Acta Cryst. (1992). B48, 790–811

Accuracy and Precision in Protein Structure Analysis: Restrained Least-Squares Refinement of the Structure of Poplar Plastocyanin at 1.33 Å Resolution

BY J. M. GUSS

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

H. D. BARTUNIK

EMBL Outstation, DESY, Hamburg, Germany

AND H. C. FREEMAN

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

(*Received* 30 *December* 1991; *accepted* 14 *April* 1992)

Abstract

The structure of the electron-transfer protein, plastocyanin (99 amino acids, one Cu atom, 10 500 Da) from poplar leaves, has been refined at 1.33 Å resolution to a residual $R = 0.15$. The space group is orthorhombic, $P2_12_12_1$, $a = 29.60$ (1), $b = 46.86$ (3), $c = 57.60$ (3) Å. The 14 303 reflections used in the refinement were obtained from a data set recorded on a four-circle diffractometer with radiation from a sealed fine-focus tube, combined with a data set measured on oscillation films exposed at the DESY synchrotron. The final model comprises 1442 (738 non-H) protein atoms, one Cu atom and 110 solvent molecules. Nine residues are described as disordered. The root-mean-square deviation from ideal bond lengths is 0.016 Å and the root-mean-square difference between the positions of the C^α atoms in this refined model and in the structure previously refined at 1.6 Å resolution is 0.11 Å. The effects of manual model adjustment, resolution, choice of standard values for geometrical parameters, inclusion of H atoms and inclusion of anomalous-scattering corrections on the copper-site geometry have been explored. The final values of the Cu–ligand bond lengths are: Cu—N(His37) 1.91, Cu—S(Cys84) 2.07, Cu—N(His87) 2.06, Cu—S(Met92) 2.82 Å.

Introduction

The copper-containing protein plastocyanin from poplar leaves (*Populus nigra* var. *italica*) can be crystallized in the form of mechanically robust crystals with excellent diffraction properties. The availability of such crystals has enabled us to extend our studies from the central problem – what is the structure and how is it related to the function of the protein? – to several other themes. In the present paper we report the refinement of the structure of plastocyanin at 1.33 Å resolution, and discuss the more detailed structural information produced. We also assess the systematic contributions of a number of crystallographic refinement variables to the accuracy and precision of the metal-site geometry.

Plastocyanin, a 'type-1' or 'blue' copper protein originally extracted from *Chlorella ellipsoidea* and identified by Katoh (1960), is now known to occur in all higher plants, green algae and some blue–green algae (Boulter, Haslett, Peacock, Ramshaw & Scawen, 1977). It functions as an electron carrier in photosynthesis between photosystems II and I, receiving an electron from the cytochrome b_6/f complex and donating the electron to the $P700^+$ photoreaction center.

The crystallization of poplar plastocyanin was reported in 1977 (Chapman, Colman, Freeman, Guss, Murata, Norris, Ramshaw & Venkatappa, 1977). The structure was solved by multiple isomorphous replacement (Colman, Freeman, Guss, Murata, Norris, Ramshaw & Venkatappa, 1978), using the known amino-acid sequence (Ambler, 1983). The structure was subsequently refined by a number of methods, culminating in restrained least-squares calculations to the 1.6 Å resolution limit of the available data (Guss & Freeman, 1983).

The molecule is a β-sandwich with seven strands predominantly in the β conformation, one irregular strand, and the copper-binding site beneath the surface at one end. The Cu atom is coordinated by the N^δ atoms of His37 and His87, the S^γ atom of Cys84, and the S^δ atom of Met92. The coordination geometry is irregular (distorted from C_{3v}) with an abnormally long bond to the methionine S^δ atom. After the 1.6 Å refinement the standard deviations in the copper–ligand bond lengths were estimated to be 0.05 Å.

