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

references