Acta Cryst. (1995). D**51**, 654–662

Structure of Hexagonal Turkey Egg-White Lysozyme at 1.65 Å Resolution

BY P. LYNNE HOWELL

Division of Biochemistry Research, Hospital for Sick Children, 555 University Avenue, Toronto, M5G 1X8, Canada, and Department of Biochemistry, Faculty of Medicine, University of Toronto, Medical Sciences Building, Toronto, M5S 1A8, Canada

(*Received* 12 *August* 1994; *accepted* 23 *November* 1994)

Abstract

The structure of hexagonal turkey egg-white lysozyme (TEWL) has been determined and refined at 1.65 Å resolution. The crystals were grown from a 150 m*M* potassium thiocyanate solution at pH 4.5 and belong to space group $P6_122$ with unit-cell dimensions $a = b = 70.96$, $c = 83.01$ Å, $\alpha = \beta = 90$, $\gamma = 120°$. The crystals were isomorphous with those of hexagonal pH 8.0 TEWL. The coordinates of PDB entry code 3LZ2 were therefore used as the initial model and subjected to rigid-body refinement, simulated annealing and least-squares refinement to a final residual of 0.20. The root-mean-square deviations from the ideal bond distances and angles were 0.016 Å and 2.2°, respectively. During the refinement, 86 water molecules and one thiocyanate ion were located in the structure. The thiocyanate ion lies close to the interface between two symmetry-related molecules. The S atom of the ion forms two direct intermolecular contacts with Arg14 and interacts indirectly *via* a network of water molecules to Arg5 of a symmetry-related molecule. The structure provides direct evidence for the mode of thiocyanate binding to arginine residues and suggests a possible mechanism for the efficiency of thiocyanate in crystallizing basic proteins.

Introduction

The structure of hexagonal turkey egg-white lysozyme (TEWL) at high pH has been determined previously at 2.5 Å resolution using data collected by the Laue method (Howell, Almo, Parsons, Hajdu & Petsko, 1992) and at 2.2 Å resolution using monochromatic data-collection techniques (Parsons, 1988; Parsons & Phillips, 1988). TEWL has also been crystallized in a monoclinic space group and the structure solved at 1.3 Å resolution (Harata, 1993). Turkey egg-white lysozyme belongs to the type C class of lysozymes which have been studied extensively using X-ray crystallographic techniques (Blake *et al.*, 1965; Moult *et al.*, 1976; Aschaffenburg *et al.*, 1980; Artymiuk & Blake, 1981; Artymiuk, Blake, Rice & Wilson, 1982; Berthou, Lifchitz, Artymiuk & Jollès, 1983; Kundrot & Richards, 1987; Kodandapani, Suresh & Vijayan, 1990; Ramanadham, Sieker & Jensen, 1989; Rao, Hogle & Sundaralingam, 1983; Parsons & Phillips, 1988; Howell, Almo, Parsons, Hajdu & Petsko, 1992; Harata, 1993; Lescar, Souchon & Alzari, 1994; Turner & Howell, 1995*a*).

In an effort to understand the process of crystallization, Riés-Kautt & Ducruix (1989, 1991) have examined the relative effectiveness of various salts on the solubility and crystallizability of hen egg-white lysozyme (HEWL). They found that for HEWL the effectiveness of the ion in crystallization experiments was linked to the inverse Hoffmeister series (*i.e.* $SCN^- > NO_3^- > Cl^- > HC_6H_5O_7^- > CH_3COO^- > H_2PO_4^- > SO_4^-$). Although thiocyanate is a well known chaotropic agent at high concentrations, at low concentrations, as little as 75 m*M*, it was found to be very effective in crystallizing HEWL. Since Pande & McMenanamy (1970) had previously postulated that thiocyanate ions would bind preferentially to arginine side chains, Riés-Kautt & Ducriux (1991) proposed that the effectiveness of the thiocyanate ion in crystallization was due to it binding to basic side chains on the protein. They provided direct evidence of the binding of a thiocyanate ion to a protein molecule in a structure determination of erabutoxin b at 1.7 Å resolution (Saludjian *et al.*, 1992) (PDB entry code: 6EBX). In this structure the N atom of the thiocyanate ion interacted with the side chains of Arg33A and Ser23B, while the S atom formed longer contacts to the carbonyl O atom of a Cys54B and a water molecule.

The molar concentration of thiocyanate used to crystallize TEWL in this study was 150 m*M*, comparable to the values used previously for HEWL and well below the typical concentrations (~2 *M*) required for crystallization with sodium chloride or ammonium sulfate. The space group of the hexagonal TEWL presented in this paper is identical to the previous structure determinations (Howell *et al.*, 1992; Parsons, 1988), but with a *c*-axis dimension that is slightly shorter in length. This shortening of the *c* axis had been noted earlier in hexagonal crystals of TEWL grown at pH 4.2 from 2.2 *M* sodium nitrate (Harata, 1993). However, the crystals grown from potassium thiocyanate diffract to a limiting resolution of 1.65 Å, considerably higher resolution than previously obtained for hexagonal TEWL. This structure determination has revealed the location of a thiocyanate ion, with a different mode of binding to the protein molecule than previously reported (Saludjian

© 1995 International Union of Crystallography
Printed in Great Britain – all rights reserved

ISSN 0907-4449 © 1995

et al., 1992). The structure presented confirms earlier hypotheses that the effectiveness of thiocyanate in crystallizing basic proteins is due to the formation of strong interactions between protein molecules and provides a possible mechanism for the effectiveness of thiocyanate in crystallizing avian lysozymes and other basic proteins.

Materials and methods

Crystallization and data collection

Turkey egg-white lysozyme (TEWL) was purchased from the Sigma Chemical Company. Crystals were grown at room temperature by the hanging-drop vapour-diffusion method from a 15 mg ml^{-1} protein solution in 50 m*M* sodium acetate buffer at pH 4.5. The 20 µl hanging drop (10 µl protein solution and 10 µl precipitating solution) was equilibrated against a 1 ml reservoir of 150 m*M* potassium thiocyanate in 50 m*M* sodium acetate buffer at pH 4.5. The crystals are hexagonal rods and grew to maximal size (1 × 0.4 × 0.4 mm) in approximately 9–13 days. From precession photography the space group was determined to be $P6_122$ or $P6_522$ with unit-cell dimensions $a = b = 70.96$, $c = 83.01$ Å, $\alpha = \beta = 90$, $\gamma = 120°$. For data collection the crystals were mounted in quartz capillaries with the long axis parallel to the tube.

