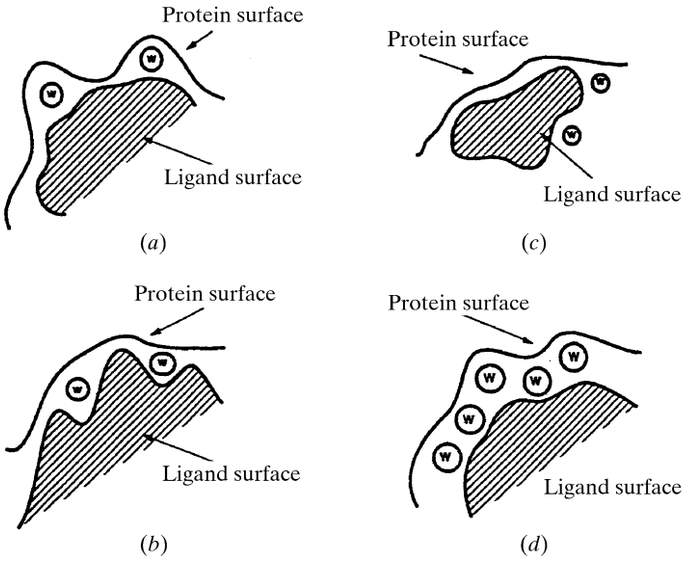

Fig. 23.4.3.5. Schematic illustration of water molecules bound in different types of grooves between protein and ligand. The hatched surfaces represent the ligand surface. (*a*) Water molecules bound in an indentation on the protein surface, where the protein surface area exposed to the water molecules 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

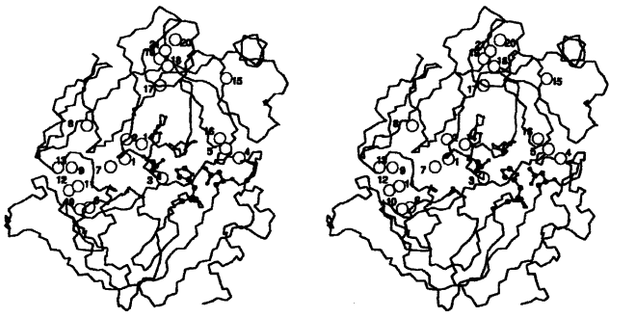

Fig. 23.4.4.1. Stereoview of the set of 21 highly conserved buried waters in eukaryotic serine proteases. The trypsin backbone is represented as a stick drawing, with the catalytic triad at the centre (filled circles). Water molecules are represented as open circles. Reprinted with permission from Sreenivasan & 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).

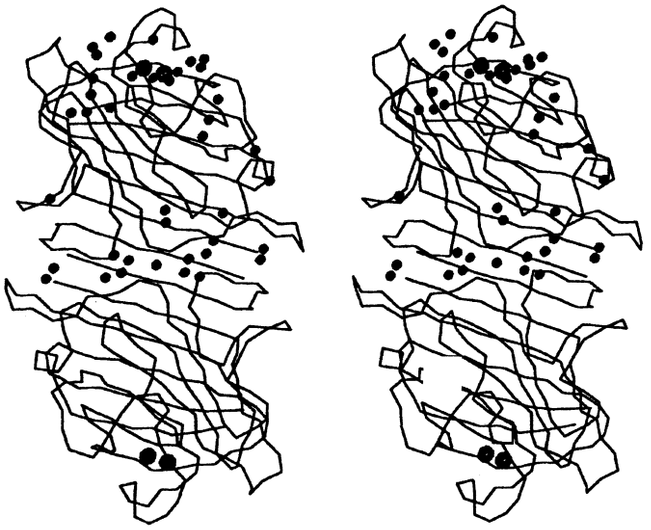

Fig. 23.4.4.2. View of the 33 conserved hydration sites in the lentil lectin crystal structures superimposed on the backbone of the lentil lectin dimer. In order to emphasize the twofold symmetry, the waters at the dimer interface are shown for both lectin monomers. Reprinted with permission from Loris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

references