X-RAY ENZYMOLOGY

C. C. F. BLAKE
Laboratory of Molecular Biophysics, University of Oxford, England

It is a brave man – or a foolish one – who claims that he can assess accurately the role of one particular technique in such a wide field of study as that represented by enzymes, particularly as most important advances are the result of using a number of methods. Nevertheless it seems reasonable, on this the tenth anniversary of the work on lysozyme, to claim that X-ray diffraction has now become one of the most powerful techniques in the study of enzymes. That this is not only the view of crystallographers can be seen by observing the prominence given to X-ray methods and results in modern textbooks and review volumes on biochemistry and enzymology. From being a relatively obscure subject in the domain of the physical sciences which had only occasional relevance to biochemistry through the structure determination of molecules such as penicillin or vitamin B_{12}, it has risen to its present important position in enzymology simply because it is the only technique that enables us to 'see' enzymes in essentially atomic detail in three dimensions. Since it is now apparent that the chemistry of an enzyme's active site depends as much on the geometry and environment of its functional groups as their individual chemistry, only by using a technique that reveals these characteristics can we hope to understand the properties of the active site. In this context X-ray studies can not only look at the enzyme itself, but also enzyme–substrate (or – pseudosubstrate) complexes and reveal the location, interaction and geometry of the bound substrate, and additionally show the enzyme's structural response to the binding in terms of reorientation of side chains and conformational changes either in the active site or remote from it. Should the enzyme's response involve the large-scale rearrangement of subunits, then X-ray studies can at least detect the change and often be used to analyse the new structure in as much detail as the original. The unique property of X-ray analysis therefore, is its comprehensiveness, the ability to reveal the complete geometry of an enzyme and to describe in detail the changes that occur in it's interaction with substrate.

Like any other technique X-ray diffraction has its limitations, and these must be clearly stated. It is now clear that X-ray analysis cannot of itself produce a sequence of a protein, which must therefore be obtained separately to interpret fully the X-ray image. Secondly, the X-ray image is a static, averaged picture of the structure which cannot show all the events that occur in the dynamic system of enzyme and substrate. However it may be possible to produce 'snap-shot' images of the dynamic system in its most important points with sufficient ingenuity, for example by freezing-out intermediates at low temperatures. A not very widely known limitation of the X-ray method is the relative lack of precision with which atoms can be placed in the structure. Hydrogen atoms cannot be located at all, except by implication, and other atoms are subject to positional errors of at least 0.25 Å even with high resolution work, that cannot be reduced to better than 0.1 Å even after exhaustive crystallographic refinement. Uncertainties of this order may be large when detailed enzyme mechanisms are being considered. It is of course possible that the basically more sensitive spectroscopic techniques may be used in the light of the X-ray results to define the fine details of the structure, provided of course that these methods live up to the claims of their protagonists.

The remainder of this essay will be concerned with describing the contribution of X-ray crystallography to the solution of some of the major problems in enzymology. If it were not for the major practical limitation of the method, the necessity of producing large crystals of the enzyme to be analysed, this contribution would be much greater.

North-Holland Publishing Company – Amsterdam

Enzyme structure

It is not possible here to describe the detailed structure of any one enzyme. The range of structures exhibited by enzymes is very wide, and each molecule has its own special characteristics. Nevertheless there are certain aspects of structure that are common to all, or nearly all enzymes, some of which are understood while others are not.

As had previously been suspected all globular enzyme molecules have a close-packed core composed almost exclusively of hydrophobic side chains, while their exteriors are largely hydrophilic. However the segregation is not quite as sharp as had been supposed; an analysis by Lee and Richards [1] has indicated that as much as 40% of the surface of proteins is composed on non-polar atoms. The exclusion of hydrophilic, and especially ionizable side chains from the core is however marked unless they are required for a particular function, such as the internal ion-pair in the serine proteases that is required for activation. In some enzymes the essentially polar main-chain traverses the core, but it is found in these cases that the hydrogen-bonding potential is always met, usually by other segments of the main-chain.

