

23. STRUCTURAL ANALYSIS AND CLASSIFICATION

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.