH and MDH structures were compared it turned out that they resemble each other to a very great extent.

The structure of two small flavoproteins, both flavodoxins, has been reported (Ludwig et al., 1972; Jensen, 1972) and it is tentatively suggested that they could have structural similarities to the dehydrogenase but this must await a closer comparison.

The conformations of the polypeptide chain in lysozyme and α-lactalbumin are probably very similar although the crystallographic evidence is not yet as convincing as is desired (Browne et al., 1969).
There is one last structural similarity I would like to mention. It concerns the non heme iron proteins which contain, besides iron, highly labile sulfur as an integral constituent of the catalytically active site. One member of this family if called HiPIP (high potential iron sulfur protein). It has an 86 residue polypeptide chain with four cysteines. These four cysteines are part of a tetrahedral cluster in the centre of the molecule (Carter et al., 1972). They are connected via their sulfur atoms to the four iron atoms of the cluster, which also contains the four labile sulfur atoms. This arrangement of four Fe atoms + four labile, inorganic S atoms + four cysteine S atoms is of considerable interest and also occurs in the ferredoxin structure which has been determined recently (Sieker et al., 1971).

I have passed quickly along a great number of protein structures stressing in particular the structural features that are the same for the members of a family and the independence of the structures from the crystal forces and from the medium from which they were crystallized. Our conclusion is that crystal structures of proteins are native structures and nobody should worry any longer.
When working on a protein problem, the biochemist should always familiarize himself with the crystal structure of his particular protein, when it is known. If it has not yet been determined he should try to crystallize the protein and get a crystallographer interested in his problem. This is essential in reaching the ultimate answer to the question: how does the protein exert its biological function? This is far from saying that X-ray crystallography alone can provide the answer to this question. On the contrary. We are approaching the day on which we shall understand how hemoglobin functions; we know that without X-ray crystallography this problem would probably never have been solved, but we also know that without the contributions of experiments in the chemical, physical and physiological field our knowledge could never have gone so far.
I have not dealt with the protein molecules of the very big type. Technical problems increase considerably with the size of the molecules when X-ray diffraction is applied. Combination with electron microscopy is promising, as the encouraging progress with TMV shows (Butler and Klug, 1971).
Binding active enzymes covalently to inert solid supports may be of advantage to industrial processes; it also functions as a stepping - stone to membrane bound proteins. The loss of flexibility in the protein caused by the rigid structure of the carrier can be studied. The major obstacle to the discovery of the structure of real membrane proteins is their transfer from the membrane into solution and the change in structure which the extraction procedure could bring about. So far X-ray crystallography has not been able to accomplish much in the membrane field. It is however worth mentioning two structures which exert their biological function at a membrane surface.
First of all cytochrome b5, which is a heme protein from the microsomal fraction of liver tissue homogenates: its function is to transport an electron, while bound to the membrane. The crystal structure of a solubilized form of the protein, which shows full activity and which has been obtained by enzymatic cleavage from the membrane, has been determined (Mathews et al., 1972). In this process about 44 residues are lost at the N-terminal part of the chain. This tail is hydrophobic in nature and is probably buried in the phospholipid layer of the membrane, with the rest of the molecule extending into the cytoplasm and acting as the electron-transport system.

The second example is insulin which -although not to be regarded as an actual membrane protein- exerts its function at a membrane surface. Careful comparison of the X-ray structure with sequence data seems to implicate four A-chain residues in the hormone's activity (Blundell et al., 1972). Complete ignorance covers the rest of the story but a close cooperation between biochemistry and X-ray crystallography will undoubtedly unravel the nature of the insulin action.

I hope to have convinced at least some of you that X-ray structures can be accepted as truly representing the structure of the native protein.

