23.4. Solvent structure

By C. Mattos and D. Ringe

23.4.1. Introduction

The unique properties of water and its role in nature have preoccupied the minds of scientists and philosophers for centuries. However, only relatively recently have the tools become available to study the specific roles that water molecules play with respect to protein structure and function. When the first crystal structure of a protein was obtained by X-ray diffraction (Kendrew, 1963), the focus was on the arrangement of the amino-acid residues into secondary and tertiary structure. Although the presence of water molecules associated with the protein was noticed, little attention was given to their structure and possible functional role. The structure of the protein itself was a great novelty, and its features were eagerly analysed. For many years, the crucial role of water molecules in maintaining both the structural integrity and the functional viability of proteins was not completely obvious, although in the 1950s Kauzmann argued correctly that water plays an important role in maintaining protein structure (Kauzmann, 1959). In the late 1970s and early 1980s, reviews began to appear focusing on the properties of water relevant to interaction with proteins (Edsall & McKenzie, 1978) and the location and role of water molecules on protein surfaces (Blake *et al.*, 1983; Edsall & McKenzie, 1983). As high-resolution structures became more easily attainable and refinement techniques improved, the importance of water molecules became increasingly apparent, and solvent structure now occupies a front seat in the realm of structural biology. There is a strong sense in the scientific community that water molecules play an integral role in many aspects of protein structure and function, and great effort is now being focused on understanding solvent effects in precise atomic detail.

In principle, water molecules can contribute both enthalpically and entropically to any process in which they are involved. The main contribution of water as solvent in the protein-folding process, for example, is entropic, driving the collapse of hydrophobic residues into the core of the protein. As is currently understood, the general shape of globular proteins is attained by this effect, with the specific structural features guided by the hydrogen bonds that define secondary structure (Hendsch & Tidor, 1994; Hendsch *et al.*, 1996). Although the solvent contribution to the protein-folding process is beyond the scope of this chapter, it is nevertheless deserving of brief mention. It is important to understand protein three-dimensional structures as having evolved in bulk water, a fact which is largely invisible to the current methods in structural biology. The size, geometry, planarity and orientational flexibility of water molecules give them structural and functional importance. All folded globular proteins have evolved with tens to hundreds of binding sites specific for water molecules, a situation contrary to that of larger ligands which usually bind in a single or small number of specific sites found in a given protein or family of proteins. In this respect, water is unique and its ubiquitous appearance is not a direct consequence of its chemical properties alone, but has an evolutionary origin. Proteins evolved in an aqueous milieu, where over time some water molecules were specifically incorporated as integral parts of the protein architecture.

At first glance, the surface of a protein determined by X-ray crystallography appears randomly populated by a layer of water molecules. A careful analysis, however, reveals that the arrangement of water molecules on protein surfaces is not random. In folded proteins, individual water molecules participate in a variety of structural and functional roles, ranging from filling small cavities that are not fully occupied by protein atoms to allowing flexibility, such as in the case of charged surface side chains that can move freely while continuously maintaining hydrogen-bonding partners. Water molecules can fill deep crevices on the protein surface, or they can play a crucial role in the thermodynamics of ligand binding. The mobility as well as the number and strength of hydrogen-bonding partners that are observed for water molecules bound to protein surfaces vary considerably, and it is becoming increasingly apparent that these factors are correlated with functional roles. The atomic coordinates for any protein should not be considered complete without those bound solvent molecules that can be observed, for they are part of the structure.

Bound water molecules have been implicated and studied in the context of substrate specificity and affinity (Quiocho *et al.*, 1989; Herron *et al.*, 1994; Ladbury, 1996), catalysis (Privé *et al.*, 1992; Singer *et al.*, 1993; Komives *et al.*, 1995), mediation of protein–DNA interactions (Clore *et al.*, 1994; Shakked *et al.*, 1994; Morton & Ladbury, 1996), cooperativity (Royer *et al.*, 1996), conformational stability (Bhat *et al.*, 1994), and drug design (Poormina & Dean, 1995*a,b,c*). One of the challenges now is to translate the structural information observed into a thermodynamic understanding of the water contribution to the various processes. In some cases, an attempt has been made to relate changes in water structure between two forms of a protein (*e.g.* ligated and unligated or native and mutant) to changes in the measured heat capacity (Holdgate *et al.*, 1997) or to measurements of enthalpy and entropy changes by titration calorimetry (Bhat *et al.*, 1994). Thermodynamic solvent isotope effects have also been reported, where the thermodynamics of association of several binding processes were evaluated calorimetrically in light and heavy water (Chervenak & Toone, 1994). In other cases, the three-dimensional structures were directly interpreted in terms of thermodynamic contributions (Quiocho *et al.*, 1989; Morton & Ladbury, 1996). Ultimately, a thorough understanding of the thermodynamics and kinetics underlying solvent structure will lead to powerful predictive methods. Theoreticians, on the one hand, have developed models based on physical principles and use experimental knowledge to assess whether their predictions are correct. Experimentalists, on the other hand, attempt to explain the observed phenomena in terms of the well established physical theories that govern the natural world. Progress is being made on both fronts, but a large gap still remains between the two. A bridge is being built from both sides of the gap and when the two sides meet at a common point, the many pieces of this complicated puzzle will have been deciphered and put in their proper places, so that a global view of molecular processes in water can be obtained from whatever perspective one wishes to take: chemical, physical, or biological.

The present chapter summarizes the empirical information gathered over the last decade or two on the structure of water molecules bound to proteins. The focus will be on structures solved by X-ray crystallography, although complementary techniques of obtaining solvent structure will be discussed briefly and, when appropriate, particular examples will be given. Section 23.4.2 is concerned with the methods by which solvent structure can be observed, Section 23.4.3 summarizes knowledge derived from database analysis of large numbers of proteins, Section 23.4.4 focuses on particular examples of groups of well studied protein structures, Section 23.4.5 discusses the contribution of protein models obtained at very high resolution to the understanding of solvent structure, and Section 23.4.6 contains an analysis of water

molecules as mediators of complex formation. Finally, Section 23.4.7 presents a conclusion and a perspective regarding the direction in which this information can lead in building a cohesive understanding of the roles played by solvent in the structural integrity and biological function of macromolecules.

23.4.2. Determination of water molecules

The most prominent method by which the structure of water molecules on the surface of macromolecules can be observed at the atomic level is X-ray crystallography. The information classically available from this methodology is on bound water molecules, characterized by a high probability density and reduced mobility relative to the bulk solvent, which results in clearly observed electron density. Information on water structure at larger distances from the protein is available in the low-resolution reflections, but difficulty in modelling the solvent in these areas has led to the common practice of discarding the very low resolution data. This chapter focuses on the water molecules for which there is information at high resolution (> 3 Å), although great progress has been made in recent years in modelling the disordered water structure at the protein–solvent interface, enabling more effective use of the low-resolution data (Badger, 1993; Jiang & Brünger, 1994; Lounnas et al., 1994). Typically, one is interested in studying solvent structure because of the effects that it has on the protein. Lounnas et al. (1994) gave a particularly interesting focus on the effect of the protein on the solvent structure surrounding it. Using a combination of molecular-dynamics simulations of explicitly solvated myoglobin and the low resolution X-ray data from myoglobin crystals, they devised a method to describe the effect of the protein on the solvent structure to a distance of 6 Å from the surface. They found that the mobility and probability density of water molecules perpendicular to the protein surface varied considerably depending on the particular composition and three-dimensional structure of the amino-acid residues at the particular area of interest (Lounnas & Pettitt, 1994; Lounnas et al., 1994).

There are a variety of criteria that have been used in placing crystallographic water molecules in electron-density maps. For tightly bound water molecules, with low B factors, the placement involves little or no subjectivity, but the choice of whether or not to include the more disordered waters (or those with low occupancy) can be rather subjective. It generally involves picking the electron-density contour level and B-factor cutoffs as well as making a choice of whether to use a simple difference electron-density map ($F_o - F_c$) or to use a higher-order difference electron-density map ($2F_o - F_c$ or $3F_o - 2F_c$). One criterion, applied consistently in placing water molecules on the surface of elastase structures, is the simultaneous presence of electron density at the 3σ contour level in an $F_o - F_c$ electron-density map and at the 1σ contour level in a $2F_o - F_c$ electron-density map. After refinement, those waters are kept that have a B factor of 50 Å^2 or less. A few exceptions do occur, where there is clear electron density for a water molecule with a B factor of up to 60 Å^2. Virtually all of the water molecules placed by these criteria have at least one hydrogen bond to a protein atom or to another water molecule and are mainly part of the first hydration shell on the protein surface.

A method that has provided information on solvent structure complementary to that obtained by X-ray crystallography is based on $D_2O - H_2O$ neutron difference maps (Shpungin & Kossiakoff, 1986). The main advantage of this methodology is in locating partially ordered water molecules whose electron-density peaks

may be at the limit of the signal-to-noise ratio allowed for confidently determining positions of water molecules by X-ray diffraction. Scattering of neutrons by H_2O and D_2O is quite different, while scattering from the protein remains the same. Therefore, difference maps based on the two data sets should average to zero where the protein is present and result in peaks only where water molecules are found. Neutron scattering is particularly suited to this because of the threefold greater scattering power of deuterated water molecules relative to light water, providing a larger signal-to-noise ratio in assigning water positions. This method is particularly useful in detecting the second hydration sphere on protein surfaces (Kossiakoff et al., 1992).

NMR spectroscopy can also serve as a complementary technique, providing dynamic information on the lifetime of interaction of a single water molecule on the protein surface. The fact that, with few exceptions, no cross-relaxation peaks are observed at the protein–water interface is an indication that the motion timescale for water molecules in contact with protein is close to that in bulk water at room temperature. The NMR data suggest that water molecules observed in crystal structures have lifetimes of the order of tens of nanoseconds or less (Bryant, 1996). A small number of relatively long-lived structural waters (with residence times in the range 10^{-2} to 10^{-8} s) can be detected by modern NMR techniques (Otting et al., 1991). Four water molecules have been detected by NMR in bovine pancreatic trypsin inhibitor (BPTI) (Otting & Wuthrich, 1989) and six have been observed in complexes of human dihydrofolate reductase with methotrexate (Meiering & Wagner, 1995). Observation of these waters in the corresponding crystal structures reveals that they are tightly bound waters, with three or four hydrogen bonds to protein atoms, and many are found to bridge between secondary-structure elements or are found to mediate protein–ligand interaction (Meiering & Wagner, 1995). It is important to understand, then, that water molecules near protein surfaces occupy energy minima favoured by hydrogen bonding and ion–dipole effects, which results in water molecules being present in these positions more often than in others. Although when looking at a crystallographic protein structure it is easy to think of a given site as being occupied by a single water molecule, it is in fact only the site that is single, with an enormous number of different individual water molecules sampling it during the time of data collection. This was qualitatively understood from the beginning, but NMR experiments have played a key role in setting quantitative upper boundaries to the residence times of water molecules on the protein surface.

Finally, mention must be made of the computational efforts invested in representing and understanding solvent structure on macromolecular surfaces. The computational work encompasses a variety of methodologies, including integral equation methods (Beglov & Roux, 1997), molecular dynamics (Brooks & Karplus, 1989; Hayward et al., 1993; van Gunsteren et al., 1994), thermodynamic understanding through free-energy simulations (Roux et al., 1996) and statistical-mechanics calculations (Lazaridis et al., 1995). The results of these studies are often complementary to the experimental information already available and provide an important component to the current insight on solvent structure (McDowell & Kossiakoff, 1995) and function (Pomes & Roux, 1996; Oprea et al., 1997). Furthermore, these techniques often provide the only means of obtaining an energetic understanding of some aspects of protein–water interaction.

The question of how the different techniques used to observe the location and properties of water molecules on the surface of proteins relate to and complement one another has been discussed in two short review articles (Levitt & Park, 1993; Karplus & Faerman, 1994). Karplus & Faerman discuss the reliability of each of the methods, illustrating their strengths and weaknesses, while Levitt &

Park present the state of our understanding of protein–water interactions as it was in 1993, based on the synthesis of results obtained from the various methods discussed above.

23.4.3. Structural features of protein–water interactions derived from database analysis

The location and nature of water interaction with protein atoms are of great interest for understanding the role played by water molecules in the structural integrity and function of macromolecules. Baker & Hubbard (1984) presented an extensive analysis of hydrogen bonding in 15 proteins. A good portion of the study focused on hydrogen bonding with water. They observed that, in general, hydrogen bonds have a certain degree of flexibility, ranging in distance between 2.4 and 3.4 Å, with angular deviation from linear of up to 60°. The authors discussed the hydrogen-bonding geometry of water itself as well as the general aspects of the hydration of protein groups. Along the protein backbone, each carbonyl group is capable of making two hydrogen bonds, while amido groups make only one. Bifurcated hydrogen bonds are relatively rare, comprising only about 4% of the main-chain amido groups and even fewer of the side chains. Baker & Hubbard (1984) observed that of all of the hydrogen bonds made by water molecules, 42% are to main-chain carbonyl oxygens, 14% to main-chain amide groups and 44% to side-chain atoms. In a subsequent review that surveyed protein–water interactions, Savage & Wlodawer (1986) pointed out some of the major problems that hinder the accurate study of the precise hydrogen-bonding geometry and chemical features of protein–water interactions: the size of the biomolecular system, the resolution of the data, and the disorder of both the biomolecule and the solvent. The review was based on a comparison of X-ray and neutron diffraction studies of water interactions in a handful of proteins solved to a resolution of 1.5 Å or better with hydration properties in crystals of small- and medium-sized molecules solved to better than 1.0 Å resolution. Although a great deal had been learned about hydrogen-bonding properties of water in crystals of small molecules that presumably can be transferred to analogous interactions with protein atoms (Savage, 1986), the authors pointed out that for biomolecules there was, at the time, no consistent method being used for solvent analysis (Savage & Wlodawer, 1986). This problem was demonstrated and analysed in a more recent review, where a comparison of three independently solved structures of interleukin-1 reveals a large variability in solvent structure (Karplus & Faerman, 1994).

The growing number of high-resolution protein crystal structures currently available in the Protein Data Bank (Berman *et al.*, 2000) allows for studies that extract statistically significant trends specific to protein–water interactions. The analysis of where and how water molecules bind to protein surfaces can be made at different levels. One can look at general properties of water interacting with each of the 20 amino-acid side chains, as well as with main-chain carbonyl oxygens and amido nitrogen atoms. At a higher level, one can study how these local interactions are modulated by the secondary-structure elements in which the residues are found. At the tertiary-structure level, one can study the location and function of water molecules as they are found in bridging secondary-structure elements and their role in the integrity of the protein architecture. At this level, studies regarding surface shape and hydrophilicity become important components of the analysis. Finally, the role of water molecules can be studied at the level of mediating protein–protein and protein–ligand interaction and their function in the affinity and specificity of these interactions. The remainder of this section summarizes information from database analysis of protein–water interactions at these various levels. The following sections then focus on individual examples to illustrate the classifications and functions of water–protein interactions.

23.4.3.1. *Water distribution around the individual amino-acid residues in protein structures*

The most comprehensive study of water molecules at the local level of binding to the individual types of amino-acid residues in protein structures was published in a series of papers (Thanki *et al.*, 1988, 1990, 1991; Walshaw & Goodfellow, 1993). The initial database consisted of 16 protein structures solved to better than 1.7 Å resolution and refined to an *R* factor of 26% or better (Thanki *et al.*, 1988). It was subsequently increased to 24 proteins using the same selection criteria (Thanki *et al.*, 1990, 1991; Walshaw & Goodfellow, 1993). All equivalent side chains as well as carbonyl or amide groups present in the database were brought to a common reference frame constructed from previously established bond lengths and bond angles (Momany *et al.*, 1975). The distribution of water molecules interacting with each of the 20 types of side chains was studied by focusing on particular atoms. Therefore, water molecules within 3.5 Å of N and O polar side-chain or main-chain atoms or within 5.0 Å of apolar side-chain carbon atoms were appropriately translated to the reference frame.

Fig. 23.4.3.1 shows the results of these superpositions for the polar main-chain amido and carbonyl groups as well as for some representative polar side chains: Ser, Tyr, Asp, Asn, Arg, His, Trp and Ala. The overall results show that despite the complex protein architecture, water molecules interact with hydroxyl, carbonyl and amide moieties, as well as with the sp^3-hybridized and ring nitrogen atoms, as expected from their known stereochemical requirements (Baker & Hubbard, 1984). Thus, there are water clusters in positions that optimize interaction with the lone-pair electrons on oxygen atoms and with the hydrogen atoms of amide and hydroxyl groups. Figs. 23.4.3.1(*a*) and (*b*) show the distribution of water molecules around the main-chain carbonyl oxygen and amido nitrogen atoms, respectively. The stereochemical requirements mentioned above are satisfied, with the distribution around the carbonyl oxygen clustered in two distinct regions peaking at an O–O distance of 2.7 Å. In contrast, there is a single water cluster interacting with the nitrogen, in line with the N—H bond at an N–O distance of about 2.9 Å. This cluster is much tighter than seen for the interactions with oxygen, reflecting a greater flexibility of water interaction with the carbonyl oxygen relative to the amido-group nitrogen atom.

Ser and Thr residues present a wide distribution of water molecules around the hydroxyl groups, presumably due to the freely rotating side chain. Fig. 23.4.3.1(*c*) shows the water-molecule distribution around Ser, which is only slightly different from that for Thr and can be representative of both. In contrast, the Tyr hydroxyl group is involved in resonance stabilization with the aromatic ring and, consequently, water molecules are clustered in the plane of the ring in well defined positions (Fig. 23.4.3.1*d*).

Fig. 23.4.3.1(*e*) shows the clustering of water molecules around the Asp side chain into four distinct groups, corresponding to the four available lone-pair electrons. The distribution around Glu is similar. Most water molecules interact with a single carbonyl oxygen, although about 11% (for Asp) and 15% (for Glu) of water molecules around these side chains interact with both oxygen atoms of a single carboxyl group. Water molecules that interact with Asn and Gln also show four clusters, with the two clusters around the carbonyl group (C=O) less distinct than those around the amido (NH$_2$) group. Fig. 23.4.3.1(*f*) shows the distribution of water-molecule sites around Asn. In the case of Gln, the difference in water clustering around the carbonyl and amido groups is much less pronounced, possibly due to a greater degree of confusion in placing this longer side chain in the correct orientation. About 6% of the
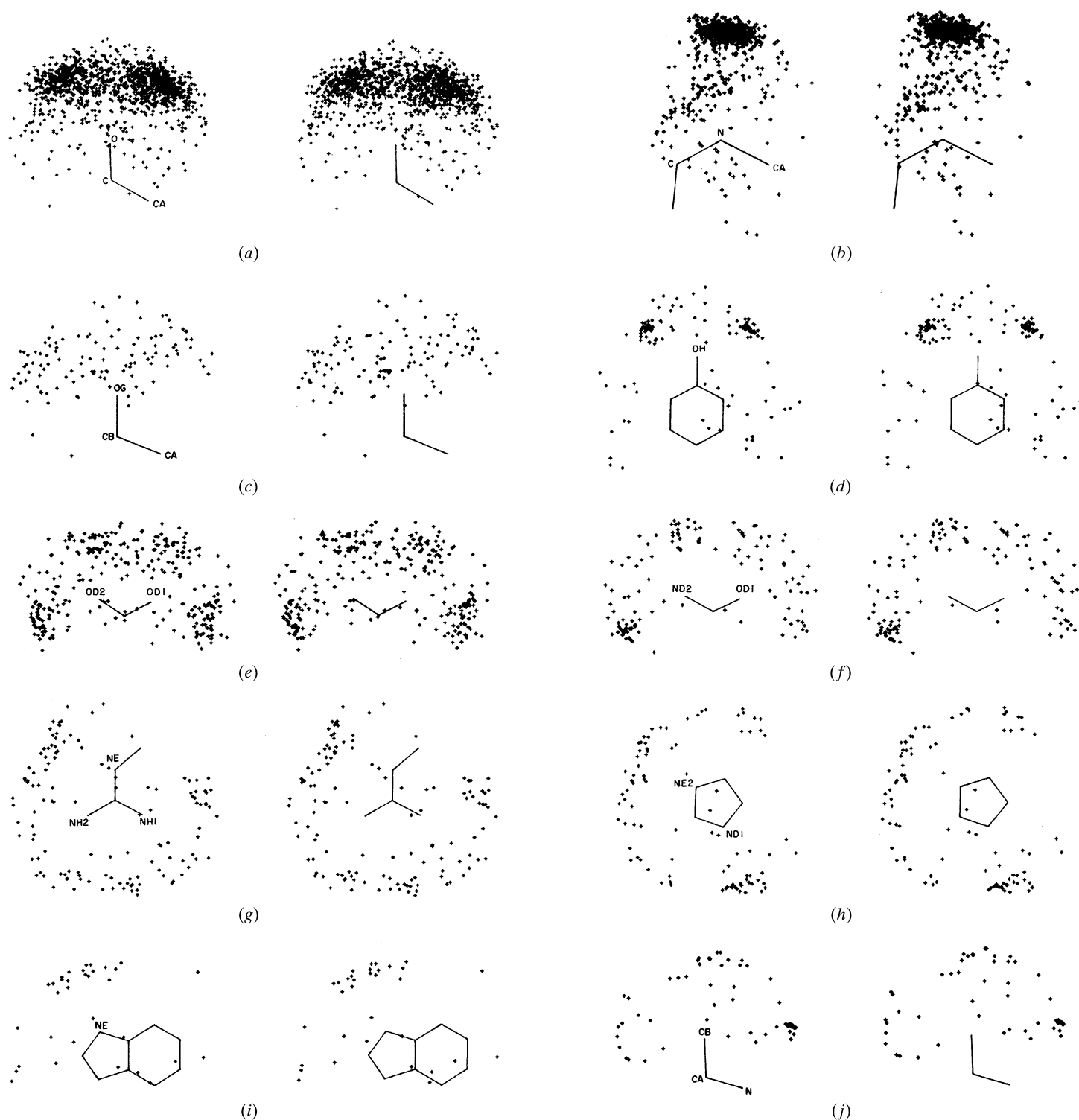
Fig. 23.4.3.1. Distribution of water-molecule sites in stereo around: (*a*) main-chain O, (*b*) main-chain N, (*c*) Ser OG, (*d*) Tyr ring, (*e*) Asp OD1 and OD2, (*f*) Asn OD1 and ND2, (*g*) Arg NH1, NH2 and NE, (*h*) His ring to 3.5 Å, (*i*) Trp ring to 3.5 Å, (*j*) Ala CB. Reprinted with permission from Thanki *et al.* (1988). Copyright (1988) Academic Press.

water molecules that interact with Asn or Gln are involved in hydrogen bonding to both the carbonyl oxygen and the amido nitrogen atoms.

