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 inaequivalvis* 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

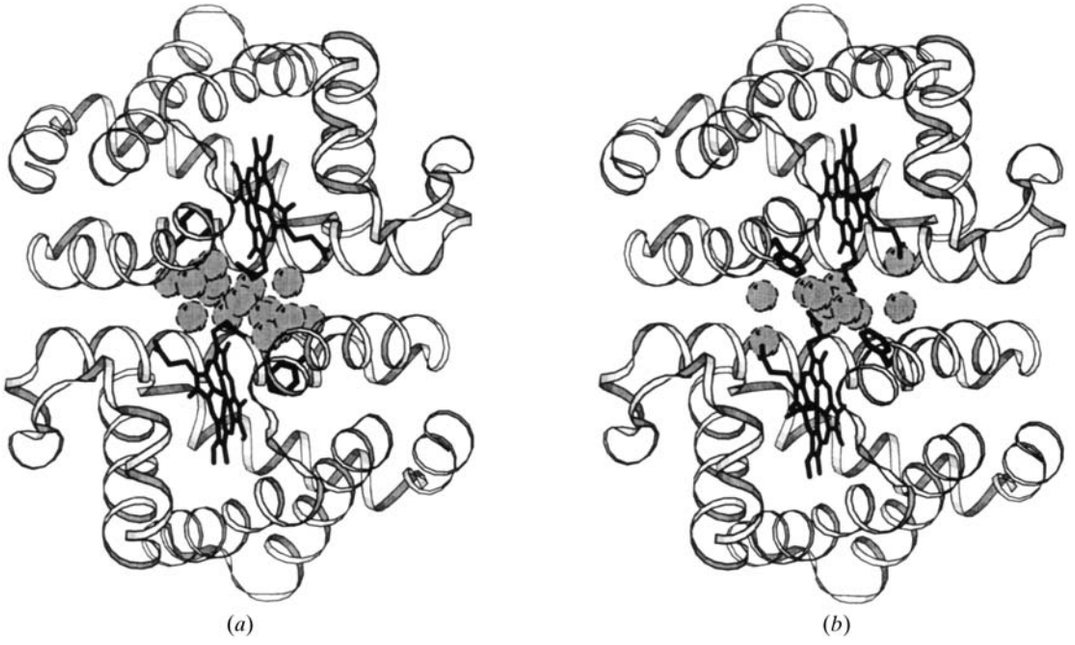

Fig. 23.4.6.1. *Scapharca* HbI interface water molecules. (*a*) Deoxy-HbI at 1.6 Å resolution (PDB code 3SDH) and (*b*) HbI-CO at 1.4 Å resolution (PDB code 4SDH). Included is a ribbon diagram showing the tertiary structure of each subunit, bond representations for the haem group and Phe97 side chain, 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 O_2, 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.

activity and cooperativity resulting from the mutation (Royer *et al.*, 1996). The authors conclude that, as a result of the mutation, the loss of two water molecules in the interface cluster is sufficient to alter the balance between the low- and high-affinity forms of the protein. This result demonstrates that water molecules are key mediators of information transfer between the haems in the two subunits in dimeric haemoglobin and that their precise positioning and interactions with protein atoms are crucial in maintaining the chemical balance required for biological function.

23.4.6.4. *Summary*

The few examples illustrated above provide diverse views of the ways in which Nature can use water molecules as integral parts of macromolecular interactions. Water molecules can be involved in specificity and recognition, in thermodynamics of binding and affinity, in the cooperative behaviour of allosteric proteins, and in catalysis. Not only do the specific examples illustrate general roles possible for water molecules in the context of a given type of macromolecule, such as proteins or nucleic acids, but they are often representative of any macromolecular system. For example, the role of water in recognition and specificity illustrated above for protein–DNA interactions has also been observed in the L-arabinose-binding protein interaction with specific sugar molecules (Quiocho *et al.*, 1989). Clearly, water molecules are involved so intimately, and in so many different ways, with the formation of molecular complexes that it is not possible to understand the formation process and the function of the complex without taking into account the role of this universal solvent.