In general the articulation of the main-chain required to bury the hydrophobic side chains that form the core is well in agreement with the principles of the conformational analysis of proteins suggested, in the first instance by G. N. Ramachandran [2]. Certainly when the main-chain dihedral angles are plotted on a Ramachandran diagram they are found to cluster closely in the minimum energy regions of the diagrams. That they do so is not entirely a consequence of the fact that the α-helix and the β-sheet, regular structures that correspond to a succession of residues having the same set of dihedral angles, are the common secondary structural elements in enzyme structures. Although an analysis by Crawford et al. [3] of seven enzymes shows that 75% of their residues are located in just three regular conformations, the α-helix, β-sheet and the reverse turn – a sequence of three peptides arranged so that the chain direction turns through 180° – the marked clustering of angles around the minimum energy values in a molecule like lysozyme that has a relatively small proportion of helix and sheet shows that each residue in so-called 'random chain' also obeys the Ramachandran rules.

An unexpected feature of the initial X-ray studies of proteins is that the α-helix and β-sheet do not assemble in the manner that had been anticipated. Neighbouring helices, for example, are never found with their axes lying parallel, but seem rather to prefer to pack at an angle of 20° or greater. Similarly the expected flat β-sheet has never been found in enzymes, instead the sheet is always found to be twisted in the same sense. The reason for this has been analysed by Chothia [4] as being a second-order effect of the Ramachandran approach in which the strands in the β-sheet are in their most stable conformation when running at about 20° to one another. Despite the presence of only these two types of major secondary structure, the number of different tertiary structures that can be constructed from them is surprisingly large. Protein molecules have been found that are composed mainly of α-helix at one extreme to structures made up almost entirely of β-sheet, either in the form of a cylindrical sheet or two twisted sheets face-to-face, at the other. Combinations of helix and sheet can give rise to more varied structures, and for that reason probably, are more common. One α-β structure that is particularly common has a central twisted β-sheet flanked on each side by helices. This sandwich structure tends to be made up of units composed of a β-strand linked to a helix that is folded back to run antiparallel to the strand. Rao and Rossmann [5] have called units of this kind 'super-secondary structures' and have suggested that they, rather than individual helices or β-strands, are the real building blocks of proteins. If repeated regularly super-secondary structures of this kind would produce proteins of particularly simple structure of the kind that have recently been observed in the enzymes of glycolysis, particularly clearly shown by triosephosphate isomerase [6] and phosphoglycerate kinase [7].

One striking feature of structure that is relatively common in enzymes, is the folding of the polypeptide chain into two globular units. These units, or domains, nearly always seem to correspond to the folding of each half of the chain and thus are only linked by a single length of chain. Remarkable in all enzymes so far discovered that have a domain structure, the active site, or perhaps more precisely the catalytic site, is situated between the domains, each of which contribute important residues to the site.

Most intracellular enzymes are polymeric. Those

that have been analysed by X-ray diffraction mostly have two or four identical subunits. They are usually found to be arranged symmetrically, although one heterologous dimer, yeast hexokinase, has been discovered. A wide variety of subunit interactions have been found involving both helices and β-sheets in hydrophobic and hydrogen-bond interactions. Generally the active sites in these molecules are found on the surface remote from one another, but in glyceraldehyde phosphate dehydrogenase, the co-operativity of substrate binding may be understood in the close pairing of the active sites [8], while the allosteric enzyme aspartate transcarbamylase has its active sites facing a central cavity in a hollow shell molecule [9].

Active sites

The most notable achievement of X-ray studies of enzymes has, of course, come in the view it provides of the structure of the active site. The first and most obvious generalization that has appeared is that the active site seems almost invariably to take the form of a slot, depression or cavity in the surface of the enzyme molecule. The shape and size of this feature is such that the substrate can only be accommodated if all the water molecules in the site are expelled on substrate binding. Thus the immediate result of the X-ray studies is to indicate that enzyme mechanisms operate in a local non-aqueous environment.