The clustering of water molecules around the planar guanidyl group of Arg is distinctly positioned around the Nε atom and on either side of the NH1 and NH2 atoms. This is shown in Fig. 23.4.3.1(*g*). The clusters peak at a distance of about 3.0 Å from the nitrogen atoms. 7% of these water molecules are shared between NH1 and NH2, and only 3% are shared between the Nε and NH1

atoms. The distribution around the Lys side chain is much broader and is qualitatively similar to the one shown for Ser in Fig. 23.4.3.1(*c*), with no particular orientational preferences, mainly due to the freely rotating nature of the Cε—Nζ bond.

His and Trp are the two residues that contain ring nitrogen atoms, which comprise the main site of interaction with water molecules for these side chains. The distributions of water molecules within 3.5 Å of these residues are shown in Figs. 23.4.3.1(*h*) and (*i*). The clustering around His shows a peak at 2.7 Å and a larger peak at

3.1 Å. The closer peak corresponds to interactions with deprotonated nitrogen (Nδ), where the lone pair of electrons renders the deprotonated nitrogen more negatively charged than the corresponding protonated nitrogen (Nε) and, therefore, the deprotonated nitrogen pulls the water molecule closer. The peak at 3.1 Å is due to water interactions with the protonated nitrogen (Nε) of His. There is a strong preference for the water molecules to lie in the plane of the ring. Relatively few water molecules exist within 3.5 Å of Trp. They mostly cluster around the Nε nitrogen at varying distances. The number of water molecules interacting with His and Trp within 5.0 Å of the ring increases greatly and peaks at a distance of about 4 Å, as discussed below for hydrophobic residues in general (Walshaw & Goodfellow, 1993).

Overall, there seem to be weaker geometric constraints on oxygen acceptors compared to nitrogen donors. Furthermore, the water interaction with oxygen atoms peaks at a distance of about 2.8 Å, while the interactions with protonated nitrogen atoms occur at a somewhat longer distance of about 3.1 Å. This is possibly due to the larger van der Waals radius of nitrogen (1.8 Å) *versus* that of oxygen (1.7 Å) (Thanki *et al.*, 1988). A more recent study of hydration around polar residues is based on seven proteins solved to better than 1.4 Å resolution (Roe & Teeter, 1993). The authors used cluster analysis to derive a predictive algorithm to locate water sites around polar side chains on protein surfaces, given the atomic coordinates of the protein alone. These more precise results confirm the general conclusions outlined above. The authors find that the water–oxygen distance is less than that of water–nitrogen by 0.07 Å and suggest the difference to be due to a van der Waals radius of 1.5 Å for nitrogen and 1.4 Å for oxygen (Roe & Teeter, 1993). Although the two groups cite different atomic radii for nitrogen and oxygen, this does not have an effect on the statistical analysis of the data. Roe & Teeter (1993) also find that the clusters associated with nitrogen atoms are approximately two times denser than those around oxygen atoms.

The analysis of the local water structure around the apolar side chains Ala, Val, Leu, Ile and Phe was extended to a distance 5.0 Å from the atom of interest, since these residues show only a few water molecules within the 3.5 Å cutoff used to analyse interactions with polar residues. The most noticeable observations from the analysis of apolar side chains are the water peak at a distance of 4 Å from the carbon atoms of interest and the presence of a polar protein atom within a hydrogen-bonding distance for 75% of these water molecules (Walshaw & Goodfellow, 1993). Phe prefers in-plane interactions and has peaks corresponding to the direction of the Cε1, Cε2, Cδ1 and Cδ2 atoms from the centre of the ring. Otherwise, any clustering observed for water molecules near apolar side chains is due to interactions with polar protein atoms and, consequently, is modulated by secondary structure.

A study of protein hydration based on atomic and residue hydrophilicity presents general results consistent with those discussed above, but also adds information that can be correlated with various experimentally and computationally derived hydrophilicity–hydrophobicity scales (Kuhn *et al.*, 1995). The authors used 10 837 water molecules found in 56 high-resolution protein crystal structures to obtain the average number of hydrations per occurrence over each amino-acid type and specific atom types. The hydration of the various amino-acid residues has already been discussed above. The atomic hydrophilicity values calculated for the different protein-atom types are of interest. Fig. 23.4.3.2 and Table 23.4.3.1 show that, regardless of where these atoms are found, neutral oxygen atoms exhibit the greatest hydration level per occurrence, closely followed by negatively charged oxygen atoms, which in turn are followed by positively charged nitrogens and neutral nitrogens, in that order. Carbon and sulfur atoms are indistinguishable in terms of hydration per occurrence and are grouped together as the least hydrated atoms (Kuhn *et al.*, 1995).
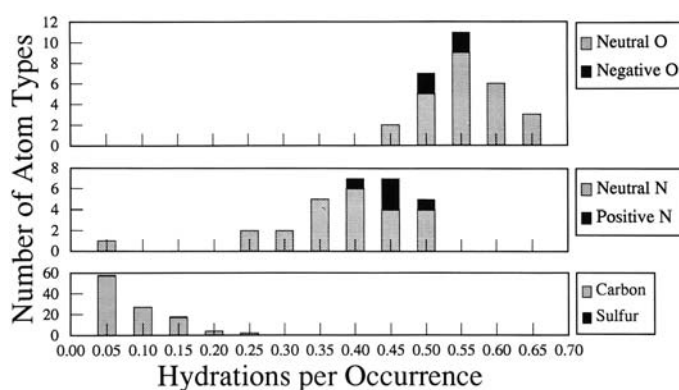

Fig. 23.4.3.2. Distribution of atomic hydration values. To determine which atoms are similar or distinct with respect to water binding, we plotted the number of atom types (*e.g.* Ala amide nitrogen, Ala Cα, . . .) at each hydration per occurrence value. Each atom type contributed one vertical unit to the graph. Oxygen atoms were the most hydrated (top graph), with negatively charged oxygen (black bars) slightly less hydrated on average than neutral oxygen (grey bars). Nitrogens (middle graph) were the next most hydrated, overlapping the oxygen distribution, and positively charged nitrogens (black bars) were somewhat more hydrated than neutral nitrogens (grey bars). Proline's amide nitrogen, with no hydrogen-bonding capacity, had the lowest nitrogen hydration value (leftmost bar). Carbon and sulfur atoms (bottom graph; note change of *y*-axis scale) were the least hydrated, with sulfur values at 0.05 and 0.15 hydrations per occurrence. Reproduced from Kuhn *et al.* (1995). Copyright (1995) Wiley-Liss, Inc. Reprinted by permission of Wiley-Liss, Inc., a division of John Wiley & Sons, Inc.

23.4.3.2. *The effect of secondary structure on protein–water interactions*

The main effect of secondary structure is on the hydration of main-chain carbonyl oxygens and amido nitrogen atoms. The clustering of water molecules around the small aliphatic apolar side chains (Walshaw & Goodfellow, 1993) and the Ser and Thr side chains (Thanki *et al.*, 1990) were also found to be guided by interactions with main-chain atoms belonging to a specific secondary structure. Other side chains are too large to have their hydration significantly affected by secondary structure. The broad solvent distribution around Ser and Thr side-chain hydroxyl oxygen atoms results from the combination of complex, but distinct, patterns that emerge when hydration around these side chains is examined separately in α-helices and β-sheets. Preferential hydrogen-bonding positions around Ser and Thr residues result from water molecules bridging between the hydroxyl group and another polar protein atom within the α-helix or β-sheet. These positions are dependent both on the χ_1 torsion angle and the type of secondary structure within which these residues are found (Thanki *et al.*, 1990).

Table 23.4.3.1. *Specific hydrophilicity values for protein atoms*

Atom type	Hydrations per occurrence *
Neutral oxygen	0.53
Negative oxygen	0.51
Positive nitrogen	0.44
Neutral nitrogen	0.35
Carbon, sulfur	0.08

* The average number of hydrations per occurrence was calculated over all atoms within each group.
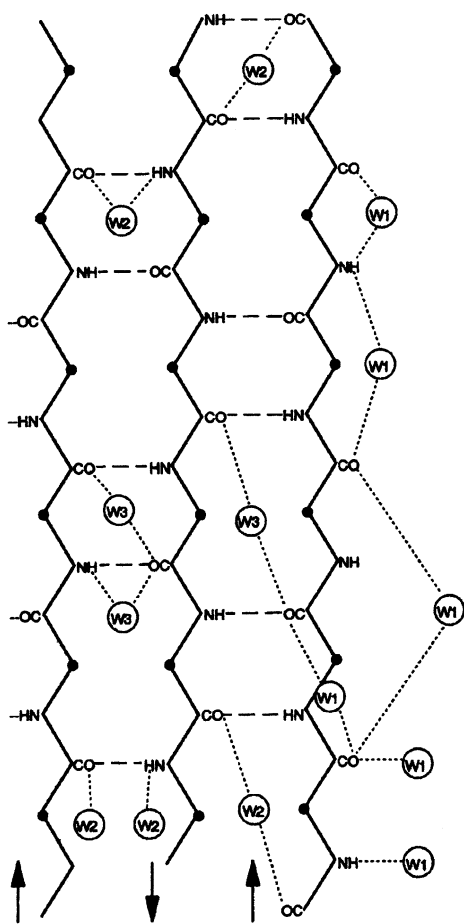
627

Fig. 23.4.3.3. Diagram of edge (W1), end (W2) and middle (W3) categories of interactions of water molecules with main-chain atoms in antiparallel β-sheets. Reprinted with permission from Thanki *et al.* (1991). Copyright (1991) Academic Press.

The analysis of main-chain hydration focused separately on hydration of β-sheets, α-helices and turns (Thanki *et al.*, 1991). In general, more water molecules were found to interact with carbonyl oxygens than with amide groups, due primarily to the fact that carbonyl oxygen atoms can accept two hydrogen bonds, whereas amido groups can donate a single one. Thus, free carbonyl oxygen atoms have the potential to interact with two water molecules, whereas those already involved in a secondary-structure interaction with the protein still have a lone pair of electrons that can accept a hydrogen bond from a water molecule. Of the free carbonyl oxygen atoms within secondary-structure elements, 45% of those in α-helices and 68% of those in β-sheets interact with water molecules. Of those that are involved in secondary-structure interactions within the protein, 21% of those in α-helices and 17% of those in β-sheets also interact with solvent. The free amide groups are well hydrated, with 38% of those in α-helices and 54% of those in β-sheets interacting with water molecules. However, virtually none (2% in helices and 6% in sheets) of the amides already involved in secondary-structure hydrogen bonding also interact with a water molecule.

Three types of interactions were observed for water molecules in the context of β-sheets (Fig. 23.4.3.3). Most (68%) of these interactions are with the edge of the β-sheet, in an extension of the secondary structure. The second most prominent type of interaction, comprising 23% of the total, is at the ends of the β-strands with either free amido or carboxyl groups. Finally, only 10% of the water molecules are found to bridge between two strands in the middle of the β-sheet.

Interactions of water molecules with α-helices are also found in three distinct positions relative to the secondary structure (Fig. 23.4.3.4): at the carbonyl terminus of the helix, at the amido terminal end and in the middle. Of those interacting at the carbonyl terminus, 48% interact with the carbonyl oxygen alone, 11% also interact with a nearby main-chain atom and 41% are involved in a water-mediated C cap, bridging a small polar side chain (Ser, Thr, Asp, or Asn) to a free carbonyl group at the end of the helix. Of water molecules interacting at the amido terminus of the helix, 25% interact with free amido groups alone, 45% bridge to local main-chain atoms and many of the remaining mediate in N-cap interactions with small polar side chains such as Ser and Asp.

In general, turns have a high exposure to solvent and therefore are found to be well hydrated. The pattern of hydration varies both with the type of turn and the location of the atoms within the turn. Not surprisingly, there are about twice as many hydrogen bonds to carbonyl groups as there are to amide groups in turns. Although the majority of the water interactions with turns are to single carbonyl oxygen or amido nitrogen atoms, bridging water molecules do appear, especially within more open turns. They occur in a variety of different patterns, bridging between two main-chain atoms in the turn or between a main chain and a small polar side chain.

Clearly, water molecules play a functional role in maintaining the integrity of the secondary-structure elements of proteins. They are often seen to extend α-helices or β-sheets, serving as an interface between these secondary-structure elements and the bulk solvent. Water molecules are also found to mediate the
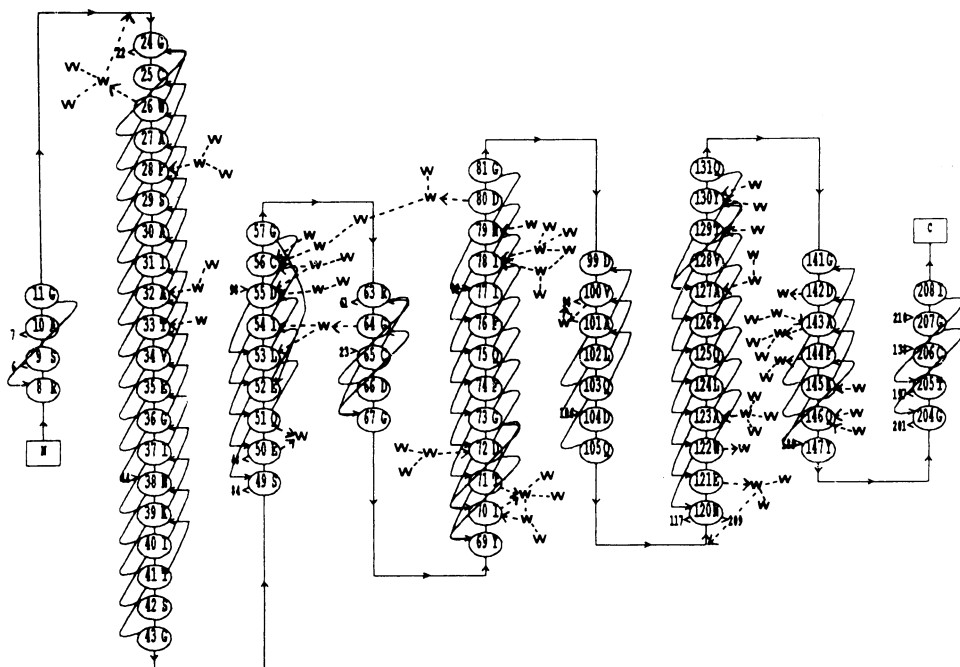

Fig. 23.4.3.4. Diagram of the hydrogen bonds in the α-helical structure in actinidin. Reprinted with permission from Thanki *et al.* (1991). Copyright (1991) Academic Press.

interaction between two protein atoms within a given secondary structure that may be too far from each other to interact directly. This may be of great importance in turns, particularly the more open ones where the protein atoms are not in ideal positions to form a tight two-residue β-turn.

23.4.3.3. *The effect of tertiary structure on protein–water interactions*

At the tertiary level, there is an interdependence between protein surface shape and the extent of water binding (Kuhn *et al.*, 1992). Kuhn *et al.* (1992) studied the binding locations of 10 837 water molecules found in 56 high-resolution crystal structures using fractal atomic density and surface-accessibility algorithms. They found strong correlations between the positions of water molecules and protein surface shape and amino-acid residue type. A probe sphere with the radius of a water molecule revealed that, in general, protein surfaces exhibit convex groove areas and concave contact surfaces. Although grooves account for approximately one quarter of a given protein surface, they bind half the water molecules. Furthermore, only within grooves was hydration found to be dependent on residue type, with charged and polar residues as well as main-chain nitrogen and oxygen atoms exhibiting a greater degree of hydration than the non-polar residues. Outside the grooves, there was a low residue-independent hydration level, with no distinction between main-chain and side-chain atoms (Kuhn *et al.*, 1992). Levitt & Park (1993) discuss the paradox between the experimental observation that water molecules are crystallographically observed primarily in crevices (Kuhn *et al.*, 1992) and the results from theoretical calculations that argue that surface tension should make crevice waters bind less strongly (Nicholls *et al.*, 1991).

While the majority of the crystallographically observed water molecules appear on the outer protein surface, the internal protein packing is not perfect, so that the three-dimensional fold usually results in a number of internal cavities that can accommodate buried water molecules. The first analysis of such cavities was based on a small set of 12 proteins for which the authors characterized such sites by their size and area, as well as by whether or not they were occupied by crystallographically observed water molecules (Rashin *et al.*, 1986). More recently, two methodologically distinct studies of intramolecular cavities used much larger databases to provide extensive and mutually consistent conclusions regarding the properties of these sites (Hubbard *et al.*, 1994; Williams *et al.*, 1994). Hubbard *et al.* (1994) analysed 121 protein chains, with no two possessing a pairwise identity greater than 40%. This study is based on a systematic method of determining the shape as well as the size of the internal cavities and categorizes each cavity as either 'solvated' (with crystallographically visible water molecules) or 'empty' (with no crystallographically visible water molecules), noting the amino-acid-residue preferences in each type. Hydrogen-bonding patterns were also noted within the solvated sites. The second study (Williams *et al.*, 1994) selected 75 non-homologous monomeric proteins, solved at 2.5 Å resolution or better. Although the authors noted the general shape, size and location of cavities, the focus of this study was on the buried water molecules and the hydrogen-bonding patterns that they form within these sites.

In general, larger proteins are able to tolerate larger cavity sizes than small proteins, and nearly all proteins with more than 100 amino-acid residues are found to have at least one cavity. These cavities are found in the protein interior at a variety of distances from the surface and reflect the difficulty of perfect packing within the core. In the database of 121 proteins (Hubbard *et al.*, 1994), 265 cavities were found to be 'solvated' and 383 were 'empty'. The solvated cavities tend to be nearer to the protein surface than the empty cavities. Nearly 60% of the solvated cavities are occupied by

a single water molecule and are of spherical shape. About 20% accommodate two water molecules, and 20% more are found to contain larger clusters (Williams *et al.*, 1994). These tend to have an elongated cigar shape. The cavity volume can be as large as 216 $\mathring{A}^3$ (an elastase cavity containing seven water molecules). The solvated cavities tend to be larger than the empty ones, with average volumes of 39.4 and 20.7 $\mathring{A}^3$, respectively (Hubbard *et al.*, 1994). The mean volume per water molecule in a cavity is 27 $\mathring{A}^3$, as compared to 30 $\mathring{A}^3$ in bulk water, suggesting that a water molecule is not favourably squeezed into a volume comparable to its own (11.5 $\mathring{A}^3$), but rather occupies similar volumes upon transfer from the bulk into the protein interior.

Solvated cavities differ from empty ones not only in location and size within the protein, but also in the constitution of the amino-acid residues lining the cavity and the secondary-structure elements that are nearby. While 50% of the total cavity molecular surface is provided by polar atoms in solvated cavities, this fraction reflects only 16% of the empty cavity surface. Polarity, not size, is the predominant factor in determining the solvation state of a cavity. Interestingly, solvated cavities have more surface area provided by coil residues than the empty cavities, often found to be lined by residues in secondary structure (Hubbard *et al.*, 1994).

There is on average one buried water molecule per 27 amino-acid residues, although there is great variation between individual proteins. These water molecules most commonly form at least three hydrogen bonds with protein atoms or other buried water molecules. Only 18% of buried water molecules make two or fewer polar contacts. Of all of the hydrogen bonds made by buried water molecules, 53% are to protein backbone atoms, 30% to protein side-chain atoms, 17% to other buried water molecules, and 3% make no visible polar contacts at all (Williams *et al.*, 1994).

The appearance of cavities in the protein core is a consequence of the optimal packing of the protein polypeptide chain as it folds into the native, functional state. Where these cavities expose polar atoms to the hydrophobic protein core, one or more buried water molecules effectively become part of the structure, serving to maintain the protein integrity by fulfilling the hydrogen-bonding potential of atoms which are more favourably solvated.

