The internal structurally conserved water molecules are unaffected by crystal contacts. Conversely, any of the surface water sites are potentially available either to be replaced by or to mediate crystal contacts, as 95% of the T4 lysozyme surface is involved in a crystal contact when all ten crystal forms are taken together.

23.4.5. The classic models: small proteins with high-resolution crystal structures

Crambin and BPTI are among the handful of proteins for which X-ray crystal structures have been obtained to 1 Å resolution or better. In general, these proteins are relatively small (BPTI, the largest in this group, has 58 amino-acid residues) and often contain at least one disulfide bond. These high-resolution crystal structures have provided structural information beyond that available for larger proteins, particularly with respect to the surface solvent structure. Their small size renders their structures accessible by NMR techniques, making it possible to assess the effect of the crystal environment on the protein and water structure. Finally, the available detail and precision of the structures, as well as their small size, make them ideal models in computational studies of protein energetics and dynamics. Both crambin and BPTI were used during the pioneering years of protein molecular-dynamics calculations. In this section, special attention is given to crambin and BPTI as representative proteins for which very high resolution structures are available. Focus is on the features of solvent structure that are not available for other proteins.

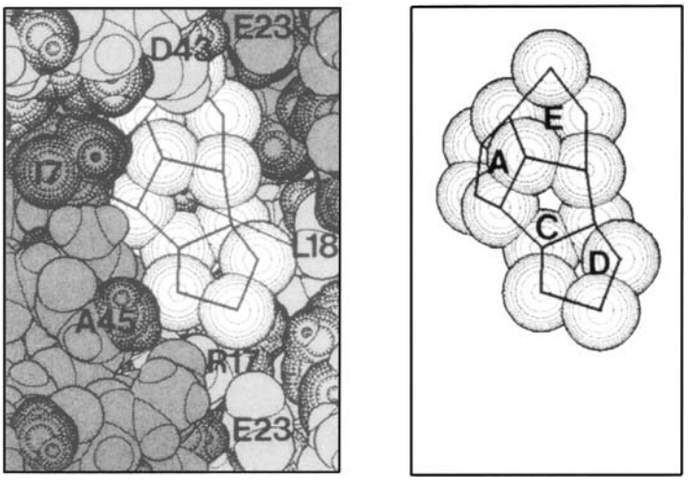

Fig. 23.4.5.1. van der Waals surface diagram of the water pentagons A, C, D and E in crambin viewed in the negative **a** direction. Rings A, C and E form a cap around leucine 18. Hydrophobic atoms are shown as dark circles, and water oxygens are shown as light circles. The methyl group of leucine 18 can be seen through the C ring. Adjacent translationally related molecules are shaded. The van der Waals radii used for the protein C, N and O atoms are 1.7, 1.4 and 1.4 Å, respectively, and for water oxygen, 1.8 Å. The larger radius is used for the water oxygens because hydrogen atoms have been omitted. Reprinted with the permission of the author from Teeter (1984).

23.4.5.1. *Crambin*

Crambin is a plant-seed hydrophobic protein of unknown function. It contains 46 amino-acid residues and was reported to form crystals that diffract to 0.88 Å resolution (Teeter & Hendrickson, 1979). The crystal structure of crambin was determined to 0.945 Å resolution directly from anomalous scattering by the six sulfur atoms involved in three disulfide bonds (Hendrickson & Teeter, 1981). Crambin is an amphipathic molecule in that the hydrophilic components (including six charged groups) are segregated from a mainly hydrophobic surface.

A total of 64 water molecules and two ethanol molecules were located in the electron-density map, despite the fact that the structure was determined in 60% ethanol. The overwhelming number of water molecules compared to ethanol is consistent with the results of the multiple-solvent crystal structures experiments described above for elastase (Mattos & Ringe, 1996).

Most of the 64 water molecules found in crambin interact with polar side chains in the typical manner described previously. The unusual information about solvent structure offered by the crambin model is that the arrangement of water molecules around hydrophobic residues is similar to that observed for clathrate hydrate structures (Teeter, 1991). Pentagonal water rings are observed to cap the Cδ2 atom of Leu18 as well as the hydrophobic methylene groups of Arg17 (Teeter, 1984, 1991). The set of five connected water rings is shown in Fig. 23.4.5.1. This ring cluster extends toward the protein, forming heterocyclic rings that are described in detail in the original article (Teeter, 1984).

