PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Organization and Function of Water in Protein Crystals [and Discussion]

J. L. Finney and D. D. Eley

Phil. Trans. R. Soc. Lond. B 1977 **278**, 3-32
doi: 10.1098/rstb.1977.0029

Phil. Trans. R. Soc. Lond. B. **278**, 3–32 (1977) [3]
Printed in Great Britain

The organization and function of water in protein crystals

By J. L. Finney
Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX

Dry proteins are dead, or at best asleep. Substitution of D_2O can drastically alter biological activity. Water is thus essential in maintaining the structural integrity of biologically active macromolecules, and is implicated in their functioning. Such water may occupy a range of dynamical states, from being strongly bound and localized, to more labile and 'liquid-like'. Spatially ordering the macromolecules aids the search for the more localized water molecules. For example, diffraction experiments on single crystals can resolve 'bound' water molecules within a protein molecule – often at active sites, coordinated to metals or ions. Less precise information is obtained on the partially occupied external water sites, which are of importance to the folding and the dynamics of the biomolecule. Orientation of fibrous molecules increases the information obtainable from n.m.r. experiments. Combination of other experimental results on disordered aggregates (e.g. in solution) with chemical and structural data on the macromolecule and water itself yields useful, if circumstantial, information. Statistical and computer techniques may help to elucidate the complex nature of water–protein interactions, and to interpret the results of more complex experiments.

1. A protein molecule is a component of a complex system

A protein molecule, be it an enzyme, a transport protein, or a structural component, is part of a *complex molecular system*. Its very existence, as well as its activity, depends upon its surroundings. If we change the surroundings beyond a certain limit, the macromolecule itself will cease to perform its function, and may even become structurally unstable. Thus, in considering the functioning of a protein molecule, we should think about it in relation to its surroundings. Ideally, we should look at the overall *system* – protein, substrate, water, salt, and possibly coenzyme – as an integrated whole.

Consider a simplified two-component system, say protein plus water. In the early stages of investigating the properties of this system, we concentrate on the protein itself, and either ignore the medium, or treat it in very general, hand-waving terms (e.g. a structureless continuum). As our understanding of the protein molecule itself improves, we begin to ask more detailed questions about the relevance of the solvent to a viable protein and its operation. We may not know how important the medium is, if at all, but we suspect we should ask questions about its possible rôle.

In a simple two-component system, we might be able to write symbolically:

$$\text{protein in solvent} \equiv \text{protein} + \text{solvent} \qquad (1)$$

considering a particular measurable property of the system, say the heat capacity. In this simple case, we assume that we can consider the contribution of the protein separately from that of the solvent in such a way that the properties of the total system can be obtained simply by adding the contributions of the individual components. The separate properties of the components can be superposed to give the properties of the complete system.

A more realistic, more general symbolic relation would be:

$$\text{protein in solvent} \equiv \text{protein} + \text{solvent} + \text{protein} \cdot \text{solvent}, \qquad (2)$$

where the third term, the 'cross-term', represents the results of the *interaction* of the two components. It represents, for example, the interdependence of the three-dimensional tertiary structure of the protein and the surrounding medium; the two 'aspects' of the system are mutually interconnected and inseparable from each other.

For a particular system, we want to know the 'magnitude' of this cross-term in comparison with the two direct terms in relation (2) above. In some cases, it may well be that the cross-term dominates, or even that the direct terms have no meaning by themselves: a protein isolated from a suitable medium is not a viable operating system. The functioning of any protein system is a molecular symbiosis. It is also in principle, a much more complex symbiosis than suggested by the simple model above, involving other components such as salt and substrate. We want to try to understand the nature of this symbiosis.

Experimental investigations of complex, interconnected systems are inherently difficult. Because of the 'cross-term' of relation (2), results are very difficult to interpret unless we can focus experimentally on a well-defined aspect of the system. Thus, measurements which average over the whole system are particularly difficult to interpret. For example, although we can make heat capacity measurements on a protein solution, any assignment of partial values to the two components, protein and solvent, are only possible on very broad assumptions about the nature of their interaction. We cannot assume with confidence

$$\text{C(protein in water)} = \text{C(protein)} + \text{C(water)}$$

nor can we easily allow for the contribution of the protein–water interactive element without assumptions about the variety of the molecular states associated with the interface.

In the terms of an impractical physicist like myself, we would like to know the space and time distribution function, $G(\boldsymbol{r}, t)$ of the protein–solvent system. Such a function would tell us the space and time behaviour of the system in macroscopically useful terms. We would like to know the behaviour of the system components over a time period commensurate with, say, one cycle of an enzyme's catalytic process. Statistical mechanics does not ask questions about the individual detailed microscopic motions of the molecules in a simpler assembly of less well-connected molecules, but describes the assembly with more useful statistical measures. So here we would like to do something similar. We do not really want to know the microscopic behaviour of every atom at any time, but how to handle the results of these molecular motions on a time scale relevant to the operation of the protein as a mechanical system. We want to predict long time scale properties of the multicomponent system on the basis of our knowledge of the short time scale events that occur at the molecular level. We cannot even do this for a 'simple', homogeneous, single component system such as water, so the multicomponent system is a long way off. But it could be an ultimate aim of experimental and theoretical studies of such complex molecular systems.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

2. What do we know in general terms about water in proteins?

No single experimental technique can statically or dynamically fully characterize water associated with a protein. However, by combining results from different methods, a 'conventional wisdom' has grown up, which can be stated in general terms as follows (Kuntz & Kauzmann 1973; Cooke & Kuntz 1974).

(1) A small number of water molecules are found internal to protein molecules. Some of these are strongly, irrotationally bound with unknown lifetimes that may be as long as seconds. Others are less strongly bound with shorter lifetimes. It is suspected that these solvent molecules may be involved in maintaining local structural stability, and some may be directly involved in protein function.

(2) A larger number of water molecules appear to be bound at specific sites on the protein surface. These are less strongly bound than the internal molecules, with probably a wide spread of relaxation times, averaging about 10^{-6} s. They act as anchoring points for:

(3) the surrounding region comprising molecules which are still less mobile than the bulk water, with a range of relaxation times perhaps 10–100 times longer than the approximately 10^{-12} s that is characteristic of bulk water. This is the region of a non-specifically 'bound' or 'partly-ordered' water shell, the nature and extent of which is uncertain. It can really only be characterized operationally in terms of the particular technique we may use to investigate it. Water in this region exhibits a lowered vapour pressure, reduced mobility, and a much-reduced freezing point; estimates of its extent depend very much on the assumptions made when interpreting the data.

A model of the 'hydration shell'

Even though this water is not 'bound' in any true sense of the word, it can be thought of as 'partially ordered' in that the characteristic relaxation time, or spread of relaxation times, is significantly longer than in bulk water. We might perhaps think of this region as low temperature, perhaps effectively supercooled water, in which the influence of the detailed geometry of, and the molecular forces operating at, the surface of the protein molecule partially stabilize certain instantaneous, local water configurations. These configurations are likely to be perturbed ones, as a result of the proximity of the protein surface. But it must be stressed that, in describing this water as 'partially ordered', we *do not* mean that we expect local, crystalline, ice-like structures to be present. The implications of a term such as 'ordering' which, in many contexts implies spatial regularity, can be misleading. The surface of a protein molecule – in common with surfaces in general – may restrict the regions of configurational space which are accessible to the water molecules in the neighbourhood, but this restriction does not extend to forcing the formation of actual crystalline arrangements, which is a very specialized local molecular organization occupying an almost infinitesimal volume in the phase space of the statistical mechanician. 'Ordering' in this situation means 'softening' diffusive and translational modes of vibration, with an increase in the lifetimes of particular (non-crystalline) molecular configurations, and an increase in the probability of occurrence of certain of these configurations at the expense of others.

The importance of the 'hydration shell'

Despite the difficulties of characterizing this region of perturbed water, it may be important that we do so. It is thought to be important in maintaining the structural stability of the protein

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

molecule, and may be a major influence on the mechanical behaviour of the molecule. Not only are changes in this region important in the folding of an extended polypeptide chain into the active, tertiary structure we call a protein, but the way in which this folded molecule fluctuates as a mechanical system – for example, during conformational changes occurring in an enzyme reaction – will depend upon its interaction with this surrounding medium. Upon the nature of the surrounding medium depends the transport of coenzyme, substrate, and protons or water molecules necessary for the protein operation. This disturbed water region is potentially of both structural and dynamical importance.

Attempts to understand the effect of this region have been limited to macroscopic energy estimates of the gross energy benefit or deficit of particular groups of atoms being exposed to solvent. By estimating the change in energy observed on transferring molecules between polar and non-polar environments, we have quantified the idea of the 'hydrophobic bond', the effect which tends to lead to the aggregation of non-polar groups away from a polar solvent. We have, however, no connection between the detailed molecular interactions which lead to this hydrophogic segregation effect, which is thought to be the major driving force in protein folding, and the final situation in static energy terms. We cannot yet connect the molecular interactions with their end result.

Thus it may be that this 'weak binding' region of solvent is that of most significance to the nature and behaviour of the complex molecular system we call a protein. It is this region in which the hydrophobic effect is 'operative' at the molecular level during the folding of the protein. It is here that the 'ordering' of water close to hydrophobic surface groups will occur, and where the effects of added salts and changes in charge distribution on the protein surface will be most felt. We want to know more about it in molecular terms. For example, what is the extent of this region? Does the degree of partial immobilization, or 'ordering' of the water molecules fall off continuously as we move away from the protein surface, or are there abrupt changes?

A major difficulty in understanding this region is our lack of understanding of bulk water itself. This 'hydration shell' is a region in which the water structure is perturbed, probably partially stabilized by the influence of the protein surface. Yet we do not understand, or at least fully agree amongst ourselves, about the nature of liquid water – the original state before the perturbation took place. Computer simulation studies (e.g. Stillinger & Rahman 1974) are shedding more light on this (simpler!) problem, not least by demonstrating the implausibility of many theories previously current. We are, however, still a long way from being able to deal theoretically with *any* liquid in detailed structural terms, let alone one with relatively complex molecular interactions such as water.

3. The use of diffraction techniques in protein–water studies

To what extent can diffraction methods verify this 'conventional wisdom' of protein water? To what extent can such methods fill in some of the details of specific and non-specific water–protein and water–water interactions in a protein–water system, and tell us something about their rôle in biological functioning?

Inelastic neutron scattering can in principle give us information concerning the space–time correlation function $G(\boldsymbol{r},t)$ for the complete protein–solvent system. However, the interpretation of inelastic neutron scattering data from such a complex system is presently beyond our capa-

bilities. We are still arguing the interpretation of inelastic scattering studies of a relatively simple system – pure bulk water (Page 1972).

If we *order* our protein molecules into a regular, repeating pattern, we can use both X-ray and elastic neutron scattering techniques to examine the molecular structure of an individual molecule. An array of molecules – no matter how large and complex – arranged in a repeating pattern acts as a diffraction grating for suitable wavelengths of radiation (*ca.* 0.1 nm). Providing we can unscramble the diffraction information effectively, we can in principle elucidate the arrangement of the component atoms within the individual molecule. The crystal acts as a *molecular amplifier*, enabling us to 'see' the detailed structure of the molecule as if through a molecular microscope.

Many molecules of biological interest will not crystallize in three dimensions, for example fibrous proteins such as collagen. In these cases, orientation parallel to the fibre axis can be thought of as one-dimensional crystallization. Useful, though less detailed information, can be obtained by diffraction and other techniques from these less-ordered systems (see, for example, Berendsen & Migchelsen 1966; Ramachandran & Chandrasekharan 1968; Migchelsen, Berendsen & Rupprecht 1968; Yonath & Traub 1969; Dehl & Hoeve 1969; Dahlborg & Rupprecht 1971; Migchelsen & Berendsen 1973; Doyle *et al.* 1975).