23.4.3.4. *Water mediation of protein–ligand interactions*

A series of three papers presents the results of an analysis of water molecules mediating protein–ligand interactions in 19 crystal structures solved to better than 2.0 Å resolution and refined to an *R* factor of at least 23% (Poormina & Dean, 1995*a,b,c*). The studies focus on hydrogen-bonding features of water molecules bridging protein–ligand complexes (Poormina & Dean, 1995*b*), on the surface shape of the protein and ligand molecules at the water-binding sites (Poormina & Dean, 1995*c*), and on the structural and functional importance of water molecules conserved at the binding sites in five sets of evolutionarily related proteins (Poormina & Dean, 1995*a*). This study was largely motivated by an attempt to distinguish between properties of water-binding sites where water molecules are displaced by ligands and those where water molecules must be considered as part of the protein surface. This type of understanding has direct implications for drug and ligand design.

In general, there is a strong correlation between the number of water molecules found to bridge any given protein–ligand complex and the number of hydrophilic groups associated with the ligand. Within this context and in agreement with the conclusions of Kuhn *et al.* (1992), the authors found that the protein shape is important in determining the location of water-binding sites at the protein–ligand interface. Fig. 23.4.3.5 illustrates the different types of grooves observed in this study. Figs. 23.4.3.5(*a*) and (*b*) represent binding of
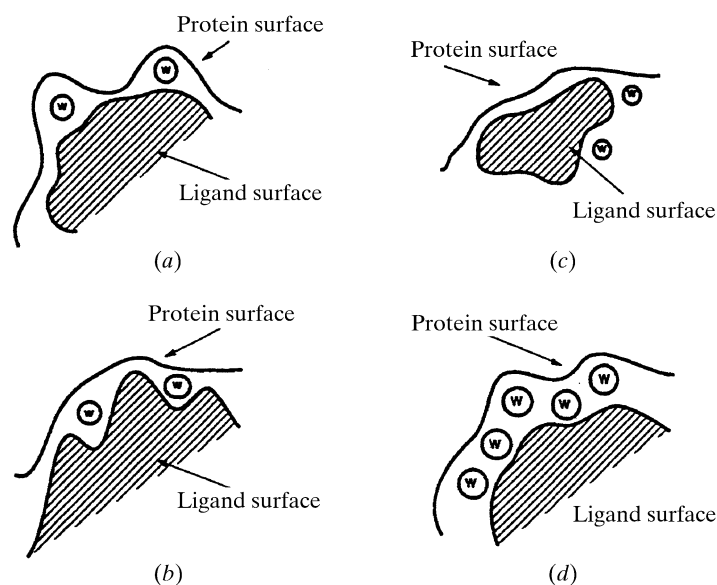
Fig. 23.4.3.5. Schematic illustration of water molecules bound in different types of grooves between protein and ligand. The hatched surfaces represent the ligand surface. (*a*) Water molecules bound in an indentation on the protein surface, where the protein surface area exposed to the water molecules is far larger than the ligand surface area; (*b*) water molecules bound in indentations on the ligand surface, where the ligand surface area exposed to the water molecule is larger than the protein surface area; (*c*) water molecules bound in shallow grooves at the protein–ligand interface and on the ligand surface; and (*d*) water molecules bound in clusters in elongated grooves with micro-grooves. Reprinted with permission from Poormina & Dean (1995*c*). Copyright (1995) Kluwer Academic Publishers.

bridging water molecules in deep grooves on the protein or on the ligand, respectively. The most common situation is illustrated in Fig. 23.4.3.5(*a*), with that in Fig. 23.4.3.5(*b*) occurring very rarely. Fig. 23.4.3.5(*c*) shows the situation where water molecules are found to interact with the ligand alone or at the periphery of the protein–ligand interface. Finally, Fig. 23.4.3.5(*d*) illustrates the situation where clusters of water molecules occupy elongated grooves, mediating the protein–ligand interaction. A striking example of this is given by the complex between chloramphenicol acetyl transferase and chloramphenicol, where two clusters of water molecules are found to form a layer between the enzyme and the ligand (Poormina & Dean, 1995*c*).

For the purposes of analysis, the authors distinguish between water molecules that interact with both protein and ligand, forming a bridge between the two, and water molecules that interact with either the protein or the ligand, but not with both. There is also a group of water molecules that interact with neither protein nor ligand, but are thought to contribute to the stability of the network of water molecules at the protein–ligand interface.

Of the 58 water molecules found to bridge between protein and ligand, 38 (nearly 80%) make three or more hydrogen bonds and satisfy tetrahedral geometry. Furthermore, they bind in deep grooves, generally interacting more strongly with the protein (Fig. 23.4.3.5*a*). The *B* factors of these bridging water molecules are comparable to those of the protein atoms with which they interact. They can, in effect, be considered an integral part of the protein structure and binding site. Many of these bridging water molecules are conserved throughout homologous proteins, even when different ligands are considered, and are clearly structurally significant in maintaining the properties of the protein binding sites.

Water molecules found to bind in shallow grooves do so either at the ligand surface or at the periphery of the protein–ligand interface.

For many of these water molecules, the surface areas of the protein and the ligand exposed to the same water molecule are nearly equal. Water molecules binding in shallow grooves are found to have zero to two polar contacts with the protein and are not particularly well conserved within families of homologous proteins.

In general, the authors conclude that water molecules that are to be considered as part of the protein binding site during the design of a new ligand are those that bind in deep grooves, making multiple hydrogen bonds to protein atoms. These water molecules tend to be conserved through families of homologous proteins. The amino-acid residues that interact with deep-groove water molecules tend to be more conserved compared with other residues interacting with the ligand. Conversely, the binding of water in shallow grooves does not seem to be influenced by any special general feature of the protein or ligand surface, and it would be difficult to select water molecules *a priori* for inclusion as part of the protein structure during the process of ligand design.

23.4.4. Water structure in groups of well studied proteins

The analysis of general features of protein–water interactions derived from large databases provides an important context for the study of solvent structure in individual proteins. The number of crystallographically visible water molecules in any one X-ray structure depends on the resolution of the data, the degree of refinement of the model, the criteria used for placement of the less well defined water molecules, and on the experience of the crystallographer. Therefore, to differentiate between water molecules that have functional roles and those that associate randomly with the protein, it is desirable to determine commonalities between several independently solved structures of the protein of interest. There are different types of functional roles that can be determined at several levels. At the global level, one can find a small number of water molecules that are essential for the structural architecture common to a given family of homologous proteins. There are also those water molecules that are structurally important for a specific protein, being present in all independently solved structures of that protein, regardless of the crystal form in which the water molecule was determined or of its interactions with ligands. Water molecules that consistently appear in crystal structures of the protein solved in a specific space group but in no others may be important for crystal packing, but not to the integrity of the protein itself. Finally, a given water molecule may be essential for mediating in a protein–ligand complex, but never appear in the native protein. At this level, all of the independently solved structures of the complex would have the water molecule present. In the examples that follow, comparative analysis between carefully selected groups of structures reveals conserved water molecules at all of these different levels and shows how they carry out particular functional roles in specific examples.

23.4.4.1. *Crystal structures of homologous proteins*

There are two families of homologous proteins for which extensive solvent-structure comparisons have revealed water molecules important in maintaining structural features common to all members of the family. In the first study presented here, 35 crystal structures of eight members of the serine protease family were analysed (Sreenivasan & Axelsen, 1992), while the second study comprises a similar analysis of 11 independently solved structures of six members of the legume lectin family (Loris *et al.*, 1994).

23.4.4.1.1. *Serine proteases of the trypsin family*

The serine proteases have an especially large number of buried water molecules. Using a probe sphere of radius 1.4 Å, an iterative
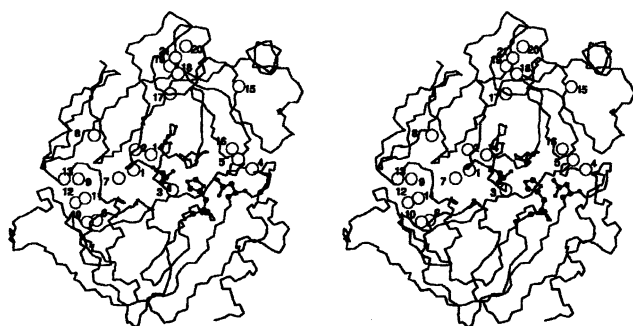
Fig. 23.4.4.1. Stereoview of the set of 21 highly conserved buried waters in eukaryotic serine proteases. The trypsin backbone is represented as a stick drawing, with the catalytic triad at the centre (filled circles). Water molecules are represented as open circles. Reprinted with permission from Sreenivasan & Axelsen (1992). Copyright (1992) American Chemical Society.

procedure was used to delete all accessible surface waters for each structure of chymotrypsin, chymotrypsinogen, trypsin, trypsinogen, elastase, kallikrein, rat tonin and rat mast cell protease. A total of 58 non-equivalent sites containing buried water molecules were found in the 35 crystal structures included in the study. Of these, 16 sites were common to all of the structures, with five additional sites common to proteins sharing the primary specificity of trypsin. A protein environment was defined for each of these 21 water sites to consist of the set of non-hydrogen protein atoms within 5 Å of the water oxygen atom. There is an average of 29 protein atoms per buried water molecule. Of these, 87% consist of main-chain atoms or conserved amino-acid side-chain atoms. The highly conserved nature of the amino-acid residues lining these water-binding sites suggests that the corresponding water molecules are important components of the protein tertiary structure and are likely to be present in all of the members of the trypsin family of serine proteases (Sreenivasan & Axelsen, 1992). Proteins in this family have two β-sheet domains, with the active site in the cleft between these domains. A large portion of the conserved buried water molecules occur in this cleft, mediating the interaction between the domains (Fig. 23.4.4.1). Conserved buried water molecules in other areas are found to bridge secondary-structure elements. These water molecules have been analysed extensively for elastase and are discussed in more detail below (Bellamacina *et al.*, 1999).

23.4.4.1.2. *Legume lectin family*

Whereas the study on serine proteases described above focused on the buried water molecules, the study on the legume lectin family included all of the conserved water molecules in the first hydration sphere. A total of 11 crystal structures were superimposed, many of them containing two independently refined monomers, making a total of 21 crystallographically independent monomers (Loris *et al.*, 1994). The six different proteins in the family (lentil lectin, pea lectin, *Lythyrus* lectin, *Griffonia* isolectin IV, *Erythrina* lectin and concanavalin A) have sequence identities ranging from 100% to 40%. Water molecules in two superimposed crystal structures were considered to occupy the same site if they were within a predefined distance of 1 Å from each other. Seven water sites were found to be conserved in all of the family members included in the study. Four of these interact with the manganese and calcium ions, and one is in the ligand-binding site. The other two stabilize secondary structures: a β-hairpin turn and a β-bulge. In all cases, the protein composition of the site was strictly conserved. A larger number of water molecules are conserved within groups of closely related members of the family. The majority of these sites are found in the interface between the two monomers that come together to form a

continuous 12-stranded β-pleated sheet and around the metal and monosaccharide binding regions (Fig. 23.4.4.2). Three crystal forms of lentil lectin were available for the study, and it was observed that of the 33 water molecules conserved between the corresponding three structures, none are involved in crystal contacts.

If one could generalize from the two studies described above, the conclusion would be that the water molecules strictly conserved across families of homologous proteins are found either at the binding site, at the interface between domains, or bridging secondary-structure elements which would otherwise not be part of the well defined protein architecture. Furthermore, it is clear that evolutionary pressure exists to maintain the composition of the amino-acid residues with which these crucial water molecules interact at their respective protein binding sites. A more recent study of conserved water molecules in a large family of microbial ribonucleases confirms the conclusions obtained in the two studies presented here (Loris *et al.*, 1999).

23.4.4.2. *Multiple crystal structures of the same protein*

Although not many studies have focused on the conserved water molecules across families of homologous proteins, there is currently a considerable amount of information on solvent structure based on groups of independently solved crystal structures of a specific protein. The comparison of multiple crystal structures is important to distinguish between the different roles played by water molecules on protein surfaces and to obtain a more complete picture of the first hydration sphere. In any one crystal structure of a given protein, it is extremely likely that the water molecules crucial to the structure or function of the protein will be seen in the electron-density map. However, the water molecules more loosely associated with the protein surface appear fortuitously in one or few structures, so that with every new structure one finds a series of water molecules not previously observed. A clear example of this is provided by a collection of eleven elastase structures solved in different organic solvents, where of a total of 1661 water molecules there are 178 molecules that are unique to one of the structures (Mattos & Ringe, 1996; Bellamacina *et al.*, 1999; Mattos *et al.*, 2000).
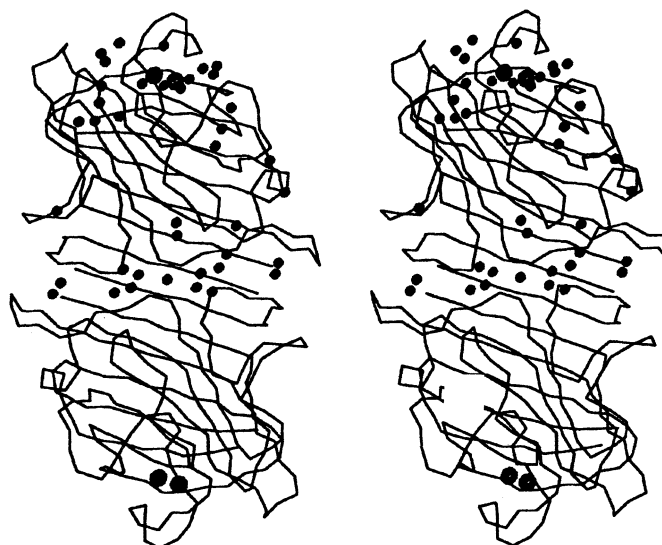

Fig. 23.4.4.2. View of the 33 conserved hydration sites in the lentil lectin crystal structures superimposed on the backbone of the lentil lectin dimer. In order to emphasize the twofold symmetry, the waters at the dimer interface are shown for both lectin monomers. Reprinted with permission from Loris *et al.* (1994). Copyright (1994) The American Society for Biochemistry & Molecular Biology.
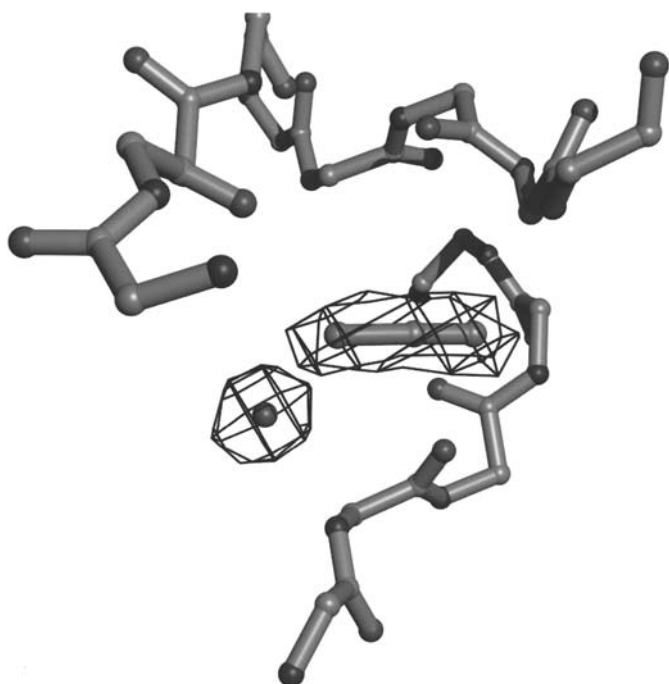
631

Fig. 23.4.4.3. A $2F_o - F_c$ electron-density map contoured at the 1.2σ level shows a distinct ellipsoidal density for acetonitrile 707 and a spherical density for a nearby water molecule. The protein backbone of the binding pocket is represented with nitrogen atoms shown in dark grey, oxygen atoms in medium grey and carbons in a lighter grey. *MOLSCRIPT* (Kraulis, 1991) was used in the preparation of this figure. Reprinted with permission from Allen *et al.* (1996). Copyright (1996) American Chemical Society.

23.4.4.2.1. *Elastase*

The crystal structure of porcine pancreatic elastase was solved in a variety of organic solvents, with the primary goal of mapping binding sites on the protein that could accommodate molecules representative of functional groups likely to be found in larger ligands (Ringe, 1995; Mattos & Ringe, 1996; Mattos *et al.*, 2000). Crystals of elastase cross-linked with glutaraldehyde were transferred to the following solutions: 100% acetonitrile, 95% acetone, 55% dimethylformamide, 80% ethanol, 40% trifluoroethanol, 80% isopropanol and 80% 5-hexene-1,2-diol (Allen *et al.*, 1996; Mattos & Ringe, 1996). In general, the crystals did not diffract in most neat organic solvents. However, in the acetonitrile case, where they did, the result was striking. In the structure of elastase solved in >99% acetonitrile, there were 126 water molecules visible in the electron-density maps, indicating that a good portion of the first hydration shell of the protein was still present. In contrast, only nine molecules of acetonitrile were clearly identified in the electron-density maps (Allen *et al.*, 1996). This is a powerful assertion of the evolutionary specificity of water molecules for protein surfaces. Fig. 23.4.4.3 shows the clear contrast between the elongated electron density of an acetonitrile molecule and the spherical electron density of a water molecule.

A similar result was obtained for all of the elastase structures solved in the mixtures of organic solvent and water mentioned above. A total of 11 structures were analysed, each containing 126–177 water molecules. The structures are listed in Table 23.4.4.1, together with the resolution of the data collected, the number of water molecules present and the number of organic solvent molecules observed in each case. The $C\alpha$ superposition of the protein atoms in the 11 structures yielded a total of 1661 individual water molecules, occupying 426 unique water-binding sites on the elastase surface. Given that elastase has a total of 240 amino-acid

residues, this represents a significant portion of the first hydration shell of the protein. This group of elastase structures served as a powerful source of information, leading to a classification of water types according to their interaction with the protein and an analysis of the specificity for water within each of the types determined (Bellamacina *et al.*, 1999).

All of the 1661 water molecules were renumbered according to the site on the protein where they were found. Any two water molecules within 1 Å of a water molecule in the cross-linked elastase structure solved in distilled water (used as the reference structure) have a common number. 39 of the 426 water-binding sites were occupied in every one of the 11 structures and were considered structurally conserved. Among these are the 16 buried water-binding sites thought to be conserved among all serine proteases (Sreenivasan & Axelsen, 1992). The 26 remaining conserved water molecules are specific to elastase and are not necessarily buried. These water molecules in general tend to have low B factors, but a few have B factors in the 30–35 Å^2 range and one conserved water molecule has a B factor of 42 Å^2.

The classification of the water sites as buried, channel, crystal contact or surface was based on the number of hydrogen-bonding interactions that a water molecule at the site could make to the protein and involved no surface-accessibility calculations (Bellamacina *et al.*, 1999). Water molecules were classified as buried if they made at least three good hydrogen-bonding interactions with protein main-chain atoms. A total of 23 buried water sites were identified in this manner, including 13 of the sites classified as buried by Sreenivasan & Axelsen (1992). One of the 16 serine protease conserved water-molecule sites is replaced by a His side chain in elastase (Sreenivasan & Axelsen, 1992). The remaining two serine protease conserved water sites were classified as channel based on the criteria used in the present study (see below). Interestingly, with the exception of these two channel water molecules, all of the buried sites found to be conserved in serine proteases are strictly conserved in all of the 11 structures in Table 23.4.4.1. The two channel water molecules are found in the aqueous structures of elastase, but are virtually absent in elastase transferred to organic solvents.