0108-7681/92/060790-22$06.00

Plastocyanin is a member of a distinctive family of proteins described collectively as the 'blue' copper proteins or cupredoxins (Adman, 1985). It has become apparent from the determination of the structures of other members of the family – azurin (Adman & Jensen, 1981; Baker, 1988), pseudoazurin (Petratos, Dauter & Wilson, 1988; Adman, Turley, Bramson, Petratos, Banner, Tsernoglou, Beppu & Watanabe, 1989) and the basic blue protein from cucumbers (Guss, Merritt, Phizackerley, Hedman, Murata, Hodgson & Freeman, 1988) – that the Cu sites exhibit both strong overall similarities and subtle structural differences. In order to produce statistically significant estimates of these differences it is obviously necessary to have individual structure determinations of the highest possible precision.

The structure of poplar plastocyanin has now been further refined using a combination of the original diffractometer data with new synchrotron X-ray photographic data.

Data collection and processing

Poplar plastocyanin was isolated, purified and crystallized as described by Chapman *et al.* (1977). The space group is orthorhombic, $P2_12_12_1$, with $a = 29.60\ (1)$, $b = 46.86\ (3)$, $c = 57.60\ (3)$ Å. The cell parameters were obtained from the averages for two crystal specimens. The unit-cell dimensions for each crystal specimen were fitted by least squares to the θ values of more than 25 high-angle reflections. The counter data used in the present work were taken from the original data recorded to 1.6 Å resolution on an Enraf–Nonius CAD-4F diffractometer operating with a fine-focus sealed-tube X-ray source (Guss & Freeman, 1983). Data to higher resolution were subsequently measured photographically using the high-intensity beam on the DORIS storage ring at the DESY synchrotron in Hamburg. The final data set was a combination of the diffractometer data for low-angle reflections and the synchrotron data for high angles.

Diffractometer data collection

Data from the first of two crystals ($1.03 \times 0.65 \times 0.40$ mm) were measured to a resolution of 1.8 Å. Unique data ($h \geq 0$, $k \geq 0$, $l \geq 0$) were measured in six shells of increasing θ. A further octant ($h \geq 0$, $k \leq 0$, $l \leq 0$) was then measured over the entire θ range. The decomposition as indicated by at least two and usually five standard reflections was 14% after the 8503 unique measurements, and 30% after the additional 8333 measurements with k and l negative. Data from a second crystal ($0.65 \times 0.40 \times 0.30$ mm) were measured in three shells of θ in the resolution range 1.9–1.6 Å. Only unique reflections in the positive octant were recorded. The data in the range $1.9 \geq d \geq 1.8$ Å were measured last. The decomposition was 18% after recording the $1.8 \geq d \geq 1.6$ Å data and 23% at the end of data collection for this crystal. A total of 4661 measurements was made from the second crystal. The data from the two crystals were placed on a common scale using the overlapping reflections, yielding a data set with 11096 independent reflections. This was the data set used in the refinement at 1.6 Å resolution.

Synchrotron oscillation film data

The film data were collected at the European Molecular Biology Laboratory (EMBL) outstation at the DESY facility. The storage ring DORIS was operated at 3.3 GeV, multi-bunch mode, 30–50 mA, with 2–4 h between reinjections. The crystal-to-film distance was calibrated using the (110) reflection from a tungsten wire ($d = 2.2336$ Å). The wavelength, calibrated using the absorption edge of a copper foil, was 1.06 Å. The beam intensity was monitored with an ion gauge and was used to give an approximately equal flux at the crystal for each exposure (Bartunik, Gehrmann & Robrahn, 1984).

The data were collected on flat film cassettes on a modified Arndt–Wonacott camera (Arndt, 1977) under the control of a PDP-11 computer (Bartunik *et al.*, 1984). The films were intentionally overexposed in order to enhance the weaker reflections at high diffraction angles. The nominal resolution at the edge of the films was 1.32 Å. Three films were used in each pack. Stainless steel foils 25 μm thick were placed between films to increase the absorbance.

