

23.4. SOLVENT STRUCTURE

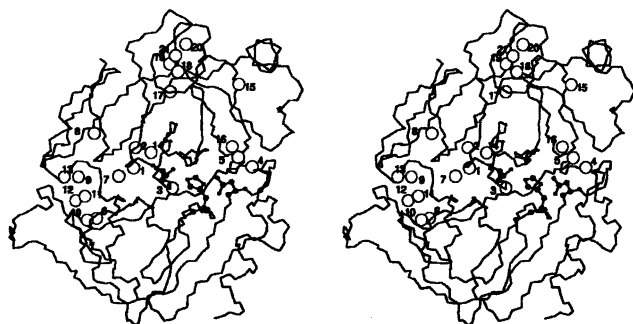


Fig. 23.4.4.1. Stereoview of the set of 21 highly conserved buried waters in eukaryotic serine proteases. The trypsin backbone is represented as a stick drawing, with the catalytic triad at the centre (filled circles). Water molecules are represented as open circles. Reprinted with permission from Sreenivasan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

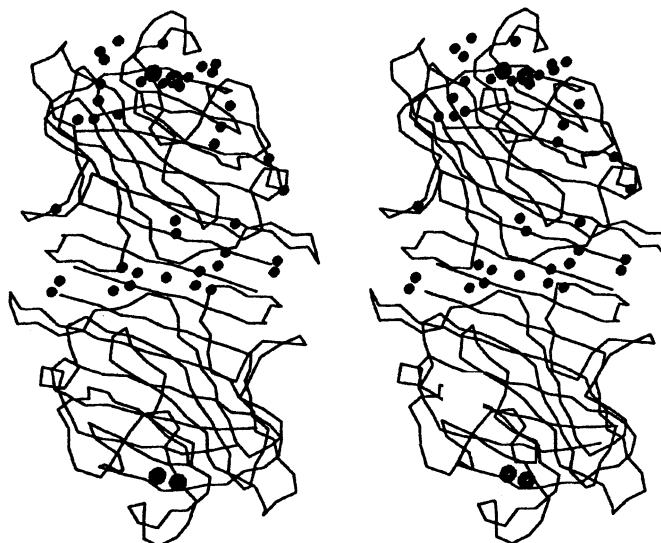


Fig. 23.4.4.2. View of the 33 conserved hydration sites in the lentil lectin crystal structures superimposed on the backbone of the lentil lectin dimer. In order to emphasize the twofold symmetry, the waters at the dimer interface are shown for both lectin monomers. Reprinted with permission from Loris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

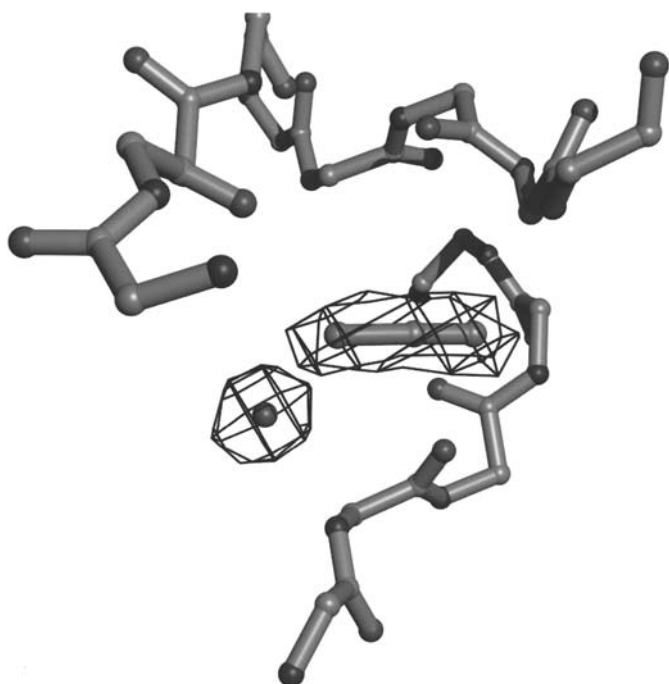


Fig. 23.4.4.3. A $2F_o - F_c$ electron-density map contoured at the 1.2σ level shows a distinct ellipsoidal density for acetonitrile 707 and a spherical density for a nearby water molecule. The protein backbone of the binding pocket is represented with nitrogen atoms shown in dark grey, oxygen atoms in medium grey and carbons in a lighter grey. *MOLSCRIPT* (Kraulis, 1991) was used in the preparation of this figure. Reprinted with permission from Allen *et al.* (1996). Copyright (1996) American Chemical Society.