The water molecules occupying the 23 buried water sites identified in this study are tightly clustered when the protein $C\alpha$ atoms are superimposed by least squares, and the interactions with the protein are conserved from structure to structure. Fig. 23.4.4.4 shows the positions of the buried water-binding sites in elastase. In general, they are found in the cleft between the two domains, in bridging elements of the secondary structure and at the base of water channels. This observation is consistent with the current

Table 23.4.4.1. *Multiple-solvent crystal structures of elastase*

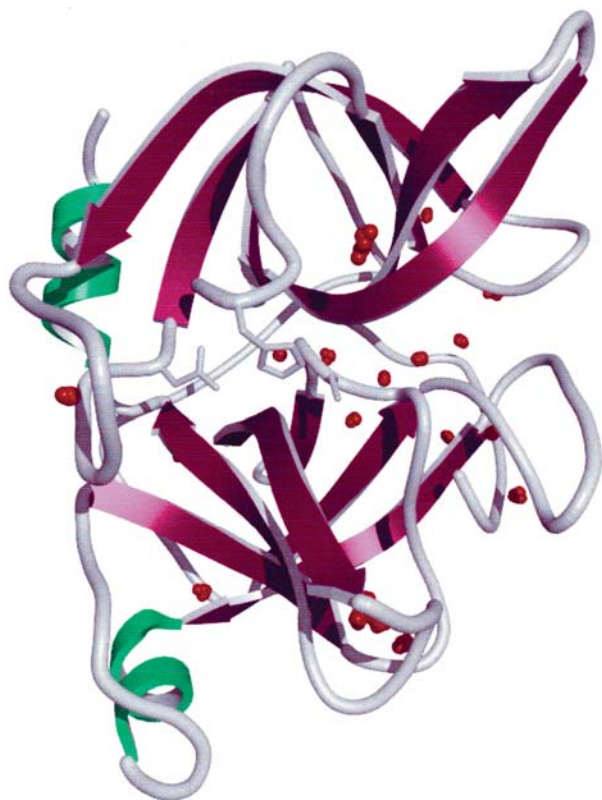| Structure	Resolution (Å)	No. of water molecules	No. of organic solvent molecules
Cross-linked	1.9	165	0
Acetonitrile	2.2	126	9
Acetone	2.0	126	6
Dimethylformamide	2.0	153	6
Ethanol	2.0	135	12
Trifluoroethanol (1)	1.9	175	4
Trifluoroethanol (2)	1.85	177	3
Isopropanol	2.2	160	4
Benzene	1.9	162	4
Cyclohexane	1.95	135	7
5-Hexene-1,2-diol	2.2	147	5

Fig. 23.4.4.4. Crystal structure of porcine pancreatic elastase represented as a ribbon diagram using *MOLSCRIPT* (Kraulis, 1991). The two α-helices are shown in green, the β-sheets are in purple and the coils are in grey. Elastase contains 240 amino-acid residues, and is composed of two β-barrel domains. The catalytic triad (Asp108, His60 and Ser203) is shown explicitly. The buried crystallographic water molecules found in 11 superimposed elastase structures solved in a variety of solvents are shown in red.
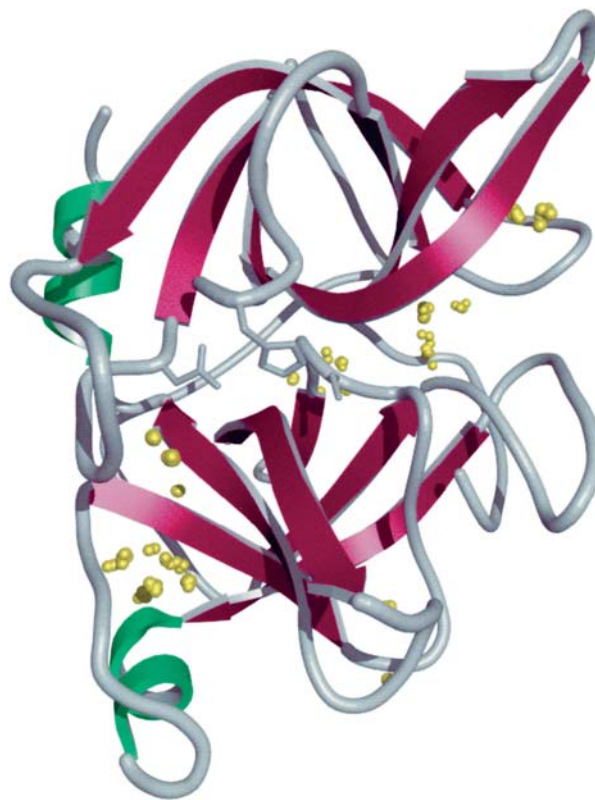
Fig. 23.4.4.5. Elastase structure represented as in Fig. 23.4.4.4. The crystallographic water molecules found in channels in 11 superimposed elastase structures solved in a variety of solvents are shown in yellow.

understanding of the functional roles played by structurally conserved water molecules as discussed above and in the following sections.

The 29 water-binding sites classified as channel contain water molecules that make hydrogen bonds with at least two other water molecules within a protein groove. The analysis of a high-resolution crystal structure of elastase (1.65 Å) revealed seven channels with a total of 32 water-binding sites (Meyer *et al.*, 1988). All of these channels were also identified in the analysis of the 11 structures in Table 23.4.4.1 (Bellamacina *et al.*, 1999). In addition, two other channels were observed. The locations of the nine elastase channels identified by the new criteria are shown in Fig. 23.4.4.5. Channels are often found in areas associated with buried water molecules, namely, at the crevice between the two domains and sandwiched between secondary-structure elements, where they lead from the surface of the protein to a buried water molecule. Fig. 23.4.4.5 also shows that the Cα superposition of the protein structures leads to a spread of water molecules within the channels. In any given structure, only two or three water molecules may be present, but the precise location and interaction with protein atoms vary so that when taken together the collection of structures gives a sense of flow inside the channels.

Of the remaining 374 water-molecule sites present within the 11 elastase structures included in this study, 56 were classified as crystal-contact sites and 318 as surface sites. Crystal-contact sites were considered to be occupied by water molecules that are within 4.0 Å of a symmetry-related protein molecule in the crystal. Fig. 23.4.4.6 shows the position of all the water molecules found to occupy these sites. The relatively large number of crystal-contact water-binding sites is a result of the somewhat broad criterion used to select them. Many of these sites are not within hydrogen-bonding distance from the nearby protein molecule, and most are not well conserved from structure to structure. Only eight of the 56 sites are occupied in the majority of the structures, and four of these make good multiple hydrogen bonds with two symmetry-related protein molecules in the crystal. These four water molecules seem to be structurally significant in the formation of the crystal contacts.

Surface water molecules were taken to be those that interact with side-chain protein atoms on the surface or make no more than two hydrogen-bonds with backbone atoms. When the 11 structures are superimposed, the surface water molecules occupying a given site are not tightly clustered. Furthermore, there is flexibility in the interactions between these water molecules and the nearby protein atoms. For example, it is often the case that all water molecules within a surface site make two or three hydrogen bonds to protein atoms, but only one of them is conserved in all of the structures where the water molecule is present at the site. Fig. 23.4.4.7 illustrates the position of all of the surface water-binding sites. Although over half of these sites are occupied in at least two of the 11 structures, a good proportion of them (178) are found in only one of the structures considered.

While crystal-contact and surface water sites were classified separately, it is important to point out that, with the exception of the four crystal-contact water-binding sites mentioned above, the crystal-contact sites exhibit very much the same traits as the surface water sites. The difference is that in the latter case, the 'surface' is provided by a single protein molecule, while in the former the interaction between two symmetry-related protein molecules constitutes the surface with which the water molecules interact.

Of the 318 surface water molecules, 21 are in the active site. The active-site water molecules were selected to be those within 4 Å of any atom belonging to either the trifluoroacetyl-Lys-Phe-*p*-
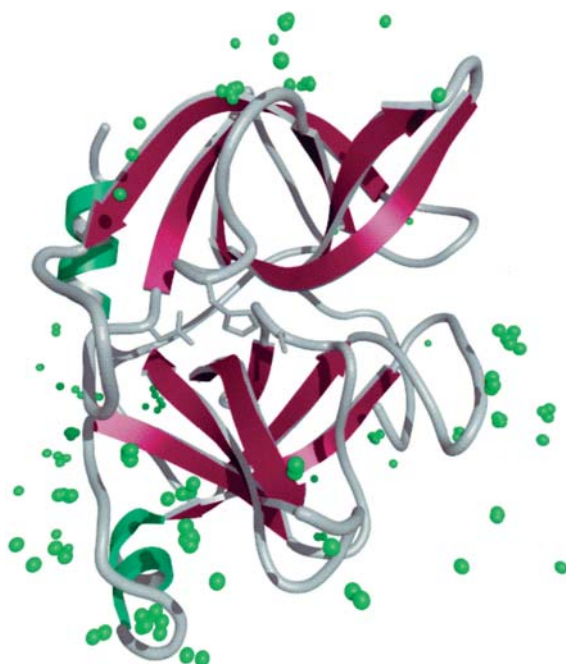
Fig. 23.4.4.6. Elastase structure represented as in Fig. 23.4.4.4. The crystallographic water molecules involved in crystal contacts in 11 superimposed elastase structures solved in a variety of solvents are shown in green.

isopropylanilide (Mattos *et al.*, 1994) or the trifluoroacetyl-Lys-Pro-*p*-trifluoromethylanilide (Mattos *et al.*, 1995) inhibitors in the structures of their complexes with elastase. These inhibitors span a large area of the active site, including an exosite not occupied by

substrate analogue inhibitors (Mattos *et al.*, 1994, 1995). The water-binding sites in the active site are not very well conserved, with most sites represented in only two to four of the 11 structures. When all of the structures are superimposed, there is at least one water molecule in each of the subsites in the elastase active site. These water molecules are displaced either by inhibitors or by organic solvent molecules in the various structures. It is not surprising that in elastase, a protein with relatively broad substrate specificity, the active site in the uncomplexed native protein is populated by many displaceable surface water molecules. With the exception of a water molecule present in the oxyanion hole, these water molecules tend to make a single hydrogen bond with the protein. This hydrogen-bonding interaction is not generally conserved between different structures where a given site is occupied in multiple structures. The displacement of these water molecules upon ligand binding is entropically favourable, as they are released into bulk solvent, without too much enthalpic cost. This relatively small enthalpic cost can be compensated by the protein–ligand interactions.

Fig. 23.4.4.8 shows all of the 1661 water molecules colour-coded by the various classifications described above. Clearly, the entire surface of the protein is well hydrated. Notice how the yellow channel waters are often followed by a red buried water molecule. In addition, there is often no obvious spatial distinction between molecules categorized as crystal contacts (green) and those categorized as surface (blue).

23.4.4.2.2. *T4 lysozyme*

Over 150 mutants of T4 lysozyme have been studied to date, and, for the majority of these, the crystal structures are available. Although most of the mutant structures crystallize isomorphously to the wild type, many of them provide a view of the molecule in different crystal environments. This collection of structures leads to
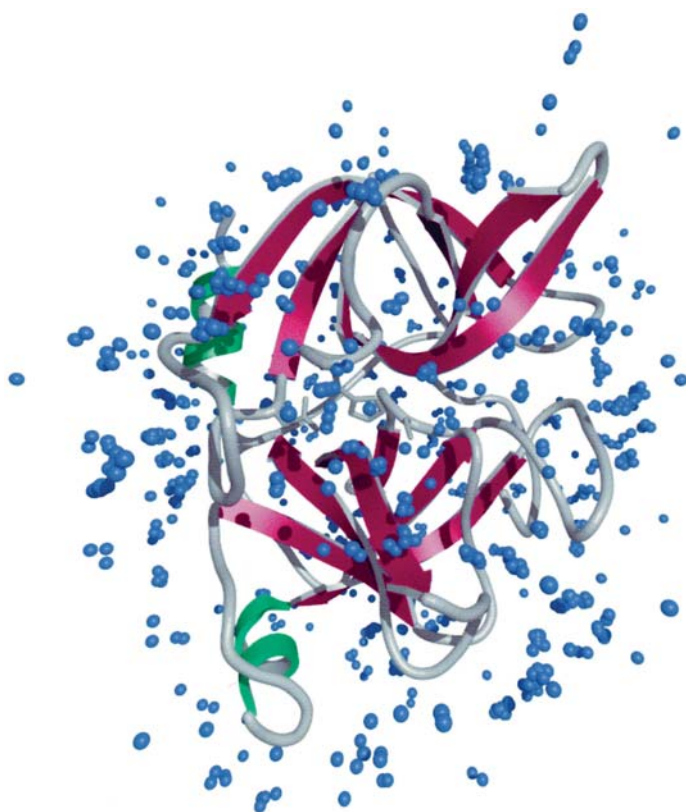

Fig. 23.4.4.7. Elastase structure represented as in Fig. 23.4.4.4. The surface crystallographic water molecules found in 11 superimposed elastase structures solved in a variety of solvents are shown in blue.
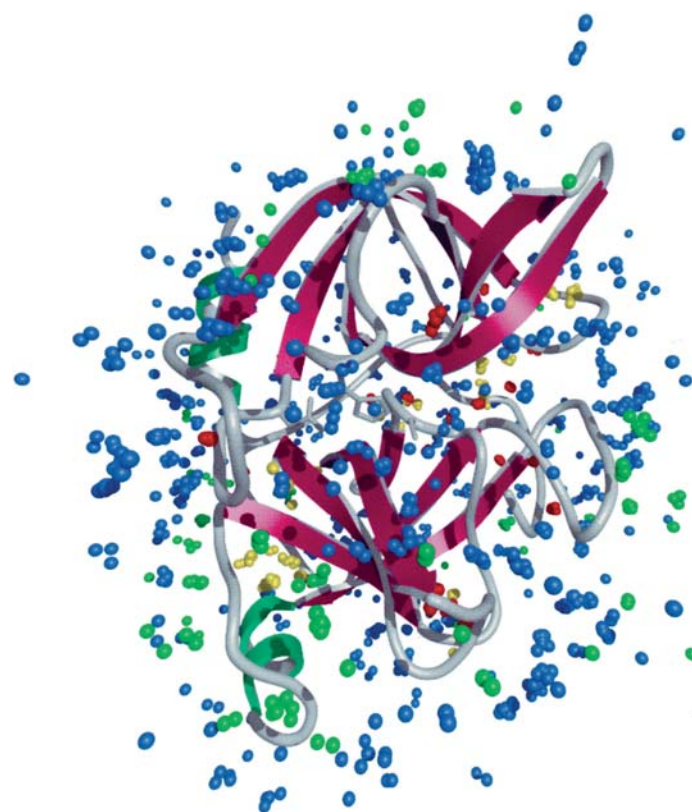

Fig. 23.4.4.8. Elastase structure represented as in Fig. 23.4.4.4. The 1661 water molecules found in 11 superimposed elastase structures of elastase are colour-coded as in Figs. 23.4.4.4–23.4.4.7.

the comparative analysis of the solvent positions in ten different crystal forms of T4 lysozyme, providing a clear picture of the effect of crystal contacts on the hydration sphere of a protein viewed by X-ray crystallography (Zhang & Matthews, 1994). The resolution and degree of refinement of the structures involved varied significantly, from 2.6 to 1.7 Å resolution, and the number of water molecules included per protein molecule ranged from 38 to 160. Nevertheless, this study revealed important features. A striking observation is that 95% of the solvent-exposed residues on T4 lysozyme were involved in at least one crystal contact in one or another of the crystal forms studied, showing that any part of the protein surface can be involved in crystal contacts. A corollary to this finding is that any of the surface water molecules can be displaced or involved in bridging protein–protein contacts in the crystal.

Of the 1675 individual water molecules observed in the 18 independently refined T4 lysozyme molecules included (Fig. 23.4.4.9), the ones that were within a sphere of radius 1.2 Å were considered to occupy the same site on the protein. As in the case of elastase described above, all of the water molecules observed upon superposition of the 18 T4 lysozyme structures represent a large portion of the first hydration shell. This reinforces the concept that multiple structures of a protein of interest provide a more complete picture of the protein hydration than possible with a single structure. There are four buried water sites that are occupied in at least 15 out of the 18 structures and are independent of crystal contacts. Two of these buried sites are at the hinge-bending region between the two helical domains and appear to play a functional role in the opening and closing of the active site (Weaver & Matthews, 1987). The other two play a structural role at the protein core. Other than the four buried water molecules, the most conserved water sites appear at the active-site cleft between the two domains and at the N-termini
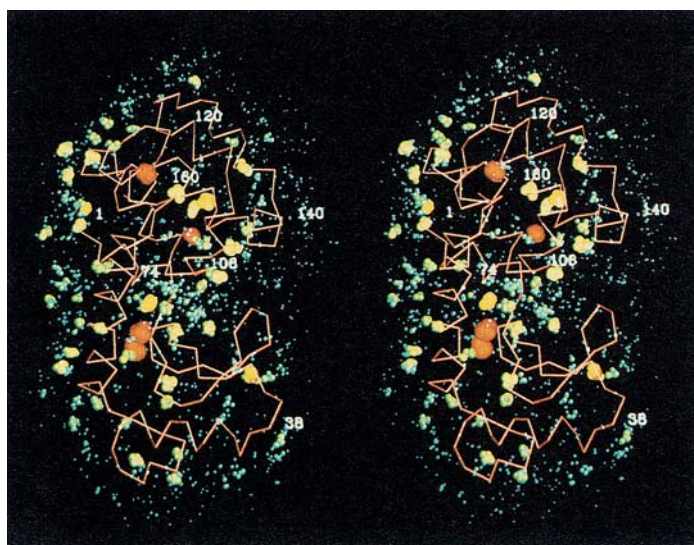

Fig. 23.4.4.9. Distribution of solvent-binding sites in 18 mutant T4 lysozymes from ten refined crystal structures. The lysozyme structures were compared to identify common sites of hydration. A total of 1675 solvent molecules were included in the comparison. Each solvent molecule is represented by a coloured sphere. The size of the sphere is proportional to the number of lysozyme structures in which solvent was observed at the same site (*i.e.* within 1.2 Å). In addition, the colour of the solvent changes from blue for the least-conserved sites to red for the most-conserved ones [*e.g.* the red spheres indicate that solvent is observed with high frequency (15–17 times) at the four internal sites]. The numbers indicate representative residue positions along the backbone of the lysozyme molecule. Reprinted with the permission of Cambridge University Press from Zhang & Matthews (1994). Copyright (1994) The Protein Society.

of α-helices. As is the case in the previous works reviewed above, the 20 most conserved water sites appear in well conserved protein environments and generally have low temperature factors. Buried or highly conserved water molecules also tend to make at least three hydrogen bonds with protein atoms or other water molecules. The less-conserved water sites appear more randomly on the protein surface and are strongly influenced by the particular crystal environment in which the structure was solved.

23.4.4.2.3. *Ribonuclease T1*

A group of four crystal structures of ribonuclease T1 in complex with guanosine, guanosine-2′-phosphate, guanylyl-2′,5′-guanosine and vanadate were used for an analysis of conserved water positions that contribute to the structural stabilization of the protein (Malin *et al.*, 1991). The four structures were obtained from isomorphous crystals and ranged in resolution from 1.7 to 1.9 Å. Conserved water molecules were considered to be those found within a sphere of 1 Å from each other in all four structures. All other water molecules were excluded from the analysis. 30 water molecules were found to be conserved. Of these, ten were observed near crystal contacts, although only one appears to be dictated by the crystal contact itself, making a single hydrogen bond with each of the symmetry-related protein molecules. Ten other water molecules form a channel that brings together an α-helix and a hairpin-like loop structure and then go on to wrap around the calcium ion, providing half of its coordination sphere. The first five of these water molecules are completely buried, holding together the two secondary-structure elements, which would otherwise collapse (Malin *et al.*, 1991). Two water molecules are found to stabilize the N and C termini, which are brought together by a disulfide bond. The remaining eight conserved water molecules hold together various elements of secondary structure or are located in the active site.

An interesting extension to this study included four additional structures: the E58A mutant in complex with guanosine-2′-monophosphate, the H92A mutant crystallized under two different conditions and wild-type RNase T1 in complex with guanosine-3′,5′-biphosphate. Two of these crystal forms were not isomorphous with the native protein crystals or with each other. Thus a total of eight structures solved in three different space groups were analysed (Pletinckx *et al.*, 1994). Although the effect of crystal packing on the three-dimensional structure of the protein is minimal, there are some significant differences in the solvent structure. In particular, there is no evidence of the calcium-binding site and its coordinating water structure in any crystal forms other than the canonical wild type. Instead, the E58A mutant has a sodium-binding site at a different position, along with three previously unobserved water molecules. It is clear that the presence of the metal ions is fortuitous and linked to the crystallization conditions.

There are 25 water molecules structurally conserved throughout the different packing arrangements studied. Ten of these are single sites, there are three clusters of two water molecules and a larger cluster originally described by Malin *et al.* (1991) to hold together the core of the protein. As was observed for the study on T4 lysozyme (Zhang & Matthews, 1994), the strictly conserved water-binding sites present in crystal structures solved across different space groups are involved in bridging protein secondary-structure elements and seem to be crucial for the integrity of the protein structure.

