Acta Cryst. (1994). D**50**, 709–730

Accuracy and Precision in Protein Crystal Structure Analysis: Two Independent Refinements of the Structure of Poplar Plastocyanin at 173 K

BY BARRY A. FIELDS

Department of Inorganic Chemistry, University of Sydney, Sydney 2006, *Australia*

HANS H. BARTSCH, HANS D. BARTUNIK AND FRANK CORDES*

Max-Planck Research Unit for Structural Molecular Biology, DESY, Notkestrasse 85, 22603 *Hamburg, Germany*

AND J. MITCHELL GUSS AND HANS C. FREEMAN

Department of Inorganic Chemistry, University of Sydney, Sydney 2006, *Australia*

(*Received* 20 *September* 1993; *accepted* 17 *March* 1994)

Abstract

The structure of the copper protein plastocyanin from poplar leaves (*Populus nigra* var. *italica*) at 173 K has been subjected to two independent refinements, using a single set of synchrotron X-ray data at 1.6 Å resolution. Energy-restrained refinement using the program *EREF* resulted in lower root-mean-square deviations from ideal geometry (*e.g.* 0.011 Å for bond lengths) but a higher residual R (0.153) than restrained least-squares refinement using the program *PROLSQ* (0.014 Å, 0.132). Electron-density difference maps in both refinements provided evidence for disorder at some side chains and solvent atoms, and the *PROLSQ* refinement made allowance for this disorder. The number of solvent sites identified at the $4\sigma(\rho)$ level was 171 in the *EREF* refinement and 189 in the *PROLSQ* refinement; 159 of the solvent sites are common to both refinements within 1 Å. The root-mean-square differences between the atomic positions produced by the two refinements are 0.08 Å for C^{α} atoms, 0.08 Å for backbone atoms and 0.12 Å for all non-H atoms (excluding six obvious outliers) of the protein molecule. The two sets of Cu–ligand bond lengths differ by up to 0.07 Å, and the ligand–Cu–ligand angles by up to 7°. At 173 K the volume of the unit cell is 4.2% smaller than at 295 K. Greater order in the solvent region is indicated by the location of 79 more solvent sites, the identification of extensive networks of hydrogen-bonded rings of solvent molecules, and a general decrease in the thermal parameters. Within the unit cell, the protein molecules are significantly translated and rotated from their positions at ambient temperature. An important structural change at low temperature is a 180° flip of the peptide group at Ser48-Gly49. Nearly all other significant differences between the structures of the protein at 173 and 295 K occur at exposed side chains. If the backbone atoms in the 173 and 295 K structures are superposed, excluding atoms involved in the peptide flip, the root-mean-square difference between the positions of 393 atoms is 0.25 Å. Two internal water molecules, not included in previous descriptions of poplar plastocyanin, have been located. The plastocyanin Cu-site geometry at 173 K is not significantly different from that at 295 K. If plastocyanin undergoes a change in Cu-site geometry at low temperature, as has been suggested on the basis of resonance Raman spectroscopic evidence, then the change is not detected within the limits of precision of the present results.

Introduction

The 'blue' copper protein plastocyanin has been well characterized by crystallographic structure analysis (Colman *et al.*, 1978; Guss & Freeman, 1983; Guss, Harrowell, Murata, Norris & Freeman, 1986; Collyer, Guss, Sugimura, Yoshizaki & Freeman, 1990; Guss, Bartunik & Freeman, 1992), NMR spectroscopy† (Moore *et al.*, 1991), electronic spectroscopy (Penfield, Gewirth & Solomon, 1985;

* Present address: Institut für Kristallographie, Freie Universität Berlin, Takustrasse 6, D-1000 Berlin 53, Germany.

† Abbreviations used: PoPc, poplar plastocyanin in the oxidized (Cu^{II}) state; PEG, polyethyleneglycol; EXAFS, extended X-ray absorption fine structure; EM, energy minimization; LT, 173 K; MD, molecular dynamics; NMR, nuclear magnetic resonance; r.m.s., root mean square; rR, resonance Raman; RT, 295 K; 1PCY, Protein Data Bank reference to atomic positional coordinate set for PoPc at RT and 1.6 Å resolution; 1PLC, Protein Data Bank reference to atomic positional coordinate set for PoPc at 1.33 Å resolution.

Gewirth, Cohen, Schugar & Solomon, 1987), vibrational spectoscopy (Woodruff, Norton, Swanson & Fry, 1984; Blair *et al.*, 1985; Han *et al.*, 1991), X-ray absorption spectroscopy (Scott, Hahn, Doniach, Freeman & Hodgson, 1982; Penner-Hahn, Murata, Hodgson & Freeman, 1989; Murphy, Hasnain, Strange, Harvey & Ingledew, 1990), elecrochemistry (Büchi, Bond, Codd, Huq & Freeman, 1992) and kinetic measurements (Jackman, McGinnis, Powls, Salmon & Sykes, 1988; Kyritsis *et al.*, 1991). A quantum-mechanical treatment of the 'blue' copper site is available (Gewirth & Solomon, 1988). The biological role of plastocyanin is electron transport between photosystems II and I in plants and some photosynthetic algae. Much of the relevant literature has been reviewed by Sykes (1990). In understanding the properties of the protein it has frequently been helpful to know that the Cu atom has a distorted tetrahedral coordination geometry and that it is bound to two N^{δ}(His) atoms at about 2.0 Å, a S(Cys) atom at about 2.1 Å and a S(Met) atom at about 2.8 Å. The Cu atom is slightly displaced from the plane of the two N^{δ}(His) atoms and the S(Cys) atom, and the long Cu—S(Met) bond is approximately perpendicular to that plane. The dimensions of the Cu site were originally revealed by the refinement of the structure of plastocyanin from poplar leaves (*Populus nigra* var. *italica*) at ambient temperature using data to 1.6 Å resolution (Guss & Freeman, 1983). This structure is now known at 1.33 Å resolution (Guss, Bartunik & Freeman, 1992).

Our interest in studying the structure of poplar plastocyanin (PoPc) at low temperature when it was already well known at ambient temperature was aroused by two observations which suggested that the geometry of the copper site changes when the protein is cooled. First, an early set of extended X-ray absorption fine structure (EXAFS) data recorded from oriented plastocyanin crystals at 175–205 K yielded a new Fourier transform feature which was provisionally assigned to a contribution from a sulfur ligand at 2.4 Å from the Cu atom (Penner-Hahn, Murata, Hodgson & Freeman, 1989). A plausible interpretation of this assignment was that the long Cu—S(Met92) bond (2.9 Å), which does not contribute to the EXAFS at ambient temperature (Scott *et al.*, 1982), is replaced by a bond about 2.4 Å long at low temperature (Penner-Hahn & Hodgson, 1986). Second, it was reported that changes in the resonance Raman (rR) spectrum of plastocyanin close to the freezing point of water can be attributed to a change in the Cu—S^{γ}(Cys84)—C^{β}—C^{α} dihedral angle, 'the formation of a weak but significant Cu—S(Met92) bond', and the resulting inhibition of 'large-amplitude motions of the Cu atom perpendicular to the N_2S(Cys) plane, which are permitted at room temperature due to the virtual absence of a Cu—S(Met) bond' (Woodruff, Norton, Swanson & Fry, 1984; Blair *et al.*, 1985). The putative short Cu—S(Met92) bond deduced from the EXAFS was subsequently shown to be an artefact caused by noise in the data (Penner-Hahn, Murata, Hodgson & Freeman, 1989), but the hypotheses derived from the rR measurements remained to be tested.

Experiments on the stabilization of PoPc crystals at low temperature were accordingly undertaken (see *Experimental*, below), and ultimately made it possible to record a set of diffraction data at 173 K on a two-dimensional detector at the DESY synchrotron. The fact that the research was a collaboration between two groups of investigators who were separated geographically by a very large distance provided an opportunity to conduct an experiment on the nature of the refinement process. It was agreed that the two groups of investigators would not communicate with each other until each was convinced that its refinement calculations were complete. Further, since the two groups were experienced in the use of two different refinement protocols, it was agreed that the Sydney group would use the restrained least-squares program *PROLSQ* (Hendrickson & Konnert, 1980) while the Hamburg group would use the energy-constrained least-squares program *EREF* (Jack & Levitt, 1978). Lists of atomic positional and thermal parameters were exchanged only when the two independent refinements had converged. The refinement protocols are summarized in Fig. 1.

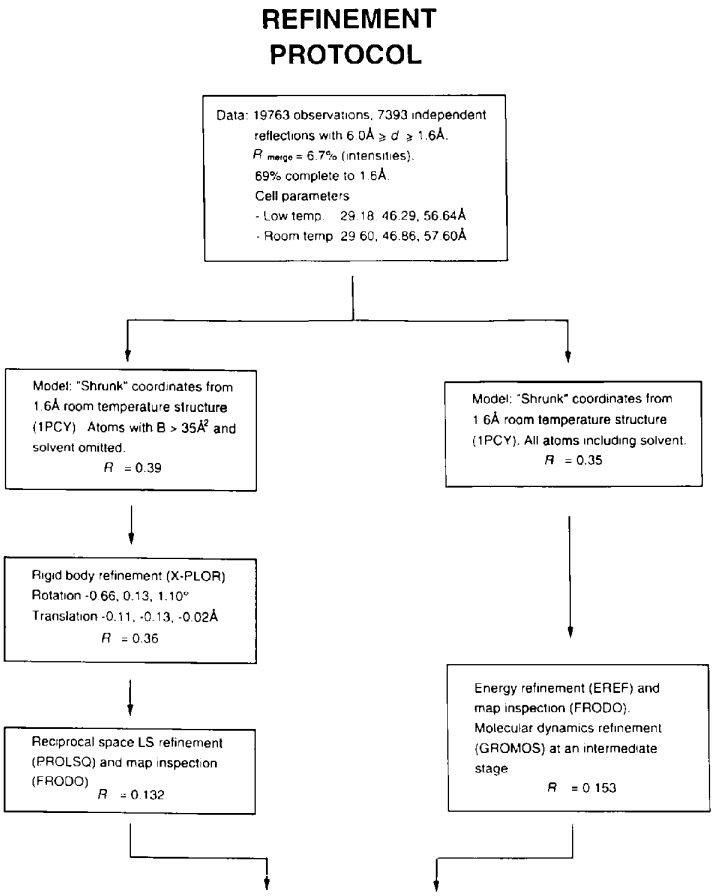
Experimental

Plastocyanin crystals at 173 K

Crystals of PoPc were grown as described by Chapman *et al.* (1977) and kept in a storage buffer containing 75% saturated ammonium sulfate and 0.1 *M* phosphate at pH 6.0. Experiments to exchange the ammonium sulfate/sodium phosphate buffer solution against a suitable cryo-solvent were successful only at temperatures above 248 K. Below this temperature, test photographs taken with synchrotron X-radiation invariably revealed powder-diffraction rings from residual ice or precipitated salts. Extrapolation from the best results suggested that in order to reach 173 K it would be necessary to use 60% methanol, but the crystals did not survive in such a solvent mixture even at 248 K.

The problems was finally overcome by a combination of solvent-exchange and snap-freezing. Immediately before the low-temperature experiment, crystals were transferred into saturated polyethylene glycol (PEG 6000) at 279 K. A single crystal was mounted with as little PEG as possible on a thin glass spatula attached to the base of a cold chamber (Bartunik &

Schubert, 1982). The crystal and the spatula were shock-frozen in liquid nitrogen. The cold chamber was then closed and mounted on an Arndt–Wonacott rotation camera modified for use with a FAST two-dimensional detector. The temperature was kept below 153 K. Subsequently the temperature was adjusted to 173 K, the variation during the data collection being less than 1 K.



	Sydney	Hamburg
Number of protein atoms	1462	739
Number of water atoms	189	171
Disordered sidechains	Ser 56, Glu 79, Ser 85	Not available in EREF
Average B for protein	5.3Å²	8.2Å²
Average B for water	17.8Å²	21.0Å²
Cu-ligand bond lengths:		
Cu-N(His 37)	1.98Å	2.05Å
-S(Cys 84)	2.14Å	2.21Å
-N(His 87)	1.95Å	1.96Å
-S(Met 92)	2.78Å	2.78Å
Rms Δ from 1.33Å structure (backbone)	0.26Å	0.26Å
Rms Δ from ideal PROLSQ bond lengths	0.014Å	0.020Å
Rms Δ from ideal EREF bond lengths		0.011Å
Rms Δ between structures (backbone)	0.08Å	

Fig. 1. Summary of the *PROLSQ* (Sydney) and *EREF* (Hamburg) refinements of the plastocyanin structure using a single set of diffraction data recorded at 173 K. The total occupancy of the protein non-H-atom sites in both models is 738. The 1462 atoms in the Sydney model comprise 747 non-H atoms, 714 H atoms, and the Cu atom. The nine additional non-H-atom sites in the Sydney model represent the alternate conformers of three disordered side chains.

Data collection

Diffraction data were collected from two crystals using the double focusing beamline X31 at HASYLAB/EMBL. The X-ray wavelength was 1.009 Å with a 0.1% band width. The data were recorded with a FAST area detector (Enraf–Nonius), the normal to the detector being inclined by 11° above the incident beam. The crystal-to-detector distance had been calibrated by multiple measurements on well characterized orthorhombic trypsin crystals. The program *MADSYN* was used for data collection as a series of 0.1° rotation exposures, orientation matrix refinement and data processing. *MADSYN* is based on the program *MADNES* (Messerschmidt & Pflugrath, 1987); it has been adapted for use with the FAST detector combined with an Arndt–Wonacott camera. Crystal number 1 (with outer dimensions 570 × 290 × 340 μm) was mounted with c parallel to the spindle axis, and crystal number 2 (460 × 350 × 250 μm) with b parallel to the spindle axis.

Under the conditions of the experiments (beam divergence and wavelength spread), the unit-cell dimensions were reproducible to within 0.1 Å. The averaged values derived from multiple measurements on crystals number 1 and number 2 were a = 29.18 (3), b = 46.29 (7), c = 56.64 (13) Å, V = 76.506 (370) Å^3; the space group was the same as at room temperature, $P2_12_12_1$. The crystals of PoPc at room temperature have unit-cell dimensions (fitted to θ values on a four-circle diffractometer) a = 29.60 (1), b = 46.86 (3), c = 57.60 (3) Å, V = 79.894 (120) Å^3. Thus the unit-cell volume was significantly (4.2%) smaller at 173 K, the reductions in the unit-cell edges a, b and c being 1.4, 1.2 and 1.7%, respectively.