Films were scanned at DESY on an Optronics P1000 scanner interfaced *via* a CAMAC crate to a VAX-11/750 minicomputer. The optical density range was 0–3 and the raster size was 100 μm. A non-linear film-response curve was determined experimentally for a series of timed exposures. Integration, background corrections and further reduction of the data were carried out using the program *FILME* (Schwager, Bartels & Jones, 1975).

Data merging and scaling

Data from different films were placed on a common scale using multiple observations of fully recorded reflections as well as the 1.6 Å diffractometer data. Observations partially recorded on adjacent films were summed using the scale factors determined from the fully recorded observations. Partial observations where even one component was flagged by *FILME* as unobserved or too intense were rejected. The scaling and subsequent Fourier calculations were carried out with the program system *PROTEIN* (Steigemann, 1974). The diffractometer data were used only in the scaling process and were

excluded from the calculation of averaged intensities. The resulting 'film only' data set was used to start the refinement of the structure.

When the initial model derived from the 1.6 Å structure (see below) had been partly refined, the quality of the data was checked by comparing the 'film only' F_{obs} values with (i) the original counter data, (ii) F_{calc} values from a model previously refined using the counter data at 1.6 Å resolution, and (iii) F_{calc} values from the model partly refined against the film data. The results led to the realisation that the 'film only' data at low-θ angles were affected by systematic errors (Fig. 1). In view of these errors (which may have been caused by the overexposure of the films and/or by an inadequate film-response curve) the film data with $d > 2.8$ Å were excluded, and a new data set was produced by merging only the high-angle film data with the original counter data. The data set used in the refinement included 76% of the measurable reflections to a resolution of 1.33 Å. The outermost shell of data, $1.38 \geq d \geq 1.33$ Å, is 30% complete. Details of the individual and merged data sets are summarized in Table 1.

Weighting scheme

Standard deviations in the structure-factor observations were determined empirically by a numerical analysis of the differences between structure amplitudes common to the diffractometer (F_{diff}) and film (F_{film}) data sets (Guss & Freeman, 1972). The mean-square differences, $\langle(|F_{diff}| - |F_{film}|)^2\rangle$, were plotted as a function of $|F|$ (Fig. 2). The variance in $|F|$, $\sigma^2(F)$, was given by a second-order function in $|F|$, $\sigma^2(F) = (144 - 1.15|F| + 0.0036|F|^2)$. The coefficients were determined by least squares. The minimum and maximum values of $\sigma(F)$ were $\sigma_{min}(F) = 7.2$ for $F_{min} = 17.4$ and $\sigma_{max}(F) = 49.0$ for $F_{max} = 968.6$. Weights $w = 1/\sigma^2(F)$ were used in the subsequent least-squares refinement.

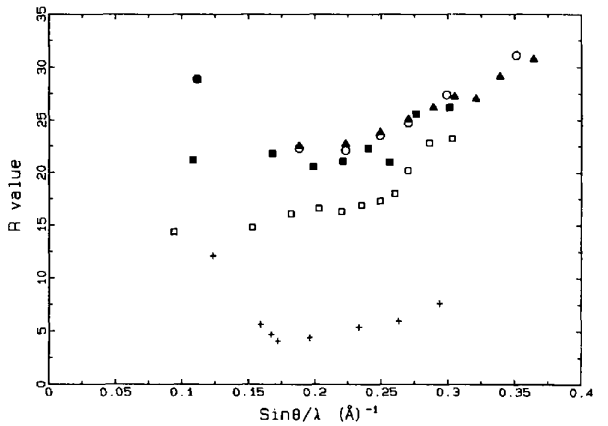

Fig. 1. Analysis demonstrating systematic errors in the film data at low θ angles. Residuals $R = \Sigma||F_A| - |F_B||/\Sigma|F_A|$ are plotted *versus* resolution. The agreement at low θ angles is significantly worse whenever F_A represents F_{obs} from the film data. + F_A, observed structure amplitude from the 1.3 Å 'film only' data; F_B, observed structure amplitude from the 1.6 Å counter data. ○ F_A, observed structure amplitude from the 1.3 Å 'film only' data; F_B, structure amplitude calculated from model refined against the 1.6 Å counter data (Guss & Freeman, 1983). ▲ F_A, observed structure amplitude from the 1.3 Å 'film only' data; F_B, structure amplitude calculated from model partially refined against the 1.3 Å 'film only' data. ■ F_A, observed structure amplitude from the 1.6 Å counter data; F_B, structure amplitude calculated from model partially refined against the 1.3 Å 'film only' data. □ F_A, observed structure amplitude from the 1.6 Å counter data; F_B, structure amplitude calculated from model refined against the 1.6 Å counter data.