Before proceeding to review the information on water in proteins which can be obtained by crystal diffraction techniques, we should stress a major drawback of the techniques. A. V. Hill summed it up by saying: 'Of course, that is not really biology at all. It is crystals, and unless it wriggles, it isn't biology'. We are interested in the behaviour of protein molecules and water in a real biological context. We want to know about the catalytic or transport mechanism of, or structural rôle of, a protein–water system *in vivo*. This problem of the applicability of results from crystal studies to molecules in solution has always been worrying. In many cases, it has been shown that differences between the conformation of a protein molecule in the crystal and in solution are small; sometimes these differences can be discerned, for example by n.m.r. techniques. In contrast, however, we have good reason to believe that the water component in protein crystals is further from biology than is the protein component. The salt concentrations and pH ranges which facilitate crystallization are often those in which biological activity is not maintained or proceeds at a much-reduced rate. These variables, as well as lattice effects, are likely to strongly perturb at least the surrounding 'hydration shell'.

With these problems in mind, we now ask what studies on protein crystals can tell us about the water phase in the protein–water system, and how its location, organization, and dynamics might be implicated in the functioning of the total system.

4. What do we know and not know about water in crystalline proteins?

The end result of a protein crystal structure determination is a set of atomic coordinates which tells us the location of (hopefully) every atom in the molecule. These coordinates may be first obtained by fitting a (known) polypeptide sequence to the electron density map which is derived experimentally from diffraction measurements on the crystals. In favourable cases, refinement procedures can result in a self-consistent set of accurate atomic coordinates.

There are three kinds of information which such diffraction studies can give us about the organization of water in crystals.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

(*a*) *Indirectly from protein coordinates*. Using only the coordinates of the constituent atoms of the protein molecule, we can find out quite precisely *where the water phase is not*. Wherever we find protein, we do not have water. This may seem at first sight a trivial statement. Indeed, it probably is, but it still contains positively useful information; excluded volume arguments such as this are often quite powerful. For example, by considering the relative locations of particular types of atoms or groups of atoms with respect to the molecular surface, we can make inferences about the influence of the solvent on the molecular structure. The information is not at a molecular level, but it does show us the *result* of molecular solvent effects on the way in which the molecule folds and is stabilized. It can also suggest explanations for some aspects of the operation of the system.

(*b*) *Directly from water coordinates*. In favourable high-resolution studies (generally at least 0.2 nm) we can actually locate water molecules which are apparently bound to the protein. Such molecules must either be strongly bound to the protein for times that are long with respect to the length of the diffraction experiment, or occupy a very localized, though exchangeable site so that some solvent molecule appears to be there throughout the experiment. The first type can truly be thought of as a permanent component of the protein.

(*c*) '*Hopeful inferences*' *from water coordinates*. Many weak peaks of low electron density are found within the solvent region between protein molecules in the crystal (see Fig. 6 of Lipscomb *et al*. 1968). These peaks, where real, correspond to regions of space where solvent is likely to be found, but is not there all the time. Their locations relate to the more weakly bound surface water molecules, and to those making up the perturbed surrounding water shell. These data are by far the most difficult to disentangle.

(*a*) *Where the water is NOT*

Early protein structures verified what had been previously thought about the relative locations in the protein of particular types of group. For example, the 'hydrophobic' side chains which preferred non-polar environments (e.g. alanine, leucine, phenylalanine) tended to pack together inside the molecule, apparently shunning contact with the solvent. In contrast, the more polar, and the ionic side-chains tended to be exposed to the solvent at the surface of the molecule. Presumably these groups were hydrated in some way, interacting directly and specifically with the surrounding solvent.

Knowing the relative exposures of particular groups to the 'solvent' – still at this stage considered as a structureless dielectric medium – calculations were made of the relative magnitudes of the various stabilizing forces holding the molecule in its three-dimensional folded structure (Kauzmann 1959). The general conclusion was that the internal hydrogen bonding between polar groups in the molecule, although extensive and with few possible hydrogen bonds unmade, made only a minor contribution to the stabilization energy. The major contribution to the stability of the tertiary structure was the tendency of the non-polar groups to aggregate together, so that they were shielded from the relatively unfavourable interaction with water. This tendency was called 'hydrophobic bonding'. Thus, it appeared that protein structures are stabilized mainly by the apparently negative philosophy of certain groups hiding from places they do not like.

So far, we can consider this hydrophobic effect only in overall, aggregated terms, without any detailed knowledge of its molecular mechanism. In the folded protein, with shielded hydrophobic groups, we see the end result of a complex set of dynamic interactions at the molecular

level, the ultimate result of the associated driving forces being isolation of particular groups from water. The molecular nature of the interaction of non-polar groups with water is not understood, although we do know that there are perturbations of the solvent structure, and changes in diffusive and translational modes of water molecules generally termed as 'ordering'. During folding, the hydrophobic interactions occurring in the solvent around the hydrophobic groups appear to be strong determining factors of the final tertiary structure. Such interactions would occur in the surrounding, 'perturbed' solvent region (§2*c*), and the water 'ordering' in this region around the equilibrium, folded structure may thus be of importance in maintaining structural stability, and controlling the dynamics of the molecule. At least, we know that disrupting this 'ordered' water region can denature the protein.

As more and more protein structures were solved, it became apparent that this 'hydrophobic in, hydrophilic out' tendency was only a generalization. Hydrophobic groups are found on the *surfaces* of many protein molecules, raising the interesting question of the molecular organization of the solvent around these patches. Studies of crystals containing both hydrophobic groups and water have shown the water surrounding such groups is strongly stabilized into a cage, or 'clathrate' structure (Jeffrey & McMullen 1967). Computer simulation studies of a single neon atom in water (Rahman & Stillinger, unpublished) showed a surrounding cage-like water structure stable over the (relatively short) time period of the calculation. Might such partially-stabilized, cage-like structures be found around these hydrophobic groups in a protein accessible to the surrounding aqueous medium? (Klotz 1958).

Hydrophobic 'patches' are often found to be associated with the surface of contact between two protein molecules in the crystal, or, more interestingly, between molecules which aggregate into polymers. This suggests that the driving force for the aggregation of multi-subunit proteins may also be hydrophobic. The insulin dimer appears to aggregate by two hydrophobic surface patches coming together, excluding the solvent from close proximity to the hydrophobic groups (Blundell *et al.* 1972; Peking Insulin Structure Research Group 1974).

However, a more detailed study of contact areas in protein–protein interactions (Chothia & Janin 1975) has shown that this generalization also is too simple and that at least some interfaces contain polar groups which interact specifically across the intermolecular boundary. Although the hydrophobic effect may be the major driving force for aggregation, it seems that specific interactions, both polar and non-polar, are important at the local level.

The surfaces of the building sites of many enzymes show a preponderance of hydrophobic groups. Here again, however, we often find some hydrophilic groups which seem possibly relevant to the specificity of the site. For example, trypsin, which hydrolyses the peptide bond next to a basic side chain such as in lysine or arginine, has a convenient negatively charged aspartic acid group situated deeply within the binding pocket. Thus, it is tempting to suggest that the driving force for many enzyme-substrate binding interactions is essentially hydrophobic in nature, although specific charged or hydrogen-bonded interactions are necessary for the geometrical specificity of the binding. Thus, the functioning of the enzyme seems to depend quite critically upon water *not* being where it is not wanted, with solvent exclusion often resulting in a very large binding energy of an enzyme to its substrate, as in trypsin (Rühlmann *et al.* 1973).

These ideas can be quantified by making actual calculations of the accessibility of solvent to various groups. Lee & Richards (1971) put numbers to the accessibility of various groups in ribonuclease-S, lysozyme and myoglobin. Their results show that a significant fraction

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

(40–50 %) of the surface area of these three molecules is non-polar. Their methods also locate internal holes large enough to accommodate water. Recently these concepts have been used in studies of protein folding (Chothia 1975) and protein interactions (Chothia & Janin 1975). In their simulation of the folding of pancreatic trypsin inhibitor, Levitt & Warshel (1975) included an approximate hydrophobic energy term which varied with the exposure of the side chain.

Thus examination of the protein part of a crystal structure in terms of the positions of various groups in relation to the solvent, gives us considerable information concerning the end result of interactions with water. In particular, an important rôle of water in protein folding, protein-substrate and protein–protein interactions is demonstrated, albeit in a non-specific way.

(*b*) *Direct location of water molecule positions*

With diffraction data of sufficiently high resolution (*ca.* 0.2 nm) we should be able to locate directly those positions which water molecules occupy (or partly occupy) in a protein crystal. There are, however, problems in interpretation, and certain reservations should therefore be made.

(i) High resolution data is essential if we are to have any confidence in assigning watermolecules to individual electron density peaks. In solving small crystal structures, the resolution is normally adequate to reveal individual atoms. For large molecules such as proteins, our electron density maps are generally some way from atomic resolution and we have to fit the polypeptide chains to them as best we can, trying to correlate particular side chains (if we know the sequence) with particular regions of electron density. Fortunately, known stereochemical limitations on the polypeptide chain reduce the possibility of erroneous fitting of the molecule to the map. The result satisfies what we think a protein should look like from other evidence.

The non-crystalline organization of the water, however, presents a more difficult problem. We do have some expectations from work on small molecule hydrates, from quantum mechanical calculations on small groups of water molecules, and even from work on bulk water, especially simulation calculations (although the rigid rotor model used, with no allowance made for non-additive contributions to the pontential function, restricts the useful information on local arrangements obtainable by Monte Carlo and molecular dynamics techniques). The resulting stereochemical restrictions are, however, insufficient to enable us to interpret with confidence many of the possible water peaks. Compared to the protein chain, there are many more relatively unconstrained degrees of freedom, and much larger areas of uncertainty, when we deal with the solvent. With advances in protein refinement techniques (Watenpaugh, Sieker, Herriott & Jensen 1973; Sayre 1974, 1975; Cutfield *et al.* 1975; Carlisle *et al.* 1975) this situation is improving, and near-atomic resolution can be achieved in favourable cases.

(ii) The X-ray scattering power of an oxygen atom is relatively low compared to that of a group of atoms in the denser protein molecule; that of hydrogen is effectively non-existent. Thus, in trying to locate water at lower than atomic resolution, we are looking for a small effect in the company of much larger ones. In neutron diffraction, the scattering powers are very much more favourable to the water problem. Those of hydrogen and dueterium are comparable with those of the other main protein constituents, C, N, and O, and moreover the scattering length of hydrogen is negative. Thus, other things being equal, neutron diffraction should be far more capable of locating water. Unfortunately, only one neutron diffraction study of a protein has so far been undertaken. Flux problems mean we require large crystals

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

(~ several mm) with which few proteins seem willing to oblige, and a nuclear reaction is expensive. A neutron diffraction study of myoglobin at 0.2 nm resolution (Schoenborn 1969, 1972) shows that good solvent information can in fact be obtained. Work in progress on triclinic lysozyme will hopefully produce further high-quality solvent data.

(iii) By no means every electron density peak in a map is real. Moreover, because of the relatively low electron density in the solvent region, this is where we would expect errors to be most prominent (for example, from series termination effects). Thus we must beware of assigning water molecules to spurious electron density. Within the protein molecule, such misleading peaks tend to be masked by the existing relatively high electron density, and can in any case often be rejected on stereochemical grounds. Our knowledge of water, the lack of adequate stereochemical constraints, and the problem of solvent disorder (the apparent partial occupancy of nearby possible sites – see §4*b*(iii) below) make it difficult to reject peaks on any sure ground.

The availability of two independent maps, preferably phased using different heavy atom derivatives, is useful in tackling this problem. One such case is that of insulin, on which preparatory work has been done in comparing possible solvent peaks between the Oxford 0.19 nm and the Peking 0.25 nm maps (Finney & Timmins, unpublished work). A gratifying number of common solvent peaks were identified on the two maps.

(iv) The surface of the protein molecule itself is sometimes poorly defined, presumably because of the flexibility of the polypeptide chain in certain regions of the molecule. If the protein molecule itself is difficult to see, we can expect to conclude little about the nearby solvent.

(v) Proteins are rarely crystallized from pure water. Thus we often cannot distinguish between salt ions and water, or water and ethanol, although the surroundings may sometimes imply one component rather than another. Moreover, the presence of salts may affect the local solvent organization. Thus, some of our solvent information may be specific to a particular salt concentration, and it may be dangerous to draw general conclusions from it.

