Review Letter

A BRIEF SURVEY OF SOME RECENT WORK IN PROTEIN CHEMISTRY

Eraldo ANTONINI
Institutes of Biological Chemistry, University of Rome and Camerino and Regina Elena Institute for Cancer Research, Rome

Received 31 August 1968

1. Introduction

An attempt to review critically recent work in protein chemistry is a difficult task even if limited to the literature which has appeared within the last few months. The number of significant papers is very large and the approaches are multiform so that it is not easy to reach general conclusions. Most of the recent work on the structure of proteins is the outcome or the development of ideas and methods which, although only a few years old, are now applied in a systematic way to a variety of protein systems. This concerns X-ray diffraction, optical rotatory dispersion, circular dichroism, resonance methods, and amino acid sequence studies. For instance, the new Atlas of Protein Sequence and Structure (editors M. O. Dayhoff and R. U. Eck, National Biomedical Research Foundation, Maryland 1966) lists the total or partial amino acid sequences of more than 200 proteins.

The advancing knowledge of protein structure stimulates correlations, still mostly empirical, of structure with function. The study of protein reactions, association–dissociation phenomena, and ligand binding, benefits from improved instrumentation and rigorous thermodynamic and kinetic treatments. Problems of protein function, especially of enzymes, are now approached directly with a wide application of methods for the study of fast reactions. However, the healthy tendency to find out more facts about protein reactions, by a variety of methods, is associated with not wholly successful attempts at generalisation and with descriptions of the behaviour of the systems in terms of simple models. In detail, the mechanisms involved in protein function are still largely unknown and there seems to be no royal road to their understanding.

In what follows, a few examples of recent significant contributions to the problems of protein structure and function will be briefly discussed.

2. Structure of proteins by X-ray analysis

The number of proteins whose structures have been elucidated by X-ray analysis at high resolution is rapidly increasing.

It is difficult to say how much these structures have in common, apart from the general overall features which were revealed by the initial work on haem proteins and lysozyme. Thus, we shall be facing shortly a sort of natural history of protein structures where each protein will have, in the framework of species variability, its morphological definition. Within this new molecular anatomy general biological principles will eventually appear just as they have appeared from classical macroscopic or microscopic anatomy.

However, it is obvious that the structures themselves, plus the available knowledge of chemical reactivity as seen in simple organic compounds, cannot directly solve problems of protein function. In this respect protein crystallographers share the same difficulties as classical anatomists who felt they could understand physiology just by looking at the morphology of organs.

A gratifying aspect of recent work is the beautiful agreement that is often seen between what was deduced from indirect chemical experiments or from physical measurements in solution and what is shown

directly by the X-ray studies; this applies to the nature of the active sites of enzymes, to the critical role of amino acid residues, and to the reactivity of specific groups. All this and other specific investigations give more and more indications that the conformation of proteins is the same in solution as in the crystals (see for instance Rupley, J. Mol. Biol. 35 (1968) 455 and Rupley and Gates, J. Mol. Biol. 35 (1968) 477).

3. Proteolytic enzymes

The structure of three proteolytic enzymes, carboxypeptidase A, chymotrypsin and papain have been reported by different groups; these are the outcome of studies at high resolution of the enzymes themselves and, in two cases, of the enzyme-substrate complexes. Of special interest have been the descriptions of the structures of the active sites, the interactions of the substrates with the proteins, and the effects of substrate binding on protein conformation. Large use is made of information from amino acid sequences and from other chemical studies; on the other hand, knowledge of the structure is used to propose hypotheses about catalytic mechanisms. The main specific results with these three enzymes are outlined below.

The structure of carboxypeptidase A (Reeke et al., Proc. Natl. Acad. Sci. 58 (1967) 2220) has been obtained at 2 A resolution so that even details at atomic level show up. The molecule, 307 residues, contains about 30% α-helix in four regions: residues 14-29, 72-88, 215-233, and 288-305. Twenty percent of the chain is in the form of a pleated sheet; the α-helical regions are mostly on the outside of the molecule on one side of the pleated sheet.

The location and the interactions of the catalytically important zinc atom have been elucidated. In addition, the paper by Reeke and his coworkers contains the description of a 2.8 A difference synthesis of the complex of the enzyme with glycyl-tyrosine (the enzyme-substrate complex). The presence of substrate induces large changes in the structure of the protein; many substrate interactions have been recognized and the proposed roles of certain residues in the catalytic activity have found their rationalisations in the structure. On the basis of the structural findings a mechanism for the hydrolysis of the substrate is proposed.

