Acta Cryst. (1998). D**54**, 522–546

Refinement of Triclinic Hen Egg-White Lysozyme at Atomic Resolution

Martin A. Walsh,[a]*† Thomas R. Schneider,[a]‡ Larry C. Sieker,[b] Zbigniew Dauter,[a,c] Victor S. Lamzin[a] and Keith S. Wilson[a,c]

[a]*European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany,* [b]*Department of Biological Structures, SM-20, University of Washington, Seattle, WA 98195, USA, and* [c]*Department of Chemistry, University of York, Heslington, York Y01 5DD, England. E-mail: walsh@anl.gov*

(*Received 15 July 1997; accepted 15 October 1997*)

Abstract

X-ray diffraction data have been collected at both low (120 K) and room temperature from triclinic crystals of hen egg-white lysozyme to 0.925 and 0.950 Å resolution, respectively, using synchrotron radiation. Data from one crystal were sufficient for the low-temperature study, whereas three crystals were required at room temperature. Refinement was carried out using the programs *PROLSQ*, *ARP* and *SHELXL* to give final conventional R factors of 8.98 and 10.48% for data with $F > 4\sigma(F)$ for the low- and room-temperature structures, respectively. The estimated r.m.s. coordinate error is 0.032 Å for protein atoms, 0.050 Å for all atoms in the low-temperature study, and 0.038 Å for protein atoms and 0.049 Å for all atoms in the room-temperature case, as estimated from inversion of the blocked least-squares matrix. The low-temperature study revealed that the side chains of 24 amino acids had multiple conformations. A total of 250 waters, six nitrate ions and three acetate ions, two of which were modelled with alternate orientations were located in the electron-density maps. Three sections of the main chain were modelled in alternate conformations. The room-temperature study produced a model with multiple conformations for eight side chains and a total of 139 water molecules, six nitrate but no acetate ions. The occupancies of the water molecules were refined in both structures and this step was shown to be meaningful when assessed by use of the free R factor. A detailed description and comparison of the structures is made with reference to the previously reported structure refined at 2.0 Å resolution.

1. Abbreviations

HEWL, hen egg-white lysozyme; LTL, low-temperature (120 K) lysozyme structure; RTL, room-temperature (295 K) lysozyme structure; F_o, F_c, observed and calculated structure-factor amplitudes, respectively; r.m.s., root-mean-square; PDB, Brookhaven Protein Data Bank; *ARP*, Automatic Refinement Procedure; ADP's, anisotropic displacement parameters; B_{eq}, B factor equivalent to a given set of ADP's; U_{eq}, isotropic mean-square displacement equivalent to a set of ADP's, $B_{eq} = 8\pi^2 U_{eq}$; Ace, acetate ion; DB plot, plot of $(3F_o - 2F_c, \alpha_c)$ density interpolated at atomic centres as a function of B_{eq}.

2. Introduction

Hen egg-white lysozyme (HEWL) was one of the first proteins to be crystallized (Steinrauf, 1959) and has been the subject of numerous crystallographic studies (see Vaney *et al.*, 1996, and references therein). This stems in part from the material being readily available, stable, quite soluble and easily crystallized in several different space groups (Steinrauf, 1959). The crystal structure of lysozyme was first determined in the tetragonal form (Blake *et al.*, 1965). Since then, crystal structures have been examined by X-ray diffraction of the monoclinic, triclinic and orthorhombic forms (Vaney *et al.*, 1996). Other crystal structure investigations have been carried out using neutron diffraction (Lehmann *et al.*, 1985) and cryogenic temperatures (Kurinov & Harrison, 1995; Young *et al.*, 1994). In addition, crystal structures of HEWL complexes with various substrate analogues (Imoto *et al.*, 1972; Kurachi *et al.*, 1976; Maenaka *et al.*, 1995; Strynadka & James, 1991) and metal ions (Kurachi *et al.*, 1975) have been determined. Lysozyme has also become one of the subjects for study of the dynamics of crystal growth, and this aspect has been studied under microgravity conditions on space flights (Vaney *et al.*, 1996). Besides investigations of the crystal structure of native HEWL, complexes with HEWL, perturbations of hydration in the crystal (Kodandapani *et al.*, 1990; Nagendra *et al.*, 1995), deuterium exchange in the crystal (Mason *et al.*, 1984) and antigenic properties of the HEWL (Chacko *et al.*, 1996; Cohen *et al.*, 1996) have been investigated. There have been numerous studies of its biochemical, biophysical and enzymatic properties (for a review see Mckenzie & White, 1991). Needless to

† To whom correspondence should be addressed at Center for Mechanistic Biology and Biotechnology, Bldg. 202, Argonne National Laboratory, Argonne IL 60439, USA.
‡ Present address: Goettingen University, Institute for Inorganic Chemistry, Tammanstrasse 4, 37077 Goettingen, Germany.

ISSN 0907-4449

say HEWL is one of the most intensively studied proteins and with the advent of genetic engineering, studies of site-specific mutants have been analysed to assess the roles played by specific amino-acid residues on structure and function (Hadfield *et al.*, 1994; Hashimoto *et al.*, 1996; Maenaka *et al.*, 1995; Wilson *et al.*, 1992).

For the majority of protein crystallographic studies, the crystallographer is at the mercy of the protein crystal which in general diffracts weakly. This is attributed to the large size and flexibility of protein molecules and the high solvent content of their crystals. Until recently, only a few small proteins which produced well ordered crystals with a low solvent content were known to diffract to atomic resolution (<1.2 Å). Important developments in recent years in crystallographic methods, detector technology and the availability of high-intensity synchrotron sources have led to a large increase in the number of protein structures determined at atomic resolution (Dauter *et al.*, 1995). At atomic resolution the number of measured reflections substantially exceeds the number of parameters to be refined. The far greater ratio of experimental observations to parameters allows, at least for the well ordered regions of the protein molecule, the refinement of the model without any stereochemical restraints. However, protein structures contain highly flexible regions which need to be restrained even at high resolution and in practice some degree of geometric restraints are applied during refinement of the protein model.

Structures at atomic resolution can be refined with high accuracy, which until recently were thought only attainable in the field of small-molecule crystallography. The increasing number of protein structures being reported at atomic resolution are providing us with a far more accurate picture of the geometric and conformational properties of proteins in general. Fast, efficient and reliable protocols are still being formulated, and/or improved. The present study enforces and extends previous protocols (Sevcik *et al.*, 1996; Sheldrick & Schneider, 1997) and provides further data for inclusion in an improved library of stereochemical target values for use in the refinement and validation of structures at lower resolution. The latter point has been the focus of attention of an EC network on the validation of three-dimensional structures. The low-temperature lysozyme structure reported here was analysed along with seven other structures refined at atomic resolution using four different validation tools, *PROCHECK* (Laskowski *et al.*, 1993), *PROVE* (Pontius *et al.*, 1996), *SQUID* (Oldfield, 1992) and *WHATCHECK* (Hooft *et al.*, 1996). The results of this work are presented in a separate paper (Wilson *et al.*, 1998).

The triclinic crystal form of HEWL is the optimal one to study at very high resolution. This particular crystal form has only one molecule in the unit cell. The molecules pack in such a way that the water content is minimal causing the crystal to have one of the lowest unit-cell volume to molecular weight ratios observed for a 14 kDa protein. The very low unit-cell volume coupled with the high crystal quality (*i.e.* low mosaic spread of the crystal with a relatively rigid protein molecule) allows the recording of diffraction data to at least 1.0 Å resolution using a conventional sealed X-ray tube. A previous crystal structure investigation by Jensen, Sieker *et al.* (Hodsdon *et al.*, 1990; Ramanadham *et al.*, 1989, 1990) aimed at elucidating the fine details of the tertiary structure was reported at 2.0 Å resolution, although diffraction data were observed to 1.0 Å. We report the *P*1 crystal structure to atomic resolution at 120 K (0.925 Å) and at 295 K (0.950 Å), abbreviated hereafter as LTL and RTL, respectively. The data have allowed a precise description of the structure in addition to highlighting the advantages of X-ray data collection using synchrotron radiation. Furthermore, comparison of the two structures at room and cryogenic temperatures provides additional support for the advantages of using cryocrystallography as a routine method in macromolecular X-ray data collection.

3. Materials and methods

3.1. *Crystallization, data collection and processing*

Crystallization of HEWL in the triclinic form (*P*1) is not generally straightforward. The *P*1 crystals develop from a 1% protein solution buffered at pH 4.5 with 100 m*M* sodium acetate in the presence of 2%(*w*/*v*) sodium nitrate at 296 K. Without seeding most of the crystals are in the monoclinic form, but about 10% of the crystals appear as the *P*1 triclinic crystal form (Sieker, 1988). Normally, monoclinic crystals appear first as thick lathe-shaped needles whereas *P*1 crystals grow later as chunky prisms. To obtain *P*1 crystals routinely in good yield, seeding the crystallizing solution with a *P*1 crystal is necessary. Seed crystals for *P*1 were obtained by the minibatch technique by adding sufficient solid sodium nitrate to make the 1% protein solution 2–2.2% in sodium nitrate and allowing the experiment to proceed until *P*1 crystals grow. Diffraction-quality *P*1 crystals were grown by the same procedure however in this case washed *P*1 seed crystals were introduced into 100 µl minibatch solutions immediately after they were set up. The crystallizing solutions were made by mixing debris-free 50 µl of 2% (20 mg ml^{-1}) protein solution with 50 µl of buffered 4% sodium nitrate. The experiments were covered and allowed to stand at 296 K.

Data were collected using synchrotron radiation from the EMBL beamlines X11 (low temperature) and BW7B (room temperature) at the DORIS storage ring, DESY, Hamburg on a 300 mm MAR Research imaging-plate scanner. For the low-temperature experiment a singe crystal suitable for X-ray diffraction was transferred to the reservoir solution to which glycerol

Table 1. *Data collection and processing statistics*

Low-temperature 120 K data				
No. of crystals	1			
Beamline	X11			
Wavelength (Å)	0.927			
Space group	*P*1			
Unit-cell dimensions (Å, °)	a = 26.65 b = 30.80 c = 33.63, α = 89.30 β = 107.40 γ = 112.20			
Resolution (Å)	20–0.925			
$R_{\rm merge}$† (I)(%)	2.8 (16.8)‡			
$\langle I\rangle/\langle\sigma(I)\rangle$	29.1 (4.9)			
Completeness (%)	90.1 (78)			
No. of observations	232156			
No. of unique reflections	58373			
Room-temperature 295 K data				
No. of crystals	3			
Beamline	BW7B			
Wavelength (Å)	0.862			
Space group	*P*1			
Unit-cell dimensions (Å, °)	a = 27.24 b = 31.87 c = 34.23, α = 88.52 β = 108.53 γ = 111.89			
	All crystals	Crystals 1+2	Crystal 3	Crystal 3
Resolution (Å)	21–0.95	2.5–0.95	5.0–1.50	21–2.45
$R_{\rm merge}$† (I) (%)	6.0 (27.3)	7.3 (27.2)	5.1 (6.0)	4.3 (5.9)
$\langle I\rangle/\langle\sigma(I)\rangle$	14.4 (4.7)	17.9 (3.5)	22.9 (20.6)	25.7 (21.0)
Completeness (%)	96.5 (94.9)	89.6 (94.7)	84.6 (95.5)	93.4 (98.2)
No. of observations	289540	268930	67251	12420
No. of unique reflections	61265	53773	13084	2937

† $R_{\rm merge} = \sum_{hl}\sum_{i=1}^{N}|\langle I^{hkl}\rangle - I_i^{hkl}|/\sum_{hkl}\sum_{i=1}^{N} I_i^{hkl}$. ‡ Values in parentheses are for the highest resolution shell.

[25%(*v*/*v*)] was added, mounted in a fibre loop (Teng, 1990) and immediately flash frozen at 120 K under a nitrogen vapour stream using an Oxford Cryosystems Cryostream (Cosier & Glazer, 1986). Three sets of data were collected at resolution cut-offs of 0.92, 1.4 and 2.4 Å to avoid saturation of the high-intensity reflections. The data were integrated, scaled and merged using the *DENZO* and *SCALEPACK* programs (Otwinowski & Minor, 1997). The overall $R_{\rm merge}$ on intensities for all data was 2.8% and the data were 90.1% complete. For the room-temperature experiment data from three crystals were combined. The crystals were mounted in glass capillaries using standard procedures (Ducruix & Giegé, 1992). Three sets of data were collected at resolution cut-offs of 0.93, 1.5 and 2.45 Å. Data from two crystals were used to collect a complete high-resolution set with an $R_{\rm merge}$ of 7.3%. The medium- and low-resolution data were collected using a third crystal. Scaling and merging resulted in an $R_{\rm merge}$ of 6.0% for all data in the resolution range 21–0.95 Å. The data were 96.5% complete.

Data-collection parameters and processing for room- and low-temperature data are summarized in Table 1. The percentage completeness of the data is shown in Fig. 1(*a*). The incompleteness in the low-temperature data was due to the blind region which is unavoidable with *P*1 symmetry when oscillating around a single axis. The lower completeness in the data collected at 120 K was due to time restraints which prevented collection of an additional data set with the crystal re-oriented to minimize the blind region. The $R_{\rm merge}$ for symmetry-equivalent reflections as a function of resolution is shown in Fig. 1(*b*). Overall *B* factors of 5.5 and 7.1 Å^2 for LTL and RTL were estimated from their respective Wilson plots (Wilson, 1942).

3.2. *Refinement protocol*

3.2.1. *Overall strategy*. Refinement was carried out against 97% of the measured data. The remaining 3% (1776, reflections for LTL and 1870 for RTL) were randomly excluded from the data set and used as a cross-validation test using the free *R* factor (Brünger, 1992). In the final steps of the refinement all data were included. Atomic coordinates and isotropic displacement parameters were refined by restrained least-squares minimization using the *CCP*4 (Collaborative Computational Project, Number 4, 1994) fast Fourier version of *PROLSQ* (Hendrickson & Konnert, 1980). The starting model for the refinements was the coordinate set 2LZT from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977), which had been refined against data to 2.0 Å resolution (Ramanadham *et al.*, 1990). All solvent molecules were removed before the refinements commenced. A total of 50 cycles were carried out. 20 further cycles were then carried out in conjunction with an automated refinement procedure (*ARP*) (Lamzin & Wilson, 1993, 1997) which models and updates solvent structure after each refinement cycle. Standard restraints as implemented in the *CCP*4 version of

PROLSQ were applied to the geometrical parameters, except in the case of displacement parameters which were assigned weaker restraints of 4 (6) and 5 Å^2 (8 Å^2) for 1–2 and 1–3 neighbour atoms (bond and angle), respectively, the values in brackets being for the side-chain equivalent. A 2:1 weight ratio of the X-ray/geometry terms was applied throughout this first stage of refinement. X-ray data in the 20–0.925 Å (LTL) and 20–0.95 Å (RTL) resolution range were used with no σ cut-off on amplitudes. No manual intervention was carried out in the isotropic refinement stage. Before commencement of anisotropic refinement with *SHELXL* (Sheldrick & Schneider, 1997) an initial ten cycles of isotropic refinement using *SHELXL* were carried out. Disordered residues were modelled and atom occupancies refined using one common occupancy value for the side-chain atoms. H atoms were not refined but included at calculated positions, introducing no additional parameters into the refinement process. Anisotropic refinement was carried out in a series of rounds with *SHELXL*. Each round consisted of ten conjugate-gradient cycles and where new atoms were introduced into the model they were always refined isotropically for two cycles before making them anisotropic. After each of these refinement rounds the model was inspected manually with $3F_o - 2F_c$ and $F_o - F_c$ Fourier maps using the program *O* (Jones *et al.*, 1991). The re-building was aided by systematically checking all electron-density peaks greater than 5σ in the $F_o - F_c$ Fourier maps.