Cooling caused a broadening of the crystal mosaicity by about 0.3°. Diffraction data were collected to 1.6 Å resolution. The entire data set consisted of 19 763 intensity measurements corresponding to 7933 unique reflections. The total exposure time was 4.8 h. Merging of the data using the program *PROTEIN* (Steigemann, 1974) yielded 7393 unique reflections with $R_{\rm merge}$ = 0.067 in I (after rejection of 6.8% of all unique reflections).* A reflection was rejected if one value of the scaled intensity differed from the mean by ≥30%. The data set included 69.3% of all possible reflections to 1.60 Å. The completeness in the last shell (1.66–1.60 Å) was 48.6%.

Structural refinements

As indicated above, the low-temperature data were used in two independent refinements of the structure. The calculations were made by means of the programs *PROLSQ* (Hendrickson & Konnert, 1980) and *EREF* (Jack & Levitt, 1978), respectively. The *EREF* refinement included an intermediate stage of molecular dynamics refinement using *GROMOS* (van Gunsteren & Berendsen, 1987). Starting phases for the *PROLSQ* and *EREF* refinements were derived from the coordinates of the room-temperature structure of PoPc refined at 1.6 Å resolution (Guss & Freeman, 1983; Protein Data Bank entry 1PCY).

PROLSQ refinement

The initial values of the calculated structure amplitudes F_c and phases φ_c included only the contributions of the protein atoms in the 1PCY coordinate set. Atoms with $B > 35$ Å^2 and all solvent atoms were omitted. In view of the significant and anisotropic shrinkage of the unit cell, the model was first subjected to 50 cycles of rigid-body refinement using the program *X-PLOR* (Brünger, Karplus & Petsko, 1989). The refinement, based on the data in the range $5.5 \geq d \geq 2.5$ Å, caused the molecule to undergo rotations of −0.66, 0.13, 1.10° and translations of −0.11, −0.13, −0.02 Å with respect to the x, y, z axes. The residual R decreased from 0.39 to 0.36.

Refinement was continued with the restrained reciprocal-space least-squares program *PROLSQ* (Hendrickson & Konnert, 1980). Models were fitted to electron-density maps by means of the program *FRODO* (Jones, 1978). A total of 50 rounds of refinement were performed, a typical 'round' consisting of several cycles of *PROLSQ* least-squares calculation followed by the calculation and inspection of $2F_o - F_c$ and/or $F_o - F_c$ electron-density maps. The range of the data included in the calculations was gradually increased from $5.5 \geq d \geq 2.5$ Å to $6.0 \geq d \geq 1.6$ Å. During the first 20 *PROLSQ* cycles the temperature factors B of all atoms were fixed at 10 Å^2. In subsequent cycles the temperature factors of non-H atoms were treated as variables. The restraints applied to the values of B are discussed later (see *Comparison between two independent refinements*, *Temperature parameters*). Protein H atoms were included in the model at calculated positions towards the end of the refinement. They were given temperature factors equal to those of the atoms to which they were bonded.

Solvent atoms were added to the model when the residual R had dropped to 0.30 for the data in the range $6.0 \geq d \geq 2.0$ Å. Peaks at or above the $4\sigma(\rho)$ level in $F_o - F_c$ maps were modeled as an O(water) atom if they had at least one plausible hydrogen-bonding partner within the range 2.5–3.3 Å. Peaks identified as solvent but making a contact shorter than 2.5 Å with another solvent atom were modeled as a disordered solvent atom and refined with occu-

* Residuals: $R = \sum\overline{||F_o| - |F_c||}/\sum|F_o|$; $R_{\rm merge} = \sum(I_i - \bar{I}_i)/\sum I_i$.

pancy 0.5. Peaks which made any other forbidden close contact were excluded. The temperature factors of solvent atoms were restrained to $B \geq 2$ Å^2.* Two neighbouring peaks buried beneath the surface of the molecule persisted throughout the refinement, having heights $>4\sigma(\rho)$ in the final $F_o - F_c$ map. Although these peaks made several favourable hydrogen-bonding contacts, they were excluded from consideration as O(water) by the rigorous application of a lower limit (3.1 Å) for contacts with C atoms. In the *EREF* refinement (see below), the lower limit was applied less rigorously and analogous peaks were modelled as solvent.

Continuous density in $(F_o - F_c)$ maps contoured at the $3\sigma(\rho)$ level revealed 12 of the 14 side chains that did not lie in density in the RT 1.6 Å structure: Glu18, Lys26, Ser45, Lys54, Glu59, Asp61, Asn64, Lys66, Ser75, Glu79, Lys95 and Asn99. The remaining two side chains, Lys30 and Glu60, were not as well defined. The electron-density features associated with them in $(F_o - F_c)$ and 'omit' maps (the latter calculated with the omission of seven side chains at a time) were significant but not continuous.

Only three side chains (Ser56, Glu79 and Ser85) were ultimately modelled as disordered, despite the fact that various types of electron-density map had suggested the presence of disorder at two more (Glu60 and Lys95). In the final model, Ser56 was refined with its O^γ atom distributed over two sites having occupancies 0.3 and 0.7, respectively. The electron density identified as O^γ(Ser56) had originally suggested a model with this atom distributed equally over three sites, but two of the sites coalesced during refinement. The side chains of Glu79 and Ser85 were each modelled as two conformers with occupancy 0.5. The difference between the two conformers of Glu79 was a $\sim 90°$ rotation about C^γ—C^δ. Attempts to refine Glu60 and Lys95 with two side-chain conformers were unsuccessful, and these residues were included in the final model as single conformers (temperature factors $\langle B \rangle = 20$, 14 Å^2).

During the concluding stages of the refinement, the persistence of residual electron density in the immediate vicinity of the Ser48-Gly49 peptide group created doubts about the correctness of the model in this region. The best way to adjust the model was not obvious, since the electron-density in this region was relatively weak. The existence of the problem was noted at the time when the atomic coordinates from the two independent refinements were exchanged. In the Hamburg analysis, the Ser48-Gly49 peptide group had been adjusted manually by about 180° with respect to the Sydney conformation at an early stage. Though the relevant maps are no longer accessible, it seems likely that the inclusion of solvent atoms in the Hamburg model (see below) had

Table 1. *Refinement statistics for PROLSQ refinement: weighting parameters, and deviations from target values*

The structural results discussed in this paper were derived from refinement calculations using the target values σ and yielding the deviations Δ listed under 'original restraints'. The σ's and deviations Δ listed under 'RT restraints' refer to a subsequent repetition of the final refinement cycles using the five times larger target values σ that had been used in the RT refinement at 1.33 Å resolution (Guss, Bartunik & Freeman, 1992), see text.

	σ	R.m.s. Δ	Worst Δ	$\sum w\Delta^2$
Bonding distances (Å)				
Bond length	0.030	0.014	0.062	
Angle distance	0.040	0.029	0.130	
Interplanar distance	0.050	0.030	0.121	0.90×10^3
X—H bond distance	0.100	0.017	0.118	
X—H angle distance	0.080	0.015	0.067	
Planar groups (Å)	0.020	0.010	0.034	0.19×10^3
Chiral volumes (Å^3)	0.150	0.124	0.458	0.79×10^2
Non-bonded contacts (Å)				
Single torsion	0.500	0.128	0.376	
Multiple torsion	0.500	0.195	0.773	0.89×10^2
Hydrogen bond ($X \cdots Y$)	0.500	0.192	0.492	
Hydrogen bond (X—H$\cdots Y$)	0.500	0.133	0.476	
Torsion angles (°)				
Peptide plane ω (0, 180°)	5.0	2.3	7.1	
Staggered (±60, 180°)	15.0	10.4	49.2	0.79×10^2
Orthonormal (±90°)	22.0	9.7	23.8	
Thermal factors (Å^2), original restraints				
Main chain (1 2)	1.0	1.6	5.7	
Main chain (1 3)	1.5	1.9	8.0	
Side chain (1–2)	1.0	2.6	10.2	0.37×10^3
Side chain (1–3)	1.5	3.6	14.7	
X—H bond	2.0	1.1	4.1	
X—H angle	2.0	1.9	10.5	
Thermal factors (Å^2), RT restraints				
Main chain (1–2)	5.0	2.3	9.7	
Main chain (1–3)	7.5	2.6	12.2	
Side chain (1–2)	5.0	4.4	25.6	0.37×10^2
Side chain (1–3)	7.5	5.3	27.8	
X—H bond	10.0	2.2	9.7	
X—H angle	10.0	3.3	20.1	
Restraints against excessive shift				
Positional parameters (Å)	0.1	—	—	
Thermal parameters (Å^2)	3.0	—	—	
Diffraction data				
Structure-factor modulus $[\sigma(F) = 13.0, w = \sigma^{-2}(F)]$	13.0	—	—	0.7376×10^4

* A referee has suggested that restraining B to ≥ 2 Å^2 may have masked the presence of heavier atoms/molecules. This hypothesis was not tested systematically, but the available evidence suggests that the assignment of the solvent sites is correct. Only nine solvent sites have the minimum value $B = 2$ Å^2, rounded to the nearest integer (Table S1; see footnote concerning Supplementary Material). Eight of these nine sites were refined with occupancy 0.5, so that the electron density would probably not be identifiable as a heavier atom/molecule even if B were lower. The ninth site, *W*104, was refined with occupancy 1.0; it requires no comment since B is *above* the minimum (2.4 Å^2).

resulted in better phases by the time when the electron density at Ser48-Gly49 was inspected. The correctness of the 'flipped' conformation was immediately confirmed by means of an 'omit' map for the Sydney model (*i.e.* an $F_o - F_c$ map where the contributions of Ser48-Gly49 to the values F_c and φ_c were omitted). The model was adjusted and the refinement was continued for nine more rounds.

The final model consisted of the Cu atom, 747 protein non-H atoms with total occupancy 738, 714 protein H atoms, and 189 O(water) atoms with total occupancy 174. The residual R was 0.132 for the 7393 independent reflections with $6.0 \geq d \geq 1.6$ Å. Other refinement statistics at listed in Table 1.*

EREF refinement

The starting model, comprising all protein and solvent atoms listed for 1PCY, had a residual $R = 0.35$ for reflections in the range $6.0 \geq d \geq 1.6$ Å. The model was improved on the basis of $2F_o - F_c$ and $F_o - F_c$ difference maps using the program *FRODO* (Jones, 1978). The main chain was adjusted in the vicinity of residue Ser48 where the angle ψ was changed by almost 180°. A number of side chains that were not or only partly defined in 1PCY were located in electron density.

The structure was subjected to combined crystallographic and energy refinement using the program suite *EREF* (Jack & Levitt, 1978). The Cu atom was left unrestrained. For all other atoms, restraints were used; parameters defining the potential energy were taken from Levitt (1974). Unrestrained isotropic temperature factors of the individual atoms were refined with the programs *DERIV* and *SHIFTS* (H. Deisenhofer, personal communication). Structure factors were calculated with a program (S. J. Remington, personal communication) based on the FFT method by Ten Eyck (1977). *EREF* minimizes the function,

$$f(E,\Delta) = E + w\Delta,$$

where E is the potential energy, $\Delta = \sum(|F_o| - |F_c|)^2$, and w is a weight factor. During most of the refinement calculations, a small value of $w = 0.0001$ was chosen in order to stress geometrical restraints; these restraints were relaxed ($w = 0.0002$) during the final stages of refinement.

The residual R decreased from 0.35 to 0.25 after one cycle of refinement. This rapid improvement resulted essentially from a reduction in the phase errors which were initially present due to the substantial difference between the cell constants of the low-temperature structure as compared to the room-temperature structure. In total about 40 alternating cycles of *EREF* and *B*-factor refinement were calculated, interspersed with re-adjustment of parts of the structure on the graphics system and location of a large number of solvent molecules, which were not present in 1PCY, in density.

Intramolecular hydrogen bonds were defined on the basis of their bond lengths (< 3.35 Å) and bond angles using the program *HYBAC* (Levitt, 1975). Hydrogen bonds between main-chain atoms were required to have, in addition, an electrostatic interaction energy more negative than -0.5 kcal mol^{-1} as calculated with the program *DSSP* (Kabsch & Sander, 1983). Peaks were identified as solvent if they were at or above the $4\sigma(\rho)$ level in $F_o - F_c$ maps and were within 2.5–3.5 Å of a potential hydrogen-bonding partner. As the refinement progressed, $3.5\sigma(\rho)$ peaks were also accepted if they met the geometrical criteria.

The final model consisted of 738 protein non-H atoms, one Cu atom and 171 water molecules. The residual R for this model was 0.153, including all reflections. Refinement statistics were subsequently generated with the program *PROCHECK* (see below).

Combined energy minimization and MD refinement

At an intermediate step of the *EREF* calculations, crystallographic refinement using molecular dynamics (MD) was carried out. The aim of the calculations was to investigate whether MD refinement could locate a number of residues in density without the need for manual adjustment using computer graphics. In particular, the side chains of six lysine residues (Lys26, 30, 54, 66, 77 and 95) were only partly defined at this stage.

The MD calculations were made with the program *GROMOS* (van Gunsteren & Berendsen, 1987) in which the energy of the molecule is defined by a semi-empirical function. The protocol was similar to that adopted by Fujinaga, Gros & van Gunsteren (1989). The integration steps were 0.002 ps over a time range of 2 ps. Restraints on the bonded interatomic distances suppressed molecular vibrations corresponding to higher frequencies. Electrostatic pair exchange interactions were calculated for distances < 8 Å between polar and charged groups with charges -1 to $+1$. The energy function contained two auxiliary terms. One of these provided a coupling to the experimentally determined X-ray

* Atomic coordinates, atomic thermal parameters and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory [References: 1PNC (*PROLSQ* refinement), 1PND (*EREF* refinement)]. These data, together with Tables S1 and S2 (water molecule environments, and additional refinement statistics as mentioned in text) have also been deposited with the IUCr (Reference: AM0011). Copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

structure factors (Brünger, 1988),

$$E_f = (1/\sigma^2)(S|F_c| - |F_o|)^2,$$

where $1/\sigma^2$ is a weighting factor and S is a scaling factor. The second auxiliary term described a harmonic potential applied to all atoms in order to keep them near their position in the initial structure. Only the side-chain atoms of the six lysine residues were exempted from this restraint.