23.4.4.2.4. *Ribonuclease A*

Ribonuclease A is not homologous to ribonuclease T1 in either sequence or structure, but both have evolved to catalyse the same reaction with specificity for different substrates (compare Figs. 23.4.4.10 and 23.4.4.11). Ribonuclease A cleaves RNA after pyrimidines, while ribonuclease T1 cleaves specifically after
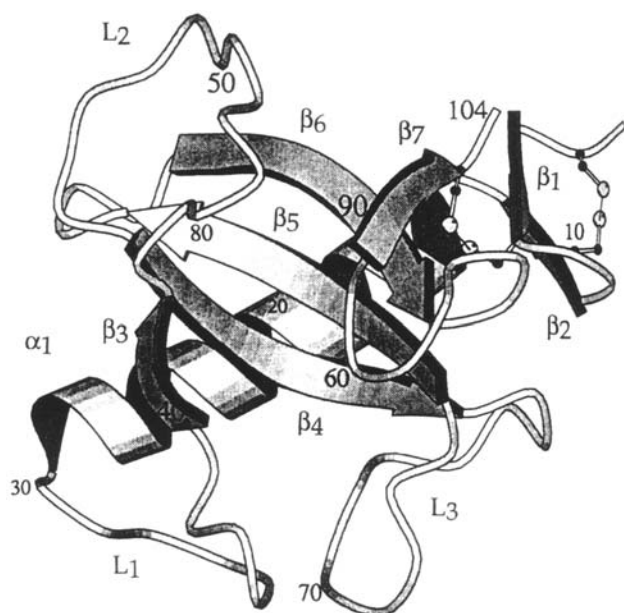
Fig. 23.4.4.10. Three-dimensional structure of RNase T1. Secondary structure is denoted as follows: α_1, α-helix; β_n, strands of β-sheet structure; L_n, loops. Drawn using *MOLSCRIPT* (Kraulis, 1991). Residue numbers indicate the beginning and end of secondary-structure elements. Reprinted with permission from Pletinckx *et al.* (1994). Copyright (1994) American Chemical Society.

guanine. Therefore, the information obtained from a study of the solvent structure in ribonuclease A is completely independent from that described above for ribonuclease T1. A collection of ten crystal structures of ribonuclease A, derived from five different crystal forms, were compared pairwise after least-squares superposition (Zegers *et al.*, 1994). 17 conserved water molecules were found to be within a sphere of 0.5 Å of each other in all of the ten structures
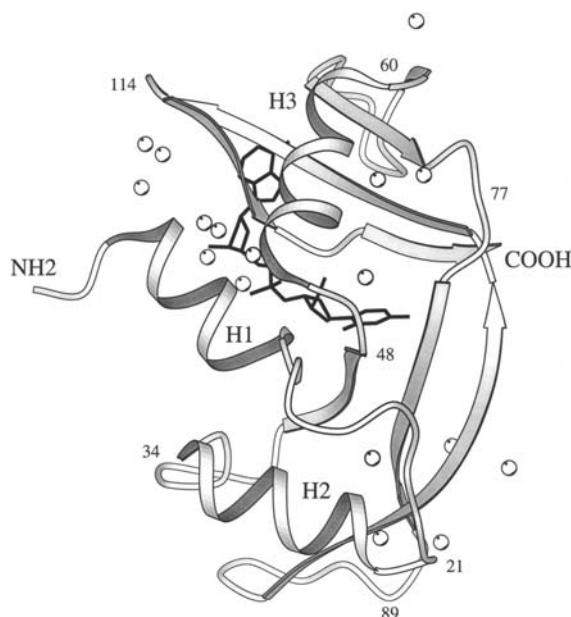

Fig. 23.4.4.11. Overall structure of RNase A. The overall structure of the d(CpA) complex of RNase A is shown as a ribbon drawing using *MOLSCRIPT* (Kraulis, 1991). The conserved water molecules are shown as white spheres and the d(CpA) inhibitor in black. The three helices are labelled H1, H2 and H3. Reprinted with the permission of Cambridge University Press from Zegers *et al.* (1994). Copyright (1994) The Protein Society.

and are shown in Fig. 23.4.4.11. These water molecules were found in small clusters of two or three or as part of a larger solvent network. Not surprisingly, they form multiple hydrogen bonds with the protein and generally have low temperature factors. Of the 17 structurally conserved sites, 13 are associated with one of the three α-helices. Most of these link the helices to one of the β-strands. Three water molecules are involved in hydrogen bonding with unpaired amido and carbonyl groups on the protein, and one is found on top of the β-pleated sheet. These interactions result in bringing together elements of secondary structure and in stabilizing distortions within these elements. Conserved water molecules are also responsible for bridging the N-terminal helix to the C-terminal β-strand, which form the two halves of the active site.

23.4.4.2.5. *Protein kinase A*

The comparative study of water molecules in seven different protein kinase A structures in complex with different ligands focused exclusively on the active site (Shaltiel *et al.*, 1998). All of the structures were solved from isomorphous crystals to resolutions ranging from 2.0 to 2.9 Å. The more lenient cutoff of 1.5 Å for the radius of the sphere within which the conserved water molecules must be found among the different structures is consistent with the relatively low resolutions of the structures included in this study. The group of structures represents the open, the closed and an intermediate conformation of the catalytic kinase domain. There is a set of six conserved water sites in the active site, in addition to the ATP molecule and the magnesium ion. The conserved water molecules coordinate to ATP, the metal ion and a conserved Tyr residue from the carboxyl terminus of the protein. Thus, the active site consists of an extended network of interactions that weave together both domains of the core, with water molecules playing an integral role in maintaining the structural features important for catalysis. Many of these water molecules associate directly with the inhibitors. In addition, five water sites are observed in positions that would be occupied by substrates or substrate analogues. These water molecules are displaced by ligand oxygen atoms that can compensate for the water hydrogen-bonding interaction with the protein.

23.4.4.3. *Summary*

Water molecules associated with proteins can be divided between those that are conserved as a result of their functional significance and those that are partially conserved or not conserved at all. The conserved water molecules are generally classified as buried or channel (by a variety of criteria). They tend to be present in the clefts between domains, are critical components of active sites, or bridge between secondary-structure elements. The water molecules that are not conserved occupy hydration sites with favourable hydrogen-bonding characteristics, where the presence of a water molecule is not essential for the structural or functional integrity of the protein.

The displacement of water molecules by organic solvent molecules in the elastase work described above showed that most displaced waters are those classified as surface or crystal-contact waters (Mattos *et al.*, 2000). In the three cases where a buried water molecule was displaced, an alcohol hydroxyl oxygen was found to replace the protein–water hydrogen-bonding interactions. This is analogous to the active-site water molecule in the HIV aspartate protease that gets replaced by a carbonyl group of a potent cyclic urea inhibitor (Lam *et al.*, 1994). In these situations, release of a tightly bound water molecule is entropically favourable, and its enthalpic interactions with the protein are compensated by similar protein–ligand interactions.

The effect of crystal contacts on the water structure was clearly illustrated in the T4 lysozyme work (Zhang & Matthews, 1994).

The internal structurally conserved water molecules are unaffected by crystal contacts. Conversely, any of the surface water sites are potentially available either to be replaced by or to mediate crystal contacts, as 95% of the T4 lysozyme surface is involved in a crystal contact when all ten crystal forms are taken together.

23.4.5. The classic models: small proteins with high-resolution crystal structures

Crambin and BPTI are among the handful of proteins for which X-ray crystal structures have been obtained to 1 Å resolution or better. In general, these proteins are relatively small (BPTI, the largest in this group, has 58 amino-acid residues) and often contain at least one disulfide bond. These high-resolution crystal structures have provided structural information beyond that available for larger proteins, particularly with respect to the surface solvent structure. Their small size renders their structures accessible by NMR techniques, making it possible to assess the effect of the crystal environment on the protein and water structure. Finally, the available detail and precision of the structures, as well as their small size, make them ideal models in computational studies of protein energetics and dynamics. Both crambin and BPTI were used during the pioneering years of protein molecular-dynamics calculations. In this section, special attention is given to crambin and BPTI as representative proteins for which very high resolution structures are available. Focus is on the features of solvent structure that are not available for other proteins.

23.4.5.1. *Crambin*

Crambin is a plant-seed hydrophobic protein of unknown function. It contains 46 amino-acid residues and was reported to form crystals that diffract to 0.88 Å resolution (Teeter & Hendrickson, 1979). The crystal structure of crambin was determined to 0.945 Å resolution directly from anomalous scattering by the six sulfur atoms involved in three disulfide bonds (Hendrickson & Teeter, 1981). Crambin is an amphipathic molecule in that the hydrophilic components (including six charged groups) are segregated from a mainly hydrophobic surface.

A total of 64 water molecules and two ethanol molecules were located in the electron-density map, despite the fact that the structure was determined in 60% ethanol. The overwhelming number of water molecules compared to ethanol is consistent with the results of the multiple-solvent crystal structures experiments described above for elastase (Mattos & Ringe, 1996).

Most of the 64 water molecules found in crambin interact with polar side chains in the typical manner described previously. The unusual information about solvent structure offered by the crambin model is that the arrangement of water molecules around hydrophobic residues is similar to that observed for clathrate hydrate structures (Teeter, 1991). Pentagonal water rings are observed to cap the Cδ2 atom of Leu18 as well as the hydrophobic methylene groups of Arg17 (Teeter, 1984, 1991). The set of five connected water rings is shown in Fig. 23.4.5.1. This ring cluster extends toward the protein, forming heterocyclic rings that are described in detail in the original article (Teeter, 1984).

Although crambin provides the clearest example of pentagonal water rings on a hydrophobic protein surface, it is not the only one. Other high-resolution crystal structures (better than 1.4 Å), such as insulin and cytochrome *c*, have also revealed pentagonal rings, but never to the extent seen in crambin (Teeter, 1984). This is very likely to be a general mode of interaction between water and hydrophobic moieties, be it in inorganic, organic, or biological molecules. The fact that it is not observed in protein structures in general may be related to the lower resolution of most X-ray
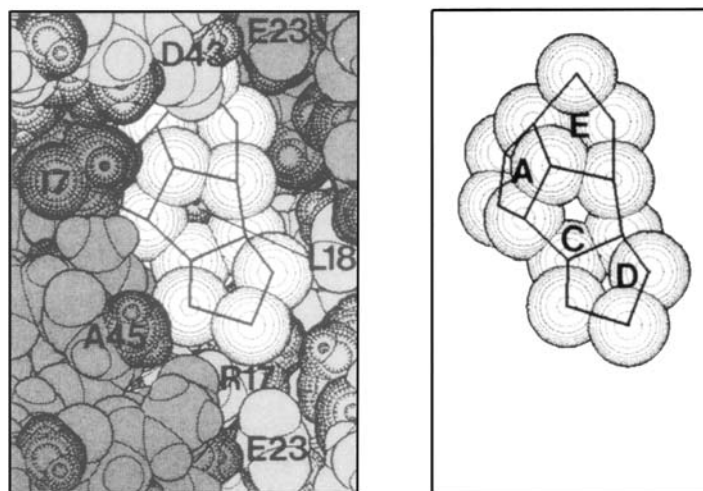

Fig. 23.4.5.1. van der Waals surface diagram of the water pentagons A, C, D and E in crambin viewed in the negative **a** direction. Rings A, C and E form a cap around leucine 18. Hydrophobic atoms are shown as dark circles, and water oxygens are shown as light circles. The methyl group of leucine 18 can be seen through the C ring. Adjacent translationally related molecules are shaded. The van der Waals radii used for the protein C, N and O atoms are 1.7, 1.4 and 1.4 Å, respectively, and for water oxygen, 1.8 Å. The larger radius is used for the water oxygens because hydrogen atoms have been omitted. Reprinted with the permission of the author from Teeter (1984).

structures, where it is not possible to model the more disordered areas where these patterns are likely to be found.

23.4.5.2. *Bovine pancreatic trypsin inhibitor*

Bovine pancreatic trypsin inhibitor (BPTI) is a protein of 58 amino-acid residues whose X-ray crystal structure was obtained in the original crystal form to 1.5 Å resolution (Deisenhofer & Steigemann, 1975). Subsequently, 1 Å X-ray data were obtained from a different crystal form, and the new model was jointly refined with 1.8 Å neutron diffraction data (Wlodawer *et al.*, 1984). Minor differences in structure between the two crystal forms of BPTI were observed (Wlodawer *et al.*, 1984). The interesting contribution of the 1 Å model to the understanding of solvent structure resulted from the ability to refine occupancy at this resolution. A total of 63 water molecules were placed in the model, 20 of them within 1 Å of a water molecule found in the structure solved in the original crystal form. During refinement against the 1 Å data set, full occupancy was assigned to all protein atoms, and water occupancy was allowed to refine. Of the 63 water-molecule positions, 29 were found to be fully occupied. The remaining 34 had partial occupancies, with 0.4 being the minimum occupancy found. Given that there are very few contacts between protein molecules in the crystal (Wlodawer *et al.*, 1984), it is reasonable to assume that this observation is representative of water occupancies on protein surfaces in general. It is likely that well over half of the water positions found on protein surfaces are less than fully occupied, although there is no definitive proof that this is true.

23.4.5.3. *Summary*

In general, small proteins serve as important models where results of X-ray crystallography, NMR and molecular-dynamics calculations can be easily compared and cross-validated, since larger proteins are more difficult to study by the latter two techniques. Small proteins are also more likely to form relatively ordered crystals, which are able to diffract X-rays to atomic resolution (of the order of 1 Å). With respect to understanding

solvent structure, the two major contributions of these very high resolution protein models are the observation of solvent structure around hydrophobic residues, where at lower resolution the water molecules 'look' disordered, and a glimpse at the pattern of water occupancy likely to occur on protein surfaces.

23.4.6. Water molecules as mediators of complex formation

The examples given in Section 23.4.4 illustrate two important roles played by water molecules at the binding sites of proteins: as structural water molecules and as displaceable water molecules. As a structural part of a binding site, water molecules are found to be strictly conserved. They are either involved in stabilizing the coming together of secondary-structure elements in a way that appropriately shapes the binding site, or they fill grooves on the protein surface, making it more specific for a given ligand. The second role involves the presence of less tightly bound, partially conserved water molecules that get displaced by the ligand upon binding. In the few examples where tightly bound water molecules are displaced by a ligand, the hydrogen-bonding interaction of the water with the protein is replaced by an atom on the ligand. A third role, not yet discussed, of water molecules in protein active sites is in the catalytic mechanism of enzymatic reactions. An extensive network of water molecules near the active site of serine proteases has been implicated in the catalytic mechanism of these enzymes (Meyer *et al.*, 1988; Meyer, 1992). If this hypothesis is indeed correct, it provides a good example of the cooperation between water molecules and protein atoms in the optimization of function. Unfortunately, it is difficult to explicitly detect catalytic water molecules crystallographically, due to the long data-collection time relative to a catalytic event. However, the development of time-resolved Laue diffraction methods has provided a view of the catalytic water molecule in some proteins, *e.g.* trypsin (Singer *et al.*, 1993), and progress is likely to continue in this area. This section focuses on a few particular examples of how water molecules mediate the formation of complexes, either in the active sites of enzymes or in the binding interface between macromolecules or protein–ligand complexes.

23.4.6.1. *Antigen–antibody association*

The X-ray crystal structures of the Fv fragment of the monoclonal antibody D1.3 and the structure of its complex with hen egg-white lysozyme were both solved to 1.8 Å resolution (Bhat *et al.*, 1994). This study revealed a significant number of water molecules contributing to the chemical complementarity at the antigen–antibody interface. There are 23 water molecules at the antigen-binding site of the free antibody fragment, while 48 are present mediating complex formation. Seven water molecules are in equivalent positions in the free and complexed antibody (within 1.5 Å). There is no net loss of water molecules at the combining site. In fact, the total number of water molecules at the antigen–antibody interface is not less, but more, than the sum of those in the free antibody combining site and in the antigenic determinant. Furthermore, there is a general decrease in *B* factors of the binding-site residues upon complex formation, implying a decrease in entropy (Bhat *et al.*, 1994). The structural results indicate that water molecules at the antigen–antibody interface play a variety of important roles. Some form an integral part of the active site, fine-tuning the shape and charge complementarity of the interaction. Others are found to be displaced during complex formation, and still others are unique to the complex, bridging between the two molecules in a variety of locations throughout the complex interface.

The structural analysis correlated well with results of calorimetric experiments that showed that complex formation is enthalpically driven, with an unfavourable entropic contribution (Bhat *et al.*, 1994). The authors suggest that water molecules play a central role in mediating complex formation and claim that the hydrophobic effect is not important in this case. This is an argument that goes contrary to the idea that affinity is contributed by hydrophobic interactions within a relatively small portion of the interface between the interacting molecules, with hydrogen-bonding and charge–charge interactions contributing primarily to specificity (Hendsch & Tidor, 1994; Clackson & Wells, 1995; Hendsch *et al.*, 1996).

23.4.6.2. *Protein–DNA recognition*

The *trp* repressor binds specifically to the target DNA sequence ACTAGT, resulting in the transcriptional control of L-tryptophan levels in bacteria. The crystal structure of the *trp* repressor/operator complex was solved to 1.9 Å resolution (Otwinowski *et al.*, 1988). Although the structure revealed hydrogen-bonding interactions between the protein and the backbone phosphate groups, no direct hydrogen bonds or non-polar contacts between the protein and DNA bases were observed. Specificity was therefore attributed to the effect of the sequence on the geometry of the phosphate backbone and to water-mediated polar contacts between protein atoms and specific DNA bases. To confirm this hypothesis, the 1.95 Å resolution crystal structure of the free decamer CCACTAGTGG was obtained, containing the recognition six-base-pair sequence (Shakked *et al.*, 1994). A comparative analysis of the free and complexed DNA showed that, when bound to the *trp* repressor, the six-base-pair region is bent by about 15° so as to compress the major groove, with concomitant expansion of the minor groove relative to the uncomplexed DNA (Shakked *et al.*, 1994). However, both free and complexed DNA are underwound, with 10.6 base pairs per turn, rather than the usual 10.0 base pairs per turn. This feature is presumably a result of the particular DNA sequence and is thought to decrease the energy barrier for the binding interaction with the *trp* repressor protein (Shakked *et al.*, 1994). Another specificity component suggested by the authors is conferred by the hydration of the consensus bases. Ten water molecules are observed to interact in the major groove at similar positions in both the free and complexed DNA. Three of these mediate in four hydrogen-bonding interactions to the protein in the complex. Interestingly, the DNA bases to which these three water molecules are bound are among the most conserved and mutationally sensitive bases of the operator. In effect, these three water molecules can be regarded as extensions of the DNA bases and part of the specific recognition elements of the target DNA sequence (Shakked *et al.*, 1994).

The idea of water molecules as mediators of interactions conferring specificity in protein–DNA associations is further supported by the co-crystal structure of the HNF-3/fork head DNA-recognition motif in complex with DNA, solved to 2.5 Å resolution (Clark *et al.*, 1993). Although the lower resolution of this protein–DNA complex may limit the unambiguous determination of water molecules to those that are tightly bound, a series of water molecules are observed in the major groove, bridging specific DNA bases to amino-acid side chains in one of the α-helices of the protein. In this case, direct hydrogen bonding between DNA bases and protein side chains also exists.

The involvement of water in specific protein–DNA recognition was further confirmed in a study of the accuracy of specific DNA cleavage by the restriction endonuclease *Eco*RI under different osmotic pressures (Robinson & Siglar, 1993). Changes in osmotic pressure, resulting from changes in osmolite concentrations, have direct effects on the number of water molecules associated with macromolecules (Rand, 1992). The *Eco*RI experiments show that

water activity affects site-specific DNA recognition, with an increase in osmotic pressure leading to a decrease in accuracy of protein–DNA recognition, as observed by DNA cleavage at sites containing an incorrect base pair (Robinson & Siglar, 1993). The results of this study strongly imply a role for one or more water molecules in recognition of specific sequences of DNA. The authors suggest that water mediation may constitute a general motif for sequence-specific DNA recognition by DNA-binding proteins (Robinson & Siglar, 1993).

The role of water molecules as mediators of sequence-specific DNA recognition may be a general motif, but not a necessary one. The solution NMR structure of the complex of erythroid transcription factor GATA-1 with the 16-base-pair DNA fragment GTTGCAGATAAACATT, containing the recognition sequence, shows that the specific interactions between GATA-1 and the major groove of the DNA are dominated by van der Waals interactions hydrophobic in nature (Omichinski *et al.*, 1993). Furthermore, NMR experiments designed to identify the location of water molecules in the complex detected clusters of water molecules bridging the protein to the DNA phosphate backbone, but showed that water was excluded from the hydrophobic interface between the protein and the DNA bases (Clore *et al.*, 1994). Although many of the existing crystal structures of protein–DNA complexes support the general view that water molecules are often integral components of the specific recognition between the protein and the target DNA, this solution structure provides an important example of exclusion of water molecules from the specificity determinants. In the GATA-1–DNA complex, however, water molecules do mediate non-specific binding of the protein to the DNA backbone. It appears, not surprisingly, that water molecules play a variety of roles in the mediation of protein–DNA interactions and that these roles are specific to each particular case.

23.4.6.3. *Cooperativity in dimeric haemoglobin*

The X-ray crystal structures of liganded and unliganded dimeric haemoglobin from *Scapharca inaequivalvis* have revealed that water molecules at the dimer interface form an integral part of the cooperativity mechanism in this system (Condon & Royer, 1994; Royer, 1994). The binding of oxygen to one of the monomers causes little rearrangement of quaternary structure. It does, instead, displace the side chain of Phe97 which, in the low-affinity deoxy form, packs in the haem pocket (Royer *et al.*, 1990). Phe97 in the deoxy form lowers the oxygen affinity by restricting movement of the iron atom into the haem plane (Royer, 1994). Upon oxygen binding, Phe97 flips to the dimer interface, removing six out of the 17 water molecules that are found in the deoxy form (Fig. 23.4.6.1). The resultant destabilization of the water clusters found between the two subunits facilitates the flipping of Phe97 in the other subunit, with a concomitant increase in oxygen affinity of the haem in the second subunit (Pardanani *et al.*, 1997; Royer *et al.*, 1997).