23.4.7. Conclusions and future perspectives

The aim of this chapter was to provide a general overview of the available crystallographic information on the roles of water molecules in their interactions with macromolecules. To achieve this aim, the focus has been on representative examples rather than an exhaustive review of the literature. The classification of water molecules according to location and frequency of occurrence among related proteins, or independently solved crystal structures of a given protein, is a crucial element in determining their functional roles.

It has become clear that water molecules involved in crystal contacts can occur virtually anywhere on the protein surface, as exemplified in the case of T4 lysozyme, and that they have properties similar to the majority of surface water molecules within the context of the crystal. Therefore, it is possible to conclude that for the majority of cases, the crystal contacts between proteins involved in the various studies discussed in this review do not significantly influence the general conclusions drawn.

Water molecules bind to proteins so as to satisfy the hydrogen-bonding potential of protein atoms that are not part of the intramolecular hydrogen-bonding pattern within the native structure. At the primary level, the hydrogen bonding is such as to follow the stereochemical requirements of the individual atom in question, in a manner similar to that occurring for the same atom in small molecules. At the secondary-structure level, these positions tend to provide extensions of α-helices or β-sheets as well as to solvate protein atoms in exposed turns. At the tertiary level, they occur more favourably in grooves or cavities within the protein.

Internal or buried water molecules are found to bridge between domains of a single monomer or bring together different secondary-structure elements within a given domain. They have also been observed in the binding sites of proteins where they fine-tune the shape or electrostatic complementarity towards the substrate or ligand. In general, buried water molecules occur in cavities within the protein, making multiple hydrogen bonds with protein atoms that are likely to be conserved among members of a given family. These water molecules are themselves conserved and must be considered an integral part of the protein architecture. They are often connected to bulk water through water channels leading to the protein surface.

Surface water molecules play important roles in protein dynamics, catalysis, thermodynamics of binding, and in mediation of cooperativity, metal binding, recognition and specificity. Representative examples of water molecules in each of these different roles are discussed in the present review. Some of the surface water molecules are found to be conserved within families of proteins, particularly when they are involved in one of the specific roles mentioned above.

In addition to the commonly observed features in crystal structures of proteins solved to around 2 Å resolution, the crystal structures of crambin and BPTI, both solved to 1 Å resolution, provide examples of the type of information only available at very high resolution. This includes the arrangement of water molecules into pentagonal rings around hydrophobic side chains and the occupancy of water molecules on protein surfaces.

A cohesive picture has emerged of the locations of well ordered water molecules on protein surfaces and their functional roles. Currently, there is a good structural view of the protein atoms as well as of the structure of water molecules associated with the protein. The question now is how this information can be used in predicting *a priori* where water molecules will be involved in important structural or functional roles. While having information on the ordered water molecules associated with protein surfaces represents significant progress toward the ultimate goal of understanding the global thermodynamic and kinetic picture of molecular processes in water, the entire system is still not understood. For instance, how does the bulk water influence the dynamic and thermodynamic processes in which the protein and ordered water molecules are involved? Furthermore, what is the importance of solutes normally found in the biological environment where proteins and other macromolecules exert their function? Knowledge must continue to expand toward an understanding of the complete system. Meanwhile, the present models can go a long way toward successful practical applications in protein engineering and ligand design. In order to improve these models, the information accumulated so far can be combined with empirical results and theoretical models to expand the understanding of the first principles underlying biological processes. Building the bridge between empirical observation and first principles is an iterative process still in its infancy.

Acknowledgements

We wish to thank Martin Karplus and Gregory A. Petsko for years of support and discussions that ultimately contributed to the integrity of this review. During the writing of this review Carla Mattos was supported by the American Cancer Society Postdoctoral Fellowship grant No. PF-4331.