Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Hydration, mobility and accessibility of lysozyme: structures of a pH 6.5 orthorhombic form and its low-humidity variant and a comparative study involving 20 crystallographically independent molecules

B. K. Biswal, N. Sukumar and M. Vijayan*

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

Correspondence e-mail: mv@mbu.iisc.ernet.in

Received 1 March 2000
Accepted 22 June 2000

PDB References: orthorhombic pH 6.5 lysozyme, 1f0w; orthorhombic pH 6.5 lysozyme, low-humidity form, 1f10.

The structure analyses of orthorhombic lysozyme grown at pH 6.5 and its low-humidity variant are reported. The structures of the same form grown at pH 9.5 and 4.5 and that of the low-humidity variant of the pH 9.5 form are available. A comparison between them shows that the changes in molecular geometry and hydration caused by changes in the amount of solvent surrounding protein molecules are more pronounced than those caused by variation in pH. In particular, the conformation and the mutual orientation of the catalytic residues Glu35 and Asp52 remain unaffected by change in pH. A comparative study involving 20 crystallographically independent lysozyme molecules, including five in the orthorhombic form, leads to the delineation of the relatively rigid, moderately flexible and highly flexible regions of the molecule. Half the binding cleft (subsites *D*, *E* and *F*) belong to the rigid region but the other half (subsites *A*, *B* and *C*) belong to a flexible region. There is no marked correlation between relative rigidity and conservation of side-chain conformation except at the binding site. The study permits the identification of seven invariant water molecules associated with the protein. Most of them are involved in important tertiary interactions, while one occurs in the active-site cleft. The study demonstrates a weak correlation between non-accessibility and rigidity. On average, the level of hydration of polar atoms increases rapidly with accessible atomic surface area, but levels off at about 15 Å^2 at a little over one ordered water molecule per polar protein atom. Only 15 N and O atoms are hydrated in all 20 molecules. 13 of these are hydrated by the seven invariant water molecules. Of the seven, only one water molecule is totally buried within the protein.

1. Introduction

Hydration and the associated flexibility of proteins are of considerable general interest (Teeter, 1984; Saenger, 1987; Sundaralingam & Shekarudu, 1989; Kodandapani *et al.*, 1990; Kramer & Berman, 1998; Sadasivan *et al.*, 1998; Nagendra *et al.*, 1998; Dong *et al.*, 1999). We have been investigating them primarily through an approach involving water-mediated transformations, in which protein crystals undergo reversible transformations with change of solvent content when the environmental humidity is systematically varied, with hen egg-white lysozyme and bovine pancreatic ribonuclease A as the model systems (Salunke *et al.*, 1985; Kodandapani *et al.*, 1990; Madhusudhan & Vijayan, 1991; Madhusudan *et al.*, 1993; Radha Kishan *et al.*, 1995; Nagendra *et al.*, 1995, 1996, 1998;

Sadasivan *et al.*, 1998; Sukumar *et al.*, 1999). These investigations established the possibility of identifying the relatively invariant features of the hydration shell and elucidating the relative mobility of different regions of the protein molecules (Kodandapani *et al.*, 1990; Madhusudhan & Vijayan, 1991; Madhusudan *et al.*, 1993; Radha Kishan *et al.*, 1995; Sadasivan *et al.*, 1998; Nagendra *et al.*, 1998; Sukumar *et al.*, 1999). Furthermore, a relationship could be established between hydration, mobility and enzyme action (Radha Kishan *et al.*, 1995; Nagendra *et al.*, 1996). A plausible explanation for the loss of activity that accompanies extreme dehydration (solvent content less than 20%) could also be provided through studies on crystal forms with unusually low solvent contents (Nagendra *et al.*, 1998). Yet another result of the studies pertains to the relation between accessibility and hydration (Sadasivan *et al.*, 1998).

Until recently, most studies on lysozyme were carried out on crystals grown at acidic pH (pH 4.6). In order to explore the structural changes that accompany variation in pH, we studied the high-resolution structure of orthorhombic lysozyme grown at the basic pH 9.5 and its low-humidity variant (Sukumar *et al.*, 1999). While the structure remains essentially the same at these two pH values, there are structural changes that accompany water-mediated transformations. Here, we report in the first part of the paper the structure of orthorhombic lysozyme grown at an intermediate pH 6.5 and its low-humidity variant in order to complete the study of structural variability of lysozyme as a function of pH.

With the present study, the structures of the native and the low-humidity forms of tetragonal, monoclinic and orthorhombic lysozyme, the last-mentioned at two different pH values, are available. Also available are the structures of these three and the triclinic crystal forms of the enzyme grown under different environmental conditions. The reported structures of uncomplexed, unmodified lysozyme account for 20 crystallographically independent copies of the molecule (after ignoring a couple to avoid repetition of structures involving near-identical conditions) situated in environments involving varying degrees of differences. They constitute the data set available to identify, with a high degree of confidence, the invariant water molecules in the hydration shell and the structural role of these water molecules, to delineate the variability in the mobility of the different parts of the molecule and to explore the relation between accessibility, hydration and mobility. The second part of the paper reports results in this area.

2. Experimental

2.1. Crystallization

Hen egg-white lysozyme was purchased from the Sigma Chemical company and was used directly for crystallization without further purification. Crystals of the enzyme were grown using the hanging-drop technique. Droplets containing 40 mg ml^{-1} of the protein in phosphate buffer (pH 6.5) with 5%(*w*/*v*) NaCl were equilibrated against the same buffer containing 20%(*w*/*v*) NaCl. Good crystals took about two weeks to grow. The low-humidity form (88%) was obtained by placing a drop of a supersaturated solution of potassium chromate (Rockland, 1960; Weast & Astle, 1980/1981) at about 1 cm from the crystal in the thin-walled glass capillary. The transformation occurred in about 24 h.

Table 1
Data-collection statistics.

Values in parentheses refer to the highest resolution shell (2.0–1.9 and 1.8–1.7 Å in the native and 88% r.h. forms, respectively). The number of unobserved reflections ($I = 0$) is rather large, especially in high-resolution shells. Their exclusion in refinement calculations is unlikely to have any major effect on the results, except that the effective resolution of the structure is likely to be less than the limit to which data have been recorded. However, $I = 0$ reflections have been included in the calculation of merging R.

	Native	88% r.h. form
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters		
a (Å)	30.47	30.58
b (Å)	59.39	55.86
c (Å)	68.78	68.58
Z	4	4
Unit-cell volume (Å^3)	124465	117148
Solvent content (%)	42.2	38.5
Data resolution (Å)	1.9	1.7
Number of observations	42921	54245
Number of unique reflections	10122 (1351)	13459 (2083)
Number of reflections with $I = 0$	3284 (899)	2823 (1064)
Completeness of data (%)	97.2 (94.7)	99.8 (99.9)
Merging R for all reflections (%)	6.4 (47.5)	5.8 (33.2)
Average $I/\sigma(I)$	18 (1.6)	25 (3.1)

2.2. Data collection

Intensity data from both the native and the low-humidity forms were collected at 293 K on a MAR Research imaging plate with X-rays generated by a Rigaku rotating-anode generator using Cu $K\alpha$ radiation. The crystal-to-detector distance was 100 mm in both cases. The data were processed using the *XDS* program package (Kabsch, 1988). Data-collection statistics for both the native and the low-humidity forms are given in Table 1.

