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 Å^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 Å^3, respectively (Hubbard *et al.*, 1994). The mean volume per water molecule in a cavity is 27 Å^3, as compared to 30 Å^3 in bulk water, suggesting that a water molecule is not favourably squeezed into a volume comparable to its own (11.5 Å^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

references