Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Triclinic lysozyme at 0.65 Å resolution

Jiawei Wang,[a] Miroslawa Dauter,[a] Randy Alkire,[b] Andrzej Joachimiak[b]* and Zbigniew Dauter[c]*

[a]SAIC-Frederick Inc., Basic Research Program, Argonne National Laboratory, Argonne, IL 60439, USA, [b]Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, IL 60439, USA, and [c]Synchrotron Radiation Research Section, MCL, National Cancer Institute, Argonne National Laboratory, Argonne, IL 60439, USA

Correspondence e-mail: andrzejj@anl.gov, dauter@anl.gov

Received 5 September 2007
Accepted 29 October 2007

PDB Reference: triclinic lysozyme, 2vb1, r2vb1sf.

This work is dedicated to the pioneers of the refinement of proteins at high resolution, in particular of triclinic lysozyme: Lyle Jensen, Larry Sieker and Larry Steinrauf, in appreciation of the services they provided to the crystallographic community.

The crystal structure of triclinic hen egg-white lysozyme (HEWL) has been refined against diffraction data extending to 0.65 Å resolution measured at 100 K using synchrotron radiation. Refinement with anisotropic displacement parameters and with the removal of stereochemical restraints for the well ordered parts of the structure converged with a conventional R factor of 8.39% and an $R_{\rm free}$ of 9.52%. The use of full-matrix refinement provided an estimate of the variances in the derived parameters. In addition to the 129-residue protein, a total of 170 water molecules, nine nitrate ions, one acetate ion and three ethylene glycol molecules were located in the electron-density map. Eight sections of the main chain and many side chains were modeled with alternate conformations. The occupancies of the water sites were refined and this step is meaningful when assessed by use of the free R factor. A detailed description and comparison of the structure are made with reference to the previously reported triclinic HEWL structures refined at 0.925 Å (at the low temperature of 120 K) and at 0.95 Å resolution (at room temperature).

1. Introduction

Hen egg-white lysozyme (HEWL) provided the first three-dimensional crystal structure of an enzyme solved by X-ray crystallography in the laboratory of D. C. Phillips (Blake *et al.*, 2001). Initial results were obtained at 6 Å resolution (Blake *et al.*, 1962) and were followed by analysis of the model at 2 Å (Blake *et al.*, 1965). This initial structure of HEWL was solved in the tetragonal crystal form with space group $P4_32_12$. Since then, the tetragonal crystals of HEWL have become a classic and have been used in various projects advancing crystallographic methodology in addition to being a very convenient test example in many macromolecular crystallography schools and courses, owing to their properties such as easy availability, high symmetry and diffraction to high resolution. HEWL contains 129 amino acids, including eight cysteines forming four disulfide bridges.

HEWL shows 'classic' crystallization polymorphism and, depending on conditions, crystallizes in many forms. The Protein Data Bank (PDB; Berman *et al.*, 2000) currently (as of September 2006) contains 173 crystal structures of HEWL in the native state or in complex with various small ligands. Of these structures, 118 are in the most common tetragonal form (space group $P4_32_12$), 14 are orthorhombic ($P2_12_12_1$), 23 and five are in two monoclinic forms (both $P2_1$), one is hexagonal ($P6_122$) and 12 are triclinic ($P1$).

© 2007 International Union of Crystallography
Printed in Denmark – all rights reserved

Table 1
Ultrahigh-resolution protein structures and *P*1 lysozyme models available in the PDB in September 2006.

Protein	PDB code	Resolution (Å)	No. of residues	*R* factor	Reference
*P*1 lysozyme					
Real-space refinement	7lyz	2.50	129	N/A	Moult *et al.* (1976)
Refined by Fourier and LSQ	1lzt	1.97	129	0.254	Hodsdon *et al.* (1990)
Refined by restrained LSQ	2lzt	1.97	129	0.124	Ramanadham *et al.* (1990)
Neutron diffraction	1lzn	1.70	129	0.204	Bon *et al.* (1999)
Space group A1†	1lks	1.10	129	0.110	Steinrauf (1998)
At 120 K	3lzt	0.93	129	0.093	Walsh *et al.* (1998)
At room temperature	4lzt	0.95	129	0.114	Walsh *et al.* (1998)
Crystallized from D_2O	1v7s	1.14	129	0.090	Harata & Akiba (2004)
After phase transition†	1v7t	1.13	129	0.113	Harata & Akiba (2004)
Native, cross-linked	2f2n	1.60	129	0.138	Salem *et al.* (2006)
Treated with urea	2f30	1.65	129	0.156	Salem *et al.* (2006)
Treated with thiourea	2f4a	2.80	129	0.167	Salem *et al.* (2006)
Treated with bromoethanol	2f4g	1.65	129	0.178	Salem *et al.* (2006)
Ultrahigh-resolution structures					
Aldose reductase	1us0	0.66	316	0.094	Howard *et al.* (2004)
Serine protease	1gci	0.78	269	0.099	Kuhn *et al.* (1998)
Trypsin	1pq7	0.80	224	0.109	Schmidt *et al.* (2003)
Hydrophobin	2b97	0.75	142	0.130	Hakanpää *et al.* (2006)
CBM36	1w0n	0.80	131	0.129	Jamal-Talabani *et al.* (2004)
PAK pilin	1x6z	0.78	123	0.143	Dunlop *et al.* (2005)
HIPIP	1iua	0.80	83	0.092	Liu *et al.* (2002)
PDZ2 domain	1r6j	0.73	82	0.075	Kang *et al.* (2004)
Antifreeze protein RD1	1ucs	0.62	64	0.137	Ko *et al.* (2003)
Rubredoxin (*Pyrococcus abyssi*)	1yk4	0.69	52	0.100	Bonisch *et al.* (2005)
Rubredoxin (*Desulfovibrio gigas*)	2dsx	0.68	52	0.099	Chen *et al.* (2006)
Crambin	1ejg	0.54	46	0.090	Jelsch *et al.* (2000)
D,L-Alpha-1	3al1	0.75	26	0.130	Patterson *et al.* (1999)

† The unit cell is different from other *P*1 structures.

The most studied form of HEWL is therefore the tetragonal form, but the triclinic form also has a long history of single-crystal X-ray structural investigations (Table 1). Triclinic crystals of HEWL were first described by Steinrauf (1959). The structure was initially solved by multiple isomorphous replacement (Dickerson *et al.*, 1962) as well as by molecular replacement (Joynson *et al.*, 1970) at a very low resolution of 6 Å and was subsequently refined by Moult *et al.* (1976) at 2.5 Å resolution. The complexes of this form of HEWL with *N*-acetylglucosamine (NAG) and *N*,*N*-diacetylchitobiose (NAG_2; Kurachi *et al.*, 1976) as well as the native structure (Hodsdon *et al.*, 1990; Ramanadham *et al.*, 1990) were refined at 2.0 Å in the laboratory of L. Jensen. The crystal phase transition from the small triclinic unit cell to a twofold larger triclinic unit cell caused by controlled dehydration was studied and the structures were refined at 1.13 and 1.14 Å resolution by Harata & Akiba (2004). Steinrauf (1998) refined the native structure at 1.1 Å resolution and Walsh *et al.* (1998) extended the resolution to 0.95 Å with data collected at room temperature and to 0.925 Å with data measured at 120 K. Salem *et al.* (2006) investigated the influence of denaturing agents on triclinic HEWL crystals cross-linked with glutaraldehyde and refined the structures of the native crystals and of crystals soaked in urea, thiourea and bromoethanol at resolutions ranging from 1.6 to 1.95 Å. They extended the earlier similar studies on the effect of denaturants performed at the Weizmann Institute (Yonath, Sielecki *et al.*, 1977; Yonath, Podjarny *et al.*, 1977; Yonath *et al.*, 1978).

Triclinic crystals of HEWL have also been studied by neutron diffraction. Lehmann and coworkers investigated the interactions of lysozyme with ethanol at 2.4 Å (Lehmann *et al.*, 1985), with dimethyl sulfoxide at 1.8 Å (Lehmann & Stansfield, 1989) and with water at 1.7 Å (Bon *et al.*, 1999).

The highest resolution structures of HEWL in the PDB are two models of the triclinic form (Walsh *et al.*, 1998), with data measured at 120 K extending to 0.925 Å (PDB code 3lzt) and at room temperature extending to 0.95 Å (code 4lzt), and one model of the tetragonal form crystallized in space with data to 0.94 Å (PDB code 1iee; Sauter *et al.*, 2001). Indeed, according to expectations (Hakanpää *et al.*, 2006), the best diffracting form is triclinic, with only one molecule and about 27% solvent in the unit cell. Here, we demonstrate that triclinic lysozyme can diffract to a resolution of 0.65 Å, well beyond the 0.8 Å limit typically used to define ultrahigh-resolution structures.

The accuracy and amount of interpretable structural details in structures solved by X-ray crystallography are related to the number of measured observables, *i.e.* the resolution of the diffraction data. Most structures in the PDB are refined at a resolution in the range 1.5–2.5 Å. However, in the last decade an increasing number of macromolecular structures have been solved at atomic resolution, beyond the limit of 1.2 Å. In September 2006, 805 atomic resolution X-ray structures had been deposited in the PDB from a total of 32 867, 13 of which reached ultrahigh resolution of 0.8 Å or beyond (Table 1). This trend results from improvements in the techniques of molecular biology and from advances in crystallographic technology (synchrotron beamlines, large area detectors, crystal cooling, computing hardware and software) as well as from the increased interest of researchers in the intricate details of protein stereochemistry and electronic structure. Such detailed features determine the chemical and biochemical properties of macromolecules and are often very important for rationalization of their particular biological characteristics.

2. Materials and methods

2.1. Crystallization and data collection

Hen egg-white lysozyme (Sigma, three times recrystallized) was used without further purification and was crystallized by the sitting-drop method under conditions analogous to those specified by Steinrauf (1959). The protein was dissolved to 35 mg ml^{-1} in 20 m*M* sodium acetate buffer pH 4.7. The well

solution comprised 0.5 *M* $NaNO_3$, 0.1 *M* sodium acetate buffer pH 4.7 and 20% ethylene glycol. The drops were prepared by mixing 1.5 µl each of protein and well solutions and were microseeded with crushed (11-year-old) crystals of triclinic lysozyme. The crystals grew at 292 K during 5 d.