REFERENCES

Achter, E.K. and Swan, I.D.A. (1971). Biochemistry 10, 2976.
Beeman, W.W. (1967). In: Small angle X-ray scattering, ed. H. Breemberger (Gordon and Breach, New York) p.197.
Blow, D.M. (1971). In: The enzymes, vol. III., ed. P.D. Boyer (Academic Press, New York and London) p.185.
Blundell, T.L., Cutfield, J.F., Dodson, E.J., Dodson, G.G., Hodgkin, D.C. and Mercola, D.A. (1972). Cold Spring Harbor Symp.Quant.Biol. 36, 233.
Browne, W.J., North, A.C.T., Phillips, D.C., Brew, K., Vanaman, T.C. and Hill, R.L. (1969). J.Mol.Biol. 42, 65.
Butler, P.J.G. and Klug, A. (1971). Nature, New Biol. 229, 47.
Carter,Jr., C.W., Freer, S.T., Xuong, Ng.H., Alden, R.A. and Kraut, J. (1972). Cold Spring Harbor Symp.Quant.Biol. 36, 381.
Dickerson, R.E., Takano, T., Eisenberg, D., Kallai, O.B., Samson, L., Cooper, A. and Margoliash, E. (1971). J.Biol.Chem. 246, 1511.
Drenth, J., Hol, W.G.J., Jansonius, J.N. and Koekoek, R. (1972). Eur.J. Biochem. 26, 177.
Drenth, J., Jansonius, J.N., Koekoek, R., Swen, H.M. and Wolthers, B.G. (1968). Nature 218, 929.
Gerig, J.T. and Rimerman, R.A. (1970). Biochem.Biophys.Res.Comm. 40, 1149.
Hartley, B.S. and Shotton, D.M. (1971). In: The enzymes, vol. III., ed. P.D. Boyer (Academic Press, New York and London) p.323.
Hendrickson, W.A. and Love, W.E. (1971). Nature, New Biol. 232, 197.
Huber, R., Epp, O., Steigemann, W. and Formanek, H. (1971). Eur.J.Biochem. 19, 42.
Huber, R., Kukla, D., Ruhlmann, A. and Steigemann, W. (1972). Cold Spring Harbor Symp.Quant.Biol. 36, 141.
Husain, S.S. and Lowe, G. (1968). Biochem.J. 108, 855.
James, M., to be published.
Jensen, L.H. (1972). Communication at Protein Crystallography Workshop, Alpbach, March 1972.
Jori, G., Gennari, G., Toniolo, C. and Scoffone, E. (1971). J.Mol.Biol. 59, 151.
Kartha, G., Bello, J. and Harker, D. (1967). Nature 213, 862.
Kraut, J. (1971). In: The enzymes, vol. III., ed. P.D. Boyer (Academic Press, New York and London) p.165.
Kronman, M.J. and Robbins, F.M. (1970). In: Fine structure of proteins and nucleic acids, eds. G.D. Fasman and S.N. Timasheff (Marcel Dekker, New York) p.271.
Lasser, N. and Feitelson, J. (1971). Biochemistry 10, 307.
Love, W.E., Klock, P.A., Lattman, E.E., Padlan, E.A., Ward,Jr., K.B. and Hendrickson, W.A. (1972). Cold Spring Harbor Symp.Quant.Biol. 36, 349.
Ludwig, M.L., Andersen, R.D., Apgar, P.A., Burnett, R.M., LeQuesne, M.E. and Mayhew, S.G. (1972). Cold Spring Harbor Symp.Quant.Biol. 36, 369.
Mathews, F.S., Argos, P. and Levine, M. (1972). Cold Spring Harbor Symp. Quant.Biol. 36, 387.

McLachlan, A.D. and Shotton, D.M. (1971). Nature, New Biol. 229, 202.
Myers, B. and Glazer, A.N. (1971). J.Biol.Chem. 246, 412.
Quiocho, F.A. and Lipscomb, W.N. (1971). In: Advances in protein chemistry, vol. 25, eds. C.B. Anfinsen, J.T. Edsall and F.M. Richards (Academic Press, New York and London) p.1.
Rossmann, M.G., Adams, M.J., Buehner, M., Ford, G.C., Hackert, M.L., Lentz, Jr., P.J., McPherson,Jr., A., Schevitz, R.W. and Smiley, I.E. (1972). Cold Spring Harbor Symp.Quant.Biol. 36, 179.
Schubert Wright, C., Alden, R.A. and Kraut, J. (1969). Nature 221, 235.
Sieker, L.C., Adman, E. and Jensen, L.H. (1972). Nature 235, 40.
Stroud, R.M., Kay, L.M. and Dickerson, R.E. (1972). Cold Spring Harbor Symp.Quant.Biol. 36, 125.
Tsernoglou, D., Hill, E. and Banaszak, L.J. (1972). Cold Spring Harbor Symp.Quant.Biol. 36, 171.
Wyckoff, H.W., Tsernoglou, D., Hanson, A.W., Knox, J.R., Lee, B. and Richards, F.M. (1970). J.Biol.Chem. 245, 305.

REPRESENTATION OF THE STRUCTURES OF LARGE MOLECULES

A.C.T. NORTH
Laboratory of Molecular Biophysics,
Department of Zoology,
University of Oxford, England.

present address:
Astbury Department of Biophysics,
University of Leeds, England.