Data were collected from a single crystal using a Siemens–Xentronics area detector with graphite-monochromated Cu $K\alpha$ radiation from a Rigaku RU-200 rotating-anode generator operating at 50 kV and 70 mA. Data processing and scaling were carried out using the *XDS* program package (Blum, Metcalf, Harrison & Wiley, 1987; Kabsch, 1988*a*,*b*) and *SORTAV* (Blessing, 1987). A total of 27 701 measured intensities were reduced to 15 265 unique reflections in the resolution range 10–1.65 Å (97% complete) with an overall merging *R* factor of 0.053 and 80% of the data in the 1.7–1.65 Å shell have $F > 2\sigma(F)$.

Structure determination and refinement

The structure of hexagonal TEWL (3LZ2) (Howell *et al.*, 1992) was taken as the initial starting model for crystallographic refinement. The initial *R* factor calculated using data between 8 and 2.5 Å resolution was 56%. Although the initial *R* factor was high, the packing of the protein molecules in the crystal was assumed to be identical to the starting model. The initial model was

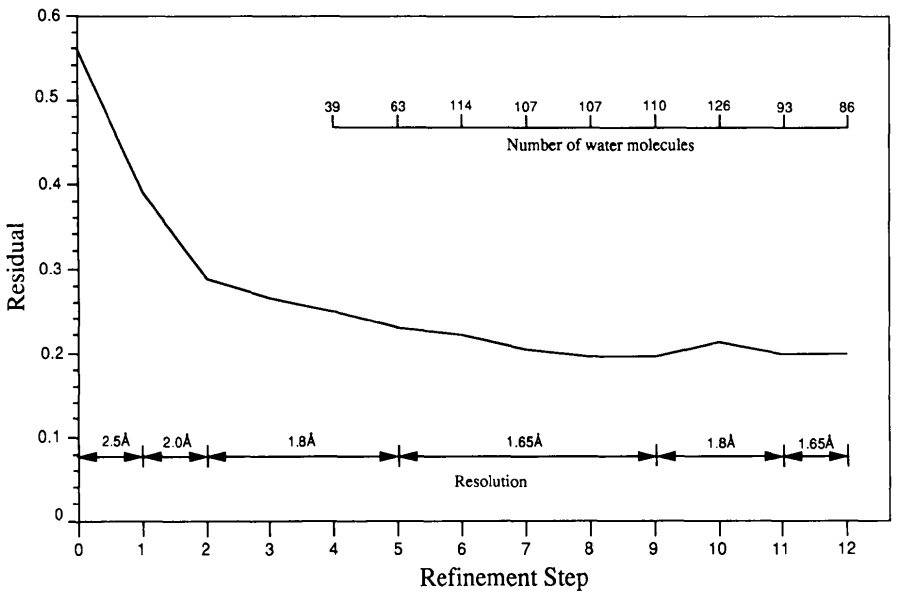

Fig. 1. Refinement of TEWL. Plot of *R* factor *versus* the refinement process. The *R* factor is the crystallographic agreement factor defined as $\sum |F_{obs} - F_{calc}| / \sum F_{obs}$ where the observed quantities are measured reflection amplitudes and the calculated quantities are computed from the atomic coordinates of the current structural model. The starting model for refinement was the previously determined pH 8 structure, PDB code: 3LZ2 (Howell, Almo, Parsons, Hajdu & Petsko, 1992). Refinement steps 1–7 used the program *X-PLOR* and subsequent refinement steps utilized the least-squares refinement program, *PROFFT*. High-resolution data were included gradually over the course of the refinement as indicated in the diagram. During steps 9–11 the data was temporarily truncated to 1.8 Å resolution. Steps 10–11 represent several cycles of refinement and manual rebuilding where the overall residual varied very little. The position of water molecule and their intermolecular interactions were being checked. The number of water molecules in Fig. 1 represents the total number of water molecules at the end of each refinement cycle.

therefore subjected to 40 cycles of rigid-body minimization using the program *X-PLOR* (Brünger, Kuriyan & Karplus, 1987; Brünger, 1992). This reduced the R factor to 39.2% for data between 8 and 2.5 Å resolution and with $F > 2\sigma(F)$. Refinement was continued using data between 8 and 2.0 Å resolution using a protocol which consisted of $C\alpha$-atom restrained minimization, a high-temperature simulated-annealing refinement, a room-temperature simulated-annealing refinement followed by further conjugate-gradient minimization using *X-PLOR*. Full details of the procedure and subsequent steps in the refinement (Fig. 1) using both *X-PLOR* (steps 1–7) and *PROFFT* (steps 8–12) (Konnert, 1976; Konnert & Hendrickson, 1980) are given below. Manual rebuilding and the addition of water molecules was undertaken after every step in the refinement process from step 4 onwards. Data with $F > 2\sigma(F)$ were used in the *X-PLOR* refinement steps, all data were used in the *PROFFT* steps.

Step 1. 40 cycles of rigid-body minimization.

Step 2. 80 conjugate-gradient minimization steps with soft repulsive potential, followed by 40 conjugate-gradient minimization steps with the non-bonded repulsions switched off. C_α restraints at 20 kcal mol^{-1} Å^{-2} (84 kJ mol^{-1} Å^{-2}), $\Delta F = 0.05$, where ΔF is the maximum movement of atoms allowed before updating the structure-factor calculation. This was followed by a 0.25 ps simulation, T = 2000 K, time step 0.5 fs, ΔF = 0.2, followed by a 0.5 ps simulation, T = 300 K, time step 0.5 fs, ΔF = 0.2 and finally 120 conjugate-gradient minimization steps.