The starting model for the MD calculations consisted of 738 protein non-H atoms, the Cu atom and 146 water molecules; the residual R was 0.235. In trial calculations, the value of R increased unless the weighting parameter σ was in the range 50–100. A starting value $\sigma = 100$ and stepwise reduction to $\sigma = 40$ resulted in an improvement of R to 0.210, and located three of the lysine residues (Lys54, 66, 77) in density without operator intervention. Some atomic positions changed by as much as 3 Å. The other three lysines required larger changes in atomic position which made manual adjustment on the graphics necessary.

In parallel with the MD calculations, the initial structure was subjected to 50 steps of 'steepest descent' energy minimization (EM) calculations. The EM refinement resulted in an r.m.s. shift of 0.22 Å in the atomic positions summed over all atoms but, in contrast with the MD refinement, moved none of the lysine side-chain atoms into density.

Further refinement statistics

Some comparative statistics for the *PROLSQ* and *EREF* refinements are shown in Table 2. The statistics were generated by means of the program *PROCHECK* which analyses a given structure in terms of 'typical' values for geometrical parameters derived from a survey of well refined structures (Laskowski, Macarthur & Thornton, 1993). The values listed in Table 2 reveal no substantial differences between the *PROLSQ* and *EREF* refinements, with the exception of the standard deviation in ω (the parameter characterizing the planarity of a peptide group). The standard deviations (*PROLSQ*, 2.2°; *EREF*, 7.0°; 'typical' value, 6.6°) and maximum deviations (*PROLSQ*, 7.1°; *EREF*, 24.5°; not shown in Table 2) in ω show that, while the *EREF* results may be described as 'typical', the restraints on peptide group planarity are more effective in *PROLSQ* than in *EREF*. A more detailed discussion is given under the heading *Effects of restraints on atomic positional coordinates*, below.

Additional refinement trials

After the conclusion of the *PROLSQ* and *EREF* refinements, a number of calculations were made in order to test the robustness of the results and to assess the extent to which the results had been influenced by the choice of refinement parameters. We refer to these calculations at relevant points in the text.

Table 2. *Comparison between PROLSQ and EREF refinements*

The following comparisons were made by means of the program *PROCHECK* (Laskowski, MacArthur & Thornton, 1993). The 'typical' values are taken from that program.

	PROLSQ value	*EREF* value	'Typical' value
Main-chain parameters			
S.d. of peptide bond planarity (ω) (°)	2.2	7.0	6.0
S.d. of C^{α} tetrahedral distortion (ζ) (Å)	2.3	2.3	3.1
Bad non-bonded contacts per 100 residues	1.0	1.0	1.4
S.d. of hydrogen-bond energies (kJ mol^{-1})	2.9	3.3	2.9
Residues in most favoured regions of Ramachandran plot (%)	91.5	91.0	87.5
Side-chain torsion angles			
S.d. of χ_1 *gauche* minus (°)	9.1	8.7	14.5
S.d. of χ_1 *trans* (°)	6.0	6.8	16.1
S.d. of χ_1 *gauche* plus (°)	10.5	9.7	14.5
S.d. of χ_1 pooled (°)	9.3	9.0	15.1
S.d. of χ_2 *trans* (°)	7.7	7.0	18.2

Comparison between two independent refinements

Several previous discussions of the plastocyanin structure have addressed the problem of estimating the accuracy and precision of the protein coordinates in general and the metal-site geometry in particular (Guss & Freeman, 1983; Guss, Harrowell, Murata, Norris & Freeman, 1986; Guss, Bartunik & Freeman, 1992). The present work provides an additional insight into this problem by comparing models derived from a single set of LT diffraction data *via* two independent refinements. The most likely causes of differences between the refined models are differences in refinement methodology, operator intervention, and/or refinement strategy.

(i) *Differences in methodology.* *EREF* and *PROLSQ* introduce restraints in different ways, give different weights to the restraints in relation to the crystallographic observations, use different values to represent ideal geometry, and have different capacities to deal with disordered side chains and partially occupied solvent sites. Such differences are obviously capable of affecting the refinements. A similar conclusion was reached, for example, by Rydel, Tulinsky, Bode & Huber (1991).

(ii) *Differences in operator intervention.* The refinement steps most likely to be affected by operator intervention are those where human pattern recognition plays an important role, *i.e.* the assignment of side-chain conformations and solvent sites. For example, the decision whether an electron-

density feature represents a single atom or a superposition of partial atoms can be influenced by operational accidents such as the order of the directions from which the electron density is viewed, or even by subjective factors such as the operator's past experience. We expect differences between the *EREF* and *PROLSQ* models to have arisen in this way, even though the effects may be difficult to quantitate. The effects of operator intervention on the outcomes of a refinement have been demonstrated previously (Guss, Bartunik & Freeman, 1992).

(iii) *Differences in refinement strategy*. The *EREF* and *PROLSQ* refinement strategies differed in the choice of starting models (*EREF*, all atoms listed in 1PCY, including solvent; *PROLSQ*, only protein atoms with $B \leq 35$ Å^2), the treatment of H atoms (*EREF*, H atoms added at calculated positions after refinement converged; *PROLSQ*, H atoms included in refinement model), and the treatmetnt of atomic temperature parameters (*EREF*, unrestrained; *PROLSQ*, restrained). The possibility that the choice of starting model in the *EREF* refinement resulted in better phases, and thus facilitated the recognition of a 'peptide flip' at an early stage of the refinement, has already been mentioned. Evidence that the differences between the starting models were not propagated to the final refinements is provided by the fate of the 44 solvent molecules in 1PCY. Despite the fact that these solvent molecules were excluded from the *PROLSQ* starting model but included in the *EREF* starting model, 29 were located in *both* refinements, and 10 were located in *neither* refinement. (Two more turned up only in the *PROLSQ* refinement, and the remaining three only in the *EREF* refinement.) Thus the possibility of bias in the refinements from the starting models can be rejected. Similarly, it is unlikely that the final results were affected significantly by differences in the treatment of H atoms: in principle the inclusion of H atoms makes the restraints on non-bonded contacts more realistic, but it has been shown elsewhere that – at least in a *PROLSQ* refinement – the effect on the positions of the non-H atoms is small (Guss, Bartunik & Freeman, 1992). The effects of treating temperature factors differently require further discussion (see *Temperature parameters*, below).

Atomic positional coordinates

The protein models derived from the *PROLSQ* (Sydney) and *EREF* (Hamburg) refinements are in close agreement (Fig. 2). The r.m.s. differences between the atomic coordinates in the two models are 0.08 Å for C$^\alpha$ atoms, 0.08 Å for backbone atoms, and 0.24 Å for side-chain atoms. The maximum difference between the positions of a backbone atoms is 0.28 Å, and 96% of the side-chain atom positions agree within 0.4 Å. The significance of the largest differences is discussed later. The slightly greater detail in the Sydney model (allowance for three disordered side chains and 27 disordered solvent sites) has no significant effect on the cited statistics.

The solvent structures in the two models are also closely similar. Cooling the crystals to 173 K has resulted in increased detail in the solvent region. The Hamburg and Sydney LT models have 171 and 189 solvent sites, respectively; 159 solvent sites are common to both models, *i.e.* occur within 1 Å of each other (Fig. 3). Table S1, comprising details of the environments of the water molecules in both models, has been deposited as Supplementary Material.*

Effect of restraints on atomic positional coordinates

The residual R from the *PROLSQ* (Sydney) refinement, 0.132, is significantly lower than that from the *EREF* (Hamburg) refinement, 0.153 (Fig. 1). The order is reversed if the r.m.s. deviations from ideal bond lengths as defined in the libraries of the respective programs are compared (*PROLSQ*, 0.014 Å; *EREF*, 0.011 Å). This illustrates the balance between two refinement objectives – to minimize the differences between the observed and calculated structure amplitudes on the one hand, and the deviations from a set of ideal geometrical criteria on the other hand. In the present work the relative weights that control this balance were chosen subjectively on the basis of past practice. The weights used in the *PROLSQ* refinement were in fact identical with those used in the original RT refinement at 1.6 Å resolution (Guss & Freeman, 1983).

* See footnote concerning Supplementary Material.

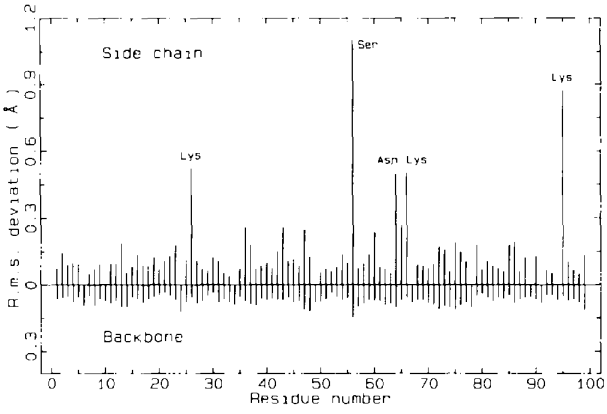


Fig. 2. Differences between the atomic positions refined by *PROLSQ* and *EREF*, respectively. The differences shown for each residue are r.m.s. values averaged over (top) the side-chain atoms, and (bottom) the backbone atoms (N, C$^\alpha$, C, O).

In order to assess the extent to which the atomic coordinates were influenced by the geometrical restraints in the two refinement programs, both the *PROLSQ* and *EREF* models were subjected to ten further cycles of *PROLSQ* refinement in which only the atomic positional parameters were varied and the restraints on them were switched off. The coordinates used for the unrestrained refinements were taken directly from the outputs of the final restrained refinement cycles (*PROLSQ*, all non-H and H atoms; *EREF*, non-H atoms). The choice of *PROLSQ* as the program for the unrestrained refinement of both models is unlikely to have affected the outcome since the temperature factors were kept constant and only minimization of $\sum w(\Delta F)^2$ was involved. The r.m.s. shifts resulting from this type of calculation can be used as conservative indicators of the errors in the atomic positions (Chambers & Stroud, 1979).

The additional cycles of unrestrained refinement of the *PROLSQ* model resulted in r.m.s. changes of 0.10, 0.11 and 0.13 Å in the positions of the protein backbone atoms, non-H atoms and side-chain atoms, respectively. The corresponding r.m.s. shifts for the *EREF* model were 0.12, 0.15 and 0.18 Å, respectively. The uncertainty calculated in this way for the backbone atom positions in the *PROLSQ* model, 0.10 Å, is consistent with an estimate derived from the r.m.s. difference between the *PROLSQ* and *EREF* models (see below). Similar values (r.m.s. shift 0.14 Å, accompanied by a decrease in R from 0.125 to 0.103) were reported for a protease-B/inhibitor complex at 1.8 Å resolution (Read, Fujinaga, Sielecki & James, 1983).

During the unrestrained refinements, the residual R for the *EREF* model decreased from 0.153 to 0.138, and that for the *PROLSQ* model from 0.132 to 0.111. The slightly smaller decrease in R for the *EREF* model when the restraints were relaxed is consistent with other evidence (see below) that most of the variables were restrained more effectively in *PROLSQ* than in *EREF*. Such comparisons are complicated by the fact that the models produced by the original refinement, and therefore used for the

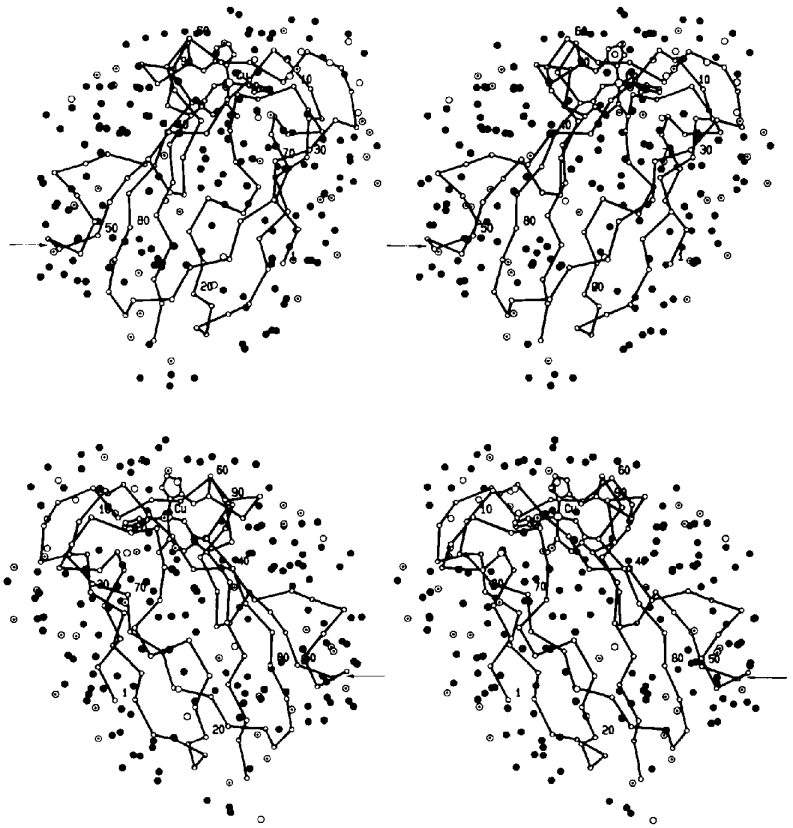


Fig. 3. The plastocyanin molecule (C^{α} atoms, Cu atom and Cu-binding side chains) and solvent atoms in the crystal at 173 K. Small circles denote solvent atoms; filled circles, solvent atoms found in both refinements; circles with dot, additional solvent atoms found in the Sydney refinement; plain circles, additional solvent atoms found in the Hamburg refinement. The arrow indicates the location of Ser48-Gly49 (see Figs. 7 and 8). The top and bottom parts of the figure are related by a rotation of 180° about the vertical axis.

unrestrained refinements, were not entirely equivalent (H atoms being included only in the *PROLSQ* model).

In a second test of the restraints, the final *EREF* model (non-H atoms) was subjected to 15 cycles of *restrained PROLSQ* refinement. Predictably, the imposition of *PROLSQ* restraints caused the *EREF* model to change towards the *PROLSQ* ideal geometry. The r.m.s. deviation between the two sets of 396 backbone-atom positions decreased from 0.08 to 0.05 Å. There were improvements in the r.m.s. and maximum deviations from all *PROLSQ* target values except those for multiple-torsion contacts, hydrogen-bonded contacts, staggered torsion angles and orthonormal torsion angles. Particularly large improvements occurred in the deviations from the target values for planar groups and for torsion angles involving peptide planes, indicating that the relevant restraints were substantially more effective in *PROLSQ* than in *EREF*. Table S2 summarizing the refinement statistics has been deposited as Supplementary Material.*

During the restrained *PROLSQ* refinement, the residual R for the *EREF* model (as calculated by *PROLSQ*) decreased to 0.151. The difference between this value and the residual $R = 0.132$ for the *PROLSQ* model could be attributed almost entirely to the effect of H atoms. The residual calculated for the *PROLSQ* model without H atoms was 0.149. Deleting the H atoms from the 1.33 Å RT model of Guss, Bartunik & Freeman (1992) resulted in similar changes (from $R = 0.141$ to 0.154 for the data to 1.6 Å, and from $R = 0.151$ to 0.163 for the data to 1.33 Å resolution).