Once most peaks greater than 5σ had been accounted for, the occupancies of the water molecules were refined. The free *R* was used to assess the validity of the resultant models. Identification of modelled water sites, with poor agreement to the experimental data was aided by use of plots of ($3F_o - 2F_c$, α_c) density interpolated at the atomic centres as a function of atomic B_{eq} (DB plots) (Sevcik *et al.*, 1996). The final stage of the protocol involved refinement of the structures against all the X-ray diffraction data. The scheme of the refinement protocol is shown in Fig. 2.

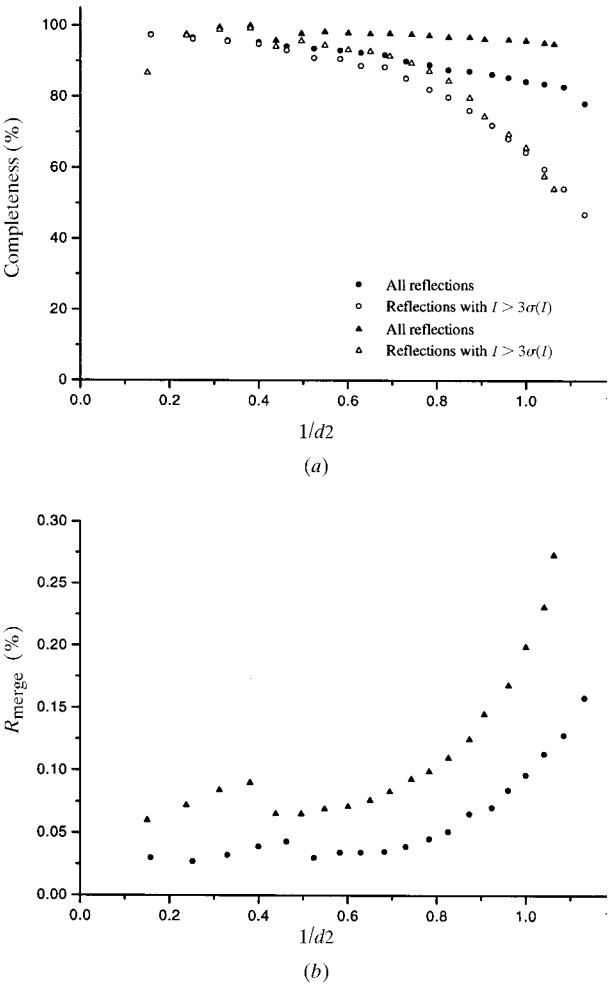


Fig. 1. The X-ray data for LTL (circles) and RTL (triangles). (*a*) The percentage completeness of the data as a function of resolution. (*b*) The merging $R(I)$ for symmetry-related reflections, $R_{merge} = \sum_{hkl} \sum_{i=1}^{N} |\langle I^{hkl} \rangle - I_i^{hkl}| / \sum_{hk} \sum_{i=1}^{N} I_i^{hkl}$ as a function of resolution.

4. Results

4.1. *Details of the refinement*

4.1.1. *Low-temperature structure*. The first stage of refinement using *PROLSQ* in conjunction with *ARP* converged to an *R* factor of 17.9% and a free *R* of 21.5%. Ten cycles of isotropic refinement with *SHELXL* using these refined coordinates resulted in a slightly lower *R* factor of 17.7%, although the free *R* remained at 21.5%. The model was inspected using $3F_o - 2F_c$ and $F_o - F_c$ Fourier maps. Only one large error in the structure was identified, Trp62 whose side-chain atoms were incorrectly placed in the previous structure. No electron density was visible for the side chain of Trp62 in the room-temperature study by the Jensen group (Hodsdon *et al.*, 1990; Ramanadham *et al.*, 1989, 1990). There was significant radiation damage to the crystal, so this probably contributed to the problem. The C-terminus of the protein (residues 128–129) was ill defined and removed from the model before the start of anisotropic refinement.

The first round of anisotropic refinement resulted in a sharp drop of the *R* factor to 13.3%, with a comparable decrease in the free *R* to 17.9%. A detailed inspection of the model was then made. Side chains that were either disordered or had multiple conformations were identified. Minor adjustments, such as placement of the correct rotamer, were made to residues Thr43, Asn106, Arg112 and Lys116. Gln121 and Arg128 had very poorly defined electron density for their side-chain atoms. Alternate conformations were modelled for Gln14 and

Asn113. Other side chains such as Arg114 clearly showed more than one conformation, although at this stage of the refinement the electron density was not sufficiently clear to model these alternate side-chain positions. Therefore, to improve the electron density in these parts of the structure, the side-chain atom occupancies were released. This resulted in an improvement in the subsequent difference Fourier, Figs. 3(*a*) and 3(*b*). This strategy was subsequently used at other sites in the protein to aid model building. Solvent sites in the vicinity of disordered residues were intuitively selected as partially occupied, their occcupancy being linked to the appropriate alternative side-chain conformation.

The refinement continued with alternate sessions of refinement and model building (Table 3). After stage 6 with an *R* factor of 11.2% and free *R* of 14.8%, H atoms were added to the model at calculated positions, lowering the *R* factors by approximately 1.0% in both cases. In the previous refinement of triclinic lysozyme (Hodsdon *et al.*, 1990; Ramanadham *et al.*, 1989, 1990), the majority of the water sites were assigned fixed occupancies of less than 1.0, see §5.5. In this study, refining the site occupancies of the water molecules directly using *SHELXL* was assessed after the refinement had converged to an *R* factor of 9.3% in stage 14. The validity of this was confirmed using the

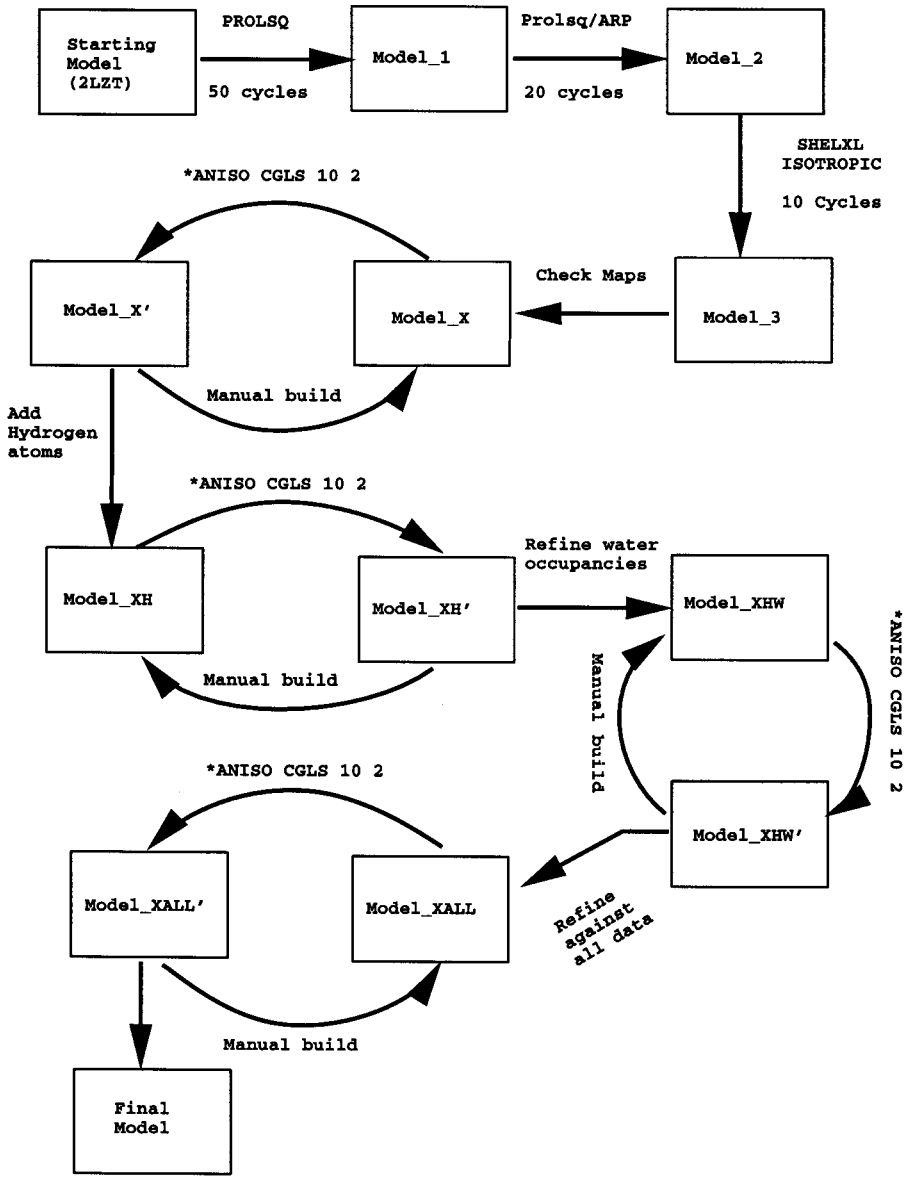


Fig. 2. A scheme of the protocol used in the refinement of triclinic lysozyme. (**SHELXL* options.)

Table 2. *Refinement statistics for triclinic lysozyme*

		PROLSQ	*PROLSQ/ARP*	*SHELXL*†
R factor‡ (%)		25.2 (23.5)§	17.9 (19.7)	8.98 (10.48)
Free *R* factor¶ (%)		28.2 (26.7)	21.5 (23.0)	11.02 (14.10)
No. of atoms (non-H)				
Protein		1001	1001	1104 (1020)
Water		–	279 (123)	250 (139)
Other††				44 (24)
R.m.s. deviations from ideal geometry				
	Target			
Bond length (1–2) (Å)	0.02	0.022 (0.025)	0.015 (0.02)	0.017 (0.017)
Angle distance (1–3)	0.04	0.05 (0.054)	0.037 (0.043)	0.044 (0.033)
Planar distance (1–4)	0.05	0.048 (0.05)	0.036 (0.046)	0.051 (0.050)
Chiral volumes (Å^3)	0.15	0.18 (0.18)	0.15 (0.17)	0.25 (0.16)
Deviations from planes (Å)	0.02	0.02 (0.02)	0.017 (0.018)	0.03 (0.03)
Planar torsion angles (°)	3.0	3.74 (3.58)	3.34 (3.35)	6.84 (6.28)
Mean B_{eq} (Å^2)				
All atoms		9.6 (13.2)	12.37 (14.23)	11.80 (16.38)
Protein atoms		9.6 (13.2)	8.63 (12.7)	8.69 (14.28)
Solvent atoms		—	25.79 (26.79)	21.97 (29.57)

† Calculated values output from *PROLSQ*, using final model from *SHELXL*. ‡ $R = 100 \sum_h \big||F_o| - |F_c|\big| / \sum_h |F_o|$. § Values in brackets are for RTL. ¶ Free *R* factor is calculated for 3% of the data which were omitted from the refinement. †† Nitrate and acetate ions.

free *R* value, see §5.5. After the last stage of refinement using all data a final *R* factor of 8.98% for 54 623 reflections with $F_o > 4\sigma(F_o)$ and 9.25% for all data, 58 367 reflections in the resolution range 20–0.925 Å was obtained.

4.1.2. *Room-temperature structure*. The same procedure as outlined in the previous section was used for the refinement of RTL. A summary of the major steps in the refinement is reported in Table 3(*b*). The isotropic refinement stage using *PROLSQ*/*ARP* resulted in far fewer solvent sites being automatically located. This is more than likely due to the poorer quality of the low-resolution X-ray data and to the fact that data from three different crystals were required for a full data set. Visual comparison of the RTL and LTL electron-density maps after the isotropic stage clearly showed the higher quality of the LTL maps in less rigid regions of the structure as would be expected, Fig. 3(*c*) and 3(*d*). The regions of the protein which showed double or multiple conformations in LTL, were for the most part not visible in RTL. One surprising exception to this was Gln121 which proved to be the most troublesome residue to fit into any reasonable electron density in LTL. Using the room-temperature data, density is clearly visible up to the CG atom although again ambiguous for the terminal atoms. No clear density for an alternate conformation was found. Ser50 was modelled in two conformations in RTL, no electron density was visible for an alternate conformation in LTL. A total of eight residues were assigned alternate conformations: Lys1, Ser24, Thr43, Ser50, Ser85, Ser86, Thr89 and Val109. The occupancies of the side-chain atoms of residues Asn19, Arg68, Asp87, Asn106 and the C-terminal atoms of Arg128 and Leu129 were refined, although no clear electron density for alternative conformations were found. Inspection of $F_o - F_c$ Fourier maps showed negative electron-density peaks in the above regions if the occupancies were fixed at unity.

Anisotropic refinement continued as for LTL, and followed a similar pattern, although the final *R* factor and free *R* converged to higher values of 11.4 and 14.3%. Refinement of water occupancies was also carried out, the *R* factor and free *R* decreased by 0.3 and 0.2%, similar to LTL. After the last stage of refinement using all data the final *R* factor was 10.48% for 56 149 reflections with $F_o > 4\sigma(F_o)$ and 10.82% for all data, 61 258 reflections in the resolution range 20–0.95 Å.