23.4.4.2.1. Elastase

The crystal structure of porcine pancreatic elastase was solved in a variety of organic solvents, with the primary goal of mapping binding sites on the protein that could accommodate molecules representative of functional groups likely to be found in larger ligands (Ringe, 1995; Mattos & Ringe, 1996; Mattos *et al.*, 2000). Crystals of elastase cross-linked with glutaraldehyde were transferred to the following solutions: 100% acetonitrile, 95% acetone, 55% dimethylformamide, 80% ethanol, 40% trifluoroethanol, 80% isopropanol and 80% 5-hexene-1,2-diol (Allen *et al.*, 1996; Mattos & Ringe, 1996). In general, the crystals did not diffract in most neat organic solvents. However, in the acetonitrile case, where they did, the result was striking. In the structure of elastase solved in >99% acetonitrile, there were 126 water molecules visible in the electron-density maps, indicating that a good portion of the first hydration shell of the protein was still present. In contrast, only nine molecules of acetonitrile were clearly identified in the electron-density maps (Allen *et al.*, 1996). This is a powerful assertion of the evolutionary specificity of water molecules for protein surfaces. Fig. 23.4.4.3 shows the clear contrast between the elongated electron density of an acetonitrile molecule and the spherical electron density of a water molecule.

A similar result was obtained for all of the elastase structures solved in the mixtures of organic solvent and water mentioned above. A total of 11 structures were analysed, each containing 126–177 water molecules. The structures are listed in Table 23.4.4.1, together with the resolution of the data collected, the number of water molecules present and the number of organic solvent molecules observed in each case. The $C\alpha$ superposition of the protein atoms in the 11 structures yielded a total of 1661 individual water molecules, occupying 426 unique water-binding sites on the elastase surface. Given that elastase has a total of 240 amino-acid

residues, this represents a significant portion of the first hydration shell of the protein. This group of elastase structures served as a powerful source of information, leading to a classification of water types according to their interaction with the protein and an analysis of the specificity for water within each of the types determined (Bellamacina *et al.*, 1999).

All of the 1661 water molecules were renumbered according to the site on the protein where they were found. Any two water molecules within 1 Å of a water molecule in the cross-linked elastase structure solved in distilled water (used as the reference structure) have a common number. 39 of the 426 water-binding sites were occupied in every one of the 11 structures and were considered structurally conserved. Among these are the 16 buried water-binding sites thought to be conserved among all serine proteases (Sreenivasan & Axelsen, 1992). The 26 remaining conserved water molecules are specific to elastase and are not necessarily buried. These water molecules in general tend to have low B factors, but a few have B factors in the 30–35 Å² range and one conserved water molecule has a B factor of 42 Å².

The classification of the water sites as buried, channel, crystal contact or surface was based on the number of hydrogen-bonding interactions that a water molecule at the site could make to the protein and involved no surface-accessibility calculations (Bellamacina *et al.*, 1999). Water molecules were classified as buried if they made at least three good hydrogen-bonding interactions with protein main-chain atoms. A total of 23 buried water sites were identified in this manner, including 13 of the sites classified as buried by Sreenivasan & Axelsen (1992). One of the 16 serine protease conserved water-molecule sites is replaced by a His side chain in elastase (Sreenivasan & Axelsen, 1992). The remaining two serine protease conserved water sites were classified as channel based on the criteria used in the present study (see below). Interestingly, with the exception of these two channel water molecules, all of the buried sites found to be conserved in serine proteases are strictly conserved in all of the 11 structures in Table 23.4.4.1. The two channel water molecules are found in the aqueous structures of elastase, but are virtually absent in elastase transferred to organic solvents.

The water molecules occupying the 23 buried water sites identified in this study are tightly clustered when the protein $C\alpha$ atoms are superimposed by least squares, and the interactions with the protein are conserved from structure to structure. Fig. 23.4.4.4 shows the positions of the buried water-binding sites in elastase. In general, they are found in the cleft between the two domains, in bridging elements of the secondary structure and at the base of water channels. This observation is consistent with the current

Table 23.4.4.1. Multiple-solvent crystal structures of elastase

Structure	Resolution (Å)	No. of water molecules	No. of organic solvent molecules
Cross-linked	1.9	165	0
Acetonitrile	2.2	126	9
Acetone	2.0	126	6
Dimethylformamide	2.0	153	6
Ethanol	2.0	135	12
Trifluoroethanol (1)	1.9	175	4
Trifluoroethanol (2)	1.85	177	3
Isopropanol	2.2	160	4
Benzene	1.9	162	4
Cyclohexane	1.95	135	7
5-Hexene-1,2-diol	2.2	147	5