Table 1. *Details of data collection and data processing*

(*a*) Counter data to 1.6 Å resolution	
X-ray source	Enraf–Nonius CAD-4F diffractometer
	Fine-focus Cu tube
Operating conditions (kV, mA)	40, 26
Take-off angle (°)	2.8
Crystal-to-detector distance (mm)	173
Method for cell dimensions	Least-squares fit to measured θ values
θ data for cell dimensions	25 reflections ($35 \leq 2\theta \leq 45°$)
Method for intensities	ω scans
Measurements: Total	21 498 reflections
Independent	11 096 reflections
$I \geq 2\sigma(I)$	8391 reflections
(*b*) Photographic synchrotron X-ray data to 1.33 Å resolution	
X-ray source	Storage ring DORIS, EMBL beamline X-11
Operating conditions (GeV, mA)	3.3, 30–50
Wavelength (Å)	1.062 (0.002)
Collimator (mm)	0.3
Crystal-to-film distance (mm)	54.8
Crystal rotation (°): Total	93
Crystal (1)	53
Crystal (2)	43
Overlap	5
Time per exposure (min)	5–10
Films	Kodak No screen (NS-9T)
	Flat cassettes
	3 films per pack
	25 μm stainless steel foils between
Measurements: Total	69 361 reflections
Fully recorded	24 582 reflections
Partial	15 709 reflections
Rejected	29 070 reflections
Independent	13 564 reflections
$R_{symmetry}$	0.12
(*c*) Merged counter and film data to 1.33 Å resolution	
Counter data	8460 reflections ($\infty \geq d \geq 1.60$ Å)
Film data	11 695 reflections ($2.8 \geq d \geq 1.33$ Å)
Combined data set	14 512 reflections
Data common to both sets	5596 reflections
R_{merge}*	0.10

* $R_{merge} = \{\Sigma_r^N[\Sigma_i^{n(r)}|I_{i,r} - \bar{I}_r|]/\Sigma_r^N[n(r)\bar{I}_r]\}$ where N is the number of unique reflections, $I_{i,r}$ is the ith estimate of the intensity of the rth reflection, and $\bar{I}_r$ is the mean value of the intensity of the rth reflection, and $n(r)$ is the number of measurements of the rth reflection in the data set.

Structure refinement

The structure was refined using the restrained reciprocal-space least-squares program *PROLSQ* (Hendrickson & Konnert, 1980; Hendrickson, 1985). The program was modified to allow for multiple kinds of anomalous scatterers. The changes were implemented in a local version of *CALC*, the subroutine that calculates structure factors and their derivatives. Electron density maps were displayed on an

Evans & Sutherland PS330 graphics system with the program *FRODO* (Jones, 1978). The initial model was that previously obtained from the refinement at 1.6 Å resolution (Guss & Freeman, 1983; file 1PCY, Protein Data Bank, Brookhaven National Laboratory, Upton, NY 11973, USA) modified as follows: (i) all solvent atoms were removed, (ii) all temperature factors were set to 15 $Å^2$ and (iii) the occupancies of atoms which were noted in the 1.6 Å structure as poorly determined were set to 0.01.* In order to restrain the geometry of the Cu site as little as possible, the Cu–ligand bond lengths and interbond angles at the Cu atom were not restrained, nor were torsion-angle restraints applied to the side chains of the Cu-binding histidine, cysteine and methionine residues.