Diffraction data were measured at the 19ID beamline of the Structural Biology Center, Advanced Photon Source, Argonne National Laboratory using an ADSC Quantum315 CCD detector. The crystal, with dimensions of 0.4 × 0.4 × 0.5 mm, was larger than the beam size of 0.2 mm^2, but was not translated during data collection. It was directly cryocooled in a stream of cold nitrogen at 100 K. Several passes of crystal rotation were used, varying the effective exposure and crystal-to-detector distance, in order to adequately measure strong low-resolution intensities as well as the weakest high-resolution intensities. All data were integrated and merged using *HKL*-2000 (Otwinowski & Minor, 1997). Details of the data-collection strategy and the statistics of diffraction data are summarized in Table 2. The highest resolution limit of 0.65 Å resulted from geometrical limitations of the detector and beamline setup, since the crystal could potentially diffract to slightly higher resolution as evidenced by the data-merging statistics. An overall *B* factor of 3.5 $Å^2$ was estimated from the Wilson plot using the *TRUNCATE* program from *CCP*4 (Collaborative Computational Project, Number 4, 1994).

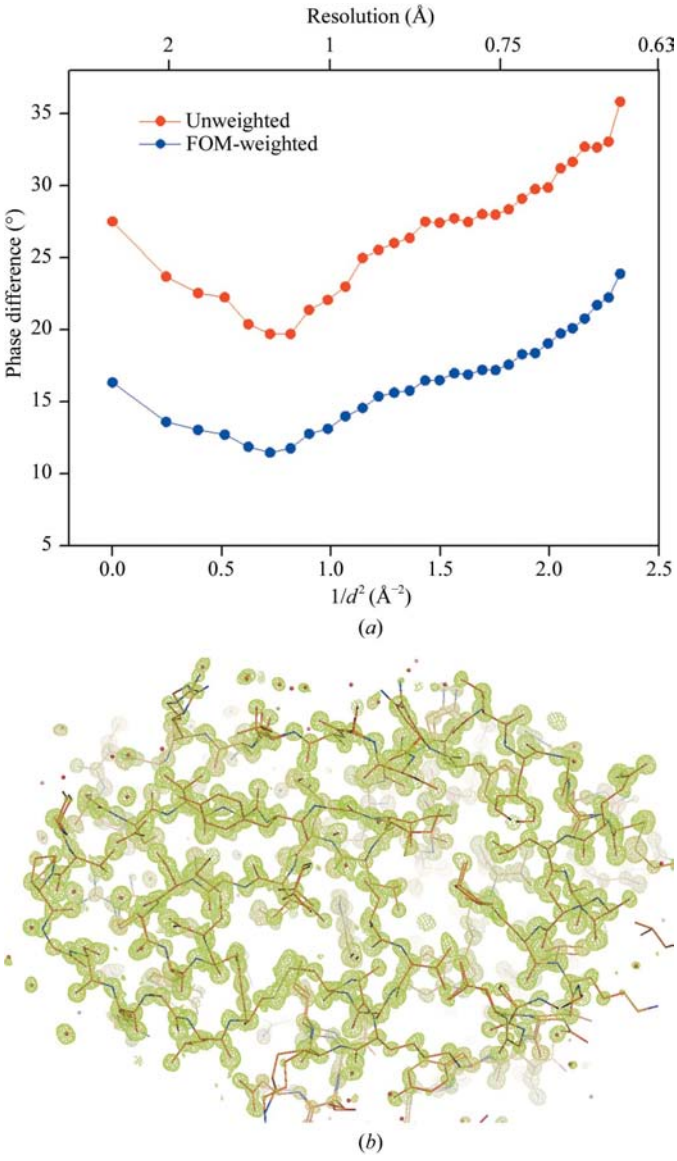

Figure 1
(*a*) Average unweighted (red dots) and FOM-weighted (blue dots) difference between the phases obtained from *ACORN* and from final refinement as a function of resolution. (*b*) Electron-density map from *ACORN* contoured at 2σ superposed with the final model.

The strategy of very high resolution data collection from triclinic crystals involves some special problems related to the issue of the blind region and the sparse distribution of symmetry-equivalent reflections in reciprocal space. The fraction of reflections lost in the blind region at the maximum resolution and therefore at the maximum diffraction angle θ_{max} [$= \sin^{-1}(\lambda/2d) = \sin^{-1}(0.6526/1.3) = 30.1^\circ$] is $(1 - \cos\theta_{max}) = 0.135$. Within all resolution ranges, this fraction amounts to $[1 - 3(4\theta_{max} - \sin 4\theta_{max})/(32\sin^3\theta_{max})] = 0.079$ (Dauter, 1999). In space group *P*1, with no rotational symmetry operations, the only way to measure the reflections missing in the blind region is to reorient the crystal on the goniostat. In order to avoid the loss of almost 8% of all reflections (up to 13.5% at the highest resolution), for two of the highest exposure passes of data collection the goniostat κ angle was set to $\pm 60^\circ$, which effectively skews the spindle axis by about 50° each way, a value larger than θ_{max}.

Data measured to high resolution with a high-flux synchrotron X-ray beam can potentially be influenced by the effects of crystal radiation damage. However, trials involving the scaling of all highest resolution data (sets *e*–*h*; see Table 2) with varying scales and *B* factors for individual diffraction images did not indicate any significant influence arising from damage. All refined values of *B* factors fluctuated within the range −0.11 to +0.27 $Å^2$ without showing any increasing trend. In the final scaling of all data the image *B* factors were set to zero and were not refined. All figures showing the model and electron density were prepared with *PyMOL* (DeLano, 2001).

2.2. Refinement protocol

2.2.1. Overall strategy. The availability of ultrahigh-resolution data to 0.65 Å makes *ab initio* phasing with *ACORN* possible (Foadi *et al.*, 2000). *ACORN* was run with 50 000 trials starting with a randomly positioned single atom. The correlation coefficient (CC) between the resulting calculated and observed *E* values was computed for each trial using the strongest 20% of the data. All trials were sorted by CC and the top 200 were then used to provide initial phase sets for dynamic density modification. Of these 200 trials, the first, with a CC of 2.7%, led to a successful structure determination after 22 cycles of density modification. The successful solution had an *R* factor of 0.42 (0.11 for the strongest *E* values) and a correlation coefficient of 58%. Fig. 1(*a*) shows the difference between the *ACORN* phase set and the final refined phases. The overall phase error was 26.7°, which indicates that the phases from *ACORN* were quite reliable. The corresponding

Table 2
Diffraction data statistics.

Data sets *a*–*h* were collected consecutively starting from the least exposed sets in order to measure the strongest intensities from the least radiation-damaged crystal. The wavelength used was 0.65 Å.

(*a*) Data collection.

Rotation pass	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*h*
Goniostat φ-angle (°)	0	180	0	180	0	180	0	0
Goniostat κ-angle (°)	0	0	0	0	0	0	−60	60
No. of images	60	60	120	120	360	360	270	270
Oscillation range (°)	3	3	1.5	1.5	0.5	0.5	0.5	0.5
Total ω rotation range (°)	180	180	180	180	180	180	135	135
Detector distance (mm)	400	400	200	200	100	100	100	100
Beam attenuation	20.6	20.6	8.9	8.9	3.1	3.1	2.4	2.4
Exposure time (s)	4	4	5	5	5	5	5	5

(*b*) Merging of all data. Values in parentheses correspond to the highest resolution range (0.67–0.65 Å).

No. of crystals	1
Beamline	SBC-CAT 19ID
Wavelength (Å)	0.6526
Space group	*P*1
Unit-cell parameters (Å, °)	a = 27.07, b = 31.25, c = 33.76, α = 88.0, β = 108.0, γ = 112.1
Resolution (Å)	30–0.65
No. of observations	1331953 (12764)
No. of unique reflections	187165 (6353)
Multiplicity	7.1 (2.7)
Completeness (%)	97.6 (67.3)
$R_{\rm merge}$ (%)	4.5 (18.4)
$\langle I\rangle/\langle\sigma(I)\rangle$	36.2 (4.2)
Wilson *B* factor (Å^2)	3.5

electron-density map was of very good quality (Fig. 1*b*), with many fine structural details such as fragments in alternate confirmations clearly identifiable in this model-phase-unbiased experimental map. The initial model was constructed by *RESOLVE* (Terwilliger, 2003*a*,*b*) with only a single chain, independent of any existing structures in the PDB, but indications of flexible conformations available in the unbiased map were referenced in the subsequent model refinement in order to minimize the negative effect of the model bias. The refinement process utilized the programs *phenix.refine* (Adams *et al.*, 2002) and *SHELXL* (Sheldrick & Schneider, 1997).

Although it is mostly the core-electron scattering that persists to high values of $\sin(\theta/\lambda)$, the low-resolution data provide important information on the solvent structure around the protein and also about the atomic valence electrons that are diffuse from atomic centres (Coppens, 1997). The low-resolution data are therefore necessary to correctly model the finest features in electron density and should be measured and used to the lowest resolution attainable, ideally 20 Å or lower (Walsh *et al.*, 1998). All reflections were therefore included without σ cutoff in the final refinement, except for the 5% of the data that were randomly selected and used for cross-validation analysis by means of the $R_{\rm free}$ factor (Brünger, 1997). Disordered residues were modeled and one common occupancy value was constrained for individual fragments of flexible main chains and side chains. The sum of occupancies for all disordered fragments was constrained to unity. H atoms were not refined but were included at idealized positions riding on their parent atoms, which introduced no additional parameters into the refinement process. H atoms in hydroxyl, amine and methyl groups were placed optimally according to the 'circular difference Fourier' as implemented in *SHELXL*. No H atoms were introduced for water molecules. Eventually, all occupancies of water molecules were refined. After each of the refinement steps, the model was inspected and rebuilt manually using the program *Coot* (Emsley & Cowtan, 2004). The main steps in the progress of refinement are summarized in Table 3.

Table 3
Outline of the refinement steps of the ultrahigh-resolution structure of HEWL to 0.65 Å resolution.

The refinement was performed with the software *phenix.refine* (P) and *SHELXL* (S). All reflections in the indicated resolution range were included in the calculation, except for a 5% test set for cross-validation with free *R*. In the final step, all data were used (187 165 reflections).