Step 3. Data were extended to 1.8 Å resolution and the following protocol run. 80 conjugate-gradient minimization steps, C_α restraints at 20 kcal mol^{-1} Å^{-2} (84 kJ mol^{-1} Å^{-2}), ΔF = 0.05. A 1 ps simulation, T = 2000 K, time step 0.5 fs, ΔF = 0.2, a 0.25 ps simulation, T = 300 K, time step 0.5 fs, ΔF = 0.2 and finally 80 conjugate-gradient minimization steps.

Step 4. 40 cycles of B-factor refinement were followed by 80 steps of conjugate-gradient minimization. First cycle of manual rebuilding. 39 water molecules were added.

Step 5. 40 cycles of conjugate-gradient minimization followed by 20 cycles of B-factor refinement. This procedure was repeated twice. Manual rebuilding enabled 63 water molecules to be located in the structure.

Step 6. 80 conjugate-gradient minimization steps, C_α restraints at 20 kcal mol^{-1} Å^{-2} (84 kJ mol^{-1} Å^{-2}), ΔF = 0.05. A 1 ps simulation, T = 2000 K, time step 0.5 fs, ΔF = 0.2, a 0.25 ps simulation, T = 300 K, time step 0.5 fs, ΔF = 0.2. 80 conjugate-gradient minimization steps were followed by 40 cycles of B-factor refinement. Manual rebuilding located 114 water molecules in the structure.

Step 7. 40 cycles of conjugate-gradient minimization, 20 cycles of B-factor refinement followed by 40 cycles of conjugate-gradient minimization. Manual rebuilding located 107 water molecules. A SCN^- ion was located in the structure.

Step 8. 19 cycles of restrained least-squares refinement were performed, including 107 water molecules and one SCN^- ion.

Step 9. Five cycles of restrained least-squares refinement were performed followed by manual rebuilding. One SCN^- and 110 waters included in the structure.

Step 10. Data truncated to 1.8 Å resolution, ten cycles of restrained least-squares refinement were performed. One SCN^- and 99 waters included in the structure. A manual rebuild was followed by four cycles of refinement with 150 water molecules included.

Step 11. Novel weighting scheme applied to the weak-resolution data at high resolution (see text). Seven cycles of refinement, with 92 water molecules, an SCN^- ion and one alternative conformation for residue 41. Manual rebuilding was followed by nine cycles of refinement with 126 water molecules included. Alternative conformations for residue 41 and 100 included in the refinement. Manual rebuild, followed by 12 cycles of refinement with 93 water molecule, one SCN^- ion and alternative conformations for residues 41 and 78. This was followed by two cycles of regularization.

Step 12. Data extended to 1.65 Å resolution and refined using two empirical weighting schemes (see text). The second empirical weighting scheme was applied to data above 2.04 Å resolution. A total of 25 cycles of refinement were performed with 86 water molecules, one SCN^- ion and an alternative conformation for residue 78.

After refinement step 7 it became clear from the examination of difference Fouriers that a thiocyanate (SCN^-) ion could be located in the structure. The SCN^- ion had been modelled in early cycles of refinement as two water molecules. The water molecule with the large positive difference peak was modelled as the S atom. The SCN^- ion was assumed be linear, with bond lengths of 1.17 and 1.65 Å for the N—C and C—S bonds, respectively (Bats & Coppens, 1977; Bats, Coppens & Kvick, 1977). Geometric parameters for the SCN^- ion were included in the ideal geometry file for the subsequent cycles of refinement using *PROFFT*.

During refinement steps 10 and 11, two new weighting schemes were used in the *PROFFT* refinement (Smith, 1995). In the previous cycles of refinement, it appeared that the empirical weighting scheme as a function of $\sin\theta/\lambda$ (Konnert & Hendrickson, 1980) was applying too much weight to the weaker data in the higher resolution ranges. Accordingly, a weighting scheme was used which gave full empirical weight to those reflections with $F_o > 2.0\sigma(F_o)$ but applied a weight equal to one quarter of the empirical weight to those data with $F_o > F_c$ and $F_o < 2.0\sigma(F_o)$; weights of zero were assigned to those data with $F_o < F_c$ and $F_o < 2.0\sigma(F_o)$. Since examination of the $\langle F_o - F_c \rangle$ differences as a function of resolution revealed that the average differences did not decrease linearly with increasing resolution but tended to

Table 1. *Final parameters after last cycle of refinement*

Resolution (Å)	8–1.65
Number of reflections	15112
Number of reflections $[F > 2\sigma(F)]$	14116
Number of protein atoms	967
Number of water molecules	86
Number of thiocyanate	3
Average B value (main chain)	23.29 (±6.72)*
Average B value (solvent molecules)	39.97 (±11.57)*
R.m.s. bond lengths (Å)	0.016 (0.02)†
R.m.s. bond angle (°)	2.2
R factor ‡	20.0§ (20.6)¶
Weighted R factor **	0.228§

* R.m.s. deviations.
† Restraint applied.
‡ R factor $= \sum |F_{obs} - F_{calc}| / \sum F_{obs}$.
§ R facor for $F > 2\sigma(F)$.
¶ R factor for all data.
** Weighted R factor $= [\sum w(F_{obs} - F_{calc})^2]^{\frac{1}{2}} / (\sum wF_{obs}^2)^{\frac{1}{2}}$.

increase significantly in the higher resolution shells, an alternative weighting scheme was applied during step 11 which consisted of a second empirical weighting scheme also a function of $\sin\theta/\lambda$ for those data above 2.04 Å resolution. The parameters for the TEWL model after the completion of the refinement are summarized in Table 1.

