Synchrotron X-ray powder diffraction study of hexagonal turkey egg-white lysozyme

Irene Margiolaki,[a]* Jonathan P. Wright,[a]* Andrew N. Fitch,[a] Gavin C. Fox[a] and Robert B. Von Dreele[b]

[a]European Synchrotron Radiation Facility, BP-220, F-38043, Grenoble CEDEX 9, France, and [b]Advanced Photon Source, Argonne, IL 60439, USA

Correspondence e-mail: margiolaki@esrf.fr, wright@esrf.fr

The structure of turkey egg-white lysozyme (TEWL) has been refined from high-resolution X-ray powder diffraction data. The sample was rapidly obtained as a polycrystalline precipitate at high protein concentration using 0.5 M NaCl solvent pH 6 and was deposited in the PDB with code 1xft. The diffraction data were collected at room temperature. Molecular replacement was shown to give a suitable starting point for refinement, illustrating that powder data can be sufficient for this approach. Crystallographic models were then refined by combined Rietveld and stereochemical restraint analysis of the powder data (d_{min} = 3.35 Å), resulting in the extraction of reliable lattice parameters and the refinement of the molecular conformation at room temperature. The structure is hexagonal [space group $P6_122$, unit-cell parameters a = 71.0862 (3), c = 85.0276 (5) Å] with 12 symmetry-related molecules in the unit cell, in agreement with previous studies. The results of our analysis are indicative of specific amino acids being disordered at this temperature. Upon cooling, a sudden drop in the lattice parameters at ~250 K is observed concurrently with the freezing of the mother liquor. The observation of severe peak broadening below this temperature indicates strain effects accompanying the freezing transition, which are found to be reversible. Finally, a correlation between the unit-cell parameters and the pH of the buffer solution is evident, in a similar manner to earlier observations on HEWL.

1. Introduction

Obtaining the atomic structure of large macromolecules such as proteins depends on the availability of good-quality single crystals. Following the recent reports of crystal structure refinement (Von Dreele, 1999) and solution (Von Dreele, 2000) of some small protein structures from powder diffraction data, the requirement for a single crystal might be relaxed. Since many materials of interest do not readily form single crystals, the availability of the powder technique widens the spectrum of samples that might be characterized crystallographically. Unfortunately, the collapse of three-dimensional reciprocal space into a one-dimensional powder diffraction pattern leads to significant loss of information. There is not only the usual phase problem; owing to significant peak overlap it is frequently not even possible to determine the intensity of individual Bragg reflections, but only their sum. Nevertheless, powder diffraction data give a range of complementary information that can be more difficult to obtain from a single crystal. The peak shapes depend on the microstructure of the material, accurate unit-cell parameters can easily be determined and the sample generally survives under more extreme conditions.

Hen egg-white lysozyme (HEWL) has been extensively studied by a wide range of crystallographic techniques, perhaps owing to the very high quality crystals that can be obtained. Turkey egg-white lysozyme (TEWL) is very similar to HEWL, with 95% sequence homology and a difference of only seven amino acids between them. However, the former does not appear to readily form crystals of such high quality. Currently, there are four entries in the PDB for hexagonal TEWL, spanning a pH range of 4.5–8 [PDB codes 1lz2 (Sarma & Bott, 1977), 2lz2 (Parsons & Phillips, unpublished work), 3lz2 (Howell *et al.*, 1992) and 1tew (Howell, 1995)]. The space group ($P6_122$) has not been previously reported for HEWL. The hexagonal space group is found both at low and high pH, although the molecular conformation is slightly modified. A monoclinic form of TEWL has also been reported, which was formed in 2.2 *M* ammonium sulfate solution (PDB code 135l; Harata, 1993).

The powder diffraction technique has developed dramatically in the last 20 years (Cheetham & Wilkinson, 1992; Harris *et al.*, 2001; Harris & Cheung, 2004; Evans & Radosavljevic Evans, 2004); however, its application to macromolecular crystallography remains in its infancy. Initially, we identified TEWL as a good candidate material for a trial powder diffraction study for several reasons. Owing to the similarity to HEWL, a great deal is known about the system already, although TEWL will hopefully be more representative of a typical microcrystalline sample in comparison to HEWL, which readily forms large crystals of excellent quality. We would like to know whether it is possible to exploit molecular-replacement techniques with powder data: the TEWL structure was originally solved in this way. Furthermore, we aim to illustrate the kind of complementary information that can be derived simply from powder data when a sample goes through a phase transition. We report refinements of the room-temperature crystal structure from powder data, the variation of the unit-cell parameters with pH of the precipitation medium and an investigation into the processes occurring when the sample is frozen.

2. Experimental

2.1. Crystallization

Turkey-egg white lysozyme (TEWL) was purchased from Sigma–Aldrich Chemical Co. (Product No. L6255, Lot 064H7230; CAS No. 12650-88-3). The crystallization procedure used here has been described previously for hen egg-white lysozyme (Von Dreele, 2001). Following this crystallization route, polycrystalline precipitate was rapidly formed after adding 20 mg of as-received dry lysozyme to 200 µl 0.5 *M* NaCl solution pH 6.0 (protein concentration = 100 mg ml^{-1}). The pH of the solution was adjusted to 6.0 with 0.05 *M* sulfuric acid prior to addition of the dry protein and measured with a pH meter. The time evolution of precipitation was observed with a polarized-light microscope (Olympus BX60) to be rapid. The sample was loaded into a 1.5 mm diameter glass capillary that was centrifuged in order to achieve better packing of precipitate. Excess mother liquor was removed and the capillary was sealed with wax to prevent protein dehydration. This will be referred to as sample *A*. An additional sample (*B*) was prepared following the same procedure using 0.5 *M* NaCl solution pH 5.0. In order to explore the effect of crystallization conditions on the structure of TEWL, crystalline powder was also prepared by the sitting-drop method. 300 µl protein solution (75 mg ml^{-1}) and 200 µl 3 *M* sodium acetate buffer solution pH 4.5 containing 6.5%(*w*/*v*) NaCl were equilibrated against a 25 ml reservoir of buffer solution. Polycrystalline precipitate was obtained in about three weeks. The latter sample was labelled *C*. Samples *A* and *B* were not buffered and evolution of pH is expected while the protein crystallizes.

2.2. X-ray data collection and processing

Powder X-ray data were collected for samples *A* and *B* with a wavelength of 0.700667 (11) Å at 295 K using the high-resolution powder diffraction beamline ID31 at the ESRF, Grenoble (Fitch, 2004). The same instrument was used for sample *C* with a wavelength of 0.600652 (13) Å. The samples were contained in spinning 1.5 mm diameter glass capillaries mounted on the axis of the diffractometer and patterns were measured with a period of 6.5 min using a beam size of 1.5 mm [1.5 mm (horizontal) × 1.0 mm (vertical)]. Appreciable changes were observed in the data after several patterns had been collected. More precisely, significant changes in the lattice parameters accompanied by a gradual increase in peak broadening are characteristic of radiation-damage effects. In order to increase the counting statistics without compromising the data quality, the capillary was translated to give a fresh region of sample as soon as radiation-damage effects were evident in the peak positions and widths. Only the data collected for fresh material were used in subsequent analysis. The normalized diffraction profile for sample *A* used in the Rietveld analysis (Rietveld, 1969) had 5789 points (0.55–12.0° 2θ, binned with a 0.002° step) showing recognisable peaks to a *d* spacing of ~3.35 Å. (Note that the term 'high resolution' in powder diffraction is defined in terms of narrow peaks with very low instrumental contribution so that the peak overlap is reduced and the information contained in the patterns is maximized.)