2.3. Structure refinement

Both the native and the low-humidity structures were refined in the same manner. The structure of the native form was refined using as input the coordinates of orthorhombic lysozyme at pH 9.5 (Sukumar *et al.*, 1999). The low-humidity form was refined using native pH 6.5 form as the starting model. Structure refinement was carried out using *X-PLOR* (Brünger, 1992). Initially, the structure was refined treating the whole molecule as a rigid body. This was followed by molecular-dynamics refinement using the simulated-annealing technique. Electron-density maps were calculated at this stage to correct and rebuild the model wherever necessary using *FRODO* (Jones, 1978). The refinement of atomic parameters was continued using *X-PLOR*. Once it converged, around an R value of 0.22, identification of water molecules began. This

Table 2
Refinement parameters.

	Native	88% r.h. form
Resolution limit used in refinement (Å)	10.0-1.9	10.0-1.7
Number of reflections with $F > 0$	6756	10564
Final R factor (%)	18.8	20.3
R_{free}† (%)	27.9	26.8
R.m.s. deviations from ideal		
Bond lengths (Å)	0.014	0.013
Bond angles (°)	1.9	1.8
Dihedral angles (°)	25.0	25.0
Improper angles (°)	1.7	1.6
Number of protein atoms	1001	1001
Number of water molecules	137	126

† 10% of the reflections were used for computing R_{free}.

Table 3
R.m.s. deviations (Å) in C^{α} positions among available native and low-humidity (lh) orthorhombic structures.

Those for side-chain atoms are given in parentheses. Cruickshank Precision Indices for coordinates (Cruickshank, 1999) of pH 4.5, pH 6.5, pH 6.5lh, pH 9.5 and pH 9.5lh forms are 0.09, 0.25, 0.14, 0.20 and 0.40 Å, respectively. The corresponding Luzzati indices (Luzzati, 1952) are 0.22, 0.23, 0.22, 0.21 and 0.26 Å, respectively.

	pH 6.5	pH 9.5	pH 6.5lh	pH 9.5lh
pH 4.5	0.18 (0.89)	0.13 (0.85)	—	—
pH 6.5	—	0.14 (0.41)	0.46 (1.21)	—
pH 9.5	—	—	—	0.43 (1.21)
pH 6.5lh	—	—	—	0.34 (1.13)

was performed in several stages based on peaks of at least 3σ in $F_o - F_c$ and 0.9σ in $2F_o - F_c$ electron-density maps. The same criteria were used for identifying water molecules in our earlier studies. Cycles of positional and B-factor refinement, correction of the model using Fourier maps and the location of water molecules continued until no significant density ($>3\sigma$ in $F_o - F_c$ and $>0.9\sigma$ in $2F_o - F_c$ maps) was left in the maps. Towards the end of the refinement, the model was checked against omit maps (Vijayan, 1980; Bhat & Cohen, 1984) to confirm the location of water molecules and to minimize the effects of model bias. The stereochemical acceptability of the structures were checked using the program *PROCHECK* (Laskowski *et al.*, 1993). Refinement converged with R values of 0.188 and 0.203 for the native and low-humidity forms, respectively. R_{free} was closely monitored in the course of refinement. The refinement parameters are given in Table 2. In the final models, 85% or more of the residues fall in the most favoured regions of the Ramachandran plot, while 14% fall in the additional allowed regions. One residue in the native form is in a generously allowed region, while none are in the disallowed regions.

2.4. Accessibility and superpositions

Solvent-accessible surface area and cavities were estimated using the program *MSP* (Connolly, 1993). A probe radius of 1.2 Å was used throughout the calculations, following the suggestions of Hubbard & Argos (1995) and Sadasivan *et al.* (1998). The program *HOMOMGR* (Rossmann & Argos, 1975) was used for superposition of different structures.

3. Results and discussion

3.1. Effect of changes in pH and solvent content on protein structure and hydration

The structure of the lysozyme molecule in the pH 6.5 orthorhombic form is very similar to that of the molecule in the other crystal forms. However, this crystal structure and those of orthorhombic forms grown at pH 4.5 (D. Carter, J. He, J. R. Ruble & B. Wright, unpublished work; PDB code 1aki) and pH 9.5 (Sukumar *et al.*, 1999) provide information on the change in molecular structure as a function of pH as the crystal packing in all the three is exactly the same. The low-humidity variants of the pH 6.5 (present study) and the pH 9.5 (Sukumar *et al.*, 1999) forms are also available. Comparison of those forms with their respective native forms provides a measure of changes in molecular structure arising from variation in the amount of surrounding water in the same packing environment as that in which the consequence of changes in pH are studied. The r.m.s. deviations in C^{α} positions between different forms (Table 3) clearly show that on average the changes in the main-chain conformation brought about by the removal of a small amount of bulk water from the surroundings of the molecule are more pronounced than those caused by variations in the pH of the medium. The variations in individual C^{α} positions between the three native forms are illustrated in Fig. 1. They are small along the polypeptide chain and are nearly uniformly distributed, although they are somewhat larger in the loops in the C-terminal segment. The variations during water-mediated transformations (Fig. 2) are larger and are more pronounced in the major loop (residues 65–73) and in the beginning of the loop following the 88–101 helix, presumably indicating the inherent flexibility of these loops. In addition, the C-terminal stretch, involving about 25 residues, exhibits larger deviations in C^{α} positions during the transformation than do the other regions.

The differences in side-chain conformation between the five forms follow the same pattern as those in the main-chain conformation (Table 3). Interestingly, the conformation and the mutual orientation of the catalytic residues Glu35 and Asp52 are essentially the same in all five forms. The torsional angles χ^1 and χ^2, which essentially define the conformation of the glutamyl side chain, vary within the ranges -60 to -88° and -59 to -70°, respectively, for Glu35 in the structures. Likewise, χ^1 for Asp52 lies within the range -53 to -62°. Around pH 5, where the activity is maximum, Glu35 is believed to be neutral and Asp52 ionized (Stryer, 1981). Both are presumably ionized at pH 6.5 and 9.5. However, the change in ionization state does not appear to have any striking structural consequences.

The hydration shell of a protein molecule is considered to be made up of ordered molecules at a distance 3.6 Å or less from a protein O or N atom (Kodandapani *et al.*, 1990; Madhusudan *et al.*, 1993; Radha Kishan *et al.*, 1995; Nagendra *et al.*, 1996). From the crystallographically located water O

Table 4
Structures used in the analysis.