A reasonable number of crystalline protein structures have now been solved to sufficiently high resolution to enable water locations to be seen. Thus, desite the problems and reservations mentioned above, it seems worthwhile to consider the available information on water and how it is relevant to the stability and functioning of the protein. Where inhibitor studies have also been made at high resolution, speculations on mechanism may be a little more valid. However, as few enzyme mechanisms are sorted out in detail to universal satisfaction, any conclusions as to involvement of water in a mechanism must be taken as highly tentative.

Table 1 lists a number of proteins that have been studied at high resolution and have yielded some specific solvent information. The treatment accorded different molecules is far from uniform, and so the information obtainable from these studies is very variable. For example, some refinements involve the tentative solvent locations; others do not. The range of proteins is reasonably wide, including enzymes, hormones, and transport proteins.

The location of water molecules in these protein crystals supports in general terms the conventional wisdom (§2). In some proteins – particularly the larger ones – there are 10–20 internal molecules which are often inaccessible to the surface. These molecules might be thought of as an integral part of the protein itself. We also find water molecules close to particular polar and charged surface sites. The solvent region between molecules sometimes shows a non-uniform electron density distribution, suggesting some 'ordering' in the hydration shell. In a

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

few cases, mainly the smaller molecules such as rubredoxin, ferredoxin, and *Chromatium* high potential iron protein (HiPIP), refinement procedures have produced coordinates of possible partially occupied water sites in this region.

The kinds of water observed can be conveniently divided into 'internal' and 'surface' molecules, and those found in the surrounding shell. This classification, though not rigid, is convenient.

TABLE 1. SELECTION OF PROTEIN STRUCTURE DETERMINATIONS GIVING SOLVENT INFORMATION

Bence–Jones protein	Epp *et al.* 1975
carbonic anhydrase	Liljas *et al.* 1972
	Kannan *et al.* 1975
carboxypeptidase A	Quiocho & Lipscomb 1971
Chromatium high potential iron protein	Carter *et al.* 1974*a*
α-chymotrypsin	Birktoft & Blow 1972
chymotrypsinogen	Wright 1973
concanavalin A	Becker *et al.* 1975
ferredoxin	Adman, Sieker & Jensen 1973, 1975
flavodoxin	Watenpaugh, Sieker & Jensen 1973*a*
insulin	Blundell, Dodson, Hodgkin & Mercola 1972
liver alcohol dehydrogenase	Brandén *et al.* 1973
	Eklund *et al.* 1975
lysozyme	Blake *et al.* 1967
myoglobin	Schoenborn 1969, 1972
papain	Drenth *et al.* 1970
	Drenth, Jasonius, Koekoek & Wolthers 1971
	Berendsen 1975
pancreatic trypsin inhibitor	Deisenhofer & Steigemann 1975
parvalbumin	Moews & Kretsinger 1975
ribonuclease A	Carlisle *et al.* 1974
	Carlisle *et al.* 1975
ribonuclease S	Richards & Wyckoff 1970
	Wyckoff *et al.* 1970
rubredoxin	Watenpaugh *et al.* 1973*b*
staphylococcal nuclease	Arnone *et al.* 1971
	Cotton & Hazen 1971
	Cotton *et al.* 1972
subtilisin	Hol 1971
	Drenth, Hol, Jasonius & Koekoek 1972
thermolysin	Matthews & Weaver 1974
triose phosphate isomerase	Banner *et al.* 1975*a*, *b*
trypsin	Rühlmann *et al.* 1973
	Stroud, Kay & Dickerson 1974
	Krieger, Kay & Stroud 1974
	Bode, Schwager & Huber 1975
	Bode, & Schwager 1975*a*
	Bode, Schwager & Fehlhammer 1975

Note: This list is not exhaustive, and the amount of solvent information varies considerably between different proteins.

(i) *Internal water*

Water as a metal ligand. In metal-containing proteins, water often occurs as one or more ligands of the metal or metals. The zinc ion of carboxypeptidase A – a proteolytic enzyme catalysing the hydrolysis of carboxy-terminal peptide bonds – binds four ligands in distorted tetrahedral coordination (figure 1). One liquid is a water molecule (or OH^- ion), the others being a glu-

tamic acid and two histidine side chains. The water ligand is probably accessible to the surface through other solvent molecules. The high dielectric constant of this channel may help dissipate the charge on the zinc ion over a larger volume, thus helping to stabilize the structure. On substrate binding, the water ligand is displaced by the carbonyl oxygen of the susceptible peptide bond (see figure 12). The environment is now hydrophobic, and proposed mechanisms suggest the apolar surroundings enhance the polarizing effect of the zinc ion on the C=O of the substrate (Quiocho & Lipscomb 1971).

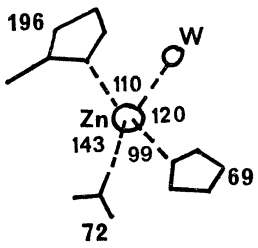

FIGURE 1. The distorted tetrahedral geometry around the Zn^{2+} ion in carboxypeptidase A. His-196, His-69 angle = 86°; Glu-72, W angle = 99°.

Single water molecules are found bound to metal ions in staphylococcal nuclease (Ca) (Arnone *et al.* 1971), carbonic anhydrase (Zn) (Liljas *et al.* 1972; Kannan *et al.* 1975), and liver alcohol dehydrogenase (Zn) (Eklund *et al.* 1975). N.m.r. techniques have demonstrated the presence of metal–water ligands in several other enzymes, including pyruvate kinase, pyruvate carboxylase, fumarase and acotinase (Mildvan 1974). More than one metal–water ligand are sometimes found. The calcium coordination sphere of bovine β-trypsin is an almost regular octahedron, with two water molecules forming bridges to other side chains and water molecules (figure 2*a*) (Bode & Schwager 1975*b*). Concanavalin A binds manganese and calcium ions in adjoining sites; their coordination polyhedra can be thought of as two edge-sharing octahedra (Becker *et al.* 1975). Two water molecules act as bridges from the calcium ion to side-chain oxygens, while of the two waters in the manganse coordination sphere, one bridges to a side-chain oxygen and hydroxyl, while the second is accessible to the solvent (figure 2*b*). In the insulin hexamer, the two zinc ions are coordinated to several solvent molecules, together with an imidazole nitrogen, in apparently octahedral coordination. The solvent networks associated with the two zincs are different (Blundell *et al.* 1972).

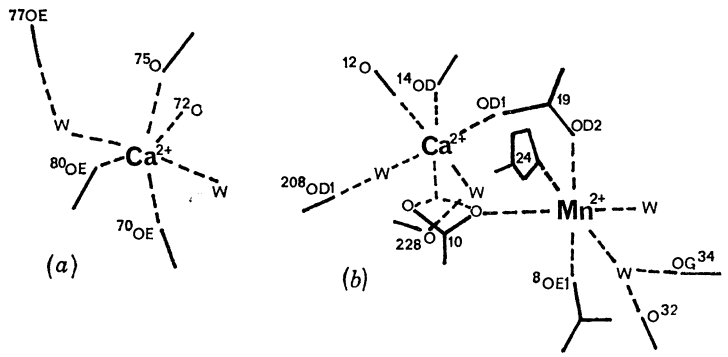

FIGURE 2. (*a*) The calcium binding site in bovine β-trypsin (after Bode & Schwager (1975*b*)). (*b*) The calcium and manganese binding sites in concanavalin A. The neighbourhoods can be thought of as two edge-sharing octahedra. After Becker *et al.* (1975). W indicates a solvent molecule.

It was suggested above that, in carboxypeptidase A, the water dipole may help to stabilize the native structure by counteracting the positive charge on the zinc ion, with other water molecules in the cleft connecting the ion to the surface of the molecule, helping to dissipate the charge concentration inside the protein. A similar situation seems to occur in other metal-containing proteins. One of the water molecules in the calcium binding site in bovine β-trypsin (figure 2*a*) connects the ion to a nearby anionic side chain (Glu-77), presumably partially discharging the positive charge on the calcium (Bode & Schwager 1975*b*). Thermolysin contains four calcium binding sites. The substitution of Eu^{3+} for Ca^{2+} in one of these sites interestingly increases the number of metal–water ligands from two to three (Matthews & Weaver 1974). In addition to this possible charge-spreading rôle, a strong hydrogen bonding of the metal to other parts of the molecule through a water bridge, as occurs in bovine β-trypsin (figure 2*a*), may help to maintain the stability of specific local geometrical arrangements (Bode & Schwager 1975*b*).

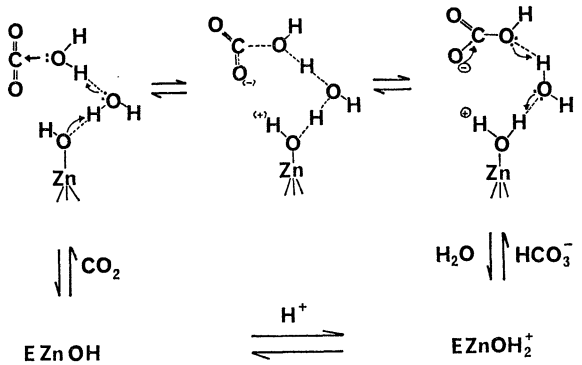

FIGURE 3. Carbonic anhydrase mechanism as proposed by Khalifah (1971). Note the linked water network facilitates proton transfer.

A direct catalytic rôle can be postulated for the water ligand in several metal proteins. As might be expected, this is particularly so for processes involving hydrolysis. Carbonic anhydrase, which catalyses the hydration of carbon dioxide and the dehydration of bicarbonate (or carbonic acid), is one example where several water-involved mechanisms have been proposed. A basic group of the enzyme appears necessary for rapid water exchange and the rapid hydration of CO_2 (Mildvan 1974). Studies of the pH dependence of the reaction indicate that the activity depends on a group in the enzyme with pK_a of 7. The basic form is required for CO_2 hydration, while the acid form takes part in the dehydration of HCO_3^-. The simplest proposal is that the zinc-bound water is the relevant group, although an ionizable imidazole group on an amino acid side chain in the active site is also a possibility. It has been suggested that the bound OH^- is by itself insufficiently reactive, but that it might be activated by a nearby imidazole (Wang 1968). Chemical modification studies, however, suggest that the basic side chains are not so involved (Whitney, Nyman & Malmström 1967).

A cooperative push-pull mechanism proposed by Khalifah (1971) involves the zinc-bound water as the origin of both donor and acceptor functions, using water bridges to facilitate rapid proton transfer along hydrogen bonds (figure 3). This would minimize the development of uncompensated charge during the reaction. Such a mechanism requires rapid proton transfer between the active site and the solvent, in order to allow the enzyme to undergo reverse protonation to return to its active form sufficiently rapidly to explain the very high turnover

rate of the process. In the early crystallographic studies of carbonic anhydrase, a path of apparently strongly bound water molecules connecting the active site to the surface was found which might have served this function (Kannan *et al.* 1971). Further work, however, showed this assignment to be erroneous, illustrating the danger involved in solvent assignment (Kannan *et al.* 1975; Notstrand, Waara & Kanna 1975). Nevertheless, such water networks may possibly serve as fast proton transfer paths in other enzyme catalyses. A different possible catalytic mechanism which would also involve water bridges has been pointed out by Kaiser & Lo (1969).

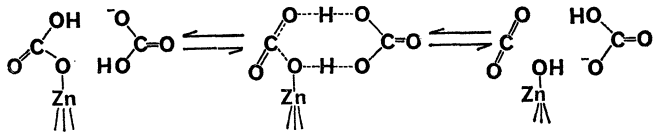

FIGURE 4. Carbonic anhydrase mechanism as proposed by Yeagle *et al.* (1975). The strongly bound bicarbonate facilitates proton transfer.

A further alternative mechanism has been proposed on the basis of ^{13}C n.m.r. studies (Yeagle, Lochmüller & Henkens 1975). Two bicarbonate binding sites are found. One is loosely bound directly to the metal centre and is the substrate binding site. The other is tightly bound in the outer coordination sphere of the metal, and is proposed to facilitate proton transfer from the zinc-bound OH. The OH^- on the metal attacks the CO_2, but with proton transfer occurring in a concerted manner via the tightly bound non-inhibitory bicarbonate (figure 4).