Data were also collected for sample *A* with 50%(*w*/*v*) PEG 3K added to provide possible cryoprotection using a Bruker Smart 6500 CCD area detector at the ID11 beamline also at the ESRF. A 0.5 mm diameter capillary was used to reduce the peak broadening arising from sample-size effects. The sample was initially cooled from 320 to 105 K at a rate of 80 K h^{-1} and then warmed back to 320 K at a rate of 160 K h^{-1}. Images of the Debye–Scherrer rings were recorded with an exposure time of 20 s and no appreciable changes in the patterns that may be attributed to radiation damage were observed. A monochromatic X-ray beam of wavelength 1.033 (3) Å and dimensions 0.05 × 0.05 mm illuminated the sample and the detector was arranged such that the beam centre was displaced from the centre of the image and the furthest edge of the

detector corresponded to a *d* spacing of ∼2.4 Å. During the data collection, the sample was oscillated about its axis by 20° to improve powder averaging. One-dimensional diffraction patterns were obtained from the image data by integrating

Figure 1
Correlation coefficients computed for molecular replacement using model 135l (monoclinic TEWL) and intensities extracted from the low-resolution data set. The top 50 rotation-function peaks were each used to generate 50 translation-function peaks with the *MOLREP* program.

Figure 2
The Rietveld fit for sample *A* at 295 K (λ = 0.700667 Å, ID31 data). Open symbols, data; red line, final calculated pattern; lower black line, difference between the experimental data and the final refined model. The bars represent the Bragg reflections corresponding to the hexagonal structural model. The background intensity has been subtracted for clarity. The profile has been expanded by a factor of 10 at Bragg angles larger than 4°. Inset: magnification of the 4–6° 2θ regime.

around the Debye–Scherrer rings using local software (*FIT2D* program; Hammersley, 1997).

3. Refinement and structure analysis

3.1. Unit-cell and space-group determination

Data analysis was initiated with the high-resolution ID31 diffraction profiles collected for sample *A*, as these had the best counting statistics. Using positions extracted for the first 20 diffraction lines, a hexagonal unit cell with dimensions of *a* = 71.1985, *c* = 85.205 Å was found using the *DICVOL* program (Boultif & Louër, 1991) with figure of merit $M(20) = 200$, a value which indicates that the indexing is substantially correct (De Wolff, 1968). A probabilistic approach for the space-group determination (Markvardsen *et al.*, 2001) incorporated into the *DASH* software package (David *et al.*, 1998) indicated that the extinction symbol '$P6_1$– –' should be chosen, giving possible space groups $P6_1/P6_5$ or $P6_122/P6_522$. Previously reported results for TEWL obtained from single crystals indicated that $P6_122$ is correct and therefore the analysis was performed using that space group. In space group $P6_122$ there is one molecule in the asymmetric unit with a Matthews coefficient of 2.18 Å^3 Da^{-1} for crystals with 43% solvent content. The same unit cell and space group were also found to be suitable for all the samples prepared under different conditions.
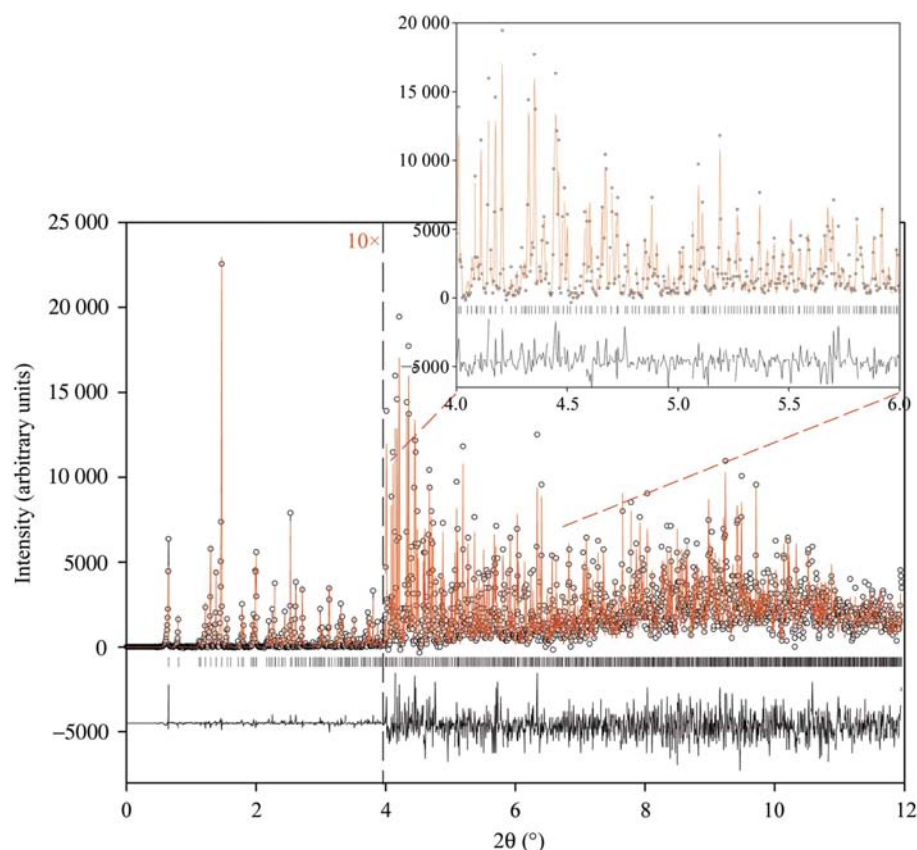
3.2. Molecular replacement

Integrated intensities were extracted from the high-resolution powder profile for sample *A* using the *PRODD* profile-refinement program (Wright & Forsyth, 2000), which is based on the *CCSL* subroutine library (Matthewman *et al.*, 1982). The profile coefficients, background and instrumental zero shift were optimized together with the values of the intensities of the Bragg peaks using a damped least-squares procedure (Wright, 2004). 2112 intensities were extracted up to an angle of 12° (*d* = 3.325 Å), leading to an excellent quality of fit ($\chi^2 = 1.5$, $R_{wp} = 4.25\%$). The algorithm used in the program is designed to equipartition overlapping peak intensities by adding a term to the minimization function that is proportional to the sum of intensities squared. These extracted intensities were then imported into the *CCP*4 software suite (Collaborative Computational Project, Number 4, 1994) to provide input into the molecular-replacement program

research papers

Table 1
Molecular-replacement results using the integrated intensities extracted from the powder data.

