*International Tables for Crystallography* (2006). Vol. F, Section 23.4.1, pp. 623–624.

# 23.4. Solvent structure

By C. Mattos and D. Ringe

## 23.4.1. Introduction

The unique properties of water and its role in nature have preoccupied the minds of scientists and philosophers for centuries. However, only relatively recently have the tools become available to study the specific roles that water molecules play with respect to protein structure and function. When the first crystal structure of a protein was obtained by X-ray diffraction (Kendrew, 1963), the focus was on the arrangement of the amino-acid residues into secondary and tertiary structure. Although the presence of water molecules associated with the protein was noticed, little attention was given to their structure and possible functional role. The structure of the protein itself was a great novelty, and its features were eagerly analysed. For many years, the crucial role of water molecules in maintaining both the structural integrity and the functional viability of proteins was not completely obvious, although in the 1950s Kauzmann argued correctly that water plays an important role in maintaining protein structure (Kauzmann, 1959). In the late 1970s and early 1980s, reviews began to appear focusing on the properties of water relevant to interaction with proteins (Edsall & McKenzie, 1978) and the location and role of water molecules on protein surfaces (Blake *et al.*, 1983; Edsall & McKenzie, 1983). As high-resolution structures became more easily attainable and refinement techniques improved, the importance of water molecules became increasingly apparent, and solvent structure now occupies a front seat in the realm of structural biology. There is a strong sense in the scientific community that water molecules play an integral role in many aspects of protein structure and function, and great effort is now being focused on understanding solvent effects in precise atomic detail.

In principle, water molecules can contribute both enthalpically and entropically to any process in which they are involved. The main contribution of water as solvent in the protein-folding process, for example, is entropic, driving the collapse of hydrophobic residues into the core of the protein. As is currently understood, the general shape of globular proteins is attained by this effect, with the specific structural features guided by the hydrogen bonds that define secondary structure (Hendsch & Tidor, 1994; Hendsch *et al.*, 1996). Although the solvent contribution to the protein-folding process is beyond the scope of this chapter, it is nevertheless deserving of brief mention. It is important to understand protein three-dimensional structures as having evolved in bulk water, a fact which is largely invisible to the current methods in structural biology. The size, geometry, planarity and orientational flexibility of water molecules give them structural and functional importance. All folded globular proteins have evolved with tens to hundreds of binding sites specific for water molecules, a situation contrary to that of larger ligands which usually bind in a single or small number of specific sites found in a given protein or family of proteins. In this respect, water is unique and its ubiquitous appearance is not a direct consequence of its chemical properties alone, but has an evolutionary origin. Proteins evolved in an aqueous milieu, where over time some water molecules were specifically incorporated as integral parts of the protein architecture.

At first glance, the surface of a protein determined by X-ray crystallography appears randomly populated by a layer of water molecules. A careful analysis, however, reveals that the arrangement of water molecules on protein surfaces is not random. In folded proteins, individual water molecules participate in a variety of structural and functional roles, ranging from filling small cavities that are not fully occupied by protein atoms to allowing flexibility, such as in the case of charged surface side chains that can move freely while continuously maintaining hydrogen-bonding partners. Water molecules can fill deep crevices on the protein surface, or they can play a crucial role in the thermodynamics of ligand binding. The mobility as well as the number and strength of hydrogen-bonding partners that are observed for water molecules bound to protein surfaces vary considerably, and it is becoming increasingly apparent that these factors are correlated with functional roles. The atomic coordinates for any protein should not be considered complete without those bound solvent molecules that can be observed, for they are part of the structure.