In each of the monomeric subunits, Thr72 is positioned to form a hydrogen bond with a water molecule at the periphery of the deoxy dimer interface (not shown in Fig. 23.4.6.1). In effect, this interaction caps the water cluster on either side of the interface, presumably helping to stabilize these well ordered water molecules. The isosteric mutation Thr72 to Val was designed to test the importance of this interaction to the stability of the water cluster in the low-affinity haemoglobin dimer and the resultant effect on ligand affinity and cooperativity (Royer *et al.*, 1996). The crystal structure of the T72V mutant was solved to 1.6 Å resolution. This crystal structure reveals that the only significant difference between the mutant and wild-type proteins is the loss of the two water molecules that directly hydrogen-bond to Thr72 in each of the wild-type subunits. Furthermore, there is a significant increase in both
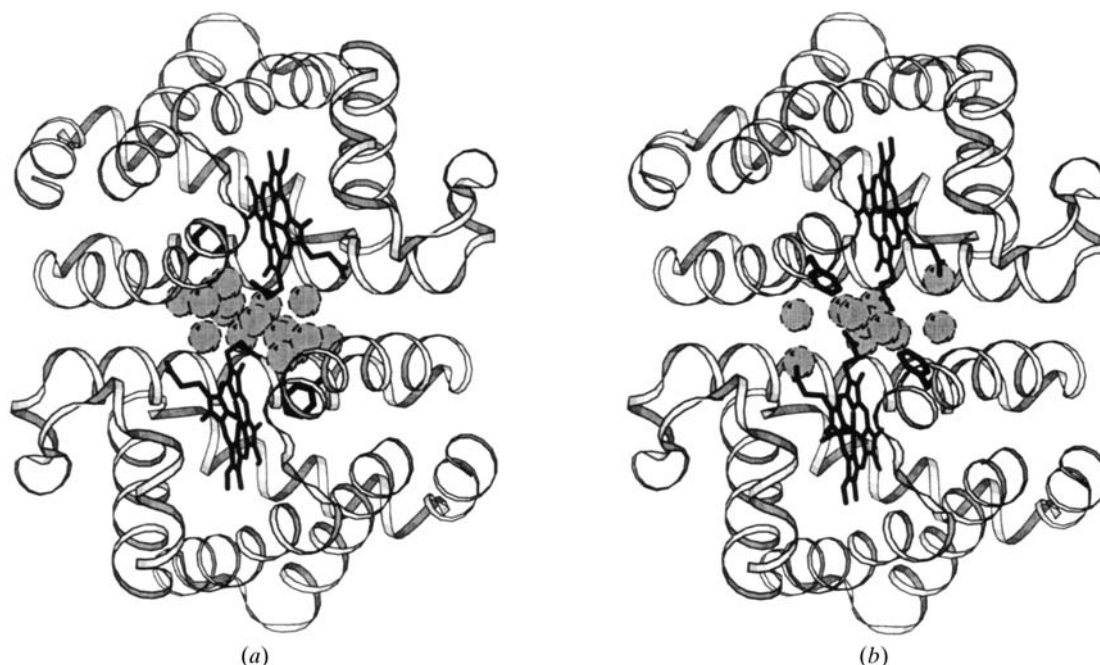

Fig. 23.4.6.1. *Scapharca* HbI interface water molecules. (*a*) Deoxy-HbI at 1.6 Å resolution (PDB code 3SDH) and (*b*) HbI-CO at 1.4 Å resolution (PDB code 4SDH). Included is a ribbon diagram showing the tertiary structure of each subunit, bond representations for the haem group and Phe97 side chain, and spheres representing the approximate van der Waals radii of oxygen atoms for core interface water molecules. Note the cluster of 17 ordered water molecules in the interface of deoxy-HbI for which Phe97 is packed in the haem pocket. Upon ligation, by either CO or O_2, Phe97 is extruded into the interface and disrupts this water cluster, expelling six water molecules from the interface. These plots were produced with the program *MOLSCRIPT* (Kraulis, 1991). Reprinted with permission from Royer *et al.* (1997). Copyright (1997) The American Society for Biochemistry & Molecular Biology.

activity and cooperativity resulting from the mutation (Royer *et al.*, 1996). The authors conclude that, as a result of the mutation, the loss of two water molecules in the interface cluster is sufficient to alter the balance between the low- and high-affinity forms of the protein. This result demonstrates that water molecules are key mediators of information transfer between the haems in the two subunits in dimeric haemoglobin and that their precise positioning and interactions with protein atoms are crucial in maintaining the chemical balance required for biological function.

23.4.6.4. *Summary*

The few examples illustrated above provide diverse views of the ways in which Nature can use water molecules as integral parts of macromolecular interactions. Water molecules can be involved in specificity and recognition, in thermodynamics of binding and affinity, in the cooperative behaviour of allosteric proteins, and in catalysis. Not only do the specific examples illustrate general roles possible for water molecules in the context of a given type of macromolecule, such as proteins or nucleic acids, but they are often representative of any macromolecular system. For example, the role of water in recognition and specificity illustrated above for protein–DNA interactions has also been observed in the L-arabinose-binding protein interaction with specific sugar molecules (Quiocho *et al.*, 1989). Clearly, water molecules are involved so intimately, and in so many different ways, with the formation of molecular complexes that it is not possible to understand the formation process and the function of the complex without taking into account the role of this universal solvent.

23.4.7. Conclusions and future perspectives

The aim of this chapter was to provide a general overview of the available crystallographic information on the roles of water molecules in their interactions with macromolecules. To achieve this aim, the focus has been on representative examples rather than an exhaustive review of the literature. The classification of water molecules according to location and frequency of occurrence among related proteins, or independently solved crystal structures of a given protein, is a crucial element in determining their functional roles.

It has become clear that water molecules involved in crystal contacts can occur virtually anywhere on the protein surface, as exemplified in the case of T4 lysozyme, and that they have properties similar to the majority of surface water molecules within the context of the crystal. Therefore, it is possible to conclude that for the majority of cases, the crystal contacts between proteins involved in the various studies discussed in this review do not significantly influence the general conclusions drawn.

Water molecules bind to proteins so as to satisfy the hydrogen-bonding potential of protein atoms that are not part of the intramolecular hydrogen-bonding pattern within the native structure. At the primary level, the hydrogen bonding is such as to follow the stereochemical requirements of the individual atom in question, in a manner similar to that occurring for the same atom in small molecules. At the secondary-structure level, these positions tend to provide extensions of α-helices or β-sheets as well as to solvate protein atoms in exposed turns. At the tertiary level, they occur more favourably in grooves or cavities within the protein.

Internal or buried water molecules are found to bridge between domains of a single monomer or bring together different secondary-structure elements within a given domain. They have also been observed in the binding sites of proteins where they fine-tune the shape or electrostatic complementarity towards the substrate or ligand. In general, buried water molecules occur in cavities within the protein, making multiple hydrogen bonds with protein atoms that are likely to be conserved among members of a given family. These water molecules are themselves conserved and must be considered an integral part of the protein architecture. They are often connected to bulk water through water channels leading to the protein surface.

Surface water molecules play important roles in protein dynamics, catalysis, thermodynamics of binding, and in mediation of cooperativity, metal binding, recognition and specificity. Representative examples of water molecules in each of these different roles are discussed in the present review. Some of the surface water molecules are found to be conserved within families of proteins, particularly when they are involved in one of the specific roles mentioned above.

In addition to the commonly observed features in crystal structures of proteins solved to around 2 Å resolution, the crystal structures of crambin and BPTI, both solved to 1 Å resolution, provide examples of the type of information only available at very high resolution. This includes the arrangement of water molecules into pentagonal rings around hydrophobic side chains and the occupancy of water molecules on protein surfaces.

A cohesive picture has emerged of the locations of well ordered water molecules on protein surfaces and their functional roles. Currently, there is a good structural view of the protein atoms as well as of the structure of water molecules associated with the protein. The question now is how this information can be used in predicting *a priori* where water molecules will be involved in important structural or functional roles. While having information on the ordered water molecules associated with protein surfaces represents significant progress toward the ultimate goal of understanding the global thermodynamic and kinetic picture of molecular processes in water, the entire system is still not understood. For instance, how does the bulk water influence the dynamic and thermodynamic processes in which the protein and ordered water molecules are involved? Furthermore, what is the importance of solutes normally found in the biological environment where proteins and other macromolecules exert their function? Knowledge must continue to expand toward an understanding of the complete system. Meanwhile, the present models can go a long way toward successful practical applications in protein engineering and ligand design. In order to improve these models, the information accumulated so far can be combined with empirical results and theoretical models to expand the understanding of the first principles underlying biological processes. Building the bridge between empirical observation and first principles is an iterative process still in its infancy.

References

23.1

Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. & Weng, J. (1987). *Protein Data Bank*. In *Crystallographic databases – information content, software systems, scientific applications*, edited by F. H. Allen, G. Bergerhoff, & R. Sievers, pp. 107–132.

Artymiuk, P. J., Mitchell, E. M., Rice, D. W. & Willett, P. (1989). *Searching techniques for databases of protein structures. J. Inf. Sci.* **15**, 287–298.

Artymiuk, P. J., Rice, D. W., Poirrette, A. R. & Willett, P. (1995). *Beta-glucosyltransferase and phosphorylase reveal their common theme. Nature Struct. Biol.* **2**, 117–120.

Barton, G. J. (1997). *3Dee: database of protein domain definitions.* http://barton.ebi.ac.uk/servers/3Dee.html.

Bennett, M. J. & Eisenberg, D. (1994). *Refined structure of monomeric diphtheria toxin at 2.3 Å resolution. Protein Sci.* **3**, 1464–1475.

Bork, P. (1992). *Mobile modules and motifs. Curr. Opin. Struct. Biol.* **2**, 413–421.

Brenner, S. E., Chothia, C., Hubbard, T. J. & Murzin, A. G. (1996). *Understanding protein structure. Using SCOP for fold interpretation. Methods Enzymol.* **266**, 635–643.

Brown, N. P., Orengo, C. A. & Taylor, W. R. (1996). *A protein structure comparison methodology. Comput. Chem.* **20**, 359–380.

Chothia, C. (1993). *One thousand families for the molecular biologist. Nature (London)*, **357**, 543–544.

Crippen, G. (1978). *The tree structural organization of proteins. J. Mol. Biol.* **126**, 315–332.

Flores, T. P., Orengo, C. A. & Thornton, J. M. (1993). *Conformational characteristics in structurally similar protein pairs. Protein Sci.* **7**, 31–37.

Gibrat, J. F., Madej, T., Spouge, J. L. & Bryant, S. H. (1997). *The VAST protein structure comparison method. Biophys. J.* **72**, MP298.

Go, M. (1981). *Correlation of DNA exonic regions with protein structural units in hemoglobin. Nature (London)*, **291**, 90–92.

Hogue, C. W., Ohkawa, H. & Bryant, S. H. (1996). *A dynamic look at structures: WWW-Entrez and the molecular modelling database. Trends Biochem. Sci.* **21**, 226–229.

Holm, L. & Sander, C. (1993). *Protein structure comparison by alignment of distance matrices. J. Mol. Biol.* **233**, 123–138.

Holm, L. & Sander, C. (1994*a*). *Searching protein structure databases has come of age. Proteins*, **19**, 165–173.

Holm, L. & Sander, C. (1994*b*). *Parser for protein folding units. Proteins*, **19**, 256–268.

Holm, L. & Sander, C. (1995). *Evolutionary link between glycogen phosphorylase and a DNA modifying enzyme. EMBO J.* **14**, 1287–1293.

Holm, L. & Sander, C. (1996). *Mapping the protein universe. Science*, **273**, 595–602.

Holm, L. & Sander, C. (1997). *Enzyme HIT. Trends Biochem. Sci.* **22**, 116–117.

Holm, L. & Sander, C. (1998). *Dictionary of recurrent domains in protein structures. Proteins*, **33**, 88–96.

Holm, L. & Sander, C. (1999). *Protein folds and families: sequence and structure alignments. Nucleic Acids Res.* **27**, 244–247.

Islam, S. A., Luo, J. & Sternberg, M. J. (1995). *Identification and analysis of domains in proteins. Protein Eng.* **8**, 513–525.

Jones, S., Stewart, M., Michie, A. D., Swindells, M. B., Orengo, C. A. & Thornton, J. M. (1998). *Domain assignment for protein structures using a consensus approach: characterisation and analysis. Protein Sci.* **7**, 233–242.

Kikuchi, T., Nemethy, G. & Scheraga, H. A. (1988). *Prediction of the location of structural domains in globular proteins. J. Protein Chem.* **88**, 427–471.

Kraulis, P. J. (1991). *MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Cryst.* **24**, 946–950.

Lesk, A. M. & Rose, G. D. (1981). *Folding units in globular proteins. Proc. Natl Acad. Sci. USA*, **78**, 4304–4308.

Levitt, M. & Chothia, C. (1976). *Structural patterns in globular proteins. Nature (London)*, **261**, 552–558.

Li, M., Dyda, F., Benhar, I., Pastan, I. & Davies, D. R. (1996). *Crystal structure of the catalytic domain of Pseudomonas exotoxin A complexed with a nicotinamide adenine dinucleotide analog: implications for the activation process and for ADP ribosylation. Proc. Natl Acad. Sci. USA*, **93**, 6902–6906.

Lionetti, C., Guanziroli, M. G., Frigerio, F., Ascenzi, P. & Bolognesi, M. (1991). *X-ray crystal structure of the ferric sperm whale myoglobin: imidazole complex at 2.0 Å resolution. J. Mol. Biol.* **217**, 409–412.

Mizuguchi, K., Deane, C. A., Blundell, T. L. & Overerington, J. P. (1998). *HOMSTRAD: a database of protein structure alignments for homologous families. Protein Sci.* **7**, 2469–2471.

Moult, J. & Unger, R. (1991). *An analysis of protein folding pathways. Biochemistry*, **30**, 3816–3824.

Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). *SCOP: a structural classification of the protein database for the investigation of sequences and structures. J. Mol. Biol.* **247**, 536–540.

Nemethy, G. & Scheraga, H. A. (1979). *A possible folding pathway of bovine pancreatic Rnase. Proc. Natl Acad. Sci. USA*, **76**, 6050–6054.

Orengo, C. A., Jones, D. T., Taylor, W. & Thornton, J. M. (1994). *Protein superfamilies and domain superfolds. Nature (London)*, **372**, 631–634.

Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B. & Thornton, J. M. (1997). *CATH – a hierarchic classification of protein domain structures. Structure*, **5**, 1093–1108.

Orengo, C. A., Pearl, F. M. G., Bray, J. E., Todd, A. E., Martin, A. C., LoConte, L. & Thornton, J. M. (1999). *The CATH database provides insights into protein structure/function relationships. Nucleic Acids Res.* **27**, 275–279.

Phillips, D. E. (1970). *British biochemistry, past and present*, p. 11. London Biochemistry Society Symposium. Academic Press.

Rashin, A. A. (1976). *Location of domains in globular proteins. Nature (London)*, **291**, 85–87.

Richardson, J. S. (1981). *The anatomy and taxonomy of protein structure. Adv. Protein Chem.* **34**, 167–339.

Rose, G. D. (1979). *Hierarchic organization of domains in globular proteins. J. Mol. Biol.* **134**, 447–470.

Rossmann, M. G. & Argos, P. (1975). *A comparison of the heme binding pocket in globins and cytochrome b5. J. Biol. Chem.* **250**, 7525–7532.

Rossmann, M. & Liljas, A. (1974). *Recognition of structural domains in globular proteins. J. Mol. Biol.* **85**, 177–181.

Russell, R. B. & Barton, G. J. (1993). *Multiple protein sequence alignment from tertiary structure comparisons. Assignments of global and residue level confidences. Proteins*, **14**, 309–323.

Sali, A. & Blundell, T. B. (1990). *The definition of general topological equivalences in proteins: a procedure involving comparison of properties and relationships through simulated annealing and dynamic programming. J. Mol. Biol.* **212**, 403–428.

Sander, C. (1981). *Physical criteria for folding units of globular proteins*. In *Structural aspects of recognition and assembly in biological macromolecules*, Vol. I. *Proteins and protein complexes, fibrous proteins*, edited by M. Balaban, pp. 183–195. Jerusalem: Alpha Press.

Sander, C. & Schneider, R. (1991). *Database of homology-derived protein structures and structural meaning of sequence alignments. Proteins*, **9**, 56–68.

Schulz, G. E. & Schirmer, H. (1979). *Principles of protein structure*, ch. 5. New York: Springer Verlag.

Siddiqui, A. S. & Barton, G. J. (1995). *Continuous and discontinuous domains: an algorithm for the automatic generation of reliable protein domain definitions. Protein Sci.* **4**, 872–884.

Sowdhamini, R., Rufino, S. D. & Blundell, T. L. (1996). *A database of globular protein structural domains: clustering of representa-*

23.1 (*cont.*)

tive family members into similar folds. Structure Fold. Des. **1**, 209–220.

Swindells, M. B. (1995). *A procedure for detecting structural domains in proteins. Protein Sci.* **4**, 103–112.

Taylor, W. R. & Orengo, C. A. (1989). *Protein structure alignment. J. Mol. Biol.* **208**, 1–22.

Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y. & Steitz, T. A. (1996). *Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. EMBO J.* **15**, 5739–5751.

Vriend, G. & Sander, C. (1991). *Detection of common three-dimensional substructures in proteins. Proteins,* **11**, 552–558.

Wernisch, L., Hunting, M. & Wodak, J. (1999). *Identification of structural domains in proteins by a graph heuristic. Proteins,* **35**, 338–352.

Wetlaufer, D. B. (1973). *Nucleation, rapid folding, and globular intrachain regions in proteins. Proc. Natl Acad. Sci. USA,* **70**, 697–701.

Wodak, J. & Janin, J. (1981). *Location of structural domains in proteins. Biochemistry,* **20**, 6544–6552.

Zehfus, M. H. (1994). *Binary discontinuous compact protein domains. Protein Eng.* **7**, 335–340.

Zehfus, M. H. (1997). *Identification of compact, hydrophobically stabilized domains and modules containing multiple peptide chains. Protein Sci.* **6**, 1210–1219.

Zehfus, M. H. & Rose, G. D. (1986). *Compact units in proteins. Biochemistry,* **25**, 5759–5765.

23.2

Åqvist, J., Luecke, H., Quiocho, F. A. & Warshel, A. (1991). *Dipoles localized at helix termini of proteins stabilize charges. Proc. Natl Acad. Sci. USA,* **88**, 2026–2030.

Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967). *On the conformation of the hen egg-white lysozyme molecule. Proc. R. Soc. London Ser. B Biol. Sci.* **167**, 365–377.

Bochkarev, A., Pfuetzner, R. A., Edwards, A. M. & Frappier, L. (1997). *Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. Nature (London),* **385**, 176–181.

Burd, C. G. & Dreyfuss, G. (1994). *Conserved structures and diversity of functions of RNA-binding proteins. Science,* **265**, 615–621.

Cheng, X. (1995). *DNA modification by methyltransferases. Curr. Opin. Struct. Biol.* **5**, 4–10.

Cleland, W. W. & Kreevoy, M. M. (1994). *Low-barrier hydrogen bonds and enzymic catalysis. Science,* **264**, 1887–1890.

Cotton, F. A., Hazen, E. E. Jr & Legg, M. J. (1979). *Staphylococcal nuclease: proposed mechanism of action based on structure of enzyme–thymidine 3′,5′-bisphosphate–calcium ion complex at 1.5-Å resolution. Proc. Natl Acad. Sci. USA,* **76**, 2551–2555.

Cusack, S., Yaremchuk, A. & Tukalo, M. (1996a). *The crystal structure of the ternary complex of T. thermophilus seryl-tRNA synthetase with tRNA(Ser) and a seryl-adenylate analogue reveals a conformational switch in the active site. EMBO J.* **15**, 2834–2842.

Cusack, S., Yaremchuk, A. & Tukalo, M. (1996b). *The crystal structures of T. thermophilus lysyl-tRNA synthetase complexed with E. coli tRNA(Lys) and a T. thermophilus tRNA(Lys) transcript: anticodon recognition and conformational changes upon binding of a lysyl-adenylate analogue. EMBO J.* **15**, 6321–6334.

Doyle, D. A., Cabral, J. M., Pfuetzner, R. A., Kuo, A. L., Gulbis, J. M., Cohen, S. L., Chait, B. T. & MacKinnon, R. (1998). *Science,* **280**, 68–77.

Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R. & Steitz, T. A. (1988). *Cocrystal structure of an editing complex of Klenow fragment with DNA. Proc. Natl Acad. Sci. USA,* **85**, 8924–8928.

Gerlt, J. A. & Gassman, P. G. (1993). *Understanding the rates of certain enzyme-catalyzed reactions: proton abstraction from carbon acids, acyl-transfer reaction, and displacement reactions of phosphodiesters. Biochemistry,* **32**, 1943–1952.

Glusker, J. P. (1991). *Structural aspects of metal liganding to functional groups in proteins. Adv. Protein Chem.* **42**, 1–76.

Goldgur, Y., Mosyak, L., Reshetnikova, L., Ankilova, V., Lavrik, O., Khodyreva, S. & Safro, M. (1997). *The crystal structure of phenylalanyl-tRNA synthetase from Thermus Thermophilus complexed with cognate tRNAPhe. Structure,* **5**, 59–68.

Harrington, R. E. & Winicov, I. (1994). *New concepts in protein–DNA recognition: sequence-directed DNA bending and flexibility. Prog. Nucleic Acid Res. Mol. Biol.* **47**, 195–270.

He, J. J. & Quiocho, F. A. (1993). *Dominant role of local dipoles in stabilizing uncompensated charges on a sulphate sequestered in a periplasmic active transport. Protein Sci.* **2**, 1643–1647.

Hibbert, F. & Emsley, J. (1990). *Hydrogen bonding and chemical reactivity. Adv. Phys. Org. Chem.* **226**, 255–379.

Hodel, A. E., Gershon, P. D. & Quiocho, F. A. (1998). *Structural basis for sequence non-specific recognition of 5′-capped mRNA by a cap-modifying enzyme. Mol. Cell,* **1**, 443–447.