Figures of merit are labeled by their peak number in the rotation function. In each case the first line represents the correct solution and the second line gives the next or highest other peak found.

Model	Data	Rotation function (Rf/σ)	Translation function (CC)
2lz2	ID31	6: 3.68	6: 0.502
		1: 4.12	6: 0.292
3lz2	ID31	1: 4.89	1: 0.457
		2: 4.42	1: 0.260
1tew	ID31	1: 4.73	1: 0.517
		2: 3.68	1: 0.271
135l	ID31	1: 5.09	1: 0.452
		2: 4.10	1: 0.262
2lz2	ID11	6: 2.60	6: 0.260
		1: 2.94	8: 0.209
3lz3	ID11	2: 2.95	2: 0.402
		1: 3.42	3: 0.359
1tew	ID11	2: 3.34	2: 0.310
		1: 3.46	2: 0.214
135l	ID11	19: 2.60	19: 0.366
		1: 3.01	12: 0.226

MOLREP (Vagin & Teplyakov, 1997). The models for both hexagonal and monoclinic TEWL, which are available in the PDB, were all tested against the extracted peak intensities. The same procedure was also carried out for the data collected using the area detector at ID11, where the peak overlap is more severe because of the lower instrumental resolution.

Table 1 and Fig. 1 summarize the results from the molecular replacement, indicating that the correct solution was always clearly distinguished from the noise peaks, even in the more challenging case of using area-detector data and the mono-clinic TEWL structure as initial model. By simply treating the data as if they were from a single crystal and ignoring the peak overlaps, the same packing arrangement as previously reported was always reproduced. Randomizing the initial orientation of the hexagonal models leads to a very slight reduction in signal to noise ($\sim$0.01 in correlation coefficient), presumably owing to different grid-sampling effects. In most cases the highest peak in the rotation function corresponds to the correct orientation of the molecule, but in the worst case the correct peak was in position 19 in the peak list and it was only when the translation functions were computed that a solution could be identified. Solutions were checked by generating all of the symmetry-related molecules using the *Swiss-PdbViewer* software (Guex & Peitsch, 1999) and comparing them with the single-crystal structures. Frequently, the models were related by the affine normaliser (0, 0, 1/2) of the space group (single-crystal models 2lz2 and 1tew are related by this translation). Using a model for tetragonal hen egg-white lysozyme (PDB code 193l; Vaney *et al.*, 1996) also leads to a structure which can be superimposed with the single-crystal structures for TEWL, apart from the differences in amino-acid sequence and side-chain conformations (top correlation coefficient, CC = 0.391, noise CC = 0.132). For comparison, we note that Nixon & North (1976) initially used only 'the strongest 35 of the 269 unique measured structure factors of human lysozyme in the $\lambda/(2 \sin \theta)$ range 10–6 Å'

when computing a rotation function and using a model previously reported for HEWL. Despite the paucity of powder data, there is still enough information to find the three positional and three orientational parameters required for the description of the molecule with respect to the unit cell.

3.3. Structure refinement for sample *A*

Further data analysis was performed with the *General Structural Analysis Software* (*GSAS*; Larson & Von Dreele, 2001) using the EXPGUI graphical user interface (Toby, 2001). In order to characterize the peak shape and background parameters, a Le Bail fit (Le Bail *et al.*, 1988) was initially carried out with sample *A* and ID31 data. The low-angle tails of the peaks were described well by the profile function of Finger *et al.* (1994), which models peak asymmetry arising from axial divergence. Parameter values were refined during the early stages of the refinement and were subsequently fixed to S/L = H/L = 2.85×10^{-3}. There was significant improvement in fit when anistotropic strain broadening of the peaks was allowed to refine [to a value of y_e = 0.932 (3); (001) direction]. The pattern-decomposition procedure converged with R_{wp} = 4.49% and χ^2 = 1.675, resulting in unit-cell parameters a = 71.0945 (3), c = 85.0371 (5) Å (space group $P6_122$). This procedure fits the data without using a crystallographic model and is commonly used for intensity extraction from powder data. Once a satisfactory description of the peak profiles had been established, refinement of the structure could proceed. Isotropic temperature factors were employed for the description of the thermal motion of all atoms and were set equal to U_{iso} = $B_{iso}/8\pi^2$ = 0.30 Å^2 (if U = 0.3 Å^2 then B = 23.7 Å^2) and not allowed to vary at any stage of the analysis. A Babinet's principle modification of all the atom scattering factors according to

$$f = f_0 - A_s \exp(-8\pi^2 B_s \sin^2 \theta/\lambda^2)$$

accounted for solvent scattering and facilitated fitting the lowest angle part of the powder diffraction data. The initial weighted R factors (R_{wp}) calculated using data between 100 and 3.35 Å resolution were 14.04, 15.13 and 18.92% for 2lz2, 3lz2 and 1tew structural models, respectively. Although the initial profile R factors were high, the packing of the protein molecules in the crystal was assumed to be correct, as indicated by the molecular-replacement results. All three initial models were subjected to several cycles of rigid-body minimization refinement that improved the fit to give R_{wp} = 13.24, 14.85 and 14.84% for the 2lz2, 3lz2 and 1tew models, respectively.

Comparison of the three refined models after rigid-body minimization, in combination with the essentially better agreement of the 2lz2 PDB entry with the experimental data, indicated that the molecular conformation in that case is most appropriate for these data. Keeping in mind that the pH for those models was 7 for 2lz2, 8 for 3lz2 and 4.5 for 1tew, it is reassuring to find the model giving the best fit to these powder data is that which was crystallized under the most similar conditions.

Table 2
Refinement results using the high-resolution profiles.

The starting model was TEWL (PDB code 1tew).

	A	B	C
Data collection			
λ (Å)	0.700667 (11)	0.700667 (11)	0.600652 (13)
2θ range (°)	0.55–12	0.55–12	0.55–12
$\Delta 2\theta$ (°) zero shift	−0.0028 (3)	−0.0028 (3)	−0.098 (12)
Combined Rietveld and stereochemical restraint refinement			
N_{ref}	2099	2093	2095
Resolution (Å)	3.35	3.35	3.35
$N_{restraints}$	5406	5389	5406
N_{steps}	5789	5788	5786
$N_{obs} = N_{restraints} + N_{steps}$	11195	11177	11192)
$N_{parameters}$	2964	2964	2964
R_{wp} (%)	8.10	2.13	6.37
R_{p} (%)	6.04	1.67	4.62
R_{F}^{2} (%)	11.29	9.99	16.51
A_{s}	6.051 (3)	5.916 (5)	5.861 (3)
B_{s}	0.992 (4)	1.278 (4)	1.374 (3)
Powder profile parameters			
y	9.3 (1)	557.4(3)	44.7(6)
y_{e}	0.9 (2)	−277.4 (2)	−29.8 (9)
PROCHECK results			
Bond error (Å)	0.0030	0.08	0.06
Angle error (°)	1.92	2.2	2.5
Residues in core region (%)	82.9	81.0	80.0
ω torsion-angle error (°)	3.5	5.0	4.0
Bad contacts per 100 residues	11.7	12.0	12.5
ζ-angle error (°)	2.1	2.24	2.2
Hydrogen-bond energy error (kJ mol^{-1})	3.64	4.19	3.89