X-ray studies of enzyme–substrate complexes have often provided a clear explanation of the known specificity of the particular enzyme in terms of the binding characteristics of the active site. All the early X-ray work was carried out on extracellular enzymes that were active against polymeric substrates. In accord with these characteristics it was found that the active sites were composed of a number of subsites, each capable of binding one unit of the polymer. The distinction between exo- and endo-activity on the polymer chain is generally apparent from the carboxypeptidase molecule (an exopeptidase) whose binding site is a groove leading to a cavity that can accommodate the terminal residue of its polypeptide substrate [10] as opposed to the open-ended slot found in for example the endopeptidase, papain [11]. At a more detailed level, a site was found in a number of enzymes that explained the particular specificity of the particular enzyme. A good example comes from lysozyme [12], where one of the six sub-sites capable of binding a saccharide unit contains a deep narrow slot that is exactly complementary in size and chemistry to the *N*-acetyl side chain, thus explaining the known specificity of the enzyme for *N*-acetyl sugars. A more remarkable result comes from the work on three homologous serine proteases where it can be seen that discrete side chain substitution in the specificity slot can account for the differing specificities of the group of enzymes. In the chymotrypsin molecule the slot is lined with hydrophobic residues and can accommodate the aromatic side chains of polypeptide substrates [13], the substitution of an aspartate at the bottom of the slot in trypsin changes the specificity through its ability to form a salt-bridge with lysine or arginine residues [14], while in elastase the blocking off of the slot by bulkier residues appears to account for its specificity towards small amino acid side chains [15].

Evidence for the significance of the specificity site in an enzyme's catalytic apparatus has come unexpectedly from X-ray work on chymotrypsinogen [16]. In view of the great difference in activity between chymotrypsin and its zymogen precursor it was anticipated that a structural explanation in terms of an unformed or blocked catalytic site would be forthcoming when the structures were available for comparison. However it was found that the charge relay system that appears to be essential to the catalytic activity of the enzyme (see below) was intact in the zymogen. The relatively localized changes in tertiary structure that result from activation were instead found to affect the specificity site, and it now appears that the zymogen is inactive not because it cannot promote bond scission but because it cannot bind and orientate the substrate in the active site [17].

In all the enzymes so far examined by X-ray diffraction the binding of substrate leads to changes in the enzyme's structure as Koshland proposed in his theory of induced fit [18]. These conformational changes seem to involve closing the active site slot or cavity, reinforcing the tendency to expel water and also concentrating and possibly orientating the catalytically important residues around the substrate. The extent of these changes varies a great deal: in the serine proteases rather small reorientations are observed, while in carboxypeptidase [10] and lactate dehydrogenase [19] side chains are observed to move by 10–20 Å

to make contact with the substrate. It is interesting to note that recent experiments with lysozyme [20], seem to confirm the observation made with carboxypeptidase [10] that occupation of the specificity site provides the trigger for the conformational changes.

Perhaps the most far-reaching conclusion that has been drawn from X-ray studies of enzymes is that catalytically important residues can be located in quite different chemical and physical environments that can profoundly alter their properties. An early example involved the two acid groups in the catalytic site of lysozyme [12]: Asp-52 was found to be located in a polar environment suggesting a low pK, while Glu-35 is in a much more non-polar environment suggestive of an anomalously high pK. These differences, subsequently confirmed by other techniques, provided one of the principle bases for the hypothesis of the enzyme's activity that came directly from the X-ray work [21]. Probably the most striking example of this effect so far discovered comes from the X-ray study of chymotrypsin [13]. Here it was found that the two known catalytic residues Ser-195 and His-57 were involved in a hydrogen-bonding network with a third residue Asp-102 whose importance had not previously been suspected. Surprisingly the acid was found to be buried in the molecule in an essentially hydrophobic environment, which suggested to Blow and Hartley [22] that its negative charge could be relayed to the serine oxygen via the hydrogen-bond network, dramatically increasing the nucleophilicity of the serine residue. This example demonstrates better than any other, the extension of understanding that is possible by being able to 'see' the active site and to observe the geometric and environmental factors that give certain side chains unusual and unpredictable chemical properties.