Temperature parameters

The variations of B as a function of residue number are similar in the Sydney and Hamburg refinements (Fig. 4). The actual B's are, however, consistently higher in the Hamburg (*EREF*) refinement.

Are the *EREF* B's systematically too high, or are the *PROLSQ* B's too low? A Wilson plot for the LT data provides qualitative evidence that the former is the case (Fig. 5). The isotropic temperature factor deduced from the Wilson plot, $B = 6.6$ Å^2, is in better agreement with the average value $\langle B \rangle = 8.0$ Å^2 from the *PROLSQ* refinement than with the average value $\langle B \rangle = 11.0$ Å^2 from the *EREF* refinement.

The systematic difference between the *PROLSQ* and *EREF* results arises from the way in which the atomic B's were calculated and refined. In *PROLSQ* the structure factors were calculated using a trigonometric expression in reciprocal space, and the B's were restrained during refinement to prevent large differences between the values for adjacent atoms. In *EREF* the structure factors were calculated with a fast Fourier transform algorithm using an atomic model in real space, and during refinement the B's were unrestrained. When the *EREF* model was subjected to further refinement by *PROLSQ*, including the application of *PROLSQ* restraints on the temperature parameters, the average values $\langle B \rangle$ for various types of atom decreased until they were close to the values obtained in the original *PROLSQ* refinement ($\langle B \rangle$ for *EREF* model after *PROLSQ* refinement, followed by $\langle B \rangle$ for *PROLSQ* model: protein atoms, 5.7, 5.3 Å^2; solvent atoms, 19.1, 17.8 Å^2; all atoms, 8.6, 8.0 Å^2).

Given the better agreement with the Wilson-plot value, and the greater chance of errors arising from sampling in the FFT calculation of structure factors affecting the B's, we conclude that the *PROLSQ* values are more likely to be correct. We are, however, left with the possibility that the B's produced by *PROLSQ* were artificially low as a result of too tight restraints (*i.e.* too low target σ's). It seemed prudent to test this hypothesis because the target σ's

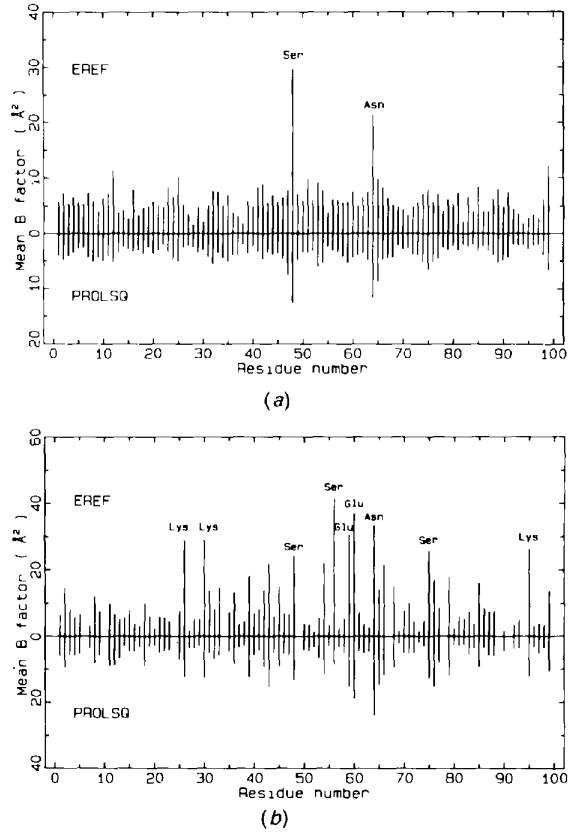


Fig. 4. Temperature factors averaged over (*a*) backbone atoms N, C$^\alpha$, C, O, and (*b*) side-chain atoms of each residue. Above line, *EREF* refinement; below line, *PROLSQ* refinement.

* See footnote concerning Supplementary Material.

for temperature factors in the LT refinement (listed under 'original restraints' in Table 1) were only one-fifth of the values used in the RT refinement of PoPc (Guss, Bartunik & Freeman, 1992). The last six *PROLSQ* refinement cycles were, therefore, repeated with the original target σ's multiplied by 5 and 50, respectively. The former value made the target σ's equal to those used in the RT refinement. The $\langle B \rangle$ for protein atoms, 5.3 Å^2, increased only to 5.5 and 5.6 Å^2 when the target σ's were multiplied by 5 and 50, respectively. In neither case was the change significant. Thus the possibility that the *PROLSQ* B's were systematically too low as a result of excessive restraints can be eliminated.

The statistics for the LT refinement with target σ's increased by a factor of 5 (*i.e.* equal to those in the RT refinement) are included in Table 1 under 'RT restraints'. It is not surprising that relaxing the restraints led to increases in the largest differences ('Worst Δ') and the r.m.s. Δ's. On the other hand, the further tenfold increase in the target σ's left the refinement statistics (not shown) almost unchanged, indicating that the LT temperature parameters were effectively unrestrained when they were refined with the RT target σ's.

Accuracy and precision of the LT model

A Luzzati plot (Luzzati, 1952) calculated with the F_{calc} values from the Sydney refinement leads to an estimated maximum average error of about 0.13 Å in the atomic positions (Fig. 6). This global estimate is likely to be too high for the positions of the protein backbone atoms, and too low for the positions of the side-chain atoms and solvent atoms: the atomic positions in the backbone are least likely – and those in exposed side chains and solvent sites most likely – to have been affected by software options for handling disorder and by operator intervention. We have already seen that agreement between the two refined models is much better along the protein backbone than at the side chains and in the solvent region.

Table 3. *Estimates of the uncertainty of atomic positions in PoPc at LT*

	Backbone atoms (Å)	Overall (Å)	Side-chain atoms (Å)
Luzzati plot	--	0.13	—
R.m.s. difference between *PROLSQ* and *EREF* models	0.08	0.12	0.24
R.m.s. change when restraints released			
PROLSQ model	0.10	0.11	0.13
EREF model	0.12	0.15	0.18
Estimated uncertainty in PoPc at RT and 1.6 Å resolution (Guss & Freeman, 1983)	0.10	–	-

For the backbone atoms a plausible estimate of the uncertainty is provided by the r.m.s. difference between their positions in the Sydney and Hamburg models, 0.08 Å. The fact that the difference is small suggests that the two refinements have determined these atomic positions equally accurately, within the limits of precision permitted by the data. The r.m.s. change during an unrestrained refinement calculation (see above) leads to a slightly higher value, 0.10 Å, consistent with the fact that the uncertainty in the atomic positions is not caused solely by differences between the two refinements. Accordingly we treat 0.10 Å as a *quasi*-e.s.d. It is equivalent to the estimated uncertainty of the backbone atom positions in the 1.6 Å RT structure analysis of PoPc, where it was possible to calculate the r.m.s. difference between two models that had been refined using independent

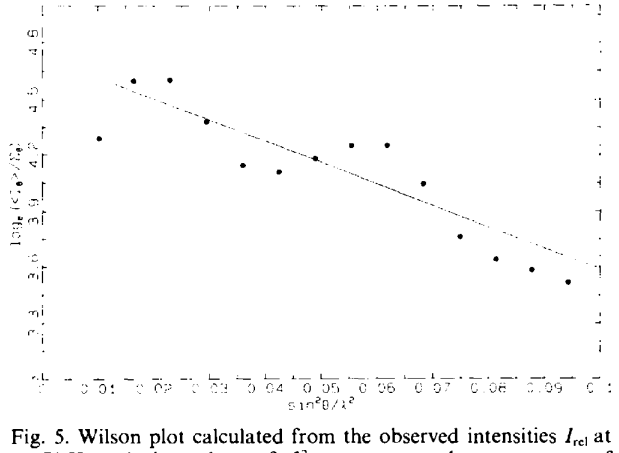


Fig. 5. Wilson plot calculated from the observed intensities I_{rel} at 173 K and the values of $f^2_{Nitrogen}$, averaged over ranges of $\sin^2\theta/\lambda^2$. The average temperature factor $B = 6.6$ Å^2 derived from the plot is in good agreement with the average $B = 8.0$ Å^2 calculated from the *PROLSQ* refinement.

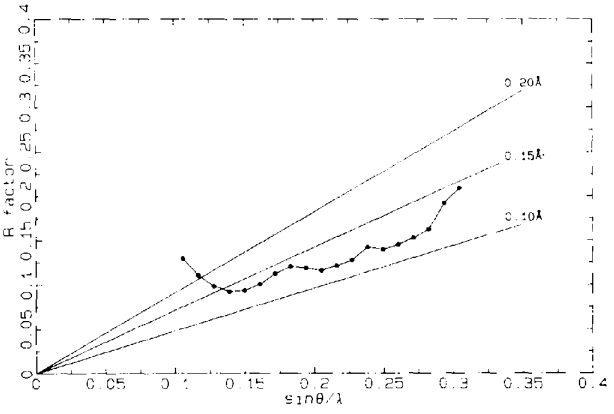


Fig. 6. Luzzati plot calculated from the values of F^2_{obs} at 173 K and the values of F^2_{calc} after the *PROLSQ* refinement.

data sets (Guss & Freeman, 1983). A summary is given in Table 3.

The indications that the present 1.6 Å LT models have approximately the same precision as the 1.6 Å RT model can be rationalized as follows: On the one hand, the mosaic spread of the LT crystals was larger and the diffraction data were less complete; on the other hand, the data were good enough to reveal additional details of the molecular structure and the solvent structure.

While it is important to establish the general precision of the molecular structure, the chemically most interesting parameters are the dimensions of the Cu site. Here the analogy between the precisions of the 1.6 Å LT and RT models is particularly helpful. The r.m.s. differences between the Cu–ligand bond lengths and bond angles in the two LT refinements are 0.05 Å and 3°, respectively. While statistics based on samples of four bond lengths and six angles must be treated with caution, the use of r.m.s. differences as estimates of the precision of the LT Cu-site geometry is supported by the fact that they are identical with the estimated standard deviations in the 1.6 Å RT model.

Atomic positions in exposed side chains and solvent region

In contrast with the positions of atoms in the polypeptide backbone and internal side chains, the positions of surface side-chain and solvent atoms have a high probability of being influenced by software options for handling disorder and by differences in operator intervention during the later stages of a refinement. In the present work there are seven differences ≥ 0.4 Å between the Sydney and Hamburg models, and all of them occur at side chains: Lys26 at C^{δ} and C^{ϵ}, Pro36 at C^{γ}, Glu43 at $O^{\epsilon 1}$, Ser56 at C^{β} and O^{γ}, Asn64 at C^{γ}, $N^{\delta 2}$ and $O^{\delta 1}$, Lys66 at N^{ζ}, and Lys95 at C^{δ}, C^{ϵ} and N^{ζ}. The differences between the final models probably originated from different interpretations of similar electon-density features by two knowledgable operators. A conservative approach is to regard these side chains as relatively poorly defined (Table 4).

Table 4. *Comparisons between side-chain conformations in PoPc at RT* (295 K) *and LT* (173 K)

Description of side chain at LT*	Description of side chain at RT†		
	Single conformer	Two conformers (disordered)	Notes
Single conformer	Glu18		(*a*)
Same as at RT	Glu59		(*a*)
	Asn99		(*a*)
Single conformer		Ser45	(*a*)
Same as one conformation at RT		Lys54	(*a*)
		Asp61	(*a*),(*h*)
		Asn64	(*a*),(*b*),(*e*)
		Ser75	(*a*)
Disordered, two conformers same as two conformers at RT		Ser56	(*b*),(*f*)
Disordered, two conformers one conformer same as at RT	Ser85		
Single conformer	Lys26		(*a*),(*b*)
not same as RT	Pro36		(*b*)
	Glu43		(*b*),(*c*)
	Ser48		(*g*)
	Ser58		(*c*),(*h*)
	Glu60		(*c*),(*d*),(*h*)
	Glu68		
	Lys95		(*a*),(*b*),(*d*)
Single conformer		Lys30	
not same as conformers at RT		Lys66	(*a*),(*b*)
		Glu71	(*i*)
Disordered, two conformers **neither** conformer same as at RT	Glu79		(*a*)

Notes: (*a*) Side chain in electron density at LT, but not at RT and 1.6 Å resolution. (*b*) Side chain with one or more atomic positions differing by ≥ 0.4 Å in the *PROLSQ* and *EREF* LT models. (*c*) Side chain difficult to locate in electron density at LT. (*d*) Evidence for disorder in LT electron density, but refined as single conformer. (*e*) Residue with largest change in backbone position on cooling (see Figs. 9 and 10). (*f*) Only residue that is disordered in both LT and RT models. (*g*) Residue involved in peptide flip (see Figs. 7, 8 and 10). (*h*) See Fig. 15. (*i*) See Fig. 16.

* The LT atomic positions are taken from the *PROLSQ* refinement.

† The RT atomic positions are taken from the refinement at 1.3 Å resolution (Guss, Bartunik & Freeman, 1992).

Description of low-temperature structure and comparison with ambient-temperature structure

Other proteins for which high-resolution structures at low temperatures have been determined include bovine pancreatic trypsinogen at 103 and 173 K (Walter *et al.*, 1982), sperm-whale metmyoglobin at 80, 115, 165 and 185 K (Parak *et al.*, 1987), flavodoxin at 123 K (Watt, Tulinsky, Swenson & Watenpaugh, 1991), rat trypsin-S195C at 123 K (Wilke, Higaki, Craik & Fletterick, 1991), ribonuclease at seven temperatures between 98 and 260 K (Tilton, Dewan & Petsko, 1992), and bovine pancreatic trypsin at 220 K (H. H. Bartsch & H. D. Bartunik, unpublished work). As in the present work, the temperature factors in these low-temperature structures are generally lower than at ambient temperature, and enhanced contrast in the electron-density maps enables additional atoms (mainly solvent) which are disordered at ambient temperature to be located.