23.4. SOLVENT STRUCTURE

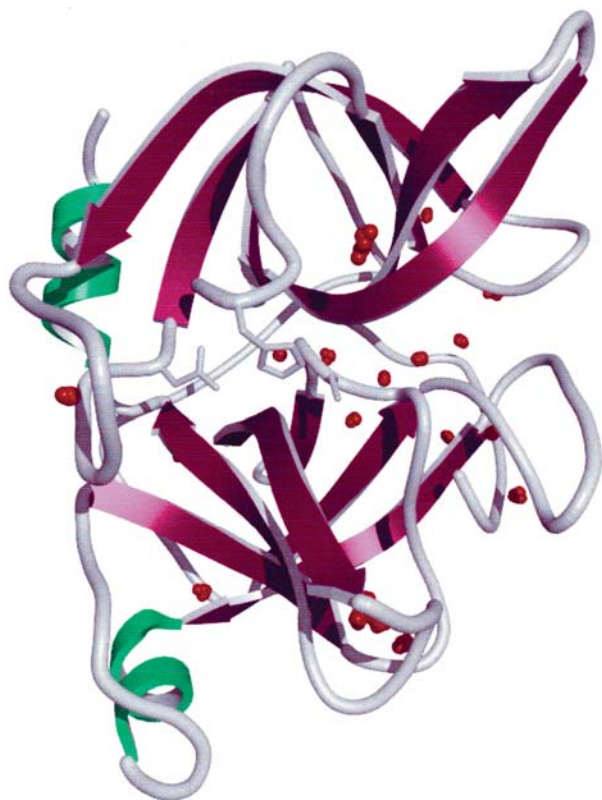


Fig. 23.4.4.4. Crystal structure of porcine pancreatic elastase represented as a ribbon diagram using *MOLSCRIPT* (Kraulis, 1991). The two α -helices are shown in green, the β -sheets are in purple and the coils are in grey. Elastase contains 240 amino-acid residues, and is composed of two β -barrel domains. The catalytic triad (Asp108, His60 and Ser203) is shown explicitly. The buried crystallographic water molecules found in 11 superimposed elastase structures solved in a variety of solvents are shown in red.

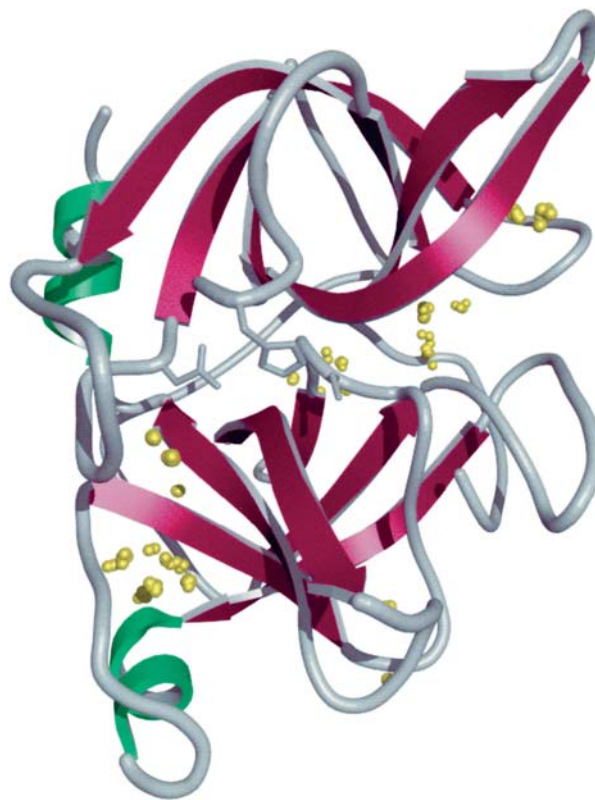


Fig. 23.4.4.5. Elastase structure represented as in Fig. 23.4.4.4. The crystallographic water molecules found in channels in 11 superimposed elastase structures solved in a variety of solvents are shown in yellow.

understanding of the functional roles played by structurally conserved water molecules as discussed above and in the following sections.

The 29 water-binding sites classified as channel contain water molecules that make hydrogen bonds with at least two other water molecules within a protein groove. The analysis of a high-resolution crystal structure of elastase (1.65 Å) revealed seven channels with a total of 32 water-binding sites (Meyer *et al.*, 1988). All of these channels were also identified in the analysis of the 11 structures in Table 23.4.4.1 (Bellamacina *et al.*, 1999). In addition, two other channels were observed. The locations of the nine elastase channels identified by the new criteria are shown in Fig. 23.4.4.5. Channels are often found in areas associated with buried water molecules, namely, at the crevice between the two domains and sandwiched between secondary-structure elements, where they lead from the surface of the protein to a buried water molecule. Fig. 23.4.4.5 also shows that the $C\alpha$ superposition of the protein structures leads to a spread of water molecules within the channels. In any given structure, only two or three water molecules may be present, but the precise location and interaction with protein atoms vary so that when taken together the collection of structures gives a sense of flow inside the channels.