Bound water molecules have been implicated and studied in the context of substrate specificity and affinity (Quiocho *et al.*, 1989; Herron *et al.*, 1994; Ladbury, 1996), catalysis (Privé *et al.*, 1992; Singer *et al.*, 1993; Komives *et al.*, 1995), mediation of protein–DNA interactions (Clore *et al.*, 1994; Shakked *et al.*, 1994; Morton & Ladbury, 1996), cooperativity (Royer *et al.*, 1996), conformational stability (Bhat *et al.*, 1994), and drug design (Poormina & Dean, 1995*a*,*b*,*c*). One of the challenges now is to translate the structural information observed into a thermodynamic understanding of the water contribution to the various processes. In some cases, an attempt has been made to relate changes in water structure between two forms of a protein (*e.g.* ligated and unligated or native and mutant) to changes in the measured heat capacity (Holdgate *et al.*, 1997) or to measurements of enthalpy and entropy changes by titration calorimetry (Bhat *et al.*, 1994). Thermodynamic solvent isotope effects have also been reported, where the thermodynamics of association of several binding processes were evaluated calorimetrically in light and heavy water (Chervenak & Toone, 1994). In other cases, the three-dimensional structures were directly interpreted in terms of thermodynamic contributions (Quiocho *et al.*, 1989; Morton & Ladbury, 1996). Ultimately, a thorough understanding of the thermodynamics and kinetics underlying solvent structure will lead to powerful predictive methods. Theoreticians, on the one hand, have developed models based on physical principles and use experimental knowledge to assess whether their predictions are correct. Experimentalists, on the other hand, attempt to explain the observed phenomena in terms of the well established physical theories that govern the natural world. Progress is being made on both fronts, but a large gap still remains between the two. A bridge is being built from both sides of the gap and when the two sides meet at a common point, the many pieces of this complicated puzzle will have been deciphered and put in their proper places, so that a global view of molecular processes in water can be obtained from whatever perspective one wishes to take: chemical, physical, or biological.

The present chapter summarizes the empirical information gathered over the last decade or two on the structure of water molecules bound to proteins. The focus will be on structures solved by X-ray crystallography, although complementary techniques of obtaining solvent structure will be discussed briefly and, when appropriate, particular examples will be given. Section 23.4.2 is concerned with the methods by which solvent structure can be observed, Section 23.4.3 summarizes knowledge derived from database analysis of large numbers of proteins, Section 23.4.4 focuses on particular examples of groups of well studied protein structures, Section 23.4.5 discusses the contribution of protein models obtained at very high resolution to the understanding of solvent structure, and Section 23.4.6 contains an analysis of water



Copyright © 2006 International Union of Crystallography

molecules as mediators of complex formation. Finally, Section 23.4.7 presents a conclusion and a perspective regarding the direction in which this information can lead in building a cohesive understanding of the roles played by solvent in the structural integrity and biological function of macromolecules.

### 23.4.2. Determination of water molecules

The most prominent method by which the structure of water molecules on the surface of macromolecules can be observed at the atomic level is X-ray crystallography. The information classically available from this methodology is on bound water molecules, characterized by a high probability density and reduced mobility relative to the bulk solvent, which results in clearly observed electron density. Information on water structure at larger distances from the protein is available in the low-resolution reflections, but difficulty in modelling the solvent in these areas has led to the common practice of discarding the very low resolution data. This chapter focuses on the water molecules for which there is information at high resolution ($>3$ Å), although great progress has been made in recent years in modelling the disordered water structure at the protein–solvent interface, enabling more effective use of the low-resolution data (Badger, 1993; Jiang & Brünger, 1994; Lounnas *et al.*, 1994). Typically, one is interested in studying solvent structure because of the effects that it has on the protein. Lounnas *et al.* (1994) gave a particularly interesting focus on the effect of the protein on the solvent structure surrounding it. Using a combination of molecular-dynamics simulations of explicitly solvated myoglobin and the low resolution X-ray data from myoglobin crystals, they devised a method to describe the effect of the protein on the solvent structure to a distance of 6 Å from the surface. They found that the mobility and probability density of water molecules perpendicular to the protein surface varied considerably depending on the particular composition and three-dimensional structure of the amino-acid residues at the particular area of interest (Lounnas & Pettitt, 1994; Lounnas *et al.*, 1994).