Round	Action taken	Program	Resolution (Å)	*R* (%)	$R_{\rm free}$ (%)
1	Simulated annealing	P	30–1.5	24.83	27.28
2	Isotropic, add solvent	P	10–1.1	15.17	16.80
3	Same atoms, isotropic	S	10–1.1	14.75	16.83
4	All-atom anisotropic	S	10–1.1	11.59	14.52
5	Add disorders (add Leu129)	S	10–1.1	10.93	14.00
6	Change resolution	S	8–1.0	11.22	13.75
7	Isotropic, water	P	8–0.65	16.78	17.24
8	Same atoms, isotropic	S	8–0.65	16.56	17.53
9	Anisotropic, add disorders	S	8–0.65	10.75	11.86
10	Isotropic, water	P	30–0.65	16.95	17.55
11	Same atoms, isotropic	S	30–0.65	16.33	17.24
12	Anisotropic, add disorders	S	30–0.65	10.71	11.69
13	Minor adjustments	S	30–0.65	10.10	11.12
14	Riding hydrogens added	S	30–0.65	9.16	10.04
15	Minor adjustments	S	30–0.65	9.00	9.95
16	Add flexible loop	S	30–0.65	8.71	9.63
17	Weighting changed	S	30–0.65	8.65	9.62
18	Restraints removed	S	30–0.65	8.48	9.59
19	Water occupancies refined	S	30–0.65	8.39	9.52
20	Free *R* removed	S	30–0.65	8.39	—

2.2.2. Details of the refinement. Leu129 and the side chain of Arg128 were ill-defined and were left out before the start of refinement. The first stage of refinement using the simulated-annealing method in *phenix.refine* converged to an *R* factor of 24.83% and an $R_{\rm free}$ of 27.28% at 1.5 Å resolution, with an isotropic displacement parameter of 6.5 Å^2. Least-squares refinement was performed with automatic water picking by *PHENIX*, resulting in an *R* and an $R_{\rm free}$ of 15.17% and 16.80%, respectively, to a resolution of 1.1 Å. Switching of isotropic refinement from *PHENIX* to *SHELXL* and a first round of anisotropic refinement resulted in a sharp decrease of the *R* factor to 11.59% and of $R_{\rm free}$ to 14.52%. A detailed inspection of the model was then made and side chains that were either disordered or had multiple conformations were identified. At the same time, Leu129, the side chain of Arg128 and several nitrate/acetate and ethylene glycol molecules were added to the model. At this stage, if the resolution was extended directly to the highest resolution of 0.65 Å the model bias in the raw model probably made the refinement procedure difficult, preventing it from detecting the minor alternate

conformations in the model-phase-biased ($F_o - F_c$, φ_c) maps. A new refinement strategy was therefore applied.

The experiences of Coppens (1997) and Pichon-Pesme *et al.* (2000) of working at ultrahigh resolution suggest that it is advantageous to omit the lowest resolution data from the initial stages of refinement in order to obtain properly converged atomic positions that are not influenced by information about the diffuse bonding electrons. The low-resolution data should be included later to model the finer features. Isotropic refinement was again performed using *PHENIX*, but with alternate conformations retained in the model and in the resolution range 8–0.65 Å. After isotropic refinement with automatically picked waters, R and R_{free} were 16.78% and 17.24%, respectively. Subsequent anisotropic refinement with *SHELXL* lowered R and R_{free} to 10.75% and 11.86%, respectively. The use of all data in the full resolution range 30–0.65 Å converged at values of 10.10% and 11.12% for the R factor and R_{free}, respectively. H atoms were added to the model according to geometrical criteria implemented in *SHELXL*. This yielded a decrease of 0.94% in the R factor and of 1.08% in R_{free} without increasing the number of parameters. Optimizing the weighting scheme in *SHELXL* and removing the restraints for the well ordered parts of the model (those in a single conformation) lowered the R factor to 8.48% and R_{free} to 9.59%. After refining the occupancies of all water molecules, R and R_{free} fell by 0.09% and 0.07%, respectively. In the final stage, all data were included in the refinement, including the R_{free} set, resulting in a final R factor of 8.11% for 175 025 reflections with $F_o > 4\sigma(F_o)$ and 8.39% for all data (187 165 reflections) in the full resolution range 30–0.65 Å.

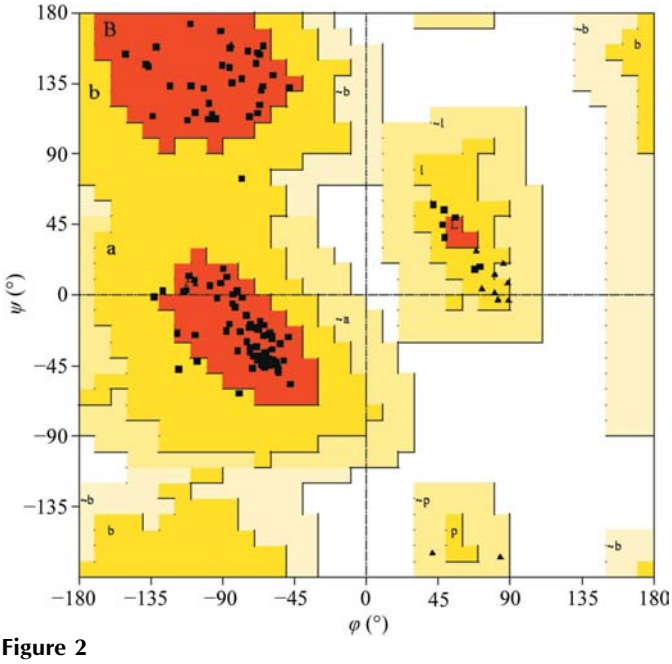

Figure 2
Ramachandran plot. Glycine residues are represented by triangles. 88.5% of the residues are located in the most favored regions, 11.5% are in the additional allowed regions and no residues are found in the generously allowed or disallowed regions of the plot.

Table 4
Final model statistics.

Resolution range (Å)	30–0.65
R factor for 187 165 reflections (%)	8.39 (8.11)†
Free R factor for 9365 reflections (%)	9.52 (9.24)†
No. of protein atoms (non-H)	1323
No. of solvent/other entity atoms	170/52
R.m.s. deviations from ideal geometry‡	
Bond distances (Å)	0.023 (0.016)
Bond angles (°)	2.4 (2.0)
Planar torsion angles (°)	6.9 (6.8)
Mean B_{eq} (Å^2)	
All atoms	7.1
Protein atoms	6.0
Solvent atoms	14.3

† R factors in parentheses are for reflections with $F_o > 4\sigma(F_o)$. ‡ Values determined using *WHATCHECK* (Hooft *et al.*, 1996) after final unrestrained refinement and (in parentheses) after final restrained refinement.

At the end of the refinement, one cycle of full-matrix minimization was calculated with all stereochemical restraints removed and damping all parameter shifts to zero using the L.S. 1/BLOC 1 combination in *SHELXL*, which permitted the estimation of uncertainties in all positional parameters.

3. Results and discussion

3.1. Quality of the structure

The final protein model is of very high quality and consists of 1323 protein-atom sites for 1001 non-H lysozyme atoms and also 170 water molecules, nine nitrate ions, one acetate ion and three ethylene glycol molecules. 45 side chains were modeled with alternate conformations as well as eight sections of main chain. The electron density for the polypeptide chain is excellent, with the exception of the C-terminus, which is quite flexible with an average B_{eq} of 25.98 Å^2 for the protein atoms of Arg128 and Leu129, compared with the average B value over the whole protein of 5.93 Å^2. A plot of the φ/ψ angles in the polypeptide backbone (Ramachandran plot) prepared with *PROCHECK* (Laskowski *et al.*, 1993) is shown in Fig. 2 (Ramachandran & Sasisekharan, 1968). 88.5% of the residues are located in the most favored regions, 11.5% of the residues are located in the additional allowed regions and no residues are found in the generously allowed and disallowed regions of the plot. The quality of the final model is summarized in Table 4.

Diffraction data to subatomic resolution permit correct identification of the atomic types directly from the electron-density map. A plot of (F_o, α_c) density interpolated at atomic centres as a function of atomic temperature factors (DB plot; Sevcik *et al.*, 1996) is given in Fig. 3(*a*) for protein atoms with unit occupancy. The S, O, N and C atoms are clustered and distinct from each other. An empirical simulation formula was proposed as $\rho^{centre}(B) = a \times (B/4\pi)^{-b}$. Least-squares fitting suggested that the parameter b is approximately a constant of 0.75 for all elements and the parameter a is 3.78, 4.49, 5.48, 14.21 for C, N, O and S, respectively, proportional to the individual atomic number of each element. The final prediction curves are shown in Fig. 3(*a*) as black dashed lines that are

in good agreement with the observed scattering points. However, because all the occupancies of water O atoms were refined, most of which deviate from unity, the density at the centre of the water molecules needs to be normalized by the atomic occupancy before comparing with the oxygen curve. Figs. 3(*b*) and 3(*c*) show the DB plots for water before and after the above-mentioned normalization. As expected, after normalization the DB plot of water sites shows an improved agreement with the theoretical prediction.

The final refinement performed using the full-matrix algorithm with the *SHELXL* keyword DAMP 0 0 provides the best estimate of the quality of the model. Fig. 4(*a*) shows a plot of coordinate error *versus* residue number for main-chain atoms, which shows almost the same pattern as that for the atomic resolution data (PDB codes 3lzt and 4lzt). The estimated average r.m.s. coordinate error is 0.018 Å for protein atoms and 0.02 Å for all atoms.

Very high resolution diffraction data collected at a very intense X-ray beam can be influenced by the effects of radiation damage. However, in the final difference Fourier map there were no features higher than the noise level in the vicinity of the S atoms involved in disulfide bridges that would indicate the breakage of these bonds, which are the most susceptible to damage. As stated previously, there were also no indications of radiation damage in the data-merging statistics. It is most likely that the effects of radiation damage were minimized by the fact that the beam cross-section was smaller than the size of the lysozyme crystal, which was characterized by a very low solvent content (about 27%).

3.2. Model stereochemistry

The high quality of the model, together with the unrestrained refinement of well ordered parts and the availability of estimates of parameter errors, allows detailed analysis of the stereochemical parameters of the protein. For well defined parts of the structure (Fig. 4*b*), individual protein atoms were identified in the ($3F_o - 2F_c$) map by resolved peaks at a density level of $\sim 2\sigma$. In the following sections, the final model is compared with the 0.925 Å low-temperature (3lzt) and 0.95 Å room-temperature (4lzt) atomic resolution structures.