4.2. *Final models*

4.2.1. *Low-temperature structure*. The final model for the 1001 protein atoms of HEWL in LTL consists of a total of 1104 protein atom sites, 250 water molecules, six nitrate ions and three acetate ions. 24 side chains were modelled in two conformations. Three sections of the main chain were modelled in two conformations, residues 103–104, 113–114 and the carbonyl O atom of Thr43. The electron density for the polypeptide chain is excellent; only three residues are poorly defined. The side-chain atoms of Arg112 have no electron density past NE at the level of 1.0σ with B_{eq}'s of 40.6, 39.5, 36.0 and 33.1 Å^2 for the NH1, NH2 CZ and NE atoms, respectively. Gln121 was modelled in two conformations, although the density for the side-chain atoms suggests multiple conformations. The C-terminus is quite flexible, with an average B_{eq} of 18.6 Å^2 for all protein atoms of Arg128 and Leu129. Arg128 was modelled in two distinct conformations, although the somewhat poor density again suggests there may be dynamic disorder or additional conformers. The occupancies of the Leu129 atoms were refined, but no

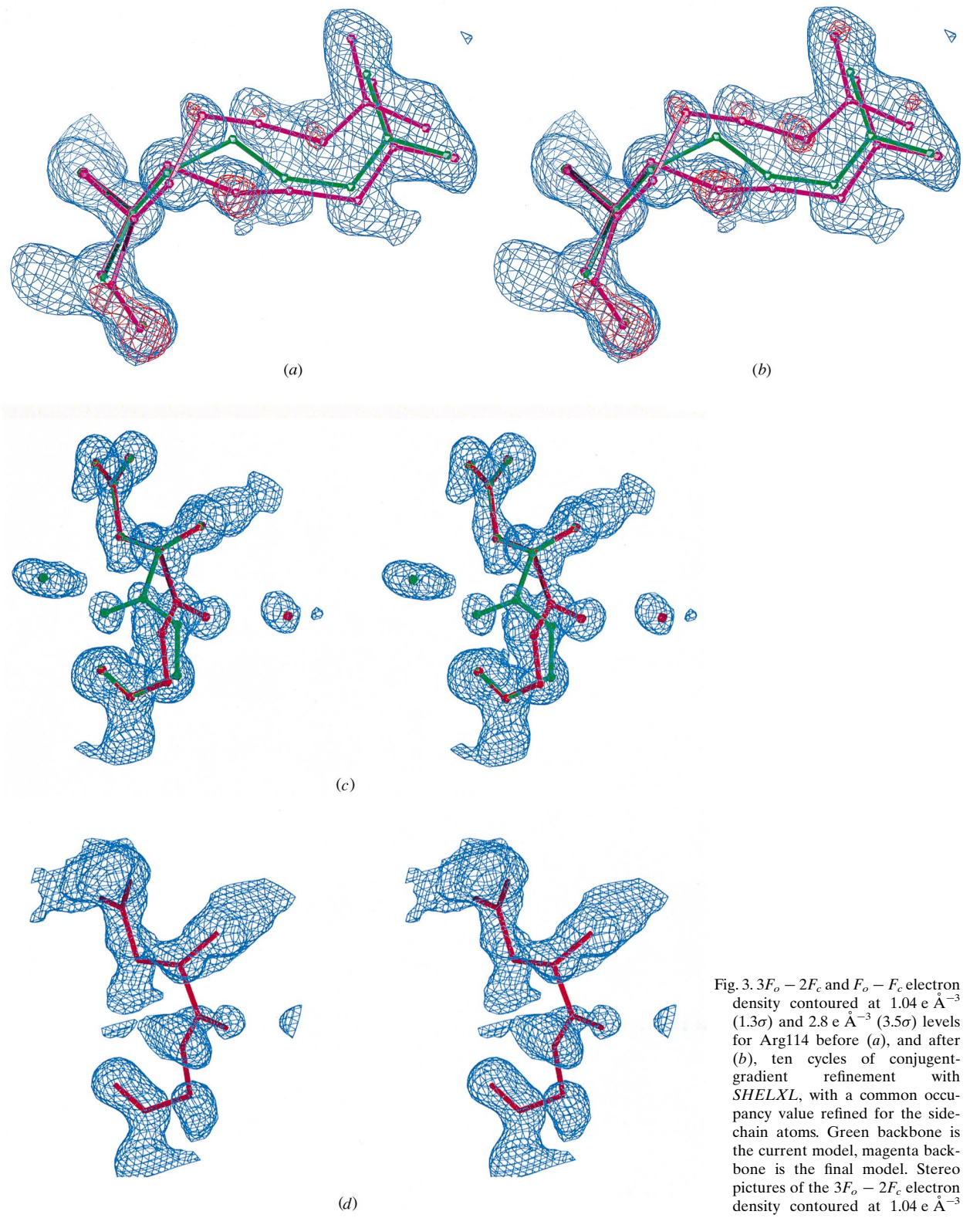
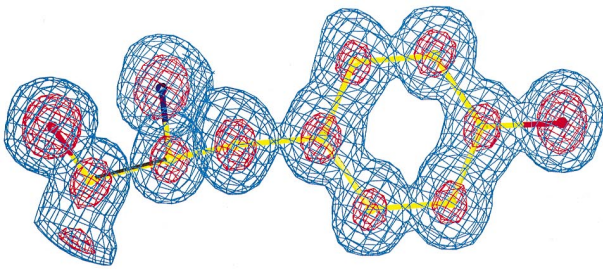


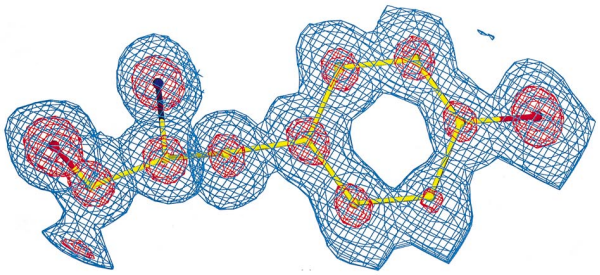
Fig. 3. $3F_o - 2F_c$ and $F_o - F_c$ electron density contoured at 1.04 e Å^{-3} (1.3σ) and 2.8 e Å^{-3} (3.5σ) levels for Arg114 before (*a*), and after (*b*), ten cycles of conjugent-gradient refinement with *SHELXL*, with a common occupancy value refined for the side-chain atoms. Green backbone is the current model, magenta backbone is the final model. Stereo pictures of the $3F_o - 2F_c$ electron density contoured at 1.04 e Å^{-3} (1.3σ) for residues 103–104 in (*c*) LTL and (*d*) RTL. Main-chain disorder is clearly visible in LTL.

alternative conformation(s) were visible in the $3F_o - 2F_c$ or $F_o - F_c$ Fourier maps.

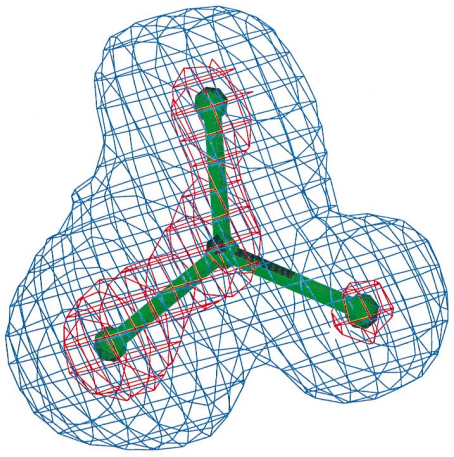
4.2.2. *Room-temperature structure*. The final model for RTL consists of 1020 protein atom sites, 139 water molecules and six nitrate ions. Eight amino-acid side chains were assigned alternate conformations. Regions of protein with high flexibility are more evident in RTL. For example, examination of the sections of the polypeptide chain modelled in two conformations in LTL showed the electron density for residues 99–104 to be broken and not fully interpretable. The electron density for Asn113–Cys115 is fully interpretable, although refined ADP's for this region of the polypeptide chain are higher than average. The main-chain protein atoms of this segment have an average B_{eq} of 13.1 Å^2 compared with 11.0 Å^2 overall. B_{eq}'s for the last 20 residues making up the C-terminus are considerably higher than for the rest of the protein as in LTL. The higher flexibility of the C-terminal end was pointed out in the previous 2 Å study. Other residues with unclear electron density were Asn106, Arg128 and Leu129. Arg128 and Leu129 were refined as partially occupied.



(*e*)



(*f*)



(*g*)

Fig. 3. (*cont.*) (*e*) and (*f*) $3F_o - 2F_c$ electron density contoured at 1.12 e Å^{-3} (1.4σ blue) and 4 e Å^{-3} (5σ red) for Tyr53 in LTL and RTL, respectively. (*g*) $3F_o - 2F_c$ electron density contoured at 1.04 e Å^{-3} (1.3σ, blue) and 3.2 e Å^{-3} (4σ, red) for nitrate-150 in LTL.

4.2.3. *Quality of models*. The quality of the final models for LTL and RTL is summarized in Table 2. A plot of the φ and ψ angles of the polypeptide chain for LTL is shown in Fig. 4 (Ramachandran & Sasisekharan, 1968). 90.3% of the residues are located in the most favoured regions, 9.7% are in the additional allowed regions and no residues are found in the generously allowed and disallowed regions of the plot. Fig. 3(*e*) shows the quality of the electron density of a well defined section of the model in LTL. Electron density is clearly visible at the 5σ level. Fig. 3(*f*) shows the equivalent section in RTL and the corresponding electron density.

Several cycles of blocked-matrix least-squares refinement containing positional parameters of 30 residues and ADP's of 15 residues per block, with two overlapping residues between successive blocks were carried out to estimate the coordinate errors in the two models. Fig. 5(*a*) and 5(*b*) shows the plot of coordinate error *versus* residue number for main-chain atoms only for both structures. The overall r.m.s. error for protein main-chain atoms with unit occupancy is 0.021 Å for

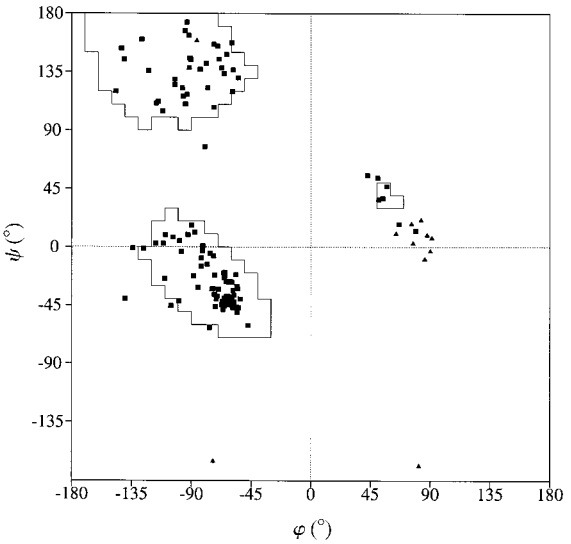


Fig. 4. Ramachandran plot for LTL. Glycine residues are represented as triangles.

Table 3. *Course of the structural refinement of lysozyme*

(*a*) At 120 K

LTL Round 1–3: no manual building of model. Round 4: Trp62 was adjusted and Leu129 assigned zero occupancy. Round 5: adjustment of residues Thr43, Asn106, Arg112, Lys116 and Gln121. Double conformations were assigned for Gln14 and Asn113. Site ccupancies of Arg114 side-chain atoms refined. Leu129 added to model. Five nitrates added. Round 6: double conformations were assigned for Asn19, Arg61, Ser85, Val109, Arg114 and the main chain at Asn103-Gly104. Site occupancies of the side-chain atoms of Arg45, Gln121 and Arg128 refined. Three nitrates added. Round 7: H atoms added. Double conformations were assigned for Thr43, Arg45, Arg61, Val99, Gln121. Site ccupancies of Arg68 side-chain atoms refined. Round 8: double conformations were assigned for Arg68, Lys97 Arg128. Three nitrates removed. Two acetate ions added. Round 9: double conformations were assigned for Asn106, the carbonyl O atoms of Thr43 and the main-chain atoms at Asn113–Cys115, the occupancies being linked to the double conformations of Asn113 and Arg114. Two acetates and one nitrate added. Round 10: double conformation assigned for Asp87. Round 11: Gln121 alternate conformation removed. Two acetates removed. One glycerol molecule added. *SWAT* option for bulk solvent. Round 12: double conformations were assigned for Ser81, Ser86. Partially occupied waters were linked to the disordered side chains of Arg45, Arg61 Arg68, Ser72, Asn106, Gln121. Round 13: weighting changed. Glycerol removed. Round 14: double conformations were assigned for Thr47 and Ser86. Round 15: water site occupancies refined. Round 16: refinement against all data.

Round	1	2	3	4	5	6	7	8
No. cycles	50	20	10	10 (2)†	10 (2)†	10 (2)†	10 (2)†	10 (2)†
No. atoms								
All (non-H)	1001	1280	1260	1260	1312	1327	1318	1355
Protein	1001	1001	993	993	1037	1053	1053	1070
Water	—	279	267	267	255	242	245	269
Other	—	—			5 nitrates	8 nitrates	5 nitrates, add H atoms	5 nitrates 2 acetates
R factor at end (%)	25.2	17.9	17.74	13.3	12.59	11.21	10.48	9.95
R_{free} at end (%)	28.2	21.5	21.5	17.9	16.35	14.8	13.77	13.04
Refinement†	ISO-PLSQ	ISO/ARP	ISO-SH	ANI-SH	ANI-SH	ANI-SH	ANI-SH	ANI-SH

Round	9	10	11	12	13	14	15	16
No. cycles	20 (2)‡	10 (2)‡	10 (2)‡+swat§	10 (2)‡	10 (2)‡	20 (2)‡	10	20 (all data)
No. atoms								
All	1396	1408	1401	1406	1407	1408	1408	1398
Protein	1097	1100	1095	1097	1099	1104	1104	1104
Water	259	268	268	271	272	258	258	250
Other	6 nitrates 4 acetates	6 nitrates 4 acetates	6 nitrates 2 acetates 1 glycerol	6 nitrates 2 acetates 1 glycerol	6 nitrates, 1 partial 3 acetates, 2 partial	6 nitrates, 1 partial 3 acetates, 2 partial	6 nitrates, 1 partial 3 acetates, 2 partial refine water occ.	6 nitrates, 1 partial 3 acetates, 2 partial
R factor at end (%)	9.68	9.49	9.37	9.27	9.28	9.27	9.03	8.98
R_{free} at end (%)	12.63	12.35	12.08	1140	11.62	11.29	11.02	—
Refinement	ANI-SH	ANI-SH	ANI-SH	ANI-SH	ANI-SH	ANI-SH	ANI-SH	ANI-SH

LTL. The overall r.m.s. coordinate error for all protein atoms is 0.032 Å. For RTL the values are 0.030 and 0.038 Å.

The O—N—C—CA atoms were restrained to lie in an unweighted least-squares plane with a standard deviation of 0.5 $Å^3$. The average standard uncertainty of the ω values as estimated by the inversion of the least-squares matrix is 0.74°. The distribution of ω shows that the peptide bond is not planar, with some peptides deviating from planarity by up to 20–25°. The distribution is not symmetric, it has a mean value of 179.8°, with a standard deviation of 7.0°, Fig 6(*a*). The largest deviation from planarity is for the peptide bond at Trp63, (CA62, C62, N63, CA63) which deviates by 24.6°. RTL has a similar distribution of ω angles with a mean value of 179.2° with a standard deviation of 6.3°, Fig. 6(*b*). The largest deviation from planarity again occurs at Trp63. The sample size is small in the above analysis, but agrees with more recent surveys of the peptide ω angles in both the Cambridge Structural Database (Allen & Kennard, 1993) and the PDB (MacArthur & Thornton, 1996).