It has become increasingly apparent that the various chemical reactions that take place in living systems rely to a considerable extent on a specific recognition of one molecule for another and that the specificity of interactions between the molecules depends crucially upon the spatial arrangement of the atoms from which they are made up. Thus the interactions between enzymes and their substrates, drugs or hormones and their receptor cells, antigens and their antibodies, depend upon a precise complementarity between the features involved - a complementarity of surface topography, of positive and negative charges, of polar and non-polar groups. It is essential for an understanding of these spatial relationships that the three-dimensional nature of the molecules should be appreciated. For example, it is known that the hydrolysis of the glycosidic bond in a cell wall polysaccharide by lysozyme takes place with retention of configuration at the C_1 atom; examination of an accurate molecular model shows that a water molecule can only approach the C_1 atom from the side remote from the enzyme, which limits the number of possible reaction mechanisms that could be responsible. There are examples in the literature of reaction mechanisms that look satisfactory in two dimensions on paper but which are seen to be impossible when a three-dimensional model is built.

Biochemists and biophysicists are fortunate in that, to a first approximation at least, many of the molecular properties in which they are interested can be faithfully represented in a model of the order of a hundred million times life-size. (They are more fortunate than cosmologists on the one hand and nuclear physicists on the other, for whom a linear scaling down or up respectively does not lead to a model that represents the required phenomena on a laboratory scale.)

X-ray crystallography is at present the only practicable method of obtaining complete three-dimensional information about molecules (I will not discuss here the often-expressed doubts as to whether the structures of biological macromolecules in the crystal can be considered to represent their structures in their natural environment; but I would point out that the structures that have at present been determined crystallographically have in no case to my knowledge been shown to be inconsistent with observations in solution, though crystallographers certainly anticipate that some molecules,such as immunoglobulins, may be essentially flexible so that the crystal may trap one out of a range of possible conformations). Crystallographers have long been used to working with molecular models and with drawings, but they are now faced with the problem of conveying information about the exact structures of very complicated molecules to other people less familiar with models.

In fact, crystallographers use molecular models for two quite distinct purposes. One is in illustrating their results; the other is in interpreting their observations and I will discuss this use briefly first.

MODELS IN THE INTERPRETATION OF CRYSTALLOGRAPHIC OBSERVATIONS

With crystals of small molecules, it is normally possible to obtain electron density maps in which individual atoms are revealed as more or less discrete peaks whose centres can be estimated accurately. The analysis of interatomic bond-lengths and inter-bond angles from such studies has led to a fairly precise knowledge of the lengths of covalent bonds and their angular separation, the distances of closest approach between non-bonded atoms, the lengths and directional properties of hydrogen bonds, etc., and only in rare circumstances are there significant departures from these norms. It is therefore possible to build up sets of standard model components, each of which may represent either a single atom or a group of atoms and which embody the bond lengths, inter-bond angles, etc., found from the crystallographic studies. Any molecular model built up from such components is therefore constrained to conform to the canonical set of dimensions.

The electron density maps derived from crystallographic studies of large molecules such as proteins do not normally have sufficiently good definition for individual atoms to be distinguished, though they often show the characteristic shapes due to particular chemical groupings, such as the branched side chain of leucine and the flattened disc shape of the aromatic ring of phenylalanine. It is normal therefore to interpret such maps with the aid of molecular models so as to lead to a solution for the structure that is consistent both with the observed electron density and with the canonical dimensions. The fit between map and model can be seen clearly in a device such as the optical comparator of Richards (1968) with which, by the aid of a half-silvered mirror, model and map are seen superimposed in space. The quality of the map and its interpretation may be adjudged by the extent to which the model and map overlap and the crystallographer may publish illustrations of the model in order to demonstrate the fit. Models used for this type of work are generally of the skeletal type in which the covalent bonds radiating from any atom are represented by rods of the correct length and set at the appropriate angles, but the atom itself is not represented physically, its centre being defined by the intersection of the bonds. Such a model must therefore be tested to ensure that non-bonded atoms do not come closer together than the sum of their van der Waals radii and a satisfactory outcome of such checks makes an essentially independent test of the model. Space-filling models in which each component is of an appropriate size to represent the van der Waals radius of the atom concerned ensure that the finished model does not contain any short contacts, but they are not generally useful in constructing an accurate model because the positions of the centres of the atoms cannot be seen or measured. For an accurate representation of the surface topography of a molecule or for investigating internal contacts, a valuable procedure is to build most of the model from skeletal components and then to "clothe" the surfaces of interest with space-filling components made to the same scale.

The crystallographer generally uses a model of the skeletal type in the course of interpreting his observations to give a set of atomic coordinates that are constrained to accepted values for covalent bond lengths and inter-bond angles. The coordinates of individual atoms may be read off the model and possibly subjected to a computer refinement procedure such as that described by Diamond (1966); the necessity for the latter procedure arises firstly because the model components may have become distorted in the course of building and secondly because errors may have arisen in the course of measuring the positions of individual atoms; the function of this computer program is to adjust the individual atomic positions so that they conform to the required constraints as nearly as possible while being moved the minimum distances from the positions measured from the model. A further refinement procedure may then be used (Diamond, 1971) in which the atomic positions are moved, still preserving the relative positional constraints, so as to give the best fit to the observed electron-density map by some objective criterion rather than the subjective one used by the model builder.