A preliminary, but very informative, report on the structure of α-chymotrypsin has been published by Sigler and his coworkers (J. Mol. Biol. 35 (1968) 143). The authors emphasize that the results are not to be taken as final due to difficulties in the interpretation of the X-ray data.

The results show the general features of the molecule and the course of the polypeptide chain. The differences between the free enzyme and the tosyl derivative, an irreversibly inhibited derivative in which the active site is sulphonylated, have been determined; they consist of small changes which involve only the region near the active centre. The paper contains also a discussion of the homology between chymotrypsin and trypsin. Finally a hypothesis is suggested for the stereochemistry of the activation process which would depend on the formation of an ion pair between isoleucine 16 and aspartate 194.

The authors point out that the overall tertiary structures of the zymogen and the enzyme are very similar; this is a case where the conclusions from studies in solution, by optical methods, have not been substantiated by the X-ray data. It shows that great caution must still be used in interpreting the results of optical rotatory dispersion and circular dichroism measurements on proteins.

The other proteolytic enzyme whose structure has been recently elucidated is papain (Drenth et al., Nature 218 (1968) 929). The study has been done at a resolution of 2.8 A. The molecule appears to be made up of two parts separated by a cleft where the active site is located. The total α-helix content is about 20%, in agreement with the conclusions of optical rotatory dispersion measurements. It has been demonstrated that the active site contains a cysteine and a histidine residue; their presence had been proposed earlier on the basis of other evidence.

4. Haemoglobin and myoglobin

Important information on the structures of these two proteins continues to be provided by the Cambridge group.

At the recent Mosbach Colloquium (April 1968) H. Watson and Nobbs have reported the results of differential X-ray analysis of oxy versus ferric myoglobin, aiming, specifically, to establish the structure

of the oxygen-iron complex. The study is admittedly not perfect, crystallographically, due mainly to the instability of the oxygenated derivative during exposure to X-rays. However, the results leave little doubt that the oxygen molecule is bound to the iron asymmetrically with the axis of the oxygen molecule inclined approximately 120° with respect to the plane of the haem, in a very similar manner to that predicted a long time ago by Pauling.

Perutz and his coworkers have published two important papers on the structure of haemoglobin.

One of these (Bolton, Cox, and Perutz, J. Mol. Biol. 33 (1968) 283) contains a three dimensional Fourier synthesis of horse deoxyhaemoglobin at 5.5 A resolution. The results are very significant because, added to the previous ones, they now permit a direct comparison of the structures of the deoxy and oxy derivatives of human and horse haemoglobin. The main conclusion is the similarity, between horse and human haemoglobin, in the structure of any of the two derivatives and hence in the structural changes associated with oxygenation.

The fact that no significant difference has been found between the structures of horse and human haemoglobin, in spite of 17 differences in the residues of the α-chain and 25 in the β-chain, might appear surprising. However, the authors point out that these differences in the sequence are not structurally important both because of the similarities in many of the residues substituted or because of their non-critical role.

Another aspect of the results concerns the difference in conformation between oxy and deoxy haemoglobin: again the X-ray data only show changes in the arrangement of the subunits but no differences in the secondary or tertiary structures of the chains in the two derivatives. This leads to the suggestion that the intrachain conformation changes, which have been postulated on the basis of several arguments, might well be so small, even if critical for the function, as to be out of reach of present methods of X-ray analysis of proteins.

The other paper by Perutz and his coworkers (Perutz et al., Nature 219 (1968) 131) contains the results of their studies at 2.8 A resolution. These provide a true atomic model of the protein, a step forward with respect to the previous 'tentative' atomic model which was based on the analysis at 6 A resolution, on the chemical sequence, and on the atomic model of myoglobin. The new model gives the expected critical details of the important parts of the protein: the interactions between the haem and the residues on the polypeptide chain, the regions of contact between the chains, and the exact location of special groups such as the reactive and 'masked' sulphydryls.

The environment of the haem (the hydrophobic pocket), although in general similar in the α- and β-chains of haemoglobin and in myoglobin, shows distinctive features in the three cases. The residues which interact with the haem are mostly invariant in the sequences of haemoglobins from different species, a finding which emphasizes the critical role and the specificity of the interactions.

In structural terms, the explanation of the cooperative interactions between the haems lies to a large extent in the region of contact between the polypeptide chains. From the model the important contacts appear to be those between an α-chain (α_1) and a β-chain (β_2) which form one of the two possible twin pairs of α,β-chains. This region of contact appears to be an ideal pathway for transmission of effects from one chain to another.