Structures†		Resolution (Å)	Solvent content (%) (Matthews, 1968)	No. of water molecules in hydration shell	PDB code
1	Tetragonal	2.00	39.40	120	—
2	Tetragonal low humidity	2.10	37.80	121	4lym
3	High-pressure tetragonal	2.00	38.80	148	3lym
4	APCF tetragonal	1.33	38.50	147	193l
5	Low-temperature tetragonal	1.75	34.60	127	1lsf
6	Triclinic	0.95	30.89	165	4lzt
7	Triclinic low temperature	0.92	25.90	249	3lzt
8	Monoclinic molecule *A*	2.00	32.00	125	1uco
9	Monoclinic molecule *B*	2.00	32.00	131	1uco
10	Monoclinic low humidity	1.75	22.00	173	1lma
11	Iodinated monoclinic molecule *A*	1.60	32.00	148	1lkr
12	Iodinated monoclinic molecule *B*	1.60	32.00	157	1lkr
13	High-temperature monoclinic molecule *A*	1.72	29.71	111	1lys
14	High-temperature monoclinic molecule *B*	1.72	29.71	128	1lys
15	Orthorhombic pH 4.5	1.50	40.86	92	1aki
16	Orthorhombic pH 6.5	1.90	42.20	113	1f0w
17	Orhorhombic pH 6.5 low humidity	1.70	38.50	125	1f10
18	Orthorhombic pH 9.5	1.90	42.42	128	1hsx
19	Orthorhombic pH 9.5 low humidity	2.00	38.00	114	1hsw
20	High-temperature orthorhombic	1.70	43.40	121	1bgi

† References: 1, D. C. Phillips, personal communication; 2, Kodandapani *et al.* (1990); 3, Kundrot & Richards (1987); 4, Vaney *et al.* (1996); 5, Kurinov & Harrison (1995); 6 and 7, Walsh *et al.* (1998); 8 and 9, Nagendra *et al.* (1996); 10, Madhusudan *et al.* (1993); 11 and 12, Steinrauf (1998); 13 and 14, Harata (1994); 15, D. Carter, J. He, J. R. Ruble and B. Wright, unpublished results; 16 and 17, present structures; 18 and 19, Sukumar *et al.* (1999); 20, Oki *et al.* (1999).

atoms in the different structures, the numbers of molecules in the hydration shells of the pH 4.5 native, pH 6.5 native, pH 6.5 low-humidity, pH 9.5 native and pH 9.5 low-humidity forms work out to be 92, 113, 127, 128 and 114, respectively. In our previous studies (Kodandapani *et al.*, 1990; Madhusudan *et al.*, 1993; Radha Kishan *et al.*, 1995; Nagendra *et al.*, 1996), a water molecule in one form and a corresponding one in another were considered to be equivalent if they interact with at least one common protein atom and if the distance between the two water molecules is less than 1.8 Å when the protein molecules along with the hydration shell are superposed. The number of such equivalent water molecules between the pH 9.5 and pH 6.5 forms is 85, while that between the native and the low-humidity forms at pH 9.5 and pH 6.5 is lower, at 56 and 57, respectively. Even though the number of identified molecules in the hydration shell in the pH 4.5 form is lower (at 92) in the pH 4.5 form, presumably on account of the stricter criterion used for locating them, 61 and 66 of them have equivalents in the pH 6.5 and pH 9.5 forms, respectively. Thus, in the case of hydration the changes arising from variation in solvent content are also more pronounced than those arising from variation in pH.

3.2. Plasticity of the lysozyme molecule

The availability of 20 crystallographically independent copies of the molecules, listed in Table 4, with widely different environments in terms of pH, crystal packing, amount of surrounding solvent and its composition, pressure and temperature, permits a meaningful delineation of the rigid and the flexible regions of the molecule. A simple computational approach, developed earlier by us (Sadasivan *et al.*, 1998), was used for this purpose. Earlier, an attempt has been made to delineate the rigid and flexible regions in the molecule through the manual examination of difference distance plots (Madhusudhan & Vijayan, 1991) using the limited set of five crystallographically independent molecules which were then available. A ratio of residues in the rigid and flexible regions similar to found in the previous attempt was obtained when a cutoff value of 0.19 Å in r.m.s.(i) (Sadasivan *et al.*, 1998) was used. As was performed earlier by Madhusudan and Vijayan using limited data, highly flexible regions were also delineated using a cut-off value of 0.35 Å. They form a subset of the flexible regions, the remainder of which may be termed as moderately flexible regions. The delineation among the relatively rigid, moderately flexible and highly flexible regions is illustrated in Fig. 3.

The calculations outlined above indicated that residues 2–3, 6–8, 11, 26–43, 51–60, 88–97 and 106 belong to the relatively rigid regions of the molecule. As is well known, the lysozyme molecule is considered to be made up of two domains: domain I, consisting of an N-terminal segment (residues 1–38) and a C-terminal segment (residues 88–129), and domain II made up of residues 39–87. Clearly the relatively rigid residues are not uniformly distributed among the three polypeptide stretches. The proportion of such residues in the C-terminal segment of domain I is comparatively low. In fact, almost the whole of the 30 residues in the C-terminal stretch is flexible. Interestingly, this is the same segment which exhibited the maximum changes when the level of hydration of the same crystal form was systematically varied (Nagendra *et al.*, 1998). The molecule contains four α-helices: 5–15 (H1), 25–36 (H2), 88–101 (H3) and 109–115 (H4). Four of the 11 residues in H1, all of those in H2 and ten of the 14 residues in H3 are in the rigid region. The whole of the C-terminal helix H4 is in a flexible region. 17 of the 28 residues in the sheets [42–63(S1) and 1–3, 38–40(S2)] are in the rigid region. Interestingly, the loops are all flexible, only one (106) residue in them having been identified as rigid. A subset of 22 residues (45–49, 67–73, 99–103, 109, 113, 121, 128–129) in the flexible regions have been delineated as highly flexible. The longest contiguous stretch (67–73) in the highly flexible regions understandably involves the main loop in the lysozyme molecule, which is known for its conformational flexibility (Nagendra *et al.*, 1998;

Sukumar *et al.*, 1999). The 99–103 stretch belonging to these regions is also known to be susceptible to conformational changes. It is gratifying that the delineation of the molecule into relatively rigid, moderately flexible and highly flexible regions is corroborated by *B* values. The *B* values of main-chain atoms in the three regions averaged over all 20 structures are 10.5, 13.6 and 20.4 Å^2, respectively. The corresponding values for the side-chain atoms are 12.3, 16.4 and 24.9 Å^2, respectively. The same trend is maintained when the averages were taken over individual structures or groups belonging to different crystal systems.

The relatively rigid residues in the polypeptide chain, belonging to α-helices and β-sheets (except 106), form the contiguous core of the molecule. The flexible regions, however, are distributed in a few patches around the rigid core. Nearly half the binding region, made up of residues 34, 35, 37, 52, 57 and 59, belongs to the rigid core. All these residues except 59 belong to the *D*, *E* and *F* subsites. Among the residues that make up these three contiguous sites, accounting for one half of the binding region, only 44 and 114 belong to a flexible region. The residues that make up the other half of the binding region (*A*, *B*, *C* subsites) are 59, 62, 63, 101, 103, 107, 108 and 109. All of them, except 59 which interacts with the acetamido side group of the C ring, belong to a flexible region.