No restraints were imposed on χ^1 angles during refinement. Analysis of χ^1 angles for all residues in LTL reveals three clusters corresponding to three possible rotamers, Fig. 6(*c*). The mean values and standard deviations in parentheses of these peaks calculated by fitting three Gaussian functions in the distribution, were 66 (4), 181 (6), and −65 (9)° when residues with double conformations are included and 67 (5), 183 (7), −67 (9) when excluded. Similarly for RTL three peaks were observed at 65 (4), 183 (10), −67 (8) when residues with double conformations are included and 67 (3), 183 (9) and −67 (8) when not, Fig. 6(*d*). These values again agree with the study by Sevcik *et al.* (1996). The rotamer torsions should not be restrained in refinement as these unrestrained para-

Table 3. (*cont.*)

(*b*) At 295 K

RTL Rounds 1–3: As for LTL. Round 4 Adjusted Gln121. Double conformations were assigned for Val109. Site occupancies refined for side-chain atoms of Arg68, Asp87, Asp106, Arg128, Leu129. Added one nitrate ion Round 5: four nitrate ions added. Round 7: add H atoms. Round 8: double conformations were assigned for Ser24, Arg68. Site occupancies refined for side-chain atoms of Asp19. Round 9: adjusted 104, Gln121, Arg128, Leu129. Double conformation assigned to Ser50. Round 10: changed weighting scheme. Round 11: water site occupancies refined. Round 12: all data. Round 13: adjusted Leu12. Added one nitrate ion. Round 14: Adjust Gln41. Model bulk solvent with *SWAT* option in *SHELXL*.

Round	1	2	3	4	5	6	7	8	9	10
No. cycles	50	20	10	10 (2)†	10 (2)†	10 (2)†	10 (2)†	10	50	20
No. atoms										
All	1001	1125	1125			1139	1139	1177	1181	1181
Protein	1001	1001	1001	1001	1016	1007	1016	1020	1020	1020
Water	—	124	124		124	116	145	141	141	141
Other	—	—	—	—	1 nitrate	4 nitrates	4 nitrates		5 nitrates	5 nitrates
R factor at end (%)	23.5	19.7	19.8	14.3	13.4	12.9	11.6	11.5	11.4	11.4
R_{free} at end (%)	26.7	23.0	21.9	17.3	16.6	16.1	14.4	14.7	14.4	14.3
Refinement	ISO-PLSQ	ISO/ARP	ISO-SH	ANI-SH	ANI-SH	ANI-SH	ANI-SH	ANI-SH	ANI-SH	ANI-SH

Round	11	12	13	14
No. cycles	10	10(2)†	10(2)†	20 (2)†
No. atoms		all data		
All	1181	1181	1183	1183
Protein	1020	1020	1020	1020
Water	141	141	139	139
Other	5 nitrates	5 nitrates	6 nitrates	6 nitrates
R factor at end (%)	11.1	11.1	10.9	10.48
R_{free} at end (%)	14.1			
Refinement	ANI-SH	ANI-SH	ANI-SH	ANI-SH

† ISO-PLSQ = isotropic refinement with *PROLSQ*; ISO-SH = isotropic refinement with *SHELXL*; ANI-SH = anisotropic refinement with *SHELXL*. ‡ Two cycles of isotropic refinement carried out where new atoms were added to the model before continuing with anisotropic refinement. § *SWAT* option in *SHELX* which models contribution of bulk solvent, see text.

Table 4. *Least-squares superpositions of lysozyme coordinate sets*

(*a*) 2LZT onto the current models at atomic resolution.

	R.m.s deviations (Å)		
	R.m.s displacement	Average displacement	Maximum displacement
Low-temperature model (LTL)			
Cα atoms only†	0.33	0.30	0.92 (Gly102 Cα)
Main-chain atoms†	0.34	0.31	1.0 (Gly102 O)
Room-temperature model (RTL)			
Cα atoms only†	0.10	0.09	0.39 (Gly102 Cα)
Main-chain atoms†	0.12	0.10	0.75 (Val99 O)

(*b*) RTL onto the LTL coordinate set

	R.m.s deviations (Å)		
	R.m.s displacement	Average displacement	Maximum displacement
Cα atoms only†	0.3	0.27	0.63 (Gly102)
Main-chain atoms†	0.3	0.27	1.1 (O113)
All atoms†	0.53	0.34	6.5 (NE2121)

† Using residues 1–127.

meters are useful in the validation of the refined structure (Wilson *et al.*, 1998). The results highlight the tighter clustering of these unrestrained parameters around the energetically preferred values as the resolution of the analysis increases.

Comparison of the backbone dihedral angles ω, φ, and ψ of LTL and RTL show good agreement and mirror the comparison of the independent molecules in the asymmetric unit of ribonuclease from *Streptomyces aureofaciens* (Sevcik *et al.*, 1996). The r.m.s. deviation between the ω angles of the two structures is 3.2°, Fig. 6(*e*). The average ω angle was closer to 180° (179.2°) than in the previous study. The correlations between the φ and ψ angles were also excellent, with r.m.s. deviations between pairs of angles in the two structures of 4.8° for φ and 4.5° for ψ, Fig. 6(*f*) and 6(*g*).

5. Description of structure and comparison with HEWL refined at 2.0 Å

5.1. *Overall structure*

In the following sections LTL at atomic resolution is described and taken as the reference structure for all

further comparisons to those determined at room temperature, *i.e.* 2LZT and RTL. The polypeptide fold of triclinic lysozyme refined against 0.925 Å data at 120 K is essentially identical to that of the room-temperature structures. The considerable increase in the number of data at this resolution has allowed a very

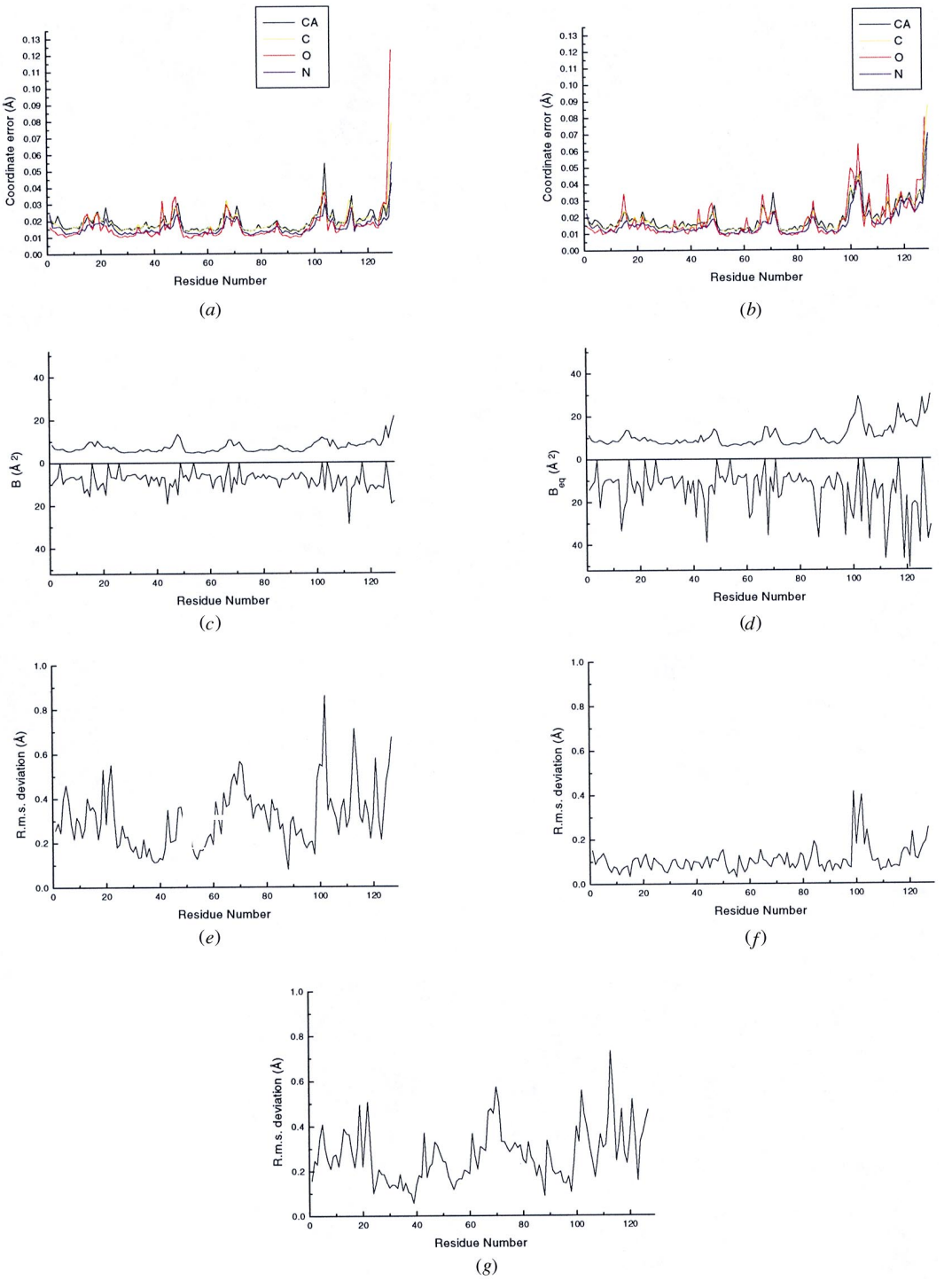


Fig. 5. Coordinate errors for main-chain atoms estimated from matrix inversion for (*a*) LTL and (*b*) RTL. Plots of B_{eq} averaged over main-chain and side-chain atoms for (*c*) LTL and (*d*) RTL. Plot of the r.m.s. deviation between the main-chain atoms of (*e*) LTL and 2LZT, (*f*) RTL and 2LZT, (*g*) LTL and RTL.

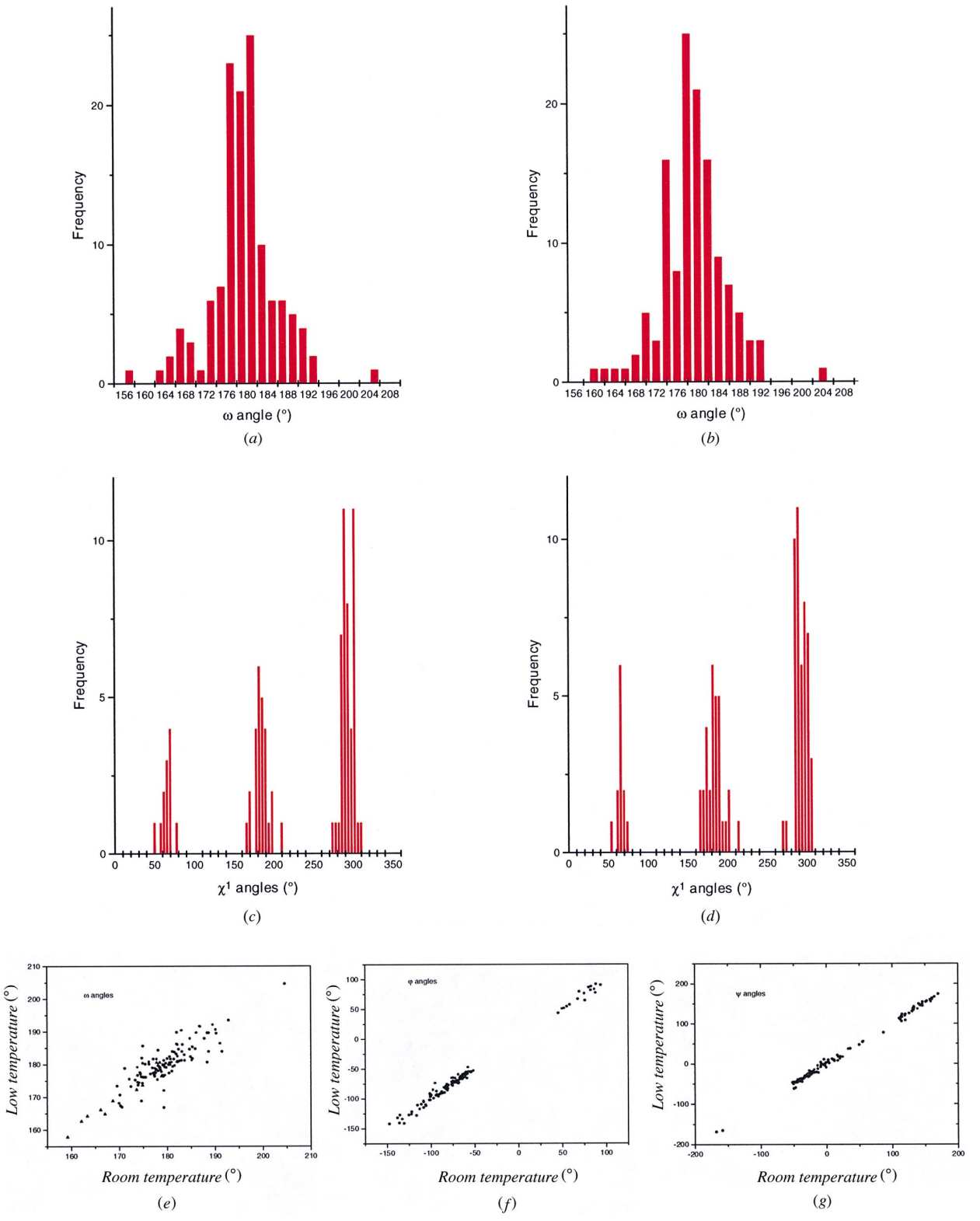


Fig. 6. Histograms of torsional angles for LTL and RTL (*a*, *b*), ω; (*c*, *d*), χ^1. Comparison of torsion angles in LTL and RTL (*e*) φ, (*f*) ψ, and (*g*) ω. Glycines and prolines represented by triangles.

detailed and precise interpretation of the electron density. Extensive modelling of disorder has been executed.

Although data to 0.950 Å have been collected at room temperature, the quality of the resulting electron-density maps is inferior to that of LTL. Consequently far less disorder has been modelled. The reason for this is twofold. Firstly the quality of the 0.95 Å room-temperature X-ray data is relatively poor, when compared with those from low temperature, Table 1 and Fig. 1(*a*) and 1(*b*). The data were collected from three crystals, although it is not clear why the data, in particular at low resolution, are not of higher quality. Secondly, collection of X-ray data at cryogenic temperatures greatly extends the lifetime of the protein crystal in the X-ray beam and usually allows collection of X-ray data to much higher resolution. The collection of the data at low temperature greatly reduces the thermal vibration of the atoms and in regions of disorder can trap the highest populated disordered conformations.

In the previously refined 2.0 Å model, no electron density was visible for Trp62 which plays a role in binding of the enzyme substrate (Blake *et al.*, 1967; Imoto *et al.*, 1972; Maenaka *et al.*, 1994). The lack of density was attributed to radiation damage during data collection; the crystals turned yellow after prolonged exposure to X-rays. Electron density was clearly visible for this residue in both LTL and RTL, although the B_{eq}'s for the side-chain atoms of Trp62 in RTL are 3–4 times higher than for Trp63. The high mobility of this residue is clearly visible in the *ORTEP* diagram shown in Fig. 7(*a*) and 7(*b*). This is probably due to the faster data acquisition, made possible by use of synchrotron radiation for RTL. However, past studies found that Trp62 could be selectively modified by ozone to give kynurenine but left Trp63 unmodified (Yamasaki *et al.*, 1979). It appears that Trp62 is sensitive to radiation damage in a manner similar to what is observed with ozone. Similar poor electron density is found in other crystal forms of HEWL. When poor electron density is observed for a section of a protein structure, the explanation for the unclear map is usually attributed to greater flexibility in this region, but as is seen here, this may arise from other causes.