Although the carbonic anhydrase mechanism is still not known with certainty, these proposals illustrate the potential involvement of water directly as the attacking group. They also suggest the possible relevance of linked water bridges as proton transfer paths, both in the active site region itself, and between the active site and the solvent.

There are other instances of possible direct water involvement. The metal ion in pyruvate carboxylase and transcarboxylase may promote the acidity of a bound water which then protonates the carbonyl oxygen of pyruvate (Mildvan 1974). In yeast enolase, the divalent ion may activate water for nucleophilic attack, with perhaps the phosphate of the substrate acting as a general base, assisting the metal in deprotonating the water (Novak, Mildvan & Kenyon 1973). Water has been proposed as the attacking nucleophile in the mechanism of the nucleotidyl transfer enzyme, staphylococcal nuclease. The calcium ion may increase the nucleophilicity of a bound water molecule, and the metal-bound water is also in a position to displace the poorer of the two possible leaving groups (Cotton & Hazen 1971; Dunn, di Bello & Anfinsen 1973; Mildvan 1974). In acotinase, a water ligand substitution on the enzyme-bound iron provides the driving force for a conformational change required by the enzyme (Glusker 1968).

Thus, metal-bound water is directly implicated in certain enzyme mechanisms, often with the metal promoting the water's activity. The examples are suggestive, but still speculative.

Near internal charged groups. X-ray diffraction studies have shown that in general, charged side chains stick out into the solvent, or at least are on the surface of the protein. This generalization is not universally true, however, and charged side chains sometimes occur internally or partly internally. Some of these internal charged groups form salt bridges with other charged groups, as between Glu-292 and Arg-272 in carboxypeptidase A (Hartsuck & Lipscomb 1971). In other examples, we find unpaired charged groups which are buried and inaccessible to solvent.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Such buried, uncompensated charged groups would be expected to confer some degree of instability on the molecule. Yet in the serine proteases, the buried Asp-102, thought to be charged at active pH, is an important part of the proposed catalytic mechanism, supplying an electron to the hydrogen-bonded charge-relay system comprising Asp-102, His-57 and Ser-195 (Birktoft, Blow, Henderson & Steitz 1970). (See also Hunkapillar, Smallcomb, Whitaker & Richards 1973, and Rogers & Bruice 1974, for critical discussion of the charge-relay mechanism.) The structure analysis of tosly-α-chymotrypsin located two water molecules forming part of an extended hydrogen-bonded network close to, though isolated from, Asp-102 (figure 5) (Birktoft & Blow 1972). These internal waters could be important in increasing the local dielectric constant, helping spread the buried charge through a larger volume, and thus stabilizing the structure around the buried group. These two water positions are also found in bovine β-trypsin (Bode & Schwager 1975*a*). It is also possible that they help to rigidly position the carboxylate group of Asp-102 in a position favourable to an effective charge-relay system.

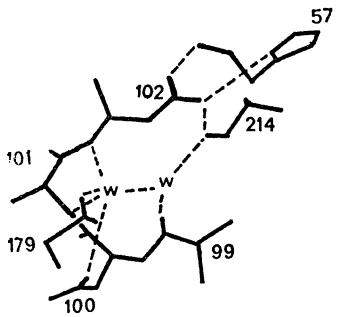

FIGURE 5. The two internal water molecules found in α-chymotrypsin, close to the buried, charged Asp-102. The high dielectric constant of this region may be important in maintaining stability of the molecule around the acid group. (After Birktoft & Blow 1972.)

It is perhaps of interest that these two waters become trapped during movement that occurs on activation of chymotrypsinogen (Wright 1973). Carboxypeptidase A has a direct water bridge mediating the buried, charged side groups of Asp-104 and Arg-59. Glu-108 interestingly sits next to a water, which is presumed to be an H_3O^+ ion hydrogen-bonded to two backbone carbonyl oxygens. The H_3O^+ presumably neutralizes the buried charge on the glutamic acid (Hartsuck & Lipscomb 1971).

Another likely example of the stabilization of internal charged groups by 'charge-spreading' is seen in the hexamer of insulin (Blundell *et al.* 1972). The hexamer is formed by aggregation of three dimers. Each aggregation step brings together pairs of negatively charged γ-carboxyl groups (residue 13 of the B chain), leading to a configuration of six in the hexamer. All these charged groups are approximately coplanar and within 0.5 nm of the threefold axis. Such a charge concentration we would expect to limit the degree of association in alkaline media (pH 8). However, above the plane of these groups, there are three loosely-bound waters arranged symmetrically about the trigonal axis, while below the plane we fine two sets of symmetrical sites (see Fig. 23 of Blundell *et al.* 1972). This arrangement of waters may well be important in shielding the apparently unfavourable charge distribution at pH values above the pK of the carboxylate groups.

Internal polar groups. In several of the larger protein structures so far solved at high resolution, many fully occupied water sites have been located, attached to the protein by hydrogen bonds

to polar groups. The kinds of organization found are varied. Some water molecules occur singly, some in pairs, some in extended networks. Most of them make two or three hydrogen bonds to polar groups or other water molecules, although some do make four. Many bind to the NH and CO groups of the main chain backbone, although some are attached to polar side chains.

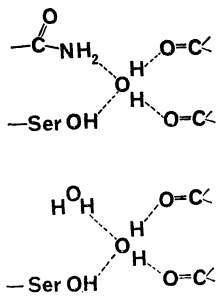

FIGURE 6. Two internal, 4-coordinated water molecules in carboxypeptidase A.

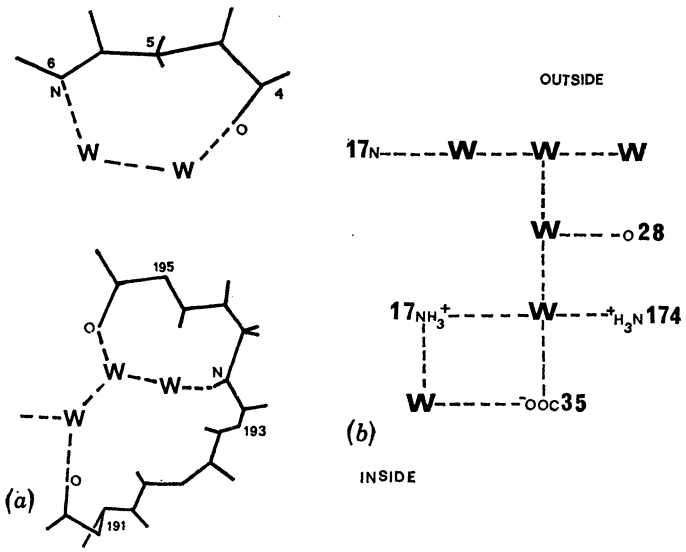

FIGURE 7. (*a*) Water bridges connecting main chain groups which are close together on the papain chain. (*b*) A water channel leading to the surface in papain. Note the water-mediated salt bridges. (After Berendsen 1975*a*.)

For example, two water molecules are found close together in lysozyme, hydrogen-bonded to glutamine and serine side chains, together with main chain NH and CO groups of other residues (Blake *et al.* 1967). Carboxypeptidase A, a much larger protein with over 300 amino acids against the 129 of lysozyme, has ten internally trapped waters, making hydrogen bonds to nearby polar (and ionic – see above, p. 16) groups (Hartsuck & Lipscomb 1971). Two of the ten make four hydrogen bonds, each tetrahedrally arranged with respect to two main chain carbonyls, a serine OH, and either a glutamine amide or another water molecule (figure 6). Seven other isolated water molecules make three hydrogen bonds each, and the remaining one, only two. Participating groups are backbone imino groups and side-chain hydroxyl, amide, and carbonyl groups. Even in the small 85 residue *Chromatium* high potential iron protein, we find one internal water molecule, hydrogen bonded to a backbone NH, a backbone CO, and a glutamine side-chain CO.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

This basic pattern of 2–4 hydrogen bonds with a variety of polar groups is borne out in other, larger molecules, where more internal waters are found. In papain, hydrogen bonding to the peptide backbone NH and CO are most frequent, with side-chain associations less so (Berendsen 1975*a*). A considerable amount of water–water bonding is observed. As in carboxypeptidase, most water molecules make two or three hydrogen bonds, with only a few making four. Groups close together on the peptide chain are often connected via water bridges, sometimes of one, two, or even three molecules (figure 7*a*). In other cases, distant regions of the chain are connected via such bridges. An interesting network of waters communicates to the surface of the molecule from the inside, with water-bridging of charged lysine and glutamic acid side chains (figure 7*b*).

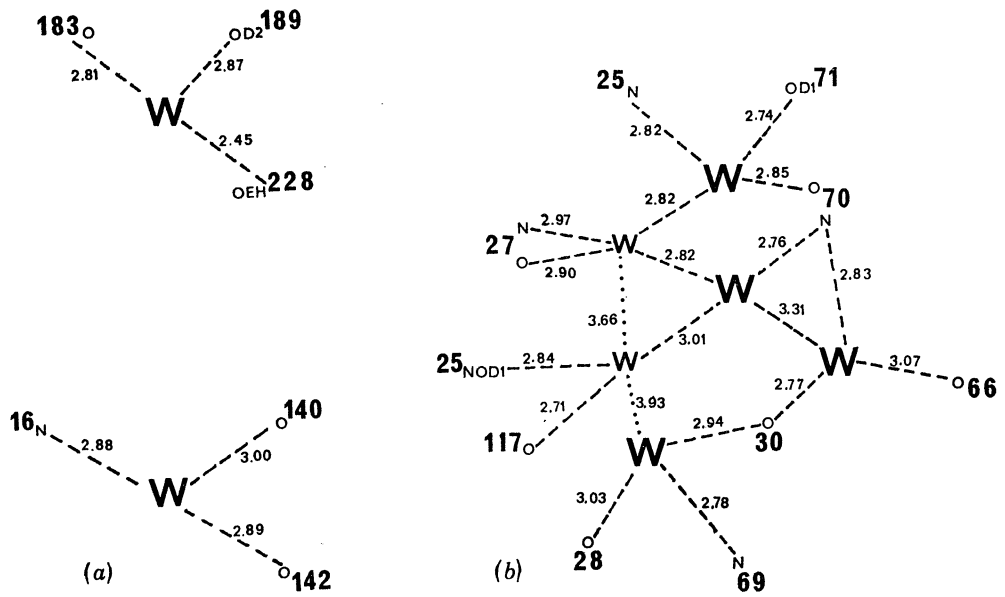

FIGURE 8. (*a*) Two isolated internal water molecules in bovine β-trypsin. (*b*) An extensive internal solvent network in bovine β-trypsin. Large W's indicate a solvent molecule is found in a similar position in α-chymotrypsin.

Figure 8*a* shows two typical internal solvent neighbourhoods in trypsin (Bode & Schwager 1975*a*). Waters are found in similar relative locations in α-chymotrypsin, although the hydrogen bonding to neighbouring groups is sometimes different in the two proteins (Birktoft & Blow 1972). Figure 8*b* shows an extensive internal water network in trypsin. The large W's identify solvent locations common to both β-trypsin and α-chymotrypsin.

It is tempting to suggest that these internal water bridges are performing a stabilization function, at least locally. Although it is generally accepted that the largest contribution to the stabilization energy of a protein molecule is the non-specific hydrophobic effect, the high degree of 'mating-up' of polar groups – both on the backbone and side chains – suggests this contribution to the folded structure is not negligible. Lumry & Biltonen (1969) emphasize the stabilizing rôle of the near-saturation of possible internal hydrogen bonds, and suggest that changing an amino acid sequence so that a few internal hydrogen bonds could no longer be made would lead to a thermodynamically unstable structure. Failure to form hydrogen bonds introduces instability of the folded structure. On the basis of their accessibility calculations, Lee & Richards (1971) suggested that the specific saturation of internal polar groups

should be a severe restriction on possible folded structures. Unpaired internal polar groups should be a destabilizing factor. If this is the case, the internal water molecule bridges might well perform a similar function. Where two or more polar groups are unfavourably placed for making direct hydrogen bonds, water molecules may bridge the gap and thus enhance at least the local stability, saturating potential hydrogen bonds that would otherwise be unsatisfied. For example, in α-chymotrypsin, many internal water molecules were found to bridge across those hydrophilic groups which were external to the two hydrophobic cylinders into which the molecule folded (Birktoft & Blow 1972). Again in the context of the serine proteases, Bode & Schwager (1975*a*) have suggested that in larger molecules, the inclusion of some rigidly-bound water molecules is necessary in order to build up a stable structure. They also suggest this may be especially relevant to proteins like trypsin where there is little organization into secondary structures of long helices or well-aligned pleated sheets.