A detailed examination of the conformational variability of the side chains was also carried out. Each side chain was assigned a conformational state (Bhat *et al.*, 1979). For example, the conformation of a lysyl side chain is defined by conformational angles χ^1, χ^2, χ^3 and χ^4. Each of these have allowed values around 60° (g^+), 180° (t) and −60° (g^-). All lysyl side chains with, say, χ^1 between 0 and −120° and each of the remaining three angles between 120° and −120°, were considered to belong to the same conformational state g^-ttt. When the rotation is about a sp^3—sp^2 bond, only two regions are distinct. For example, in the case of histidine, χ^{21} can have values in the neighborhood of 90° (0 to 180°) or −90° (−180 to 0°). In phenylalanine and tyrosine, there is no physical distinction between the two sets of values as the branching involves the same type of atoms; the conformational state is therefore essentially defined by χ^1. The terminal amide or carboxyl groups in asparagine, glutamine, asparatic acid and glutamic acid were assumed to have no conformational preferences (Bhat *et al.*, 1979), although a recent report based on very high resolution structures indicate that they might (Lovell *et al.*, 1999).

Understandably, none of the 11 arginyl residues conserves its side-chain conformation among the 20 crystallographically independent molecules. The same is true for lysine, except that Lys96 has the same energetically favourable g^-ttt conformation in 19 out of 20 molecules; χ^4 is g^- in the low-temperature tetragonal form. At the other extreme, again understandably, the conformations of cysteines are conserved in all the molecules. Nine (27, 37, 39, 46, 59, 65, 74, 93, 113) of the 14 asparaginyl residues have conserved conformations; in Asn106, χ^1 corresponds to g^- in all except one case. Among the seven aspartyl residues, 52 and 66 have conserved conformations, while the conformations of 48 and 119 are conserved except in one case each. The conformations of only Gln57 and Glu35 are conserved among the three glutaminyl and two glutamic acid residues. The sole histidine residue (15) in the structure has the most favoured g^- conformation ($\chi^1 \simeq -60$) in all structures, but χ^{21} has different values. The conformation of seven of the ten seryl residues (24, 36, 50, 60, 81, 91, 100) are conserved. Interestingly, five of these have the g^+ conformation, which facilitates a side-chain···main-chain hydrogen bond (Bhat *et al.*, 1979). Of the seven threonyl residues, however, the conformations of Thr40 and Thr89 are conserved; those of Thr69 and Thr47 are conserved in all except one molecule each. The aromatic residues (three Phe, three Tyr and six Trp) exhibit a high degree of conservation; the only variation exhibited is in χ^{21} of Trp62. Hydrophobic residues with aliphatic side chains (six Val, six Ile, eight Leu), on the other hand, exhibit considerable conformational variability. Only Val2, Val29 and Ile58 have conformations which are conserved in all the molecules. The conformations of Val92 and Ile98 are conserved except in one molecule in each case. Of the two methionyl residues, the conformation of one (Met12) is conserved in all molecules, while that of the other (Met105) is conserved in all but one.

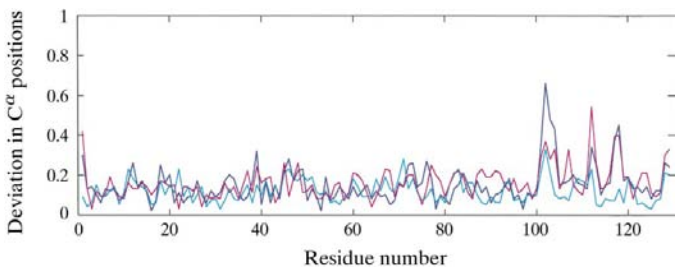

Figure 1
Deviations of C^α positions between native pH 9.5 and pH 4.5 forms (cyan), pH 9.5 and pH 6.5 forms (magenta) and pH 6.5 and pH 4.5 forms (blue).

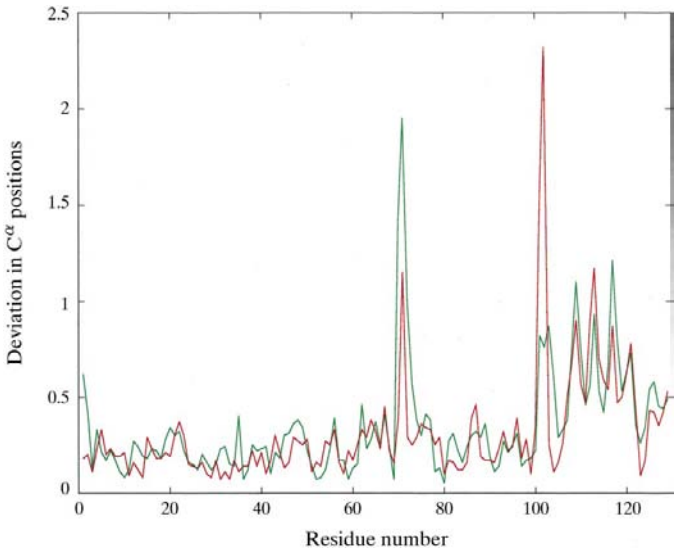

Figure 2
Deviations of C^α positions between the native and low-humidity variant of the pH 9.5 form (red) and the pH 6.5 form (green).

Of the 103 amino-acid residues other than Gly, Ala and Pro in lysozyme, 45 have totally conserved side-chain conformations. Of these, only 21 belong to the rigid region of the molecule, as determined by deviations in C^{α} atoms. Therefore, there is no direct relationship between rigidity in main-chain conformation and the conservation of side-chain conformation. The situation is somewhat different in the binding site. As mentioned earlier, residues 34, 35, 37, 52, 57 and 58 belonging to subsites *D*, *E* and *F* are in the rigid region. All of them have conserved side-chain conformation. In particular, the conformations of the catalytic residues Glu35 and Asp52 are conserved in all 20 molecules. Asn44, which interacts with subsite *E*, and Arg114, which interacts with subsite *F*, belong to a flexible region and have variable side-chain conformations. The residues on the other part of the binding region, namely 101, 103, 107, 108 and 109 on the one lip, and 62 and 63 on the other lip, belong to a flexible region. Residue 107 is Ala. Of the remaining six residues, all but 63 and 108 have variable side-chain conformations. 63 and 108 are tryptophans and a change in their side-chain conformation would necessitate a major structural rearrangement on account of the bulkiness of the indole group. Thus, on the whole, the rigidity of the main chain and the conservation of side-chain conformation correlate well in the binding region. Hence, half of the binding region, made up sub-sites *D*, *E* and *F*, is comparatively rigid, while the other half, comprising sub-sites *A*, *B* and *C*, is comparatively flexible in terms of main-chain as well as side-chain conformation. The possible correlation between the variation in the plasticity of the binding site and the accessibility of different subsites for ligands in the crystal structures was explored. This was performed primarily by superposing the positions of NAM-NAG-NAM and tri-*N*-acetylchitotriose in their tetragonal lysozyme complexes (Kelly *et al.*, 1979; Cheetham *et al.*, 1992) on the lysozyme molecule in the other crystal forms. These calculations indicate that sites *A*, *B*, *C* and *D* are accessible in tetragonal and triclinic crystals. They are inaccessible in the monoclinic crystal. Sites *B* and *D* appear to be definitely inaccessible in the orthorhombic form, while sites *A* and *C* are also likely to be inaccessible. No crystal structure involving sugar binding at subsites *E* and *F* has thus far been determined. An examination of the relevant intermolecular contacts, however, appears to indicate that they are likely to be inaccessible for sugar binding in all the crystal forms, although it is difficult to be definite about the accessibility of a site in the absence of relevant experimental data. In any case, an explanation for the observed relative flexibility of the different subsites could not be obtained on the basis of their accessibility or possible stabilization through intermolecular interactions; it appears to be intrinsic to the binding site.