The determination of the extent of distortion or change in conformation of the substrate that takes place on binding is appreciably more difficult than defining the response of the enzyme. This is partly because it is difficult to work with true enzyme substrate complexes, and partly because the changes to the substrate are usually small. However there are two clear examples from the X-ray work of distortion of substrate on binding. The work on lysozyme seems to have thrown up examples of all the general effects found in the later work and so it is here [12]. In lysozyme the extension by model building of the observed binding of a trisaccharide inhibitor to the true hexasaccharide substrate could only be achieved by distorting the fourth sugar residue from its expected chair conformation. The distortion became an important element in the hypothesis of the enzyme action and its presence has been supported by other studies. However the latest X-ray work [23,24] suggests that the distorted conformation is not a half-chair as had originally been proposed, but possibly a sofa or boat conformation. The other evidence for substrate distortion has been obtained from the very accurate structure determination of the complex between trypsin and the pancreatic trypsin inhibitor [25]. In this complex it is found that the inhibitor is bound with the lysine side chain of residue 15 in the specificity pocket and 15–16 peptide lying across the enzyme's catalytic site. The peptide unit is strongly deformed towards a tetrahedral conformation at its carbonyl carbon, with the oxygen of Ser-195 at the apex at a distance of 2.3 Å. It is tempting to interpret this conformation in terms of the proposed tetrahedral intermediate of the chymotrypsin mechanism, as a locked transition state.

Enzyme regulation

The extrinsic regulation of enzyme activity is clearly one of the most important considerations in the study of enzymes. In principle X-ray diffraction is an ideal tool for showing how regulation is achieved in specific cases, by comparing native structures with those combined with effectors. However, regulatory enzymes are usually relatively large and it is only comparatively recently that such enzymes have begun to be investigated by X-ray techniques. Although the allosteric behaviour of no enzyme has yet been explained in detailed structural terms, Perutz [26] has been able to describe a structural model for the co-operative binding of oxygen by haemoglobin in terms that are likely to be relevant to enzymes.

Perutz's model is derived from the structural analysis of the compact deoxy-form of the tetramer and the oxy-form in which the subunits have become rearranged in a looser form. The trigger for the allosteric changes seems to be the haem-iron which undergoes a change in spin-state and a decrease in ionic radius on binding oxygen. The decrease in radius on oxygenation permits the iron, that is out of the haem plane in the deoxy-

form, to move 0.75 Å into the plane, taking the haem-linked histidine with it and causing small changes to the tertiary structure. These changes require the breaking of salt bridges that link the subunits together in the deoxy-state and, at the same time, change the environment of ionizable groups so that protons (the Bohr protons) are released. Diphosphoglycerate, a known effector of haemoglobin is found to bind in a central cavity of the tetrameric molecule when it is in the deoxy-conformation, with its two phosphate groups linking the β-chains together by salt bridges. No binding is found, nor can it be expected, to the oxy-form of the molecule because the central cavity becomes too constricted. In agreement with the hypotheses of allostery, the substrate and effectors (0_2, H^+ and 2,3-diphosphoglycerate) are bound at spatially distinct sites and the interplay between them is mediated by changes in quaternary structure. However the allosteric behaviour of the haemoglobin molecule does not agree in detail wit either the model proposed by Monod et al. [27] nor with Koshland's [28], but has features of both. The essence of the haemoglobin mechanism is that the interaction energy arises through step-by-step release of the constraints on the unreactive deoxy-form, which changes the equilibrium in favour of the oxy-form and diminishes the work required to change the tertiary structure of each subunit from the unreactive to the reactive form.

Although the allosteric mechanism of an enzyme has yet to be worked out in this detail, some preliminary observations have been made on yeast hexokinase [29]. This enzyme has two chemically identical subunits which strikingly dimerize through heterologous interactions involving non-integral rotation about an axis with translation along the axis. Binding of substrate to this asymmetrical dimer shows an interesting interdependence. Glucose substrates bind to one subunit preferentially, causing extensive conformational changes. In the absence of sugars, nucleotides do not bind, but in their presence ADP and β-γ-imido ATP bind predominantly to a site located between the subunits, and in doing so cause the sugar binding to become equivalent. Presumably the intersubunit nucleotide binding site, which because of the heterologous subunit interactions is made up from different parts of the two monomers, is only fully formed by the conformational changes induced by sugar binding. In turn the nucleotide binding induces further changes that fully form both sugar binding sites. This intersubunit site appears to be the regulatory site because there is in addition a further nucleotide binding site on each subunit sufficiently near the hexose sites for it to represent the location of the phosphorylating nucleotide. This is clearly a complex system with complex behaviour, the details of which will take some time to work out. It is however difficult to see how any other technique can provide this kind of information.