Of the remaining 374 water-molecule sites present within the 11 elastase structures included in this study, 56 were classified as crystal-contact sites and 318 as surface sites. Crystal-contact sites were considered to be occupied by water molecules that are within 4.0 Å of a symmetry-related protein molecule in the crystal. Fig. 23.4.4.6 shows the position of all the water molecules found to occupy these sites. The relatively large number of crystal-contact

water-binding sites is a result of the somewhat broad criterion used to select them. Many of these sites are not within hydrogen-bonding distance from the nearby protein molecule, and most are not well conserved from structure to structure. Only eight of the 56 sites are occupied in the majority of the structures, and four of these make good multiple hydrogen bonds with two symmetry-related protein molecules in the crystal. These four water molecules seem to be structurally significant in the formation of the crystal contacts.

Surface water molecules were taken to be those that interact with side-chain protein atoms on the surface or make no more than two hydrogen-bonds with backbone atoms. When the 11 structures are superimposed, the surface water molecules occupying a given site are not tightly clustered. Furthermore, there is flexibility in the interactions between these water molecules and the nearby protein atoms. For example, it is often the case that all water molecules within a surface site make two or three hydrogen bonds to protein atoms, but only one of them is conserved in all of the structures where the water molecule is present at the site. Fig. 23.4.4.7 illustrates the position of all of the surface water-binding sites. Although over half of these sites are occupied in at least two of the 11 structures, a good proportion of them (178) are found in only one of the structures considered.

While crystal-contact and surface water sites were classified separately, it is important to point out that, with the exception of the four crystal-contact water-binding sites mentioned above, the crystal-contact sites exhibit very much the same traits as the surface water sites. The difference is that in the latter case, the 'surface' is provided by a single protein molecule, while in the former the interaction between two symmetry-related protein molecules constitutes the surface with which the water molecules interact.

Of the 318 surface water molecules, 21 are in the active site. The active-site water molecules were selected to be those within 4 Å of any atom belonging to either the trifluoroacetyl-Lys-Phe-*p*-

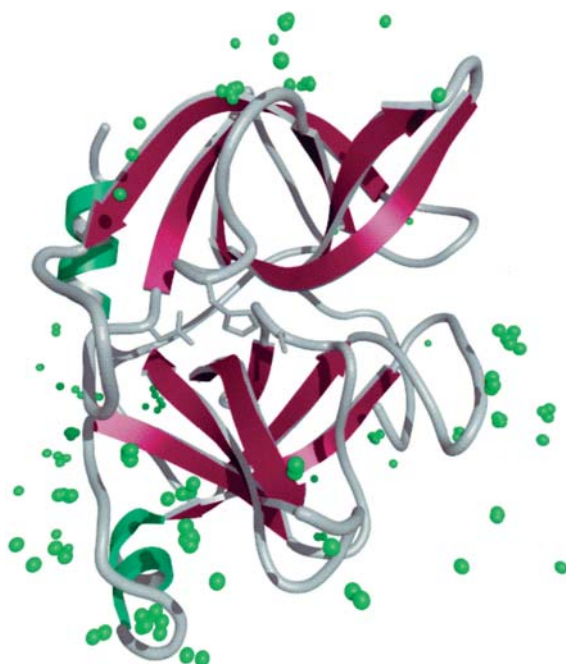


Fig. 23.4.4.6. Elastase structure represented as in Fig. 23.4.4.4. The crystallographic water molecules involved in crystal contacts in 11 superimposed elastase structures solved in a variety of solvents are shown in green.

isopropylanilide (Mattos *et al.*, 1994) or the trifluoroacetyl-Lys-Pro-*p*-trifluoromethylanilide (Mattos *et al.*, 1995) inhibitors in the structures of their complexes with elastase. These inhibitors span a large area of the active site, including an exosite not occupied by

substrate analogue inhibitors (Mattos *et al.*, 1994, 1995). The water-binding sites in the active site are not very well conserved, with most sites represented in only two to four of the 11 structures. When all of the structures are superimposed, there is at least one water molecule in each of the subsites in the elastase active site. These water molecules are displaced either by inhibitors or by organic solvent molecules in the various structures. It is not surprising that in elastase, a protein with relatively broad substrate specificity, the active site in the uncomplexed native protein is populated by many displaceable surface water molecules. With the exception of a water molecule present in the oxyanion hole, these water molecules tend to make a single hydrogen bond with the protein. This hydrogen-bonding interaction is not generally conserved between different structures where a given site is occupied in multiple structures. The displacement of these water molecules upon ligand binding is entropically favourable, as they are released into bulk solvent, without too much enthalpic cost. This relatively small enthalpic cost can be compensated by the protein–ligand interactions.