The extent to which this stability argument should be taken is presently uncertain. In their computer simulation of pancreatic trypsin inhibitor folding, Levitt & Warschel (1975) used very much averaged potential functions, including an averaged hydrophogic energy, and neglected hydrogen bonding. Although it is not clear how close their folded structure is to the native conformation, or whether there are any very high energy barriers to be crossed in passing from one structure to the other, the two structures do appear to be 'topologically' similar. The regions in which hydrogen bonding is extensive in the native structure are somewhat loose in the simulated folded conformation, but the overall structure is still there. This suggests the overall folding can indeed be considered in these apparently crude terms, but that as we approach the folded structure, we must begin to 'switch on' the more specific interactions, to 'shuffle' the molecule into its local energy minimum with the help of hydrogen bonds. If direct hydrogen bonds between polar groups are significant to the stability of the protein, then so also are the water bridges. They may only be relevant to maintain local structure – such as the rigid positioning of the carboxylate group of Asp-102 in trypsin (see above, p. 16) – but as enzymatic activity may well depend strongly on the conformation of critical groups, these interactions are presumably vital, although their fractional contribution to the energy of stabilization *with respect to the extended chain structure* may be relatively small. Their contribution to stability close to the active conformation may well be very significant.

(ii) *Surface bound water*

Many water molecules are attached to polar and hydrophilic groups on the surface of the molecule. Some of these are hydrogen-bonded to one group only, and presumably connect with the surrounding solvent, acting as 'anchoring points' for the perturbed solvent shell. Other molecules make more than one hydrogen bond to the same protein molecule, others bridge across to neighborouring molecules in the crystal, and others are found in active site clefts and other channels running into the molecule.

No clear pattern emerges concerning preference for particular kinds of water-attachment site. Of the hydrogen bonds to surface water molecules in concanavalin A, 20% are to main chain amides, 51% to main chain carbonyls, and 29% to side chains (Becker *et al.* 1975). Trypsin shows a similar pattern with 25% to main chain amides, 39% to main chain carbonyls, and 36% to side chains (Bode & Schwager 1975*a*). The neutron diffraction work on myoglobin, however, shows only 10% to main chain amides, 36% to main chain carbonyls, and 55% to side chains (Schoenborn 1972). The myoglobin side-chain figures may be inflated by the

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

possibility of many of the 'solvent' molecules bound to charged groups being salt ions. A preference for main chain carbonyls over amides is, however, implied by these figures, supporting suggestions that the hydrogen bond to CO may be stronger than that to NH.

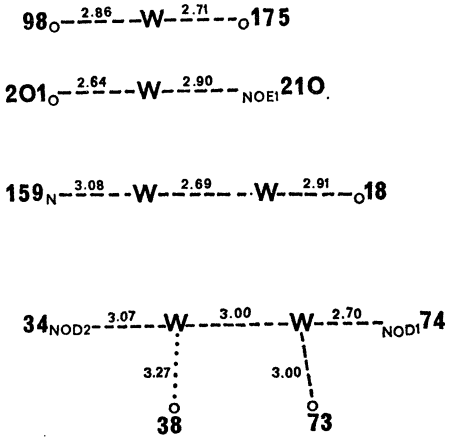

FIGURE 9. Typical 'simple' surface bridges found in bovine β-trypsin.

Double or multiple attachments to the same protein molecule. These water molecules are presumably important in maintaining the local structure of the molecule, in a way similar to that discussed in §4*b*(i) above. In α-chymotrpysin, such molecules are found in surface crevices; a similarly placed water in trypsin is expelled on inhibitor binding (Krieger *et al.* 1974). In bovine β-trypsin, 23 external waters are located. As figure 9 shows, some of these are single bridges across two polar atoms, some are parts of double bridges linked by one or more hydrogen bonds to the surface. Although many of the internal water molecules in trypsin can be identified with ones found in α-chymotrypsin, there are significant differences between the surface molecule locations in the two proteins. Even more complex multiple bridges can be found, as in myoglobin (figure 10).

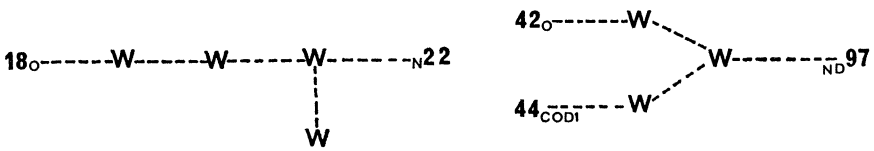

FIGURE 10. Two myoglobin surface bridges.

In *Chromatium* high potential iron protein, a small oxidation-reduction protein of 85 residues, surface water locations have been invoked to explain the large number of hairpin turns found at the protein–water interface (Carter *et al.* 1974*a*). Thirty-one of the 85 residues are near these turns, and the backbone atoms of these residues bind 28 out of the 50 water molecules bound by the main chain. It is suggested that the hairpin turn is a good way of introducing a structure, one end of which is complementary to water, but which connects chain segments that will eventually be inside the molecule. These turns create most of the protein interface.

Thus, surface water bridges are found extensively, and presumably assist in local structure stabilization. Upon any conformational changes of the protein, these bridges are liable to experience strain, and thus must be considered a part of the mechanical system that is the protein. We do not know how far these bridges may be artefacts of the crystalline state; when the

molecule is in solution, the removal of lattice forces may reduce the stability of some bridges, perhaps transforming them into single point attachments.

It is noteworthy that no extensive hydration structures have been found around hydrophobic groups, although this is hardly surprising considering the limitation of crystal diffraction techniques.

Bridges between subunits or neighbouring molecules. Many examples of intermolecular bridges are now known. The detailed myoglobin neutron diffraction study shows single, double, and

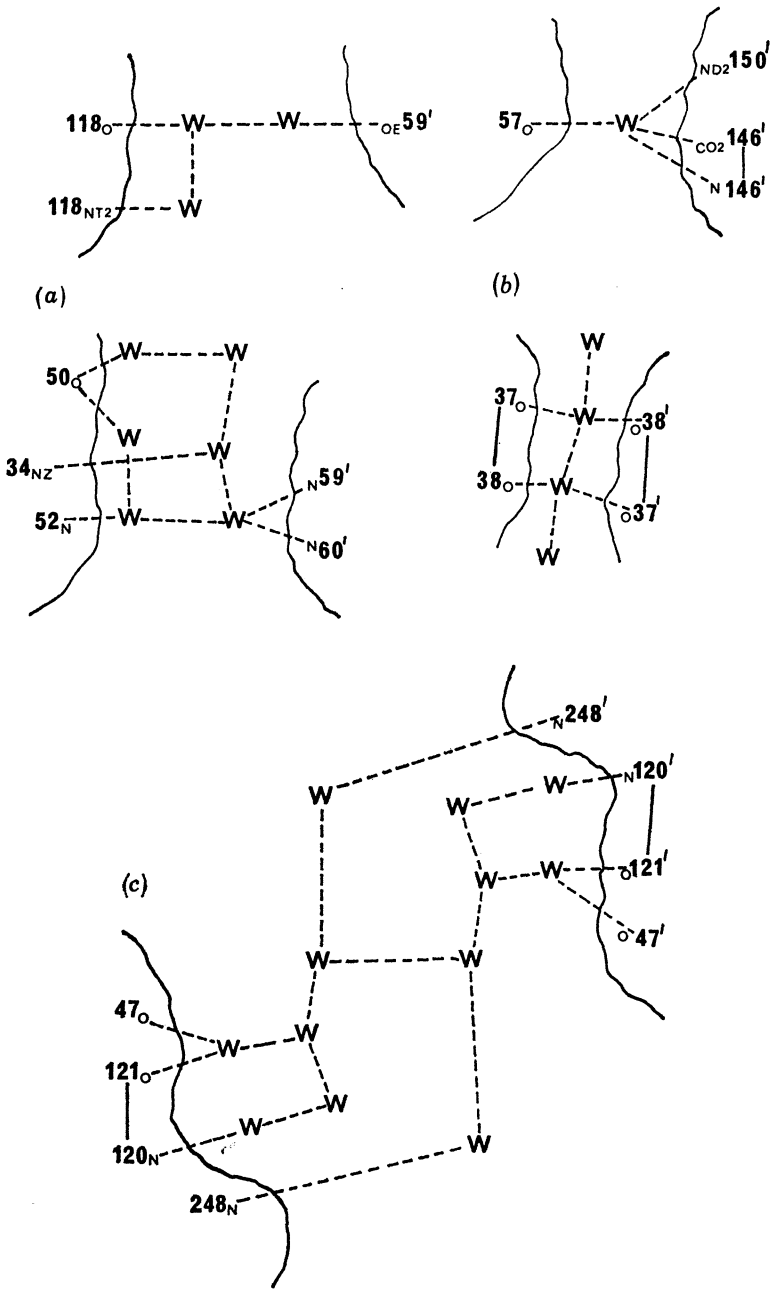

FIGURE 11. (*a*) Two intermolecular bridges in myoglobin. (*b*) and (*c*) Simple and complex intermolecular bridges in α-chymotrypsin. The larger two networks have approximately twofold symmetry 'induced' by the symmetry of the crystal. As such, they would not remain thus near the protein molecule in solution.

complex bridges between neighbouring molecules (Schoenborn 1972), while triose phosphate isomerase has single water bridges between arginine and glutamine side chains in each of the subunits related by a diad axis (Banner *et al.* 1975*b*). α-Chymotrypsin shows single, double and more complex intermolecular bridges around the contact region (Birktoft & Blow 1972), while eight intermolecular water bridges have been found in *Chromatium* high potential iron protein (Carter *et al.* 1974*a*). Figure 11 shows the types of intermolecular bridges found.

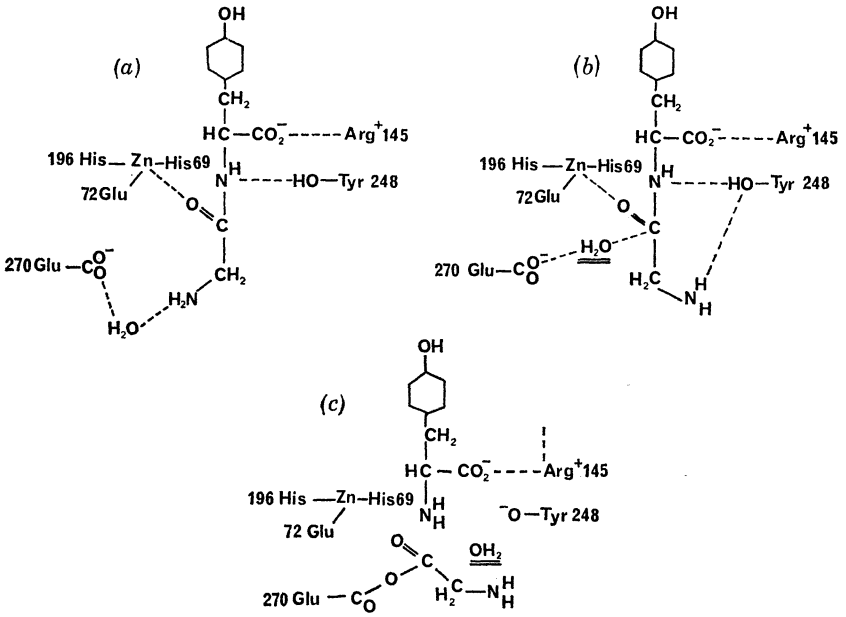

FIGURE 12. (*a*) Binding of Gly-Tyr to carboxypeptidase A. (*b*) General base mechanism for carboxypeptidase A. (*c*) Nucleophilic pathway for carboxypeptidase A. (After Quiocho & Lipscomb 1971.)