5. New tools in protein chemistry

High resolution proton nuclear magnetic resonance (NMR) is rapidly becoming a major tool in the study of protein structure and function. This is testified, for instance, by several papers on NMR spectra of proteins which appeared in the May and June issues of the Proceedings of the National Academy of Sciences (nos. 1 and 2 of volume 60). These are: an NMR study of the active site of carboxypeptidase A (Navon et al., Proc. Natl. Acad. Sci. 60 (1968) 86), an NMR study of lysozyme (Cohen and Jardetzky, Proc. Natl. Acad. Sci. 60 (1968) 92), and one of sperm whale cyanmetmyoglobin (Wüthrich, Shulman and Peisach, Proc. Natl. Acad. Sci. 60 (1968) 373).

NMR spectra of proteins contain a great amount of information due to signals from many hundreds of protons; however the spectral assignments are often difficult and at present a good deal of effort goes into correlations of the detailed structure of the NMR spectrum with information from X-ray analysis and chemical studies.

In myoglobin the NMR spectra give a detailed pic-

ture of the electronic situation in and around the haem group. From the spectra as a function of temperature, proton resonances lying outside the range usually covered by resonances of diamagnetic compounds could be related to hyperfine interactions with the paramagnetic haem group and to interactions with local magnetic fields produced by ring currents in porphyrin and imidazole. NMR contact shifts seem to be a probe of the electronic wave functions of the haem group in the native protein; thus the systematic NMR study of different myoglobin and haemoglobin derivatives looks most promising in unravelling structure-function relationships in these proteins.

The NMR spectra of hen egg-white lysozyme have been obtained after denaturation of the enzyme in various ways and in the presence of the inhibitor *N*-acetylglucosamine. The sharpest peaks are obtained at 65°C in 8 M urea and excess 2-mercaptoethanol, which breaks the four disulphide bonds. Under these conditions it is possible to assign the major peaks to the phenylalanine and tryptophan protons.

In the experiments performed after heat denaturation the spectra reflect the rigidity of the molecule due to the unbroken disulphide bonds. In 8 M urea no important conformational changes occur. Experiments with the inhibitor show that aromatic residues, particularly tryptophan, are implicated in the binding site.

6. Isozymes

A volume of the New York Academy of Sciences (volume 151 (1968) pp. 1-689, consulting editor E. S. Vesell) is dedicated to 'Multiple Molecular Forms of Enzymes'. The amount of information now available on isozymes is impressive. This may depend as Kaplan points out (Kaplan, Ann. N. Y. Acad. Sci. 151 (1968) 382) on the fact that it now appears that enzymes existing in only one molecular form are exceptions. However, apart from the true genetic variants, multiple forms of enzymes may appear in electrophoresis due to a number of causes which should be always looked for before drawing conclusions. Especially revealing in this respect is the analysis of Cann and Goad (Ann. N. Y. Acad. Sci. 151 (1968) 638) that shows clearly how apparent electrophoretic heterogeneity (multiplicity of zones) might arise from interactions with solvent molecules or from isomerisation phenomena. Although these effects have often been discussed in the past, it seems appropriate to emphasize once again that the appearance of multiple zones in electrophoresis might not be due to inherent heterogeneity.

Although the topic is not directly related to the problem of isozymes, in the same volume of the Annals of the New York Academy of Sciences Margoliash and Fitch (Ann. N. Y. Acad. Sci. 151 (1968) 359) discuss several aspects of the variability of primary structure of cytochrome c in different species. They emphasize the evolutionary implications of the sequence data and arrive, through a reconstruction of the phylogenetic tree, at the sequence of ancestral cytochrome c. This study is a good example of how knowledge of protein structure can be used to solve general biological problems.

7. Ligand binding to proteins

Interactions between binding sites are still the most fashionable aspects of studies of ligand binding to proteins. Their significance in regulatory phenomena continues to be emphasized and the number of systems where such effects appear increases steadily.

However, there seems to be some confusion in the whole field: this confusion arises partly from the use of magic words which, having lost their original, sometimes purely descriptive, meaning, serve to provide an *ad hoc* explanation for all sorts of phenomena and are taken to represent general biological principles. Typical in this respect is the case of the word 'allosteric' which, initially introduced to indicate a special kind of enzyme inhibition, has then been used for any kind of effect mediated by conformational changes of the protein and finally for a specific, restricted, highly speculative model for interactions between binding sites in proteins and enzymes. This new sort of dialect does not seem to be justified and it seems highly desirable to return, as far as possible, to standard physical and chemical terminology. In addition it is also felt that a better separation should be achieved in this field between the results of experiments and their interpretation so that experimental findings might still have a clear meaning *per se*, quite apart from the particular model in the framework of

which they were collected.