Fig. 23.4.4.8 shows all of the 1661 water molecules colour-coded by the various classifications described above. Clearly, the entire surface of the protein is well hydrated. Notice how the yellow channel waters are often followed by a red buried water molecule. In addition, there is often no obvious spatial distinction between molecules categorized as crystal contacts (green) and those categorized as surface (blue).

23.4.4.2.2. *T4 lysozyme*

Over 150 mutants of T4 lysozyme have been studied to date, and, for the majority of these, the crystal structures are available. Although most of the mutant structures crystallize isomorphously to the wild type, many of them provide a view of the molecule in different crystal environments. This collection of structures leads to

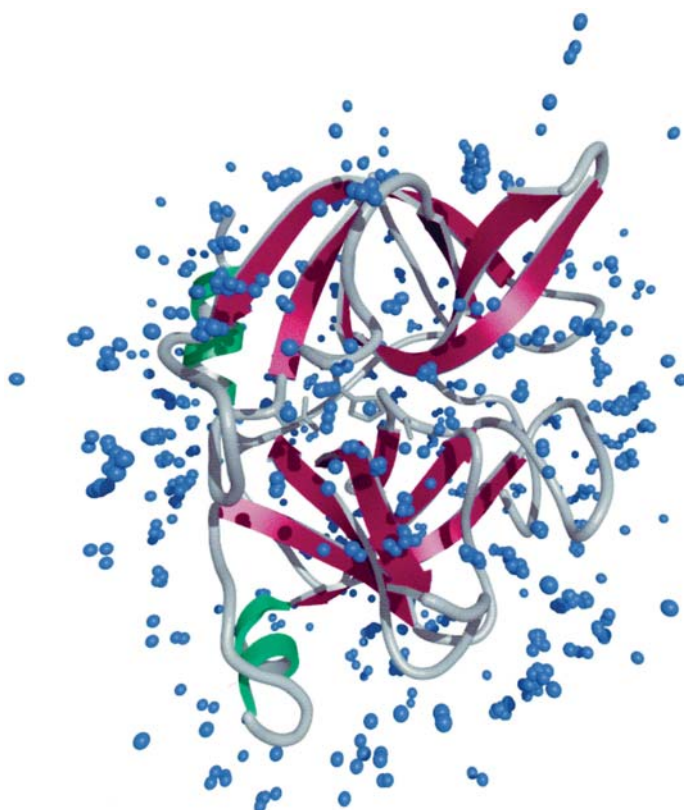


Fig. 23.4.4.7. Elastase structure represented as in Fig. 23.4.4.4. The surface crystallographic water molecules found in 11 superimposed elastase structures solved in a variety of solvents are shown in blue.

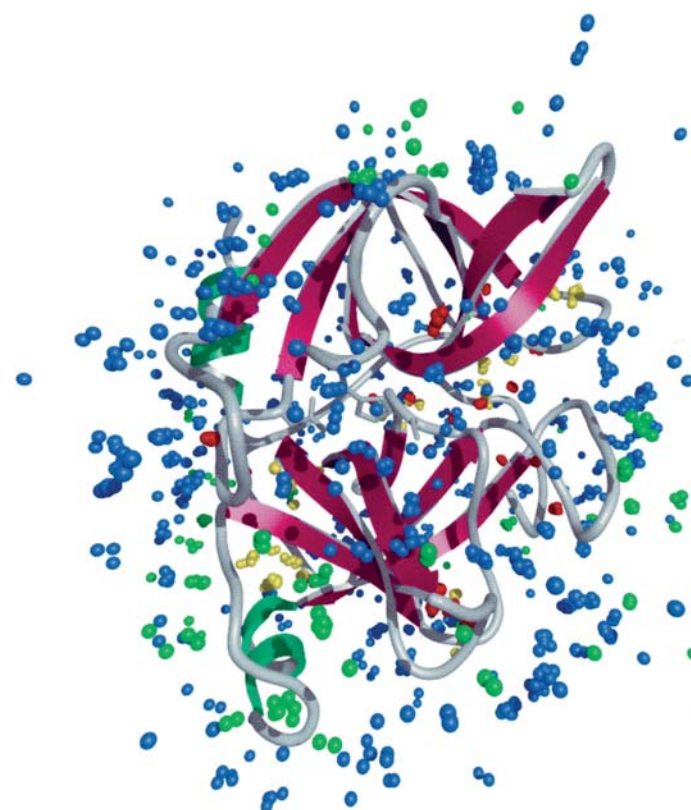


Fig. 23.4.4.8. Elastase structure represented as in Fig. 23.4.4.4. The 1661 water molecules found in 11 superimposed elastase structures of elastase are colour-coded as in Figs. 23.4.4.4–23.4.4.7.