The main-chain···main-chain hydrogen bonds found in all 20 structures are illustrated in Fig. 4. There are 58 such hydrogen bonds in total, involving 116 donor and acceptor atoms. Of these atoms, 58 belong to the rigid region of the molecule, indicating a weak correlation between the rigidity and the ability to retain hydrogen bonds. The conserved hydrogen bonds involving side chains are listed in Table 5. They involve a total of seven side-chain atoms, six of them belonging to side chains which have the same conformation in all the structures. Understandably, a change in side-chain conformation results in considerable movements of the atoms in them and conservation of conformation is necessary for the retention of hydrogen bonds involving side-chain atoms.

3.3. Invariant water molecules in the hydration shell

Protein molecules, except perhaps those embedded in membranes, are surrounded by interacting water molecules which constitute its hydration shell. The shell is not a contiguous network of water molecules, but is made up of patches (Madhusudan *et al.*, 1993). Most of the ordered water molecules are found to interact with N and O atoms. It is also believed that the hydration shell is almost an integral part of the protein (Sukumar *et al.*, 1999). It is recognized that positions of the attached water molecules are liable to substantially higher variations than the protein atoms themselves. In this context, it is interesting to identify water molecules which have relatively fixed positions with respect to the protein molecule and the possible sites of these invariant water molecules. Admittedly, the number of molecules located in the electron-density map depends upon the criteria for location adopted by the investigator and the resolution of the map. However, the smaller subset of molecules with fixed positions are likely to be located in all cases where the resolution is reasonably high as in the structures under consideration. Furthermore, the number of molecules identified as invariant is likely to decrease as the number of independent structures considered increases (Sadasivan *et al.*, 1998). However, it is felt that the present analysis involving as many as 20 crystallographically independent copies of the protein molecule and its hydration shell, the structures of which have been determined under different environmental condi-

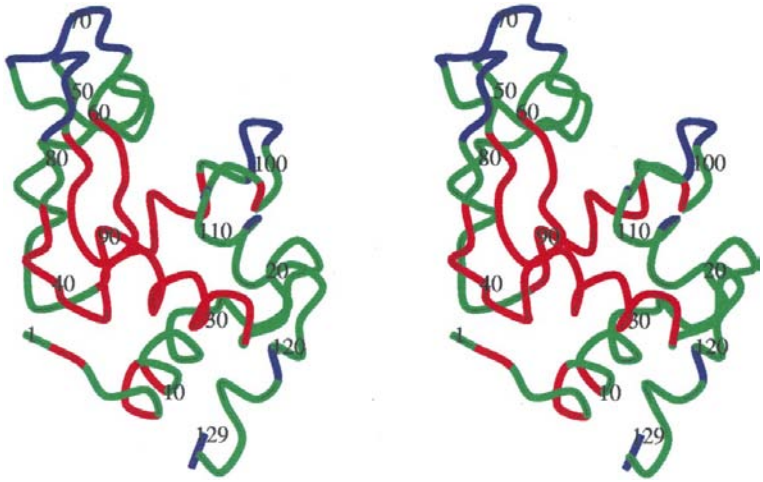

Figure 3
Stereo diagram illustrating the relatively rigid (red), moderately flexible (green) and highly flexible (blue) regions in the lysozyme molecule.

Table 5
Conserved hydrogen bonds involving side-chain atoms.

Donor atom	Acceptor atom
Lys1 N	Thr40 OG1
Trp28 NE1	Tyr23 O
Gln41 N	Asn39 OD1
Thr51 N	Ser60 OG
Lys96 NZ	His15 O
Met105 N	Tyr23 OH
Trp108 NE1	Leu56 O

Table 6
Invariant water molecules in the hydration shell of HEW lysozyme.

The numbering of water molecules corresponds to that in the triclinic form (Walsh *et al.*, 1998).

Water	Protein atoms with which the water molecule interacts in all the structures	Additional protein atoms with which the water molecule interacts in at least one structure
W1001	53 O, 56N, Ser91 OG	55 N, 57N
W1002	49 O, Thr51 OG1	Arg68 NE, Arg45 NE, 51N, Asp66 OD2, Thr69 OG1, Arg45 NH1, Arg68 NH1, Arg68 NH2, Arg45 NH2
W1007	87 O, 91N, Ser91 OG,	83 O, Ser85 OG, 90N, 82 O, 88 O
W1013	22 O, 24N	Ser24 OG, Asn27 ND2
W1016	84 O	1 N, Thr40 OG1, Gln41 NE2, 86N, Ser86 OG
W1020	90 O	94 N
W1030	Asp52 OD2, Gln57 OE1	Asn44 OD1, Glu35 OE1, Asn44 ND2

tions, is likely to lead to a meaningful identification of invariant water molecules. The fact that a series of calculations with different subsets of 18 molecules and their hydration shells as the database led to nearly the same set of invariant water molecules reinforces this conclusion.

For identifying invariance, all the remaining 19 molecules along with their hydration shells were superposed on the molecule in the triclinic form, a structure determined at a high resolution. A water molecule is considered invariant if it has at least one common protein contact (3.6 Å) with a protein O or N atom in all protein molecules and if the distance between equivalent water molecules in every pair of structures is less than 1.8 Å. The seven invariant water molecules so identified and their interactions are listed in Table 6. It is interesting to enquire into the roles of these invariant molecules.

W1001 and W1007 interact with three protein atoms each in all 20 molecules. One interaction, that involving Ser91 OG, is common to the two sets. Several other contacts occur in one or more molecules. The interactions that occur in the molecule in the triclinic form are illustrated in Fig. 5. Clearly, the two water molecules have an important structural role in holding together the main β-sheet (residues 52–63) and the 80–84 and 88–101 helices. Interestingly, water-mediated interactions involving these three regions exist even in a crystal form of lysozyme with an abnormally low solvent content of 9%, except that the interactions are mediated by a single water molecule instead of the two water molecules in the normally hydrated forms (Nagendra *et al.*, 1998). W1002 lends additional stability to the main β-sheet and also connects it to the main loop (65–73). W1013 interconnects the N-terminal region of the 25–36 helix and the interesting structural feature involving two consecutive β-turns in the 19–23 stretch. W1020 is attached to the 88–101 helix, but its structural role is not immediately obvious. W1030 is the only invariant water molecule in the active site. It forms part of a network that connects the two catalytic and neighbouring residues.