The PoPc unit cell at 173 K

The changes in unit-cell dimensions when PoPc crystals are cooled from room temperture to 173 K correspond to a shrinkage of 4.2% in the unit-cell volume [1.7% (1.0 Å) along *c*, 1.4% (0.4 Å) along *a*,

and 1.3% (0.6 Å) along *b*]. There is clear evidence from the *X-PLOR* rigid-body refinement preceding the Sydney *PROLSQ* calculations that the protein molecules adjust to the decreased volume by rotating about 1° with respect to the unit-cell axes.

The reduction of the unit-cell volume of PoPc by 4.2% lies in the range of shrinkages which have been reported for bovine pancreatic trypsinogen at 173 K (0.4%), sperm-whale met-myoglobin at 185 K (2.8%), flavodoxin at 123 K (5.3%), and rat trypsin-S195C at 123 K (7.5%). More of the unit-cell shrinkage can be attributed to the solvent than to the protein. If the radius of gyration of the PoPc molecule, R_{gyr}, is calculated from the coordinates of the 396 polypeptide backbone atoms, its value decreases from 12.31 Å at RT to 12.22 Å at LT. The change in R_{gyr} corresponds to a decrease of ~2% in the molecular volume. Since the decrease in the unit-cell volume is 4.2%, and if the fraction of the unit cell occupied by solvent is 36% as in the RT structure (Chapman *et al.*, 1977), then it follows that the decrease in the volume occupied by solvent is ~8%. For flavodoxin [60% solvent (*v*/*v*)], the protein volume decreases by ~3% and the solvent volume by ~7% (Watt, Tulinsky, Swenson & Watenpaugh, 1991). For rat trypsin-S195C, an even larger decrease in R_{gyr} from 15.8 to 15.5 Å (5%) is observed (Wilke, Higaki, Craik & Fletterick, 1991).

Whether the volume occupied by solvent is smaller at LT because there are fewer solvent molecules, or whether the same number of solvent molecules are packed more efficiently, remains to be determined. The greater detail seen in the solvent region suggests that the solvent is more ordered at low temperature, but does not prove that the average volume per solvent molecule is reduced.

The Cu-*site geometry at* 173 K

The dimensions of the Cu site at low and ambient temperatures are compared in Table 5. Neither the differences between the Sydney and Hamburg dimensions at LT, nor the differences between the LT models at 1.6 Å resolution and the RT model at 1.33 Å resolution, are significant in terms of the estimated standard deviations [0.05 Å and 3° at 173 K (present work), 0.04 Å and 2.5° at 295 K (Guss, Bartunik & Freeman, 1992)]. The difference between the two Cu—N(His) bond lengths, which is significant in the RT model at 1.33 Å resolution, is not observed at 1.6 Å resolution in the LT or RT models.

Apparent discrepancies between the values of some dihedral angles in the three refinements may be ignored. They are associated with the absence of restraints on the deviations of the Cu atom from the imidazole ring planes of His37 and His87 (B. A. Fields, J. M. Guss & H. C. Freeman, unpublished work).

Table 5. *Dimensions of the* Cu *site in poplar plastocyanin at low and ambient temperatures*

SYD, structure at 173 K refined by *PROLSQ* (Sydney); HAM, structure at 173 K refined by *EREF* (Hamburg); 1.33 Å, structure at 295 K refined by *PROLSQ* (Guss, Bartunik & Freeman, 1992).

	SYD	HAM	1.33 Å
Cu–ligand bond lengths (Å)			
Cu—N(His37)	1.98	2.05	1.91
Cu—N(His87)	1.95	1.95	2.06
Cu—S(Cys84)	2.14	2.20	2.07
Cu—S(Met92)	2.78	2.78	2.82
Deviations of Cu from planes of best fit (Å)			
Cu···N(His37),S(Cys84),N(His87)	0.37	0.47	0.36
Cu···N(His37),S(Cys84),S(Met92)	0.69	0.70	0.65
Cu···S(Cys84),N(His87),S(Met92)	0.74	0.74	0.70
Hydrogen bond to S(Cys84) (Å)			
(Asn38)N—H···S(Cys84)	3.55	3.47	3.51
Bond angles at Cu atom (°)			
N(His37)—Cu—N(His87)	103	102	97
N(His37)—Cu—S(Cys84)	132	128	132
N(His37)—Cu—S(Met92)	88	95	88
N(His87)—Cu—S(Cys84)	114	115	121
N(His87)—Cu—S(Met92)	107	107	101
S(Cys84)—Cu—S(Met92)	108	107	110
Bond angles at ligand atoms (°)			
C^{γ}—N(His37)—Cu	129	132	129
C^{ϵ}—N(His37)—Cu	122	112	123
C^{γ}—N(His37)—C^{ϵ}	108	110	106
C^{γ}—N(His87)—Cu	129	129	126
C^{ϵ}—N(His87)—Cu	126	123	124
C^{γ}—N(His87)—C^{ϵ}	105	107	105
C^{β}—S(Cys84)—Cu	106	107	111
C^{β}—S(Cys84)···N(Asn38)	112	111	108
Cu—S(Cys84)···N(Asn38)	107	108	109
C^{γ}—S(Met92)—Cu	131	130	127
C^{ϵ}—S(Met92)—Cu	98	97	96
C^{γ}—S(Met92)—C^{ϵ}	96	98	98
Dihedral angles at Cu site (°)			
Cu—N(His37)—C^{γ}—$C^{\delta 2}$	171	150	165
Cu—N(His37)—C^{ϵ}—N^{ϵ}	−172	−157	−166
N(His37)—Cu—S(Cys84)—C^{β}	−107	−114	−110
N(His37)—Cu—S(Met92)—C^{γ}	157	157	161
N(His37)—Cu—S(Met92)—C^{ϵ}	51	51	56
N(His37)—Cu—N(His87)—C^{γ}	148	157	167
Cu—N(His87)—C^{γ}—C^{β}	175	174	159
Cu—N(His87)—C^{ϵ}—N^{ϵ}	−176	−174	−159
N(His87)—Cu—S(Cys84)—C^{β}	115	116	113
N(His87)—Cu—S(Met92)—C^{γ}	−100	−99	−102
N(His87)—Cu—S(Met92)—C^{ϵ}	154	155	153
N(His87)—Cu—N(His37)—C^{γ}	139	153	147
Cu—S(Cys84)—C^{β}—C^{α}	−171	−169	−168
S(Cys84)—C^{β}—C^{α}—N	172	169	169
Cu—S(Met92)—C^{γ}—C^{β}	71	76	68
C^{ϵ}—S(Met92)—C^{γ}—C^{β}	177	−178	173
S(Met92)—C^{γ}—C^{β}—C^{α}	170	172	166
S(Met92)—Cu—S(Cys84)—C^{β}	−4	−2	−3

A 'peptide flip' at residues 48–49

The most notable difference between the molecular structures at LT and RT occurs at residues 48–49

(Table 6, Figs. 7 and 8). The LT structure corresponds to a type I reverse turn (Venkatachalam, 1968) with $\varphi_{48} = -47°$, $\psi_{48} = -30°$, $\varphi_{49} = -94°$, $\psi_{49} = -5°$. At RT this reverse turn is a type II turn (Venkatachalam, 1968) with $\varphi_{48} = -55°$, $\psi_{48} = 133°$, $\varphi_{49} = 90°$, $\psi_{49} = -14°$. Both types of turn have a O···H—N bond between residues 47 and 50. A plausible explanation for the interconversion is that the two conformations of the molecule lie in shallow free-energy minima with approximately equal free energies, and that small changes in molecular pack-

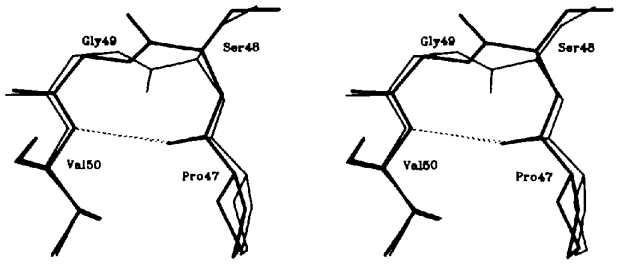


Fig. 7. 'Peptide flip' at Ser48-Gly49. The conformation at residues 48–49 is shown at 173 K (heavy lines) and at 295 K (light lines).

Table 6. *Protein and solvent contacts* (< 3.35 Å) *in the vicinity of Ser*48-*Gly*49 *at* 173 K (*LT*) *and at* 295 K (*RT*)

Residue	Contacts	Residue	Distance (Å) LT	RT
Ser48	N···O	Ile46	3.0	—†
	N···O_w	133	3.0	2.8
	O···O_w	102	2.8	—
	O···O_w	186	—	2.6
	O···O_w	203	—	3.2
	O···O_w	244^{vii}	2.8	*
	C···N	Val50	—	3.3
	O^{γ}···OT	Asn99^{iv}	2.8	3.1
	O^{γ}···O_w	157	3.1	3.3
	C^{β}···O_w	300	3.3	*
	C^{β}···O_w	203	—	3.3
Gly49	C^{α}···O	Glu59^{viii}	3.3	3.2
	C^{α}···O_w	299	3.0	*
	N···O_w	299	2.4	*
	N···O_w	128	—	3.0
	N···C	Pro47	3.3	—
	N···O	Pro47	3.1	—
	O···O^{γ}	Ser85^{vii}	2.5	2.7
	O···C^{β}	Ser85^{vii}	3.2	3.3
	O···N	Pro86^{vii}	3.3	—
	O···O^{γ}	Ser#85^{vii}	2.5	—
	O···C^{β}	Ser#85^{vii}	3.1	—
	O···O_w	129	—	3.2

Symmetry codes: (iv) $\frac{1}{2} + x, \frac{1}{2} - y, 1 - z$; (vii) $1 - x, -\frac{1}{2} + y, \frac{1}{2} - z$.

* Solvent atoms not located in the structure at RT.

† Dashes (—) denote distances ≥ 3.35 Å.

Symbol denotes a conformer of a disordered residue.

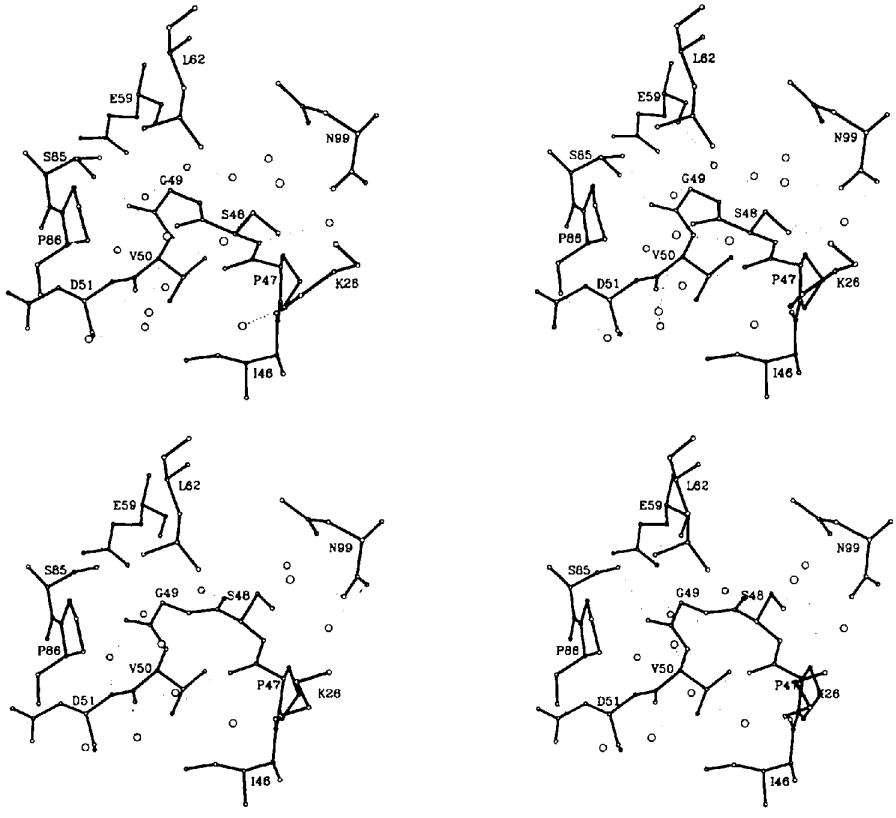


Fig. 8. Environment of Ser48-Gly49 in the crystal structure of poplar plastocyanin (top) at 173 K, and (bottom) at 295 K.

ing (such as those associated with changes in the ordering of the solvent and in the volume of the unit cell) suffice to freeze out one conformation at LT and another at RT. Peptide flips in response to changes in molecular packing have been reported previously [for example, at Gly78 in two orthorhombic crystal forms of trypsin (Bartunik, Summers & Bartsch, 1989)].

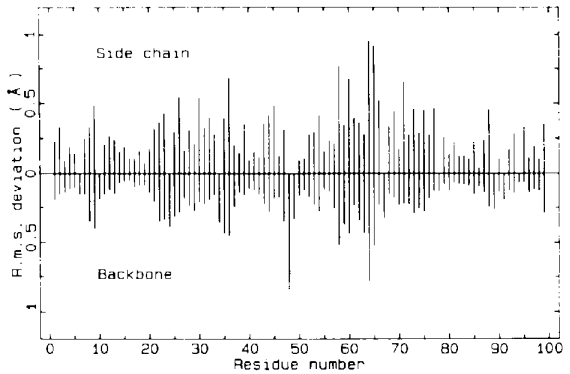


Fig. 9. Differences between atomic positions at 295 K (1.33 Å structure) and 173 K (*PROLSQ* refinement). Differences are r.m.s. deviations averaged over (top) side-chain atoms, and (bottom) backbone atoms, respectively, of each residue.

Two independent observations on plastocyanin support this hypothesis. (i) In the structure analysis of cucumber plastocyanin (Garrett, 1987), one of the two molecules in the asymmetric unit has the same conformation as poplar plastocyanin at RT. The second molecule of the asymmetric unit has a disordered Ser48-Gly49 peptide group, the two conformers differing by a 180° peptide flip. (ii) In a NMR structure analysis of French bean plastocyanin (Moore *et al.*, 1991), about 60% of a family of structures generated from the NMR data in solution have the same conformation at residues 48–49 as poplar plastocyanin in the crystal structure at RT. The remaining 40% of the structures in solution have the same 'peptide flipped' conformation as found in the present work.