23.4. SOLVENT STRUCTURE

the comparative analysis of the solvent positions in ten different crystal forms of T4 lysozyme, providing a clear picture of the effect of crystal contacts on the hydration sphere of a protein viewed by X-ray crystallography (Zhang & Matthews, 1994). The resolution and degree of refinement of the structures involved varied significantly, from 2.6 to 1.7 Å resolution, and the number of water molecules included per protein molecule ranged from 38 to 160. Nevertheless, this study revealed important features. A striking observation is that 95% of the solvent-exposed residues on T4 lysozyme were involved in at least one crystal contact in one or another of the crystal forms studied, showing that any part of the protein surface can be involved in crystal contacts. A corollary to this finding is that any of the surface water molecules can be displaced or involved in bridging protein-protein contacts in the crystal.

Of the 1675 individual water molecules observed in the 18 independently refined T4 lysozyme molecules included (Fig. 23.4.4.9), the ones that were within a sphere of radius 1.2 Å were considered to occupy the same site on the protein. As in the case of elastase described above, all of the water molecules observed upon superposition of the 18 T4 lysozyme structures represent a large portion of the first hydration shell. This reinforces the concept that multiple structures of a protein of interest provide a more complete picture of the protein hydration than possible with a single structure. There are four buried water sites that are occupied in at least 15 out of the 18 structures and are independent of crystal contacts. Two of these buried sites are at the hinge-bending region between the two helical domains and appear to play a functional role in the opening and closing of the active site (Weaver & Matthews, 1987). The other two play a structural role at the protein core. Other than the four buried water molecules, the most conserved water sites appear at the active-site cleft between the two domains and at the N-termini

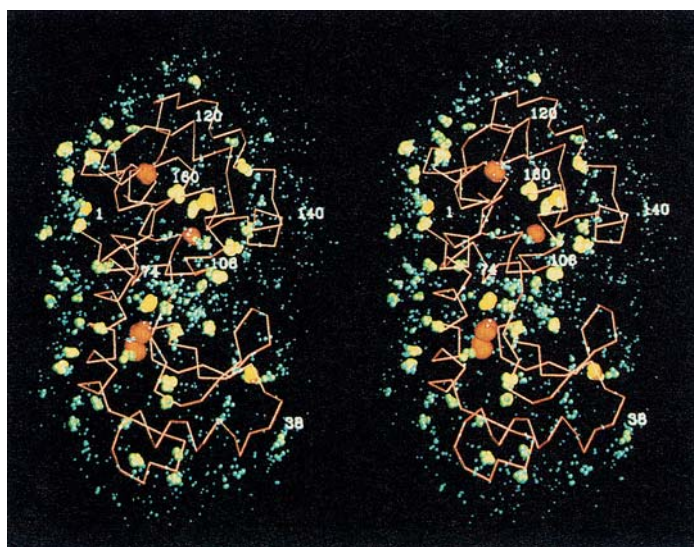


Fig. 23.4.4.9. Distribution of solvent-binding sites in 18 mutant T4 lysozymes from ten refined crystal structures. The lysozyme structures were compared to identify common sites of hydration. A total of 1675 solvent molecules were included in the comparison. Each solvent molecule is represented by a coloured sphere. The size of the sphere is proportional to the number of lysozyme structures in which solvent was observed at the same site (*i.e.* within 1.2 Å). In addition, the colour of the solvent changes from blue for the least-conserved sites to red for the most-conserved ones [*e.g.* the red spheres indicate that solvent is observed with high frequency (15–17 times) at the four internal sites]. The numbers indicate representative residue positions along the backbone of the lysozyme molecule. Reprinted with the permission of Cambridge University Press from Zhang & Matthews (1994). Copyright (1994) The Protein Society.

of α -helices. As is the case in the previous works reviewed above, the 20 most conserved water sites appear in well conserved protein environments and generally have low temperature factors. Buried or highly conserved water molecules also tend to make at least three hydrogen bonds with protein atoms or other water molecules. The less-conserved water sites appear more randomly on the protein surface and are strongly influenced by the particular crystal environment in which the structure was solved.

