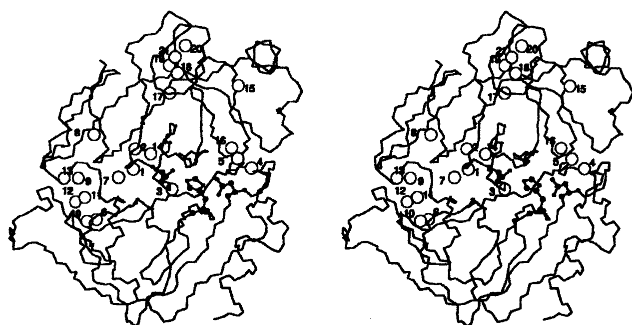

Fig. 23.4.4.1. Stereoview of the set of 21 highly conserved buried waters in eukaryotic serine proteases. The trypsin backbone is represented as a stick drawing, with the catalytic triad at the centre (filled circles). Water molecules are represented as open circles. Reprinted with permission from Sreenivasan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

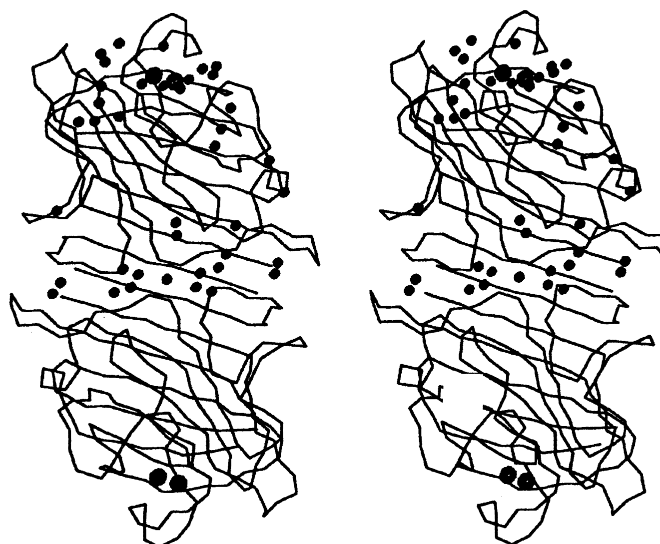

Fig. 23.4.4.2. View of the 33 conserved hydration sites in the lentil lectin crystal structures superimposed on the backbone of the lentil lectin dimer. In order to emphasize the twofold symmetry, the waters at the dimer interface are shown for both lectin monomers. Reprinted with permission from Loris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

Copyright © 2006 International Union of Crystallography

references