Root-mean-square deviations between the structures based on least-squares superposition of the 3lzt, 4lzt and current

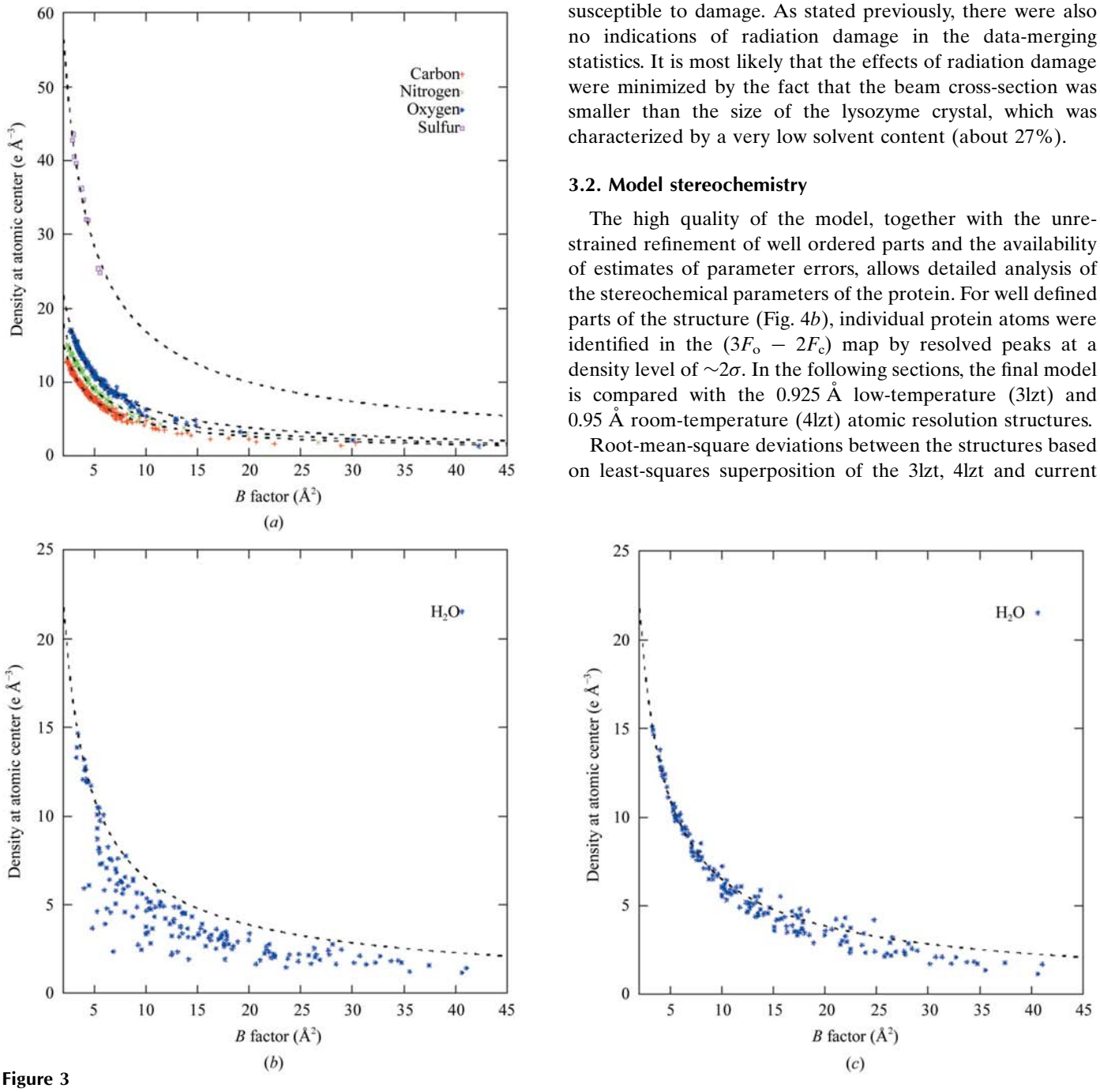

Figure 3
(*a*) (F_o, α_c) electron densities at atomic centres as a function of temperature factor (DB plot) for all protein atoms with unit occupancy. (*b*) DB plot for water molecules before their densities were normalized by occupancies. (*c*) The same as (*b*) except that normalization has been performed.

Table 5
Least-squares superpositions of lysozyme coordinate sets 3lzt and 4lzt onto the current model.

Only residues in single conformations were used in the statistics. The root-mean-square and average deviations are given in angstroms.

	3lzt			4lzt		
	No. of atoms	R.m.s. deviation	Average deviation	No. of atoms	R.m.s. deviation	Average deviation
C^{α} atoms	74	0.16	0.13	78	0.24	0.20
Main chain	295	0.16	0.13	312	0.24	0.21
All atoms	525	0.34	0.17	567	0.42	0.26

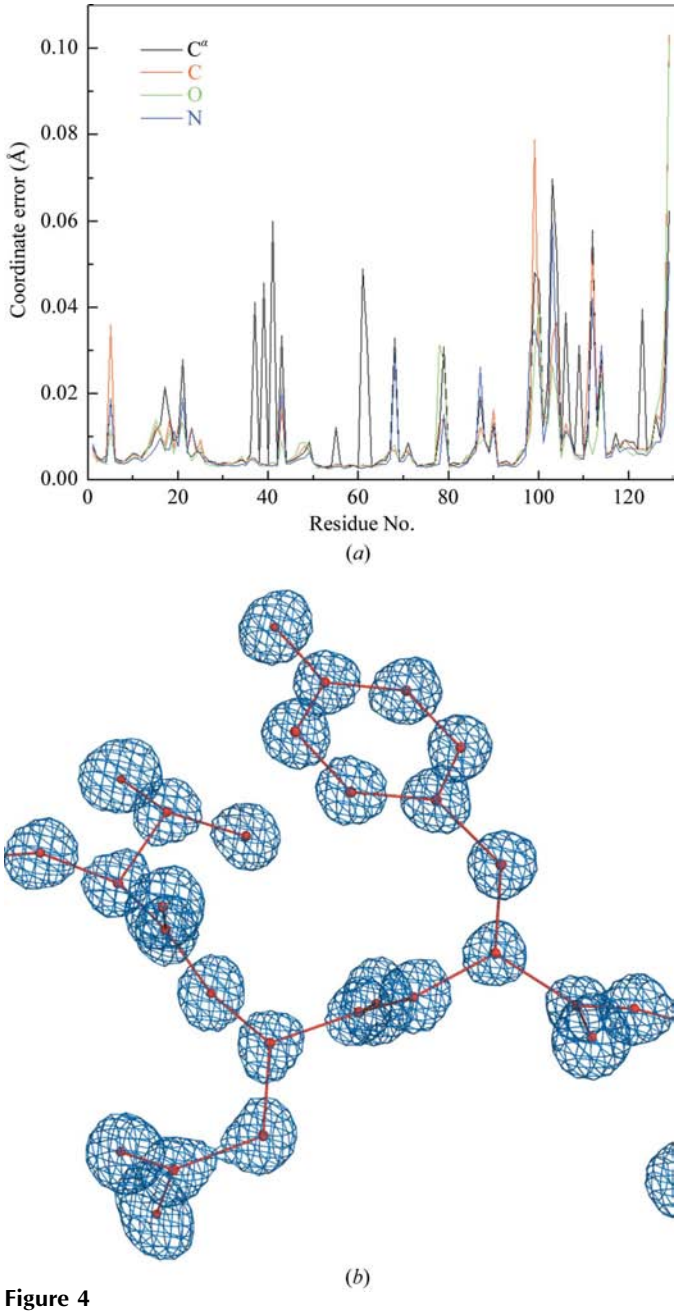

Figure 4
(*a*) Coordinate errors for main-chain atoms in the major conformation estimated from inversion of the least-squares matrix. (*b*) $3F_o - 2F_c$ electron density around Asp52, a residue that participates in the catalytic activity of the enzyme, contoured at a level of 1.8σ.

model are given in Table 5. Despite the fact that only atoms in single conformations were used for the superposition of these models, the level of deviation of the main-chain atoms exceeds the accuracy of the atomic positions obtained (for 3lzt and the current structure) from the inversion of the least-squares matrix by about tenfold. In side chains, this level of conformational variation is about twofold higher. Clearly, the presence of approximately one-third of the structure in double conformations confirms that even a rather rigid molecule such as lysozyme is characterized by a considerable degree of flexibility and can adopt somewhat different conformation depending on small differences in the protocols used for crystallization, cryocooling and crystal handling.

3.2.1. N—C^{α}—C angles. The main-chain N—C^{α}—C valence-angle restraint is implemented in *SHELXL via* the 1–2 (N—C^{α}, C^{α}—C) bond distances and the 1–3 (N—C) angle distance, which gives a target value of 111.2° for this angle. However, it has been noted that the distribution of the N—C^{α}—C angle is quite wide (Esposito *et al.*, 2000) and a large deviation from the target value has sometimes been interpreted as being correlated with the conformation of the main chain. In Fig. 5(*a*), the N—C^{α}—C angles in the present structure are analyzed, where red column bars represent the distribution of all N—C^{α}—C angles and blue bars refer only to the unrestrained single-conformation angles. The 93 valence angles in the blue histogram were not restrained to the target value and they are rather widespread, with a mean value of 111.8° and a standard deviation of 2.2°. The outliers present in the red histogram correspond to fragments with main-chain disorder, especially the largest deviation for Asn103 (N*A*—C^{α}*A*—C*A*), which has an angle of 122.4°. Inspection of the model indicated that partial peptide flipping occurred at this residue. The two outliers in 3lzt (Fig. 5*b*) are also correlated with Asn103 owing to its peptide flipping; a single site with C^{α} and peptide-flipped C atoms produced stress in the main chain with abnormal values of 140.1° and 100.4° (Fig. 5*c*). As shown in Fig. 5(*d*), in the current structure not only the C^{α} atom of Asn103 but also its side chain was modeled with double conformations. However, the ND2 atom of Asn103 is hydrogen-bonded to Ile98 O, which drags the main chain of Ile98–Asp101, which is part of the α-helix, into alternate conformations. According to the refinement of the occupancies, Ile98*A* should be hydrogen-bonded to Asn103*A* (green line in Fig. 5*d*), as their refined occupancies are very similar, with values of 0.56 and 0.53, respectively, even though they were refined separately. The peptide-flipped Asn103*B* is also linked to Asn106*B* *via* a water molecule, with similar occupancies of 0.47 and 0.40. In 4lzt, Asn103 was modeled with a single conformation and there is no outlier for this residue. The only outlier in 4lzt (not shown) is the C-terminal residue Leu129, which is quite flexible, as mentioned earlier.

3.2.2. Planarity of peptide bonds. The distribution of the ω angles, which characterize the configuration of the peptide bonds, is shown in Fig. 6(*a*). The blue histogram represents the distribution of 75 ω angles of the unrestrained single-conformation residues and the red bars refer to all residues. It is clear that the distribution of the ω angles is very wide

whether the double conformations are included or not and shows that some peptide groups deviate from planarity by up to 27°. The mean value and standard deviation of the distribution were 179.8° and 7.3°, respectively, when residues with double conformations were included, and 178.4° and 5.8° when they were excluded. The latter values are unbiased, as this angle was not restrained in the refinement by the usual requirement of planarity for the well ordered peptide. Although planarity restraints were kept for Trp62 in connection with its alternate conformations discussed below in detail, the largest deviation from planarity occurred at this peptide bond ($C^{\alpha}62A-C62-N63-C^{\alpha}63$), with an ω angle differing from the ideal *trans* conformation by 27.3°. The average standard uncertainty of the ω values as estimated by the inversion of the least-squares matrix is 0.29° (1.11° with double conformations included).