In order to achieve a better agreement between the observed and calculated profiles, the crystal structure of TEWL (sample A) was subjected to a combined Rietveld and stereochemical restraint least-squares refinement using previously described techniques (Von Dreele, 2000, 2001, 2005) and the *GSAS* software. Protein refinement was achieved by employing the band-matrix approximation implemented in the program with a matrix bandwidth of 50 parameters, while 5406 stereochemical restraints had to be imposed. Rietveld refinement involving 986 protein atoms in the asymmetric unit eventually yielded residuals of R_{wp} = 8.10%, R_{p} = 6.04%, R_{F}^{2} = 11.29%. The resulting protein stereochemistry was periodically evaluated with the *PROCHECK* (Laskowski *et al.*, 1993), *SFCHECK* (Vaguine *et al.*, 1999), *ERRAT* (Colovos & Yeates, 1993) and *WHATCHECK* (Hooft *et al.*, 1996) programs. Graphical examination and manual rebuilding of the protein molecule was performed using the *Swiss-PdbViewer* package (Guex & Peitsch, 1999). Details of this refinement are listed in Table 2 and the resulting fitted powder diffraction profile is shown in Fig. 2.

Refinement of all three of the models quickly leads to improvements in the profile R factors and these improvements must be balanced against the resulting stereochemistry of the molecule. By altering the weighting of the data and the restraints in the refinement process, a particular model is selected. Despite the initially better agreement indices for model 2lz2 after rigid-body refinement, the higher resolution

Table 3
R.m.s. deviations (Å) between the current refined model and previously reported models for TEWL.

| Model	1tew	2lz2	3lz2	135l
A	0.96	1.02	1.00	1.06
1tew	—	0.45	0.445	0.66
2lz2	—	—	0.41	0.70
3lz2	—	—	—	0.74

1tew model gives a preferable stereochemistry when refined to the same values for the profile R factors. This might reflect the higher quality of the initial model, which was refined from a higher resolution data set (1.65 Å for 1tew *versus* 2.2 Å for 2lz2). The final refinement reported here is that based on the 1tew starting model.

3.4. Molecular structure

In the present study, polycrystalline precipitate was obtained from amorphous lyophilized powder by preparing solutions of relatively high protein and salt concentrations. These conditions are quite different to those used in the preparation of single-crystal specimens and thus some structural modifications might be expected. We believe that the high-resolution powder profiles presented here offer sufficient information for a structural analysis. Since the quality of fit is significantly better than that found in single-crystal refinements at comparable resolution, overfitting cannot be ruled out. However, manual rebuilding of the structure generally leads to large increases in the profile R factor, giving a clear indication that the data are sensitive to fine conformational details within the structure. In this case, the manual rebuilding steps mainly consisted of selecting different rotamers for certain side chains in order to avoid clashes and alleviate bad contacts. The restrained least-squares procedure has been described in further detail elsewere (Von Dreele, 2005). Furthermore, to test the validity of the final model, refinements were carried out starting from each of the possible models [1tew (Howell, 1995), 3lz2 (Howell *et al.*, 1992), 2lz2 (Parsons & Phillips, unpublished work)]. Refinements involving 1tew converged to better final R values than those using 2lz2 and 3lz2, suggesting that the refinement was also sensitive to the initial model used. Only the 1tew model which resulted in the best quality of fit is presented here. In the original 1tew structure (Howell, 1995), as in other TEWL structures, several flexible loop regions proved difficult to model and the same regions presented problems during the refinement of the powder structure.

3.5. Structure description and comparison

TEWL is a relatively compact structure comprised of eight helices (residues 5–8, 12–15, 26–35, 60–63, 81–83, 89–99, 109–114 and 121–123) and four antiparallel β-sheets (residues 1–3, 38–40, 44–46 and 50–52). The overall topology and secondary structure of the powder result are essentially the same as in the previously determined single-crystal structures (see Table 3 for r.m.s. deviations). However, one notable secondary-
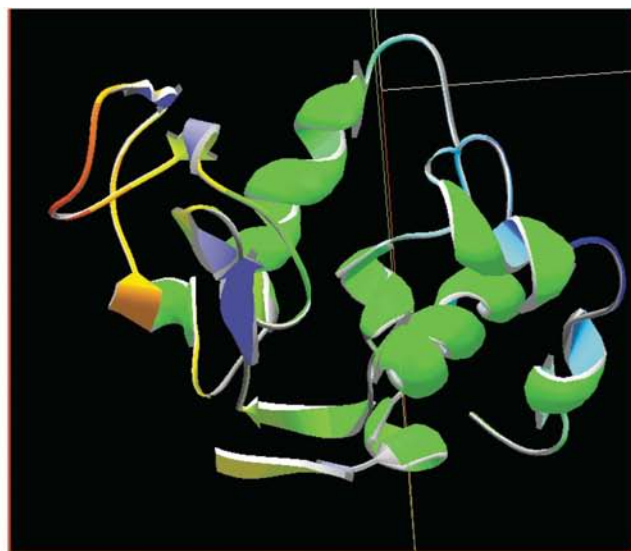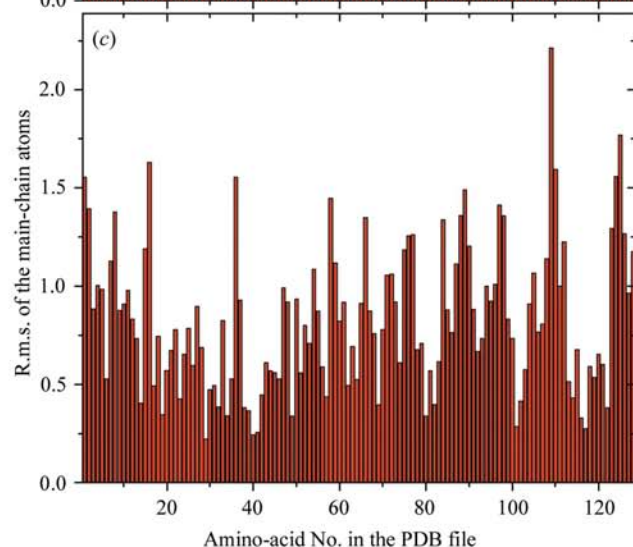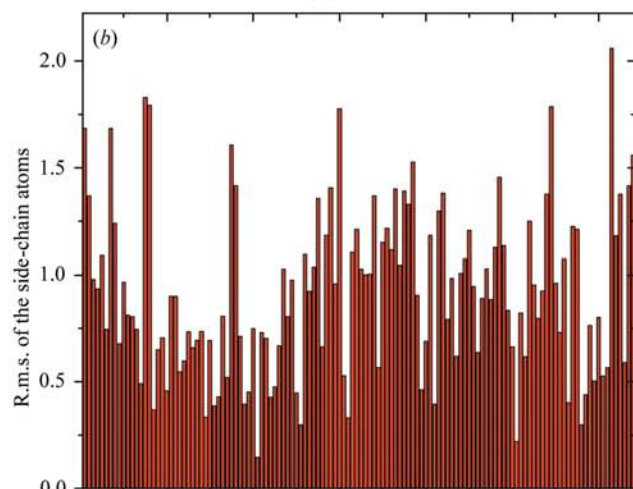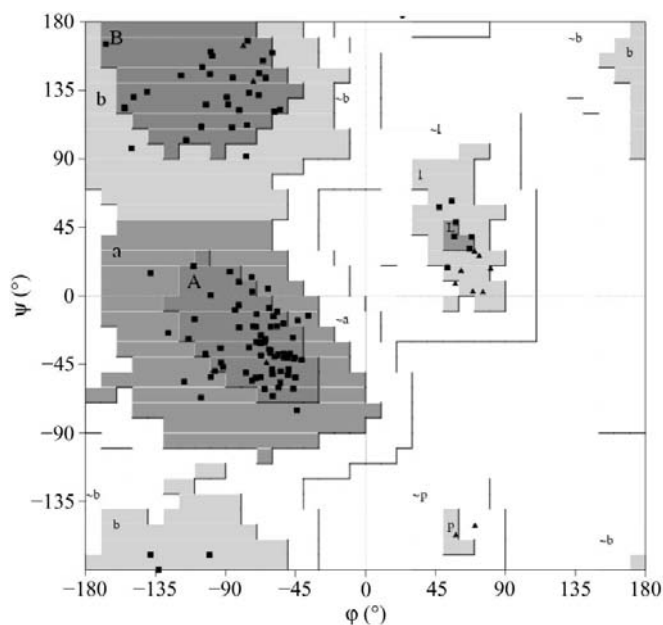
Figure 3