Evolution of enzymes

The final important aspect of enzymology to which X-ray studies can make an important contribution is the discovery and definition of evolutionary pathways amongst enzymes, to see how new functions can arise from modification of structure. Much of the work in this field has depended, and will continue to depend, on sequence analysis. However, X-ray analysis has two contributions to make: first, by defining the real nature of a relationship and second, to use the clear observation that tertiary structures are more stable in evolution than primary structure to discover evolutionary relationships that are beyond the horizon of sequence homology.

An obvious example of the first contribution of X-ray studies, is the definition of the reason for the specificity changes in the serine proteases that has already been discussed. Another that is currently in progress [30] involves a comparison of the structurally homologous enzymes, lysozyme and α-lactalbumin to discover the changes that have been made to the active site of lysozyme to produce the quite different function that lactalbumin has as part of the lactose synthetase system.

X-ray analysis can also make an important contribution when sequence analysis and functional analysis lead to an ambiguous result as they have with the bacterial protease, subtilisin. The X-ray work showed that although the subtilisin molecule has an entirely different tertiary structure from the mammalian serine proteases its active site contains a copy of the serine protease charge relay system [31]. On the other hand recent work [32] on the SGPB protease from *Streptomyces griseus* has revealed that it has a clear structural relationship to the mammalian proteases even though its primary structure is only remotely

related showing numerous large deletions and insertions in its polypeptide chain in relation to its mammalian counterparts. These are prime examples of convergent evolution (albeit of active site structure, *not* of tertiary structure), and divergent evolution from a common ancestor respectively.

On the basis of the evidence for the stability of tertiary structure to evolutionary pressures produced from many examples including those given above, it seems reasonable to ascribe a history of divergent evolution from a common ancestor to molecules that are found to have similar tertiary structures, even though their sequences may not be homologous, provided there is collateral evidence of functional similarity. A fascinating instance of this kind has come from the X-ray work on four enzymes, lactate-, malate-, alcohol-, and glyceraldehyde phosphate-dehydrogenases [33–36]. The first two which are structurally closely homologous, are related to the others in a surprising manner. All the enzymes are binuclear with one domain functioning as a binding unit for the NAD co-factor while the other is involved in binding the specific substrate. The co-factor binding site has a very similar basic structure [37] in all four enzymes (showing differences no greater than those between the ancestrally homologous SGPB protease and chymotrypsin molecules), but the specific substrate binding part has a quite distinct structure in each enzyme. Interestingly a similar structure to the NAD binding unit has been found in flavodoxin [5] and phosphoglycerate kinase [7], where it is also involved in nucleotide co-factor binding. The proposal that this nucleotide binding unit has evolved from a common ancestral protein is reinforced by the observation that the various nucleotides (NAD, FMN and ATP) are bound at essentially the same site on the common structural unit [38]. Results of this kind seem to provide a basis for tracing the evolution of enzymes in the acquisition of new functions that underpins the evolutionary processes in organisms.