3.2.3. The χ^1 angles. No restraints were imposed on χ^1 angles during refinement with *SHELXL*. The χ^1 torsion angles of side chains are clustered into three populations corresponding to three preferred rotamers (Fig. 6*b*). The mean values and standard deviations (in parentheses) of these peaks are 66.7 (5.1), 181.7 (7.8) and 293.0 (8.6) when residues with double conformations were excluded and 66.4 (13.2), 184.3 (10.5) and 293.6 (10.5) when they were included.

3.2.4. Conformation of Trp62. Trp62 plays an important role in binding the enzyme substrate in the sugar B subsite (Blake *et al.*, 1967). It is stacked against the glucopyranose ring of *N*-acetyl-chitotriose (GlcNAc) when GlcNAc binds to site B of the active-site cleft (Blake *et al.*, 2001). However, owing to the flexibility of this residue, poor electron density is usually observed in all crystal forms of native HEWL. No electron density is visible in the refined 2 Å resolution triclinic lysozyme (Hodsdon *et al.*, 1990). The threefold to fourfold higher B_{eq} values in the 4lzt model than the 3lzt model also indicate the high mobility of this residue (Walsh *et al.*, 1998). However, in the subatomic structure a double conformation is clearly

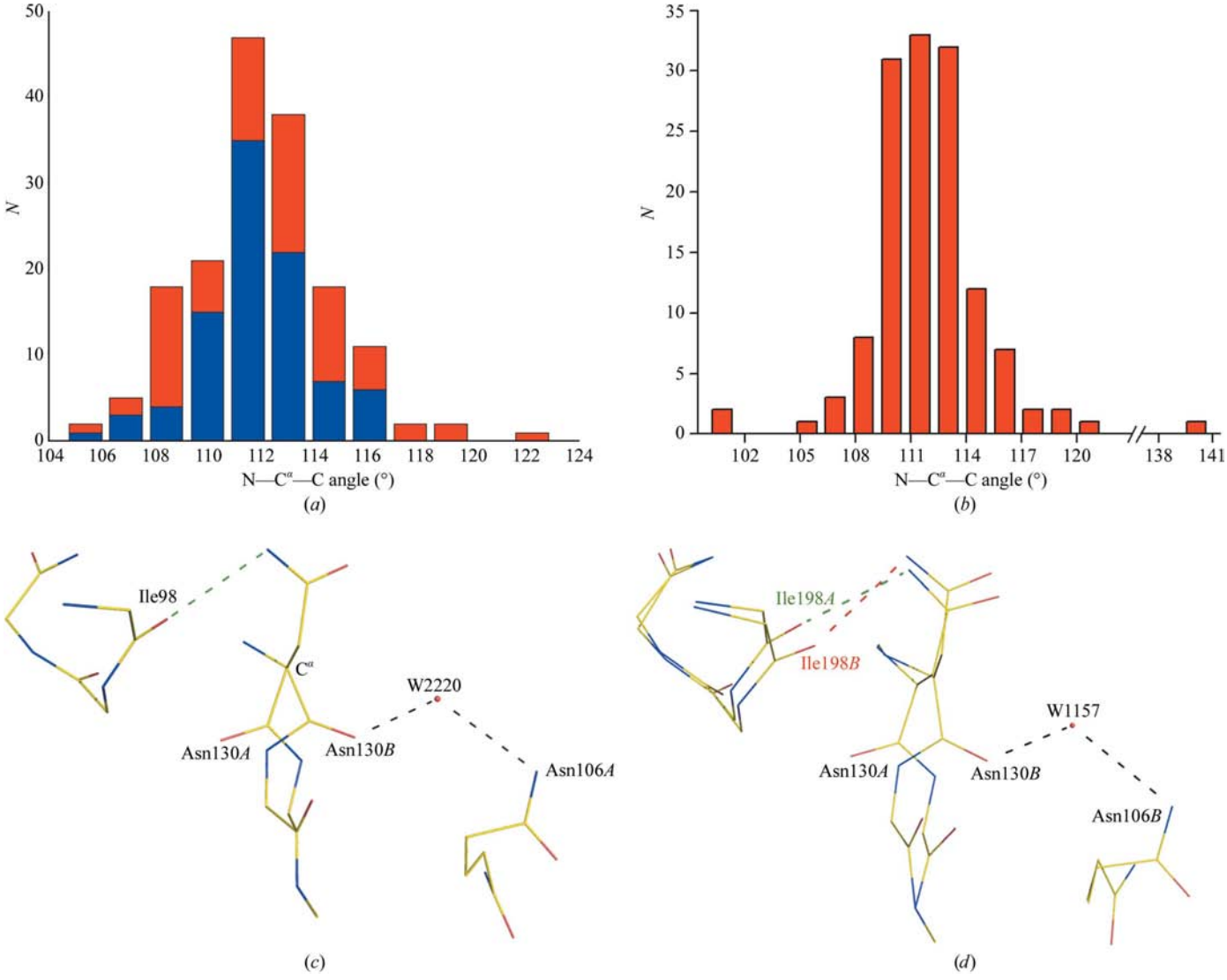

Figure 5
(*a*) Distribution of the N—C^{α}—C angles for the current model and (*b*) for 3lzt. (*c*) Modeling the peptide flip around Asn103 in 3lzt and (*d*) in the current model. Blue and red colours in (*a*) correspond to residues in single and double conformations, respectively.

present for Trp62 (Fig. 6*c*) with a dihedral angle of 25.4° between the two indole planes. The disorder of the Trp62 side chain is also responsible for the large deviation in the distribution of ω angles.

3.3. Displacement parameters

The pattern of atomic displacement parameters along the protein chain is very similar to that in the atomic resolution structures (3lzt and 4lzt), but the equivalent *B* factors are much lower (Fig. 6*d*). The mean *B* factors of the main-chain, side-chain and solvent atoms are 4.82, 7.15 and 14.32 Å^2, respectively. Table 6 shows the B_{eq} statistics for 4lzt, 3lzt and the current model. Although more atoms were modeled in the current structure, with 624 main-chain atom sites modeled compared with 530 sites in 3lzt and 517 sites in 4lzt, the average *B* value is 4.82 Å, which is much lower than that for 3lzt (7.36 Å) and for 4lzt (11.0 Å). The average B_{eq} values of other HEWL structures in the PDB are in the range 12–25 Å^2.

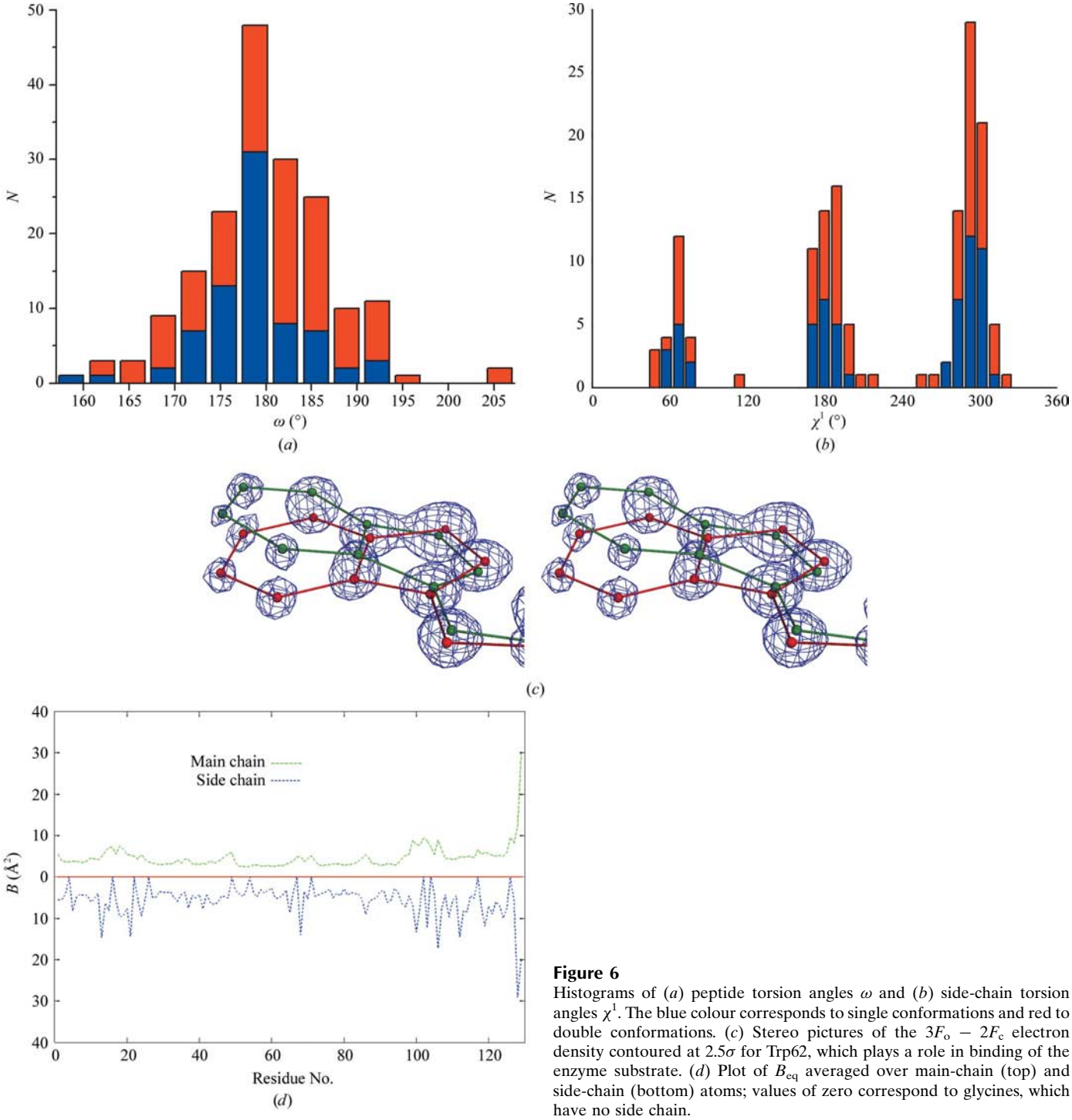

Figure 6
Histograms of (*a*) peptide torsion angles ω and (*b*) side-chain torsion angles χ^1. The blue colour corresponds to single conformations and red to double conformations. (*c*) Stereo pictures of the $3F_o - 2F_c$ electron density contoured at 2.5σ for Trp62, which plays a role in binding of the enzyme substrate. (*d*) Plot of B_{eq} averaged over main-chain (top) and side-chain (bottom) atoms; values of zero correspond to glycines, which have no side chain.