The function of these water molecules is presumably in partial stabilization of intermolecular contacts, be they in multimeric proteins such as triose phosphate isomerase or insulin, or in assisting stable molecular aggregation in the crystalline state. As such, they can be discussed in terms similar to those used in discussing the stabilizing rôle of internal water molecules (§4*b*(i) above). As discussed in §4*a* above, the major driving force for aggregation is likely to be hydrophobic, although specific interactions between complementary groups may occur across the molecular interface. As with internal hydrogen bonding within the protein, solvent bridges might be expected to fill in where direct hydrogen bonding is not feasible.

Water in active site clefts, and behaviour on binding and catalysis. We saw in §4*a* that binding sites tend to be largely hydrophobic with a potentiality for specific interactions to explain substrate specificity. For example, the bottom of the trypsin pocket contains an aspartate group, which is thought might be responsible for the enzyme's specificity toward the charged side groups arginine and lysine (Krieger *et al.* 1974; Bode & Schwager 1975*a*).

Crystallographic studies have shown that in the native enzyme, binding pockets often contain ordered or partially ordered water molecules. In carboxypeptidase A, some, but not all, of these waters are displaced on binding glycyl-L-tyrosine, a poor substrate for the enzyme; the end of the pocket is thereby converted into a hydrophobic region. The zinc-bound water is displaced by the carbonyl oxygen of the susceptible peptide bond. Other binding interactions (figure 12*a*) are a salt link between the C-terminal carboxylate and Arg-145, and a water

bridge between the free amino group of Gly-Tyr and the carboxylate of Gly-270. This solvent bridge is possible only with dipeptides; it may be one reason for the relatively slow rate of cleavage of dipeptides by CPA (Quiocho & Lipscomb 1971).

Two mechanisms have been proposed for the enzyme (Quiocho & Lipscomb 1971). In both, the zinc ion, previously bound to the water which we postulated earlier (§4*b*(i)) helped dissipate the charge through a previously water-occupied region of high dielectric constant, polarizes the substrate carbonyl to render the carbon atom of the group more prone to nucleophilic attack. The expulsion of the water from this region results in a non-polar, low dielectric constant region, probably concentrating the zinc charge of the carbonyl group. Thus, not only does the change in solvent structure contribute to the driving force of the binding, but it also enhances the general acid function of the zinc.

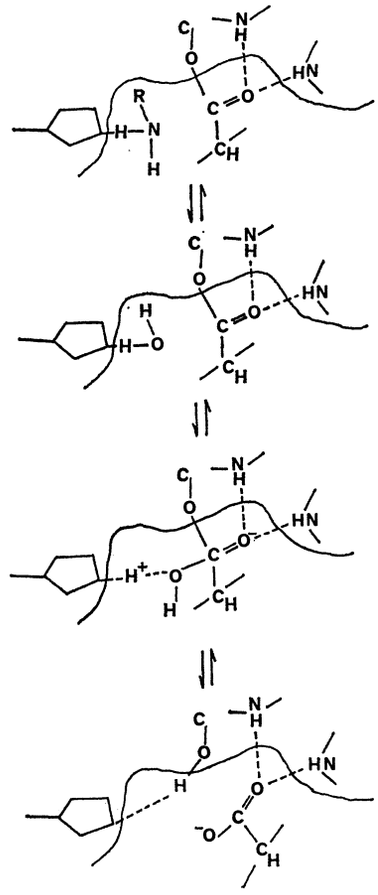

FIGURE 13. Mechanism proposed for the deacylation step of α-chymotrypsin. A water molecule displaces the leaving group, and the reaction proceeds by general base-catalysed removal of the water proton, and the formation and subsequent breakdown of a tetrahedral intermediate. (After Henderson, Wright, Hess & Blow 1972.)

Figures 12(*b*) and (*c*) show the two alternative mechanisms proposed from here on. In both, water may be directly implicated in some part of the process. In the general base path (*b*), the Glu-270 promotes the attack of a lone pair of a water molecule on susceptible carbonyl carbon, the Tyr 48 OH providing a proton for the NH group. In the nucleophilic pathway (*c*), the water is not involved, the Glu-270 directly attacking the carbonyl carbon. The acyl-enzyme

intermediate is then attacked by either a water molecule, which is promoted by the tyrosine O^- obtained after its proton has been transferred to the NH group, or directly by the OH^- formed by regeneration of the tyrosine OH.

On the basis of high resolution crystallographic studies of several complexes of α-chymotrypsin, a detailed stereochemical model for its hydrolytic mechanism has been built up. A specific rôle for water as an attaching nucleophile is included (Henderson 1970).

On binding of virtual substrates, several ordered, or partially ordered water molecules are expelled from the binding pocket. In the case of the tosyl-α-chymotrypsin complex, the tosyl group does not entirely fill the pocket, and two waters remain bound to polar groups at the bottom. The large binding energy is again explained by the large hydrophobic component of the driving force.

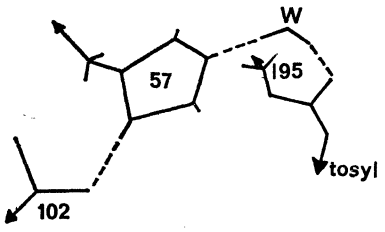

FIGURE 14. The binding of a water in the active site in tosyl-α-chymotrypsin. The water seems well-placed to initiate deacylation of a good substrate. (After Henderson 1970.)

Figure 13 illustrates schematically the deacylation step postulated for the hydrolysis of an amide bond. In the 'native' enzmye, a 'solvent' molecule, originally thought to be water, is hydrogen bonded to both the $N^{\epsilon 2}$ of His-57, and the O^{γ} of Ser-195. This solvent molecule is in the position that is occupied by an amino N or ester O of the peptide or ester substrate to which a proton is transferred in the acylation step. The virtual substrate studies show two waters bonded between the substrate and enzyme. One of these occupies a site close to the original solvent molecule W in the native enzyme, and forms a hydrogen bond between the sulphonyl or carbonyl group of the tosyl or indoleacrylol group and the $N^{\epsilon 2}$ of His-57 (figure 14). It is this water which eliminates the leaving group in the acylation step and which is positioned to attack the carbonyl of the acyl-enzyme. In a 'good' substrate – as against the 'poor' substrates that can be studied crystallographically – the carbonyl oxygen of the substrate must swing out of line (from the virtual substrate position) to allow this water molecule, one proton of which has been partly removed by the charge-relay system acting as a general base, to initiate nucleophilic attack of the carbonyl of the acyl-enzyme. Examination of the geometry of the active site shows the substrate carbonyl may be stabilized in this attack-prone position both by direct hydrogen bonds to other parts of the enzyme molecule (the main chain NH groups of Ser-195 and Gly-193) and a second water molecule located between the carbonyl oxygen and the peptide carbonyl of Phe-41. The deacylation step can then proceed by general base catalysed removal of the water proton, and the formation and subsequent breakdown of a tetrahedral intermediate.

Studies on the chymotrypsin-pancreatic trypsin inhibitor complex (Blow *et al.* 1972) suggest the stability of this inhibited complex is in fact partly due to the inaccessibility of the acyl group to water. The susceptible group is thus shielded from the attacking nucleophile, water.

High resolution studies on the closely related serine protease trypsin and its inhibitor complexes generally support the rôle of the located water in the deacylation step. Studies by Stroud

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

et al. (1974) on benzamidine and di-isopropylfluorophosphate–trypsin complexes at 0.27 nm resolve solvent molecules around the active site and in the specificity binding pocket, as do studies of Huber *et al.* on benzamidine and pancreatic trypsin inhibitor complexes (PTI). The latter studies have recently been taken to very high resolution (0.19–0.15 nm), providing much data on the solvent both in native and complexed structures (Bode & Schwager 1975*a*, *b*; Bode *et al.* 1975; Deisenhofer & Steigemann 1975).

The specificity pocket of active trypsin has three well-ordered water molecules, and probably several more mobile ones, which are displaced on complexing with PTI. Again, it seems likely that the desolvation of the specificity pocket and the active site provides the main entropy contribution stabilizing the complex.

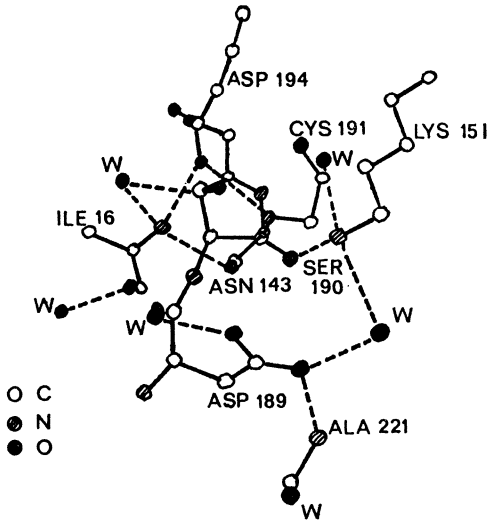

FIGURE 15. Two water molecules in the bottom of the specificity pocket of the pancreatic trypsin inhibitor-trypsin complex. The lysine NH_3^+ connects to the Asp-189 via a water bridge. (After Bode, Schwager & Huber 1975.)

In the trypsin–PTI complex, eight water molecules are found linking the enzyme and inhibitor. Six of these are at the surface of the complex, mostly having three hydrogen bond linkages to main chain amido and carbonyl groups, polar side chains, and surface water molecules. These solvent bridges are all single, and are thought to give added stability to the complex (Bode & Schwager 1975*a*). In addition, two water molecules remain in the bottom of the inhibitor pocket (figure 15). The most interesting of these mediates a salt bridge between lysine 15 of the inhibitor and Asp-189 of the enzyme. Thus, it appears that the specificity of trypsin for charged side groups is *not* explained by a direct salt bridge with Asp-189 as might be expected. Such appears to be not directly possible in this case, and so a water molecule, originably present in the free trypsin, mediates. The second internal water links the same lysine with two main chain carbonyl oxygens.

In benzamidine trypsin, two solvent molecules are located near the catalytic site (Stroud *et al.* 1974; Bode *et al.* 1975). The one close to the corresponding water molecule in chymotrypsin (see above) which is thought to initiate deacylation (Henderson 1970) shows a broad peak, suggesting it is somewhat mobile. This water is too far from the serine O^{γ} to form a hydrogen bond. It is suggested by Bode & Schwager (1975*a*) for trypsin, and by Vandlen & Tulinsky (1973) and Tulinsky & Wright (1973) for chymotrypsin, that it is in fact a sulphate

on with high occupancy. This solvent is in any case displaced by a good substrate on binding, so does not affect the above discussion. It does, however, illustrate the problems involved in making solvent assignments.

In any hydrolysis reaction or its reverse, we might expect water to be involved somewhere in the overall reaction scheme. Carbonic anhydrase and the serine proteases we have already discussed. In lysozyme, the glycosyl enzyme undergoes nucleophilic attack by water from the solvent (Vernon 1967; Phillips 1966), while in papain, deacylation of the postulated acyl-enzyme intermediate probably occurs through water attack (Drenth *et al.* 1970). The binding of 3′ CMP to ribonuclease involves several water bridges (figure 16) (Richards & Wyckoff 1970), and the metal-bound water in liver alcohol dehydrogenase is directly implicated in one proposed mechanism (Eklund *et al.* 1975). Water is also directly implicated in proposed mechanisms for electron-transfer and other proteins, for example flavodoxin (Ludwig *et al.* 1972; Watenpaugh *et al.* 1973*a*), and *Chromatium* high potential iron protein (Carter *et al.* 1974*b*).

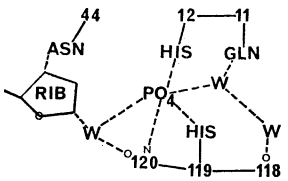

FIGURE 16. Water bridges in the binding of 3′CMP to ribonuclease. (After Richards & Wyckoff 1970.)