R.m.s. and maximum deviations between the structures based on various least-squares superposition of CA atoms are given in Table 4. The two room-temperature structures are essentially identical with r.m.s. deviation between CA atoms of 0.12 Å or less. A maximum displacement of 1.19 Å occurs at the C-terminal residue Leu129; the electron density for this residue was not clearly defined in RTL. LTL and RTL have larger differences, with an r.m.s. deviation between CA atoms of the order of 0.3 Å, which is due to the shrinkage of the unit cell on freezing and concomitant movements in the protein molecule. The r.m.s. deviations between corresponding main-chain atoms based on least-squares superposition of these atoms of LTL/2LZT, RTL/2LZT and LTL/RTL are plotted as a function of residue number in Figs. 5(*e*), 5(*f*) and 5(*g*).

5.2. *Regions of the protein structure with more than one conformation*

In LTL the side chains of 24 residues were modelled with double conformations, in particular five out of the ten serines, and six out of 11 arginines, Table 5. Several polar side chains on the surface, were found to have more than one conformation. However, the side chains of some buried residues such as valine and threonine were also assigned double conformations.

For RTL the side chains of only eight residues were modelled in two conformations, Table 5. Two of these, Ser50 and Thr43, had only a single conformation in LTL. The additional side chains modelled in double conformations in LTL, had no clear electron density for alternate conformations in RTL; however the density was in general of inferior quality to parts of the structure which were more rigid. In addition in RTL, the occupancy of the side chains of a further six residues was refined, but with a single conformation only.

Arginines with only one conformation, Arg5, Arg14 Arg21 and Arg73 are in well defined density. Arg112 had poorly defined electron density for the terminal side-chain atoms, with B_{eq} values of 35–40 Å^2. The electron density showed no clear density for an alternate conformation. The electron density for Arg128 allowed two possible conformations to be modelled, but it was clear that further conformations were highly likely. Attempts were made to interpret the electron density in this region of the structure more thoroughly, with additional conformations of this residue, but no satisfactory agreement was found. These arginines in RTL were modelled in one conformation, although the electron density was not as clear and suggested that more than one conformation of the side chain was present, except for Arg73 which had well defined electron density. The electron density for Gln121 proved difficult to model in LTL and again only two conformations of many possible were modelled.

In addition to alternate conformations of side chains, parts of the main chain were found to have clearly more than one conformation. The first section of the polypeptide chain where two conformations were modelled, is located at the peptide bond between residues 103 and 104, where in simple terms a peptide flip occurs. The carbonyl group of Asn103 is rotated by approximately 180° between the two conformations, with the backbone N and CA of Gly104 refined in two slightly different positions, Fig. 3(*c*). The loop region at residues 113–115 shows more extensive disorder, due primarily to the flexibility of the side chains of Asn113 and Arg114, Fig. 7(*c*). The third section of main-chain disorder is located

at the carbonyl O atom of Thr43 and is linked to the disorder of its side chain, Fig. 7(*d*).

The disordered residues were for the most part found to interact through hydrogen-bonding networks, either directly or indirectly *via* water molecules, to other disordered regions in the protein. Only Ser72 and Ser86 had no direct or indirect interactions with other disordered side chains. Taking each alternate conformation separately, a number of examples of larger networks can be easily traced in LTL, Fig. 8. The first example consists of the *A* conformations of Asn19, Arg114, the *A* and *B* conformations of Asp18 and Gln41 and acetate ion 201 (Ace201). The OD1 atom of *A*18 interacts *via* hydrogen bonds between water w2120 to *A*41 NE2 and *B*41 OE1. *A*41 NE2 in turn hydrogen bonds to OD1 of *A*19. *B*18 OD1 hydrogen bonds to *A*114 NH2, and *A*114 NH1 hydrogen bonds to Ace201. The OD1 atom of *A*19 hydrogen bonds to NE2 of *A*41. The second example consists of the *B* conformations of Asp19, Ser81, and the *A* conformations of Ser85 and Asp87. The OD1 atom of Asp87 interacts with the OG atom of Ser85 *via* two tightly bound water molecules, w2018 and w2056. Ser85 interacts with the ND2 atom of Asn19 *via* w2281. The ND2 of Asn19 makes a direct hydrogen bond to the OG atom of Ser81. The third example consists of the *A* conformers of Ser24, Thr43, Arg68, and the *A* and *B* conformers of Arg45. The OG of Thr43 hydrogen bonds to OG of Ser24 and to the *A* and *B* conformers of Arg45 *via* w2022. Arg45 in turn hydrogen bonds *via* w2022 to Arg68. The fourth example consists of the *A* conformer

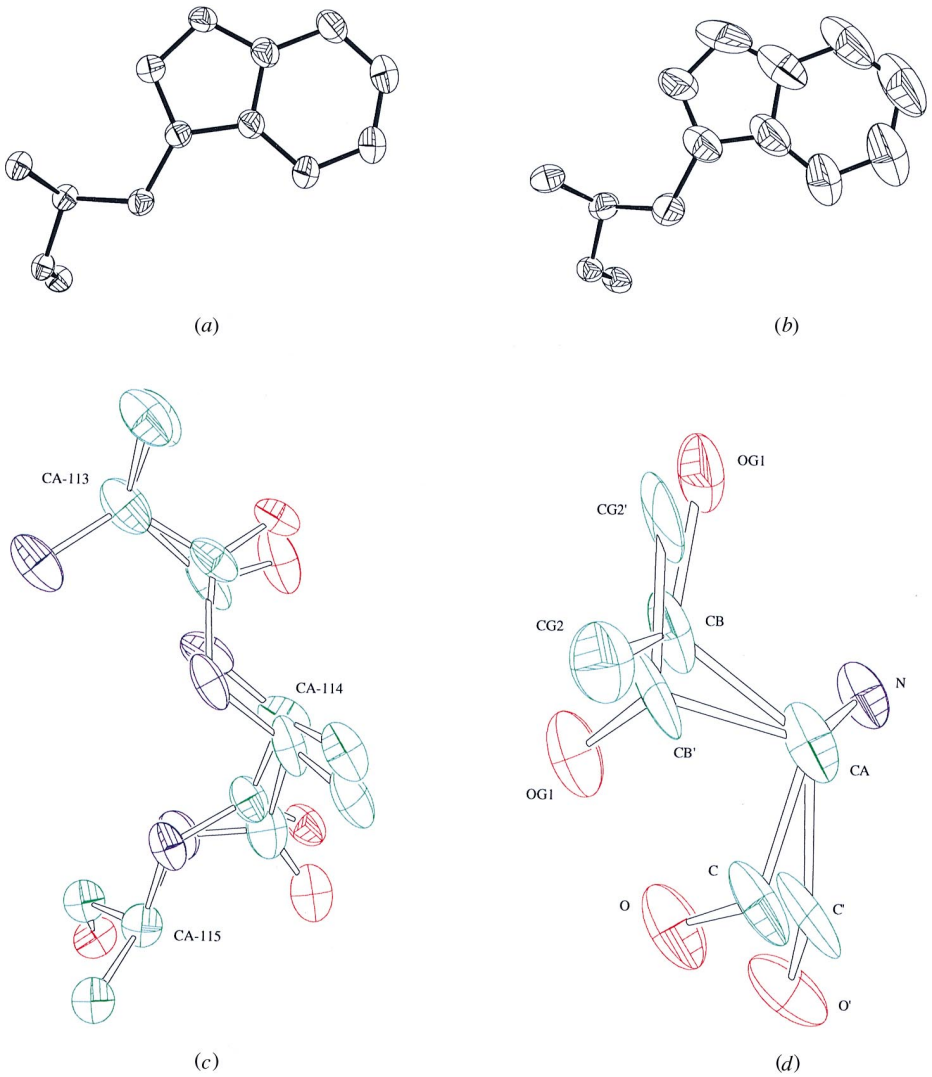


Fig. 7. *ORTEP* drawings (20% ellipsoids) of Trp62 in (*a*) LTL and (*b*) RTL; (*c*) main-chain disorder at 113–115 in LTL, only CB atoms of side chains are shown for clarity. (*d*) Main/side-chain disorder at Thr43 in LTL. The figures were produced using the program *ORTEX* (McArdle, 1994).

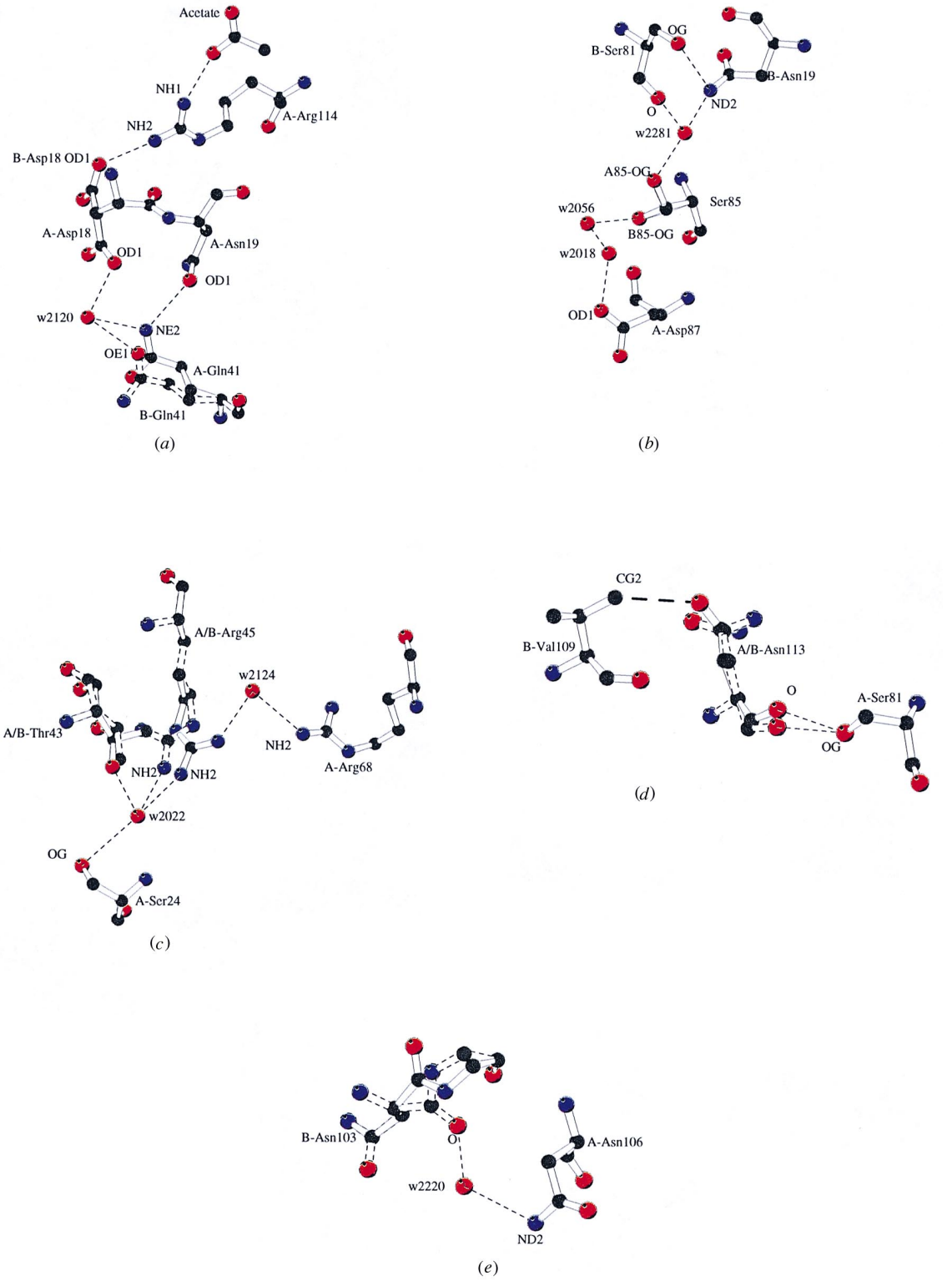


Fig. 8. Ball-and-stick representations of five interacting networks of disordered regions of the protein structure in LTL, see text for details. The figure was produced with the program *MOLSCRIPT* (Kraulis, 1991).

Table 5. *Residues with alternative conformations*

Occupancies of the side-chain atoms of Asn19, Arg68, Asp87, Asn106 and Arg128, Leu129 were refined in RTL but no alternate conformation(s) were assigned.

(*a*) For LTL

Residue	Occupancy	Atoms occupancy refined	Residue	Occupancy	Atoms occupancy refined
Lys1	0.78	CE, NZ	Ser85	0.64	CB, OG
Asp18	0.78	CB, CG, OD1, OD2	Ser86	0.79	CB, OG
Asn19	0.82	CB, CG, OD1 ND2	Asp87	0.78	CB, CG, OD1, OD2
Ser24	0.68	CB, OG	Lys97	0.5	CE, NZ
Gln41	0.59	CB, CG, CD, OE1, NE2	Val99	0.85	CB, CG1, CG2
Thr43	0.65	CB, CG2, OG1	Asn106	0.55	CB, CG, OD1, ND2
Arg45	0.63	CD, NE, CZ,NH1 NH2	Val109	0.79	CB, CG1, CG2
Thr47	0.67	CB, CG2, OG1	Asn113	0.54	CB, CG, OD1, ND2
Arg61	0.88	CD, NE, CZ, NH1, NH2	Arg114	0.6	CA, CB, CG, CD, NE, CZ, NH1, NH2
Arg68	0.54	CG, CD,NE, CZ, NH1, NH2	Gln121	0.59	CB, CG, CD, OE1, NE2
Ser72	0.8	CB, OG	Arg125	0.67	NE, CZ, NH1, NH2
Ser81	0.69	CG, OG	Arg128	0.54	CB, CG, CD, NE, CZ, NH1, NH2

(*b*) For RTL

Residue	Occupancy	Atoms occupancy refined	Residue	Occupancy	Atoms occupancy refined
Lys1	0.67	CE, NZ	Ser85	0.63	CB, OG
Ser24	0.57	CB, OG	Ser86	0.54	CB, OG
Thr43	0.50	CB, CG2, OG1	Thr89	0.62	CB, CG2, OG1
Ser50	0.86	CB, OG	Val109	0.63	CB, CG1, CG2

of Ser81, the main chain at Asn113 and the *B* conformer of Val109. Ser81 OG hydrogen bonds to the disordered carbonyl group of Asn113. CG2 of Val109 is at a distance of 3.1 Å from the OD1 atom of Asn113. The last example involves the *A* conformer of Asn106, a water and the *B* conformation of the disordered main chain at 103–104. The ND2 atom of 106 hydrogen bonds to water w2220 which in turn bonds to the carbonyl O atom of 103.