Kinetic and thermodynamic aspects of ligand binding in complex protein systems have been discussed in two lucid papers in the newly published Quarterly Reviews of Biophysics. Eigen (Q. Rev. Biophys. 1 (1968) 3) describes the possibilities of 'relaxation spectrophotometry' in enzyme reactions and in general in ligand binding reactions of proteins. Wyman (Q. Rev. Biophys. 1 (1968) 35) discusses linkage theory, cooperative effects in binding, site-site interactions, and the role played in these phenomena by the protein conformation, as exemplified by haemoglobin. Both authors point out the important general application of specific models to real systems.

8. Complex kinetic behaviour of enzyme-catalysed reactions and the effect of activators and inhibitors

The proposed models for explaining sigmoid curves for initial velocities versus substrate concentrations generally apply equilibrium treatments to the binding of substrate molecules to the enzyme. The sigmoid curve is taken to result from interactions between several catalytic sites in the molecule located in different subunits. The sites are implicitly assumed to be intrinsically equivalent. The interactions between binding sites for similar or dissimilar ligands (i.e. between different substrate binding sites or between binding sites for substrates and activators) would be physically mediated by conformation changes in the protein. Starting from these general premises a number of restricted models may be and have been proposed. The specific models may be derived from general kinetic and thermodynamic considerations with the imposition of additional limitations and assumptions which seem to be met by real systems. A comparison of the features and possibilities of the various models has recently been given by Koshland and his coworkers (Kirtley and Koshland, J. Biol. Chem. 242 (1967) 4192; Haber and Koshland, Proc. Natl. Acad. Sci. 58 (1967) 2087).

In regard to their validity the case of aspartate transcarbamylase is interesting: a series of papers (Changeux, Gerhart and Schachman, Biochemistry 7 (1968) 531; Gerhart and Schachman, Biochemistry 7 (1968) 538; Changeux and Rubin, Biochemistry 7 (1968) 553) appears to confirm that the behaviour of the system fits remarkably well one of the 'allosteric' models. However, in another paper on this celebrated enzyme McClintock and Markus (J. Biol. Chem. 243 (1968) 2855) state that the experimental findings do not agree entirely with either one of the two most used specific models (the Monod-Wyman-Changeux or the Koshland-Nemethy-Filmer models).

On the other hand, in the case of enzyme kinetics it is more and more realized that such effects, i. e. departure from simple Michaelis-Menten behaviour and sigmoid shapes for velocity-substrate concentration curves, might be present even in the absence of cooperative interactions between substrate binding sites. A recent paper (Sweeny and Fisher, Biochemistry 7 (1968) 561) illustrates well this point. Some experimental evidence for the real existence of situations of this kind in AMP deaminase has been recently obtained in the present author's laboratory (Suelter, Kovacs and Antonini, J. Biol. Chem., in press).

9. Reactions of haemoglobin and myoglobin with ligands – the status of the problem

Haemoglobin continues to play a central role in protein chemistry. The continuously advancing knowledge of its structural and functional properties parallels attempts to formulate models and theories for phenomena such as the haem-haem interaction and the Bohr effect, the mechanisms of which, however, are still largely unknown in detail. This, which is evident from the current literature, came out clearly in the course of a conference held recently on haemoglobin and myoglobin (Second 'La Cura Conference on Haem Proteins', La Cura, Italy, June 1968). The conference which was attended by about 20 people and which lasted for about 2 weeks, was essentially devoted to the mechanisms of the reaction with ligands of these two proteins: several new findings were presented and older ones were analysed and discussed in great detail. However, there often was no general agreement in the interpretation of the experiments and in the main conclusions; previous, apparently firmly established, statements were withdrawn; in addition, several situations seemed difficult to rationalize so that, contrary to the previous La Cura Conference, a somewhat pessimistic atmo-

sphere often dominated the meetings. Some of the most debated problems were: whether the primary unit of function in haemoglobin is the $\alpha\beta$ dimer or the full $\alpha_2\beta_2$ tetramer; the way protein conformation gains control on the reactivity of the haem; the problem of haemoglobin dissociation, in particular equilibria involving single chain molecules; the evidence for intrachain conformation changes accompanying ligand binding; the identification of the Bohr effect groups.