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

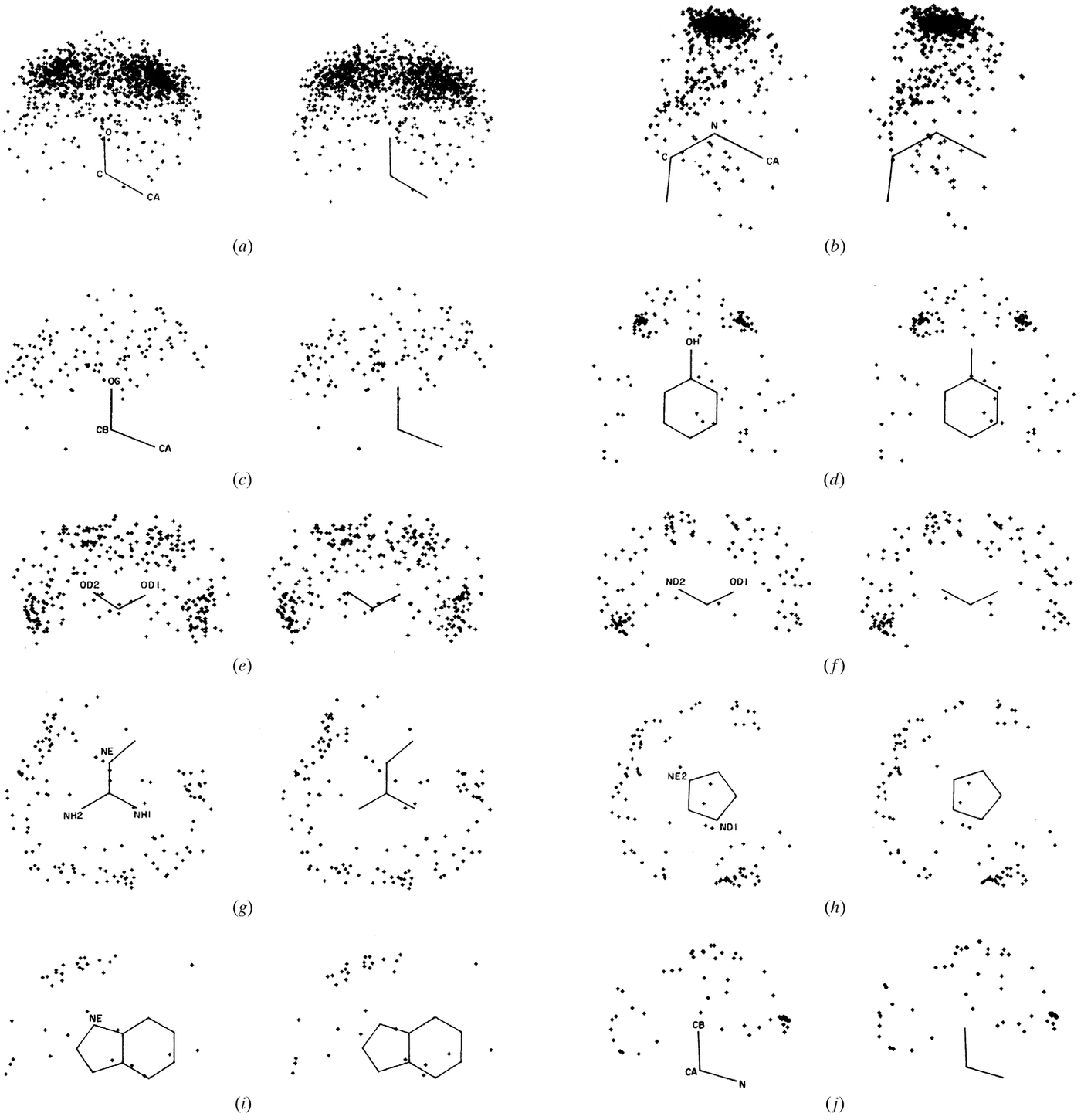

Fig. 23.4.3.1. Distribution of water-molecule sites in stereo around: (*a*) main-chain O, (*b*) main-chain N, (*c*) Ser OG, (*d*) Tyr ring, (*e*) Asp OD1 and OD2, (*f*) Asn OD1 and ND2, (*g*) Arg NH1, NH2 and NE, (*h*) His ring to 3.5 Å, (*i*) Trp ring to 3.5 Å, (*j*) Ala CB. Reprinted with permission 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 10837 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).

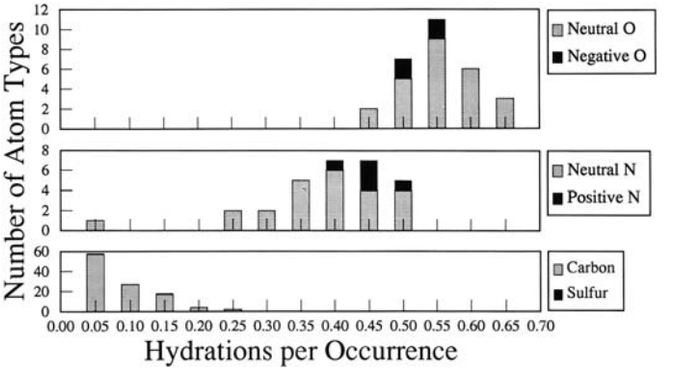

Fig. 23.4.3.2. Distribution of atomic hydration values. To determine which atoms are similar or distinct with respect to water binding, we plotted the number of atom types (*e.g.* Ala amide nitrogen, Ala Cα, . . .) at each hydration per occurrence value. Each atom type contributed one vertical unit to 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.

references