Recent relatively high resolution structure determinations of HEWL crystals in the tetragonal form grown on earth and in space, have also modelled disordered residues, but to a far lesser extent (Vaney *et al.*, 1996). One serine residue was located that had two distinct conformations, Ser86, which correlates with the present study where two conformations of this residue have been modelled. No other serine residues were modelled or observed to have more than one conformation in that study. Three other residues were modelled in two conformations, Lys1, Val109 and Asn59, although the latter was only observed in crystals of HEWL grown on earth. Two of the six valines in the triclinic lysozyme were modelled with two conformations, Val99 and Val109, the former with occupancies of 0.85 and 0.15, the low occupancy value of the alternate conformation of Val99 could be a possible reason for this not being observed in the above study. However, this residue has poorly defined side-chain electron density in RTL. One asparagine was modelled with two conformations in the triclinic structure but it was Asn19 and not Asn59 as in the earth-grown tetragonal HEWL. The side-chain atoms of Asn59 are located in well defined electron density in both LTL and RTL. The terminal side-chain atoms (CE, NZ) of Lys1 and Lys97, had clearly interpretable density for two conformations in the present study, again showing some correlation with the tetragonal structure. From the results of the present study, it is clear that the acquisition of atomic resolution X-ray diffraction data in combination with cryocrystallographic techniques allows better modelling of the flexible parts of the protein structure, by freezing out the disorder, and/or trapping the major conformations of the flexible parts of the protein structure and thus providing albeit a static model of the flexible sections of the protein structure.

5.3. *Displacement parameters*

The variation of the average main-chain and side-chain B_{eq}'s along the polypeptide chain for both LTL and RTL is shown in Figs. 5(*c*) and 5(*d*). Table 6 shows the B_{eq} statistics for LTL, RTL and 2LZT. The average B_{eq} value for LTL main-chain protein atoms shows only a small decrease in overall value, when compared with RTL and 2LZT. However, a more significant drop is observed for the protein side chains. The average B_{eq} value for waters is again comparable, the average for LTL is lower than RTL as expected, even though it has 111 additional waters. The occupancies of the water molecules were refined in the present atomic resolution study. In 2LZT, 35 waters were assigned unit occupancy, the rest being partially occupied. The occupancies of the waters were not refined, but fixed at values less than unity assigned on the basis of peak height from a difference map where the height of a carbonyl O atom was used for calibration.

The average B_{eq}'s of other HEWL structures in the PDB reveal values in the range 15–25 Å^2. A low-temperature study of monoclinic and tetragonal HEWL (Young *et al.*, 1993, 1994) had an average B_{eq} value of

Table 6. *Comparison of displacement parameters*

	2LZT		0.95 Å (RTL)		0.925 Å (LTL)	
Atoms	Total	$\langle B_{eq} \rangle$	Total	$\langle B_{eq} \rangle$	Total	$\langle B_{eq} \rangle$
Protein main chain	517	9.11	517	11.0	530	7.36
Protein side chain	484	11.93	503	17.65	574	9.91
Protein atoms (all)	1001	10.47	1020	14.28	1104	8.69
Water	249	28	139	28.99	250	23.34
Other	20	26.15	24	32.96	44	14.15
All	1270	14.24	1183	16.38	1398	11.80

Table 7. *Bound nitrate ions in triclinic lysozyme*

LTL	RTL	2LZT
Nitrate 150	Water 1062	Water 192
Nitrate 151	Nitrate 203	Nitrate 132
Nitrate 152	Nitrate 202	Nitrate 131
Nitrate 153	Nitrate 206	w183; w353
Nitrate 154	Nitrate 201	Nitrate 130
Nitrate 155	Nitrate 204	Nitrate 133
Gln121	Nitrate 205	Nitrate 134

approximately 8 Å^2 although data only extend to 2.0 Å. This conflicts with a more recent study where the average B_{eq} value for the protein atoms was reported as 22.2 Å^2 for data collected at 120 K on tetragonal HEWL (Kurinov & Harrison, 1995). Young *et al.* (1993) also report an average B_{eq} value of 15.2 Å^2 in a room-temperature study of tetragonal HEWL. In an independent study of tetragonal HEWL at 1.5 Å, we observed average B_{eq} values in the range 20–24 Å^2 (to be published).

However, Young *et al*. refined their structures against data restricted to the 6–2 Å range. In our view, the refinement of protein structures with such a low-resolution cut-off is to be discouraged. This practice only results in cosmetically lower *R* factors and not in better structures. The low-resolution data provide important information on the solvent structure around the protein. They should be measured and used to the lowest resolution attainable, ideally 20 Å or less. The use of all data in the refinement protocol results in higher quality electron-density maps with little or no discontinuity. The difference in the overall average *B* values in the above studies is almost certainly a direct result of truncating the low-resolution data.

5.4. *Bound ions*

5.4.1. *Nitrate ions*. A total of six nitrate ions were located in LTL, with one of these ions being modelled as partially occupied. All the nitrate ions are bound to the surface of the protein and interact directly or indirectly *via* waters with another protein molecule. Fig. 3(*g*) shows the final electron density for nitrate 150.

There are six nitrate sites in LTL and RTL, but only five in 2LZT, Table 7. Four sites are common to all three structures. The two atomic resolution structures, LTL and RTL, share an additional site, but the sixth nitrate lies in a different position. The sixth nitrate in RTL, not found in LTL, corresponds to the fifth site in the other room-temperature structure 2LZT.

The sixth and unique nitrate ion in LTL is tightly bound to the protein *via* hydrogen bonds to the backbone amide N atom of Asn106 and the hydroxyl group of Tyr23 *via* atoms O3 and O2 of the nitrate ion, respectively. The O1 atom is in hydrogen-bonding distance to the side-chain atoms OD1 and ND2 of the *B* conformation of Asn106. The high flexibility of Asn106 in RTL may explain why this nitrate was not identified in 2LZT. The other anomaly in the nitrate sites is the modelling of electron density close to Gln121 as a nitrate ion in both of the room-temperature structures, and as a possible alternate conformation for the Gln121 side chain in LTL. RTL shows Gln121 to have far clearer and interpretable electron density than in LTL, although no density is visible for the end of the side chain. During refinement of LTL the electron density at this site was accounted for, firstly by the Gln121 side-chain atoms and secondly by placing a nitrate ion in this position. After inspection of the resulting difference Fouriers for both cases, the model with the Gln121 side-chain atoms accounting for the electron density was chosen. However, it is also feasible that the site may be partially occupied by a nitrate ion in both LTL and RTL.

The crystal structure of a low-humidity monoclinic form of HEWL has two nitrates bound, again at the interface of two molecules (Kodandapani *et al.*, 1990). The nitrate sites in this form are in no way related to the nitrate sites found in the present study. The only similarity is that nitrate 131 ($P2_1$ form) is in close proximity to the binding site of nitrate 155 in the triclinic form.

5.4.2. *Acetate ions*. A total of three acetate ions were located in LTL, two of which were modelled in two discrete but overlapping sites. No acetate ions were found in RTL or 2LZT. The first two partially occupied acetate ions in LTL were placed in the model in electron density which was first thought to resemble a bound glycerol molecule from the cryoprotectant. This glycerol site was reassessed to be more likely due to acetate ions after consideration of the chemical environment and the fact that after further refinement the resultant $F_o - F_c$ electron-density maps showed a strong negative electron-density peak for the glycerol molecule. The other acetate ions were placed in sites which were initially modelled as nitrate ions. Ace201 and Ace204 are located

at the surface of the protein and are bound at the interface of two molecules. Ace203 is found close to the surface of the protein in a crevice close to the 60's loop region of the protein backbone. The equivalent site of the first acetate ion (Ace201) in LTL is occupied by a water atom (205) in 2LZT. No waters or ions were placed at this position in RTL. There is one water site in both room-temperature models at one of the carboxylate O atoms of Ace203 which was modelled in two discrete sites. Two waters at approximately the same positions as the carboxylate O atoms of the acetate ion 204 are present in both the room-temperature structures.

5.4.3. *Other cations and anions*. An Na^+ ion was located in the recently refined structure of the tetragonal form of HEWL (Vaney *et al.*, 1996). This ion binds in the crevice formed by the 60–75 loop region. This loop has been shown to be highly flexible and differs in the triclinic structure, mainly as a result of a peptide flip at Arg73. This conformational change was compared in all the HEWL structures at better than 2.0 Å to assess if the well defined loop in their structure traps an Na^+ or water molecule and thus holds the conformation in place. In the present structure the space provided by the peptide flip is filled by a water molecule (w2010) which hydrogen bonds to the amide N atom of Asn74 and the OG atom

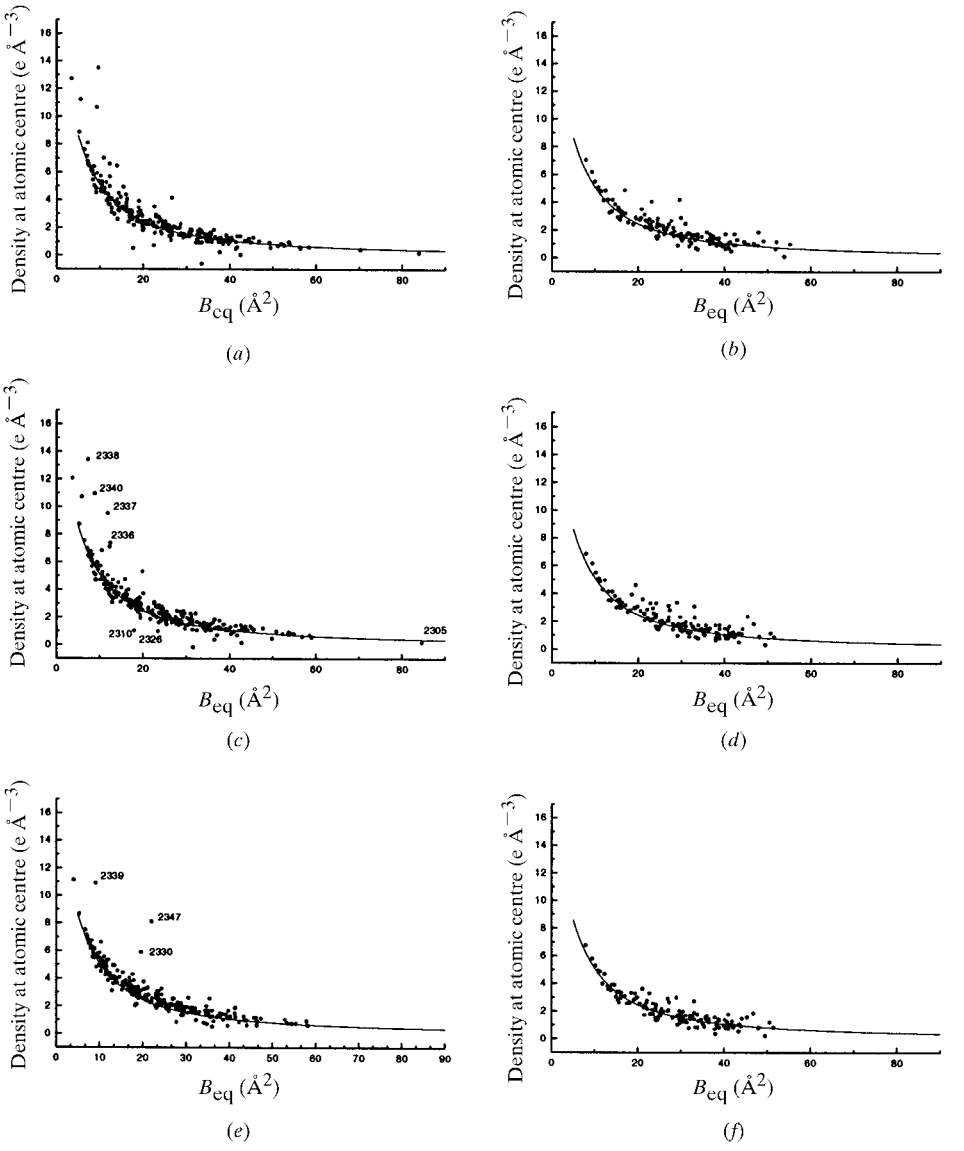


Fig. 9. ($3F_o - 2F_c$, α_c) electron densities [on an absolute scale, $F(000)$ contribution included], at atomic centres as a function of B_{eq}. Water molecules before refinement of occupancies using *SHELXL*, (*a*) LTL and (*b*) RTL. Refined water molecules occupancies in (*c*) LTL and (*d*) RTL. Waters indicated were deleted before refinement against all data. (*e*) and (*f*) as (*c*) and (*d*) but after refinement against all data.

of Ser72. The bound Na^+ ion observed in the tetragonal form is not present, even though sodium acetate and sodium nitrate were used in the crystallization buffer. There is a chloride ion bound between the interface of two molecules in the above tetragonal form interacting with the hydroxyl group of Tyr23 and the ND2 of Asn112 of another protein molecule. No chloride was present in the crystallization buffer of triclinic lysozyme, but the equivalent site in the triclinic form is filled by the side-chain atoms of Arg68 of another protein molecule, although nitrate 153 does interact with the hydroxyl group of Tyr23 from the other side of the functional group. This nitrate group interacts with another molecule *via* w2109 and w2124.

5.5. *Water structure*

5.5.1. *Low-temperature structure*. Water peaks were first located automatically during isotropic refinement using *ARP*, where 279 electron-density peaks were attributed to water molecules. During refinement with *SHELXL*, water/solvent peaks were manually added or subtracted using as a guide peaks greater than 5σ in the difference Fouriers ($F_o - F_c$ maps). An approximate calculation of the non-protein volume of the triclinic cell provides enough room for 269 full occupancy water molecules. Modelling of the diffuse solvent using Babinet's principle (Moews & Kretsinger, 1975) as implemented in the *SWAT* option in *SHELXL* was therefore delayed until the final rounds of refinement; since the majority of the total number of possible solvent sites had been accounted for in the earlier stages of the refinement. Use of the *SWAT* option in the final stages of refinement resulted in a drop of 0.15% in the *R* factor and 0.04% in the free *R*, and thus contributed very little to the agreement of the low-resolution data.

In the later stages of refinement during modelling of alternate side-chain conformations, 29 waters which were clearly associated with one conformation of a disordered side-chain residue had their occupancies linked to the refined value of the protein side chain in question. 20 waters were assigned fixed occupancies, less than unity, where they were located as peaks greater than 5σ in the $F_o - F_c$ electron-density maps. The *R* factor for the model with 210 fully occupied waters and the 49 waters with occupancies less than unity was 9.27%. Visual inspection of the electron density, suggested that many water sites were probably only partially occupied.