(a) Refined conformation of TEWL (sample A) at 295 K illustrated as ribbon. The colors represent the r.m.s. variation at different regions of the main chain, with dark blue and red corresponding to the minimum and the maximum values, respectively. (b) R.m.s. deviations of the main-chain atoms and (c) of the side-chain atoms between the refined (sample A) and the original 1tew model plotted against residue number.

structural difference occurs in the first helix, which incorporates a short loop region between residues Leu8 and Ala10 in the powder structure. Fig. 3 shows the backbone structure and illustrates the r.m.s. deviation between the refined model for the A sample at room temperature and the hexagonal TEWL structure 1tew. The average r.m.s. deviation of the C^α-atom positions after aligning the structures in *Swiss-PdbViewer* was determined to be 0.95 Å over 128 residues. The deviation between the structures may arise in part as a consequence of the differences in the range and type of data used (d_{min} = 1.65 Å for 1tew and d_{min} = 3.35 Å for the powder structure). Overall, the refined molecular structure and geometry are in good agreement with previous structures. The Ramachandran plot for the main-chain torsion angles (φ, ψ) (Ramakrishnan & Ramachandran, 1965) (Fig. 4) indicates that there are no torsion angles falling in disallowed regions. The mean positional error in the coordinates was estimated from a Luzzati plot of R factor against $1/d$ (Luzzati, 1952) to be 0.227 Å. In comparing the powder structure with previous TEWL structures, the main differences occur in three flexible loop regions, notably in the β-hairpin between residues 46 and 50 but also in loops between residues 60–62 and 101–107. The region between Arg46 and Gly50 has been reported to be flexible (Howell *et al.*, 1992), as demonstrated by much higher B values in the monoclinic structure (Harata, 1993). The regions spanning residues 60–62, 70–74 and 101–107 are also flexible and are involved in making intermolecular contacts that vary slightly between the structures. Difficulties were encountered in locating the side chains of two tryptophan residues (Trp62
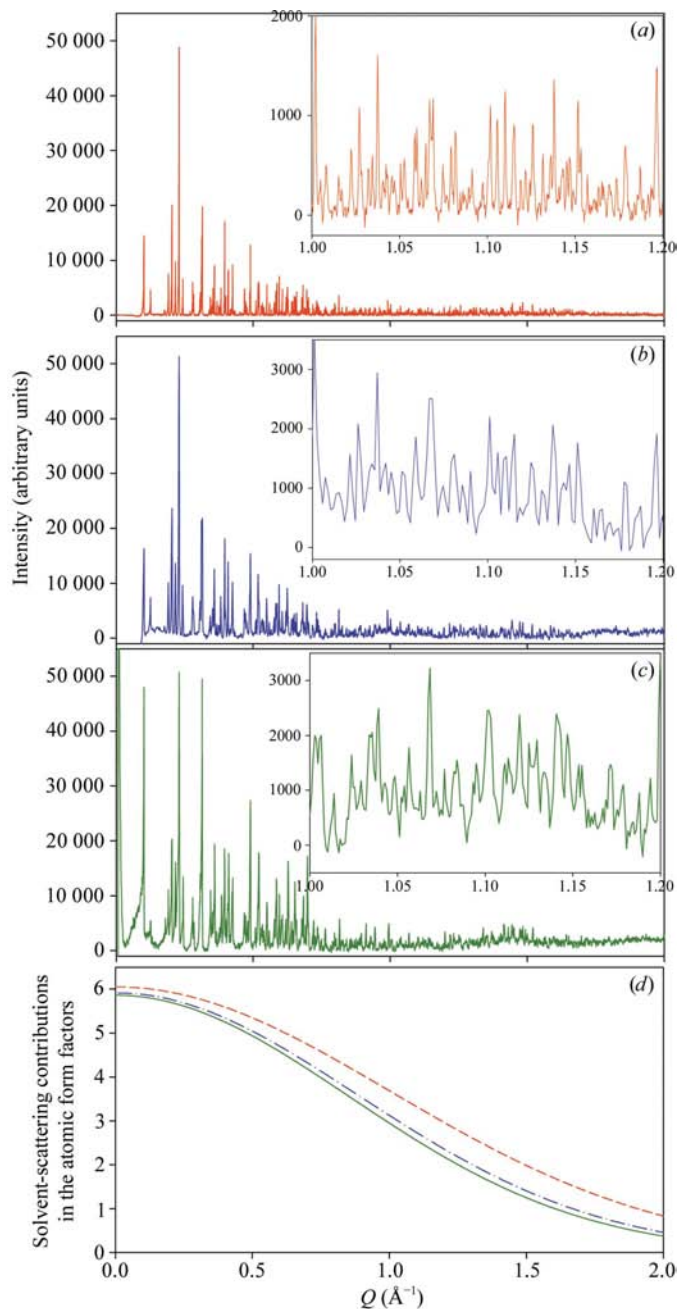

Figure 4
Ramachandran plot for main-chain torsion angles (φ, ψ). Glycine residues are represented by triangles. The shading indicates allowed regions of conformational space; the darker the shading, the more favourable the region is. The shaded areas are labeled with respect to the secondary-structural element associated with the region and how favourable the region is: A, a or ~a represents the α-helical region, with A being the most favourable; B, b or ~b similarly represent the β-sheet region and l or ~l loops.