(iii) *Water organization in the perturbed solvent shell*

The use of X-ray and neutron diffraction of crystals in elucidating solvent organization in this 'partially ordered' region is limited but not negligible. Provided the 'ordering' effect of the protein surface is sufficient to partially localize water molecules in certain regions of space, we may be able to draw some conclusions about the nature of the instantaneous water networks surrounding the molecule. Imagine a perfect diffraction experiment, in which the electron density map we obtain corresponds one to one with the real electron density of the system (figure 17*a*). That from a completely disordered organization of solvent molecules would be completely flat. If the protein interface exerted a fully ordering effect in which solvent molecules were 'frozen' in fixed positions, then we would see electron density peaks at each such site, with intensities corresponding to full occupation (figure 17*b*). In the real situation of a partly ordered solvent, the molecules' mobility is restricted to particular regions of space by the interplay of the forces exerted by the protein interface and the nature of the solvent molecules themselves. We would see more, weaker peaks, whose intensities corresponded to the *probability* of finding a water molecule at that position at a given time. The electron density map for a disordered situation like this is really a *probability density map*. If we could perform an instantaneous diffraction experiment, we would obtain an instantaneous electron density map which would tell us where all our solvent molecules are at a given instant. An experiment at a different time would give a different map corresponding to a different solvent network. The map we actually obtain over a lengthy diffraction experiment is the time average of all these networks. We want to obtain some information concerning the instantaneous networks, and their temporal relationship to each other.

Undoubtedly, data are available from refinements of several, mainly small, protein structures. The carboxypeptidase maps at 0.2 nm resolution show distinct electron densities for

about 50 % of the solvent region, implying that there is some significant 'ordering' effect of the protein surface (Lipscomb *et al.* 1968). The neutron studies of myglobin located 106 solvent peaks of at least 50 % occupancy (Schoenborn 1972). Refinement of rubredoxin resulted in 104 sites with occupancies down to 0.3, in addition to the 23 surface molecules with 100 % occupancy (Watenpaugh *et al.* 1973*b*). Ferredoxin, after similar treatment, shows 108 variously occupied sites (Adman *et al.* 1975). *Chromatium* high potential iron protein includes 74 solvent positions (Carter *et al.* 1974*a*), while concanavalin A shows in the order of 100 (Becker *et al.* 1975). As refinement procedures and experimental techniques improve, more and more reliable data will no doubt be forthcoming. Of particular interest will be further neutron diffraction studies, and low temperature measurements where disorder may be reduced.

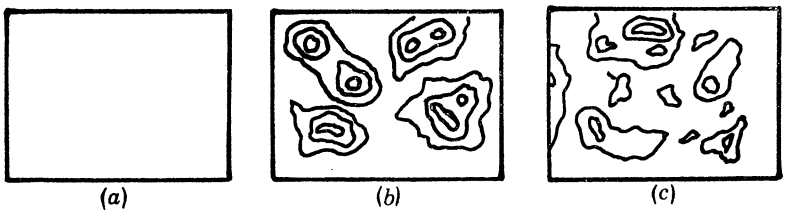

FIGURE 17. Hypothetical electron density maps for (*a*) totally disordered solvent; (*b*) totally ordered solvent; (*c*) partially-ordered solvent.

Even assuming the data are 100 % accurate, rationalizing such sets of partially occupied sites is no easy problem. One reason is our lack of knowledge of the stereochemical constraints we should put on possible water networks. We know already to expect them to be non-crystalline, showing no evidence of extended ice-like structures, and to be approximately tetrahedrally coordinated. Computer simulation calculations of liquid water, and now of simple solutions, are beginning to tell us in more detail what stereochemical constraints to expect, but examination of some of the resulting instantaneous networks suggest the situation is still very complex. In particular, some neighbouring molecules which appear to be relatively well-placed for a strong hydrogen bond, being close to mutual tetrahedral relationships, are, perhaps because of the details of the potential, barely attractive or even slightly repulsive (Finney, Heathman & Nicholas, unpublished work). Thus the detailed molecular organization may be more critically dependent upon the details of the potential function than are macroscopic averaged properties. The rigid rotor, pair-additive potential function may be insufficient to give us reliable localized information.

Nevertheless, it seems worth while to make some initial attempts to eluciate the solvent shell in this way, perhaps starting with smaller molecules. In a study of nucleic acid–mutagen interactions, the crystal structure of adenyl-3′,5′-uridine plus 9-aminoacridine (Seeman, Day & Rich 1975), a solvent-filled cylinder containing 15 water molecules distributed among 21 loci was found. Knowing that two-thirds of these waters were in fully occupied sites led to a description of the water in terms of two reasonable, different networks (Seeman, unpublished work). Some initial work has been done on the vitamin B_{12} coenzyme (Finney & Timmins, unpublished work) with the coordinates for 38 sites available, many of low occupancy (Lenhert 1968). It will be necessary to use other methods, one of which is computer simulation, to look at solvent behaviour at model interfaces. We still know very little about the solvent organization in the neighbourhood of a non-polar group, or its consequences in terms of an effective potential. Computer simulation studies may throw some light on this particular aspect of the

problem. Such simulation studies of themselves involve several unsolved technical problems, for the relatively long time events expected to occur in such heterogeneous systems will require either unrealistic amounts of computer time, or improvements in techniques (Berendsen 1975*b*). Our knowledge of intermolecular potential functions for such systems is also still inadequate, especially for extracting reliable local geometrical information. At the other end of the scale of simplicity and time, model studies will have their uses. Calculations of distances between polar groups on protein surfaces show a slight bias to water-bridging distances (Berendsen 1968; Finney & Timmins, unpublished work). These calculations might be extended in a statistical way from the anchoring points into the immediate surrounding solvent region. There are also possibilities in more theoretical network approaches to non-crystalline systems currently being developed in connection with simpler systems (Finney & Hiley, in preparation).

(*c*) *The problems of the crystal*

Although we may with reasonable confidence say that the essential conformations of many protein molecules are similar to those found in solution, we cannot say the same for the solvent. The close approaches of different molecules in the crystal will of necessity distort the surrounding solvent shell, though it would have less effect on the internal molecules, and those at the surface distant from such contacts.

Even if we do manage to sort out a set of solvent networks for a crystal, we would not expect these networks to have a one to one correspondence with the hydration shell in solution.

It goes without saying, of course, that the *in vivo* surroundings are considerably more complex than this, and the solvent shell is likely to vary with, for example, pH.

Nevertheless, such information would be useful. It could give us some ideas about the driving forces at the protein–solvent interface under relatively well-controlled conditions. It would provide us with reliable spatial information in circumstances in which the nature of the perturbing influence is well characterized. Such could provide a more reliable base for interpreting data on more complex, disordered systems, gained from other experimental techniques (e.g. a protein in solution, a membrane surface).

5. Summary

Proteins are complex systems which involve water in maintaining their stability, controlling their dynamics, and sometimes specifically in their catalytic activity. The nature of the surrounding solvent is of relevance to the diffusion in and out of substrate, product, and coenzyme, and also will affect any necessary transport of water or protons. In many cases, water might be thought of as a coenzyme itself.

Diffraction studies of crystalline proteins shows how water is located within and close to the protein molecule. Water molecules may be bound to metal ions, or to internal charged side chains, aiding stability by spreading uncompensated charge. Where internal hydrogen bonds cannot be made directly by polar groups, water molecules often add stability by bridging across them. Such water bridges may be important in stabilizing local specific geometry close to that required for optimal catalytic activity. Surface bridges may also be structurally significant. Between molecules, and sometimes between enzyme and substrate, water bridges mediate hydrogen bonds that would otherwise be geometrically impossible. They may also be important in binding site specificity, as well as in fixing the detailed structure of a protein–protein inter-

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

face in multi-subunit proteins. The exclusion of water from otherwise hydrophobic regions provides a large component of the driving force in protein folding, protein–substrate, and protein–protein interactions.

In some enzyme systems, water is directly implicated in postulated catalytic mechanisms. It may act as a general acid or general base, sometimes activated by nearby charged groups, including metal ions. In hydrolytic processes, water may be the attacking nucleophile in the breakdown of an intermediate. Rapid transport of protons to and from an active site region may be facilitated by linked water networks.

Many relevant questions are unanswered, and some probably as yet unasked. The effects of the perturbed, partly ordered solvent shell on protein dynamics, and on substrate and coenzyme access to the protein, are not well understood. Is solvent tunnelling likely to be a significant effect, as suggested by recent experiments on electron transport? (Kroh & Stradowski 1973; Miller 1975). If so, how critical is the solvent organization in the surrounding region? More information on solvent organization is needed to help answer some of these questions. It is not easy information to obtain.

Some of the unpublished work quoted was done in collaboration with Peter Timmins, Stephen Heathman, and John Nicholas. Peter Timmins, Bob Bywater, and Christine Slingsby patiently answered the chemical and biochemical queries of an ignorant physicist.

I am grateful to Professor Dorothy Hodgkin, F.R.S., and her colleagues for access to the Oxford and Peking insulin maps, to Dr R. Huber for pre-publication data on trypsin and trypsin complexes, and to Herman Berendsen, Michael Levitt, Ned Seeman, and Arieh Warshel, for discussions.

References (Finney)