Hodel, A. E., Gershon, P. D., Shi, X., Wang, S. M. & Quiocho, F. A. (1997). *Specific protein recognition of an mRNA cap through its alkylated base. (Letter.) Nature Struct. Biol.* **4**, 350–354.

Jacobson, B. L. & Quiocho, F. A. (1988). *Sulphate-binding protein dislikes protonated oxyacids: a molecular explanation. J. Mol. Biol.* **204**, 783–787.

Joachimiak, A., Schevitz, R. W., Kelley, R. L., Yanofsky, C. & Sigler, P. B. (1983). *Functional inferences from crystals of Escherichia coli trp repressor. J. Biol. Chem.* **258**, 12641–12643.

Kissinger, C. R., Liu, B. S., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. (1990). *Crystal structure of an engrailed homeodomain–DNA complex at 2.8 Å resolution: a framework for understanding homeodomain–DNA interactions. Cell,* **63**, 579–590.

Labahn, J., Scharer, O. D., Long, A., Ezaz-Nikpay, K., Verdine, G. L. & Ellenberger, T. E. (1996). *Structural basis for the excision repair of alkylation-damaged DNA. Cell,* **86**, 321–329.

Ledvina, P. S., Tsai, A.-H., Wang, Z., Koehl, E. & Quiocho, F. A. (1998). *Dominant role of local dipolar interactions in phosphate binding to a receptor cleft with an electronegative charge surface potential: equilibrium, kinetic and crystallographic studies. Protein Sci.* **7**, 2550–2559.

Ledvina, P. S., Yao, N., Choudhary, A. & Quiocho, F. A. (1996). *Negative electrostatic surface potential of protein sites specific for anionic ligands. Proc. Natl Acad. Sci. USA,* **93**, 6786–6791.

Lindahl, T. (1982). *DNA repair enzymes. Annu. Rev. Biochem.* **51**, 61–87.

Luecke, H. & Quiocho, F. A. (1990). *High specificity of a phosphate transport protein determined by hydrogen bonds. Nature (London),* **347**, 402–406.

Marcotrigiano, J., Gingras, A. C., Sonenberg, N. & Burley, S. K. (1997). *X-ray studies of the messenger RNA 5′ cap-binding protein (eIF4E) bound to 7-methyl-GDP. Nucleic Acids Symp. Ser. pp.* 8–11.

Meador, W. E., George, S. E., Means, A. R. & Quiocho, F. A. (1995). *X-ray analysis reveals conformational adaptation of the linker in functional calmodulin mutants. (Letter.) Nature Struct. Biol.* **2**, 943–945.

Medveczky, N. & Rosenberg, H. (1971). *Phosphate transport in Escherichia coli. Biochim. Biophys. Acta,* **241**, 494–506.

Nagai, K. (1996). *RNA–protein complexes. Curr. Opin. Struct. Biol.* **6**, 53–61.

Nagai, K., Oubridge, C., Ito, N., Jessen, T. H., Avis, J. & Evans, P. (1995). *Crystal structure of the U1A spliceosomal protein complexed with its cognate RNA hairpin. Nucleic Acids Symp. Ser.* 1–2.

Nicholls, A., Sharp, K. & Honig, B. (1991). *Protein folding and association; insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins,* **11**, 281–296.

Orgel, L. E. (1966). *An introduction to transition-metal chemistry. Ligand-field theory,* 2nd ed. London: Methuen and New York: Wiley.

23.2 (*cont.*)

Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. & Sigler, P. B. (1988). *Crystal structure of trp repressor/operator complex at atomic resolution. Nature (London)*, **335**, 321–329; erratum (1988), **335**, 837.

Pabo, C. O. & Sauer, R. T. (1992). *Transcription factors: structural families and principles of DNA recognition. Annu. Rev. Biochem.* **61**, 1053–1095.

Pardee, A. B. (1966). *Purification and properties of a sulfate-binding protein from Salmonella typhimurium. J. Biol. Chem.* **241**, 5886–5892.

Pavletich, N. P. & Pabo, C. O. (1991). *Zinc finger–DNA recognition: crystal structure of a Zif268–DNA complex at 2.1 Å. Science*, **252**, 809–817.

Pflugrath, J. W. & Quiocho, F. A. (1985). *Sulphate sequestered in the sulphate-binding protein of Salmonella typhimurium is bound solely by hydrogen bonds. Nature (London)*, **314**, 257–260.

Quiocho, F. A. (1986). *Carbohydrate-binding proteins: tertiary structures and protein–sugar interactions. Annu. Rev. Biochem.* **55**, 287–315.

Quiocho, F. A., Sack, J. S. & Vyas, N. K. (1989). *Substrate specificity and affinity of a protein modulated by bound water molecules. Nature (London)*, **340**, 404–407.

Quiocho, F. A., Wilson, D. K. & Vyas, N. K. (1987). *Stabilization of charges on isolated charged groups sequestered in proteins by polarized peptide units. Nature (London)*, **329**, 561–564.

Rademacher, T. W., Parekh, R. B. & Dwek, R. A. (1988). *Glycobiology. Annu. Rev. Biochem.* **57**, 785–838.

Rould, M. A., Perona, J. J. & Steitz, T. A. (1991). *Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. Nature (London)*, **352**, 213–218.

Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976). *Sequence-specific recognition of double helical nucleic acids by proteins. Proc. Natl Acad. Sci. USA*, **73**, 804–808.

Shimon, L. J. & Harrison, S. C. (1993). *The phage 434 OR2/R1-69 complex at 2.5 Å resolution. J. Mol. Biol.* **232**, 826–838.

Steitz, T. A. (1990). *Structural studies of protein–nucleic acid interaction: the sources of sequence-specific binding. Q. Rev. Biophys.* **23**, 205–280.

Sutton, R. B., Fasshauer, D., Jan, R. & Brunger, A. T. (1998). *Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. Nature (London)*, **395**, 347–353.

Tucker, P. W., Hazen, E. E. Jr & Cotton, F. A. (1979). *Staphylococcal nuclease reviewed: a prototypic study in contemporary enzymology. III. Correlation of the three-dimensional structure with the mechanisms of enzymatic action. Mol. Cell. Biochem.* **23**, 67–86.

Ueda, H., Iyo, H., Doi, M., Inoue, M. & Ishida, T. (1991). *Cooperative stacking and hydrogen bond pairing interactions of fragment peptide in cap binding protein with mRNA cap structure. Biochim. Biophys. Acta*, **1075**, 181–186.

Ueda, H., Iyo, H., Doi, M., Inoue, M., Ishida, T., Morioka, H., Tanaka, T., Nishikawa, S. & Uesugi, S. (1991). *Combination of Trp and Glu residues for recognition of mRNA cap structure. Analysis of m7G base recognition site of human cap binding protein (IF-4E) by site-directed mutagenesis. FEBS Lett.* **280**, 207–210.

Varani, G. (1997). *A cap for all occasions. Structure*, **5**, 855–858.

Volbeda, A., Fontecilla-Camps, J. C. & Frey, M. (1996). *Novel metal sites in protein structures. Curr. Opin. Struct. Biol.* **6**, 804–812.

Vyas, N. K. (1991). *Atomic features of protein–carbohydrate interactions. Curr. Opin. Struct. Biol.* **1**, 732–740.

Vyas, N. K., Vyas, M. N. & Quiocho, F. A. (1988). *Sugar and signal-transducer binding sites of the Escherichia coli galactose chemoreceptor protein. Science*, **242**, 1290–1295.

Wang, Z., Luecke, H., Yao, N. & Quiocho, F. A. (1997). *Nature Struct. Biol.* **4**, 519–522.

Weis, W. I. & Drickamer, K. (1996). *Structural basis of lectin–carbohydrate recognition. Annu. Rev. Biochem.* **65**, 441–473.

Werner, M. H., Gronenborn, A. M. & Clore, G. M. (1996). *Intercalation, DNA kinking, and the control of transcription. Science*, **271**, 778–784; erratum (1996), **272**, 19.

Worm, S. H. van den, Stonehouse, N. J., Valegard, K., Murray, J. B., Walton, C., Fridborg, K., Stockley, P. G. & Liljas, L. (1998). *Crystal structures of MS2 coat protein mutants in complex with wild-type RNA operator fragments. Nucleic Acids Res.* **26**, 1345–1351.

Yao, N., Ledvina, P. S., Choudhary, A. & Quiocho, F. A. (1996). *Modulation of a salt link does not affect binding of phosphate to its specific active transport receptor. Biochemistry*, **35**, 2079–2085.

23.3

Altona, C., Geise, H. J. & Romers, C. (1968). *Conformation of non-aromatic ring compounds, XXIV. On the geometry of the perhydrophenanthrene skeleton in some steroids. Tetrahedron*, **24**, 13–32.

Altona, C. & Sundaralingam, M. (1972). *Conformational analysis of the sugar ring in nucleosides and nucleotides. J. Am. Chem. Soc.* **94**, 8205–8212.

Ansevin, A. T. & Wang, A. H. (1990). *Evidence for a new Z-type left-handed DNA helix. Nucleic Acids Res.* **18**, 6119–6126.

Arnott, S. (1970). *The geometry of nucleic acids. Prog. Biophys. Mol. Biol.* **21**, 265–319.

Babcock, M. S. & Olson, W. K. (1994). *The effect of mathematics and coordinate system on comparability and 'dependencies' of nucleic acid structure parameters. J. Mol. Biol.* **237**, 98–124.

Babcock, M. S., Pednault, E. & Olson, W. (1993). *Nucleic acid structure analysis: a users guide to a collection of new analysis programs. J. Biomol. Struct. Dyn.* **11**, 597–628.

Babcock, M. S., Pednault, E. & Olson, W. (1994). *Nucleic acid structure analysis. Mathematics for local Cartesian and helical structure parameters that are truly comparable between structures. J. Mol. Biol.* **237**, 125–156.

Basham, B., Eichman, B. F. & Ho, P. S. (1998). *The single-crystal structures of Z-DNA.* In *Oxford handbook of nucleic acid structure*, edited by S. Neidle, ch. 7, pp. 200–252. Oxford University Press.

Berman, H. M. (1996). *Crystal studies of B-DNA: the answers and the questions. Biopolymers Nucleic Acid Sci.* **44**, 23–44.

Bugg, C. E., Thomas, J. M., Sundaralingam, M. & Rao, S. T. (1971). *Stereochemistry of nucleic acids and their constituents. X. Solid-state base-stacking patterns in nucleic acid consituents and polynucleotides. Biopolymers*, **10**, 175–219.

Crick, F. H. C. & Watson, J. D. (1954). *The complementary structure of deoxyribonucleic acid. Proc. R. Soc. London Ser. A*, **223**, 80–96.

Crothers, D. M. & Drak, J. (1992). *Global features of DNA structure by comparative gel electrophoresis. Methods Enzymol.* **212**, 46–71.

Crothers, D. M., Haran, T. E. & Nadeau, J. G. (1990). *Intrinsically bent DNA. J. Biol. Chem.* **265**, 7093–7096.

Davies, D. B. (1978). *Conformations of nucleosides and nucleotides. Prog. Nucl. Magn. Reson. Spectros.* **12**, 135–186.

Dickerson, R. E. (1972). *The structure and history of an ancient protein. Sci. Am.* **226** (April), 58–72.

Dickerson, R. E. (1983). *The DNA helix and how it is read. Sci. Am.* **249** (December), 94–111.

Dickerson, R. E. (1985). *Helix polymorphism and information flow in DNA.* In *Proceedings of the Robert A. Welch Foundation conferences on chemical research XXIX: Genetic chemistry; the molecular basis of heredity*, pp. 38–79.

Dickerson, R. E. (1992). *DNA structure from A to Z. Methods Enzymol.* **211**, 67–111.

Dickerson, R. E. (1997a). *Obituary: Irving Geis, 1908–1997. Structure*, **5**, 1247–1249.

Dickerson, R. E. (1997b). *Irving Geis, molecular artist, 1908–1997. Protein Sci.* **6**, 2843–2844.

Dickerson, R. E. (1997c). *Biology in pictures: molecular artistry. Curr. Biol.* **7**, R720–R741.

23.3 (*cont.*)

Dickerson, R. E. (1998*a*). *Sequence-dependent B-DNA conformation in crystals and in protein complexes*. In *Structure, motion, interaction and expression of biological macromolecules*, edited by R. H. Sarma & M. H. Sarma, pp. 17–36. New York: Adenine Press.

Dickerson, R. E. (1998*b*). *Helix structure and molecular recognition by B-DNA*. In *Oxford handbook of nucleic acid structure*, edited by S. Neidle, ch. 7, pp. 145–197. Oxford University Press.

Dickerson, R. E. (1998*c*). *DNA bending: the prevalence of kinkiness and the virtues of normality. Nucleic Acids Res.* **26**, 1906–1926.

Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C.-S., von Kitzing, E., Wang, A. H.-J. & Zhurkin, V. B. (1989). *Definitions and nomenclature of nucleic acid structure components. EMBO J.* **8**, 1–4; *J. Biomol. Struct. Dyn.* **6**, 627–634; *Nucleic Acids Res.* **17**, 1797–1803; *J. Mol. Biol.* **206**, 787–791.

Dickerson, R. E. & Chiu, T. K. (1997). *Helix bending as a factor in protein/DNA recognition. Biopolymers Nucleic Acid Sci.* **44**, 361–403.

Dickerson, R. E. & Geis, I. (1969). *The structure and action of proteins*. New York: Harper & Row and Menlo Park: W. A. Benjamin Co.

Dickerson, R. E. & Geis, I. (1976). *Chemistry, matter and the universe*. Menlo Park: Benjamin/Cummings Co.

Dickerson, R. E. & Geis, I. (1983). *Hemoglobin: structure, function, evolution, and pathology*. Menlo Park: Benjamin/Cummings Co.

Dickerson, R. E., Goodsell, D. & Kopka, M. L. (1996). *MPD and DNA bending in crystals and in solution. J. Mol. Biol.* **256**, 108–125.

Dickerson, R. E., Goodsell, D. S., Kopka, M. L. & Pjura, P. E. (1987). *The effect of crystal packing on oligonucleotide double helix structure. J. Biomol. Struct. Dyn.* **5**, 557–579.

Dickerson, R. E., Goodsell, D. S. & Neidle, S. (1994). . . . *the tyranny of the lattice. . .. Proc. Natl Acad. Sci. USA*, **91**, 3579–3583.

El Hassan, M. A. & Calladine, C. R. (1997). *Conformational characteristics of DNA: empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps. Philos. Trans. R. Soc. London A*, **355**, 43–100.

Feigon, J. (1996). *DNA triplexes, quadruplexes & aptamers*. In *Encyclopedia of nuclear magnetic resonance*, edited by D. M. Grant & R. K. Harris, pp. 1726–1731. New York: Wiley.

Franklin, R. E. & Gosling, R. G. (1953). *The structure of sodium thymonucleate fibres. I. The influence of water content. Acta Cryst.* **6**, 673–677.

Haschmeyer, A. E. V. & Rich, A. (1967). *Nucleoside conformation: an analysis of steric barriers to rotation about the glycosidic bond. J. Mol. Biol.* **27**, 369–384.

Herbert, A. & Rich, A. (1996). *The biology of left-handed Z-DNA. J. Biol. Chem.* **271**, 11595–11598.

Ho, P. S. & Mooers, B. H. M. (1996). *Z-DNA crystallography. Biopolymers Nucleic Acid Sci.* **44**, 65–90.

Hoogsteen, K. (1963). *The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. Acta Cryst.* **16**, 907–916.

Hunter, C. A. & Sanders, J. K. M. (1990). *The nature of π–π interactions. J. Am. Chem. Soc.* **112**, 5525–5534.

Juo, Z. S., Chiu, T. K., Leiberman, P. M., Baikalov, I., Berk, A. J. & Dickerson, R. E. (1996). *How proteins recognize the TATA box. J. Mol. Biol.* **261**, 239–254.

Kendrew, J. C. (1961). *The three-dimensional structure of a protein molecule. Sci. Am.* **205** (December), 96–110.

Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993). *Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature (London)*, **365**, 520–527.

Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). *Crystal structure of a yeast TBP/TATA-box complex. Nature (London)*, **365**, 512–520.

Koo, H.-S., Drak, J., Rice, J. A. & Crothers, D. M. (1990). *Determination of the extent of DNA bending by an adenine-thymine tract. Biochemistry*, **29**, 4227–4234.

Koo, H.-S., Wu, H.-M. & Crothers, D. M. (1986). *DNA bending at adenine-thymine tracts. Nature (London)*, **320**, 501–506.

Kostrewa, D. & Winkler, F. K. (1995). Mg^{2+} *binding to the active site of EcoRV endonuclease: a crystallographic study of complexes with substrate and product DNA at 2 Å resolution. Biochemistry*, **34**, 683–696.

Langridge, R., Marvin, D. A., Seeds, W. E., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F. & Hamilton, L. D. (1960). *The molecular configurations of deoxyribonucleic acid. II. Molecular models and their Fourier transforms. J. Mol. Biol.* **2**, 38–64.

Lavery, R. & Sklenar, H. (1988). *The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. J. Biomol. Struct. Dyn.* **6**, 63–91.

Lavery, R. & Sklenar, H. (1989). *Defining the structure of irregular nucleic acids: conventions and principles. J. Biomol. Struct. Dyn.* **6**, 655–667.

Leslie, A. G. W., Arnott, S., Chandrasekaran, R. & Ratliff, R. L. (1980). *Polymorphism of DNA double helices. J. Mol. Biol.* **143**, 49–72.

Levitt, M. & Warshel, A. (1978). *Extreme conformational flexibility of the furanose ring in DNA and RNA. J. Am. Chem. Soc.* **100**, 2607–2613.

Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G. & Lu, P. (1996). *Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. Science*, **271**, 1247–1254.

Marini, J. C., Levene, S. D., Crothers, D. M. & Englund, P. T. (1982). *Bent helical structures in kinetoplast DNA. Proc. Natl Acad. Sci. USA*, **79**, 7664–7668.

Nikolov, D. B., Chen, H., Halay, E. D., Hoffman, A., Roeder, R. G. & Burley, S. K. (1996). *Crystal structure of a human TATA box-binding protein/TATA element complex. Proc. Natl Acad. Sci. USA*, **93**, 4862–4867.

Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y. W., Ebright, R. H. & Berman, H. M. (1996). *Structure of the CAP–DNA complex at 2.5 angstroms resolution: a complete picture of the protein–DNA interface. J. Mol. Biol.* **260**, 395–408.

Pelton, J. G. & Wemmer, D. E. (1989). *Structural characterization of a 2:1 distamycin A/d(CGCAAATTGGC) complex by two-dimensional NMR. Proc. Natl Acad. Sci. USA*, **86**, 5723–5727.

Pelton, J. G. & Wemmer, D. E. (1990). *Binding modes of distamycin-A with d(CGCAAATTTGCG)2 determined by two-dimensional NMR. J. Am. Chem. Soc.* **112**, 1393–1399.

Phillips, D. C. (1966). *The three-dimensional structure of an enzyme molecule. Sci. Am.* **215** (November), 78–90.

Pohl, F. M. (1976). *Polymorphism of a synthetic DNA in solution. Nature (London)*, **260**, 365–366.

Pohl, F. M. & Jovin, T. M. (1972). *Salt-induced co-operative conformational change of a syhnthetic DNA: equlibrium and kinetic studies with poly(dG-dC). J. Mol. Biol.* **67**, 375–396.

Rice, P. A., Yang, S.-W., Mizuuchi, K. & Nash, H. A. (1996). *Crystal structure of an IHF–DNA complex: a protein-induced DNA U-turn. Cell*, **87**, 1295–1306.

Saenger, W. (1984). *Principles of nucleic acid structure*. New York, Berlin, Heidelberg and Tokyo: Springer-Verlag.

Schneider, B., Neidle, S. & Berman, H. M. (1997). *Conformations of the sugar–phosphate backbone in helical DNA crystal structures. Biopolymers*, **42**, 113–124.

Schultz, S. C., Shields, G. C. & Steitz, T. A. (1991). *Crystal structure of a CAP–DNA complex: the DNA is bent by 90 degrees. Science*, **253**, 1001–1007.

Schumacher, M. A., Choi, K. Y., Zalkin, H. & Brennan, R. G. (1994). *Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. Science*, **266**, 763–770.

Schwartz, T., Rould, M. A., Lowenjaupt, K., Herbert, A. & Rich, A. (1999). *Crystal structure of the Zα domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. Science*, **284**, 1841–1845.

23.3 (cont.)

Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976). *Sequence-specific recognition of double helical nucleic acids by proteins. Proc. Natl Acad. Sci. USA*, **73**, 804–808.

Sklenár, V. & Feigon, J. (1990). *Formation of a stable triplex from a single DNA strand. Nature (London)*, **345**, 836–838.

Sprous, D., Young, M. A. & Beveridge, D. L. (1999). *Molecular dynamics studies of axis bending in* $d(G_5\text{-}(GA_4T_4C)_2\text{-}C_5)$ *and* $d(G_5\text{-}(GT_4A_4C)_2\text{-}C_5)$: *effects of sequence polarity on DNA curvature. J. Mol. Biol.* **285**, 1623–1632.