23.4.4.2.3. Ribonuclease T1

A group of four crystal structures of ribonuclease T1 in complex with guanosine, guanosine-2'-phosphate, guanylyl-2',5'-guanosine and vanadate were used for an analysis of conserved water positions that contribute to the structural stabilization of the protein (Malin *et al.*, 1991). The four structures were obtained from isomorphous crystals and ranged in resolution from 1.7 to 1.9 Å. Conserved water molecules were considered to be those found within a sphere of 1 Å from each other in all four structures. All other water molecules were excluded from the analysis. 30 water molecules were found to be conserved. Of these, ten were observed near crystal contacts, although only one appears to be dictated by the crystal contact itself, making a single hydrogen bond with each of the symmetry-related protein molecules. Ten other water molecules form a channel that brings together an α -helix and a hairpin-like loop structure and then go on to wrap around the calcium ion, providing half of its coordination sphere. The first five of these water molecules are completely buried, holding together the two secondary-structure elements, which would otherwise collapse (Malin *et al.*, 1991). Two water molecules are found to stabilize the N and C termini, which are brought together by a disulfide bond. The remaining eight conserved water molecules hold together various elements of secondary structure or are located in the active site.

An interesting extension to this study included four additional structures: the E58A mutant in complex with guanosine-2'-monophosphate, the H92A mutant crystallized under two different conditions and wild-type RNase T1 in complex with guanosine-3',5'-biphosphate. Two of these crystal forms were not isomorphous with the native protein crystals or with each other. Thus a total of eight structures solved in three different space groups were analysed (Pletinckx *et al.*, 1994). Although the effect of crystal packing on the three-dimensional structure of the protein is minimal, there are some significant differences in the solvent structure. In particular, there is no evidence of the calcium-binding site and its coordinating water structure in any crystal forms other than the canonical wild type. Instead, the E58A mutant has a sodium-binding site at a different position, along with three previously unobserved water molecules. It is clear that the presence of the metal ions is fortuitous and linked to the crystallization conditions.

There are 25 water molecules structurally conserved throughout the different packing arrangements studied. Ten of these are single sites, there are three clusters of two water molecules and a larger cluster originally described by Malin *et al.* (1991) to hold together the core of the protein. As was observed for the study on T4 lysozyme (Zhang & Matthews, 1994), the strictly conserved water-binding sites present in crystal structures solved across different space groups are involved in bridging protein secondary-structure elements and seem to be crucial for the integrity of the protein structure.

23.4.4.2.4. Ribonuclease A

Ribonuclease A is not homologous to ribonuclease T1 in either sequence or structure, but both have evolved to catalyse the same reaction with specificity for different substrates (compare Figs. 23.4.4.10 and 23.4.4.11). Ribonuclease A cleaves RNA after pyrimidines, while ribonuclease T1 cleaves specifically after

23. STRUCTURAL ANALYSIS AND CLASSIFICATION

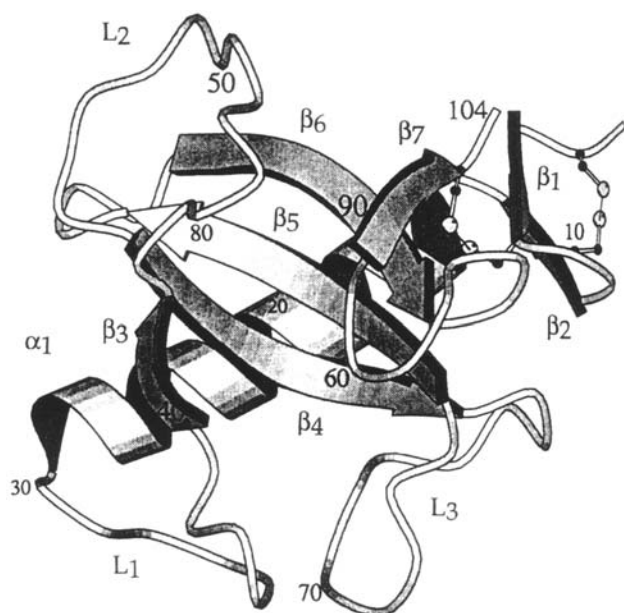


Fig. 23.4.4.10. Three-dimensional structure of RNase T1. Secondary structure is denoted as follows: α_1 , α -helix; β_n , strands of β -sheet structure; L_n , loops. Drawn using *MOLSCRIPT* (Kraulis, 1991). Residue numbers indicate the beginning and end of secondary-structure elements. Reprinted with permission from Pletinckx *et al.* (1994). Copyright (1994) American Chemical Society.

guanine. Therefore, the information obtained from a study of the solvent structure in ribonuclease A is completely independent from that described above for ribonuclease T1. A collection of ten crystal structures of ribonuclease A, derived from five different crystal forms, were compared pairwise after least-squares superposition (Zegers *et al.*, 1994). 17 conserved water molecules were found to be within a sphere of 0.5 Å of each other in all of the ten structures