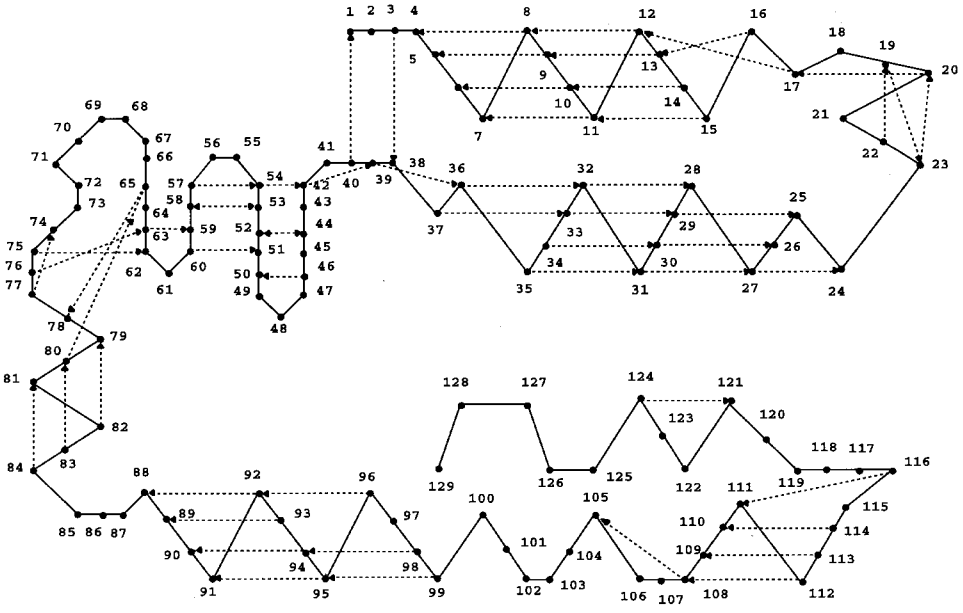

Figure 4
Invariant main-chain···main-chain hydrogen bonds in the lysozyme molecule. The criteria prescribed by Baker & Hubbard (1984) as modified by Kodandapani *et al.* (1990) and followed in the work on low-humidity monoclinic lysozyme by Madhusudan *et al.* (1993) were used for identifying hydrogen bonds between protein atoms.

3.4. Accessibility, hydration and mobility

Accessibility and hydration are obviously related. The relation between accessibility and mobility is yet to be fully explored.

The lysozyme molecule contains a total of 378 N and O atoms. In the 20 molecules considered, the number of these atoms which are totally inaccessible (zero accessibility) varies between 130 and 148.

Table 7
N and O atoms hydrated in all structures.

The number of water molecules attached to each atom in the 20 structures is also given.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1 N	2	2	2	2	2	2	3	3	2	3	3	2	3	3	1	3	2	3	2	3
22 O	2	2	5	3	2	1	2	2	3	2	1	2	2	1	1	2	2	2	2	3
24 N	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	2	1	2	1	1
27 OD1	1	1	1	1	1	1	2	1	1	1	1	1	2	2	1	1	1	1	1	1
52 OD2	3	3	2	2	3	2	3	1	1	1	2	1	1	1	1	1	1	1	2	1
53 O	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
56 N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
65 OD1	3	2	3	4	1	2	4	2	1	1	3	1	2	2	3	3	4	5	2	3
83 O	1	2	1	1	1	1	2	1	1	1	1	1	1	1	1	2	1	1	1	1
84 O	1	1	1	1	1	2	2	2	1	1	2	1	2	1	1	2	1	1	1	1
86 OG	1	3	2	2	1	1	3	3	1	4	4	1	1	4	1	2	2	2	4	2
87 O	2	2	2	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
90 O	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
91 N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
91 OG	3	3	3	3	2	2	3	3	2	2	2	3	3	2	3	3	3	3	3	2

94 have zero accessibility in all molecules. Understandably, only three of these belong to side chains, the remainder being main-chain N and O atoms. Of these 91, 47 are found in the relatively rigid region of the molecule. There are 19 residues in which the main-chain N as well as O atoms are totally inaccessible in all 20 molecules, 12 of them belonging to the relatively rigid region. Considering that this region accounts for only about a third of the total number of amino-acid residues, the above numbers indicate a clear correlation between accessibility of the main-chain polar atoms and mobility, although it is not very strong. Accessibility of each residue as a whole was also considered. There are 22 residues for which less than 5% of the total surface area is accessible in all 20 structures. They are Leu8, Ala9, Met12, Leu17, Trp28, Val29, Cys30, Ala31, Ala32, Thr40, Gly54, Ile55, Leu56, Ile58, Cys80, Leu83, Val92, Cys94, Ala95, Met105, Trp108 and Cys115. Understandably, almost all of them are hydrophobic. 14 of the 22 belong to the rigid region of the molecule. This again indicates a correlation, albeit moderate, between rigidity and non-accessibility. Among the 22, five are alanine or glycine, for which side-chain conformation is irrelevant. Of the remaining 17, ten have conserved side-chain conformation, indicating no strong correlation between non-accessibility and conservation of side-chain conformation.

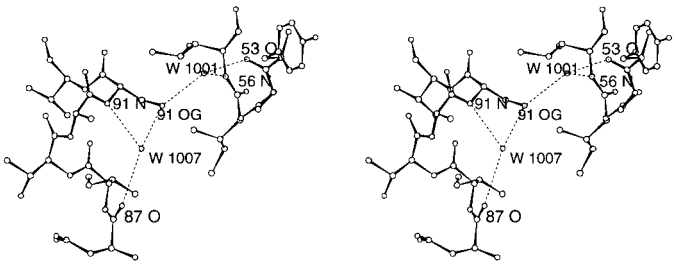

Figure 5
Stereo diagram illustrating the interactions involving the invariant water molecules W1001 and W1007.

When dealing with the relation between accessibility and hydration, the reduction in accessibility caused by crystal packing also needs to be taken into account. The relation between solvent content and the loss of accessible surface area of the protein owing to crystal packing is illustrated in Fig. 6. As is to be expected, accessible surface area reduces when the solvent content decreases. However, the magnitude of reduction depends on the crystal system or, in other words, the nature of molecular packing. For example, the accessible surface area lost for a given solvent content is much higher in orthorhombic lysozyme than in tetragonal lysozyme.

Crystal packing naturally reduces atomic accessibility in many cases, but it is altogether abolished only in a very few

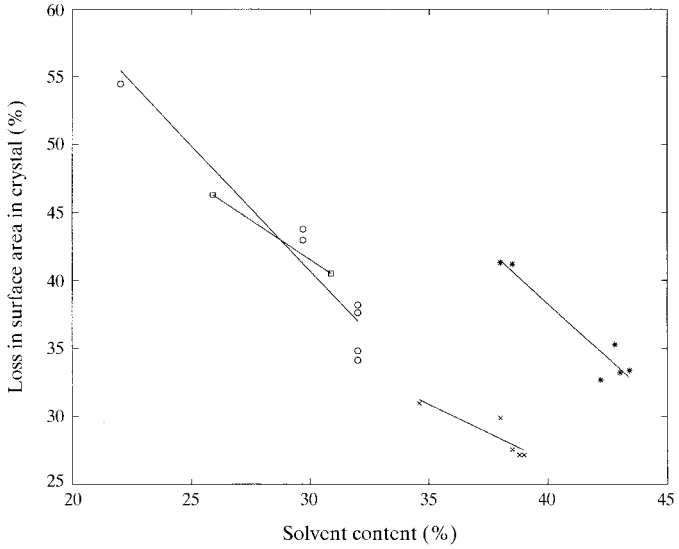

Figure 6
Relation between solvent content and the loss of accessible surface area owing to crystal packing in tetragonal (cross), monoclinic (circle), orthorhombic (star) and triclinic (square) forms of lysozyme.