The molecular structure of PoPc at 173 K: *polypeptide backbone*

The peptide flip is not the only significant difference between the LT and RT structures. Superposition of 393 backbone N, C^{α}, C and O atoms (*PROLSQ* LT model, 1.33 Å RT model, three atoms omitted at residue 48) yields a r.m.s difference of

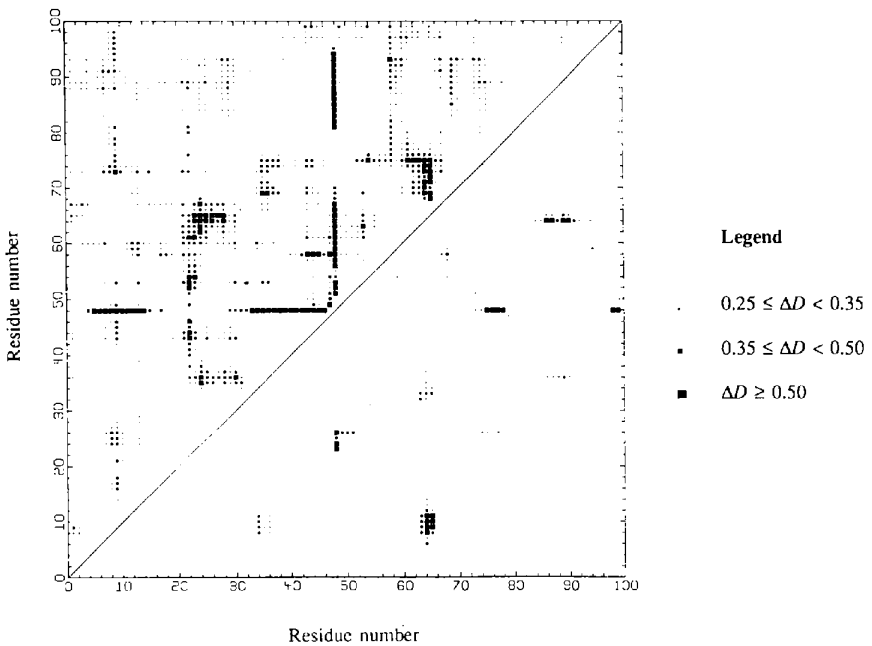


Fig. 10. A matrix plot (Berghuis & Brayer, 1992) of the differences between $C^{\alpha}\cdots C^{\alpha}$ distances at 173 and 295 K. The value at the point i,j represents $\Delta D_{P-Q} = (C_{ij})_P - (C_{ij})_Q$ where $(C_{ij})_P$ and $(C_{ij})_Q$ are the $C^{\alpha}\cdots C^{\alpha}$ distances between the ith and jth residues at temperatures P and Q, respectively. The values plotted in the upper and lower triangles are $\Delta D_{295-173\ K}$ and $\Delta D_{173-295\ K}$, respectively. Only positive values of ΔD are shown. The plot is based on the *PROLSQ* (Sydney) positions of the C^{α} atoms at 173 K. The most prominent individual effects occur at Ser48 and Asn64. The fact that there is a greater density of positive ΔD's in the upper triangle indicates that the $C^{\alpha}\cdots C^{\alpha}$ distances at 173 K are systematically smaller than at 295 K, consistent with other evidence that cooling causes a slight shrinkage of the molecule (see text).

0.25 Å. In terms of the *quasi*-e.s.d. discussed earlier, 0.10 Å, individual differences larger than the r.m.s. value 0.25 Å are highly significant. Figs. 9 and 10 are complementary illustrations of the changes in the polypeptide-atom positions.

Fig. 9 is a conventional line diagram in which the changes averaged over the backbone atoms $NC^{\alpha}CO$ of each residue are plotted against residue number. The majority of the atoms undergoing substantial movements when the crystals are cooled occur at loops near the north and south ends of the molecule [8–9, 21–26, 34–36, 47–49 (non-β strand 5), 58–60, 62–67, 69–70 (β-strand 6), 73–76 (β-strand 6), 88–90]. A more detailed analysis not elaborated here shows that the movements at residues 8–9, 21–26, 58–60, 66–67, 69–70 and 73–76 are towards the centre of mass of the molecule; only the movement at 62–65 is away from the centre of mass. The largest individual changes occur at Ser48 (obviously due to the peptide flip) and at Asn64, a residue with a large temperature factor (Fig. 12) and other symptoms of flexibility (Table 4). The locations of the cited residues in the molecule can be inferred from the atomic labels in Fig. 3.

Fig. 10 is a matrix plot (Berghuis & Brayer, 1992) showing the differences between $C^{\alpha}\cdots C^{\alpha}$ distances for all possible pairs of C^{α} atoms. A coherent row or column of positive or negative differences indicates that a particular C^{α} atom has moved with respect to a series of other C^{α} atoms, *i.e.* that its position has changed significantly. In agreement with Fig. 9, Ser48 and Asn64 are identified as the residues undergoing the largest changes in position, and concerted movements are seen to occur at the other groups of residues mentioned above. An important additional conclusion from Fig. 10 is that cooling causes a slight shrinkage of the molecule (see figure caption). Thus the conclusion drawn earlier from the decrease in the unit-cell volume is supported.

Two buried solvent molecules

An aspect of the protein structure not previously noted in PoPc crystals is the existence of two buried (internal) water molecules, *W*333 and *W*334 (Table S1; see footnote concerning Supplementary Material).

In the Hamburg refinement these two water molecules were included at an early stage. In the Sydney refinement the same electron-density features were observed but were not included in the model because they made short contacts with several C atoms. Two analogous electron-density peaks [$>3\sigma(\rho)$] in the 1.33 Å RT analysis had earlier been rejected as solvent for the same reason (Guss, Bartunik & Freeman, 1992). The unequivocal identification of two solvent molecules in the same region of the oleander plastocyanin structure (Tong, 1991) strongly indicates that the internal solvent molecules in poplar plastocyanin are real. The omission of *W*333 and *W*334 from the refinement of the Sydney model did not lead to any significant differences from the Hamburg model in the surrounding protein structure.

As shown in Fig. 11, the two internal solvent molecules act as bridges between backbone amide groups of residues 22–25 at the 'southern' ends of β-strands 2 and 3, and residues 72–74 on β-strand 6. Thus they fill the gap in a classic β-bulge (Richardson, Getzoff & Richardson, 1978) between two anti-parallel β-strands. We note that the residues that are in contact with the internal solvent molecules are among those whose polypeptide backbone atoms undergo concerted changes in position when the crystals are cooled (see above).

The molecular structure of PoPc at 173 K: *side chains*

When all 738 non-H atoms in the LT and RT models of the molecule are superposed, 26 atoms

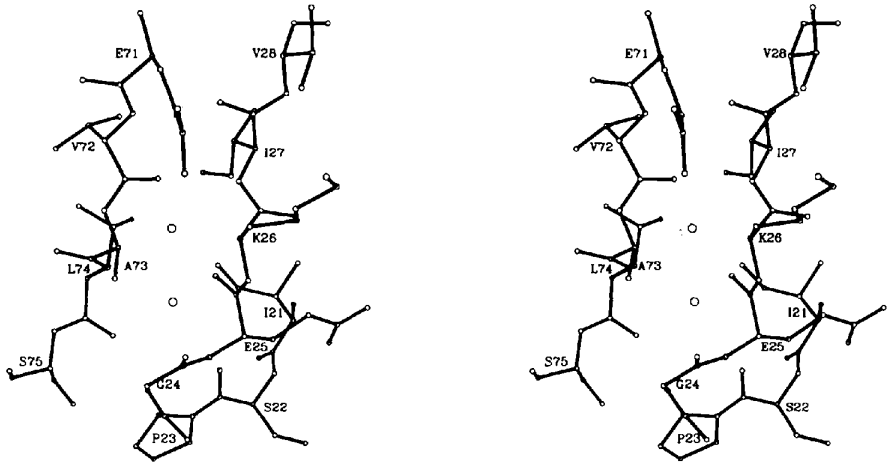


Fig. 11. Two water molecules *W*333 and *W*334, partially buried in the gap between a β-bulge in strand 3 and the anti-parallel strand 6 of the polypeptide. The distance between *W*333 and *W*334 is 2.8 Å (Hamburg model). Hydrogen bonds: *W*333 to Glu25 O, 2.7 Å; Val72 O, 2.7 Å; Leu74 N, 2.8 Å. *W*334 to Ser22 O, 3.2 Å; Glu25 O, 2.7 Å; Glu25 N, 3.1 Å; Leu74 O, 2.7 Å; Leu74 N, 2.7 Å. Short contacts: *W*333 to Ile27 $C^{\delta 1}$, 3.0 Å; Ile27 $C^{\gamma 1}$, 3.0 Å; Ala73 C^{α}, 2.9 Å; Ala73 C, 3.2 Å. *W*334 to Ile21 $C^{\delta 1}$, 3.0 Å.

have deviations ≥ 1 Å.† If these atoms are excluded, superposition of the remaining 712 protein atoms yields an r.m.s. difference of 0.29 Å, not much larger than the value for the polypeptide backbone atoms (see above).

An important difference between the descriptions of the side chains at LT and RT is that there are only three disordered side chains in the LT model (Ser56, Glu79 and Ser85), but nine disordered side chains in the RT model (Lys30*, Ser45*, Lys54*, Ser56*, Asp61*, Asn64*, Lys66*, Glu71* and Ser75*). Only one side chain, Ser56, is described as disordered and having similar conformers in both models. Five side chains in the LT model have the same conformation as one of the two conformers of the corresponding side chains in the RT model (Ser45, Lys54, Ser75, Asp61 and Asn64). At 12 side chains there are substantial changes in conformation (Lys26, Lys30, Pro36, Glu43, Ser48, Ser58, Glu60, Lys66, Glu68, Glu71 and Glu79). The existence of individual differences ≥ 1 Å between the LT and RT atom positions in some of these side chains has already been noted. Comparisons between side-chain conformations at LT and RT are summarized in Table 4.

Only two of the differences in side-chain conformation are associated with differences in intramolecular hydrogen bonding. At RT, both conformers of Lys30* form a long intramolecular salt bridge (4.1 Å) with Asp2 $O^{\delta 1}$; at LT, Lys30 N^{ζ} is hydrogen bonded to Gly67 O. At RT, Ser58 O^{γ} forms an intramolecular hydrogen bond to Glu60 $O^{\epsilon 2}$; at LT, Glu60 $O^{\epsilon 1}$ is hydrogen bonded to Asn76 $N^{\delta 2}$ in a neighbouring molecule (see Fig. 15, below). The interactions ascribed to the side chains of Lys26 and Lys66* at RT remain intact despite the changes in conformation, since the positions of the N^{ζ} atoms remain unchanged.

The structure of PoPc at 173 K: *temperature parameters*

The average value $\langle B \rangle = 6.6$ Å^2 derived from the Wilson plot for the data at 173 K (Fig. 5) is significantly lower than the value $\langle B \rangle = 15.5$ Å^2 calculated from the deposited (1PLC) B's at 295 K. While the individual B's are lower at 173 K, the general trends in the temperature factors plotted as functions of residue number are similar at 173 and 295 K (Fig. 12). The effects of cooling are, however, not uniform throughout the molecule. At 295 K, the backbone atoms with the highest temperature factors occur at Gly24, Glu25, Asn64* and Ser75*. At 173 K, the backbone atoms with the highest temperature factors occur at Pro47, Ser48, Asn64 and Ala65. As noted by Cusack (1992), while the *overall* behaviour of crystallographic B factors is typically biphasic, *individual* residues within a protein exhibit a wide range of temperature dependences indicating local variations in dynamical behaviour.

*Increased order in solvent region**

There are significant differences between the solvent structures at LT and RT. The number of indentifiable solvent sites has increased by 70% from 110 at RT to 189 at LT. The shrinkage of the unit cell and the significant re-orientation of the protein molecules are associated with changes in the positions of many solvent sites. Only 63 of the solvent molecules in the LT model lie within 1.0 Å of a solvent site with occupancy ≥ 0.50 in the RT structure. The average B for these 63 solvent molecules has

† The differences ≥ 1 Å occur at side-chain atoms of Lys26, Lys30*, Lys54*, Ser58, Glu60, Asp61*, Asn64*, Lys66*, Glu68, Glu71*, Glu79, Lys 95, and at all atoms except N of Ser48. Asterisks denote residues that are disordered in the RT structure at 1.33 Å resolution (Guss, Bartunik & Freeman, 1992).

* In the description of the LT solvent structure, it is assumed that the solvent molecules are water. The Sydney (*PROLSQ*) coordinates are used except where noted otherwise.

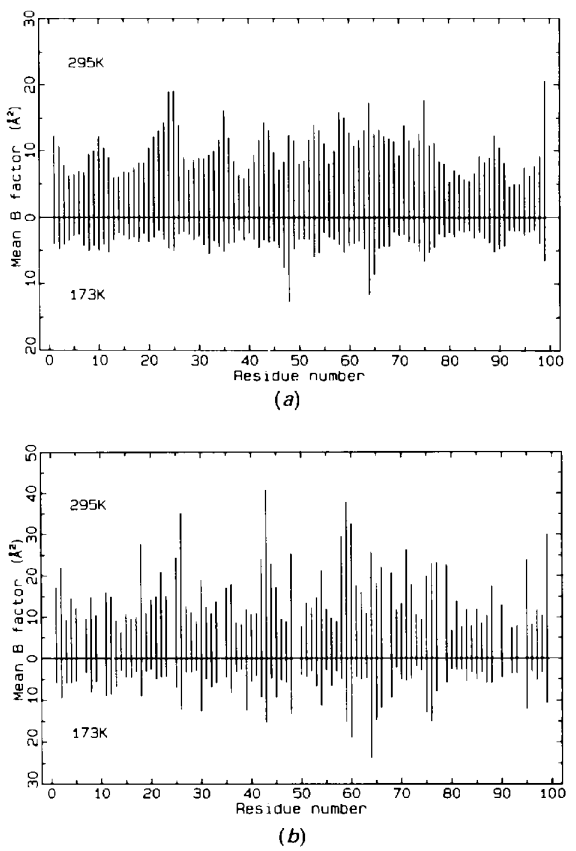


Fig. 12. Temperature factors $\langle B \rangle$ averaged over (*a*) backbone atoms, and (*b*) side-chain atoms of each residue. Above line, 295 K; below line, 173 K.

decreased from 28 Å^2 at RT to 11 Å^2 at LT. As we shall show below, the number of solvent molecules hydrogen bonded only to other solvent molecules has increased more (from 16 to 58, 260%) than the number hydrogen bonded to protein atoms (from 94 to 131, 40%).