The crystallographer has now reached the stage of having a model, more or less accurate, and a set of coordinates of the individual atoms and he wishes to convey this information to other people. In this paper I wish to distinguish between distinct situations, lectures, publications and research.

ILLUSTRATION OF LECTURES

It is quite obvious that a list of coordinates is not an appropriate way of conveying information about the arrangement of atoms in a molecule to an audience and photographs or diagrams must be used.

Single photographs of molecular models tend to look so confused that they may be practically useless. However, with care, useful photographs may be obtained of quite limited regions of a model; attention must be paid to lighting just those features that are of interest and it is often helpful to use a high aperture, giving a small depth of focus so that the features that are not of interest are blurred. If more than a localised region of the model is to be illustrated, it is almost essential to use some method of creating the illusion of three dimensions. One method is to make a film of the model while it is rotated or oscillated back and forth. The three-dimensional impression is quite clear while the model is moving, but is only retained for a short while when the movement ceases; it is really rather difficult for a lecturer and his audience to concentrate on a specific feature while it is in continual motion. An alternative method is by the use of stereoscopic pairs of photographs taken from two slightly different angles; the photographs must then be projected so that one eye of the observer sees one image and the other eye sees the second image. Two systems of image separation are in common use. In one, polaroid filters are placed in the light paths from the two images and members of the audience wear polaroid spectacles. It is necessary to use a metallized screen so that the direction of polarisation is not lost. The requirement for a special projector, screen and spectacles is rather inconvenient, but it is possible to give satisfactory reproduction of black and white and of full colour photographs. In the second system, image separation is achieved by the use of images of different colour (red and green (or blue)) and coloured spectacles; the projector and screen may be conventional and only the special spectacles

are required. Unfortunately it is not possible to show full colour photographs with the red and green system and it is in fact quite difficult to achieve adequate colour separation even from black and white drawings.To prepare slides from drawings of black lines on a white background, the best procedure appears to be as follows: (1) take separate photographic negatives (i.e. white on black) of the left- and right-eye views, (2) doubly expose a colour film to the left-eye negative with red light and the right-eye negative with green light. It may be necessary to use different exposures for the two colours in order to achieve the best results. The slides so produced show the drawings as bright lines on a black background. (The spectacles should be used with the left eye looking through the red filter.) It is absolutely essential that the two images on the slide should be parallel and in exact vertical register. Registration between the two images is also important when the polarising system is used; although stereo projectors have facilities for horizontal and vertical adjustment of one image with respect to the other, it is excrutiatingly painful for an audience to look at a stereo slide whose images are out of vertical alignment and great care ought therefore to be taken in mounting slides so as to minimise the vertical adjustment required as well as to make sure that the two images are exactly parallel. Many audiences have received an unfavourable impression of the value of stereo slides after seeing badly mounted ones and it is confusing if stereo slides cover too great a depth of field. The correct angular separation between the images can only be achieved at one distance from the screen and, while in practice considerable tolerance in the angular separation can be accepted, stereo slides are not really suitable for large audiences and they have a further limitation in the quite significant number of people who do not have effective stereoscopic vision.

In referring above to the preparation of slides from drawings, I have foreshadowed my next point, which is that photographs of models are often extremely confused (and confusing) because of the supporting rods, clamps and screws required to hold them in place and for many purposes diagrams are much more revealing than photographs of skeletal models. I have already referred to the difficulty of making accurate space-filling models, but they are undoubtedly invaluable for showing the surface topography of a molecule and difficult to better by means of drawings. However, drawings are usually clearer than photographs of skeletal models and are more flexible in focussing attention on the features of interest.

The desirability of producing a large number of drawings to illustrate different aspects of a structure makes it worthwhile writing computer programs to prepare drawings on a digital graph plotter. With one such program, written by the author, a "dictionary" of connectivities is incorporated so that for any type of amino-acid residue, the program knows which atom is connected to which. Once the atomic coordinates of a protein molecule have been fed in, followed by a list of simple directives specifying the direction of viewing, the scale of the drawing, sizes of circles representing atoms, thickness of lines representing bonds, whether side chains are to be drawn, the numbers of the residues to be drawn, etc., the computer generates a drawing of sufficient quality for the preparation of slides or figures for papers. The one program permits considerable variety of styles (see figures 1 - 3); stereo pairs of drawings may be prepared by rotating the direction of viewing suitably or the illusion of depth may be given by varying the thickness of lines