X-ray data to atomic resolution allow the identification of the atomic type directly from the electron-density map. A plot of ($3F_o - 2F_c$, α_c) density interpolated at atomic centres as a function of B_{eq}'s (DB plot) for the water molecules is shown in Fig. 9(*a*). This shows a proportion of the water molecules to be below the theoretical distribution for an O atom, again suggesting some water sites are less than fully occupied. Therefore, the occupancies of all water molecules apart from those specifically linked to the refined occupancies of disordered side chains were released. After ten cycles of anisotropic refinement, the *R* factor and free *R* dropped by 0.24 and 0.27%, respectively. The more or less equivalent drop in both the conventional *R* factor and free *R*, convincingly indicated the validity of this step. The DB plot after this stage showed improved agreement with the theoretical prediction, although a number of solvent sites deviated from the theoretical distribution, Fig. 9(*c*). The waters highlighted in the latter plot were deleted before continuing with refinement. w2310 and w2372 had no visible electron density in $3F_o - 2F_c$ maps contoured at the 0.8σ (0.6 e $Å^{-3}$) level. w2305 had a $B_{eq} > 80$ $Å^2$. w2336, w2337 and w2338 were situated at the same positions of the O atoms of a partially occupied nitrate ion (nitrate 155) and their occupancies had been linked to the occupancy of this nitrate. As they deviated significantly from the DB plot they were deleted. Finally, water w2340 had been placed at a possibly split water peak. After refinement against all data there were three water sites that deviated significantly from the theoretical distribution; w2339 is located at the end of a 'tear-drop'-shaped split water peak, w2347 is part of a smeared peak of electron density from a water network and w2330 which is

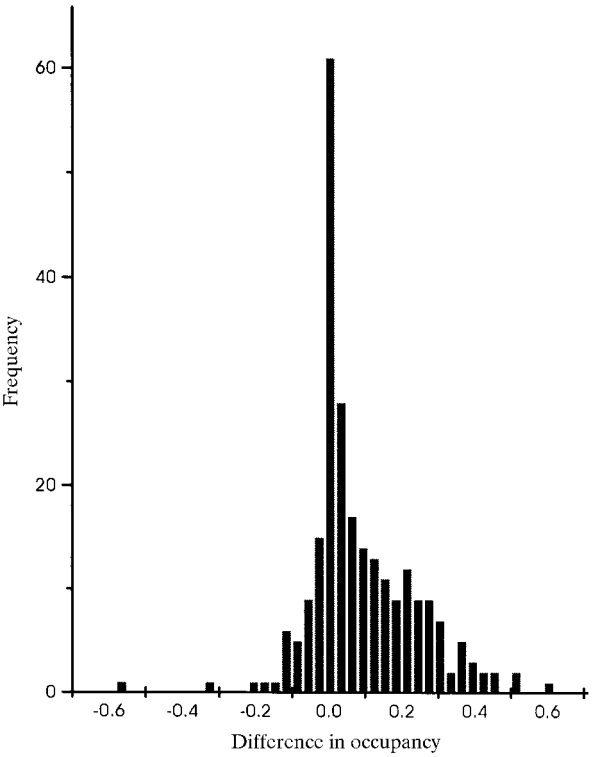


Fig. 10. Histogram of the distribution of differences in water molecules occupancies in LTL derived from refined values with *SHELXL* and those estimated from the DB plot, [($OCC_{db} - OCC_{sh}$) see text].

located in a clear electron-density peak. The latter water was assigned an arbitrary occupancy of 0.5, before the commencement of occupancy refinement with *SHELXL*; refinement to a lower occupancy value by *SHELXL* and the size of the peak reveals it is probably not due to a water molecule, although no other source for the electron-density peak was obvious from the chemical environment. The DB plots are a useful tool in identifying dubious water sites, and for locating possible metal-ion binding sites such as sodium which are hard to distinguish from electron-density maps alone.

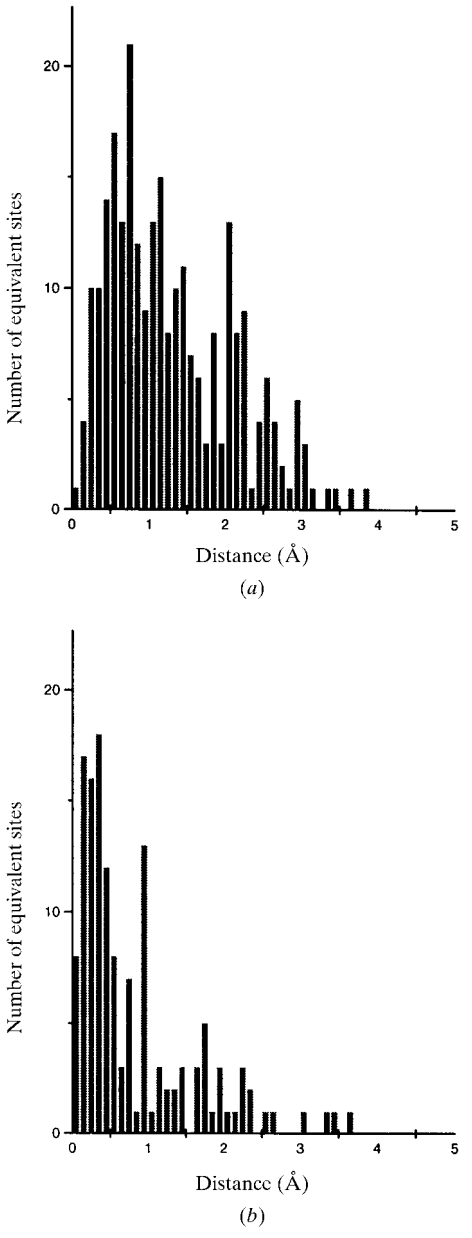


Fig. 11. Histogram of number of equivalent water molecules between (*a*) LTL and 2LZT, and (*b*) RTL and 2LZT.

Water occupancies were calculated directly from the DB plots computed before the introduction of partial occupancies and refinement carried out with these assigned values in *SHELXL*. After 15 cycles of refinement the *R* factor was 9.09% and the free *R* 10.99%, a decrease of 0.18 and 0.3%, respectively. This is a similar result to the previous refinement, although this time the number of parameters refined were less. Comparison of the occupancy values of the waters estimated from the DB plot with those output from direct refinement against the experimental data showed good agreement. A histogram of the difference ($OCC_{db} - OCC_{sh}$) where OCC_{db} is the occupancy calculated from the DB plot and OCC_{sh} from refinement with *SHELXL* is shown in Fig. 10. The distribution is not symmetric, OCC_{db} being generally greater than OCC_{sh}. A mean difference of 0.028 with a standard deviation of 0.076 was derived from fitting a Gaussian function in the distribution. The outliers with occupancy differences greater than 40% (eight water sites) were inspected with $3F_o - 2F_c$ electron-density maps and these large differences were due as expected to poor density for five of the sites, the other three sites had good density. Three sites differed in occupancy by more than 50% and these were clearly badly defined water sites, the highest difference of 0.6 was for w2272. Although further analyses with other protein data sets at atomic resolution need to be performed, these results clearly show estimation of water occupancies from DB plots can provide a more accurate picture of the solvent structure of proteins at lower resolution.

A total of 250 water molecules were included in the final model. The average occupancy for all waters was 0.8. Taking site occupancies greater than 0.95 to be unity gives 71 fully occupied water sites. 19 waters had site occupancies of less than 0.5. The total water stoichiometry for the final model is 197.9. Taking account of the space taken up by the nitrate and acetate ions leaves room for approximately 30–35 additional solvent sites based on the approximate estimate of space being available for 269 fully occupied water molecules. Electron-density peaks at the 5σ level or greater in $F_o - F_c$ maps was used as the cut-off value for selection of solvent peaks in the refinement.

Comparison of the LTL solvent structure to that of 2LZT shows some agreement of water sites, although a better agreement between the water sites from the two structures was expected, Fig. 11(*a*). After least-squares superposition of the two coordinate sets using CA atoms only, 50 water sites deviate by less than 0.5 Å and 142 sites less than 1.2 Å between the two models. 51 sites deviate by more than 2 Å. This comparison is made difficult because of the two different temperatures used in the collection of the respective data sets. The main-chain atoms of LTL and 2LZT deviate by an r.m.s. of 0.3 Å. Assuming waters that agree within 1 Å between the two structures to be equivalent results in 48% (119

sites) of the waters in LTL being paired with waters from the room-temperature 2LZT.

The solvent structure at the sugar binding site in the triclinic form was compared with the crystal structure of HEWL in the tetragonal form complexed with tri-*N*-acetylchitotriose, coordinate set 1LMB (Maenaka *et al.*, 1995) from the PDB (Bernstein *et al.*, 1977), to assess if the solvent structure mimics the bound substrate in any way. The *N*-acetyl functional group of the sugar bound at sub-site *C* of the six-site binding cleft was found to superimpose on one of the bound acetate ions (Ace203) in LTL. The only other site where the solvent appears to mimic the substrate is a water molecule (w2179) which is more or less positioned at the O3 atom of the sugar ring bound at sub-site *C*. The CG, CD, NE, CZ atoms of Arg5 from a symmetry-related molecule take up or mimic the positions of the O3—C3—C4—O4 atoms of the sugar ring bound at sub-site *A*. The other solvent sites at the binding cleft showed no similarity to the bound substrate of the above crystal structure.

5.5.2. *Room-temperature structure*. The same protocol that was used for the location of solvent sites in LTL was applied, but located only 124 water molecule sites during isotropic refinement. A total of 139 waters were included in the final RTL model. This amounts to only 56% of the total number of water sites found in LTL and 2LZT. Analysis of the Fourier peaks greater than 4σ in the final $F_o - F_c$ electron-density maps, showed no discernible peaks that could be attributed to water molecules.

The overall quality of the room-temperature X-ray data is poor in comparison with that of LTL. The overall $R_{\rm merge}$ of 6% is high for $P1$ symmetry. Scaling of the low-resolution data (21–2.5 Å) gave an overall $R_{\rm merge}$ 4.3%. One would expect better statistics for triclinic data. This is reflected in the electron-density maps, which are not as well defined as one would expect for data to this resolution.

Refinement of the water site occupancies was also carried out in the last stages of the refinement. This

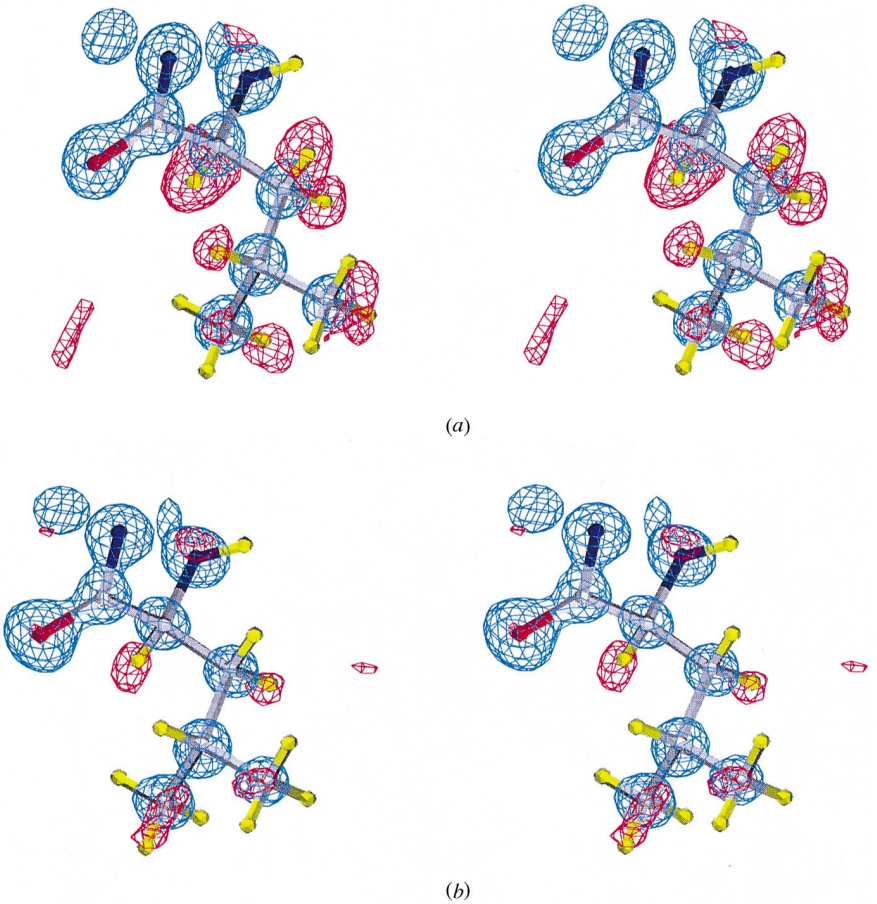


Fig. 12. Stereo pictures of $3F_o - 2F_c$ (blue) and $F_o - F_c$ (green) electron density for Leu8 contoured at 2.39 (3.0σ) and 0.21 e Å^{-3} (2.4σ), respectively, for (*a*) LTL and (*b*) RTL. The maps were calculated after randomizing the final model coordinates by 0.3 Å, and re-refining (H atoms omitted) with *SHELXL* until convergence (15 cycles). H atoms are coloured yellow.

resulted in a 0.3% drop in R factor and 0.21% drop in free R. Taking site occupancies greater than 0.95 to be unity, results in 44 fully occupied water sites. Nine waters have occupancies of less than 0.5. The total water stoichiometry for the final model is 111.96

Comparison of the solvent models of RTL and 2LZT is shown in Fig. 11(b). In this case there is a higher concentration of equivalent sites than for the LTL/2LZT comparison, in spite of the lower number of water sites in RTL. 77 water molecules deviate by less than 0.5 Å and 104 by less than 1.1 Å. 35 water sites (25%) from RTL deviate by more than 1.1 Å when compared with 2LZT.

Comparison of the water structure at the substrate binding site again showed some correlation with the polar functional groups of the bound tri-*N*-acetylchitotriose in tetragonal HEWL 1LZB. There is one water at the *N*-acetyl functional group of the *N*-acetyl-D-glucosamine molecule bound at sub-site *C* and two waters are found close to the hydroxyl groups O1 and O3. Waters in the binding crevice show good agreement between the two structures.

5.6. *Visualization of H atoms*

Although H atoms were added to the model at their theoretically calculated positions, electron-density peaks were visible for these atoms before they were added to the LTL and RTL models. Randomization of the final coordinates and re-refinement until convergence (H atoms omitted) produced even clearer difference electron-density peaks for the H atoms. Electron density for Leu8 which is situated in a well defined region of the protein structure is illustrated in Fig. 12. Electron-density peaks for the CA H atoms are visible for both LTL and RTL in the $F_o - F_c$ maps contoured at 0.21 e Å^{-3} (2.4σ). The CG and CB H atoms are visible in LTL, whereas only one of the CB H atoms is visible in RTL. Two of the six methyl H atoms of the CD C atoms are visible in LTL, RTL shows no peaks at the contour level shown in Fig. 12. Lowering the contour level to 1.9σ (0.17 e Å^{-3}) revealed electron density for the amide N atoms and all of the methyl H atoms of CD2 in LTL. In general H atoms were visible far more clearly in LTL than in RTL. Apart from the use of cryo-temperatures, the collection of the X-ray data from three crystals for RTL is another probable cause for the lower number of electron-density peaks for H atoms.