Results and discussion

Quality of molecular structure

Lysozyme is a small compact protein of 129 amino acids. The secondary-structure assignment using the program *Vadar* (Wishart, Willard & Sykes, 1994) has defined six helices (residues 4–15, 24–35, 79–83, 88–100, 108–114 and 119–124), a region of β-pleated sheet (residues 41–46, 50–53, 59–64) and a type II turn between residues 19 and 22. Fig. 2 is a stereo C_α representation of the molecule. As expected, the structure is very similar to the high-pH Laue structure (3LZ2) used as the initial model. The average r.m.s. deviation in the main-chain atoms (N, C_α, C) between the final model and the Laue model for residues 1–128 was 0.445 Å, determined using the least-squares fitting program *PROFIT* (Smith, 1994). Fig. 3 shows the Ramachandran plot for the main-chain torsion angles (φ,ψ) (Ramakrishnan & Ramachandran, 1965). No torsion angle falls 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 0.15–0.25 Å.*

Examination of the electron-density maps during and after completion of the refinement revealed several important features. While the electron density for the polypeptide backbone was continuous for most of the protein, very little density could be found for Leu129 which has, therefore, been omitted from the model. In common with previous structure determinations of avian lysozymes, there are a number of flexible loop regions which have ill defined main-chain density. These regions, especially residues 46–50, were difficult to build and the resultant temperature factors are larger than the overall average temperature factor for the protein, indicative of their increased flexibility. The electron density for several arginine residues was found to be truncated. No electron density was found for the side chain of Arg128, for Arg45 (except C_β) and Arg68 (beyond C_δ). These

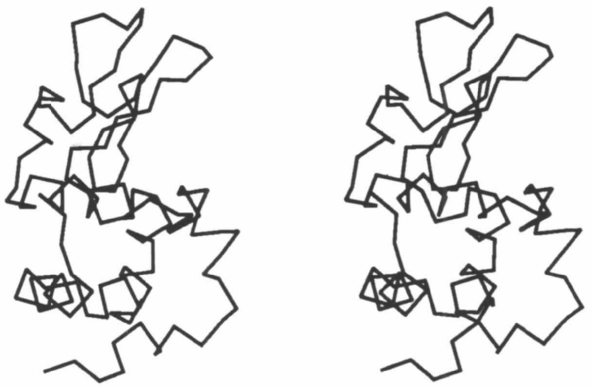

Fig 2. Stereoview of the C_α representation of the refined structure of turkey egg-white lysozyme. The molecule is oriented looking down the first helix.

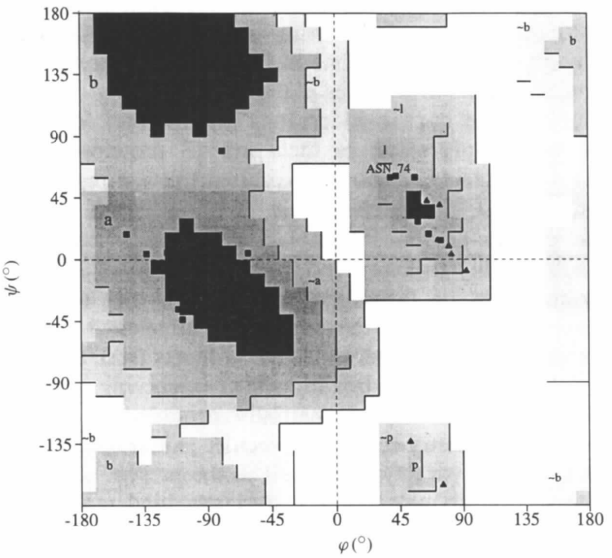

Fig. 3. Ramachandran plot for main-chain torsion angles (φ,ψ). Glycine residues are represented triangles. The shading indicates allowed regions of conformational space, the darker the shading the more favorable the region (Laskowski, MacArthur, Moss & Thornton, 1993). The shaded areas are labeled with 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, similarily for B, b or ~b representing the β-sheet region, and l or ~l for loops.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1TEW, R1TEWSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: AM020).

residues have, therefore, been truncated appropriately in the model. Significant difficulty was encountered in modelling Trp62, which sits at the interface between two protein molecules and interacts with a symmetry-related Trp62. Trp62 is an important residue in the catalytic mechanism making strong hydrogen-bond and van der Waal interactions to two of the sugar residues (B and C) in the active site. When a saccharide is bound in the active site this residue has been shown to undergo a large conformational change (see next section for description) (Strynadka & James, 1991; Cheetham, Artymuik & Phillips, 1992; Turner & Howell, 1995*a*). As it is presently modelled, the shortest distance between the Trp62 and its symmetry-related residue is 1.0 Å, much closer than the sum of the van der Waals radii. There is no evidence from either a simulated-annealing $2F_o - F_c$ omit map or difference Fourier of an alternative conformation for the tryptophan residue. The residue must, therefore, have multiple alternative conformations. The residue was modelled at 50% occupancy. The side chain of Ile78 has been modelled as two alternative conformations.

Table 2. *Intermolecular distances less than* 3.5 Å *for this work and* 3LZ2 (*the starting model*)

Protein–protein interactions	This work	3LZ2*
O 15 (Leu15)–NH1 (Arg114)	3.27	—
NH1/NH2 (Arg21)–O (Gly67)	2.54	2.61
NH2 (Arg21)–(Asn77)†	3.47 (ND2)	2.62 (OD1)
NZ (Lys33)–OD2 (Asp87)	—	3.27
NZ (Lys33)–OG1 (Thr89)	2.52	2.69
ND2 (Asn37)–OD2 (Asp87)	3.10	—
NE2 (His41)–Ne2 (His41)	3.22	3.42
ND2 (Asn65)–ND2(Asn65)	—	2.70
O (Gly67)–(Asn77)†	2.97 (ND2)	3.24 (OD1)
O (Gly71)–NZ (Lys116)	—	2.62
O (Ser72)–OD2(Asp103)	3.05	—
O (Lys73)–OD2 (Asp103)	3.47	—
O (Asn74)—O (Gly101)	—	3.00
O (Asn74)–N (Gly102)	2.68	3.16
OD1 (Asn77)–O (Ser100)	—	3.49
NZ (Lys96)–NH1 (Arg114)	3.26	—
Ne (Arg112)–NE (Arg112)	2.57	—
NH1 (Arg112)–NE (Arg 112)	2.87	—
NH1 (Arg125)–O (Arg128)	—	2.82

* Howell, Almo, Parsons, Hajdu & Petsko (1992).

† A single atom on the Asn77 side chain interacts with both NH2 of Arg21 and carbonyl O atom of Gly67. The orientation of the Asn77 side chain is different in both structures.