References

[1] Lee, B. and Richards, F. M. (1971) J. Mol. Biol. 55, 379.
[2] Ramakrishnan, C. and Ramachandran, G. N. (1965) Biophysical J. 5, 909.
[3] Crawford, J. L., Lipscomb, W. N. and Schellman, C. G. (1973) Proc. Nat. Acad. Sci. USA 70, 538.
[4] Chothia, C. (1973) J. Mol. Biol. 75, 295.
[5] Rao, S. T. and Rossmann, M. G. (1973) J. Mol. Biol. 76, 241.
[6] Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I. and Wilson, I. A. (1975) Nature 255, 609.
[7] Blake, C. C. F. (1975) Essays in Biochemistry 11, 37.
[8] Buehner, M., Ford, C. G., Moras, D., Olsen, K. W. and Rossmann, M. G. (1974) J. Mol. Biol. 90, 25.
[9] Wiley, D. C., Evans, D. R., Warren, S. G., McMurray, C. H., Edwards, B. F. P., Franks, W. A. and Lipscomb, W. N. (1971) Symp. Quant. Biol. 36, 285.
[10] Hartsuck, J. A. and Lipscomb, W. N. (1971) The Enzymes, 3rd Edn. Vol. III, 1.
[11] Drenth, J., Jansonius, J. N., Koekoek, R. and Wolthers, B. G. (1971) The Enzymes, 3rd Edn. Vol. III, 485.
[12] Blake, C. C. F., Johnson, L. N., Mair, G. A., North, A. C. T., Phillips, D. C. and Sarma, V. R. (1967) Proc. Roy. Soc. B167 378.
[13] Blow, D. M. (1971) The Enzymes, 3rd Edn. Vol. III, 185.
[14] Stroud, R. M., Kay, L. M. and Dickerson, R. E. (1971) Symp. Quant. Biol. 36, 125.
[15] Hartley, B. S. and Schotton, D. M. (1971) The Enzymes, 3rd Edn. Vol. III, 323.
[16] Kraut, J. (1971) The Enzymes, 3rd Edn. Vol. III, 165.
[17] Wright, H. T. (1973) J. Mol. Biol. 79, 1; ibid 13.
[18] Koshland, D. R. (1958) Proc. Nat. Acad. Sci. USA 44, 98.
[19] Adams, M. J., Buehner, M., Chandrasekhar, K., Ford, G. C., Hackert, M. L., Liljas, A., Rossmann, M. G., Smiley, I. E., Allison, W. S., Everse, J., Kaplan, N. O. and Taylor, S. S. (1973) Proc. Nat. Acad. Sci. USA 70, 1968.
[20] Snape, K. W., Thesis, D. Phil, Oxford University.
[21] Phillips, D. C. (1967) Proc. Nat. Acad. Sci. 57, 484.
[22] Blow, D. M., Birktoft, J. J. and Hartley, B. S. (1969) Nature 221, 337.
[23] Ford, L. O., Johnson, L. N., Machin, P. A., Phillips, D. C. and Tjian, R. (1974) J. Mol. Biol. 88, 349.
[24] Beddell, C. R., Blake, C. C. F. and Oatley, S. J. (1975) J. Mol. Biol. 97, 643.
[25] Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J. and Steigemann, W. (1974) J. Mol. Biol. 89, 73.
[26] Perutz, M. F. (1970) Nature 228, 726.
[27] Monod, J., Wyman, J. and Changeux, J. P. (1965) J. Mol. Biol. 12, 88.
[28] Koshland, D. E., Nemethy, G. and Filmer, D. (1966) Biochemistry 5, 365.
[29] Anderson, W. F., Fletterick, R. J. and Steitz, T. A. (1974) J. Mol. Biol. 86, 261; Fletterick, R. J., Bates, D. J. and Steitz, T. A. (1975) Proc. Nat. Acad. Sci. USA 72, 38.
[30] Phillips, D. C. and Smith, S. J., unpublished results.
[31] Kraut, J., Robertus, J. D., Birktoft, J. J., Alden, R. A., Wilcox, P. E. and Powers, J. C. (1972) Symp. Quant. Biol. 36, 117.
[32] Delbaere, L. T. J., Hutcheon, W. L. B., James, M. N. G. and Thiessen, W. E. (1975) Nature 257, 758.
[33] Rossmann, M. G., Adams, M. J., Buehner, M., Ford, G. C., Hackert, M. L., Lenz, P. J., McPherson, A., Schvitz, R. W. and Smiley, I. E. (1972) Symp. Quant. Biol. 36, 179.
[34] Hill, E., Tsernoglou, D., Webb, L. and Banasyak, L. J. (1972) J. Mol. Biol. 72, 179.

[35] Branden, C-I., Eklund, H., Nordstrom, B., Boiwe, T., Soderlund, G., Zepperzauer, E., Ohlsson, J. and Akeson, A. (1973) Proc. Nat. Acad. Sci. USA 70, 2439.
[36] Buehner, M., Ford, G. C., Moras, D., Olsen, K. W. and Rossmann, M. G. (1973) Proc. Nat. Acad. Sci. USA 70, 3052.
[37] Rossmann, M. G., Moras, D. and Olsen, K. W. (1974) Nature 250, 194.
[38] Ohlsson, I., Nordstrom, B. and Branden, C-I. (1974) J. Mol. Biol. 89, 339.