Sprous, D., Zacharias, W., Wood, Z. A. & Harvey, S. C. (1995). *Dehydrating agents sharply reduce curvature in DNAs containing A-tracts. Nucleic Acids Res.* **23**, 1816–1821.

Sundaralingam, M. (1975). *Principles governing nucleic acid and polynucleotide conformations.* In *Structure and conformation of nucleic acids and protein–nucleic acid interactions*, edited by M. Sundaralingam & S. T. Rao, pp. 487–524. Baltimore: University Park Press.

Thomas, K. A., Smith, G. M., Thomas, T. B. & Feldmann, R. J. (1982). *Electronic distributions within protein phenylalanine aromatic rings are reflected by the three-dimensional oxygen atom environments. Proc. Natl Acad. Sci. USA*, **79**, 4843–4847.

Voet, D. & Voet, J. G. (1990). *Biochemistry.* New York: John Wiley & Sons.

Voet, D. & Voet, J. G. (1995). *Biochemistry*, 2nd edition. New York: John Wiley & Sons.

Wahl, M. C. & Sundaralingam, M. (1996). *Crystal structures of A-DNA duplexes. Biopolymers Nucleic Acid Sci.* **44**, 45–63.

Wahl, M. C. & Sundaralingam, M. (1998). *A-DNA duplexes in the crystal.* In *Oxford handbook of nucleic acid structure*, edited by S. Neidle, ch. 5, pp. 117–144. Oxford University Press.

Watson, J. D. & Crick, F. H. C. (1953). *Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. Nature (London)*, **171**, 737–738.

Winkler, F. K., Banner, D. W., Oefner, C., Tsernoglou, D., Brown, R. S., Heathman, S. P., Bryan, R. K., Martin, P. D., Petratos, K. & Wilson, K. S. (1993). *The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. EMBO J.* **12**, 1781–1795.

Wu, H.-M. & Crothers, D. M. (1984). *The locus of sequence-directed and protein-induced DNA bending. Nature (London)*, **308**, 509–513.

Yang, W. & Steitz, T. A. (1995). *Crystal structure of the site-specific recombinase gamma delta resolvase complexed with a 34 bp cleavage site. Cell*, **82**, 193–207.

23.4

Allen, K. N., Bellamacina, C. R., Ding, X., Jeffery, C. J., Mattos, C., Petsko, G. A. & Ringe, D. (1996). *An experimental approach to mapping the binding surfaces of crystalline proteins. J. Phys. Chem.* **100**, 2605–2611.

Badger, J. (1993). *Multiple hydration layers in cubic insulin crystals. Biophys. J.* **65**, 1656–1659.

Baker, E. N. & Hubbard, R. E. (1984). *Hydrogen bonding in globular proteins. Prog. Biophys. Mol. Biol.* **44**, 97–179.

Beglov, D. & Roux, B. (1997). *An integral equation to describe the solvation of polar molecules in liquid water. J. Phys. Chem.* **101**, 7821–7826.

Bellamacina, C., Mattos, C., Griffith, D., Ivanov, D., Stanton, M., Petsko, G. A. & Ringe, D. (1999). Unpublished results.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *The Protein Data Bank. Nucleic Acids Res.* **28**, 235–242.

Bhat, T. N., Bentley, G. A., Boulot, G., Greene, M. I., Tello, D., Dall'Acqua, W., Souchon, H., Schwarz, F. P., Maiuzza, R. A. & Poljak, R. J. (1994). *Bound water molecules and conformational stabilization help mediate an antigen–antibody association. Proc. Natl Acad. Sci. USA*, **91**, 1089–1093.

Blake, C. C. F., Pulford, W. C. A. & Artymiuk, P. J. (1983). *X-ray studies of water in crystals of lysozyme. J. Mol. Biol.* **167**, 693–723.

Brooks, C. L. & Karplus, M. (1989). *Solvent effects on protein motion and protein effects on solvent motion. Dynamics of the active site region of lysozyme. J. Mol. Biol.* **208**, 159–181.

Bryant, R. G. (1996). *The dynamics of water–protein interactions. Annu. Rev. Biophys. Biomol. Struct.* **25**, 29–53.

Chervenak, M. C. & Toone, E. J. (1994). *A direct measure of the contribution of solvent reorganization to the enthalpy of ligand binding. J. Am. Chem. Soc.* **116**, 10533–10539.

Clackson, T. & Wells, J. T. (1995). *A hot spot of binding energy in a hormone–receptor interface. Science*, **267**, 383–386.

Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. (1993). *Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature (London)*, **364**, 412–420.

Clore, G. M., Bax, A., Omichinski, J. G. & Gronenborn, A. M. (1994). *Localization of bound water in the solution structure of a complex of the erythroid transcription factor GATA-1 with DNA. Curr. Biol.* **2**, 89–94.

Condon, P. & Royer, W. (1994). *Crystal structure of oxygenated Scapharca dimeric hemoglobin at 1.7 Å resolution. J. Biol. Chem.* **269**, 25259–25267.

Deisenhofer, J. & Steigemann, W. (1975). *Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 Å resolution. Acta Cryst.* B**31**, 238–250.

Edsall, J. T. & McKenzie, H. A. (1978). *Water and proteins. I. The significance and structure of water; its interaction with electrolytes and non-electrolytes. Adv. Biophys.* **10**, 137–207.

Edsall, J. T. & McKenzie, H. A. (1983). *Water and proteins. II. The location and dynamics of water in protein systems and its relation to their stability and properties. Adv. Biophys.* **16**, 53–183.

Gunsteren, W. F. van, Luque, F. J., Timms, D. & Torda, A. E. (1994). *Molecular mechanics in biology: from structure to function, taking account of solvation. Annu. Rev. Biophys. Biomol. Struct.* **23**, 847–863.

Hayward, S., Kitao, A., Hirata, F. & Go, N. (1993). *Effect of solvent on collective motions in globular protein. J. Mol. Biol.* **234**, 1207–1217.

Hendrickson, W. A. & Teeter, M. M. (1981). *Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. Nature (London)*, **290**, 107–113.

Hendsch, Z. S., Jonsson, T., Sauer, R. T. & Tidor, B. (1996). *Protein stabilization by removal of unsatisfied polar groups: computational approaches and experimental tests. Biochemistry*, **35**, 7621–7625.

Hendsch, Z. S. & Tidor, B. (1994). *Do salt bridges stabilize proteins? A continuum electrostatic analysis. Protein Sci.* **3**, 211–226.

Herron, J. N., Terry, A. H., Johnston, S., He, S.-M., Guddat, L. W., Voss, E. W. & Edmundson, A. B. (1994). *High resolution structures of the 4-4-20 Fab–fluorescein complex in two solvent systems: effects of solvent on structure and antigen-binding affinity. Biophys. J.* **67**, 2167–2175.

Holdgate, G., Tunnicliffe, A., Ward, W. H. J., Weston, S. A., Rosenbrock, G., Barth, P. T., Taylor, I. W. F., Pauptit, R. A. & Timms, D. (1997). *The entropic penalty of ordered water accounts for weaker binding of the antibiotic Novobiocin to a resistant mutant of DNA gyrase: a thermodynamic and crystallographic study. Biochemistry*, **36**, 9663–9673.

Hubbard, S. J., Gross, K.-H. & Argos, P. (1994). *Intramolecular cavities in globular proteins. Protein Eng.* **7**, 613–626.

Jiang, J.-S. & Brünger, A. (1994). *Protein hydration observed by X-ray diffraction. Solvation properties of penicillopepsin and neuraminidase crystal structures. J. Mol. Biol.* **243**, 100–115.

Karplus, P. A. & Faerman, C. (1994). *Ordered water in macromolecular structure. Curr. Opin. Struct. Biol.* **4**, 770–776.

Kauzmann, W. (1959). *Some factors in the interpretation of protein denaturation. Adv. Protein Chem.* **14**, 1–63.

Kendrew, J. C. (1963). *Myoglobin and the structure of proteins. Science*, **139**, 1259–1266.

Komives, E. A., Lougheed, J. C., Liu, K., Sugio, S., Zhang, Z., Petsko, G. A. & Ringe, D. (1995). *The structural basis for pseudoreversion of the E165D lesion by the secondary S96P*

23.4 (*cont.*)

mutation in triosephosphate isomerase depends on the positions of active site water molecules. Biochemistry, **34**, 13612–13621.

Kossiakoff, A. A., Sintchak, M. D., Shpungin, J. & Presta, L. G. (1992). *Analysis of solvent structure in proteins using neutron $D_2O - H_2O$ solvent maps: pattern of primary and secondary hydration of trypsin. Proteins Struct. Funct. Genet.* **12**, 223–236.

Kraulis, P. J. (1991). *MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Cryst.* **24**, 946–950.

Kuhn, L., Siani, M. A., Pique, M. E., Fisher, C. L., Getzoff, E. D. & Tainer, J. A. (1992). *The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures. J. Mol. Biol.* **228**, 13–22.

Kuhn, L. A., Swanson, C. A., Pique, M. E., Tainer, J. A. & Getzoff, E. D. (1995). *Atomic and residue hydrophilicity in the context of folded protein structures. Proteins Struct. Funct. Genet.* **23**, 536–547.

Ladbury, J. E. (1996). *Just add water! The effect of water on the specificity of protein–ligand binding sites and its potential application to drug design. Chem. Biol.* **3**, 973–980.

Lam, P. Y. S., Jadhav, P. K., Eyermann, C. J., Hodge, C. N., Ru, Y., Bacheler, L. T., Meek, J. L., Otto, M. J., Rayner, M. M., Wong, Y. N., Chang, C.-H., Weber, P. C., Jackson, D. A., Sharpe, T. R. & Erickson-Vitanen, S. (1994). *Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. Science*, **263**, 380–384.

Lazaridis, T., Archontis, G. & Karplus, M. (1995). *Enthalpic contribution to protein stability: insights from atom-based calculations and statistical mechanics. Adv. Protein Chem.* **47**, 231–306.

Levitt, M. & Park, B. H. (1993). *Water: now you see it, now you don't. Structure*, **1**, 223–226.

Loris, R., Langhorst, U., De Vos, S., Decanniere, K., Bouckaert, J., Maes, D., Transue, T. R. & Steyaert, J. (1999). *Conserved water molecules in a large family of microbial ribonucleases. Proteins Struct. Funct. Genet.* **36**, 117–134.

Loris, R., Stas, P. P. G. & Wyns, L. (1994). *Conserved waters in legume lectin crystal structures. The importance of bound water for the sequence–structure relationship within the legume lectin family. J. Biol. Chem.* **269**, 26722–26733.

Lounnas, V. & Pettitt, B. M. (1994). *Distribution function implied dynamics versus residence times and correlations: solvation shells of myoglobin. Proteins Struct. Funct. Genet.* **18**, 148–160.

Lounnas, V., Pettitt, B. M. & Phillips, G. N. Jr (1994). *A global model of the protein–solvent interface. Biophys. J.* **66**, 601–614.

McDowell, R. S. & Kossiakoff, A. A. (1995). *A comparison of neutron diffraction and molecular dynamics structures: hydroxyl group and water molecule orientations in trypsin. J. Mol. Biol.* **250**, 553–570.

Malin, R., Zielenkiewicz, P. & Saenger, W. (1991). *Structurally conserved water molecules in ribonuclease T1. J. Biol. Chem.* **266**, 4848–4852.

Mattos, C., Bellamacina, C., Amaral, A., Peisach, E., Vitkup, D., Petsko, G. A. & Ringe, D. (2000). *The application of the multiple solvent crystal structures method to porcine pancreatic elastase.* In preparation.

Mattos, C., Giammona, D. A., Petsko, G. A. & Ringe, D. (1995). *Structural analysis of the active site of porcine pancreatic elastase based on the X-ray crystal structures of complexes with trifluoroacetyl-dipeptide-anilide inhibitors. Biochemistry*, **34**, 3193–3203.

Mattos, C., Rasmussen, B., Ding, X., Petsko, G. A. & Ringe, D. (1994). *Analogous inhibitors of elastase do not always bind analogously. Nature Struct. Biol.* **1**, 55–58.

Mattos, C. & Ringe, D. (1996). *Locating and characterizing binding sites on proteins. Nature Biotech.* **14**, 595–599.

Meiering, E. M. & Wagner, G. (1995). *Detection of long-lived bound water molecules in complexes of human dihydrofolate reductase with methotrexate and NADPH. J. Mol. Biol.* **247**, 294–308.

Meyer, E. (1992). *Internal water molecules and H-bonding in biological macromolecules: a review of structural features with functional implications. Protein Sci.* **1**, 1543–1562.

Meyer, E., Cole, G., Radhakrishnan, R. & Epp, O. (1988). *Structure of native porcine pancreatic elastase at 1.65 Å resolution. Acta Cryst.* B**44**, 26–38.

Momany, F. A., McGuire, R. F., Burgess, A. W. & Scheraga, H. A. (1975). *Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occuring amino acids. J. Phys. Chem.* **79**, 2361–2381.

Morton, C. J. & Ladbury, J. E. (1996). *Water mediated protein–DNA interactions: the relationship of thermodynamics to structural detail. Protein Sci.* **5**, 2115–2118.

Nicholls, A., Sharp, K. A. & Honig, B. (1991). *Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins Struct. Funct. Genet.* **11**, 281–296.

Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S. J. & Gronenborn, A. (1993). *NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. Science*, **261**, 438–446.

Oprea, T. I., Hummer, G. & Garcia, A. E. (1997). *Identification of a functional water channel in cytochrome P450 enzymes. Proc. Natl Acad. Sci. USA*, **94**, 2133–2138.

Otting, G., Liepinsh, E. & Wuthrich, K. (1991). *Protein hydration in aqueous solution. Science*, **254**, 974–980.

Otting, G. & Wuthrich, K. (1989). *Studies of protein hydration in aqueous solution by direct NMR observation of individual protein-bound water molecules. J. Am. Chem. Soc.* **111**, 1871–1875.

Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. & Sigler, P. B. (1988). *Crystal structure of trp repressor/operator complex at atomic resolution. Nature (London)*, **335**, 321–329.

Pardanani, A., Gibson, Q. H., Colotti, G. & Royer, W. E. (1997). *Mutation of residue Phe97 to Leu disrupts the central allosteric pathway in Scapharca dimeric hemoglobin. J. Biol. Chem.* **272**, 13171–13179.

Pletinckx, J., Steyaert, J., Zegers, I., Choe, H.-W., Heinemann, U. & Wyns, L. (1994). *Crystallographic study of Glu58Ala RNase T1-2'-guanosine monophosphate at 1.9 Å resolution. Biochemistry*, **33**, 1654–1662.

Pomes, R. & Roux, B. (1996). *Structure and dynamics of a proton wire: a theoretical study of H^+ translocation along the single-file water chain in the gramicidin A channel. Biophys. J.* **71**, 19–39.

Poormina, C. S. & Dean, P. M. (1995*a*). *Hydration in drug design. 3. Conserved water molecules at the ligand-binding sites of homologous proteins. J. Comput.-Aided Mol. Des.* **9**, 521–531.

Poormina, C. S. & Dean, P. M. (1995*b*). *Hydration in drug design. 1. Multiple hydrogen-bonding features of water molecules in mediating protein-ligand interactions. J. Comput.-Aided Mol. Des.* **9**, 500–512.

Poormina, C. S. & Dean, P. M. (1995*c*). *Hydration in drug design. 2. Influence of local site surface shape on water binding. J. Comput.-Aided Mol. Des.* **9**, 513–520.

Privé, G. G., Milburn, M. V., Tong, L., DeVos, A. M., Yamaizumi, Z., Nishimura, S. & Kim, S. H. (1992). *X-ray crystal structures of transforming p21 ras mutants suggest a transition-state stabilization mechanism for GTP hydrolysis. Proc. Natl Acad. Sci.* **89**, 3649–3653.

Quiocho, F. A., Wilson, D. K. & Vyas, N. K. (1989). *Substrate specificity and affinity of a protein modulated by bound water molecules. Nature (London)*, **340**, 404–407.

Rand, R. P. (1992). *Raising water to new heights. Science*, **256**, 618.

Rashin, A. A., Iofin, M. & Honig, B. (1986). *Internal cavities and buried waters in globular proteins. Biochemistry*, **25**, 3619–3625.

Ringe, D. (1995). *What makes a binding site a binding site? Curr. Opin. Struct. Biol.* **5**, 825–829.

23.4 (*cont.*)

Robinson, C. R. & Siglar, S. G. (1993). *Molecular recognition mediated by bound water. A mechanism for star activity of the restriction endonuclease EcoRI. J. Mol. Biol.* **234**, 302–306.

Roe, S. M. & Teeter, M. M. (1993). *Patterns for prediction of hydration around polar residues in proteins. J. Mol. Biol.* **229**, 419–427.

Roux, B., Nina, M., Pomes, R. & Smith, J. C. (1996). *Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: a molecular dynamics free energy pertubation study. Biophys. J.* **72**, 670–681.

Royer, W. (1994). *High-resolution crystallographic analysis of a co-operative dimeric hemoglobin. J. Mol. Biol.* **235**, 657–681.

Royer, W. E., Fox, R. A. & Smith, F. R. (1997). *Ligand linked assembly of Scapharca dimeric hemoglobin. J. Biol. Chem.* **272**, 5689–5694.

Royer, W. E. Jr, Hendrickson, W. A. & Chiancone, E. (1990). *Structural transitions upon ligand binding in a cooperative dimeric hemoglobin. Science*, **249**, 518–521.

Royer, W. E., Pardanani, A., Gibson, Q. H., Peterson, E. S. & Friedman, J. M. (1996). *Ordered water molecules as key allosteric mediators in a cooperative dimeric hemoglobin. Proc. Natl Acad. Sci. USA*, **93**, 14526–14531.

Savage, H. (1986). *Water structure in vitamin B$_{12}$ coenzyme crystals. Biophys. J.* **50**, 967–980.

Savage, H. & Wlodawer, A. (1986). *Determination of water structure around biomolecules using X-ray and neutron diffraction methods. Methods Enzymol.* **127**, 162–183.

Shakked, Z., Guzikevich-Guerstein, G., Frolow, F., Rabinovich, D., Joachimiak, A. & Sigler, P. B. (1994). *Determinants of repressor/operator recognition from the structure of the trp operator binding site. Nature (London)*, **368**, 469–473.

Shaltiel, S., Cox, S. & Taylor, S. (1998). *Conserved water molecules contribute to the extensive network of interactions at the active site of protein kinase A. Proc. Natl Acad. Sci. USA*, **95**, 484–491.

Shpungin, J. & Kossiakoff, A. A. (1986). *A method of solvent structure analysis for proteins using $D_2O - H_2O$ neutron difference maps. Methods Enzymol.* **127**, 329–342.

Singer, P., Smalas, A., Carty, R. P., Mangel, W. F. & Sweet, R. M. (1993). *The hydrolytic water molecule in trypsin, revealed by time-resolved Laue crystallography. Science*, **259**, 669–673.

Sreenivasan, U. & Axelsen, P. H. (1992). *Buried water in homologous serine proteases. Biochemistry*, **31**, 12785–12791.

Teeter, M. M. (1984). *Water structure of a hydrophobic protein at atomic resolution: pentagon rings of water molecules in crystals of crambin. Proc. Natl Acad. Sci. USA*, **81**, 6014–6018.

Teeter, M. M. (1991). *Water–protein interactions: theory and experiment. Annu. Rev. Biophys. Biophys. Chem.* **20**, 577–600.

Teeter, M. M. & Hendrickson, W. A. (1979). *Highly ordered crystals of the plant seed protein crambin. J. Mol. Biol.* **127**, 219–223.

Thanki, N., Thornton, J. M. & Goodfellow, J. M. (1988). *Distribution of water around amino acid residues in proteins. J. Mol. Biol.* **202**, 637–657.

Thanki, N., Thornton, J. M. & Goodfellow, J. M. (1990). *Influence of secondary structure on the hydration of serine, threonine and tyrosine residues in proteins. Protein Eng.* **3**, 495–508.

Thanki, N., Umrania, Y., Thornton, J. M. & Goodfellow, J. M. (1991). *Analysis of protein main-chain solvation as a function of secondary structure. J. Mol. Biol.* **221**, 669–691.

Walshaw, J. & Goodfellow, J. M. (1993). *Distribution of solvent molecules around apolar side-chains in protein crystals. J. Mol. Biol.* **231**, 392–414.

Weaver, L. & Matthews, B. (1987). *Structure of bacteriophage T4 lysozyme refined at 1.7 Å resolution. J. Mol. Biol.* **193**, 189–199.

Williams, M. A., Goodfellow, J. M. & Thornton, J. M. (1994). *Buried waters and internal cavities in monomeric proteins. Protein Sci.* **3**, 1224–1235.

Wlodawer, A., Walter, J., Huber, R. & Sjolin, L. (1984). *Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II. J. Mol. Biol.* **180**, 301–329.

Zegers, I., Maes, D., Dao-Thi, M.-H., Poortmans, F., Palmer, R. & Wyns, L. (1994). *The structures of RNase A complexed with 3′-CMP and d(CpA): active site conformation and conserved water molecules. Protein Sci.* **3**, 2322–2339.

Zhang, X.-J. & Matthews, B. W. (1994). *Conservation of solvent-binding sites in 10 crystal forms of T4 lysozyme. Protein Sci.* **3**, 1031–1039.

references