

23. STRUCTURAL ANALYSIS AND CLASSIFICATION

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 *EcoRI* 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 *EcoRI* experiments show that

23.4. SOLVENT STRUCTURE

water activity affects site-specific DNA recognition, with an increase in osmotic pressure leading to a decrease in accuracy of protein–DNA recognition, as observed by DNA cleavage at sites containing an incorrect base pair (Robinson & Siglar, 1993). The results of this study strongly imply a role for one or more water molecules in recognition of specific sequences of DNA. The authors suggest that water mediation may constitute a general motif for sequence-specific DNA recognition by DNA-binding proteins (Robinson & Siglar, 1993).

The role of water molecules as mediators of sequence-specific DNA recognition may be a general motif, but not a necessary one. The solution NMR structure of the complex of erythroid transcription factor GATA-1 with the 16-base-pair DNA fragment GTTGCAGATAAACATT, containing the recognition sequence, shows that the specific interactions between GATA-1 and the major groove of the DNA are dominated by van der Waals interactions hydrophobic in nature (Omichinski *et al.*, 1993). Furthermore, NMR experiments designed to identify the location of water molecules in the complex detected clusters of water molecules bridging the protein to the DNA phosphate backbone, but showed that water was excluded from the hydrophobic interface between the protein and the DNA bases (Clore *et al.*, 1994). Although many of the existing crystal structures of protein–DNA complexes support the general view that water molecules are often integral components of the specific recognition between the protein and the target DNA, this solution structure provides an important example of exclusion of water molecules from the specificity determinants. In the GATA-1–DNA complex, however, water molecules do mediate non-specific binding of the protein to the DNA backbone. It appears, not surprisingly, that water molecules play a variety of roles in the mediation of protein–DNA interactions and that these roles are specific to each particular case.

23.4.6.3. Cooperativity in dimeric haemoglobin

The X-ray crystal structures of liganded and unliganded dimeric haemoglobin from *Scapharca inaequalvis* have revealed that water molecules at the dimer interface form an integral part of the cooperativity mechanism in this system (Condon & Royer, 1994; Royer, 1994). The binding of oxygen to one of the monomers causes little rearrangement of quaternary structure. It does, instead, displace the side chain of Phe97 which, in the low-affinity deoxy form, packs in the haem pocket (Royer *et al.*, 1990). Phe97 in the deoxy form lowers the oxygen affinity by restricting movement of the iron atom into the haem plane (Royer, 1994). Upon oxygen binding, Phe97 flips to the dimer interface, removing six out of the 17 water molecules that are found in the deoxy form (Fig. 23.4.6.1). The resultant destabilization of the water clusters found between the two subunits facilitates the flipping of Phe97 in the other subunit, with a concomitant increase in oxygen affinity of the haem in the second subunit (Pardanani *et al.*, 1997; Royer *et al.*, 1997).

In each of the monomeric subunits, Thr72 is positioned to form a hydrogen bond with a water molecule at the periphery of the deoxy dimer interface (not shown in Fig. 23.4.6.1). In effect, this interaction caps the water cluster on either side of the interface, presumably helping to stabilize these well ordered water molecules. The isosteric mutation Thr72 to Val was designed to test the importance of this interaction to the stability of the water cluster in the low-affinity haemoglobin dimer and the resultant effect on ligand affinity and cooperativity (Royer *et al.*, 1996). The crystal structure of the T72V mutant was solved to 1.6 Å resolution. This crystal structure reveals that the only significant difference between the mutant and wild-type proteins is the loss of the two water molecules that directly hydrogen-bond to Thr72 in each of the wild-type subunits. Furthermore, there is a significant increase in both

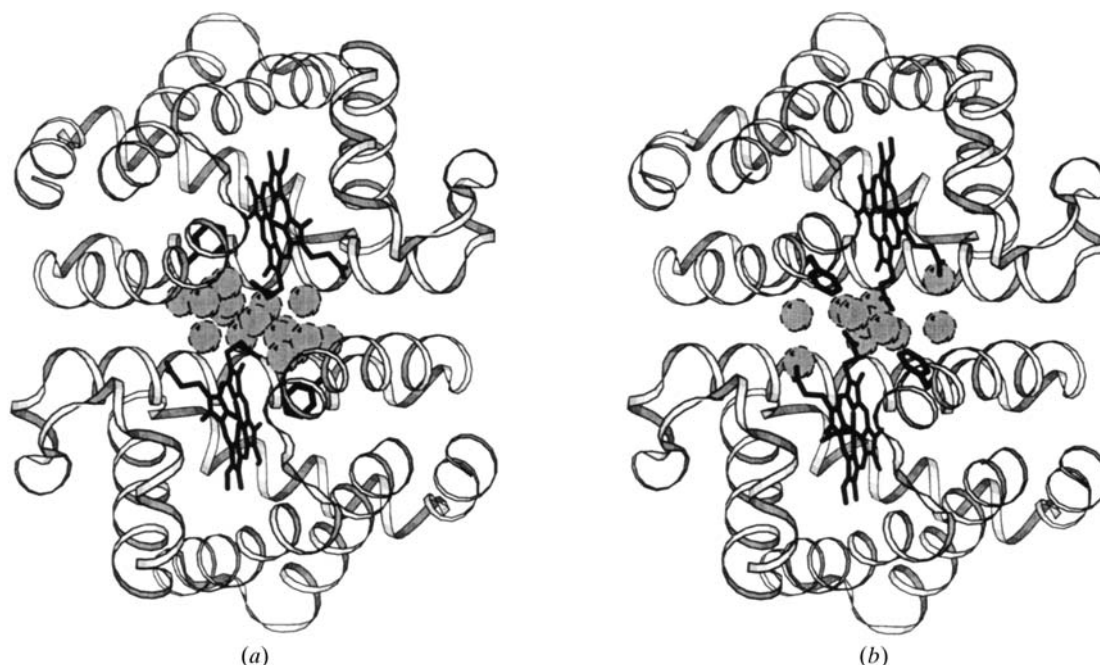


Fig. 23.4.6.1. *Scapharca* HbI interface water molecules. (a) Deoxy-HbI at 1.6 Å resolution (PDB code 3SDH) and (b) HbI-CO at 1.4 Å resolution (PDB code 4SDH). Included is a ribbon diagram showing the tertiary structure of each subunit, bond representations for the haem group and Phe97 side chain, and spheres representing the approximate van der Waals radii of oxygen atoms for core interface water molecules. Note the cluster of 17 ordered water molecules in the interface of deoxy-HbI for which Phe97 is packed in the haem pocket. Upon ligation, by either CO or O₂, Phe97 is extruded into the interface and disrupts this water cluster, expelling six water molecules from the interface. These plots were produced with the program *MOLSCRIPT* (Kraulis, 1991). Reprinted with permission from Royer *et al.* (1997). Copyright (1997) The American Society for Biochemistry & Molecular Biology.