3.4. Residues in more than one conformation

The current structure has revealed eight sections of main chain with alternate conformation as well as 45 side chains, which account for approximately 35% of the whole protein chain. The lowest occupancy of about 10% was identified and refined for the fragment consisting of the main chain and side chains of residues 87–90.

3.4.1. Alternate conformations of the main chain. In spite of the overall rigidity of lysozyme, eight parts of the main chain were modeled with double conformations: Arg5, Asp18–Leu25, Thr43, Ile78–Pro79, Asp87–Ala90, Ile98–Ser100, Asn103–Gly104 and Trp111–Arg114. These eight flexible main-chain regions could be catalogued into three classes. Arg5, Thr43, Ile78–Pro79 and Trp111–Arg114 belong to the first class, in which main-chain atoms were modeled with alternate conformations mainly as a consequence of stress imposed by disordered side chains. The second class revealed the nature of the flexibility of the loop area, such as Asp18–Leu25, which links the two α-helices (Fig. 7*a*). The third class specifically arises from the surrounding hydrogen bonds, as mentioned in §3.2.1, where peptide flipping in Asn103 caused the movement of the main chain of Ile98–Ser100 which is connected through a hydrogen bond. Another disordered

Table 6
Comparison of displacement parameters.

	0.95 Å (4lzt)		0.925 Å (3lzt)		Current work	
Atoms	Total	$\langle B_{eq}\rangle$ (Å^2)	Total	$\langle B_{eq}\rangle$ (Å^2)	Total	$\langle B_{eq}\rangle$ (Å^2)
Protein main chain	517	11.0	530	7.36	624	4.82
Protein side chain	503	17.65	574	9.91	699	7.15
Protein atoms (all)	1020	14.28	1104	8.69	1323	6.05
Water	139	28.99	250	23.34	170	14.32
Other	24	32.96	44	14.15	52	10.00
All	1183	16.38	1398	11.80	1545	7.10

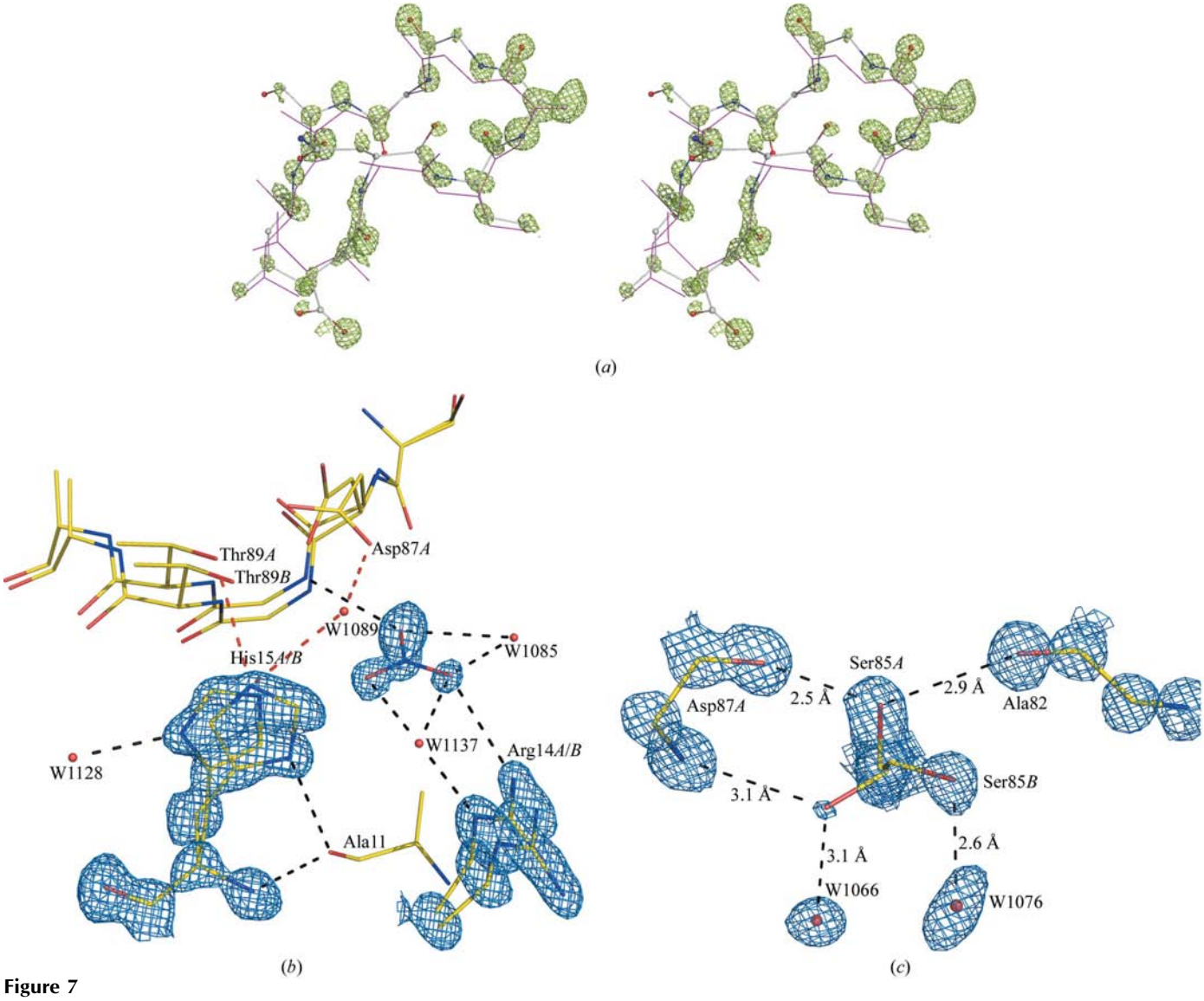

Figure 7
(*a*) Stereo plot of the OMIT map for the flexible-loop area (18–25) contoured at 3.5σ, with the major conformation represented in pink. Electron-density map and model of two more flexible fragments: (*b*) His15 and (*c*) Ser85 in three side-chain conformations.

main chain Asp87–Ala90 also belongs to this class, as illustrated in Fig. 7(*b*). The ND1 and N atoms in the major conformation of His15 with an occupancy of 0.65 are hydrogen-bonded to the O atom of Ala11, while NE2 is linked to Thr89 and Asp87 *via* a water molecule. His15 is not only modeled in a major conformation similar to that in 3lzt and 4lzt, but also in a minor conformation with the imidazole flipped, which causes ND1 to hydrogen bond to another water molecule. At the same time, NE2 of the major His15 conformation dragged the main chain of Asp87–Ala90 *via* hydrogen bonding to OG1 of the minor Thr89 with an occupancy of 0.13.

3.4.2. Alternate conformations of side chains. The current structure has revealed a total of 45 alternate-conformation side chains, 14 of which belong to main-chain fragments that also display alternate conformations. Side-chain disorder involved various types of residues, in particular ten out of 11 arginines, underlying the ability of arginine to make interactions (Brinda & Vishveshwara, 2005). As shown previously in §3.2.4, the side chain of Trp62 displayed a double conformation owing to the lack of restriction by binding substrate in the native lysozyme. In addition to the more common double conformations, a triple conformation is illustrated in Fig. 7(*c*), where Ser85 OG of forms hydrogen bonds with three neighboring atoms. Only part of Lys1 was modeled with two conformations.

3.5. H atoms

Although 1355 H atoms were added to the model at theoretically calculated positions and refined as 'riding' on their parent atoms, electron-density peaks were visible for most of these atoms before they were added to the model. In order to remove the refinement memory, the final coordinates were randomized by 0.1 Å along the *x*, *y* and *z* axes with *MOLEMAN*2 (Kleywegt *et al.*, 2001), which resulted in an average coordinate shake of 0.096 Å. Re-refinement until convergence without inclusion of H atoms produced peaks in a difference map for 416 of 1355 (30.7%) H atoms at the 3σ level within a radius of less than 0.5 Å from their calculated positions. Electron density for Thr51 is illustrated in Fig. 8(*a*) and all the electron-density peaks for the H atoms of this residue are visible in the $F_o - F_c$ map contoured at 3.2σ.

The H-OMIT map for Tyr53 is shown in Fig. 8(*b*). In addition to the normal H atoms, covalent-bond density can also be observed for the chemical bonds. Interestingly, the H atom of the hydroxyl of Tyr53 also manifested two possible positions, forming two hydrogen bonds with adjacent O atoms.

3.6. Solvent region

3.6.1. Water structure. Water peaks were first located automatically during isotropic refinement using *PHENIX*, in which 218 electron-density peaks were attributed to water molecules. During refinement with *SHELXL*, water/solvent peaks were manually added or subtracted using peaks greater than 5σ as a guide in the difference Fourier ($F_o - F_c$) maps. Modeling of the diffuse solvent used the Babinet principle as implemented in the *SWAT* option in *SHELXL*.