Although crambin provides the clearest example of pentagonal water rings on a hydrophobic protein surface, it is not the only one. Other high-resolution crystal structures (better than 1.4 Å), such as insulin and cytochrome *c*, have also revealed pentagonal rings, but never to the extent seen in crambin (Teeter, 1984). This is very likely to be a general mode of interaction between water and hydrophobic moieties, be it in inorganic, organic, or biological molecules. The fact that it is not observed in protein structures in general may be related to the lower resolution of most X-ray structures, where it is not possible to model the more disordered areas where these patterns are likely to be found.

23.4.5.2. *Bovine pancreatic trypsin inhibitor*

Bovine pancreatic trypsin inhibitor (BPTI) is a protein of 58 amino-acid residues whose X-ray crystal structure was obtained in the original crystal form to 1.5 Å resolution (Deisenhofer & Steigemann, 1975). Subsequently, 1 Å X-ray data were obtained from a different crystal form, and the new model was jointly refined with 1.8 Å neutron diffraction data (Wlodawer *et al.*, 1984). Minor differences in structure between the two crystal forms of BPTI were observed (Wlodawer *et al.*, 1984). The interesting contribution of the 1 Å model to the understanding of solvent structure resulted from the ability to refine occupancy at this resolution. A total of 63 water molecules were placed in the model, 20 of them within 1 Å of a water molecule found in the structure solved in the original crystal form. During refinement against the 1 Å data set, full occupancy was assigned to all protein atoms, and water occupancy was allowed to refine. Of the 63 water-molecule positions, 29 were found to be fully occupied. The remaining 34 had partial occupancies, with 0.4 being the minimum occupancy found. Given that there are very few contacts between protein molecules in the crystal (Wlodawer *et al.*, 1984), it is reasonable to assume that this observation is representative of water occupancies on protein surfaces in general. It is likely that well over half of the water positions found on protein surfaces are less than fully occupied, although there is no definitive proof that this is true.

23.4.5.3. *Summary*

In general, small proteins serve as important models where results of X-ray crystallography, NMR and molecular-dynamics calculations can be easily compared and cross-validated, since larger proteins are more difficult to study by the latter two techniques. Small proteins are also more likely to form relatively ordered crystals, which are able to diffract X-rays to atomic resolution (of the order of 1 Å). With respect to understanding

solvent structure, the two major contributions of these very high resolution protein models are the observation of solvent structure around hydrophobic residues, where at lower resolution the water molecules 'look' disordered, and a glimpse at the pattern of water occupancy likely to occur on protein surfaces.

23.4.6. Water molecules as mediators of complex formation

The examples given in Section 23.4.4 illustrate two important roles played by water molecules at the binding sites of proteins: as structural water molecules and as displaceable water molecules. As a structural part of a binding site, water molecules are found to be strictly conserved. They are either involved in stabilizing the coming together of secondary-structure elements in a way that appropriately shapes the binding site, or they fill grooves on the protein surface, making it more specific for a given ligand. The second role involves the presence of less tightly bound, partially conserved water molecules that get displaced by the ligand upon binding. In the few examples where tightly bound water molecules are displaced by a ligand, the hydrogen-bonding interaction of the water with the protein is replaced by an atom on the ligand. A third role, not yet discussed, of water molecules in protein active sites is in the catalytic mechanism of enzymatic reactions. An extensive network of water molecules near the active site of serine proteases has been implicated in the catalytic mechanism of these enzymes (Meyer *et al.*, 1988; Meyer, 1992). If this hypothesis is indeed correct, it provides a good example of the cooperation between water molecules and protein atoms in the optimization of function. Unfortunately, it is difficult to explicitly detect catalytic water molecules crystallographically, due to the long data-collection time relative to a catalytic event. However, the development of time-resolved Laue diffraction methods has provided a view of the catalytic water molecule in some proteins, *e.g.* trypsin (Singer *et al.*, 1993), and progress is likely to continue in this area. This section focuses on a few particular examples of how water molecules mediate the formation of complexes, either in the active sites of enzymes or in the binding interface between macromolecules or protein–ligand complexes.