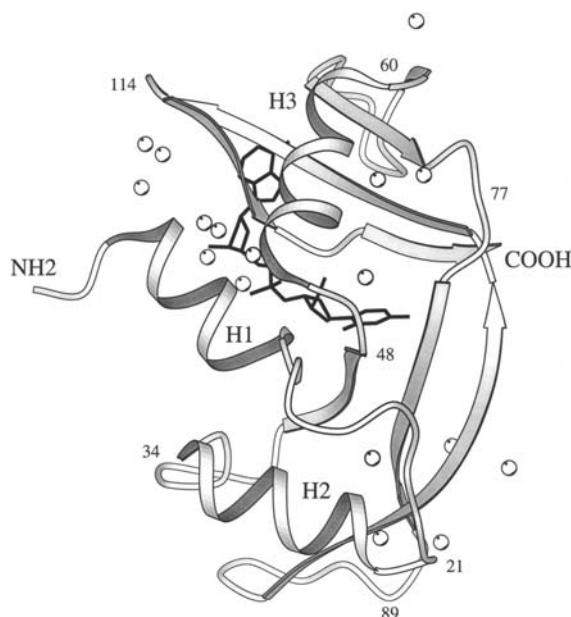


Fig. 23.4.4.11. Overall structure of RNase A. The overall structure of the d(CpA) complex of RNase A is shown as a ribbon drawing using *MOLSCRIPT* (Kraulis, 1991). The conserved water molecules are shown as white spheres and the d(CpA) inhibitor in black. The three helices are labelled H1, H2 and H3. Reprinted with the permission of Cambridge University Press from Zegers *et al.* (1994). Copyright (1994) The Protein Society.

and are shown in Fig. 23.4.4.11. These water molecules were found in small clusters of two or three or as part of a larger solvent network. Not surprisingly, they form multiple hydrogen bonds with the protein and generally have low temperature factors. Of the 17 structurally conserved sites, 13 are associated with one of the three α -helices. Most of these link the helices to one of the β -strands. Three water molecules are involved in hydrogen bonding with unpaired amido and carbonyl groups on the protein, and one is found on top of the β -pleated sheet. These interactions result in bringing together elements of secondary structure and in stabilizing distortions within these elements. Conserved water molecules are also responsible for bridging the N-terminal helix to the C-terminal β -strand, which form the two halves of the active site.

23.4.4.2.5. Protein kinase A

The comparative study of water molecules in seven different protein kinase A structures in complex with different ligands focused exclusively on the active site (Shaltiel *et al.*, 1998). All of the structures were solved from isomorphous crystals to resolutions ranging from 2.0 to 2.9 Å. The more lenient cutoff of 1.5 Å for the radius of the sphere within which the conserved water molecules must be found among the different structures is consistent with the relatively low resolutions of the structures included in this study. The group of structures represents the open, the closed and an intermediate conformation of the catalytic kinase domain. There is a set of six conserved water sites in the active site, in addition to the ATP molecule and the magnesium ion. The conserved water molecules coordinate to ATP, the metal ion and a conserved Tyr residue from the carboxyl terminus of the protein. Thus, the active site consists of an extended network of interactions that weave together both domains of the core, with water molecules playing an integral role in maintaining the structural features important for catalysis. Many of these water molecules associate directly with the inhibitors. In addition, five water sites are observed in positions that would be occupied by substrates or substrate analogues. These water molecules are displaced by ligand oxygen atoms that can compensate for the water hydrogen-bonding interaction with the protein.

23.4.4.3. Summary

Water molecules associated with proteins can be divided between those that are conserved as a result of their functional significance and those that are partially conserved or not conserved at all. The conserved water molecules are generally classified as buried or channel (by a variety of criteria). They tend to be present in the clefts between domains, are critical components of active sites, or bridge between secondary-structure elements. The water molecules that are not conserved occupy hydration sites with favourable hydrogen-bonding characteristics, where the presence of a water molecule is not essential for the structural or functional integrity of the protein.

The displacement of water molecules by organic solvent molecules in the elastase work described above showed that most displaced waters are those classified as surface or crystal-contact waters (Mattos *et al.*, 2000). In the three cases where a buried water molecule was displaced, an alcohol hydroxyl oxygen was found to replace the protein–water hydrogen-bonding interactions. This is analogous to the active-site water molecule in the HIV aspartate protease that gets replaced by a carbonyl group of a potent cyclic urea inhibitor (Lam *et al.*, 1994). In these situations, release of a tightly bound water molecule is entropically favourable, and its enthalpic interactions with the protein are compensated by similar protein–ligand interactions.

The effect of crystal contacts on the water structure was clearly illustrated in the T4 lysozyme work (Zhang & Matthews, 1994).