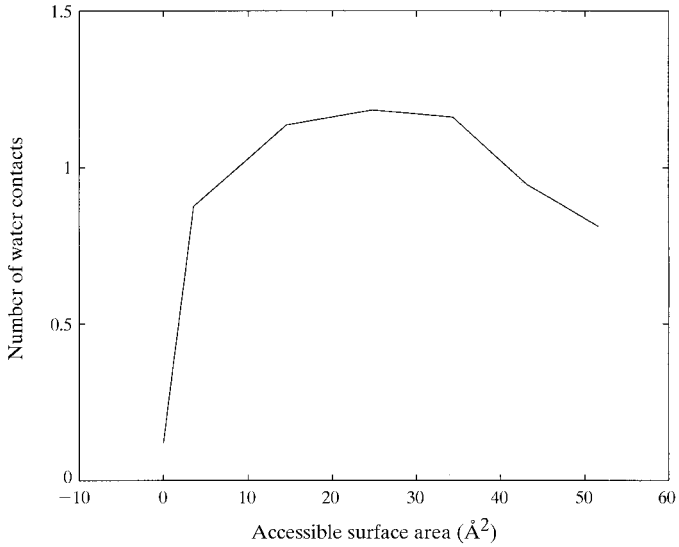

Figure 7
Relation between accessible atomic surface area and number of water contacts of N and O atoms. The area and contacts have been averaged over all the concerned atoms in the 20 structures in a given bin.

instances. For example, 230–248 N or O atoms have non-zero accessibility in the 20 independent lysozyme molecules. When crystal packing is also taken into account, these numbers become 219–239. The numbers of hydrated atoms are generally still lower. The actual number would of course depend upon the number of ordered water molecules located in the crystal structure concerned. Thus, the number of hydrated polar atoms in low-temperature triclinic lysozyme, with a hydration shell made up of 249 water molecules, is 224, while this number is 141 in pH 4.5 orthorhombic lysozyme, in which the number of water molecules is the lowest at 92. The number varies between 135 and 201 in the other structures. The number of water molecules interacting with a hydrated polar atom varies between one and five. The number of accessible polar atoms which become totally inaccessible (zero accessibility) owing to crystal packing in one crystal form or another is only 54. Many of them are common to several structures.

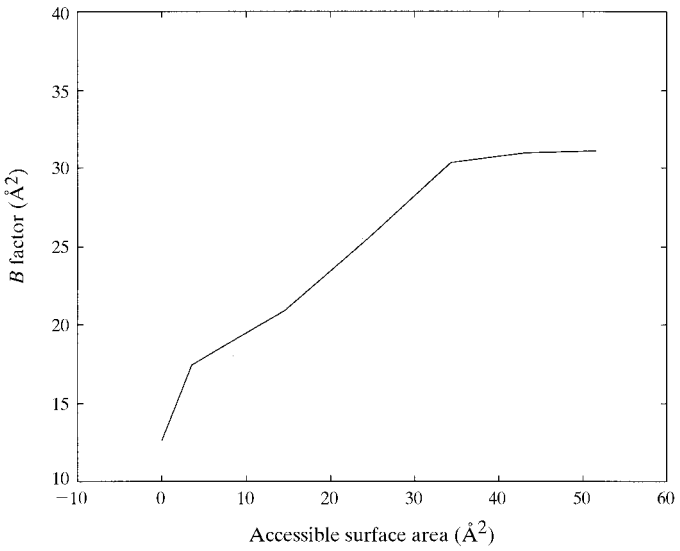

Figure 8
Relation between accessible atomic surface area and displacement parameters of N and O atoms. The averaging was performed in the same way as in Fig. 7.

Thus, the effect of crystal packing on hydration does not appear to be large. The relationship between accessibility and hydration of polar atoms, as obtained from the average values calculated from the 20 structures, is illustrated in Fig. 7. The level of hydration increases rapidly with accessible surface area, but stops increasing at about 15 Å^2. Beyond 30 Å^2, the level actually begins to decrease. The relation between accessible surface area and average displacement parameter for the same structures, shown in Fig. 8, provides a partial explanation for this behaviour. Highly accessible atoms are most often likely to belong to flexible side chains with static and dynamic disorder. The movement of the side chains is likely to render the water molecules attached to them less visible in electron-density maps. Furthermore, water molecules which are disordered as a consequence of disorder in the protein atoms to which they are attached are less likely to be located in electron-density maps. The behaviour of the displacement parameter of course does not completely explain the relation between hydration and accessibility. Fig. 7 clearly shows that a water molecule can interact with the protein atom even when only a small area on its surface is accessible. However, it would appear that on average it is more difficult for a second water molecule to interact with an already hydrated atom, probably because of steric problems. The fact that the relationship illustrated in Fig. 7 is qualitatively similar to that in a set of ribonuclease A crystals (Sadasivan *et al.*, 1998) seems to indicate that this inference is of general validity.

Two cavities (Connolly, 1993) occur in all 20 structures. Another occurs in 19 structures and yet another in 18 structures. Of these four, three are hydrophobic in nature and are apparently empty. Only one, occurring in all structures, is lined with polar atoms (53 O, 56 N and 91 OG). The invariant water molecule W1001 is within this cavity (Fig. 9). Indeed, W1001 appears to be the only buried water molecule in the structure.

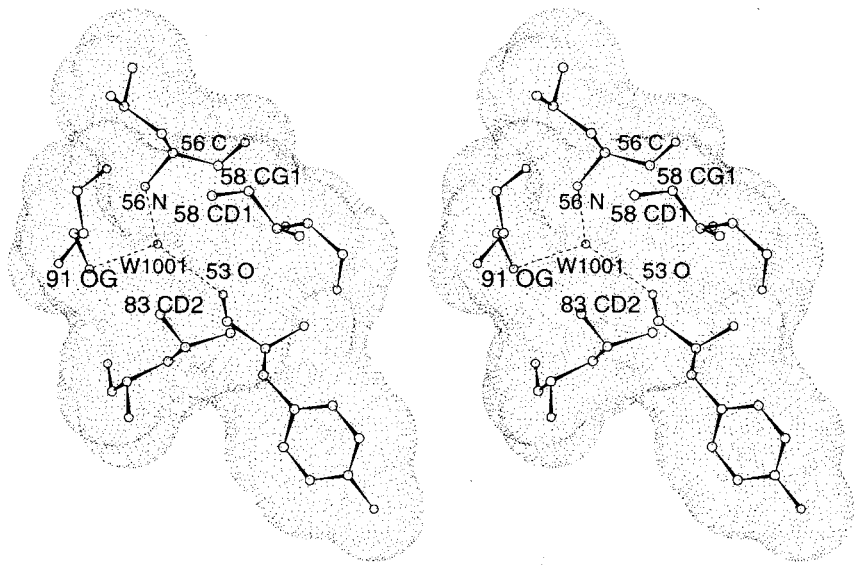

Figure 9
Stereo diagram illustrating the environment of the buried water molecule W1001.

