Acta Cryst. (1994). D**50**, 663-666

Water Structure Associated with Proteins and its Role in Crystallization

By Michel Frey

Laboratoire de Cristallographie et de Cristallogénèse des Protéines, Institut de Biologie Structurale, 41 Avenue des Martyrs, 38027 Grenoble CEDEX 01, France

(*Received* 10 *September* 1993; *accepted* 26 *November* 1993)

Abstract

X-ray or neutron diffraction studies have shown, at the atomic level, that water molecules occupy well determined sites inside or at the surface of biological macromolecules. These water molecules are constitutive of biomolecules and play a crucial role in their structural and functional properties. Upon crystallization some water molecules are either desolvated or involved in crystal packing. The role of water in determining crystal packing has been experimentally confirmed by several X-ray analyses.

1. Introduction

Water molecules are known to play an essential role in the structural and functional properties of biological macromolecules (proteins and nucleic acids; for reviews see Finney, 1979; Edsall & McKenzie, 1983; Baker & Hubbard; 1984; Savage & Wlodawer, 1986; Saenger, 1987; Teeter, 1991; Rupley & Carreri, 1991; Timasheff, 1993; Westhof, 1993). They are seen as mobile molecules which preferentially occupy sites either in the interior of the macromolecules ('internal waters') or in crevices distributed over the surface. During their residence on these sites, water molecules establish hydrogen bonds with the biomolecule (hence their name 'bound waters') thus contributing to the stability of the three-dimensional structure. Water molecules are also directly involved in the catalytic activity of enzymes such as serine proteases (*e.g.* James, Sielecki, Brayer, Delbaere & Bauer, 1980).

Most of the data, at the atomic level, concerning the water structure associated with biological macromolecules have been obtained by X-ray or neutron diffraction studies of crystals. As we know, protein and nucleic acid crystals are made up of a three-dimensional lattice of molecules which delimits channels of variable dimensions filled with solvent (mostly water molecules; Matthews, 1968). This characteristic feature is used to prepare isomorphous heavy-atom derivatives, which have been, and still are, essential to the determination of new three-dimensional structures. Moreover, the presence of solvent within crystals allows determination of the structure of 'bound water' sites and, in many cases, the study of the biological activity [see, for example, the review by Rossi (1992)] under conditions which are close to those found in solution. The aim of this paper is to review briefly the essential structural properties of water molecules associated with crystallized proteins, their involvement in crystal packing and their possible role in crystallization.

2. Water in and around protein molecules

Water molecules constitute the sole hydrogen-bond partners of *ca* 30% of the main-chain C—O and N—H bonds and *ca* 60% of the polar side-chain groups of a protein molecule. These water molecules have a strong preference to be bound to protein O atoms rather than to N atoms as they use their protons rather than their lone pairs of electrons in hydrogen bonding. This accounts for the greater geometrical flexibility of hydrogen bonding on O atoms, as compared with that for N atoms. O atoms are found in greater numbers at the protein surface and may be involved in more than one hydrogen bond (Baker & Hubbard, 1984). The distribution of water molecules around main-chain polar atoms and polar or apolar side chains is non-random (Walshaw & Goodfellow, 1993, and references therein). NMR studies of bovine pancreatic trypsin inhibitor (BPTI) have shown that the residence times of water molecules range from about 10 ms to 10 ns for those located at the interior of the protein (Otting, Liepinsh & Wüthrich, 1991) and from 200 to 19 ps for those found at the surface (Brunne, Liepinsh, Otting, Wüthrich & van Gunsteren, 1993).

Water molecules help to stabilize the three-dimensional structure of the protein by bridging polar or charged groups, or both, the bonding potentials of which would otherwise have been unsatisfied. Furthermore, surface accessibility and fractal density measurements suggested that water molecules influence protein surface topography (Kuhn *et al.*, 1992). Water molecules also help to stabilize the local structure of catalytic sites by 'spreading' buried electrostatic charges over a large volume, and maintaining a local optimal stereochemistry for catalytic activity (*e.g.* Finney, 1979; James *et al.*, 1980; Blake, Pulford & Artymiuk, 1983). On substrate binding these water molecules are - wholly or partially - easily displaced, thus contributing to the binding energy between the protein and its substrate. Water molecules may be otherwise involved in catalysis

© 1994 International Union of Crystallography
Printed in Great Britain all rights reserved

ISSN 0907-4449 © 1994

directly (*e.g.* Singer, Smalås, Carthy, Mangel & Sweet, 1993*a*).

To summarize, it should be emphasized that water molecules are a constitutive part of the protein, as supported by the fact that many of them have been found to occupy the same sites in crystals of different forms of the same molecule or of homologous proteins (*e.g.* Blake *et al.*, 1983). Along the same lines, recent NMR studies have shown that internal water molecules of BPTI also occupy the same sites in solution as in the crystal structure (Otting *et al.*, 1991).

3. Crystal structure of water molecules

3.1. *Discrete water molecules*

Protein crystals contain solvent volumes which may range from *ca* 22 to over 70% of their total volume (Matthews, 1968). Electron-density maps derived from X-ray studies at high resolution (*i.e.* 1.8 Å or better) show numerous peaks, located at the interior or at the surface of the protein under study, which can be modelled as water molecules. In several cases, peaks located at the molecular surface can be alternatively modelled as water molecules or as side chains. Some surface peaks which are too close to each other cannot be occupied simultaneously in the same crystal cell, as shown by their relatively weaker electron density (Smith, Hendrickson, Honzatko & Sheriff, 1986). These so-called multisite configurations reflect, again, the highly dynamical properties of some surface side chains and water molecules associated with proteins.

In crystals, where protein molecules are closely packed, over 90% of the water molecules have been found to occupy discrete sites (*e.g.* Teeter, 1984; Adman, Sieker & Jensen, 1991; Dauter, Sieker & Wilson, 1992; Madhusudan, Kodandapani & Vijayan, 1993). However, in general, ordering of water sites beyond 4-5 Å (the so-called 'first hydration shell') has only been observed when water molecules are able to form a network with this 'first hydration shell' (*e.g.* Jensen & Watenpaugh, 1986). It is of interest to note that non-random arrangements of water sites extending over several layers from the molecular surface have been proposed from an X-ray study of cubic insulin crystals which contain 65% solvent by volume (Badger & Caspar, 1991). These latter results have been questioned in view of the possible impact of experimental errors and model uncertainties on the electron density of the solvent region (Jensen, 1991). However, a recent study of the same insulin crystals equilibrated with glucose (at concentrations of 1-3 *M*) indicates that the glucose distribution in the bulk solvent is also non-random (Kapulsky, Badger & Caspar, 1993). This again suggests that water molecule sites may be somewhat ordered beyond the vicinity of the protein molecule, at least in a crystalline environment.

3.2. *Water molecules and crystal packing*

An analysis of the molecular packing of several protein crystals following the periodic bond-chain method (PBC; Hartman & Perdok, 1955) has been carried out. This showed that the crystallizing particle consists of the protein and 'bound waters' and that, consequently, upon crystallization, some water molecules are either desolvated or involved in intermolecular contacts (Frey, Genovesio-Taverne & Fontecilla-Camps, 1988). Moreover, it has been observed that water-mediated bridges and direct polar interactions are involved in similar numbers of crystal-packing molecular interactions (Frey *et al.*, 1988; Crosio, Janin & Jullien, 1992). Thus, the nature and the contribution of the forces which are involved in crystal packing appear, not unexpectedly, similar to those which contribute to the stability of protein-protein complexes (Chothia, 1975; Janin, Miller & Chothia, 1988; Janin & Chothia, 1990).

Taking into account these crystal-packing forces, the attachment energy of the crystallizing particle onto the different faces of a rubredoxin crystal - as predicted by the PBC theory - has been calculated. It has been shown that there is a good inverse correlation between the calculated attachment energy and the growth of most of the faces of the protein crystal (Frey, Genovesio-Taverne & Fontecilla-Camps, 1991).

4. Water molecules and crystallization

4.1. *Influence on the solvent structure*

Because of the highly dynamic character of solvent (mostly water molecules), there are continuous exchanges between the 'first hydration shell', the channels of bulk solvent and the mother liquor. Consequently, the structure and the chemical contents of solvent sites might depend highly on physico-chemical properties of the crystallization medium such as pH, salt concentration, precipitant *etc.*

Measurements of crystal density, combined with crystallographic data analysis may provide, prior to the X-ray analysis, correct information on the respective fractions of 'bound' and bulk solvent within crystals, and, in some cases, on the precise composition of 'bound' solvent (see, for example, Scanlon & Eisenberg, 1975; Hagler & Moult, 1978). Selenate-exchanged (SO_4^{2-}, SeO_4^{2-}) mother liquor is also an efficient way of identifying the sulfate ions although, to our knowledge, this method has only been reported for α-chymotrypsin crystals (Tulinsky & Wright, 1973; Blevins & Tulinsky, 1985). In all cases analysis of crystal contents could be of major interest and even essential (Blevins & Tulinsky, 1985) for a reliable chemical assignment of solvent sites observed in the electron-density maps.

The most suitable crystallization medium allowing examination of the interactions between protein and water molecules as they occur *in vivo* should involve

physiological pH, salt content and concentrations, although comparison of the structures of bacteriophage T4 lysozyme, determined with crystals grown at different salt concentrations, suggests that the presence of high salt only has a limited effect on the structure of 'bound waters' (Bell *et al.*, 1991).

4.2. *Influence on crystal packing*

The involvement of water molecules in crystal packing has been exemplified experimentally by X-ray analyses of several crystal forms of lysozyme, kept in various conditions of humidity. Thus, crystals of the tetragonal and monoclinic forms, grown under the usual conditions (Alderton, Ward & Fevold, 1945; Steinrauf, 1959), were kept in sealed thin-walled capillaries with a few drops of a saturated solution of potassium chromate (K_2CrO_4) for 15-20 h. For both forms: (i) the protein molecules were found to be displaced from their positions in the native crystals along with the most strongly 'bound' water molecules and (ii) the most significant changes occurred in the inhibitor-binding region (Kodandapani, Cheravakkattumadhom & Vijayan, 1990; Madhusudan *et al.*, 1993). It is of interest to note that the cell volume of the 'low-humidity' monoclinic form was more reduced than those of the tetragonal form. This resulted in: (i) an exceptionally close packing of the protein molecules within the monoclinic crystal (22% solvent by volume); (ii) a high-resolution diffraction pattern (beyond 1.75 Å resolution) and (iii) the location of 148 (out of 175 possible) water molecules (Madhusudan *et al.*, 1993).

Similar experiments have been performed on triclinic lysozyme crystals, cross-linked with glutaraldehyde, with $CaCl_2$ solutions (98-32% humidity range) or silica gel (~0.01% humidity). These experiments have shown a good correlation between the loss of water and the cell-volume shrinkage (Kachalova, personal communication). A preliminary comparison of the crystal structures of 'dry' (~0.01% humidity) and 'wet' (98% humidity) forms suggested a compression of the molecule upon drying of about 4-6% which might explain the loss of activity at this hydration level. The determination of the crystal structures of several forms obtained at intermediate humidity levels (70, 39, 30%) is underway (Kachalova *et al.*, 1991, and personal communication).

Along the same lines Pelletier & Kraut (1992) have observed that crystals of a complex between cytochrome *c* and cytochrome *c* peroxidase diffracted better if they were standing in sealed capillary tubes for an extended period of time (3 weeks to 4 months). Since this involved a decrease of the unit-cell volume, they concluded that their crystals had partially dried out (Pelletier & Kraut, 1992). A similar phenomenon has also been observed as a result of the slow diffusion of cryo-protective agents into the crystals (Schick & Jurnak, 1994).

5. Perspectives

The attainment of reliable X-ray models of proteins, including the structure of their associated water molecules, appears essential, more than ever, to the understanding of their structural, catalytic, dynamical and thermodynamical properties (Brändén & Jones, 1990). This could be crucial for drawing reliable conclusions concerning the actual influence of a site-directed mutation on the biological function (*e.g.* Loll & Lattman, 1990) or for studying the electrostatic forces and the dynamic properties which play a major role in the interactions between proteins and substrates or between proteins and any of their molecular functional partners.

Careful X-ray structure determinations require above all a better knowledge of the physico-chemical properties of the solvent in and around the crystals under study: pH, ion content, humidity *etc.* Moreover, control of the solvent content either through dehydration or through diffusion of cryo-protective agents seems most promising to produce more close packed - or at least better ordered - crystals and thus to extend the resolution of X-ray diffraction patterns (Salunke, Veerapandian, Kodandapani & Vijayan, 1985; Kachalova *et al.*, 1991; Schick & Jurnak, 1994)

Collection of X-ray data at wavelengths around 1 Å and at a low temperature should also allow the major causes of experimental errors, namely absorption and dynamic disorder, to be limited. Inclusion of both high-resolution and low-resolution X-ray data should lead to a better estimation of the scale factor and to more reliable electron-density maps in the critical zone which extends from 'bound waters' to bulk solvent. A systematic search and stereochemical analysis of the solvent peaks within electron-density maps would be also most helpful (*e.g.* Berghuis *et al.*, 1993).

Of particular interest are also time-resolved Laue X-ray studies which might provide an insight into the dynamics of water molecules in the different steps of a catalytic reaction. In this respect the recent study of the role of a water molecule in the hydrolysis of a transient form of bovine trypsin is most promising, although the conclusions of this experiment are still highly debated (Singer *et al.*, 1993*a*,*b*; Perona, Craik & Fletterick, 1993).

I thank my colleagues J. Fontecilla, D. Housset, F. Vellieux and A. Volbeda for critical reading of this manuscript. I greatly acknowledge Professor Lyle Jensen and his colleagues for the stimulating days spent at Seattle. I am very much indebted to Professor Eric Westhof in Strasbourg for his encouragement and for many helpful discussions.

References

Adman, E. T., Sieker, L. C. & Jensen, L. H. (1991). *J. Mol. Biol.* **217**, 337-352.

ALDERTON, G., WARD, W. H. & FEVOLD, H. L. (1945). *J. Biol. Chem.* **157**, 43-58.

BADGER, J. & CASPAR, D. L. D. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 622-626.

BAKER, E. N. & HUBBARD, R. E. (1984). *Prog. Biophys. Mol. Biol.* **44**, 97-179.

BELL, J. A., WILSON, K. P., ZHANG, X.-J., FABER, H. R., NICHOLSON, H. & MATTHEWS, B. W. (1991). *Proteins Struct. Funct. Genet.* **10**, 10-21.

BERGHUIS, A., TONG, H., CHEN, J., LUO, Y., GUSS, M., FREEMAN, H. & BRAYER, G. (1993). Am. Crystallogr. Assoc. Annu. Meet., Albuquerque, Abstract PB16.

BLAKE, C. C. F., PULFORD, W. C. A. & ARTYMIUK, P. J. (1983). *J. Mol. Biol.* **167**, 693-723.

BLEVINS, R. A. & TULINSKY, A. (1985). *J. Biol. Chem.* **260**, 8865-8872.

BRÄNDÉN, C. I. & JONES, T. A. (1990). *Nature (London)*, **343**, 687-689.

BRUNNE, R. M., LIEPINSH, E., OTTING, G., WÜTHRICH, K. & VAN GUNSTEREN, W. F. (1993). *J. Mol. Biol.* **231**, 1040-1048.

CHOTHIA, C. (1975). *Nature (London)*, **254**, 304-308.

CROSIO, M.-P., JANIN, J. & JULLIEN, M. (1992). *J. Mol. Biol.* **228**, 243-251.

DAUTER, Z., SIEKER, L. C. & WILSON, K. S. (1992). *Acta Cryst.* **B48**, 42-59.

EDSALL, J. T. & MCKENZIE, H. A. (1983). *Adv. Biophys.* **16**, 53-183.

FINNEY, J. (1979). *Water. A Comprehensive Treatise*, pp. 47-122. New York, London: Plenum Press.

FREY, M., GENOVESIO-TAVERNE, J. C. & FONTECILLA-CAMPS, J. C. (1988). *J. Cryst. Growth*, **90**, 245-258.

FREY, M., GENOVESIO-TAVERNE, J. C. & FONTECILLA-CAMPS, J. C. (1991). *J. Phys. D*, **24**, 105-110.

HAGLER, A. T. & MOULT, J. (1978). *Nature (London)*, **272**, 222-226.

HARTMAN, P. & PERDOK, W. G. (1955). *Acta Cryst.* **8**, 4-52.

JAMES, M. N. G., SIELECKI, A. R., BRAYER, G. D., DELBAERE, L. T. J. & BAUER, C. A. (1980). *J. Mol. Biol.* **144**, 43-88.

JANIN, J. & CHOTHIA, C. (1990). *J. Biol. Chem.* **265**, 16027-16030.

JANIN, J., MILLER, S. & CHOTHIA, C. (1988). *J. Mol. Biol.* **204**, 155-164.

JENSEN, L. H. (1991). Am. Crystallogr. Assoc. Annu. Meet., Toledo, Abstract G06.

JENSEN, L. H. & WATENPAUGH, K. D. (1986). *Trans. Am. Crystallogr. Assoc.* **22**, 89-96.

KACHALOVA, G. S., MOROZOV, V. N., MOROZOVA, T. YA., MYACHIN, E. T., VAGIN, A. A., STROKOPYTOV, B. V. & NEKRASOV, YU. V. (1991). *FEBS Lett.* **284**, 91-94.

KAPULSKY, A., BADGER, J. & CASPAR, D. L. D. (1993). Am. Crystallogr Assoc. Annu. Meet., Albuquerque, Abstract PI12.

KODANDAPANI, R., CHERAVAKKATTUMADHOM, G. S. & VIJAYAN, M. (1990). *J. Biol. Chem.* **265**, 16126-16131.

KUHN, L. A., SIANI, M. A., PIQUE, M. E., FISHER, C. L., GETZOFF, E. & TAINER, J. A. (1992). *J. Mol. Biol.* **228**, 13-22.

LOLL, P. J. & LATTMAN, E. E. (1990). *Biochemistry*, **29**, 6866-6873.

MADHUSUDAN, KODANDAPANI, R. & VIJAYAN, M. (1993). *Acta Cryst.* **D49**, 234-245.

MATTHEWS, B. W. (1968). *J. Mol. Biol.* **33**, 491-497.

OTTING, G., LIEPINSH, E. & WÜTHRICH, K. (1991). *Science*, **254**, 974-980.

PELLETIER, H. & KRAUT, J. (1992). *Science*, **258**, 1748-1755.

PERONA, J. J., CRAIK, C. S. & FLETTERICK, R. J. (1993). *Science*, **261**, 620-621.

ROSSI, G. L. (1992). *Curr. Opin Struct. Biol.* **2**, 816-820.

RUPLEY, J. A. & CARRERI, G. (1991). *Adv. Protein Chem.* **41**, 38-173.

SAENGER, W. (1987). *Annu. Rev. Biophys. Biophys. Chem.* **16**, 93-114.

SALUNKE, D. M., VEERAPANDIAN, B., KODANDAPANI, R. & VIJAYAN, M. (1985). *Acta Cryst.* **B41**, 431-436.

SAVAGE, H. & WLODAWER, A. (1986). *Methods Enzymol.* **127**, 162-183.

SCANLON, W. J. & EISENBERG, D. (1975). *J. Mol. Biol.* **98**, 485-502.

SCHICK, B. & JURNAK, F. (1994). *Acta Cryst.* **D50**, 563-568.

SINGER, P. T., SMALÅS, A., CARTHY, R. P., MANGEL, W. F. & SWEET, R. (1993*a*). *Science*, **259**, 669–673.

SINGER, P. T., SMALÅS, A., CARTHY, R. P., MANGEL, W. F. & SWEET, R. (1993*b*). *Science*, **261**, 621–622.

SMITH, J. L., HENDRICKSON, W. A., HONZATKO, R. B. & SHERIFF, S. (1986). *Biochemistry*, **25**, 5018-5027.

STEINRAUF, L. K. (1959). *Acta Cryst.* **12**, 77-79.

TEETER, M. M. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 6014-6018.

TEETER, M. M. (1991). *Annu. Rev. Biophys. Biophys. Chem.* **20**, 577-600.

TIMASHEFF, S. N. (1993). *Annu. Rev. Biophys. Struct.* **22**, 67-97.

TULINSKY, A. & WRIGHT, L. H. (1973). *J. Mol. Biol.* **81**, 47-56.

WALSHAW, J. & GOODFELLOW, J. M. (1993). *J. Mol. Biol.* **231**, 392-414.

WESTHOF, E. (1993). *Water and Biological Macromolecules.* London: MacMillan.