23.4.6.1. *Antigen–antibody association*

The X-ray crystal structures of the Fv fragment of the monoclonal antibody D1.3 and the structure of its complex with hen egg-white lysozyme were both solved to 1.8 Å resolution (Bhat *et al.*, 1994). This study revealed a significant number of water molecules contributing to the chemical complementarity at the antigen–antibody interface. There are 23 water molecules at the antigen-binding site of the free antibody fragment, while 48 are present mediating complex formation. Seven water molecules are in equivalent positions in the free and complexed antibody (within 1.5 Å). There is no net loss of water molecules at the combining site. In fact, the total number of water molecules at the antigen–antibody interface is not less, but more, than the sum of those in the free antibody combining site and in the antigenic determinant. Furthermore, there is a general decrease in *B* factors of the binding-site residues upon complex formation, implying a decrease in entropy (Bhat *et al.*, 1994). The structural results indicate that water molecules at the antigen–antibody interface play a variety of important roles. Some form an integral part of the active site, fine-tuning the shape and charge complementarity of the interaction. Others are found to be displaced during complex formation, and still others are unique to the complex, bridging between the two molecules in a variety of locations throughout the complex interface.

The structural analysis correlated well with results of calorimetric experiments that showed that complex formation is enthalpically driven, with an unfavourable entropic contribution (Bhat *et al.*, 1994). The authors suggest that water molecules play a central role in mediating complex formation and claim that the hydrophobic effect is not important in this case. This is an argument that goes contrary to the idea that affinity is contributed by hydrophobic interactions within a relatively small portion of the interface between the interacting molecules, with hydrogen-bonding and charge–charge interactions contributing primarily to specificity (Hendsch & Tidor, 1994; Clackson & Wells, 1995; Hendsch *et al.*, 1996).

23.4.6.2. *Protein–DNA recognition*

The *trp* repressor binds specifically to the target DNA sequence ACTAGT, resulting in the transcriptional control of L-tryptophan levels in bacteria. The crystal structure of the *trp* repressor/operator complex was solved to 1.9 Å resolution (Otwinowski *et al.*, 1988). Although the structure revealed hydrogen-bonding interactions between the protein and the backbone phosphate groups, no direct hydrogen bonds or non-polar contacts between the protein and DNA bases were observed. Specificity was therefore attributed to the effect of the sequence on the geometry of the phosphate backbone and to water-mediated polar contacts between protein atoms and specific DNA bases. To confirm this hypothesis, the 1.95 Å resolution crystal structure of the free decamer CCACTAGTGG was obtained, containing the recognition six-base-pair sequence (Shakked *et al.*, 1994). A comparative analysis of the free and complexed DNA showed that, when bound to the *trp* repressor, the six-base-pair region is bent by about 15° so as to compress the major groove, with concomitant expansion of the minor groove relative to the uncomplexed DNA (Shakked *et al.*, 1994). However, both free and complexed DNA are underwound, with 10.6 base pairs per turn, rather than the usual 10.0 base pairs per turn. This feature is presumably a result of the particular DNA sequence and is thought to decrease the energy barrier for the binding interaction with the *trp* repressor protein (Shakked *et al.*, 1994). Another specificity component suggested by the authors is conferred by the hydration of the consensus bases. Ten water molecules are observed to interact in the major groove at similar positions in both the free and complexed DNA. Three of these mediate in four hydrogen-bonding interactions to the protein in the complex. Interestingly, the DNA bases to which these three water molecules are bound are among the most conserved and mutationally sensitive bases of the operator. In effect, these three water molecules can be regarded as extensions of the DNA bases and part of the specific recognition elements of the target DNA sequence (Shakked *et al.*, 1994).

The idea of water molecules as mediators of interactions conferring specificity in protein–DNA associations is further supported by the co-crystal structure of the HNF-3/fork head DNA-recognition motif in complex with DNA, solved to 2.5 Å resolution (Clark *et al.*, 1993). Although the lower resolution of this protein–DNA complex may limit the unambiguous determination of water molecules to those that are tightly bound, a series of water molecules are observed in the major groove, bridging specific DNA bases to amino-acid side chains in one of the α-helices of the protein. In this case, direct hydrogen bonding between DNA bases and protein side chains also exists.

The involvement of water in specific protein–DNA recognition was further confirmed in a study of the accuracy of specific DNA cleavage by the restriction endonuclease *Eco*RI under different osmotic pressures (Robinson & Siglar, 1993). Changes in osmotic pressure, resulting from changes in osmolite concentrations, have direct effects on the number of water molecules associated with macromolecules (Rand, 1992). The *Eco*RI experiments show that

references