Crystal packing

The overall packing of the protein molecules in the crystal is similar to that found previously for the TEWL (Parsons, 1988; Howell *et al.*, 1992) and guinea fowl egg-white lysozyme (GEWL) (Lescar *et al.*, 1994). The small decrease in the length of the *c* axis in the low-pH crystals is accommodated by the protein rotating 3.86° relative to the starting model. As a result of this rotation and variations in the side-chain conformations, there are differences in the intermolecular contacts found in the hexagonal crystals at high and low pH. In the low-pH structure there are 13 protein–protein intermolecular contacts *versus* 12 in the starting structure (3LZ2), of which six are common. Table 2 lists the protein–protein intermolecular contacts less than 3.5 Å for this structure and the starting model (3LZ2). Examination of the active site reveals that residues involved in binding the saccharide at sites *A* and *B* are all involved in intermolecular contacts (residues 101 and 103). Saccharide-binding site *C* (involving residues 59, 62, 63 and 107) is partially occluded by Lys73 from a symmetry-related molecule. Although, Trp62 interacts unfavorably with itself in the present structure (see section above), it is not sterically hindered and is free to move in a manner comparable to that seen in the inhibitor complexes of MurNAc–GlcNAc–MurNAc (Strynadka & James, 1991), tri-GlcNAc with HEWL (Cheetham *et al.*, 1992) and tri-GlcNAc with partridge lysozyme (PEWL) (Turner & Howell, 1995*a*) where it undergoes a rotation of approximately 17° and an overall shift of 1 Å towards the active-site cleft. The protein in the high-pH crystal structure has been shown to be active in the crystalline state (Howell, Warren, Amatayakul-Chantler, Petsko & Hajdu, 1992) and work is in progress to determine if the crystals grown from potassium thiocyanate are also active and of use in time-resolved Laue diffraction experiments.

Thiocyanate ion

During the course of the refinement it became apparent that a thiocyanate ion could be located in the electron density. Figs. 4(*a*)–4(*c*) are a series of difference Fouriers calculated to determine the correct orientation of the thiocyanate ion; all maps are contoured at 3σ and in all calculations the C atom was omitted. Fig. 4(*d*) is a $2F_o - F_c$ map of the thiocyanate ion. During the initial refinement, the thiocyanate ion was modelled as two independent water molecules. However, the $F_o - F_c$ map examined after refinement cycle 7, (see Fig. 1) gave a large oblate positive difference-density peak around one water molecule (see Fig. 4*a*) and, therefore, the possibility that the water molecules could represent a thiocyanate ion was investigated.

The large positive difference-density peak was assumed to be the S atom. Fig. 4(*b*) shows the positive difference density for the thiocyanate ion when it was modelled as an N and S atom. The small spherical positive difference peak was assumed to be the thiocyanate's C atom. For an internal check, the position of the S and O atoms were reversed, and as expected the difference Fouriers showed large positive density in a position similar to that observed when the thiocyanate ion had been modelled as two water molecules and a large negative difference-density peak around the S atom.

An examination of the intermolecular contacts in the structure of TEWL reveals that the thiocyanate ion

sits in a pocket close to the interface between two protein molecules, along the 6_1 screw axis. The pocket is lined with charged residues, including two arginines (Arg14, Arg5*), one lysine (Lys33) and an aspartic acid residue (Asp87). The S atom of the thiocyanate ion interacts directly with the NH2 (3.28 Å) and NE (3.42 Å) of Arg14, with a third, longer interaction to a water molecule (H2O130, 3.57 Å) (see Fig. 5). The water molecule (H2O130) is hydrogen bonded to the main-chain N atom of Ile88, OD1 of Asp87 and a second water molecule (H2O132). This water molecule (H2O132) makes a hydrogen-bond contact with H2O197* from the second protein molecule and is in turn hydrogen bonded to NE of Arg5*. The three water molecules 130, 132 and 197* form a hydrogen-bonded network between the two protein molecules.

It was expected from accurate low-temperature X-ray and neutron structure determinations and electron-density studies of Na and NH_4 thiocyanate complexes (Bats & Coppens, 1977; Bats, Coppens & Kvick, 1977) that the negative charge will reside on the S atom, with the S atom's lone-pair electrons in a diffuse ring around the S atom, approximately perpendicular to the C—S bond axis. The structure of NH_4SCN (Bats, Coppens & Kvick, 1977) has four hydrogen bonds, two N—H···S and two N—H···N, with bond lengths of 3.378 Å (0.005) and 2.954 Å (0.005), respectively. The hydrogen-bond distances found between the S atom and Arg14 are, therefore, in good agreement with previous small-molecule studies. In addition, the angles between the C—S···NH2 and NE, 92.35 and 98.49°, respectively, are also close to the accepted range (95–100°). A search of the Cambridge Crystallographic Database (Allen, Kennard & Taylor, 1983) for intermolecular contacts between the S atom in a thiocyanate ion and an O atom revealed a wide range of distances from 3.20 to 3.78 Å with an average distance of 3.51 (0.17) Å. The contact that the S atom makes with water H2O130 is well within this range. An examination of the environment around the thiocyanate's N atom reveals no strong electrostatic interactions, since the closest atoms that could form potential hydrogen bonds (NZ of Arg33*, H2O197*) are both approximately 4.0 Å away. However, there are several potential van der Waals interactions with the C atoms of residues Arg5* (3.79 Å) and Phe38* (3.99 Å). The temperature factors associated with the thiocyanate (average value 40.17 $Å^2$) are comparable to the average temperature factor for all solvent molecules in the structure (39.97 $Å^2$). Examination of the regions near to other arginine or lysine residues in the structure failed to locate any additional thiocyanate molecules.

Thiocyanate has previously been found to bind to charged residues in the structure of erabutoxin b at 1.7 Å resolution (Saludjian *et al.*, 1992). In this structure determination, the crystallization from thiocyanate yielded a new orthorhombic crystal form. The N atom of the thio-

* Symmetry-related molecules.