There are a variety of criteria that have been used in placing crystallographic water molecules in electron-density maps. For tightly bound water molecules, with low $B$ factors, the placement involves little or no subjectivity, but the choice of whether or not to include the more disordered waters (or those with low occupancy) can be rather subjective. It generally involves picking the electron-density contour level and $B$-factor cutoffs as well as making a choice of whether to use a simple difference electron-density map ($F_o - F_c$) or to use a higher-order difference electron-density map ($2F_o - F_c$ or $3F_o - 2F_c$). One criterion, applied consistently in placing water molecules on the surface of elastase structures, is the simultaneous presence of electron density at the $3\sigma$ contour level in an $F_o - F_c$ electron-density map and at the $1\sigma$ contour level in a $2F_o - F_c$ electron-density map. After refinement, those waters are kept that have a $B$ factor of 50 Å$^2$ or less. A few exceptions do occur, where there is clear electron density for a water molecule with a $B$ factor of up to 60 Å$^2$. Virtually all of the water molecules placed by these criteria have at least one hydrogen bond to a protein atom or to another water molecule and are mainly part of the first hydration shell on the protein surface.

A method that has provided information on solvent structure complementary to that obtained by X-ray crystallography is based on $D_2O - H_2O$ neutron difference maps (Shpungin & Kossiakoff, 1986). The main advantage of this methodology is in locating partially ordered water molecules whose electron-density peaks may be at the limit of the signal-to-noise ratio allowed for confidently determining positions of water molecules by X-ray diffraction. Scattering of neutrons by $H_2O$ and $D_2O$ is quite different, while scattering from the protein remains the same. Therefore, difference maps based on the two data sets should average to zero where the protein is present and result in peaks only where water molecules are found. Neutron scattering is particularly suited to this because of the threefold greater scattering power of deuterated water molecules relative to light water, providing a larger signal-to-noise ratio in assigning water positions. This method is particularly useful in detecting the second hydration sphere on protein surfaces (Kossiakoff *et al.*, 1992).

NMR spectroscopy can also serve as a complementary technique, providing dynamic information on the lifetime of interaction of a single water molecule on the protein surface. The fact that, with few exceptions, no cross-relaxation peaks are observed at the protein–water interface is an indication that the motion timescale for water molecules in contact with protein is close to that in bulk water at room temperature. The NMR data suggest that water molecules observed in crystal structures have lifetimes of the order of tens of nanoseconds or less (Bryant, 1996). A small number of relatively long-lived structural waters (with residence times in the range $10^{-2}$ to $10^{-8}$ s) can be detected by modern NMR techniques (Otting *et al.*, 1991). Four water molecules have been detected by NMR in bovine pancreatic trypsin inhibitor (BPTI) (Otting & Wuthrich, 1989) and six have been observed in complexes of human dihydrofolate reductase with methotrexate (Meiering & Wagner, 1995). Observation of these waters in the corresponding crystal structures reveals that they are tightly bound waters, with three or four hydrogen bonds to protein atoms, and many are found to bridge between secondary-structure elements or are found to mediate protein–ligand interaction (Meiering & Wagner, 1995). It is important to understand, then, that water molecules near protein surfaces occupy energy minima favoured by hydrogen bonding and ion–dipole effects, which results in water molecules being present in these positions more often than in others. Although when looking at a crystallographic protein structure it is easy to think of a given site as being occupied by a single water molecule, it is in fact only the site that is single, with an enormous number of different individual water molecules sampling it during the time of data collection. This was qualitatively understood from the beginning, but NMR experiments have played a key role in setting quantitative upper boundaries to the residence times of water molecules on the protein surface.

Finally, mention must be made of the computational efforts invested in representing and understanding solvent structure on macromolecular surfaces. The computational work encompasses a variety of methodologies, including integral equation methods (Beglov & Roux, 1997), molecular dynamics (Brooks & Karplus, 1989; Hayward *et al.*, 1993; van Gunsteren *et al.*, 1994), thermodynamic understanding through free-energy simulations (Roux *et al.*, 1996) and statistical-mechanics calculations (Lazaridis *et al.*, 1995). The results of these studies are often complementary to the experimental information already available and provide an important component to the current insight on solvent structure (McDowell & Kossiakoff, 1995) and function (Pomes & Roux, 1996; Oprea *et al.*, 1997). Furthermore, these techniques often provide the only means of obtaining an energetic understanding of some aspects of protein–water interaction.

The question of how the different techniques used to observe the location and properties of water molecules on the surface of proteins relate to and complement one another has been discussed in two short review articles (Levitt & Park, 1993; Karplus & Faerman, 1994). Karplus & Faerman discuss the reliability of each of the methods, illustrating their strengths and weaknesses, while Levitt &



references