In the later stages of refinement, during the modeling of alternate side-chain conformations, 28 waters which were clearly associated with only one part of a disordered side chain had their occupancies linked to the refined value of the protein side chain in question. The *R* factor for the model with 140 fully occupied waters and 30 waters with occupancies less than unity was 8.48%. Visual inspection of electron density suggested that many water sites were probably only partially occupied. Therefore, the occupancies of all water molecules were refined. After 20 cycles of anisotropic refinement, *R* and

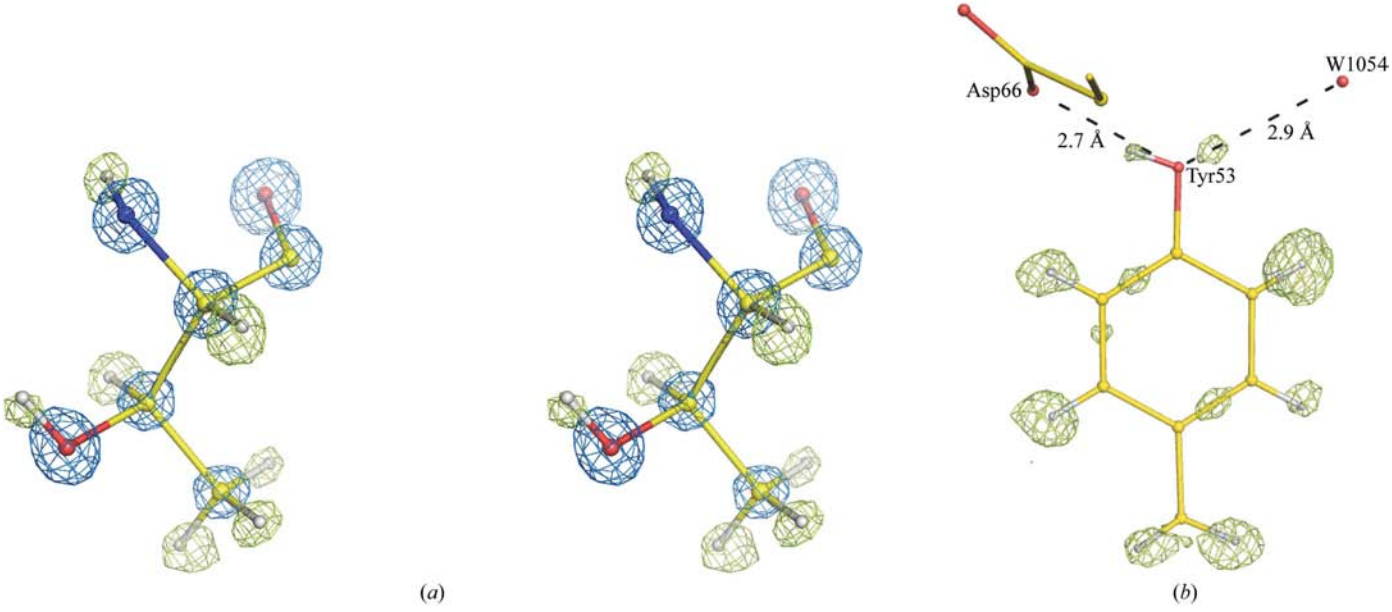

Figure 8
(*a*) Stereo picture of $3F_o - 2F_c$ (blue) and $F_o - F_c$ (green) electron density for Thr51 contoured at 4σ and 3.2σ, respectively. The map was calculated after randomizing the final model coordinates by 0.096 Å and re-refining until convergence. (*b*) Hydrogen-OMIT map for the side chain of Tyr53.

R_{free} fell by 0.09% and 0.07%, respectively. The more or less equivalent drop in both the conventional R factor and the free R convincingly indicated the validity of this step. All water occupancies were refined with better than 10% accuracy. Finally, a total of 170 water molecules were included in the final model; the average occupancy for all waters was 0.78. Taking site occupancies greater than 0.95 to be unity gives 34 fully occupied water sites. 18 waters have site occupancies of less than 0.5.

Comparison of the solvent structure in the current model to that of 3lzt shows that 87 water sites deviate by less than 0.5 Å after least-squares superposition of the two coordinate sets using all the protein atoms. Assuming waters that agree within 1 Å between the two structures to be equivalent results in 72% (122 sites) of the waters in the current model being paired with waters from 3lzt. The r.m.s. distance between paired water sites is about 0.45 Å, with an isotropic displacement parameter difference of 9.26 Å^2. The displacement parameters of the paired waters correlate well if their values are low ($B < 10$ Å^2), but there is no significant correlation when the B_{eq} values are higher.

3.6.2. Nitrate and acetate ions. A total of nine nitrate ions were located, with eight of these ions being modeled 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 in the packing interfaces. Nitrate ions were distinguished from acetate ions on the basis of their bond lengths, electron density and involvement of their atoms in hydrogen bonds.

Six nitrate sites were identified in both 4lzt and 3lzt (Table 7). Four sites are common to all three structures. The two low-temperature structures (3lzt and the current model) share an additional site, nitrate 407. The site of nitrate 404, which was not found in 4lzt, was occupied by a water molecule

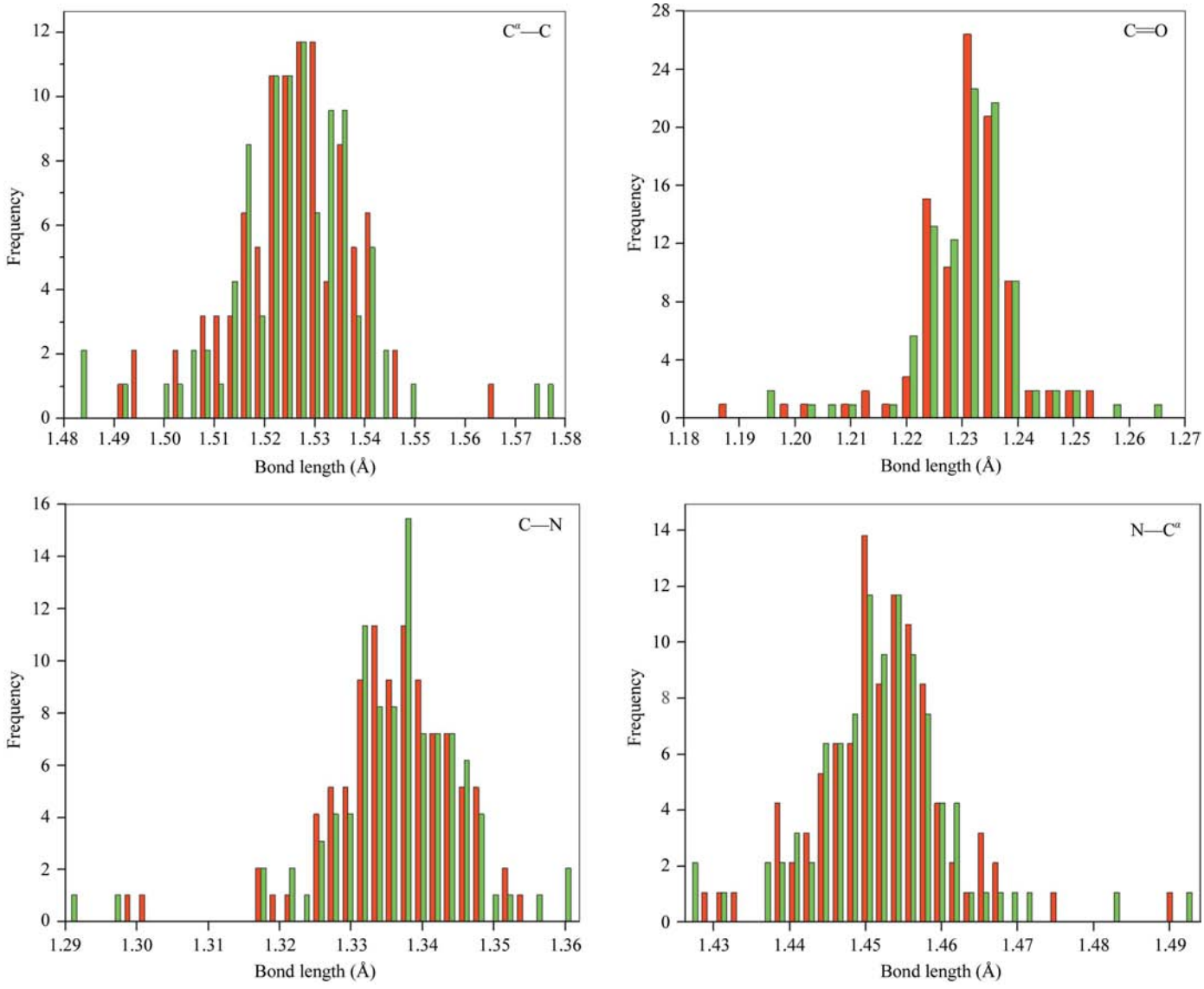

Figure 9
Distributions of peptide-bond lengths for well ordered parts of the model. Red columns represent restrained refinement and green columns unrestrained refinement.

Table 7
Bound nitrate and acetate ions and ethylene glycol molecules in triclinic lysozyme.

Current model	0.95 Å (4lzt)	0.925 Å (3lzt)
Nitrate 401	Nitrate 205	Gln121
Nitrate 402	Water 1043	Acetate 204
Nitrate 403	Nitrate 204	Nitrate 155
Nitrate 404	—	Water 2179
Nitrate 405	Nitrate 203	Nitrate 151
Nitrate 406	Nitrate 206	Nitrate 153
Nitrate 407	—	Nitrate 150
Nitrate 408	Nitrate 201	Nitrate 154
Nitrate 409	—	Acetate 201A
Acetate 201	Nitrate 202	Nitrate 152
Glycol 301	—	Water 2053
Glycol 302	Waters 1025 and 1113	Acetate 203
Glycol 303	Water 1126	Waters 2298 and 2140

in 3lzt. According to the surrounding chemical environment, site 402 should contain a nitrate ion, not an acetate ion as in 3lzt. Interestingly, the side chain of Gln121 in 3lzt stretched into the position of nitrate 401; however, the same position was modeled as a nitrate in 4lzt.

Only one acetate ion was located in the current structure. The current acetate 201 was replaced by a nitrate ion in 3lzt and 4lzt.

3.6.3. Ethylene glycol molecules. Three ethylene glycol molecules were located in the model, all of which come from the cryoprotectant. However, in 3lzt and 4lzt glycerol was used as the cryoprotectant rather than ethylene glycol, so that there are no ethylene glycols in the corresponding positions. Interestingly, EDO302 occupies the position of the acetamide group of GlcNAc at the centre of the active-site cleft, the O2 atom of which hydrogen bonds to the carboxyl O atom of Ala107.

4. Refinement with removed/relaxed restraints

The number of reflections per parameter in the final refinement was about 13.4, which is comparable to the situation encountered in small-molecule refinement. With such a high observation-to-parameter ratio, it is possible to carry out refinement without any stereochemical restraints. However, such a radical approach would be inappropriate for those main/side chains for which owing to high mobility or disorder there is not enough information in the diffraction data to adequately define the atomic parameters. The restraints were therefore retained for the disordered main-chain/side-chain fragments and the rest of the restraints were completely removed, except for the NZ atom of Lys13. The NZ atom of Lys13 is involved in an N—H···O hydrogen-bond system with the flexible C-terminal residue Leu129; although it is modeled only in a single conformation, negative density clearly appeared around the NZ atom. Therefore, restraints on the 1–2 bond-length NZ—CE and the 1–3 NZ—CD distances were kept for the NZ atom, but its final B_{eq} value (26.69 Å^2) was still higher than those for CE (19.94 Å^2) and CD (11.22 Å^2).