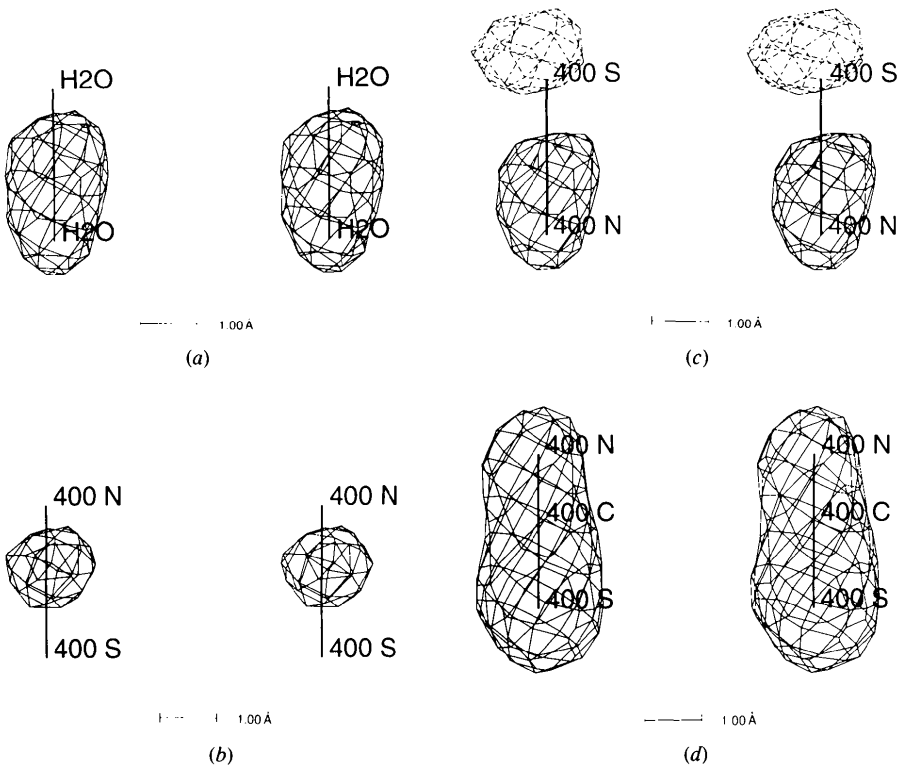

Fig. 4. Stereoviews of difference Fouriers and a $2F_o - F_c$ electron-density map of the thiocyanate ion used to determine the correct orientation of the ion. For (*a*)–(*c*) the maps were calculated with the carbon atom excluded and are contoured at 3σ. Positive difference density is represented by a solid line and negative difference density by dashed line. The difference Fourier maps were calculated with (*a*) the N and S atoms represented as two water molecules, (*b*) with N and S atoms in the correct orientation and (*c*) the position of the N and S atoms reversed. (*d*) A $2F_o - F_c$ electron-density map of the thiocyanate ion.

cyanate ion interacted with the side chains of Arg33A (3.26 Å) and Ser23B (2.68 Å), while the S atom formed longer contacts to the carbonyl O atom of a Cys54B (3.31 Å) and a water molecule (3.43 Å). The thiocyanate was located in a polar cave delimited by three molecules in the vicinity of the intermolecular β-sheet. However, in this study, the packing of the molecules is identical to the previous forms of hexagonal TEWL grown from 3 *M* NaCl (Bott & Sarma, 1976; Howell *et al.*, 1992; Parsons, 1988) and, therefore, allows us to speculate on the efficiency of thiocyanate in crystallizing basic proteins. No water molecules had been built into the starting model (3LZ2) because of the incompleteness of the data set collected *via* the Laue diffraction technique and resolution limit (2.5 Å). An initial examination of the monochromatic structure 2LZ2 (Parsons, 1988; Parsons & Phillips, 1988), failed to locate any water molecules in the vicinity of the thiocyanate ion or the water network. In order to examine the differences in solvent structure that may result from the presence of the thiocyanate ion, we have solved the structure of the pH 8 hexagonal TEWL using monochromatic data at 2.3 Å resolution (Turner & Howell, 1995*b*). The water molecules were located independently by different investigators and were found in similar positions to the SCN^- ion and two of the three water molecules in the network. The only water molecule that does not have a comparable water in the pH 8 structure, is H2O132, a secondary hydration shell water which forms the hydrogen-bond interactions with water molecules 130 and 197*. Although, it cannot be ruled out that this water molecule could not be located due to the difference in resolution range of the data in the pH 8 structure, it does suggests that the ability of thiocyanate ions to crystallize TEWL at low molar concentrations could be related to the difference in size of the thiocyanate ion, its negative charge and an increased effectiveness in stabilizing the polar environment in which it is located in comparison with solvent molecules.

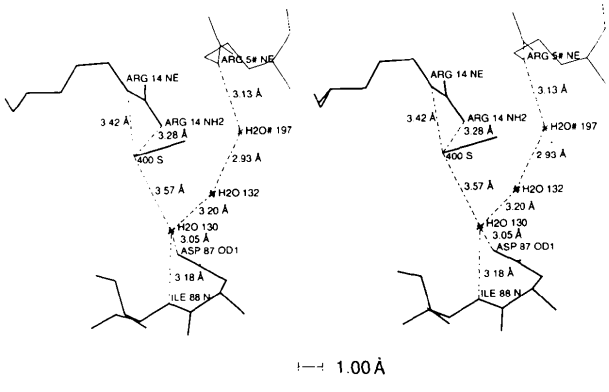

Fig. 5. Stereoview of the electrostatic interactions and hydrogen-bonded water network around the thiocyanate ion. Distances are in Å and water molecules are represented by a cross. The primary protein molecule is draw with a think solid line and the symmetry-related molecule with a thin line.

Comparison with previously determined TEWL structures

The overall differences between this structure and the previously determined structures of TEWL structure are fairly small. Table 3 tabulates the r.m.s. deviations in main-chain atoms for the monoclinic structure, 1LZ3 (Harata, 1993), the monochromatic pH 7 structure, 2LZ2, (Parsons, 1988; Parsons & Phillips, 1988), the Laue pH 8.0 structure 3LZ2 (Howell *et al.*, 1992) and this work. As can been seen from Fig. 6(*a*) the main variations are the result of large differences in three loop regions, residues 46–50, 68–73 and 97–104. Residues 45–50 are involved in a turn between two regions of β-pleated sheet, residues 68–73 are part of a loop between a β-sheet and an α-helix and residues 97–104 connect two α-helices. For loops, 46–50 and 68–73, the r.m.s. differences correlate well with the larger *B* factors found for these regions (Bott & France, 1990) [see Fig. 6(*b*) and Fig. 5 of Harata (1993)] and the differences are most likely attributable to the difficulty of fitting the backbone atoms to poor electron density arising from the conformational flexibility of the loops. Although loop 97–104 is a region that has been shown to be conformationally flexible in triclinic HEWL (Ramanadham, Sieker & Jensen, 1990) and in molecular dynamics simulations of tetragonal HEWL (Post *et al.*, 1986; Ichiye, Olafson, Swaminathan & Karplus, 1986), the differences seen in various TEWL structures do not correlate well with increased thermal motion as represented by the temperature factor (Fig. 6*b*). In TEWL structures, the conformational variability in this loop is due in part to the differences in the intermolecular contacts found in the monoclinic and hexagonal TEWL crystals.

Solvent molecules

A careful comparison has been made of the solvent molecules in this structure and those in the monoclinic structure (1LZ3). Examination of the coordinates from the PDB for 1LZ3 found only 75 water molecules and, therefore, corresponds to the structure refined at 1.5 Å resolution. We restricted our comparison to these structures as it was considered that the placement of water molecules would be more accurate than the lower resolution structures (2LZ2 and 3LZ2). In addition, we have omitted from the comparison, 11 water molecules from the monoclinic structure and 19 waters from the hexagonal structure that could not be present due to differences in the protein–protein interactions that result as a consequence of the different space-group symmetries.

We found 32 water molecules in the primary hydration shell were within 1.0 Å of each other in the two structures. This represents approximately 49% of the solvent molecules in both structures that are independent of packing interactions. The intermolecular contacts made

Table 3. *Average r.m.s. deviations in main-chain atoms* (Cα—N—C) *calculated using PROFIT* (*Smith*, 1994*b*)

	3LZ2*	1LZ3†	2LZ2‡
This work	0.445	0.434	0.451
3LZ2	—	0.500	0.410
1LZ3	—	—	0.509

* Howell, Almo, Parsons, Hajdu & Pesko (1992).
† Harata (1993).
‡ Parsons & Phillips (1988).

by these water molecules were examined further to establish whether the contacts made were similar or not. A contact was considered similar if the distance between the water molecule and the atom was less than 3.5 Å in both structures. We found that contacts made to backbone atoms were more highly conserved that those to side-chain atoms or other water molecules. 89% of the contacts made to backbone atoms were found to be identical, *versus* 64 and 55% for side-chain and water–water contacts, respectively. The present results are in keeping with previous results on T4 lysozyme crystals (Bell *et al.*, 1991), where it was found that solvent hydrogen-bond patterns were not sensitive to the presence of high salt.

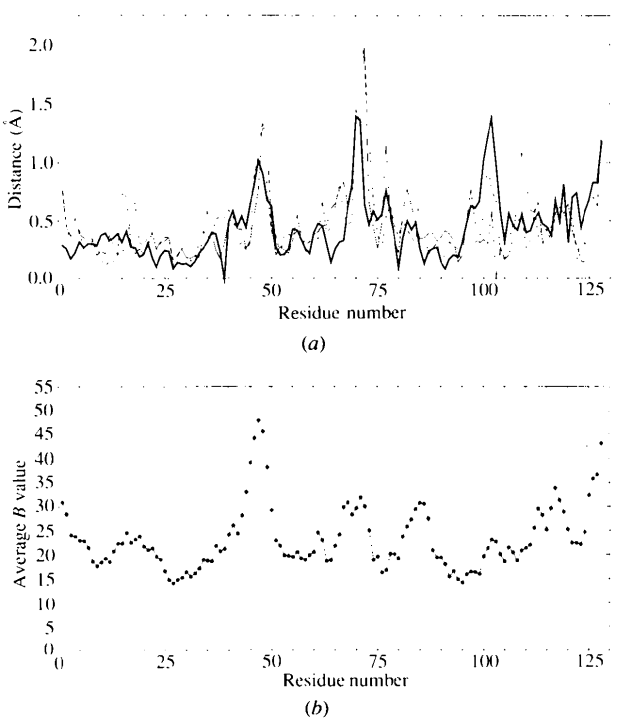

Fig. 6. (*a*) Plot of the average r.m.s. deviations between this work and the previously determined TEWL structures. Thick line, 1LZ3 (monoclinic structure); thin solid line, 2LZ2 (monochromatic pH 7); dashed line, 3LZ2 (Laue structure). (*b*) Plot of average main-chain temperature factors *versus* residue number for the structure presented in this paper.

Concluding remarks

The 1.65 Å structure of hexagonal TEWL crystallized at pH 4.5 from a potassium thiocyanate solution has enable us to examine the efficiency of the SCN^- ion in crystallizing basic proteins at low molar concentrations of thiocyanate. As predicted from earlier work the thiocyanate ion binds preferentially to the side chains of arginine residues. However, in contrast to work on erabutoxin b where the N atom interacts with the positively charged arginine, in this study it is the S atom that directly interacts with the NE and NH2 of the arginine side chain (Arg14). In this study the N atom appears to makes no strong electrostatic interactions with the protein molecule. Examination of the solvent networks in the pH 8 and pH 4.5 structures suggests that the efficiency of thiocyanate in crystallizing basic proteins could be related to the difference in size of the thiocyanate ion *versus* a water molecule, its negative charge and an increased effectiveness in stabilizing the polar environment in which it is located. The structure determination and location of thiocyanate ions in additional crystal structures of basic proteins grown from thiocyanate solutions would enable this hypothesis to be tested.

This work is supported by a grant from the Medical Research Council of Canada. The author would like to thank Graham Bentley for assistance in the data collection and G. David Smith for stimulating discussions and providing numerous computer programs and refinement protocols.

References

Allen, F. H., Kennard, O. & Taylor, R. (1983). *Acc. Chem. Res.* **16**, 146–153.
Artymiuk, P. J. & Blake, C. C. F. (1981). *J. Mol. Biol.* **152**, 737–762.
Artymiuk, P. J., Blake, C. C. F., Rice, D. W. & Wilson, K. S. (1982). *Acta Cryst.* B**38**, 778–783.
Aschaffenburg, R., Blake, C. C. F., Dickie, H. M., Gayen, S. K., Keegan, R. & Sen, A. (1980). *Biochim. Biophys. Acta*, **625**, 64–71.
Bats, J. W. & Coppens, P. (1977). *Acta Cryst.* B**33**, 1542–1548.
Bats, J. W., Coppens, P. & Kvick, A. (1977). *Acta Cryst.* B**33**, 1534–1542.
Bell, J. A., Wilson, K. P., Zhang, X.-J., Faber, H. R., Nicholson, H. & Matthews, B. W. (1991). *Proteins Struct. Funct. Genet.* **10**, 10–21.
Berthou, J., Lifchitz, A., Artymiuk, P. J. & Jolles, P. (1983). *Proc. R. Soc. London Ser. B*, **217**, 471–489.
Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1965). *Nature (London)*, **206**, 757–763.
Blessing, R. H. (1987). *Crystallogr. Rev.* **1**, 3–58.
Blum, M., Metcalf, P., Harrison, S. C. & Wiley, D. C. (1987). *J. Appl. Cryst.* **20**, 235–242.
Bott, R. & France, J. (1990). *Protein Eng.* **3**, 649–657.
Bott, R. & Sarma, R. (1976). *J. Mol. Biol.* **106**, 1037–1046.
Brünger, A. T. (1992). *X-PLOR Manual*, Version 3.0. Yale Univ., New Haven, CT, USA.
Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.

CHEETHAM, J. C., ARTYMIUK, P. J. & PHILLIPS, D. C. (1992). *J. Mol. Biol.* **224**, 613–628.

HARATA, K. (1993). *Acta Cryst.* D**49**, 497–504.

HOWELL, P. L., ALMO, S. C., PARSONS, M. R., HADJU, J. & PETSKO, G. A. (1992). *Acta Cryst.* B**48**, 200–207.

HOWELL, P. L., WARREN, C., AMATAYAKUL-CHANTLER, S., PETSKO, G. A. & HAJDU, J. (1992). *Proteins Struct. Funct. Genet.* **12**, 91–99.

ICHIYE, T., OLAFSON, B. D., SWAMINATHAN, S. & KARPLUS, M. (1986). *Biopolymers*, **25**, 1909–1937.

KABSCH, W. (1988*a*). *J. Appl. Cryst.* **21**, 67–71.

KABSCH, W. (1988*b*). *J. Appl. Cryst.* **21**, 916–924.

KODANDAPANI, R., SURESH, C. G. & VIJAYAN, M. (1990). *J. Biol. Chem.* **256**, 16126–16131.

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

KONNERT, J. H. & HENDRICKSON, W. A. (1980). *Acta Cryst.* A**36**, 344–350.

KUNDROT, C. E. & RICHARDS, F. M. (1987). *J. Mol. Biol.* **193**, 1157–170.

LASKOWSKI, R. A. MACARTHUR, M. W., MOSS, D. S. & THORNTON, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

LESCAR, J., SOUCHON, H. & ALZARI, P. M. (1994). *Protein Sci.* **3**, 788–798.

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

MOULT, J., YONATH, A., TRAUB, W., SMILANSKY, A., PODJARNY, A., RABINOVICH, D. & SAYA, A. (1976). *J. Mol. Biol.* **100**, 179–195.

PANDE, C. S. & MCMENAMY, R. H. (1970). *Arch. Biochem. Biophys.* **136**, 260–267.

PARSONS, M. R. (1988). PhD thesis, Univ. of Leeds, England.

PARSONS, M. R. & PHILLIPS, S. E. V. (1988). Protein Data Bank, entry code 2LZ2.

POST, C. B., BROOKS, B. R., KARPLUS, M., DOBSON, C. M., ARTYMIUK, P. J., CHEETHAM, J. C. & PHILLIPS, D. C. (1986). *J. Mol. Biol.* **190**, 455–479.

RAMAKRISHNAN, C. & RAMACHANDRAN, G. N. (1965). *Biophys. J.* **5**, 909–933.

RAMANADHAM, M., SIEKER, L. C. & JENSEN, L. H. (1989). *The Immune Response to Structurally Defined Proteins: The Lysozyme Model*, edited by S. J. SMITH-GILL & E. SERCARZ, pp 15–24. New York: Adenine Press.

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

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

RIÉS-KAUTT, M. M. & DUCRUIX, A. F. (1989). *J. Biol. Chem.* **264**, 745–748.

RIÉS-KAUTT, M. M. & DUCRUIX, A. F. (1991). *J. Cryst. Growth*, **110**, 20–25.

SALUDJIAN, P., PRANGE, T., NAVAZA, J., MÉNEZ, R., GUILLOTEAU, J. P., RIÉS-KAUTT, M. M. & DUCRUIX, A. (1992). *Acta Cryst.* B**48**, 520–531.

SMITH, G. D. (1995). *J. Appl. Cryst.* In preparation.

SMITH, G. D. (1994). *PROFIT. Program for Least Squares Analysis of Protein Structures.* Hauptman-Woodward Medical Research Institute, Buffalo, New York, USA.

STRYNADKA, N. C. J. & JAMES, M. N. G. (1991). *J. Mol. Biol.* **220**, 401–424.

TURNER, M. A. & HOWELL, P. L. (1995*a*). *Protein Sci.* In the press.

TURNER, M. A. & HOWELL, P. L. (1995*b*). In preparation.

WISHART, D., WILLARD, L. & SYKES, B. (1994). *Vadar: A Program for Analysis of Protein Structures.* Univ. of Alberta, Canada.