Electron density was visible for 77 out of 127 of the peptide H atoms in LTL at the 1.8σ level (0.16 e Å^{-3}). A similar percentage of CB H atoms was also visible. H atoms were visible in a rather random fashion for other atoms in the structure. As expected, well defined rigid regions of the structure were more likely to have visible electron density for H atoms. This last point is strikingly

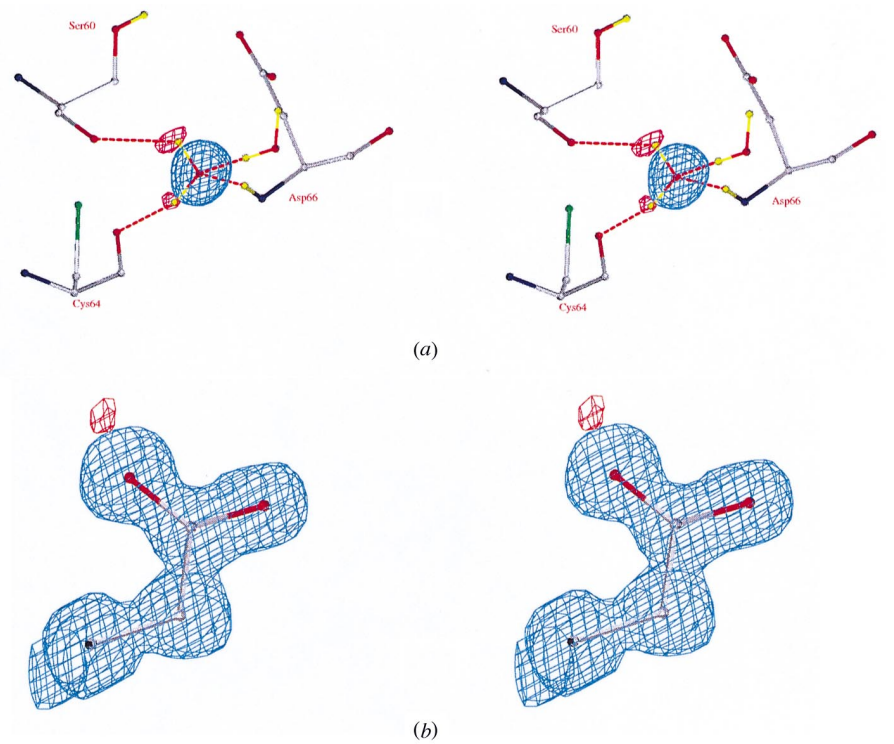


Fig. 13. (a) $3F_o - 2F_c$ (blue) and $F_o - F_c$ (red) electron density for w2005 contoured at 1.04 (1.3σ) and 0.22 e Å^{-3} (2.5σ), respectively. (b) $3F_o - 2F_c$ (blue) and $F_o - F_c$ (red) electron density for Glu7 contoured at 1.04 (1.3σ) and 0.18 e Å^{-3} (2.0σ), respectively. The maps were calculated as in Fig. 12

highlighted by Fig. 13(*a*) which clearly shows two electron-density peaks for the two H atoms of a tightly bound water molecule in LTL.

Protonation of the carboxylate groups of aspartate and glutamate side chains was investigated. HEWL contains seven aspartate and two glutamate residues. Each CG—OD1, CG—OD2 and CD—OE1, CD—OE2 bond length was measured for the aspartate and glutamate residues, respectively. To improve the accuracy of the measurements, the refined coordinates were randomized by 0.3 Å and re-refined until convergence (15 cycles). The above analysis indicated that the carboxylate groups of Asp18, Asp101 and the two glutamate residues are protonated. Electron density was visible for the H atom of Asp101 and Glu7, but no density was visible at the catalytic residue Glu35. Fig 13(*b*) shows the electron-density peak for the protonated Glu7 residue.

6. Conclusions

The present study describes the protocol used for the refinement of HEWL at atomic resolution, and highlights the merits of collecting data at cryogenic temperatures. The protocol shows that once a relatively reliable lower resolution model is available, the amount of manual intervention at least in the isotropic stages of refinement should be minimal or non-existent. The strategy implemented in the refinement of both the room- and low-temperature structures consists of a number of well defined steps which have been described previously (Sevcik *et al.*, 1996; Sheldrick *et al.*, 1997) but with some minor alterations. Addition of double conformations was performed during the anisotropic refinement. Identification of some minor alternate conformations was aided by first refining the atom occupancies of the major conformation, resulting in well defined difference density peaks for the alternate position of the side chain. Modelling of solvent structure was for the most part treated in an automated manner. The vast majority of water sites were automatically selected using *ARP* in the early isotropic stage. Waters hydrogen bonded to residues in alternate conformations had their occupancies linked to the relevant conformation. Refinement of the remaining water site occupancies was cross validated by the use of the free *R* and was shown to improve the agreement of the distribution of the water sites in the resulting DB plot. Estimation of water occupancies from their DB plots gave good correlation with the experimentally refined values.

The refinement of lysozyme at both room and low temperature reveals the structures to be essentially identical. The use of cryotechniques allowed complete data to be collected on one lysozyme crystal. The resolution of the cryogenic experiment was limited by geometrical constraints. The $I/\sigma(I)$ ratio of about 5 for the reflection intensities in the higher resolution shell clearly indicates that the crystals diffract beyond 0.9 Å resolution.

Although triclinic HEWL is tightly packed, 24 disordered side-chain residues were identified as well as three sections of main-chain disorder, accounting for approximately 20% of the protein structure. Identification of side chains with different rotamer conformations in the RTL was not so clear-cut. Only a third of the disordered residues in LTL were located. Modelling of the solvent structure was made easier by the use of cryogenic data. The resultant electron-density maps were far superior to those of the room-temperature analyses. Acetate ions were modelled in LTL, these ions were not detected in previous studies or in RTL. Trp62 which had no distinguishable electron density in the previous room-temperature structure, had well defined electron density for both the room- and low-temperature structures. The use of synchrotron radiation and current two-dimensional detector technology allowed the data to be collected in a far shorter time than in the previous study. The effects of prolonged exposure to X-radiation are considerably reduced or effectively eliminated by cryo-cooling of the crystal as is reported in other cryocrystallographic studies. The use of short less damaging wavelength radiation at the synchrotron provides an additional advantage.

The drop in the overall displacement parameter (B_{eq}) for the protein at low and room temperature in triclinic lysozyme is not large. The protein molecules are well packed and display much lower than average B_{eq}'s than is observed for other proteins at room temperature. Nevertheless, the results confirm that collection of X-ray data at cryogenic temperatures results in better data, clearer electron-density maps and allows both identification and modelling of the flexible parts of the structure.

X-ray data collected to this resolution provides the crystallographer with an extremely accurate model of the protein structure being refined. The large number of observations reduces model bias and allows better estimates of the errors in the refined structure. This has important consequences when assessing small but important perturbations in structure on for example, binding of a substrate or cofactor. Determination of the protonation states of charged amino acids which are involved in the catalytic mechanism of an enzyme either by direct measurement of the appropriate bond lengths and/or visualization of the actual H atom should provide invaluable evidence and clues for the system being studied. Features such as disorder in both the main- and side-chain atoms are easily identified and modelled at high resolution. Modelling of flexible loop regions is of importance as these regions frequently play a role in protein function and in protein–protein interactions. Extensive modelling of the solvent regions in the structure is also made possible. In this paper we have

shown that it is possible to refine water site occupancies, thus providing a more precise picture of the dynamics of the protein solvent interactions. This may allow the identification and assessment of both structurally and functionally important water sites in the protein. As further atomic resolution structures are refined, the resulting stereochemical information will provide better estimates of the types of restraints that should be used in the refinement of protein structures both at atomic and more importantly lower resolutions.

The authors are grateful to M. Dauter for crystallization. MAW was supported by an EC Institutional Fellowship, contract CT 930485. This work was supported in part by the EC BIOTECHNOLOGY contracts CT92–0524 and CT96–0189. Atomic coordinates and observed structure factors have been deposited with the PDB.†

References

Allen, F. H. & Kennard, O. (1993). *Chem. Des. Autom. News*, **8**, 31–37.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Blake, C. C. F., Johnson, L. N., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. (1967). *Proc. R. Soc. London Ser. B*, **167**, 378–388.

Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1965). *Nature (London)*, **206**, 757–761.

Brünger, A. T. (1992). *Nature (London),* **355**, 472–475.

Chacko, S., Silverton, E. W., Smith-Gill, S. J., Davies, D. R., Shick, K. A., Xavier, K. A., Willson, R. C., Jeffrey, P. D., Chang, C. Y. Y., Sieker, L. C. & Sheriff, S. (1996). *Proteins Struct. Funct. Genet.* **26**, 55–65.

Cohen, G. H., Sheriff, S. & Davies, D. R. (1996). *Acta Cryst.* D**52**, 315–326.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cosier, J. & Glazer, A. M. (1986). *J. Appl. Cryst.* **19**, 105–107.

Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1995). *Curr. Opin. Struct. Biol.* **5**, 784–790.

Ducruix, A. & Giegé, R. (1992). Editors. *Crystallization of Nucleic Acids and Proteins*, 1st ed. Oxford: IRL Press.

Hadfield, A. T., Harvey, D. J., Archer, D. B., MacKenzie, D. A., Jeenes, D. J., Radford, S. E., Lowe, G., Dobson, C. M. & Johnson, L. N. (1994). *J. Mol. Biol.* **243**, 856–872.

Hashimoto, Y., Yamada, K., Motoshima, H., Omura, T., Yamada, H., Yasukochi, T., Miki, T., Ueda, T. & Imoto, T. (1996). *J. Biochem.* **119**, 145–150.

† Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 4LZT, R4LZTSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: LI0269).

Hendrickson, W. A. & Konnert, J. H. (1980). *Incorporation of Stereochemical Information into Crystallographic Refinement*, pp. 13.01–13.23. Bangalore: Indian Academy of Sciences.

Hodsdon, J. M., Brown, G. M., Sieker, L. C. & Jensen, L. H. (1990). *Acta Cryst.* B**46**, 54–62.

Hooft, R. W. W., Sander, C., Vriend, G. & Abola, E. E. (1996). *Nature (London),* **381**, 272–272.

Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C. & Rupley, J. A. (1972). *The Enzymes,* 3rd ed. Vol. 7, edited by P. D Boyer, pp. 665–868. New York: Academic Press.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kodandapani, R., Suresh, C. G. & Vijayan, M. (1990). *J. Biol. Chem.* **265**, 16126–16131.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Kurachi, K., Sieker, L. C. & Jensen, L. H. (1975). *J. Biol. Chem.* **256**, 7663–7667.

Kurachi, K., Sieker, L. C. & Jensen, L. H. (1976). *J. Mol. Biol.* **101**, 11–24.

Kurinov, I. V. & Harrison, R. W. (1995). *Acta Cryst.* D**51**, 98–109.

Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.

Lamzin, V. S. & Wilson, K. S. (1997). *Methods Enzymol.* **277**, 269–305.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lehmann, M. S., Mason, S. A. & McIntyre, G. J. (1985). *Biochemistry.* **24**, 5862–5869.

McArdle, P. (1994). *J. Appl. Cryst.* **27**, 438.

MacArthur, M. W. & Thornton, J. M. (1996). *J. Mol. Biol.* **264**, 1180–1195.

Mckenzie, H. A. & White, F. H. (1991). *Adv. Protein Chem.* **41**, 173–315.

Maenaka, K., Kawai, G., Watanabe, K., Sunada, F. & Kumagai, I. (1994). *J. Biol. Chem.* **269**, 7070–7075.

Maenaka, K., Matsushima, M., Song, H., Sunada, F., Watanabe, K. & Kumagai, I. (1995). *J. Mol. Biol.* **247**, 281–293.

Mason, S. A., Bentley, G. A. & McIntyre, G. (1984). *Neutrons in Biology,* Vol. 27, edited by B. P. Schoenborn, pp. 323–334. New York: Plenum Press.

Moews, P. C. & Kretsinger, R. H. (1975). *J. Mol. Biol.* **91**, 202–228.

Nagendra, H. G., Sudarsanakumar, C. & Vijayan, M. (1995). *Acta Cryst.* D**51**, 390–392.

Oldfield, T. J. (1992). *J. Mol. Graphics*, **10**, 247–252.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276** 307–326.

Pontius, J., Richelle, J. & Wodak, S. (1996). *J. Mol. Biol.* **264**, 121–136.

Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Protein Chem.* **23**, 283–437.

Ramanadham, M., Sieker, L. C. & Jensen, L. H. (1989). *The Immume Response to Structurally Defined Proteins: The Lysozyme Model*, edited by S. J. Smith-Gill & E. E. Sercarz, pp. 15–24. New York: Adenine.

Ramanadham, M., Sieker, L. C. & Jensen, L. H. (1990). *Acta Cryst.* B**46**, 63–69.
Sevcik, J., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1996). *Acta Cryst.* D**52**, 327–344.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Sieker, L. C. (1988). *J. Cryst. Growth*, **90**, 31–38.
Steinrauf, L. K. (1959). *Acta Cryst.* **12**, 77–79.
Strynadka, N. C. J. & James, M. N. G. (1991). *J. Mol. Biol.* **220**, 401–424.
Teng, T. Y. (1990) *J. Appl. Cryst.* **23**, 387–391.
Vaney, M. C., Maignan, S., Riès-Kautt, M. & Ducruix, A. (1996). *Acta Cryst.* D**52**, 505–517.
Wilson, A. J. C. (1942). *Nature (London)*, **150**, 151–152.
Wilson, K. P., Malcom, B. A. & Matthews, B. W. (1992). *J. Biol. Chem.* **267**, 10842–10849.
Wilson, K. S., Dauter, Z., Lamzin, V. S., Walsh, M., Wodak, S., Richelle, J., Pontius, J., Vaguine, A., Hooft, R. W. W., Sander, C., Vriend, G., Thornton, J. M., Laskowski, R. A., MacArthur, M. W., Dodson, E. J., Murshudov, G., Oldfield, T. J., Kaptein, R. & Rullmann, J. A. C. (1998). *J. Mol. Biol.* **276**, 417–436.
Yamasaki, N., Tsujita, T., Eto, T., Masuda, S., Mizuno, K. & Sakiyama, F. (1979). *J. Biochem. (Tokyo)*, **86**, 1291–1300.
Young, A. C. M., Dewan, J. C., Nave, C. & Tilton, R. F. (1993). *J. Appl. Cryst.* **26**, 309–319.
Young, A. C. M., Tilton, R. F. & Dewan, J. C. (1994). *J. Mol. Biol.* **235**, 302–317.