The generally lower B's of the solvent molecules and the fact that a larger number of solvent sites can be located are symptoms of greater order in the solvent region at LT. Two extended networks of water molecules can be identified. Both networks occur in regions of close approach between protein molecules. One network includes several five- and six-membered rings as well as three-, four- and >six-membered rings (Fig. 13). In the other network, an almost planar five-membered ring is wedged between Phe35 of one protein molecule and Phe70 of another (Fig. 14). The tendency of the water structure to be ordered near hydrophobic regions of the protein surface is well known (Teeter, 1991). Examples of similar hydrogen-bonded solvent clusters in other proteins are 'four pentagonal rings ... along a hydrophobic surface at an intermolecular contact' in crambin (Teeter, 1984), and 'a five-membered ring system covering an exposed non-polar side chain (Val)' in 2-Zn insulin (Baker, Dodson, Dodson, Hodgkin & Hubbard, 1987).

Some statistics of the types of contact made by the solvent molecules in the LT crystals (RT values in parentheses) are: 4 (18) are hydrogen bonded only to protein atoms, 127 (76) are hydrogen bonded both to protein atoms and to other water molecules, and the remaining 58 (16) are hydrogen bond only to other water molecules. The 131 (94) water molecules making contacts with the protein include 46 (22) that are hydrogen bonded only to side chains, 49 (40) that

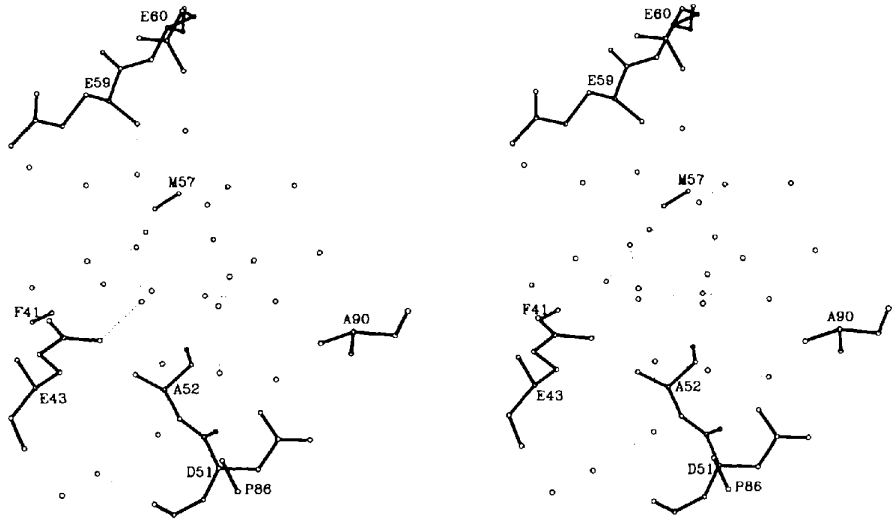


Fig. 13. Ordered water structure including one three-membered ring, four pentagons and five hexagons at an intermolecular contact region. Near the bottom of the figure, a water molecule (with $B = 5$ Å^2) is within hydrogen-bonding distance of two peptide O atoms, one peptide N atom and two other water O atoms. For clarity only the C=O groups of Phe41, Met57 and Pro86 are shown.

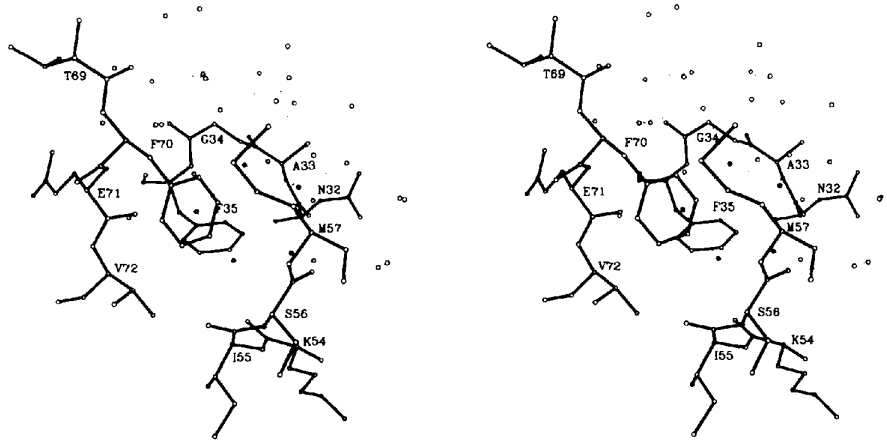


Fig. 14. Pentagonal ring of hydrogen-bonded water molecules between two exposed hydrophobic residues, Phe35 and Phe70, in neighbouring protein molecules. The pentagon is almost planar, and the water molecules in it have low temperature factors ($B = 6$, 6, 8, 10 and 10 Å^2). Three atoms in the pentagon each form a hydrogen bond to a protein atom.

Table 7. *Intermolecular hydrogen bonds in poplar plastocyanin crystals at* 173 K (*LT*) *and* 295 K (*RT*)

Contacts‡				$d_{X\cdots Y}$ (Å) LT	RT
Asp8	O	···Ser17^{v}	O^{γ}	2.7	2.9
Ser20	O^{γ}	···Asp61^{xi}	$O^{\delta1}$	2.5	2.6
Gly24	O	···Asn64^{i}	$N^{\delta2}$	3.3	3.2
Asp44	$O^{\delta1}$	···Thr69^{viii}	$O^{\gamma1}$	2.9	3.1
Asp44	$O^{\delta2}$	···Glu71^{viii}	$O^{\epsilon1}$	*	3.2
Asp44	$O^{\delta2}$	···Glu71^{viii}	$O^{\epsilon2}$	*	3.0
Asp44	$O^{\delta2}$	···Glu71^{viii}	$O^{\epsilon2}$	*	3.2†
Ser48	O^{γ}	···Asn99^{ix}	OT	2.8	3.1
Gly49	O	···Ser85^{vii}	O^{γ}	2.5	2.7
Asp51	$O^{\delta1}$	···Gly89^{vii}	N	2.9	2.9
Lys54	N^{ζ}	···Gly88^{vii}	$O^{\epsilon1}$	2.6	2.9
Glu60	$O^{\epsilon1}$	···Asn76^{ix}	$N^{\delta2}$	3.0	*
Glu60	O	···Asn99^{xi}	$N^{\delta2}$	3.0	2.7

Symmetry codes: (i) $-x, -\frac{1}{2}+y, \frac{1}{2}-z$; (iv) $\frac{1}{2}+x, \frac{1}{2}-y, 1-z$; (v) $-\frac{1}{2}+x, 1\frac{1}{2}-y, 1-z$; (vii) $1-x, -\frac{1}{2}+y, \frac{1}{2}-z$; (viii) $1+x, y, z$; (ix) $1-x, \frac{1}{2}+y, \frac{1}{2}-z$; (xi) $\frac{1}{2}-x, 1-y, -\frac{1}{2}+z$; (xii) $\frac{1}{2}-x, 1-y, \frac{1}{2}+z$.

* At LT, Asp44 and Glu71 are bridged by two water molecules (see text and Fig. 7).

† Alternate conformer of Glu71 at RT.

‡ At RT, this distance is 3.5 Å, and Glu60 $O^{\epsilon2}$ is hydrogen bonded to Ser58 O^{γ} (see text and Fig. 16).

are hydrogen bonded only to backbone amide groups, and 36(32) that are hydrogen bonded both to side chains and to backbone amide groups. There are water–protein hydrogen bonds at *all* side chains with charged or polar functional groups, with the exception of Asn31, Asn38, Ser56 and the four Cu-binding residues His37, Cys84, His87 and Met92. In the original 1.6 Å RT structure, the exposed N(imidazole) atom of His87 was thought to be hydrogen bonded to a solvent molecule. The absence of this solvent molecule from the LT structure is consistent with the 1.33 Å RT structure. However, as discussed elsewhere (Guss, Bartunik & Freeman, 1992), the lack of a hydrogen-bonded solvent molecule dose not alter the description of $N^{\epsilon2}$(His87) as a protonated N(imidazole) atom.

The total numbers of water–O(peptide) and water–N(peptide) contacts are 90 (69) and 34 (32), respectively. The ratio 90:34 = 2.65:1 agrees with the value 2.8:1 obtained by Finney (1979) for a sample of ten refined protein structures. The apparently global validity of this statistic reflects the proportion of H to O atoms in H_2O, the ability of O(peptide) atoms to accept more than one hydrogen bond whereas N(peptide) atoms can donate only one proton, and the fact that the O(peptide) atom but not the N(peptide) atom of Pro residues can participate in hydrogen bonds.

Direct and solvent-bridged intermolecular contacts

Differences between the direct intermolecular protein–protein contacts at LT and RT are only minor (Table 7). Seven intermolecular hydrogen bonds are shorter, and only three are longer, at LT than at RT. At LT, Glu60 forms an intermolecular hydrogen bond (Glu60 $O^{\epsilon1}\cdots N^{\delta2}$ Asn76, 3.0 Å). At RT, this interaction is non-existent or very weak (3.5 Å), and – as mentioned earlier – Glu60 forms an intramolecular hydrogen bond to Ser58 O^{γ} (Fig. 15).

In contrast with the relatively small changes in *direct* intermolecular protein–protein interactions, cooling causes a substantial rearrangement of con-

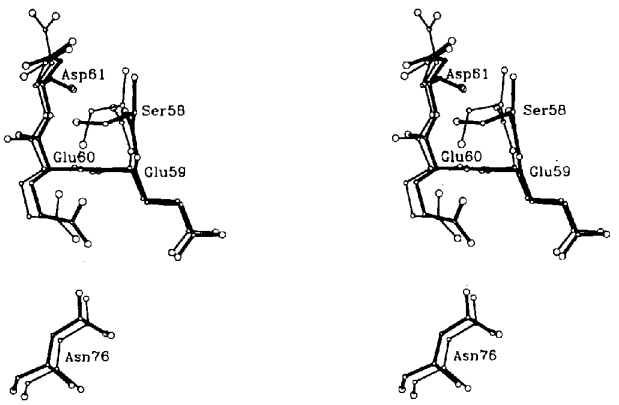


Fig. 15. A significant change in intra- and intermolecular hydrogen bonding. At 295 K (light bonds), the OH group of Ser58 is hydrogen bonded to the COO^- group of Glu60. At 173 K (heavy bonds), the intramolecular hydrogen bond from Ser58 is made with the backbone >NH of Asp61, while the COO of Glu60 forms an additional hydrogen bond with the side-chain amide group of Asn76 in a neighbouring molecule.

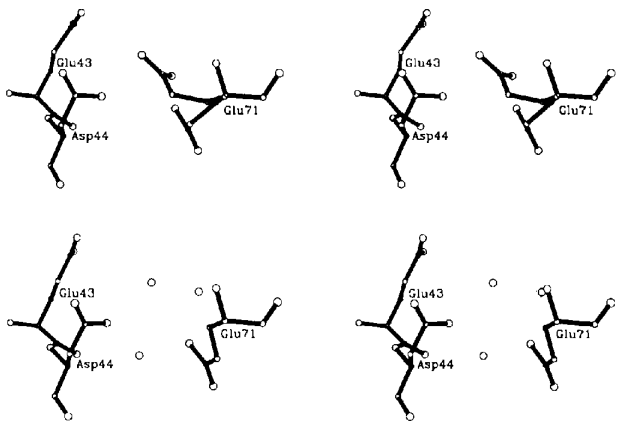


Fig. 16. Increased order at 173 K. In the structure at 295 K (top), the side chain of Glu71 is disordered. A direct $COO^-\cdots H^+\cdots$ OOC interaction with the side chain of Asp44 in an adjacent molecule can be inferred (Guss, Bartunik & Freeman, 1992). At 173 K (bottom), the intermolecular link between the COO^- groups of the same residues is made *via* two bridging water molecules.

tacts *via* bridging solvent molecules. The *number* of water molecules identified in intermolecular bridges remains almost constant (36 *versus* 33). The *positions* of the bridging water molecules, however, are significantly different. 12 bridging water sites are unique to the LT structure. Conversely, ten bridging water sites are unique to the RT structure. Among the 45 bridging protein–solvent hydrogen bonds that are common to the LT and RT structures, 30 are shorter at LT. The fact that the average temperature factor of the bridging waters at LT is very low ($\langle B \rangle = 13$ Å^2) is consistent with the greater ordering of the solvent in the LT structure.

The LT structure provides one example of ordered solvent replacing the second conformer of a disordered side chain in the 1.33 Å RT structure. In the LT model, the carboxylate group of Glu71 is hydrogen bonded to three solvent molcules (*W*234, *W*240 and *W*267). *W*234 and *W*267 act as bridges to the carboxylate of Asp44 in a neighbouring molecule (Fig. 16, bottom). A high degree of order is indicated by low temperature factors at the bridging solvent oxygens (12 and 15 Å^2) and the side chains of Asp44 and Glu71 ($\langle B \rangle = 4.9$ and 5.4 Å^2). At LT there is thus no hint of disorder at the side chain of Glu71. In the RT structure, the Glu71 side chain was modelled as disordered and making a hydrogen-bonded —$COO^- \cdots H^+ \cdots {}^-OOC$— contact with the side chain of Asp44 (Fig. 16, top). The effect of cooling the crystal has been the replacement of the second conformer of the Glu71 side chain by *W*234 and *W*240; the occupancy of the latter is, however, only partial (Table S1; see footnote concerning Supplementary Material).

Is there a change in Cu-site geometry at low temperature?

The present work was undertaken largely to test a hypothesis based on resonance Raman experiments that the Cu-site geometry in plastocyanin changes significantly when the protein is cooled to below ~273 K (Woodruff, Norton, Swanson & Fry, 1984; Blair *et al.*, 1985). Three changes in the Cu-site geometry were proposed: (i) a decrease in the Cu—S(Met92) bond length, (ii) a consequential inhibition of large-amplitude vibrations of the Cu atom perpendicular to the N_2S(Cys) plane, and (iii) a change in the Cu—S^{γ}(Cys84)—C^{β}—C^{α} dihedral angle. Whether the proposed structural changes were derived from a correct assignment of rR frequencies is a separate question. Our concern was (and remains) whether the LT crystal structure analysis is capable of casting any light on the rR spectroscopic changes that were reported to occur at ~273 K.