Adman, E. T., Sieker, L. C. & Jensen, L. H. 1973 *J. biol. Chem.* **248**, 3987–3996.
Adman, E. T., Sieker, L. C. & Jensen, L. H. 1975 *Acta Cryst.* A **31** (S3), 534.
Arnone, A., Bier, C. J., Cotton, F. A., Hazen, E. E., Richardson, D. C., Richardson, J. S. & Yonath, A. 1971 *J. biol. Chem.* **246**, 2302–2316.
Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D. & Waley, S. G. 1975*a* *Nature, Lond.* **255**, 609–614.
Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Rivers, P. S. & Wilson, I. A. 1975*b* *Acta Cryst.* A **31** (S3), 527.
Becker, J. W., Reeke, G. N., Wang, J. L., Cunningham, B. A. & Edelman, G. M. 1975 *J. biol. Chem.* **250**, 1513–1514.
Berendsen, H. J. C. 1968 In *Theoretical and experimental biophysics* (ed. A. Cole), vol. 1, pp. 1–76. New York: Dekker.
Berendsen, H. J. C. 1975*a* In *Water: a comprehensive treatise* (ed. F. Franks), vol. 5, pp., 293–330. New York: Plenum.
Berendsen, H. J. C. (ed.) 1975*b* *Report of CECAM Workshop on Molecular Dynamics of Long Time-Scale Events*, CECAM, Orsay, France.
Berendsen, H. J. C. & Migchelsen, C. 1966 *Fed. Proc.* **25**, 998–1002.
Birktoft, J. J. & Blow, D. M. 1972 *J. molec. Biol.* 68, 187–240.
Birktoft, J. J., Blow, D. M., Henderson, R. & Steitz, T. A. 1970 *Phil. Trans. R. Soc. Lond.* B **257**, 67–176.
Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. 1967 *Proc. R. Soc. Lond.* B **167**, 365–377.
Blow, D. M., Wright, C. S., Kukla, D., Rühlmann, A., Steigemann, W. & Huber, R. 1972 *J. molec. Biol.* **69**, 137–144.
Blundell, T., Dodson, G., Hodgkin, D. & Mercola, D. 1972 *Adv. Prot. Chem.* **26**, 279–402.
Bode, W. & Schwager, P. 1975*a* *J. molec. Biol.* (in the press).
Bode, W. & Schwager, P. 1975 *b* *FEBS Lett.* **56**, 139–143.
Bode, W., Schwager, P. & Fehlhammer, H. 1975 (in preparation).
Bode, W., Schwager, P. & Huber, R. 1975 *Proc. FEBS Congress*, Paris 1975 (in the press).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Brandén, C-I., Eklund, H., Nordström, B., Boiwe, T., Söderland, G., Zeppezauer, E., Ohlsson, I. & Åkeson, Å. 1973 *Proc. Nat. Acad. Sc. U.S.A.* **70**, 2439–2442.
Carlisle, C. H., Gorinsky, B. A., Moss, D. S., Palmer, R. A., de Rango, C., Mauguen, Y. & Tsoucaris, G. 1975 *Acta Cryst.* A **31** (S3), 521.
Carlisle, C. H., Palmer, R. A., Mazumdar, S. K., Gorinsky, B. A. & Yeates, D. G. R. 1974 *J. molec. Biol.* **85**, 1–18.
Carter, C. W., Kraut, J., Freer, S. T., Xuong, N. H., Alden, R. A. & Bartsch, R. G. 1974*a* *J. biol. Chem.* **249**, 4212–4225.
Carter, C. W., Kraut, J., Freer, S. T. & Alden, R. A. 1974*b* *J. biol. Chem.* **249**, 6339–6346.
Chothia, C. 1975 *Nature, Lond.* **254**, 304–308.
Chothia, C. & Janin, J. 1975 *Nature, Lond.* **256**, 705–708.
Cooke, R. & Kuntz, I. D. 1974 *Ann. Rev. Biophys. Bioeng.* **3**, 95–126.
Cotton, F. A. & Hazen, E. E. 1971 *Enzymes* (ed. P. D. Boyer), **4**, 153–175.
Cotton, F. A., Bier, J. C., Day, V. W., Hazen, E. E. & Larsen, S. 1972 *Cold Spring Harbour Symp. quant. Biol.* **36**, 243–255.
Cutfield, J. F., Dodson, E. J., Dodson, G. G., Hodgkin, D. C., Isaacs, N. W., Sakabe, K. & Sakabe, N. 1975 *Acta Cryst.* A **31** (S3), 521.
Dahlborg, U. & Rupprecht, A. 1971 *Biopolymers* **10**, 849–863.
Dehl, R. E. & Hoeve, C. A. J. 1969 *J. Chem. Phys.* **50**, 3245–3251.
Deisenhofer, J. & Steigemann, W. 1975 *Acta Cryst.* B **31**, 238–250.
Doyle, B. B., Haas, J., Hulmes, D. J. S., Ibel, K., Jenkins, G. T., Miller, A., Timmins, P. & White, J. W. 1975 *Brookhaven Symp. Neutrons in Biol.* (in the press).
Drenth, J., Hol, W., Jasonius, J. N. & Koekoek, R. 1972 *Cold Spring Harbor Symp. Quant. Biol.* **36**, 107–123.
Drenth, J., Jasonius, J. N., Koekoek, R., Sluyterman, L. A. A. & Wolthers, B. G. 1970 *Phil. Trans. R. Soc. Lond.* B **257**, 231–236.
Drenth, J., Jasonius, J. N., Koekoek, R. & Wolthers, B. G. 1971 *Adv. Prot. Chem.* **25**, 79–115.
Dunn, B. M., di Bello, C. & Anfinsen, C. B. 1973 *J. biol. Chem.* **248**, 4769–4774.
Eklund, H., Nordström, B., Zeppezauer, E., Söderland, G., Ohlsson, I., Boiwe, T., Söderberg, B-O & Brandén, C-I. 1975 *Acta Cryst.* A **31** (S3), 529.
Epp, O., Lattman, E. E., Schiffer, M., Huber, R. & Palm, W. 1975 (in preparation).
Glusker, J. P. 1968 *J. molec. Biol.* **38**, 149–162.
Hartsuck, J. A. & Lipscomb, W. N. 1971 *Enzymes* (ed. P. D. Boyer), **3**, 1–56.
Henderson, R. 1970 *J. molec. Biol.* **54**, 341–354.
Henderson, R., Wright, C. S., Hess, G. P. & Blow, D. M. 1972 *Cold Spring Harbor Symp. Quant. Biol.* **36**, 63–70.
Hol, W. 1971 Thesis, University of Groningen.
Hunkapillar, M., Smallcomb, S., Whitaker, D. & Richards, J. H. 1973 *Biochemistry* **12**, 4732–4743.
Jeffrey, G. A. & McMullan, R. K. 1967 *Prog. inorg. Chem.* **8**, 43–108.
Kaiser, E. T. & Lo, K-W. 1969 *J. Am. chem. Soc.* **91**, 4912–4918.
Kannan, K. K., Liljas, A., Waara, I., Bergstén, P-C., Lövgren, S., Strandberg, B., Bengtsson, U., Carlbom, U., Fridborg, K., Järup, L. & Petef, M. 1971 *Cold Spring Harbor Symp. Quant. Biol.* **36**, 221–231.
Kannan, K. K., Notstrand, B., Fridborg, K., Lövgren, S., Ohlsson, A. & Petef, M. 1975 *Proc. natn. Acad. Sci. U.S.A.* **72**, 51–55.
Kauzmann, W. 1959 *Adv. Prot. Chem.* **14**, 1–63.
Khalifah, R. G. 1971 *J. biol. Chem.* **246**, 2561–2573.
Klotz, I. M. 1958 *Science, N.Y.* **128**, 815–822.
Krieger, M., Kay, L. M. & Stroud, R. M. 1974 *J. molec. Biol.* **83**, 209–230.
Kroh, J. & Stradowski, C. Z. 1973 *Int. J. Radiat. Phys. Chem.* **5**, 243–245.
Kuntz, I. D. & Kauzmann, W. 1973 *Adv. Prot. Chem.* **28**, 239–345.
Lee, B. & Richards, F. M. 1971 *J. molec. Biol.* **55**, 379–400.
Lenhert, P. G. 1968 *Proc. R. Soc. Lond.* A **303**, 45–84.
Levitt, M. & Warschel, A. 1975 *Nature, Lond.* **253**, 694–698.
Liljas, A., Kannan, K. K., Bergsten, P-C., Waara, I., Fridborg, K., Strandberg, B., Carlbom, U., Järup, L., Lövgren, S. & Petef, M. 1972 *Nature New Biol.* **235**, 131–137.
Lipscomb, W. N., Hartsuck, J. A., Reeke, G. N., Quiocho, F. A., Bethge, P. H., Ludwig, M. L., Steitz, T. A., Muirhead, H. & Coppola, J. C. 1968 *Brookhaven Symp. Biol.* **21**, 24–90.
Ludwig, M. L., Anderson, R. D., Apgar, P. A., Burnett, R. M., Le Quesne, M. E. & Mayhew, S. G. 1972 *Cold Spring Harbor Symp. Quant. Biol.* **36**, 369–380.
Lumry, R. & Biltonen, R. 1969 In *Structure and stability of biological macromolecules* (eds S. N. Timasheff & G. D. Fasman), pp. 65–212. New York: Dekker.
Matthews, B. W. & Weaver, L. H. 1974 *Biochemistry* **13**, 1719–1725.
Migchelsen, C. & Berendsen, H. J. C. 1973 *J. Chem. Phys.* **59**, 296–305.
Migchelsen, C., Berendsen, H. J. C. & Rupprecht, A. 1968 *J. molec. biol.* **37**, 235–237.
Mildvan, A. S. 1974 *Rev. Biochem.* **43**, 357–399.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Miller, J. R. 1975 *Science, N.Y.* **189**, 221–222.
Moews, P. C. & Kretsinger, R. H. 1975 *J. molec. Biol.* **91**, 201–228.
Notstrand, B., Waara, I. & Kannan, K. K. 1975 *Isozymes* **1**, 575–599. New York: Academic Press.
Novak, T., Mildvan, A. S. & Kenyon, G. L. 1973 *Biochemistry* **12**, 1690–1701.
Page, D. I. 1972 In *Water: a comprehensive treatise* (ed. F. Franks), **1**, pp. 333–362. New York: Plenum.
Peking Insulin Structure Research Group 1974 *Sci. Sin.* **17**, 752–778.
Phillips, D. C. 1966 *Sci. Am.* **215** (11), 78–90.
Quiocho, F. A. & Lipscomb, W. N. 1971 *Adv. Prot. Chem.* **25**, 1–78.
Ramachandran, G. N. & Chandrasekharan, R. 1968 *Biopolymers* **6**, 1649–1658.
Richards, F. M. & Wyckoff, H. W. 1970 *Enzymes* (ed. P. D. Boyer), **4**, 647–806.
Rogers, G. A. & Bruice, T. 1974 *J. Am. Chem. Soc.* **96**, 2473–2481.
Rühlmann, A., Kukla, D., Schwager, P., Bartels, K. & Huber, R. 1973 *J. molec. Biol.* **77**, 417–436.
Sayre, D. 1974 *Acta Cryst.* A **30**, 180–184.
Sayre, D. 1975 *Acta Cryst.* A **31** (S3), S20.
Schoenborn, B. P. 1969 *Nature, Lond.* **224**, 143–146.
Schoenborn, B. P. 1972 *Cold Spring Harbor Symp. Quant. Biol.* **36**, 569–575.
Seeman, N. C., Day, R. O. & Rich, A. 1975 *Nature, Lond.* **253**, 324–326.
Stroud, R. M., Kay, L. M. & Dickerson, R. E. 1974 *J. molec. Biol.* **83**, 185–208.
Stillinger, F. H. & Rahman, A. 1974 *J. Chem. Phys.* **60**, 1545–1557.
Tulinsky, A. & Wright, L. H. 1973 *J. molec. Biol.* **81**, 47–56.
Vandlen, R. L. & Tulinsky, A. 1973 *Biochemistry* **12**, 4193–4200.
Vernon, C. A. 1967 *Proc. R. Soc. Lond.* B **167**, 389–401.
Wang, J. H. 1968 *Science, N.Y.* **161**, 328–334.
Watenpaugh, K. D., Sieker, L. C., Herriott, J. R. & Jensen, L. H. 1973*a* *Acta Cryst.* B **29**, 943–956.
Watenpaugh, K. D., Sieker, L. C. & Jensen, L. H. 1973*b* *Proc. natn. Acad. Sci. U.S.A.* 70, 3857–3680.
Whitney, P. L., Nyman, P. O. & Malmström, B. G. 1967 *J. biol. Chem.* **242**, 4122–4220.
Wright, H. T. 1973 *J. molec. Biol.* **79**, 1–11.
Wyckoff, H. W., Tsernoglou, D., Hansen, A. W., Knox, J. R., Lee, B. & Richards, F. M. 1970 *J. biol. Chem.* **245**, 305–328.
Yeagle, P. L., Lochmüller, C. H. & Henkens, R. W. 1975 *Proc. natn. Acad. Sci. U.S.A.* **72**, 454–458.
Yonath, A. & Traub, W. 1969 *J. molec. Biol.* **43**, 461–472.

Discussion

D. D. Eley, F.R.S. (*Department of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD*). For proteins such as bovine serum albumin and haemoglobin in the dry state, the room temperature resistivity of about 10^{17} Ω cm is reduced to around 10^{8} Ω cm by 20% of adsorbed water. Assuming electronic conduction and a mobility of 40 cm^2/V s. this corresponds to about 10^{9} conduction electrons/cm^3 which is about 1 electron/10^{10} protein molecules (Eley 1973). Similar figures will apply to many proteins, since the water adsorption isotherms and the electrical effects referred to are rather unspecific. If the electrons are to be injected into the protein from adsorbed water molecules we need some mechanism for perhaps one water molecule per protein molecule to be strongly chemisorbed, i.e. to insert in a C=O..HN bond of the α-helical or β-sheet hydrogen bonding system, to give $\mathrm{C{=}O..H{-}O..HN}$.

Is there any evidence from X-ray diffraction for the occasional adsorbed water molecule being taken up in such a way at favoured sites, e.g. at bends in an α-helix?

Reference

D. D. Eley 1973 Proc. 3rd Int. Conf. From Theoretical Physics to Biology (ed. M. Marois) Versailles 1971, p. 147. Basel: Karger.

J. L. Finney. I do not think we have enough data to draw reliable conclusions about specifically favoured water adsorption sites apart from the generalizations concerning the chemical nature of the groups available for hydrogen bonding at the protein molecule surface. The

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

α-helix itself is essentially a hydrophobic entity, and thus any direct water bonding to it would be a destabilizing influence upon it. There is some evidence in the *Chromatium* high potential iron protein that water is involved with sharp turns at the surface (Carter *et al.* 1974). I have not considered the water location in terms of its relation to secondary structure; bearing in mind the above reservations it might be useful to re-examine the available information in relation to this.

Reference

Carter, C. W., Krant, J., Freer, S. T., Xuong, N. H., Alden, R. A. & Bartsch, R. G. 1974 *J. biol. Chem.* **249**, 4212–4225.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES