Acta Cryst. (1993). D**49**, 234–245

Protein Hydration and Water Structure: X-ray Analysis of a Closely Packed Protein Crystal with Very Low Solvent Content

By Madhusudan, R. Kodandapani and M. Vijayan

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, *India*

(*Received* 9 *April* 1992; *accepted* 16 *June* 1992)

Abstract

Low-humidity monoclinic lysozyme, resulting from a water-mediated transformation, has one of the lowest solvent contents (22% by volume) observed in a protein crystal. Its structure has been solved by the molecular replacement method and refined to an R value of 0.175 for 7684 observed reflections in the 10–1.75 Å resolution shell. 90% of the solvent in the well ordered crystals could be located. Favourable sites of hydration on the protein surface include side chains with multiple hydrogen-bonding centres, and regions between short hydrophilic side chains and the main-chain CO or NH groups of the same or nearby residues. Major secondary structural features are not disrupted by hydration. However, the free CO groups at the C terminii and, to a lesser extent, the NH groups at the N terminii of helices provide favourable sites for water interactions, as do reverse turns and regions which connect β-structure and helices. The hydration shell consists of discontinuous networks of water molecules, the maximum number of molecules in a network being ten. The substrate-binding cleft is heavily hydrated, as is the main loop region which is stabilized by water interactions. The protein molecules are close packed in the crystals with a molecular coordination number of 14. Arginyl residues are extensively involved in intermolecular hydrogen bonds and water bridges. The water molecules in the crystal are organized into discrete clusters. A distinctive feature of the clusters is the frequent occurrence of three-membered rings. The protein molecules undergo substantial rearrangement during the transformation from the native to the low-humidity form. The main-chain conformations in the two forms are nearly the same, but differences exist in the side-chain conformation. The differences are particularly pronounced in relation to Trp 62 and Trp 63. The shift in Trp 62 is especially interesting as it is also known to move during inhibitor binding.

Introduction

The importance of water in the structure and action of proteins as indeed in those of other biomolecules has been very well recognized. A substantial part of the current understanding of the geometrical features of water around protein molecules is derived from high-resolution X-ray crystallographic studies of protein crystals (Finney, 1979; Sakabe, Sakabe & Sasaki, 1980; Blake, Pulford & Artymiuk, 1983; Blundell, Barlow, Borakakoti & Thornton, 1983; Teeter, 1984; Baker & Hubbard, 1984; Wlodawer, Deisenhofer & Huber, 1987; Baker *et al.*, 1988; Thanki, Thornton & Goodfellow, 1988, 1990; Kodandapani, Suresh & Vijayan, 1990). These studies suggest that each protein is associated with a characteristic hydration shell a substantial part of which is invariant with respect to environmental effects and species variation (Blake, Pulford & Artymiuk, 1983; Wlodawer, Deisenhofer & Huber, 1987; Kodandapani, Suresh & Vijayan, 1990; Madhusudan & Vijayan, 1991; Malin, Zielenkiewicz & Saenger, 1991). The results of these studies have also been relevant to water structure in general (Teeter, 1984).

Low-humidity monoclinic hen egg-white (HEW) lysozyme, resulting from a water-mediated transformation (Salunke, Veerapandian & Vijayan, 1984; Salunke, Veerapandian, Kodandapani & Vijayan, 1985) of the native crystals when the environmental humidity is systematically reduced, has perhaps the lowest solvent content (22% by volume) in any protein crystal. The transformation, which involves substantial structural rearrangement, results in great improvement in the diffraction pattern, and the low-humidity crystals, in which the protein molecules appear to be close packed (Kitaigorodsky, 1955, 1973), diffract beyond 1.75 Å. Consequently, about 90% of the solvent (mainly water) molecules, which are well ordered, could be readily located in the crystals. The structure thus provides an ideal system for investigating protein hydration and the water structure associated with proteins. Some of the features observed in the crystal are also of general validity. Furthermore, the low-humidity monoclinic form has been analysed at a resolution higher than those at which the other crystal forms of lysozyme have been reported so far (Imoto, Johnson, North, Phillips & Rupley, 1972; Kundrot & Richards, 1987;

0907-4449/93/020234-12$06.00 © 1993 International Union of Crystallography

Kodandapani, Suresh & Vijayan, 1990; Ramanadham, Sieker & Jensen, 1990; Kodandapani & Vijayan, 1991) and the analysis reported here provides the most accurate geometrical description to date of the enzyme molecule.

Experimental

HEW lysozyme was commercially obtained from Sigma Chemical Company, USA. The monoclinic crystals of the enzyme were grown using the well known method described in the literature (Steinrauf, 1959). The relative humidity around the crystals was maintained at 88% by placing a few drops of saturated K_2CrO_4 solution about 1 cm away from the crystal in a sealed thin-walled capillary (Rockland, 1960; Salunke, Veerapandian, Kodandapani & Vijayan, 1985). The transformation to the low-humidity form was usually complete within 15–20 h and this was ascertained by examination of precession X-ray photographs. The transformed crystals were then used for data collection.

Data collection and processing

Screenless oscillation photography (Arndt & Wonacott, 1977) was used for recording photographic data. The crystals were mounted along the b^* axis with a crystal-to-film distance of 45.4 mm. A sealed-tube copper-target generator operated at 40 kV and 25 mA was used as the X-ray source. Indu X-ray films manufactured by Hindustan Photo Films Co. Ltd, Ooctamund, India, were used for data collection. The oscillation range was varied between 4 and 3.5°. A 20% overlap between adjacent ranges was maintained throughout data collection to ensure that the entire data set was available as fully recorded reflections. A total of 18 crystals were used to record 62 film packs covering a 180° rotation range. Typically an exposure time of 20 h was used for each oscillation photograph. The films were digitized using a Joyce Loebl Scandig film scanner controlled by a PDP 11/44 system at intervals of 100 μm. The data were processed and scaled using the program originally written by Rossmann (1979, 1985), subsequently modified by Dr I. Rayment and adapted for the PDP 11/44 system by Dr M. R. N. Murthy. Partially recorded reflections were used for the post refinement (Rossmann, Leslie, Abdel-Meguid & Tsukihara, 1979) of setting angles, mosaicity and cell parameters. However, only fully recorded reflections were used in structure determination and refinement. The refined cell parameters, the solvent content (Matthews, 1968) and the statistical details pertaining to the photographic data are given in Table 1.

Table 1. *Crystal data and details of photographic data collection*

Space group	$P2_1$
a	26.9 Å
b	59.0 Å
c	31.3 Å
β	111.9°
Solvent volume	22%
Total number of reflections measured	43889
Total number of unique reflections (including overloads)	8508
Total number of unique reflections with $I > 2\sigma(I)$	8288
Merging R for all reflections*	0.065
Merging R for reflections with $I > 2\sigma(I)$	0.061

$R = \sum_h \sum_i |\langle I_h \rangle - S_i I_{hi}| / \sum_h \sum_i I_{hi}$ where $\langle I_h \rangle$ is the average intensity of reflection h. I_{hi} is the intensity of the reflection in the ith measurement, and S_i is the scale factor.

The 1.75 Å data set contained as many as 347 reflections for which only overloads were available. Therefore, a second data set, up to a nominal resolution of 2 Å, was collected with low exposure times (85 min for each range). All the data were collected from a single crystal on 45 film packs, each pack corresponding to an oscillation range of 4.5°. The procedures used for data processing were the same as those used with the main 1.75 Å data set. The new low-exposure data set was free of overloads. Of the 347 overloaded reflections in the main data set, only 345 were found in the second data set. However, 17 additional reflections, not present in the main data set, were available in the new data set. These 362 new reflections were incorporated into the main data set after scaling their intensities using reflections common to both data sets. The Wilson plot, computed using the final merged data set, yielded a B value of 10 Å^2.

Structure solution

The native monoclinic lysozyme crystals contain two protein molecules, referred to as molecules A and B, in the asymmetric unit (Rao, Hogle & Sundaralingam, 1983). During the transformation to the low-humidity form they become equivalent which necessitates an appropriate unit-cell transformation (Salunke, Veerapandian, Kodandapani & Vijayan, 1985). The choice of the unit cell in the low-humidity form was such that the two molecules in it, which are related to each other by a crystallographic 2_1 screw axis, were located approximately at the positions corresponding to molecule A and a translational equivalent of molecule B in the native unit cell. The known coordinates of the two molecules in the native structure, after appropriate transformations, were successively used in attempts to refine the structure of the low-humidity form by employing the restrained least-squares method. Both the models failed to refine properly. This failure appeared to reflect the substantial rearrangement of molecules in

the crystal during the transformation from the native to the low-humidity form.

The structure was solved using the molecular replacement method (Rossmann, 1972) with the lysozyme molecule in low-humidity tetragonal lysozyme (Kodandapani, Suresh & Vijayan, 1990) as the search model. Crowther's rotation function (Crowther, 1972), calculated using the *MERLOT* package (Fitzgerald, 1988) adapted for the PDP 11/44 system by Dr K. Suguna, contained a single significant peak. An *R*-factor search (Nixon & North, 1976) for the orientation of the molecule corresponding to this peak, using a locally developed program (Mande, 1991), readily yielded the solution to the translation problem.

Refinement

The model obtained using the molecular replacement method was used as the starting model in the refinement of the structure by employing the Hendrickson–Konnert restrained least-squares program (Konnert, 1976; Hendrickson & Konnert, 1980*a*,*b*) as modified by E. Dodson to incorporate Agarwal's fast Fourier transform method (Agarwal, 1978). All main-chain atoms were restrained for peptide geometry. Temperature factors and contact distances involving solvent molecules were not restrained. The contact distances were, however, prevented from going below 2.3 Å. Occupancy factors were kept constant at unity for all atoms. Initial refinement was carried out using reflections in the 8–3 Å resolution shell. The resolution shell was expanded in stages to 8–2.2, 8–2 and 10–1.75 Å. The correctness of the model was checked, parts of it were rebuilt and solvent molecules located using $F_o - F_c$, $2F_o - F_c$ and 'omit' type maps (Vijayan, 1980; Bhat & Cohen, 1984) calculated at different stages. The electron density for the side chain of Gln 121 was extremely poor, and only the C^β atom of the side chain was included in the refinement. A total of 172 refinement cycles were carried out. The final *R* factor was 0.175 for 7684 reflections with $F > 4\sigma(F)$ in the 10–1.75 Å resolution shell.* The weighting parameters for stereochemical constraints used in the final cycles, and the r.m.s. deviations from ideality on completion of refinement, are listed in Table 2.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1LMA, R1LMASF), and are available in machine-readable form from the Protein Data Bank at Brookhaven. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37067 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England. At the request of the authors, the atomic coordinates and structure factors will remain privileged until 30 December 1993.

Table 2. *Weighting parameters for stereochemical terms and deviations from standard values*

	Number	σ	R.m.s. deviation
Bonding distance (Å)			
Bond distance	1021	0.02	0.010
Angle distance	1384	0.04	0.034
Planar distance	370	0.06	0.040
Planar groups			
Deviation from plane (Å)	179	0.02	0.009
Chiral volume (Å^3)	144	0.12	0.091
Non-bonded contacts (Å)			
Single torsion		0.5	0.180
Multiple torsion		0.5	0.286
Possible hydrogen bond		0.5	0.194
Torsion angles (°)			
Planar angle	139	5	1.8
Staggered angle	175	20	17.3
Orthonormal	13	20	23.3

Results

Accuracy and definition of the structure

The final refined model contains 997 of a total of 1001 protein atoms. Assuming the solvent density and partial specific volume of the protein to be 1 g cm^{-3} and 0.74 cm^3 g^{-1} respectively, the number of water molecules in the crystal asymmetric unit works out to be 175 if the solvent is entirely water. In the X-ray analysis 148 water molecules and two nitrate ions could be located. The Luzzati plot (Luzzati, 1952) indicates the average coordinate error to be between 0.1 and 0.15 Å. The final electron-density map is indeed of high quality as can be seen from Fig. 1. The *B* values also often provide a reliable measure of the definition of the structure. In the present structure, the average *B* values for all protein atoms and all solvent atoms are 10.8 and 29.3 Å^2 respectively. The corresponding values for the main-chain and side-chain atoms are 8.1 and 13.8 Å^2 respectively. These values are low in comparison to those found in most other structures and therefore indicate that the overall definition of the structure is good.

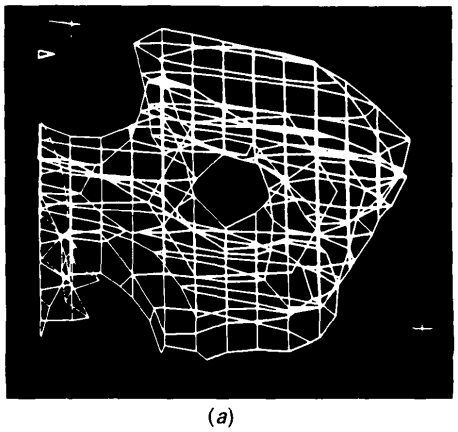
Delineation of hydrogen bonds and solvent interactions

The criteria suggested by Baker & Hubbard (1984) as modified by Kodandapani, Suresh & Vijayan (1990) were used in delineating hydrogen bonds. The distance component of these criteria involves a cut-off value of 2.6 Å for H···O distances in cases were hydrogen atoms could be fixed from geometrical considerations and a cut-off value of 3.6 Å for O(N)···O distances in cases where the hydrogen atoms could not be fixed. The angle criteria were liberal and ensured that no genuine hydrogen bond is omitted from consideration. Only the distance criter-

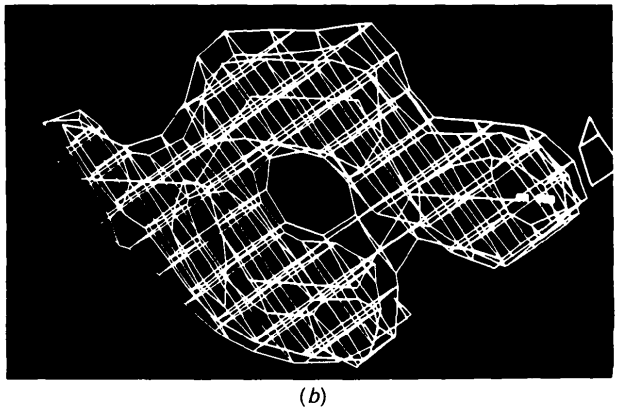
ion (< 3.6 Å) was used in identifying interactions of water molecules with oxygen and nitrogen atoms in protein molecules or among themselves.

Overall features

The Ramachandran plot (Ramachandran, Ramakrishnan & Sasisekharan, 1963; Ramachandran & Sasisekharan, 1968) for the refined structure exhibits the expected behaviour (Richardson & Richardson, 1989; Weaver, Tronrud, Nicholson & Matthews, 1990). The distribution of side-chain conformation angles is in agreement with that reported for other protein structures (Janin, Wodak, Levitt & Maigret, 1978; Bhat, Sasisekharan & Vijayan, 1979; McGregor, Islam & Sternberg, 1987). The main-chain–main-chain hydrogen bonds, which together with the φ and ψ angles completely define the tertiary fold of the protein molecule, are illustrated in Fig. 2. Of the 77 such bonds, 46 are in helices, nine in β-structures and 14 in reverse turns. Thus, well defined secondary structural elements account for more than 85% of the main-chain–main-chain hydrogen bonds. It is also interesting to note that 117 of a total of 129 residues belong to helices, β-structure or turns. Of the remaining 12, nine belong to the main loop region (residues 65 to 73) while three belong to a minor loop (residues 85 to 87). The details of the tertiary structure, including side-chain conformation and hydrogen-bonded interactions, are very similar to those reported earlier for lysozyme (Imoto, Johnson, North, Phillips & Rupley, 1972; Kodandapani, Suresh & Vijayan, 1990; Kodandapani & Vijayan, 1991). The present structure, however, provides a more precise description of the geometry because of the higher resolution X-ray analysis and improved order in the crystals. In addition, in the low-humidity monoclinic form the main loop region of the molecule is well ordered unlike in the tetragonal forms (Imoto, Johnson, North, Phillips & Rupley, 1972; Kundrot & Richards, 1987; Kodandapani, Suresh & Vijayan, 1990). Also, Trp 62 is better defined than in other forms of lysozyme.



(*a*)



(*b*)

Fig. 1. Final electron density corresponding to the side chains of (*a*) Phe 38 and (*b*) Tyr 53.

The hydration shell

15 of the 148 water molecules located in the asymmetric unit interact only with other water molecules while one does not have any interaction at all. Of the remaining 132 water molecules, 94 interact with one protein molecule, 36 are involved in bridging two protein molecules each and two interact with three protein molecules each. Consequently, each protein molecule is surrounded by, and interacts with, 172 water sites. These water sites may be taken to constitute the hydration shell of the protein molecule. Of the 274 protein–water interactions, 88 involve main-chain CO groups, 48 main-chain NH groups, 64 side-chain oxygen atoms and 74 side-chain nitrogen atoms. As in the case of other proteins (Baker & Hubbard, 1984), water interactions with protein oxygen atoms are more numerous than those with protein nitrogen atoms, but the difference is less pronounced in low-humidity monoclinic lysozyme.

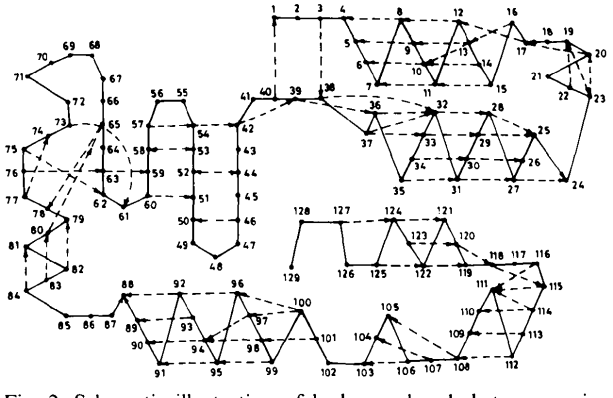


Fig. 2. Schematic illustration of hydrogen bonds between main-chain atoms.

Water bridges and favourable sites of hydration

Of the 172 water sites in the hydration shell, about two-thirds (115) have only one interaction each with the protein molecule while about a sixth (31) have two interactions each. The number of water molecules with three, four, five and six interactions each with the protein molecule are 14, seven, four and one respectively. The water molecules in the hydration shell with two or more interactions each give rise to 170 water bridges altogether. Of these a majority (106) involve protein atoms separated by five residues or less; 23 of them involve atoms of the same residue. These 'local' bridges presumably contribute to the stability of the local structure. The water bridges also lead to the identification of what may be described as favourable sites of hydration. For example, as has been noted earlier (Kodandapani, Suresh & Vijayan, 1990), side chains with multiple hydrogen-bonding centres often provide such sites. Altogether there are 12 such sites in the structure made up of guanidyl, amide or carboxylate groups, and all of them involve specific interactions (Salunke & Vijayan, 1981). The region between short hydrophilic side chains and the main-chain CO or NH group of the same or nearby residue also constitutes a favourable site of hydration. A few examples of such sites are illustrated in Fig. 3.

Water bridges connecting widely separated atoms along the polypeptide chain could be important for stabilization of the tertiary structure. These 'non-local' water bridges contain a lower proportion (25%) of bridges involving exclusively main-chain atoms than local bridges do (34%). Again, as in the case of local ones, short hydrophilic side chains are extensively involved in the bridges. The hydration shell contains 18 well defined ($B < 20$ Å^2) water molecules with three or more interactions with the protein. Most of them connect regions widely separated along the polypeptide chain. Many also remain invariant with respect to environmental changes and species variation (Madhusudan & Vijayan, 1991).

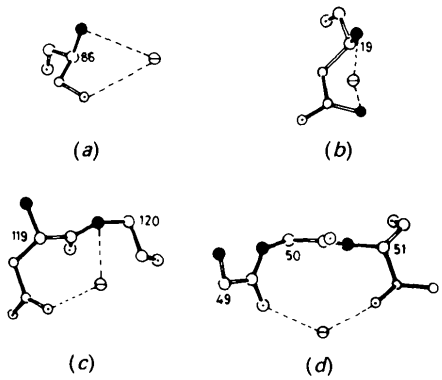


Fig. 3. Examples of favourable sites of hydration involving a short hydrophilic and (*a*) an NH group of the same residue, (*b*) a CO group of the same residue, (*c*) an NH group of the adjacent residue and (*d*) a CO group of the next-nearest residue. In this and the subsequent figures, carbon, nitrogen, oxygen atoms and water molecules are represented by open, filled, dotted and dashed circles; only relevant side chains are shown and residue numbers are given, where necessary, close to the appropriate C^{α} positions.

Secondary structure and hydration

In the present structure, the main-chain hydrogen bonds that stabilize the α-helices are mostly unaffected by hydration, although water molecules interact with a few CO and NH groups in the helices. Four water molecules are involved in such interactions with the 5 to 15 helix. Two of them interact with the free CO groups of residues 13 and 15 at the C terminus of the helix while one interacts with the free NH group of residue 5 at the N terminus. The remaining water molecule interacts with the CO group of residue 7. However, this interaction does not disrupt the 11–7 hydrogen bond. As will be shown later, the main-chain NH groups of residues 5, 6 and 7 are hydrogen bonded to a nitrate ion. The main-chain oxygen and nitrogen atoms in the 25 to 36 and 88 to 101 helices are hydrated to a somewhat larger extent, with seven water molecules involved in the interactions in each case. These interactions are illustrated in Fig. 4. In the 25 to 36 helix, the interactions are confined to helix terminii, six with the free CO groups and one with a free NH group. In the 88 to 101 helix, three water molecules interact with C-terminal CO groups and two with an N-terminal NH group. In the central region of the helix 90 O and 93 O are hydrated. The hydration of 93 O might account for the absence of a helical hydrogen bond at this position. Six water molecules are attached to the 109 to 115 helix. Four of them interact with C-terminal CO groups of residues 113, 114 (two water molecules) and 115, while the remaining two interact with the NH groups of the N-terminal residues 109 and 110. The hydration pattern outlined above clearly demonstrates that the free CO groups at the C terminii and, to a lesser extent, the NH groups at the N terminii of α-helices provide favourable sites for water interactions.

The two 3_{10} helical stretches (80 to 84 and 120 to 124) in the molecule are shorter and also proportionately more heavily hydrated than the α-helices. A water molecule each interacts with 81 O, 82 O, 83 O, 84 O, 80 N and 81 N accounting for six water molecules interacting with the main-chain atoms of the five residues in the 80 to 84 3_{10} helix. In the 120 to 124 3_{10} helix, a water molecule each interacts with 120 N, 121 N, 121 O and 124 O. Though less pronounced than in α-helices, the propensity of N-terminal NH groups and C-terminal CO groups to

interact with water molecules is discernible in 3_{10} helices also.

The hydration of the main-chain atoms in the major β-structure region involves 20 water molecules, as illustrated in Fig. 5. None of the main-chain–main-chain hydrogen bonds which stabilize the β-structure appears to have been disrupted by water molecules. The main-chain atoms hydrated are largely, but not entirely, those not involved in such hydrogen bonds and point away from the β-structure region. The same comments are true for the minor β-structure region also. The region includes the N-terminal residue. The terminal amino group is heavily hydrated and two of the water bridges at this point additionally stabilize the region.

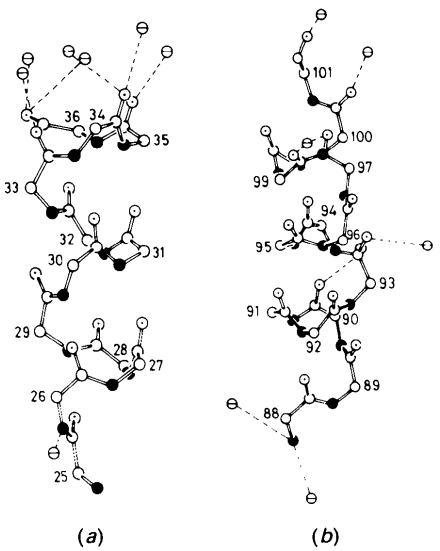


Fig. 4. The hydration of main-chain atoms in (*a*) the 25 to 36 and (*b*) the 88 to 101 helices.

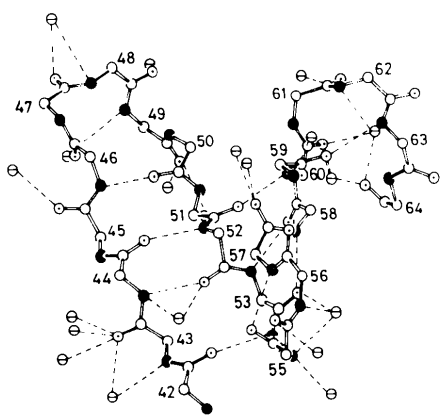


Fig. 5. Hydration of the main-chain atoms in the major β-structure region.

Understandably the reverse turns in the structure are heavily hydrated, as most of them occur on the protein surface. As reported earlier (Kodandapani, Suresh & Vijayan, 1990), these water molecules appear to lend additional stability to the turns. The 46 to 49 turn presents an interesting case, as can be seen from Fig. 6, in which a $4 \rightarrow 1$ hydrogen bond is substituted by a $4 \rightarrow 1$ water bridge. Turns of this type and other water-mediated turns have been reported earlier in small peptides (Aubury, Vituox, Boussard & Marraud, 1981; Nair & Vijayan, 1980).

Hydration of side chains

As observed in earlier studies (Blake, Pulford & Artymiuk, 1983; Baker & Hubbard, 1984) the number of interacting water molecules is least for Trp. Of the frequently occurring residues, Asn is the most heavily hydrated. Asp, Arg and Lys side chains are hydrated nearly to the same extent. It was observed that the variation, among different types of residues, in the average number of water molecules per hydrogen-bonding site is much lower than the variation in the number of molecules per side chain. The same trend is seen in other proteins (Baker & Hubbard, 1984).

Water networks in the hydration shell and heavily hydrated regions on the protein surface

Water molecules in the hydration shell are not uniformly distributed on the surface of the protein, nor do they form a continuous network. In the hydration shell there are 55 water molecules which do not interact with other water molecules belonging to the same shell. The number of interacting pairs of water molecules in the hydration shell is 19 and there are seven networks containing three water molecules each. The number of networks containing four, five and six water molecules is three, three and two respectively. There are also two more-extensive networks in the hydration shell, one containing nine water molecules and the other containing ten water molecules. The continuity of the hydration shell would, of course, be affected by the contacts of the protein molecule with its neighbours in the crystal.

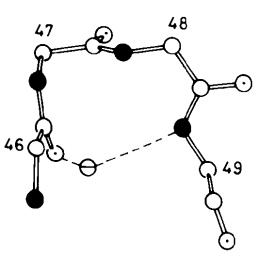


Fig. 6. A water-mediated β-turn.

However, that does not appear to explain entirely the discontinuous nature of the hydration shell. Discontinuities in the hydration shell are possibly a general feature of protein hydration.

The networks on the protein surface contain linear, branched and cyclic arrangements of water molecules. All the observed cyclic arrangements are triangular and six such triangular arrangements of water molecules are observed on the protein surface.

It is interesting to note that the most extensive network on the protein surface is of water molecules (ten in number) associated with the active site. Heavy hydration of active site is also observed in other enzymes [for example, ribonuclease A (Wlodawer, Svensson, Sjölin & Gilliland, 1988)]. The water structure associated with the protein groups that are involved in substrate binding (Johnson & Phillips, 1965; Imoto, Johnson, North, Phillips & Rupley, 1972; Ford, Johnson, Machin & Phillips, 1974; Perkins, Johnson, Machin & Phillips, 1978; Kelly, Sielecki, Sykes, James & Phillips, 1979), which encompasses the ten-membered network in addition to other water molecules, is illustrated in Fig. 7. The cleft involved in substrate binding is indeed a heavily hydrated region of the protein molecule. As many as 24 water molecules are associated with this region. The second most extensive network (involving nine water molecules) is associated with parts of the main β-structure and the loop region. One of the six-membered networks is also associated primarily with the loop region. Indeed, the loop region, made up of residues 65 to 73, is perhaps the most heavily hydrated stretch of polypeptide chain in the structure. The hydration of this region is illustrated in Fig. 8. As many as 23 water molecules interact with the nine residues comprising this region. 12 water molecules interact with seven CO groups and five water molecules with an equal number of main-chain NH groups. There are no main-chain–main-chain interactions within the loop to stabilize it. As can be seen from Fig. 8, many water molecules bridge different parts of the loop region and the loop appears to be stabilized primarily by interactions involving these water molecules. In addition to the two regions of conspicuous heavy hydration mentioned above, heavy hydration generally occurs in the regions which connect β-structure and helices, and terminii of these major secondary structural features. This is not surprising, as such regions are rich in reverse turns and are normally exposed to the solvent.

Molecular packing and intermolecular solvent bridges

With a solvent content as low as 22%, the structure can be considered to consist of close-packed protein molecules with interstices filled with solvent molecules. It is, therefore, interesting to examine the molecular packing in low-humidity monoclinic lysozyme crystals in the light of the approach developed by Kitaigorodsky (1955, 1973) for packing of organic molecules in their crystals.

The lysozyme molecules in the structure pack in layers perpendicular to the crystallographic b axis. In each layer, one molecule is surrounded by six other molecules, as schematically illustrated in Fig. 9. This arrangement corresponds to the closest packing of organic molecules in a layer. Adjacent molecular layers stacked along b are related by 2_1 screw axes. The stacking of the layers is such that four molecules in the layer above and four from the layer below come into contact with the reference molecule (Fig. 9). Thus each molecule is surrounded by, and is in contact with, 14 other molecules. A coordination number of 14 also corresponds to close packing of organic molecules (Kitaigorodsky, 1955, 1973).

Of a total of 172 intermolecular protein–protein contacts (<3.6 Å), only 86 are crystallographically

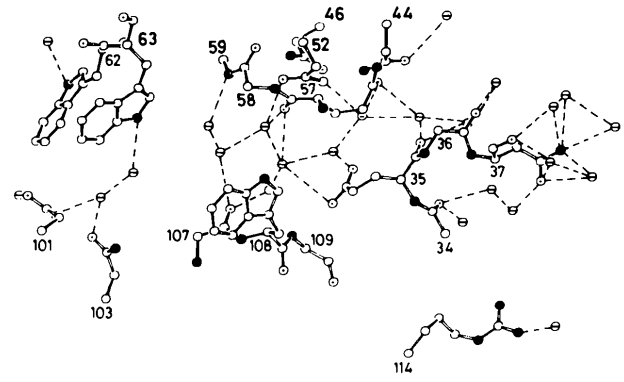


Fig. 7. Hydration of the hexasaccharide binding site.

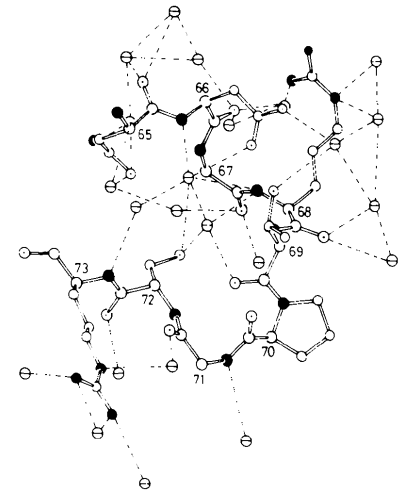


Fig. 8. Hydration of the loop region.

independent, as the surrounding molecules are symmetrically arranged in pairs about the reference molecule. Of the 86, 12 are hydrogen bonds. Interestingly nine of these intermolecular hydrogen bonds involve arginine and one lysine and there is only one main-chain–main-chain interaction among them. The interactions involving arginine include two salt bridges.

There are 76 water molecules, half of them crystallographically independent, that bridge the reference molecule with the surrounding 14 molecules. 72 of them bridge two protein molecules each, while the remaining four bridge three protein molecules each. There is considerable variation in the number of such water bridges between pairs of molecules: for example, while there are 13 bridges between the reference molecule and (4) [and between (7) and the reference molecule], there is only one bridge between the reference molecule and (10) [and between (14) and the reference molecule]. (See Fig. 9 for the numbering of protein molecules.) Altogether, 94 crystallographically independent protein atoms are involved in these bridges. 39 of them are main-chain atoms while the remaining 55 are side-chain atoms. Interestingly, 23 of the side-chain atoms belong to arginyl residues and all ten arginyl side chains in the structure are involved in such water bridges. The other side chains which frequently occur in these bridges are those of asparginyl (8) and aspartyl (5) residues.

There are several bridges, each involving two water molecules in series. There are almost as many such bridges as those involving one water molecule. There are also intermolecular bridges with three water molecules, but often it is difficult to delineate individual bridges from the network of water molecules in protein–protein interfaces. However, each water molecule in these networks interacts either directly with protein atoms or with the water molecules in the hydration shell.

The two nitrate ions in the structure are also involved in cross-linking protein molecules. The ions have well defined electron densities with average B values of 27.2 and 20.3 Å^2 which compare favourably with an average B value of 29.3 Å^2 for water molecules. The interactions involving the two ions are illustrated in Fig. 10. One of the nitrate ions (Fig. 10*a*) is involved in a type B specific interaction containing two hydrogen bonds (Salunke & Vijayan, 1981) with the guanidyl group of Arg 14 in the reference molecule and a hydrogen bond with the peptide nitrogen of Ser 72 in molecule (11). Furthermore, a water molecule bridges the nitrate ion and the CO group of Ser 72 of the same molecule. The imidazole group of His 15 in the reference molecule is in the immediate vicinity of the nitrate ion although no hydrogen bond exists between them. The guanidyl and the imidazole groups presumably provide a positive environment for the negatively charged nitrate ion. The second nitrate ion, the environment of which is shown in Fig. 10(*b*), is primarily anchored at the N terminus of the 5 to 15 α-helix in the reference molecule. It forms hydrogen bonds with the peptide NH groups of residues 5, 6

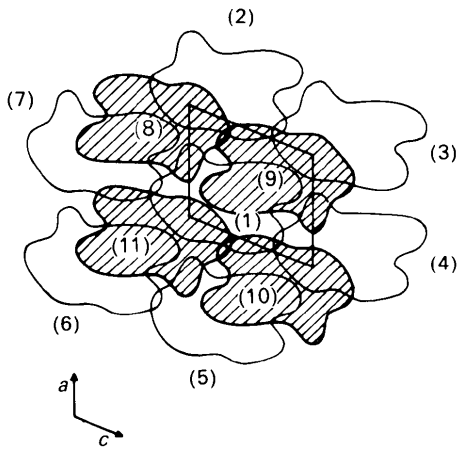


Fig. 9. Schematic representation of the molecular coordination in the structure. The reference molecule [(1) x, y, z] and the surrounding molecules [(2) $1+x, y, z$; (3) $x+1, y, z+1$; (4) $x, y, z+1$; (5) $x-1, y, z$; (6) $x-1, y, z-1$; (7) $x, y, z-1$] lie in the ac plane. The four screw-related molecules [(8) $1-x, y+\frac{1}{2}, -z$; (9) $1-x, y+\frac{1}{2}, 1-z$; (10) $-x, y+\frac{1}{2}, 1-z$; (11) $-x, y+\frac{1}{2}, -z$] in the layer above that of the reference molecule, and which are in contact with (1), are shaded. Four molecules in the layer below [(12) $1-x, y-\frac{1}{2}, -z$; (13) $1-x, y-\frac{1}{2}, 1-z$; (14) $-x, y-\frac{1}{2}, 1-z$; (15) $-x, y-\frac{1}{2}, -z$] are also in contact with (1).

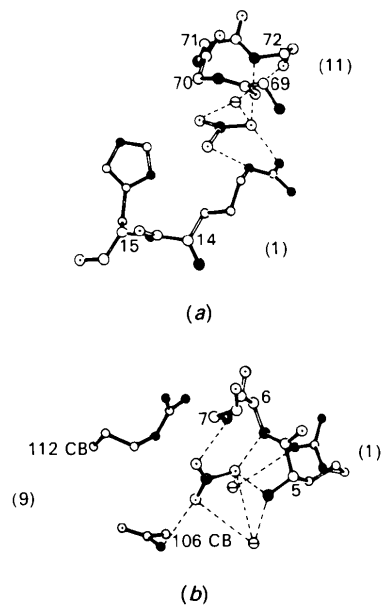


Fig. 10. (*a*), (*b*) Interactions involving the two nitrate ions. Figures in parentheses refer to the molecular numbering (Fig. 9).

and 7. It has water bridges with the main-chain NH and the side-chain guanidyl groups of Arg 5. The ion also forms a hydrogen bond with the side-chain amide nitrogen of Asn 106 in molecule (9). In addition, the guanidyl group of Arg 112 in this molecule is located as close as 3.7 Å to the nitrate ion. The partial positive charges at the N terminus of the 5 to 15 α-helix (Hol, 1985) and the guanidyl group presumably compensate the negative charge on the nitrate ion. Thus, not surprisingly, the two nitrate ions occupy positively charged niches in the interstices between protein molecules.

Water clusters and their geometrical features

A close examination of water molecules in the crystal structure appears to indicate that, unlike in many other protein structures such as insulin (Baker *et al.*, 1988), there are no continuous water channels in the crystal. Instead there are discrete water clusters, or pools, in the interstices among the protein molecules or on the protein surface. Those occurring in the interstices contain water molecules belonging to the hydration shells of the surrounding protein molecules as well as water molecules which do not directly interact with protein atoms. The largest such cluster of ordered water, illustrated in Fig. 11, contains 42 molecules and occurs in the interface produced by protein molecules (1) (reference molecule), (3), (4) and (13). The next largest cluster has 21 water molecules and is in contact with protein molecules (1), (2), (3) and (9). Then there are clusters with 18, 13, eight and five water molecules each. There are two four-membered clusters, two three-membered clusters and four two-membered clusters. The remaining 19 water molecules do not interact with other water molecules.

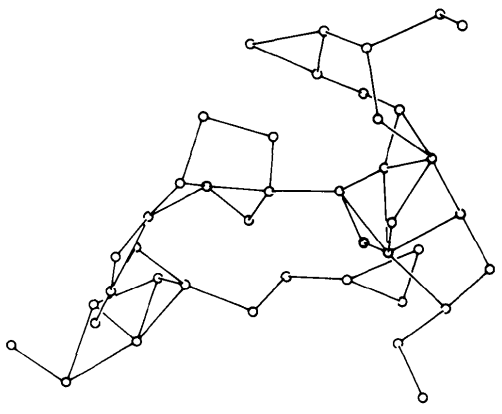


Fig. 11. The largest water cluster in the structure involving 42 water molecules.

The large clusters of water molecules observed in the structure are of considerable interest in relation to the water structure in biological systems. The organization of molecules in the clusters cannot be explained on the basis of any simple model. However, as can be seen from Fig. 11, there is a preponderance of three-membered rings. In fact, there are altogether 26 such rings in the structure. The number of four-membered rings is as low as four. The number of five-membered rings, which have been reported to occur extensively in crambin (Teeter, 1984), is still lower at three.

The frequent occurrence of three-membered rings is also reflected in the distribution of water–water–water angles shown in Fig. 12(*a*). The distribution has a peak at about 60°. Such a peak has also been observed for other structures (Blake, Pulford & Artymiuk, 1983; Baker *et al.*, 1988), although the frequent occurrence of three-membered rings has not been explicitly recognized. The other peak in the distribution is around 100–120° which obviously reflects the trigonal and the tetrahedral nature of the distribution of interactions around many water molecules. Yet another distribution of interest concerns that of water–water distances illustrated in Fig. 12(*b*). The distribution is rather broad, but peaks at around 2.9 Å. Finally, the distribution of the number of hydrogen-bonded water neighbours per water molecule is given in Fig. 12(*c*). As in the crystal

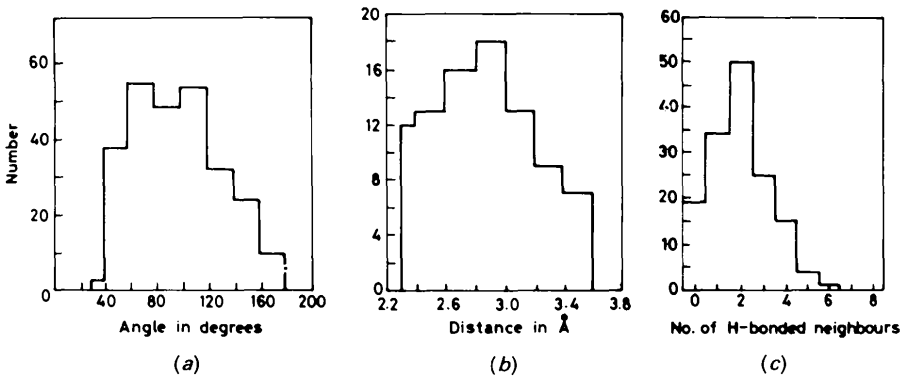


Fig. 12. Distribution of (*a*) water–water–water angles, (*b*) water–water distances and (*c*) number of hydrogen-bonded neighbours per ordered water molecule.

structures of tortoise and human lysozyme (Blake, Pulford & Artymiuk, 1983), the distribution has a peak at two. The mean value, also, works out to be two.

Comparison with the native structure

The structure of native monoclinic lysozyme is available at a resolution of 2.5 Å (Rao, Hogle & Sundaralingam, 1983). Only protein atoms have been located and the refinement with an R value of 26% for 6045 reflections does not appear to be complete. The structure, therefore, is much less accurate than that of the low-humidity form. This difference in accuracy is partially due to intrinsic factors. As noted earlier, contrary to expectations and in contrast to the situation observed in tetragonal lysozyme (Kodandapani, Suresh & Vijayan, 1990), the low-humidity form of monoclinic lysozyme diffracts to a much higher resolution than the native form does. Consequently, the mean temperature factor of the main-chain atoms in the low-humidity form is 5.9 Å^2 lower than that in the native form. As the resolution and the accuracy of the two structures are not comparable, the comparison here is confined to gross features. As water molecules have not been located in the native crystals, the question of comparison between hydration shells does not arise.

Molecular orientations and packing

The two crystallographically indepedent molecules in the asymmetric unit in the native crystals, which are related to each other by pseudo B centering, become equivalent in the low-humidity form. This transformation is illustrated in Fig. 13. With native cell dimensions of $a = 28.0$, $b = 62.9$, $c = 60.5$ Å and $\beta = 90.8°$, the transformed low-humidity form would have cell dimensions of $a = 28.0$, $b = 62.9$, $c = 33.5$ Å and $\beta = 114.8°$. However, the transformation is accompanied by loss of water and the actual observed cell dimensions of the low-humidity form are $a = 26.9$, $b = 58.9$, $c = 31.3$ Å and $\beta = 111.9°$. In addition to shrinkage of the three cell translations, the monoclinic angle changes by approximately 3°. Nevertheless, it may be noted that the **a**, **b** and **c*** directions in the native and the low-humidity forms are nearly coincident. The difference between the calculated and actual monoclinic angle, nearly 3°, may be considered to set the upper limit for the deviation between the coordinate systems. Therefore, for purposes of comparison, coordinates of molecules A and B in the native form and those of the molecule in the low-humidity form were referred to a, b and c^* of the respective crystals.

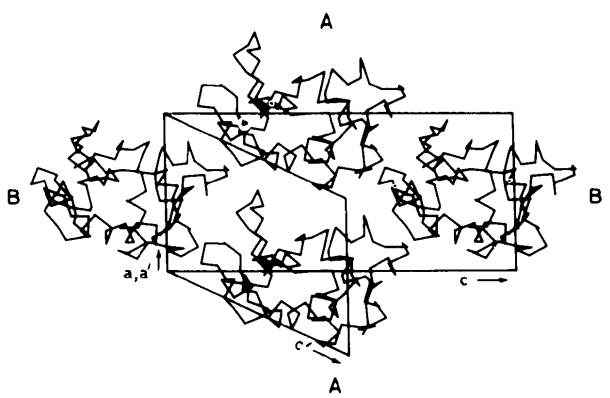


Fig. 13. Illustration of the gross relationship between the native and the low-humidity forms. Only molecules in a layer are given. The layers above and below are related to the reference layer by 2_1 symmetry. Molecules A and B in the native structure are indicated. They become equivalent in the low-humidity form. The unprimed and primed symbols correspond to the native and the low-humidity forms respectively.

The mutual orientations of different molecules, based on C^α positions, were determined using a method developed by Rossmann & Argos (1975) and adapted for the DEC 1090 computer by Dr M. R. N. Murthy. Molecule B in the native crystals can be oriented in the same way as molecule A using the orientation matrix

$$\begin{matrix} 0.9724 & -0.2008 & 0.1193 \\ 0.2131 & 0.9717 & -0.1019 \\ -0.0955 & 0.1245 & 0.9876 \end{matrix}$$

which corresponds to Eulerian angles $\theta_1 = -40.5$, $\theta_2 = 9.0$ and $\theta_3 = 52.5°$ as defined in the zyz system. The matrices and angles that relate molecule A and molecule B respectively to the molecule in the low-humidity form are

$$\begin{matrix} 0.9866 & 0.1627 & -0.0096 \\ -0.1620 & 0.9855 & 0.0513 \\ 0.0178 & -0.0491 & 0.9986 \end{matrix}$$

with angles $\theta_1 = 100.6$, $\theta_2 = 3.0$ and $\theta_3 = -70.1°$, and

$$\begin{matrix} 0.9951 & 0.0399 & 0.0908 \\ -0.0463 & 0.9965 & 0.0694 \\ -0.0877 & -0.0733 & 0.9935 \end{matrix}$$

with angles $\theta_1 = 37.6$, $\theta_2 = 6.5$ and $\theta_3 = -40.1°$. It may be noted that the differences in the orientations are not as large as would appear initially from Eulerian angle values. When θ_2 is close to zero, θ_1 and θ_3 represent rotations about almost the same axis and $\theta_1 + \theta_3$ is relevant. The differences in orientations are yet quite substantial.

The difference in molecular packing between the two forms is best described in terms of the distances of a given protein molecule from the 14 neighbours that surround it. The surrounding molecules are related to the reference molecule by crystallographic

symmetry elements in the low-humidity form. Also, they are symmetrically disposed about the reference molecule in pairs. The situation is different in the native form because of the presence of two crystallographically independent molecules in the unit cell. The surroundings of the two molecules are very similar but not identical. Each molecule *A* is surrounded by eight molecules *B* and six molecules *A*. Likewise each molecule *B* is surrounded by eight molecules *A* and six molecules *B*. In each set, molecules of one type are related by crystallographic symmetry elements while they are related to those of the other type by pseudo centering or pseudo 2_1 axes.

The distances of the surrounding molecules from the reference molecule in the three cases are given in Table 3. On average, these distances are considerably shorter in the low-humidity form than in the native crystals, although there are a few instances where a distance in the native form is marginally shorter than the corresponding distance in the low-humidity form. The reduction in these distances during the transformation to the low-humidity form could be as high as 7 Å. On average, differences in intermolecular distances are comparatively low when the centroids of the molecules have the same, or nearly the same, *y* coordinates. Large differences are associated with screw- or pseudo-screw-related molecules. This is not surprising as the shrinkage of the unit cell along the **b** direction is much larger than that along the other directions. It should, however, be emphasized that this is only an average effect; the actual differences in intermolecular distances are highly anisotropic and they cannot be explained in terms of simple shrinkage.

It is clear that the molecules undergo very substantial rearrangement, rotational as well as translational, during the transformation. This accounts for the inability to refine the low-humidity structure with molecule *A* or molecule *B* as the starting model. It is also interesting to note that the substantial rearrangement improves the order in the crystals leading to a much better diffraction pattern.

Table 3. *Distances* (Å) *of the centroid of the reference molecule from the centroids of the* 14 *neighbouring molecules*

See Fig. 9 for numbering of molecules.

Molecule	Low-humidity form	Native form Mol. *A*	Native form Mol. *B*
2	26.9	28.0	28.0
3	32.8	32.7	33.7
4	31.3	32.3	34.8
5	26.9	28.0	28.0
6	32.8	33.7	32.7
7	31.3	34.8	32.3
8	39.0	38.1	39.7
9	32.7	33.8	43.5
10	37.4	42.7	44.4
11	35.0	37.1	38.8
12	39.0	39.7	38.1
13	32.7	33.8	34.5
14	37.4	42.7	44.4
15	35.0	38.8	37.1

Changes in molecular structure

For comparison, molecules *A* and *B* in the native crystals were superimposed on the molecule in the low-humidity form using appropriate transformation matrices. The r.m.s. differences in the positions of the main-chain atoms in molecules *A* and *B* with respect to those in the molecule in the low-humidity form are 0.61 and 0.67 Å respectively. The corresponding difference between those in molecules *A* and *B* is 0.73 Å. The main-chain conformations of the three molecules are almost the same although differences in detail exist. Larger differences exist in the side-chain conformation. The differences are particularly pronounced, as illustrated in Fig. 14, in Trp 62 and Trp 63. Shifts in these side chains, particularly in that of Trp 62, always accompany inhibitor binding (Johnson & Phillips, 1965; Imoto, Johnson, North, Phillips & Rupley, 1972; Ford, Johnson, Machin & Phillips, 1974; Perkins, Johnson, Machin & Phillips, 1978; Kelly, Sielecki, Sykes, James & Phillips, 1979). The side chain of Trp 62 was also found to shift substantially during water-mediated transformations in tetragonal lysozyme (Kodandapani, Suresh & Vijayan, 1990).

We thank Drs K. Suguna and S. C. Mande for their help in the structure solution by the molecular replacement method. Our thanks are also due to Professor Sir David Phillips for encouragement. The refinement program used in this investigation was supplied by Eleanor Dodson, the data processing and Fourier programs by Professor M. R. N. Murthy and the HP 1000 version of the *ORTEP* (Johnson, 1970) program which was used for preparing some of the figures by Professor C. Ramakrishnan. We thank them all and also the Supercomputer Education and Research Centre of the Institute, for

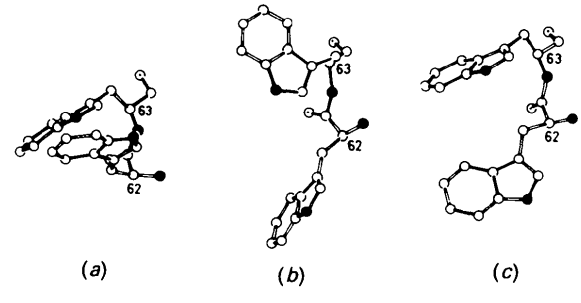


Fig. 14. Disposition of Trp 62 and Trp 63 in (*a*) molecule *A* in the native crystals, (*b*) molecule *B* in the native crystals and (*c*) the molecule in the low-humidity form, as viewed along comparable directions.

their help. Our thanks are also due to C. Govindswamy, James Paul and K. B. Shobana for technical help. Financial support from the Department of Science and Technology, Government of India, is acknowledged. M and RK have been recipients of fellowships from the Council of Scientific and Industrial Research, India.

References

AGARWAL, R. C. (1978). *Acta Cryst.* A**34**, 791–809.

ARNDT, U. W. & WONACOTT, A. J. (1977). *The Rotation Method in Crystallography*. Amsterdam: North-Holland.

AUBURY, A., VITUOX, B., BOUSSARD, G. & MARRAUD, M. (1981). *Int. J. Pept. Protein Res.* **8**, 195–202.

BAKER, E. N., BLUNDELL, T. L., CUTFIELD, J. F., CUTFIELD, S. M., DODSON, E. J., DODSON, G. G., HODGKIN, D. C., HUBBARD, R. E., ISAACS, N. W., REYNOLDS, C. D., SAKABE, K., SAKABE, N. & VIJAYAN, M. (1988). *Philos. Trans. R. Soc. London*, **319**, 369–456.

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

BHAT, T. N. & COHEN, G. H. (1984). *J. Appl. Cryst.* **17**, 244–248.

BHAT, T. N., SASISEKHARAN, V. & VIJAYAN, M. (1979). *Int. J. Pept. Protein Res.* **13**, 170–184.

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

BLUNDELL, T. L., BARLOW, D., BORAKAKOTI, N. & THORNTON, J. (1983). *Nature (London)*, **306**, 281 283.

CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173–178. New York: Gordon and Breach.

FINNEY, J. L. (1979). *Water, a Comprehensive Treatise*, Vol. 6, edited by F. FRANKS, pp. 47–122. New York: Plenum Press.

FITZGERALD, P. M. D. (1988). *J. Appl. Cryst.* **21**, 273–278.

FORD, L. O., JOHNSON, L. N., MACHIN, P. A. & PHILLIPS, D. C. (1974). *J. Mol. Biol.* **88**, 349–371.

HENDRICKSON, W. A. & KONNERT, J. H. (1980*a*). *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 13.01–13.26. Banglore: Indian Academy of Sciences.

HENDRICKSON, W. A. & KONNERT, J. H. (1980*b*). *Biomolecular Structure, Function, Conformation and Evolution*, Vol. 1, edited by R. SRINIVASAN, pp. 43–57. Oxford: Pergamon Press.

HOL, W. G. J. (1985). *Prog. Biophys. Mol. Biol.* **45**, 149–195.

IMOTO, T., JOHNSON, L. N., NORTH, A. C. T., PHILLIPS, D. C. & RUPLEY, J. A. (1972). *The Enzymes*, Vol. 7, edited by P. D. BOYER, pp. 665–808. New York: Academic Press.

JANIN, K., WODAK, S., LEVITT, M. & MAIGRET, B. (1978). *J. Mol. Biol.* **125**, 357–386.

JOHNSON, C. K. (1970). *ORTEP*. Report ORNL-3794. Oak Ridge National Laboratory, Tennessee, USA.

JOHNSON, L. N. & PHILLIPS, D. C. (1965). *Nature (London)*, **206**, 761–763.

KELLY, J. A., SIELECKI, A. R., SYKES, B. D., JAMES, M. N. G. & PHILLIPS, D. C. (1979). *Nature (London)*, **282**, 875–878.

KITAIGORODSKY, A. I. (1955). *Organic Chemical Crystallography*. New York: Consultants Bureau.

KITAIGORODSKY, A. I. (1973). *Molecular Crystals and Molecules*. New York: Academic Press.

KODANDAPANI, R., SURESH, C. G. & VIJAYAN, M. (1990). *J. Biol. Chem.* **265**, 16126–16131.

KODANDAPANI, R. & VIJAYAN, M. (1991). *Molecular Conformation and Biological Interactions*, edited by P. BALARAM & S. RAMASESHAN, pp. 149–170. Bangalore: Indian Academy of Sciences.

KONNERT, J. H. (1976). *Acta Cryst.* A**32**, 614–617.

KUNDROT, C. E. & RICHARDS, F. M. (1987). *J. Mol. Biol.* **193**, 157–170.

LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.

MCGREGOR, M. J., ISLAM, S. A. & STERNBERG, M. J. E. (1987). *J. Mol. Biol.* **198**, 295–310.

MADHUSUDAN & VIJAYAN, M. (1991). *Curr. Sci.* **60**, 165–170.

MALIN, R., ZIELENKIEWICZ, P. & SAENGER, W. (1991). *J. Biol. Chem.* **266**, 4848–4852.

MANDE, S. C. (1991). PhD thesis, Indian Institute of Science, Bangalore, India.

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

NAIR, C. M. & VIJAYAN, M. (1980). *J. Chem. Soc. Perkin Trans.* 2, pp. 1800–1804.

NIXON, P. E. & NORTH, A. C. T. (1976). *Acta Cryst.* A**32**, 320–325.

PERKINS, S. J., JOHNSON, L. N., MACHIN, P. A. & PHILLIPS, D. C. (1978). *Biochem. J.* **A173**, 607–616.

RAMACHANDRAN, G. N., RAMAKRISHNAN, C. & SASISEKHARAN, V. (1963). *J. Mol. Biol.* **7**, 95–99.

RAMACHANDRAN, G. N. & SASISEKHARAN, V. (1968). *Adv. Protein Chem.* **3**, 284–438.

RAMANADHAM, M., SIEKER, L. C. & JENSEN, L. H. (1990). *Acta Cryst.* B**46**, 63–69.

RAO, S. T., HOGLE, J. & SUNDARALINGAM, M. (1983). *Acta Cryst.* C**39**, 237–240.

RICHARDSON, J. S. & RICHARDSON, D. C. (1989). *Prediction of Protein Structure and Principles of Protein Conformation*, edited by G. D. FASMAN, pp. 1–98. New York: Plenum Press.

ROCKLAND, L. B. (1960). *Anal. Chem.* **32**, 1375–1376.

ROSSMANN, M. G. (1972). *The Molecular Replacement Method*. New York: Gordon and Breach.

ROSSMANN, M. G. (1979). *J. Appl. Cryst.* **12**, 225–238.

ROSSMANN, M. G. (1985). *Methods Enzymol.* **14**, 237–280.

ROSSMANN, M. G. & ARGOS, P. (1975). *J. Biol. Chem.* **250**, 7525–7532.

ROSSMANN, M. G., LESLIE, A. G. W., ABDEL-MEGUID, S. S. & TSUKIHARA, T. (1979). *J. Appl. Cryst.* **12**, 570–581.

SAKABE, K., SAKABE, N. & SASAKI, K. (1980). *Water and Metal Cations in Biological Systems*, edited by B. PULLMAN & K. YAGI, pp. 117–128. Tokyo: Japan Scientific Society Press.

SALUNKE, D. M., VEERAPANDIAN, B., KODANDAPANI, R. & VIJAYAN, M. (1985). *Acta Cryst.* B**41**, 431–436.

SALUNKE, D. M., VEERAPANDIAN, B. & VIJAYAN, M. (1984). *Curr. Sci.* **53**, 231–235.

SALUNKE, D. M. & VIJAYAN, M. (1981). *Int. J. Pept. Protein Res.* **18**, 348–351.

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

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

THANKI, N., THORNTON, J. M. & GOODFELLOW, J. M. (1988). *J. Mol. Biol.* **202**, 637–657.

THANKI, N., THORNTON, J. M. & GOODFELLOW, J. M. (1990). *Protein Eng.* **3**, 495–508.

VIJAYAN, M. (1980). *Acta Cryst.* A**36**, 295–298.

WEAVER, L. H., TRONRUD, D. E., NICHOLSON, H. & MATTHEWS, B. W. (1990). *Curr. Sci.* **59**, 833–837.

WLODAWER, A., DEISENHOFER, J. & HUBER, R. (1987). *J. Mol. Biol.* **193**, 145–156.

WLODAWER, A., SVENSSON, L. A., SJÖLIN, L. & GILLILAND, G. A. (1988). *Biochemistry*, **27**, 2705–2717.