At first sight, the absence of a significant change in the crystallographically determined Cu-site geometry between 295 and 173 K is in conflict with the LT rR results. However, we cannot eliminate the possibility that structural changes which are detected by rR spectroscopy are lost in the noise of the crystal structure analysis. In commenting on the relationship between the two types of experiment we labour under the difficulty that we can estimate the precision of the crystallographic results, but that we have no estimates for the magnitudes of the structural changes that would be required to account for the reported temperature dependence of the rR spectra. Our conclusions are as follows.

(i) *Change in the* Cu—S(*Met*92) *bond length*

Within the limits of precision, the hypothesis that the Cu—S(Met92) bond becomes shorter when the protein is cooled below about 273 K can be eliminated. This conlusion is based not merely on Table 5 but also on the following computational experiment. The conformation of the Met92 side chain was modified in a plausible way using *FRODO* to reduce the Cu—S(Met92) distance to 2.4 Å. Six cycles of *PROLSQ* least-squares refinement were carried out. The Met92 side chain returned to its original conformation with Cu—$S^{\delta}_{92} = 2.77$ (5) Å. The other three Cu–ligand bond lengths changed by less than 0.01 Å and the residual *R* remained unchanged (0.132).

An alternative hypothesis – that, as the temperature is reduced, the protein contains an increasing proportion of molecules with a shorter Cu—S(Met92) distance – cannot be eliminated. To test this hypothesis, the Met92 side chain was replaced by two plausible conformers, each with occupancy 0.5, having Cu—$S^{\delta}_{92} = 2.4$ and 3.2 Å, respectively. It is to be noted that the positions of S^{δ}_{92} in the two conformers differed by only 0.8 Å, half the limit of resolution of the diffraction data. After 15 cycles of *PROLSQ* refinement, both conformers had changed in the direction of the unmodified conformation but had not converged (Cu—$S^{\delta}_{92} = 2.65$, 2.93 Å). Again the other three Cu–ligand bond lengths and the residual *R* were unchanged.

(ii) *Change in anisotropic vibrations of the* Cu *atom*

The crystallographic data neither support nor eliminate the hypothesis that at LT the large-amplitude thermal vibrations of the Cu atom perpendicular to the plane of the three strong Cu–ligand bonds are inhibited by a strengthening of the Cu···S(Met92) interaction.

The isotropic temperature factor of the Cu atom decreased from $B = 8.8$ Å^2 at RT to $B = 4.6$ Å^2 at LT. An $F_o - F_c$ map using the LT diffraction data had negative peaks along the Cu—N(His37) and Cu—S(Cys84) bonds, and a positive peak between

the Cu atom and S(Met92). An $F_o - F_c$ map calculated with the 1.33 Å RT data was qualitatively similar. The maps showed that the Cu-atom vibrations are anisotropic at RT and remain so at LT, but they did not eliminate the hypothesis that the amplitude or frequency of vibration undergoes a change.

An attempt to test for a change in anisotropy at the Cu site by calculating an $F_{LT,\varphi LT} - F_{RT,\varphi RT}$ electron-density difference map was unsuccessful. The movement of the molecules within the unit cell swamped any changes in the temperature factors at the Cu site. The displacement of the Cu site produced a large positive electron-density difference at the LT Cu position and a corresponding negative feature at the RT Cu position.

(iii) *Change in the Cys84 side-chain conformation*

Within the limits of precision, the LT structure analysis provides no evidence for a change in the Cu—S^{γ}(Cys84)—C^{β}—C^{α} dihedral angle (Table 5). The conformation of the Cu-binding Cys side chain is highly conserved in 'blue' copper proteins and may indicate a favoured pathway for electron transfer (Han *et al.*, 1991).

(iv) *Summary*

There is no crystallographic evidence to support the hypothesis that the plastocyanin Cu site undergoes a structural change involving the side chains of Cys84 and/or Met92 at LT. We recall that LT EXAFS measurements on solutions and crystals of poplar plastocyanin likewise yielded no evidence of a shorter Cu—S(Met92) distance; temporary doubts about this conclusion were caused by an artefact (Penner-Hahn, Murata, Hodgson & Freeman, 1989). On the other hand, it is possible that rR measurements are sensitive to temperature-dependent structural changes that are simply undetectable by X-ray diffraction and X-ray spectroscopy.

Concluding remarks

There is close agreement between models of poplar plastocyanin obtained by two independent refinements using a single set of LT diffraction data. Differences that appear to be statistically significant in terms of the precision of the individual analyses can in general be attributed to the different force fields used in the refinement programs, and to different interpretations of electron-density maps.

Cooling the crystals of poplar plastocyanin from 295 to 173 K causes a significant reduction in the volume of the unit cell, a rotation of the protein molecules with respect to the unit-cell axes, and an increase of order in the solvent region. Additional solvent sites have been identified. The temperature factors in the solvent region are substantially lower than at RT.

Within the limits of precision, the Cu-site geometry remains unchanged. There are only minor changes in the polypeptide backbone conformation, except for a peptide flip at Ser48-Gly49. Independent evidence from NMR spectra and from the structure analyses of other plastocyanins indicates that the two conformations at residues 48–49 are easily interconvertible. Side chains with significantly different conformations at LT occur at four Lys residues (26, 30, 66, 95), five Glu residues (43, 60, 68, 71, 79), two Ser residues (48, 58), and Pro36. At five other side chains, the LT model has a single conformer with the same conformation as one of two conformers modelled at RT (Ser45, Lys54, Asp61, Asn64 and Ser75). One residue is modelled as disordered at LT but not at RT, one of the LT conformers having the same conformation as in the RT model. An unexpected discovery is two water molecules, buried beneath the protein molecular surface and omitted from earlier descriptions of the plastocyanin structure.

This work was supported by the Australian Research Council (grant A28930307 to HCF and JMG) and by the German Federal Minister for Research and Technology (BMFT, under the contract number 05 180MP B 0). Helpful discussions with Drs W. H. Woodruff and T. M. Loehr are gratefully acknowledged.

References

BAKER, E. N., DODSON, E., DODSON, G., HODGKIN, D. & HUBBARD, R. (1987). In *Crystallography in Molecular Biology*, edited by D. MORAS, J. DRENTH, B. STRANDBERG, D. SUCK & W. WILSON, pp. 179–192. New York: Plenum Press.

BARTUNIK, H. D. & SCHUBERT, P. (1982). *J. Appl. Cryst.* **15**, 227–231.

BARTUNIK, H. D., SUMMERS, L. J. & BARTSCH, H. H. (1989). *J. Mol. Biol.* **210**, 813–828.

BERGHUIS, A. M. & BRAYER, G. D. (1992). *J. Mol. Biol.* **223**, 959–976.

BLAIR, D. F., CAMPBELL, G. W., SCHOONOVER, J. R., CHAN, S. I., GRAY, H. B., MALMSTRÖM, B. G., PECHT, I., SWANSON, B. I., WOODRUFF, W. H., CHO, W. K., ENGLISH, A. M., FRY, H. A., LUM, V. & NORTON, K. A. (1985). *J. Am. Chem. Soc.* **107**, 5755–5766.

BRÜNGER, A. T. (1988). In *Crystallographic Computing* 4: *Techniques and New Technologies*, edited by N. W. ISAACS & M. R. TAYLOR, pp. 126–140. Oxford Univ. Press.

BRÜNGER, A. T., KARPLUS, M. & PETSKO, G. A. (1989). *Acta Cryst.* A**45**, 50–61.

BÜCHI, F. N., BOND, A. M., CODD, R. M., HUQ, L. N. & FREEMAN, H. C. (1992). *Inorg. Chem.* **31**, 5007–5014.

CHAMBERS, J. L. & STROUD, R. M. (1979). *Acta Cryst.* B**35**, 1861–1874.

CHAPMAN, G. V., COLMAN, P. M., FREEMAN, H. C., GUSS, J. M., MURATA M., NORRIS, V. A., RAMSHAW, J. A. M. & VENKATAPPA, M. P. (1977). *J. Mol. Biol.* **110**, 187–189.

COLLYER, C. A., GUSS, J. M., SUGIMURA, Y., YOSHIZAKI, F. & FREEMAN, H. C. (1990). *J. Mol. Biol.* **211**, 617–632.

COLMAN, P. M., FREEMAN, H. C., GUSS, J. M., MURATA, M., NORRIS, V. A., RAMSHAW, J. A. M. & VENKATAPPA, M. P. (1978). *Nature (London)*, **272**, 319–324.

CUSACK, S. (1992). *Curr. Biol.* **2**, 411–413.

FINNEY, J. L. (1979). In *Water, a Comprehensive Treatise*, Vol. 6, edited by F. FRANKS, pp. 47–122. New York: Plenum Press.

FUJINAGA, M., GROS, P. & VAN GUNSTEREN, W. F. (1989). *J. Appl. Cryst.* **22**, 1–8.

GARRETT, T. P. J. (1987). PhD thesis, Univ. of Sydney, Australia.

GEWIRTH, A. A., COHEN, S. L., SCHUGAR, H. J. & SOLOMON, E. I. (1987). *Inorg. Chem.* **26**, 1133–1145.

GEWIRTH, A. A. & SOLOMON, E. I. (1988). *J. Am. Chem. Soc.* **110**, 3811–3819.

GUSS, J. M., BARTUNIK, H. D. & FREEMAN, H. C. (1992). *Acta Cryst.* B**48**, 790–811.

GUSS, J. M. & FREEMAN, H. C. (1983). *J. Mol. Biol.* **169**, 521–563.

GUSS, J. M., HARROWELL, P. R., MURATA, M., NORRIS, V. A. & FREEMAN, H. C. (1986). *J. Mol. Biol.* **192**, 361–387.

HAN, J., LOEHR, T. M., FREEMAN, H. C., CODD, R., HUQ, L., BEPPU, T., ADMAN, E. T. & SANDERS-LOEHR, J. (1991). *Biochemistry*, **30**, 10904–10913.

HENDRICKSON, W. A. & KONNERT, J. H. (1980). In *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 13.01–13.23. Bangalore: Indian Academy of Sciences.

JACK, A. & LEVITT, M. (1978). *Acta Cryst.* A**34**, 931–935.

JACKMAN, M. P., MCGINNIS, J., POWLS, R., SALMON, G. A. & SYKES, A. G. (1988). *J. Am. Chem. Soc.* **110**, 5880–5887.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

KABSCH, W. & SANDER, C. (1983). *Biopolymers*, **22**, 2577–2637.

KYRITSIS, P., LUNDBERG, L. G., NORDING, M., VANNGÅRD, T., YOUNG, S., TOMKINSON, N. P. & SYKES, A. G. (1991). *J. Chem. Soc. Chem. Commun.* pp. 1441–1442.

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

LEVITT, M. (1974). *J. Mol. Biol.* **82**, 393–420.

LEVITT, M. (1975). Personal communication cited by C. CHOTHIA [*Nature (London)* (1975), **254**, 304–308].

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

MESSERSCHMIDT, A. & PFLUGRATH, J. W. (1987). *J. Appl. Cryst.* **20**, 306–315.

MOORE, J. M., LEPRE, C. A., GIPPERT, G. P., CHAZIN, W. J., CASE, D. A. & WRIGHT, P. E. (1991). *J. Mol. Biol.* **221**, 533–555.

MURPHY, L. M., HASNAIN, S. S., STRANGE, R. W., HARVEY, I. & INGLEDEW, W. J. (1990). In *X-ray Absorption Fine Structure*, edited by S. S. HASNAIN, pp. 152–154. Chichester: Ellis Horwood.

PARAK, F., HARTMANN, H., AUMANN, K. D., REUSCHER, H., RENNEKAMP, G., BARTUNIK, H. D. & STEIGEMANN, W. (1987). *Eur. Biophys. J.* **15**, 237–249.

PENFIELD, K. W., GEWIRTH, A. A. & SOLOMON, E. I. (1985). *J. Am. Chem. Soc.* **107**, 4519–4529.

PENNER-HAHN, J. E. & HODGSON, K. O. (1986). In *Structural Biological Applications of X-ray Absorption, Scattering, and Diffraction*, edited by H. D. BARTUNIK & B. CHANCE, pp. 35–47. New York: Academic Press.

PENNER-HAHN, J. E., MURATA, M., HODGSON, K. O. & FREEMAN, H. C. (1989). *Inorg. Chem.* **28**, 1826–1832.

READ, R. J., FUJINAGA, M., SIELECKI, A. R. & JAMES, M. N. G. (1983). *Biochemistry*, **22**, 4420–4433.

RICHARDSON, J. S., GETZOFF, E. D. & RICHARDSON, D. C. (1978). *Proc. Natl Acad. Sci. USA*, **254**, 2574–2578.

RYDEL, T. J., TULINSKY, A., BODE, W. & HUBER, R. (1991). *J. Mol. Biol.* **221**, 583–601.

SCOTT, R. A., HAHN, J. E., DONIACH, S., FREEMAN, H. C. & HODGSON, K. O. (1982). *J. Am. Chem. Soc.* **104**, 5364–5369.

STEIGEMANN, W. (1974). PhD thesis, Technische Univ. München, Germany.

SYKES, A. G. (1990). *Struct. Bonding*, **75**, 175–224.

TEETER, M. M. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 6014–6018.

TEETER, M. M. (1991). *Annu. Rev. Biophys. Chem.* **20**, 577–600.

TEN EYCK, L. F. (1977). *Acta Cryst.* A**33**, 486–492.

TILTON, R. F., DEWAN, J. C. & PETSKO, G. A. (1992). *Biochemistry*, **31**, 2469–2481.

TONG, H. (1991). PhD thesis, Univ. of Sydney, Australia.

VAN GUNSTEREN, W. F. & BERENDSEN, H. J. C. (1987). *GROMOS, Groningen Molecular Simulation Library*. Groningen, The Netherlands: BIOMOS b.v.

VENKATACHALAM, C. M. (1968). *Biopolymers*, **6**, 1425–1436.

WALTER, J., STEIGEMANN, W., SINGH, T. P., BARTUNIK, H. D., BODE, W. & HUBER, R. (1982). *Acta Cryst.* B**38**, 1462–1472.

WATT, W., TULINSKY, A., SWENSON, R. P. & WATENPAUGH, K. D. (1991). *J. Mol. Biol.* **218**, 195–208.

WILKE, M. E., HIGAKI, J. N., CRAIK, C. S. & FLETTERICK, R. J. (1991). *J. Mol. Biol.* **219**, 511–523.

WOODRUFF, W. H., NORTON, K. A., SWANSON, B. I. & FRY, H. A. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 1263–1267.