23.4. SOLVENT STRUCTURE

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).
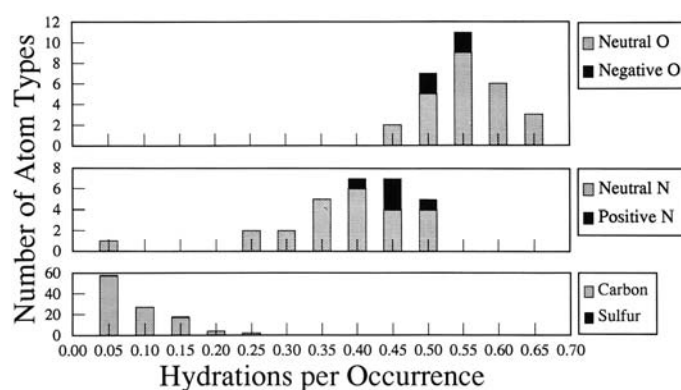

Fig. 23.4.3.2. Distribution of atomic hydration values. To determine which atoms are similar or distinct with respect to water binding, we plotted the number of atom types (*e.g.* Ala amide nitrogen, Ala Cα, . . .) at each hydration per occurrence value. Each atom type contributed one vertical unit to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

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
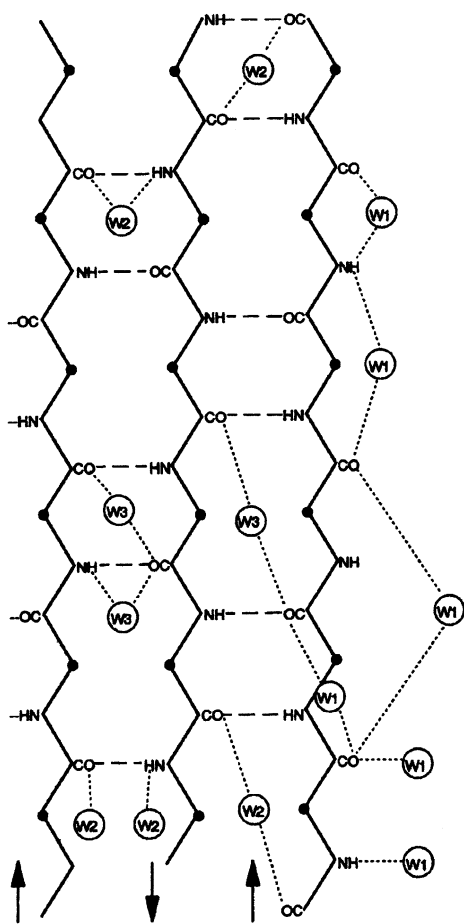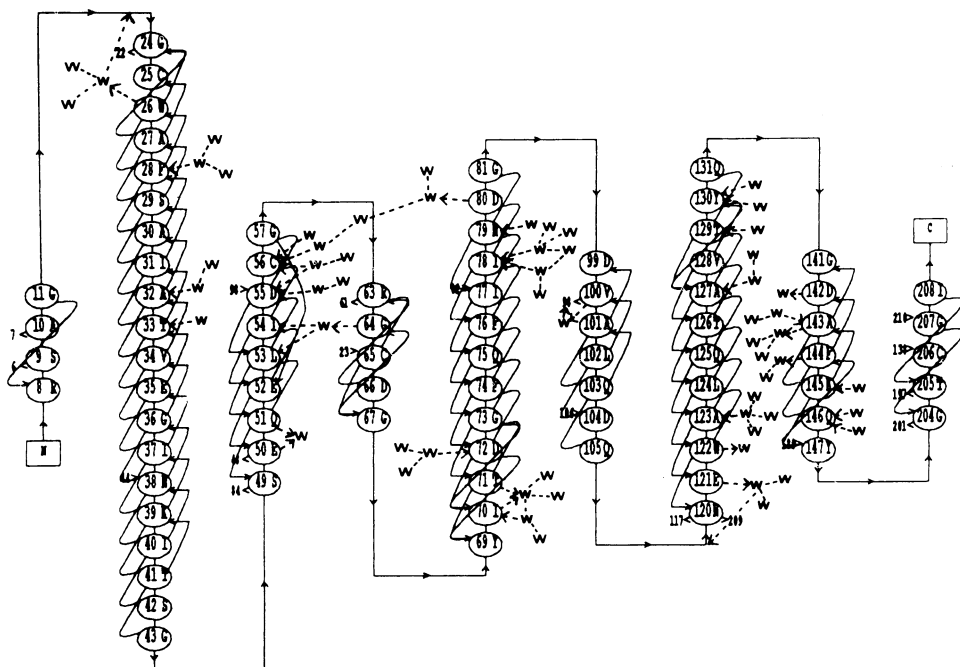

Fig. 23.4.3.3. Diagram of edge (W1), end (W2) and middle (W3) categories of interactions of water molecules with main-chain atoms in antiparallel β-sheets. Reprinted with permission from Thanki *et al.* (1991). Copyright (1991) Academic Press.

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