Although the number of water molecules that surround and interact with the protein molecule has similar values in most of the 20 structures, there is considerable variation in the atoms hydrated. In fact, only 15 N and O atoms are hydrated in all 20 molecules. The level of hydration in them is given in Table 7. Interestingly, 13 of these 15 are involved in interactions with the seven invariant water molecules (Table 6). As mentioned earlier, five of these seven are involved in water bridges that stabilize the structure and, as illustrated in Fig. 10, one (W1030) is part of a network of water molecules connecting the catalytic residues Glu35 and Asp52. W1030 disappears when the water content is

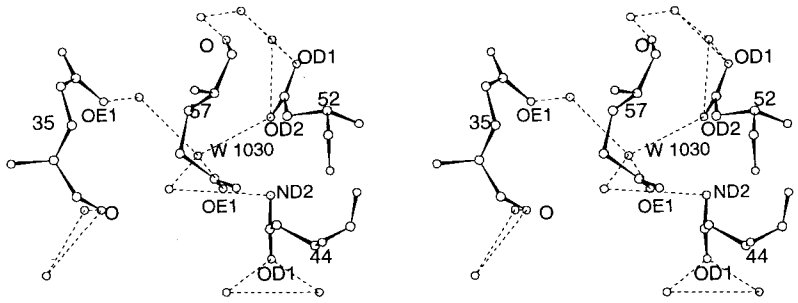

Figure 10
Stereo diagram illustrating the water network involving the invariant water molecule W1030 connecting the catalytic residues Glu35 and Asp52 in the triclinic form.

reduced to well below 20%, the amount of water required for activity (Rupley *et al.*, 1983; Rupley & Careri, 1991), as in the 9% solvent-content form (Nagendra *et al.*, 1998). In this form, the mutual disposition of the catalytic residues deviates from that found in all other structures primarily on account of a change in the conformation of Glu35. Thus, invariance in protein hydration is not without structural and functional significance.

X-ray intensity data were collected using the National Area Detector facility at the Indian Institute of Science supported by the Departments of Science & Technology (DST) and Biotechnology (DBT), Government of India. Facilities at the Indian Institute of Science Supercomputer Education and Research Centre of the Institute and the Interactive graphic facility (supported by the DBT) were used for computation and model building. The work has been carried out under a DST supported research project.

References

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.
Brünger, A. T. (1992). *X-PLOR. Version 3.1. A System for X-ray Crystallography and NMR.* Yale University, Connecticut, USA.
Cheetham, J. C., Artymiuk, P. J. & Phillips, D. C. (1992). *J. Mol. Biol.* **224**, 613–628.
Connolly, M. L. (1993). *J. Mol. Graph.* **11**, 139–141.
Cruickshank, D. W. J. (1999). *Acta Cryst.* D**55**, 583–601.
Dong, J., Boggon, T. J., Chayen, N. E., Raftery, J., Bi, R. & Helliwell, J. R. (1999). *Acta Cryst.* D**55**, 745–752.
Harata, K. (1994). *Acta Cryst.* D**50**, 250–257.
Hubbard, S. J. & Argos, P. (1995). *Protein Eng.* **8**, 1011–1015.
Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 916–924.
Kelly, J. A., Sielecki, A. R., Sykes, B. D., James, M. N. G. & Phillips, D. C. (1979). *Nature (London)*, **282**, 875–878.
Kodandapani, R., Suresh, C. G. & Vijayan, M. (1990). *J. Biol. Chem.* **265**, 16126–16131.
Kramer, R. Z. & Berman, H. M. (1998). *J. Biomol. Struct. Dyn.* **16**, 367–380.
Kundrot, C. E. & Richards, F. M. (1987). *J. Mol. Biol.* **193**, 157–170.
Kurinov, I. V. & Harrison, R. W. (1995). *Acta Cryst.* D**51**, 98–109.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lovell, S. C., Word, J. M., Richardson, J. S. & Richardson, D. C. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 400–405.
Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.
Madhusudan, Kodandapani, R. & Vijayan, M. (1993). *Acta Cryst.* D**49**, 234–245.
Madhusudan & Vijayan, M. (1991). *Curr. Sci.* **60**, 165–170.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Nagendra, H. G., Sudarsanakumar, C. & Vijayan, M. (1995). *Acta Cryst.* D**51**, 390–392.
Nagendra, H. G., Sudarsanakumar, C. & Vijayan, M. (1996). *Acta Cryst.* D**52**, 1067–1074.
Nagendra, H. G., Sukumar, N. & Vijayan, M. (1998). *Proteins Struct. Funct. Genet.* **32**, 229–240.
Oki, H., Matsuura, Y., Komtsu, H. & Chernov, A. A. (1999). *Acta Cryst.* D**55**, 114–121.
Radha Kishan, K. V., Chandra, N. R., Sudarsanakumar, C., Suguna, K. & Vijayan, M. (1995). *Acta Cryst.* D**51**, 703–710.
Rockland, L. B. (1960). *Anal. Chem.* **32**, 1375–1376.
Rossmann, M. G. & Argos, P. (1975). *J. Biol. Chem.* **250**, 7525–7532.
Rupley, J. A. & Careri, G. (1991). *Adv. Protein Chem.* **41**, 37–172.
Rupley, J. A., Gratton, E. & Careri, G. (1983). *Trends Biochem. Sci.* **8**, 18–22.
Sadasivan, C., Nagendra, H. G. & Vijayan, M. (1998). *Acta Cryst.* D**54**, 1343–1352.
Saenger, W. (1987). *Annu. Rev. Biophys. Chem.* **16**, 93–114.
Salunke, D. M., Veerapandian, B., Kodandapani, R. & Vijayan, M. (1985). *Acta Cryst.* B**41**, 431–436.
Steinrauf, L. K. (1998). *Acta Cryst.* D**54**, 767–779.
Stryer, L. (1981). *Biochemistry.* San Francisco: W. H. Freeman & Co.
Sukumar, N., Biswal, B. K. & Vijayan, M. (1999). *Acta Cryst.* D**55**, 934–937.
Sundaralingam, M. & Shekarudu, Y.C. (1989). *Science*, **244**, 1333–1337.
Teeter, M. M. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 6014–6018.
Vaney, M. C., Maignan, S., Ries-Kautt, M. & Ducruix, A. (1996). *Acta Cryst.* D**52**, 505–517.
Vijayan, M. (1980). *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkateshan, pp 19.01–19.26. Bangalore: Indian Academy of Sciences.
Walsh, M. A., Schneider, T. R., Seiker, L. C., Dauter, Z., Lamzin, V. S & Wilson, K. S. (1998). *Acta Cryst.* D**54**, 522–546.
Weast, R. C. & Astle, M. J. (1980/1981). Editors. *CRC Handbook of Chemistry & Physics*, p. E46. Boca Raton: CRC Press.