Getting rid of restraints effectively lowered the R and R_{free} factors by 0.17% and 0.1%, respectively. The differences in main-chain bond lengths between the restrained refinement and the nonrestrained refinement for the well ordered parts of the structure only are shown in Fig. 9. The average values of the C^{α}—C, C═O, C—N and N—C^{α} bond lengths were 1.527 (14), 1.232 (11), 1.337 (10) and 1.453 (10), respectively. The respective values recommended by Engh & Huber (1991) are 1.525 (21), 1.232 (20), 1.329 (14) and 1.458 (19) and those recommended by Jaskolski *et al.* (2007) are 1.527 (14), 1.234 (12), 1.334 (13) and 1.454 (12). In spite of the low number of observations, the currently obtained values are more consistent (except for the C═O bond) with those of Jaskolski *et al.* (2007).

5. Conclusions

The present study describes subatomic resolution data collection from a crystal of triclinic HEWL and the protocol used for structural refinement at 0.65 Å resolution. For a low-symmetry space group, $P1$ in this case, the sparse distribution of symmetry-equivalent reflections in reciprocal space and the effect of the blind region make several passes of crystal rotation with various effective exposures and crystal-to-detector distances necessary in order to measure strong low-resolution as well as the weakest high-resolution intensities. At such high resolution, which is usually only available in small-molecular crystallography, *ab initio* phasing with *ACORN* produced a high-quality electron-density map with some main-chain and side-chain alterations observable. This model-unbiased map was also used to distinguish the nitrate and acetate ions from each other in the refinement.

The strategy implemented consists of a number of well defined steps which have been described previously, but with some minor alterations. Addition of double conformations was performed during the anisotropic refinement. After identification of these minor alternate conformations, the more detailed model was refined again with isotropic displacement parameters and high diffraction-angle data so that the coordinates of the model became more accurate. Subsequent refinement with anisotropic ADPs allowed us to identify more disorder in the main chain or side chains. This protocol was confirmed by the cross-validated R_{free} factor in the refinement. Diffraction data to such high resolution permits the identification of the atomic types in the DB plot for the majority of atoms with B factors lower than 20 Å^2. After refinement of water-site occupancies, the DB plot of occupancy-normalized density for water molecules was comparable with the curve obtained for the protein O atoms.

Although HEWL is thought of as a rigid protein, the high-quality final model contains 1323 protein-atom sites for 1001 non-H atoms, in which 45 disordered side-chain residues were identified as well as eight sections of main-chain disorder, accounting for approximately 35% of the protein structure. The flexibility of loop regions was also clearly observed in the density. After modeling this loop ensemble, the crystallographic R and R_{free} factors fell considerably. The flexibility of

the side chain of Trp62 in native HEWL usually resulted in poor or undistinguishable electron density in previous structures. However, this movement was clearly frozen in this study, with double alternate conformations observable for the indole part of Trp62.

X-ray data collected to this resolution provide the crystallographer with an extremely accurate model of the protein structure being refined. It is expected that new parameters for these restraints could be extracted from future unrestrained subatomic structures.

HEWL is considered as a very rigid molecule that is stabilized by four disulfide bridges. Its *P*1 crystal form is very tightly packed, with an unusually small amount of solvent. However, comparison of lysozyme models refined at atomic resolution, for which uncertainties of atomic positions are available from full-matrix least-squares refinement, shows differences in the conformations of the main and side chains that far exceed the accuracy of the atomic positions. This suggests a considerable level of conformational flexibility and variability of lysozyme molecules that most probably results from subtle differences in the way the crystals were prepared and handled. A level of variability and adaptability of at least 0.2 Å or more can therefore be expected for other less rigid proteins. This should be taken into account in the modeling studies of protein–ligand complexes of potential medical relevance.

This work was supported in part by Federal funds from the National Cancer Institute, National Institutes of Health contract No. NO1-CO-12400 and the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does the mention of trade names, commercial products or organizations imply endorsement by the US Government. Diffraction data were collected at the SBC 19-ID beamline at the Advanced Photon Source, Argonne National Laboratory. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. W-31-109-Eng-38.

References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissing, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Blake, C. C. F., Fenn, R. H., Johnson, L. N., Koening, D. F., Mair, G. A., North, A. C. T., Oldham, J. W. H., Philips, D. C., Poljak, R. J., Sarma, V. R. & Vernon, C. A. (2001). *International Tables for Crystallography*, edited by M. G. Rossmann & E. Arnold, Vol. *F*, pp. 745–772. Dordrecht: Kluwer Academic Publishers.

Blake, C. C. F., Fenn, R. H., North, A. C. T., Phillips, D. C. & Poljak, R. J. (1962). *Nature (London)*, **196**, 1173–1176.

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

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

Bon, C., Lehmann, M. S. & Wilkinson, C. (1999). *Acta Cryst.* D**55**, 978–987.

Bönisch, H., Schmidt, C. L., Bianco, P. & Ladenstein, R. (2005). *Acta Cryst.* D**61**, 990–1004.

Brinda, K. V. & Vishveshwara, S. (2005). *BMC Bioinformatics*, **6**, 296.

Brünger, A. T. (1997). *Methods Enzymol.* **277**, 366–396.

Chen, C.-J., Liu, Y.-H., Chen, Y.-T. & Liu, M.-Y. (2006). *Biochem. Biophys. Res. Commun.* **349**, 79–90.

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

Coppens, P. (1997). *X-ray Charge Densities and Chemical Bonding*, p. 11. Oxford University Press.

Dauter, Z. (1999). *Acta Cryst.* D**55**, 1703–1717.

DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, CA, USA.

Dickerson, R. E., Reddy, J. M., Pinkerton, M. & Steinrauf, L. K. (1962). *Nature (London)*, **196**, 1178.

Dunlop, K. V., Irvin, R. T. & Hazes, B. (2005). *Acta Cryst.* D**61**, 80–87.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Esposito, L., Vitagliano, L., Sica, F., Sorrentino, G., Zagari, A. & Mazzarella, L. (2000). *J. Mol. Biol.* **297**, 713–732.

Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Jia-xing, Y. & Chao-de, Z. (2000). *Acta Cryst.* D**56**, 1137–1147.

Hakanpää, J., Linder, M., Popov, A., Schmidt, A. & Rouvinen, J. (2006). *Acta Cryst.* D**62**, 356–367.

Harata, K. & Akiba, T. (2004). *Acta Cryst.* D**60**, 630–637.

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

Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. (1996). *Nature (London)*, **391**, 272.

Howard, E. I., Sanishvili, R., Cachau, R. E., Mitschler, A., Chevrier, B., Barth, P., Lamour, V., Van Zandt, M., Sibley, E., Bon, C., Moras, D., Schneider, T. R., Joachimiak, A. & Podjarny, A. (2004). *Proteins*, **55**, 792–804.

Jamal-Talabani, S., Boraston, A. B., Turkenburg, J. P., Tarbouriech, N., Ducros, V. M.-A. & Davies, G. J. (2004). *Structure*, **12**, 1177–1187.

Jaskolski, M., Gilski, M., Dauter, Z. & Wlodawer, A. (2007). *Acta Cryst.* D**63**, 611–620.

Jelsch, C., Teeter, M. M., Lamzin, V., Pichon-Pesme, V., Blessing, R. H. & Lecomte, C. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 3171–3176.

Joynson, M. A., North, A. C. T., Sarma, V. R., Dickerson, R. E. & Steinrauf, L. K. (1970). *J. Mol. Biol.* **50**, 137–142.

Kang, B. S., Devedjiev, Y., Derewenda, U. & Derewenda, Z. S. (2004). *J. Mol. Biol.* **338**, 483–493.

Kleywegt, G. J., Zou, J.-Y., Kjeldgaard, M. & Jones, T. A. (2001). *International Tables for Crystallography*, edited by M. G. Rossmann & E. Arnold, Vol. *F*, pp. 353–356. Dordrecht: Kluwer Acadamic Publishers.

Ko, T.-P., Robinson, H., Gao, Y.-G., Cheng, C.-H., DeVries, A. L. & Wang, A. H.-J. (2003). *Biophys. J.* **84**, 1228–1237.

Kuhn, P., Knapp, M., Soltis, M., Ganshaw, G., Thoene, M. & Bott, R. (1998). *Biochemistry*, **37**, 13446–13452.

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

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.

Lehmann, M. S. & Stansfield, R. D. F. (1989). *Biochemistry*, **28**, 7028–7033.

Liu, L., Nogi, T., Kobayashi, M., Nozawa, T. & Miki, K. (2002). *Acta Cryst.* D**58**, 1085–1091.
Moult, J., Yonath, A., Traub, W., Smilansky, A., Podjarny, A., Rabinovitch, D. & Saya, A. (1976). *J. Mol. Biol.* **100**, 179–195.
Otwinowski, Z. & Minor, M. (1997). *Methods Enzymol.* **276**, 307–326.
Patterson, W. R., Anderson, D. H., DeGrado, W. F., Cascio, D. & Eisenberg, D. (1999). *Protein Sci.* **8**, 1410–1422.
Pichon-Pesme, V., Lachekar, H., Souhassou, M. & Lecomte, C. (2000). *Acta Cryst.* B**56**, 728–737.
Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Protein Chem.* **23**, 283–437.
Ramanadham, M., Sieker, L. C. & Jensen, L. H. (1990). *Acta Cryst.* B**46**, 63–69.
Salem, M., Mauguen, Y. & Prangé, T. (2006). *Biochim. Biophys. Acta*, **1764**, 903–912.
Sauter, C., Otálora, F., Gavira, J.-A., Vidal, O., Giegé, R. & García-Ruiz, J. M. (2001). *Acta Cryst.* D**57**, 1119–1126.
Schmidt, A., Jelsch, C., Ostergaard, P., Rypniewski, W. & Lamzin, V. S. (2003). *J. Biol. Chem.* **278**, 43357–43362.
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.
Steinrauf, L. K. (1959). *Acta Cryst.* **12**, 77–79.
Steinrauf, L. K. (1998). *Acta Cryst.* D**54**, 767–779.
Terwilliger, T. C. (2003*a*). *Acta Cryst.* D**59**, 38–44.
Terwilliger, T. C. (2003*b*). *Acta Cryst.* D**59**, 45–49.
Walsh, M. A., Schneider, T. R., Sieker, L. C., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1998). *Acta Cryst.* D**54**, 522–546.
Yonath, A. J., Podjarny, A., Honing, B., Sielecki, A. & Traub, W. (1977). *Biochemistry*, **16**, 1418–1424.
Yonath, A. J., Podjarny, A., Honing, B., Traub, W., Sielecki, A., Herzberg, O. & Moult, J. (1978). *Biophys. Struct. Mech.* **4**, 27–36.
Yonath, A., Sielechi, A., Moult, J., Podjarny, A. & Traub, W. (1977). *Biochemistry*, **16**, 1413–1417.