and Trp63) in the interface between two symmetry-related molecules. The electron density is ill-defined in this region, indicative of a higher conformational flexibility; however, Trp62 appears to make an intermolecular contact. Trp62 has been implicated in the catalytic mechanism and is proposed to form interactions with sugar residues in the active site during catalysis. In the structures of several inhibitor complexes with avian lysozymes (Strynadka & James, 1991; Cheetham *et al.*, 1992; Turner & Howell, 1995), Trp62 is repositioned and drawn into the active-site cleft when a saccharide molecule is bound. Other problematic regions that were difficult to build included the N-terminal region and the C-terminus residues Arg128 and Leu129. As in previous reports, Leu129 could not be located in the density and was thus omitted from the final model. The crystal packing is similar to that observed in the previous hexagonal TEWL structures (Howell, 1995; Howell *et al.*, 1992; Parsons & Phillips, unpublished work) and the intermolecular contacts mirror closely those observed in the 3lz2 structure (Howell *et al.*, 1992). An extensive network of contacts, presumed to be hydrogen bonds, connect the regions spanning Gly71–Asn77 and Ser100–Lys116 in symmetry-related molecules in the powder structure. Intermolecular interactions involving Asp103, Ser72 and Lys73 partially block the saccharide-binding site; however, the active site remains open to the bulk solvent and important catalytic residues including Glu35 and Asp52 are largely accessible.

3.6. Variation with pH

In contrast to data collected for sample *A*, the high-resolution diffraction profiles for samples *B* and *C* are of relatively inferior quality, possibly owing to an incomplete crystallization process or a higher percentage of liquid solution, which resulted in elevated backgrounds and a poorer signal-to-noise ratio. However, reliable lattice parameters could be extracted for the two data sets. All the reflections in

Figure 5
X-ray powder diffraction profiles of TEWL samples (*a*) *A*, (*b*) *B* and (*c*) *C* (ID31 at 295 K). (*d*) Solvent-scattering contribution to atomic scattering factors calculated for the three different samples. The red (dashed), blue (dot/dash) and green (solid) lines correspond to samples *A*, *B* and *C*, respectively.
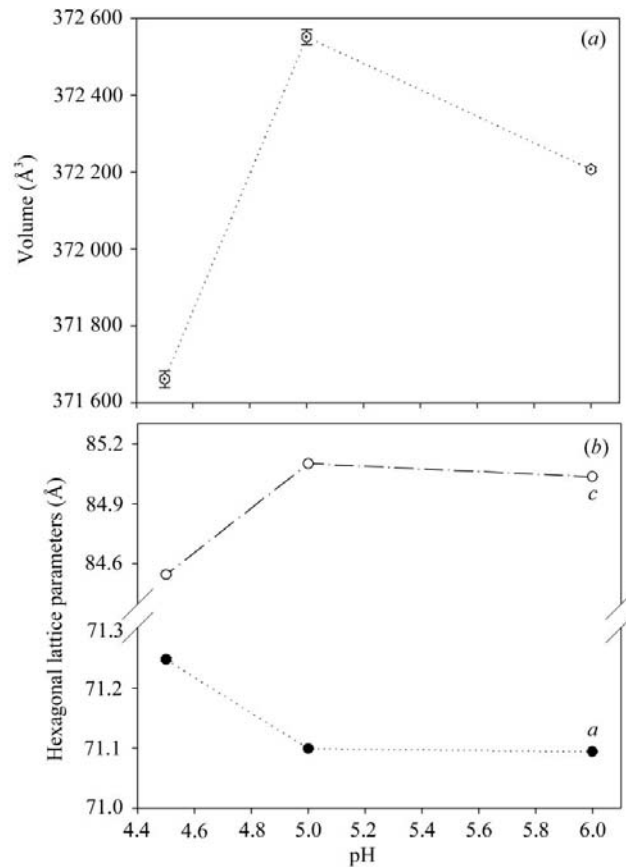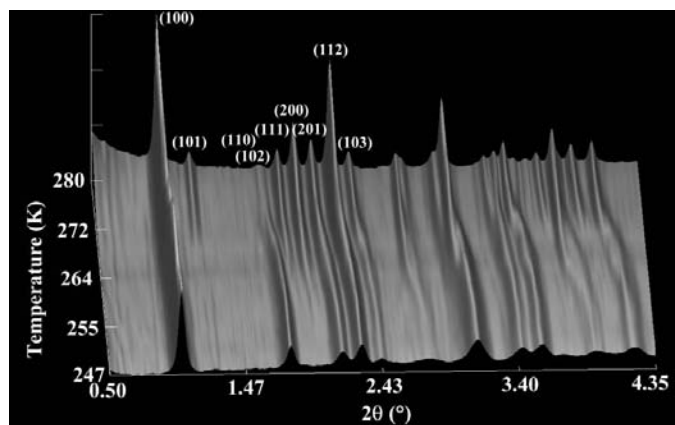

Figure 6
Evolution of refined lattice parameters (*a*) and volume of the hexagonal unit cell (*b*) of TEWL with pH.

the three diffraction profiles collected for samples B and C were indexed with essentially the same unit cell, implying the absence of any structural transitions with respect to the structure of sample A. However, the high-resolution patterns are quite distinct, with the intensities of the low-angle peaks varying significantly between samples (Fig. 5) and a correlation between unit-cell parameters and crystallization conditions is evident. The former effect can be explained in terms of altered solvent-scattering contribution to the atomic form factors being more pronounced at low angles, as shown in Fig. 5(*d*). Precisely, the lattice parameters obtained for sample A at room temperature [$a = 71.0862$ (3), $c = 85.0276$ (5) Å] are very close to the values given for single crystals formed in pH 7.0 solutions ($a = 71.000$, $c = 84.900$ Å). The lattice parameters corresponding to sample B [$a = 71.093$ (46), $c = 85.100$ (6) Å] are larger than those of A by 0.009 and 0.085% along the a and c crystallographic axes, respectively. For sample C [$a = 71.236$ (3), $c = 84.534$ (4) Å] the changes are $\Delta a/a = 0.21\%$ and $\Delta c/c = -0.58\%$. These small changes in volume of the unit cell indicates that it is in the ratio of unit-cell parameters where most of the changes occur (Fig. 6). Rietveld refinements for all samples yield a small microstrain broadening but no measurable particle-size broadening effects. The variations of the unit cell with pH indicate that there is a significant change between the low-pH and high-pH structures, probably arising from differing intermolecular interactions. This is in contrast to Von Dreele (2000, 2001), where little or no microstrain and only particle size were observed.

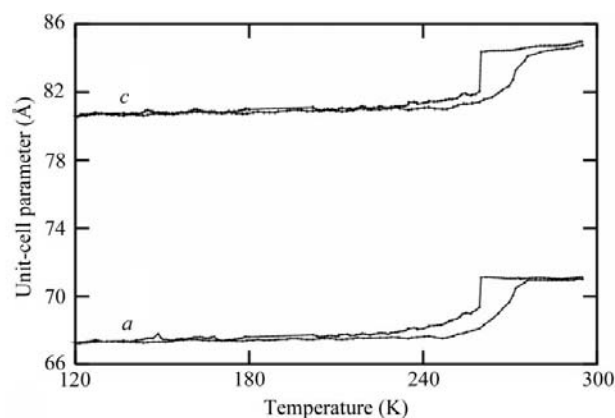### 3.7. Variation with temperature

Additional measurements were performed on sample A at the ID11 beamline of the ESRF in order to observe the evolution of the unit cell with temperature. Despite the fact that a cryoprotectant (PEG) had been added to the solution, ice formation was observed upon cooling. The data quality obtained from the area detector was found to have significantly poorer peak resolution compared with that from the parallel beam geometry at ID31; however, the counting

statistics are much improved and no radiation-damage problems were observed during the course of the experiment (Fig. 7). Upon cooling, the sample displays a sharp phase transition accompanied by the appearance of peaks that could be accounted for as ice (unit-cell parameters $a = 4.51$, $c = 7.35$ Å, hexagonal). Peaks arising from the protein were observed to be significantly broadened upon freezing and could only be resolved to a resolution of $d_{min} \simeq 8$ Å, compared with $d_{min} \simeq 3$ Å for the room-temperature data. The data remained similar down to a temperature of 120 K, with the only significant change below the freezing point being the intensity of the (101) reflection, which is slightly enhanced at lower temperatures. Upon returning the sample to room temperature, a continuous transition from the frozen state back to the room-temperature state was observed.

Owing to the paucity of the data for the low-temperature state, only rigid-body Rietveld refinements were attempted in order to model the low-temperature structure. Unit-cell and peak-shape parameters could be extracted and these were essentially identical when a pattern-decomposition technique was used instead of a structural model. The variation of the unit-cell and peak-shape parameters upon cooling are shown in Figs. 8 and 9, where the model for the peak shape was a Gaussian having width parameterized *versus* 2θ as follows,
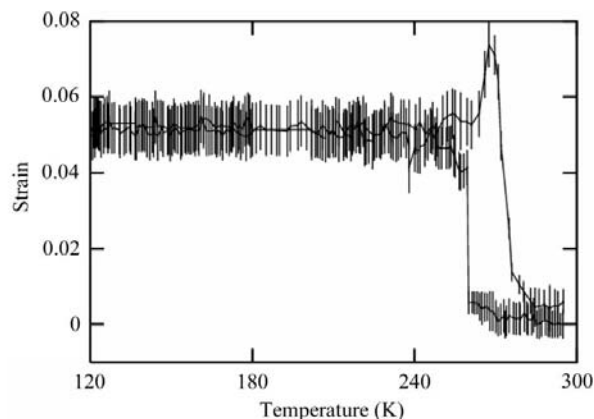

Figure 7
Diffraction profiles plotted as a function of temperature collected while warming the TEWL sample through the freezing transition. The sample was initially cooled from 320 to 105 K at a rate of 80 K h^{-1} and then warmed back to 320 K at a rate of 160 K h^{-1}.

Figure 8
Variation of unit-cell parameters with temperature.

Figure 9
Variation of peak-shape parameters with temperature.

$$\mathrm{FWHM}^2 = U\tan^2\theta + P/\cos^2\theta.$$

This formulation was sufficient to fit the data and has the advantage that the parameters U and P can easily be related to the microstructure of the material (Larsen & Von Dreele, 2001; *GSAS* manual). Strain broadening in real space gives $\Delta d/d$ as a constant, which can be related to the U parameter *via*

$$\mathrm{Strain} = \frac{\pi}{18000}[(8\ln 2)(U - U_i)]^{1/2}\,\%,$$

where U_i is the instrumental contribution. For particle-size broadening effects the relationship is

$$\mathrm{Size} = \frac{18000K\lambda}{\pi(8\ln 2)^{1/2}P}\,\mathrm{nm},$$

although the numerical values for the particle-size broadening effects were not significant owing to the poor instrumental resolution. By comparison to the high-resolution data collected on ID31, the patterns in the room-temperature state represent the instrumental contributions to the peak-shape function and allow the broadening at low temperatures to be quantified. The change in unit-cell parameters is $\Delta a/a \simeq 5\%$ and $\Delta c/c \simeq 5\%$, which is in good agreement with the measured 'strain' parameter of 5% for the low-temperature state. Assuming there is negligible transport into or out of the protein unit cell during freezing, the density of the material must increase substantially.

The results regarding the crystal structure from the rigid-body refinements are less satisfactory. Owing to the collapse of the unit cell, the molecules begin to overlap with each other and the structure is physically unreasonable. Attempts to include an 'anti-bump' restraint into the refinement prevented the model from converging or explaining the data. Nevertheless, if the structure is interpreted only at very low resolution (of ~20 Å), it appears that the diameter of the channels which lie parallel to the c crystallographic axis is increased at low temperature (~7.7 Å) compared with room temperature (~5.3 Å), although this change is small compared with the effective resolution. Given the overall contraction of the unit cell, this result is surprising. It would seem that the bulk of the contraction takes place in parts of the unit cell occupied by protein molecules. The severe peak broadening suggests the low-temperature phase is more disordered, with a variety of conformations being adopted in different unit cells.

Because of the difficulties related to molecular overlap in the rigid-body refinement of the low-temperature data, a comparison based on the intensities of the Bragg reflections was also carried out. The largest change in intensity is the suppression of the (200) peak, which almost disappears compared with room-temperature data, where it is fairly intense. At room temperature the (100) and (200) reflections both have the same phase (180°) and their intensities are influenced mainly by the channels along the c axis. The suppression of the (200) peak is accounted for in the rigid-body refinement by increasing the diameter of this channel. Generating two-dimensional Fourier maps of the structure projected along the c axis gives a more direct view of the changes occurring upon freezing. By assuming that the phases of the reflections (100), (110) and (200) are all unchanged on passing through the phase transition, a low-resolution map can be produced where the only uncertainty is in the extraction of the intensity values. The conservation of phase seems reasonable, since these centric reflections would need to pass through zero during the continuous phase transition and they do not appear to do so. Fig. 10 shows the generated Fourier map, which illustrates that the changes in intensity can be accounted for by a smoothing out of the electron density, which makes the apparent channel diameter seem larger. A more reasonable hypothesis for the contraction of the unit cell is that the formation of an ice phase in the bulk solvent leads to a rapid increase in the osmotic potential of the crystallization buffer. It is conceivable that this would lead to transport of liquid water out of the crystallites and would be manifested in a rapid contraction of the solvent channels and, by extension, the unit cells. While this explains the contraction, the changes in peak intensities are difficult to predict under those circumstances.

4. Conclusions

Powder diffraction methods can provide a wealth of complementary information about the structure and properties of protein crystals. We have shown that when a good model for the three-dimensional structure of a protein molecule is available, molecular-replacement techniques can be used to find out how the molecule packs into a unit cell for a particular structural modification. The precise structure of the molecule may also be refined by using stereochemical restraints in addition to the usual Rietveld method for profile refinement. While significant improvements in fit and in the actual stereochemistry can be produced by this procedure, it remains unclear how far the structure is determined by the data and how much of a problem of model bias remains. Owing to the collapse of three-dimensional reciprocal space into the one dimension of a powder pattern and the increasing peak overlap at higher angles, the information content of the data is not reflected by the minimum d-spacing to which data are considered. In this case, we used data to a d-spacing of 3.35 Å, but the real structural resolution is closer to 4–5 Å in terms of information content and this is reflected in the multiple minima found during the course of refinements. We believe that with high-resolution data the technique would be powerful enough to be applied routinely to small proteins as a confirmatory tool when models for the folding of the protein are available.

By examining the variations of the structure with pH and when it is frozen, we hope to have illustrated those aspects of the powder technique that can be applied most easily to protein samples. The variation of unit-cell parameters with pH indicates a marked change for samples A and B (unbuffered pH 6 and 5, respectively) *versus* C (pH 4.5). The results from the low-temperature experiments indicate how strategies for cryocooling and cryoprotecting samples might be more easily developed with powdered samples than with more valuable
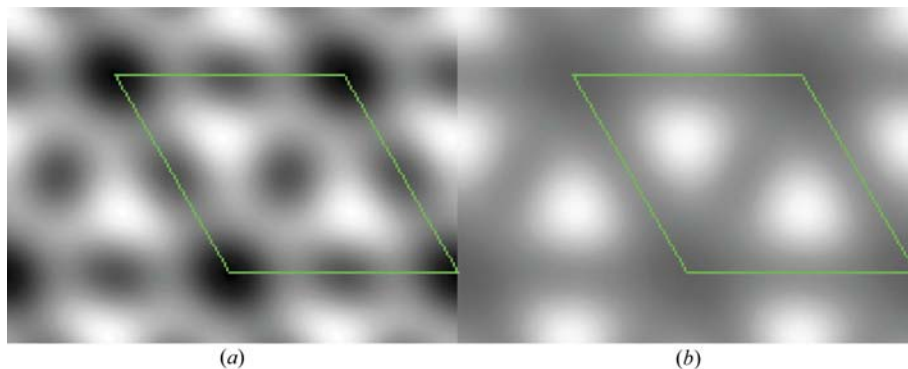
Figure 10
Projections of the electron density in the *ab* plane for TEWL at room temperature (*a*) and low temperature (*b*). The maps are constructed from the three reflections (100), (110) and (200), assuming the phase angles do not change on passing through the freezing transition. The dark- and light-colored areas correspond to low and high electron density, respectively.

and rare single crystals. These strategies would depend on the quantities of sample available; however, preliminary results show that data can be obtained in some cases using as little as 35 μg of crystalline material. Microcrystalline material can be screened for quality of diffraction while increasing the concentration of cryoprotectant in fine increments using a resealable capillary. In this specific case, we have shown that TEWL underwent a dramatic contraction upon freezing concurrent with the freezing of the mother liquor and a large decrease in diffraction quality.

References

Boultif, A. & Louër, D. (1991). *J. Appl. Cryst.* **24**, 987–993.
Cheetham, A. K. & Wilkinson, A. P. (1992). *Angew. Chemie Int. Ed. Engl.* **31**, 1557–1570.
Cheetham, J. C., Artymiuk, P. J. & Phillips, D. C. (1992). *J. Mol. Biol.* **224**, 613–628.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
Colovos, C. & Yeates, T. O. (1993). *Protein Sci.* **2**, 1511–1519.
David, W. I. F., Shankland, K. & Shankland, N. (1998). *Chem. Commun.*, pp. 931–932.
De Wolff, P. M. (1968). *J. Appl. Cryst.* **1**, 108–113.
Evans, J. S. O. & Radosavljevic Evans, I. (2004). *Chem. Soc. Rev.* **33**, 539–547.
Finger, L. W., Cox, D. E. & Jephcoat, A. P. (1994). *J. Appl. Cryst.* **27**, 892–900.
Fitch, A. N. (2004). *J. Res. Natl Inst. Stand. Technol.* **109**, 133–142.
Guex, N. & Peitsch, M. C. (1999). *Swiss-PdbViewer.* Glaxo Wellcome Experimental Research. http://www.expasy.ch/spdbv.
Hammersley, A. P. (1997). ESRF Internal Report ESRF97HA02T.
Harata, K. (1993). *Acta Cryst.* **D49**, 497–504.
Harris, K. D. M. & Cheung, E. Y. (2004). *Chem. Soc. Rev.* **33**, 526–538.
Harris, K. D. M., Tremayne, M. & Kariuki, B. M. (2001). *Angew. Chemie Int. Ed.* **40**, 1626–1651.
Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. (1996). *Nature (London)*, **381**, 272.
Howell, P. L. (1995). *Acta Cryst.* **D51**, 654–662.
Howell, P. L., Almo, S. C., Parsons, M. R., Hajdu, J. & Petsko, G. A. (1992). *Acta Cryst.* **B48**, 200–207.
Larson, A. C. & Von Dreele, R. B. (2004). *General Structure Analysis System (GSAS).* Los Alamos National Laboratory Report LAUR 86-748.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Le Bail, A., Duroy, H. & Fourquet, J. L. (1988). *Mater. Res. Bull.* **23**, 447–452.
Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.
Markvardsen, A. J., David, W. I. F., Johnson, J. C. & Shankland, K. (2001). *Acta Cryst.* **A57**, 47–54.
Matthewman, J. C., Thompson, P. & Brown, P. J. (1982). *J. Appl. Cryst.* **15**, 167–173.
Nixon, P. E. & North, A. C. T. (1976). *Acta Cryst.* **A32**, 320–325.
Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.
Rietveld, H. M. (1969). *J. Appl. Cryst.* **2**, 65–71.
Sarma, R. & Bott, R. (1977). *J. Mol. Biol.* **113**, 555–565.
Strynadka, N. C. J. & James, M. N. G. (1991). *J. Mol. Biol.* **220**, 401–424.
Toby, B. H. (2001). *J. Appl. Cryst.* **34**, 210–213.
Turner, M. A. & Howell, P. L. (1995). *Protein Sci.* **4**, 442–449.
Vagin, A. A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* **D55**, 191–205.
Vaney, M. C., Maignan, S., Riès-Kautt, M. & Ducruix, A. (1996). *Acta Cryst.* **D52**, 505–517.
Von Dreele, R. B. (1999). *J. Appl. Cryst.* **32**, 1084–1089.
Von Dreele, R. B. (2000). *Acta Cryst.* **D56**, 1549–1553.
Von Dreele, R. B. (2001). *Acta Cryst.* **D57**, 1836–1842.
Von Dreele, R. B. (2005). *Acta Cryst.* **D61**, 22–32.
Wright, J. P. (2004). *Z. Kristallogr.* **219**, 791–802.
Wright, J. P. & Forsyth, J. B. F (2000). Rutherford Appleton Laboratories Report RAL-TR-2000-012.