The refinement was carried out in 'rounds' which usually comprised (i) a number of cycles of refinement, (ii) the calculation of electron density difference maps and/or 'omit' electron density maps in which the contribution of selected atoms was omitted from the calculated structure amplitudes and phases, and (iii) adjustment of the model, including the addition of O atoms representing solvent. Details of the 23 rounds of refinement, comprising 186 least-squares cycles, are presented in Table 2.

H-atom coordinates were added to the atom parameter file after round 3. The H-atom positions were calculated using the program *HAFFIX* [distributed as part of the June 1982 and subsequent versions of the *PROLSQ* program package (Hendrickson, 1985)]. Methyl and amine groups were placed in staggered conformations. Hydroxyl protons of serine, threonine and tyrosine were adjusted manually to make plausible hydrogen-bonded contacts where possible. The inclusion of H atoms did not markedly affect the agreement between the observed and calculated structure amplitudes, but experience suggests that the presence of hydrogens improves the restraint specifications, especially for non-bonded and hydrogen-bonded contacts (Getzoff, Tainer, Simpson, Bell & Hallewell, 1989).

Solvent sites were identified as electron density difference peaks having an amplitude ≥ 3 e.s.d.'s of the electron density difference *and* having a location appropriate for the formation of hydrogen bonds to one or more plausible donor or acceptor atoms. If a potential solvent site was too close to a previously identified site to permit both sites to be occupied simultaneously, the new site was flagged so that restraints against close contacts with it were not generated. Occupancies were refined as described below. The positions of solvent atoms were checked after most rounds of refinement to confirm that their geometry was still reasonable and that they were still located in significant electron density. After rounds 14 and 18 of the refinement, series of 'omit' maps were calculated with 30% of the solvent removed. Solvent positions not coinciding with significant electron density difference features were deleted.

Electron density difference syntheses also provided evidence that several residues were present in more than one conformation. The disorder was confirmed in each case by refining each conformer in the absence of the other and then observing the presence of the second conformer in a difference synthesis. The atoms of both conformers were subsequently included with fixed fractional occupancies. An 'omit' map for all residues with multiple conformations was computed prior to round 18 to check the disorder.

After round 14 of the refinement, an attempt was made to simulate the effect of the disordered ('non-discrete') solvent by modifying the scattering curves (Langridge, Marvin, Seeds, Wilson, Hooper, Wilkins & Hamilton, 1960; Fraser, Macrae & Suzuki, 1978). An explicit temperature parameter was incorporated in the correction term. While the solvent simulation – as expected – significantly reduced the residual R for data in the range $\infty \geq d \geq 8$ Å, the agreement at high angles became significantly worse. No further attempt was made to model the disordered solvent. The 110 reflections with $d \geq 8$ Å were discarded.

Individual isotropic temperature factors for the protein atoms were introduced early in the refinement (round 1, cycle 4). After this, whenever

* An occupancy of precisely zero would have been treated by the program *PROLSQ* as 1.0. Atoms given low occupancies in the initial model were: Asp2, $C^{\gamma+}$; Glu18, $C^{\gamma+}$; Ser22, O^{γ}; Glu25, $C^{\gamma+}$; Lys26, $C^{\gamma+}$; Lys30, $C^{\gamma+}$; Asp42, $C^{\gamma+}$; Glu43, $C^{\gamma+}$; Ser45, O^{γ}; Ser48, O^{γ}; Lys54, $C^{\gamma+}$; Ser58, O^{γ}; Glu59, $C^{\gamma+}$; Glu60, $C^{\gamma+}$; Asp61, $C^{\gamma+}$; Asn64, $C^{\gamma+}$; Lys66, $C^{\gamma+}$; Glu68, $C^{\gamma+}$; Glu71, $C^{\gamma+}$; Ser75, O^{γ}; Lys77, $C^{\gamma+}$; Glu79, $C^{\gamma+}$; Met92, C^{ϵ}; Lys95, $C^{\gamma+}$; Asn99, $C^{\gamma+}$. The plus symbol indicates that the atom and all atoms further along the side chain were given low occupancies.

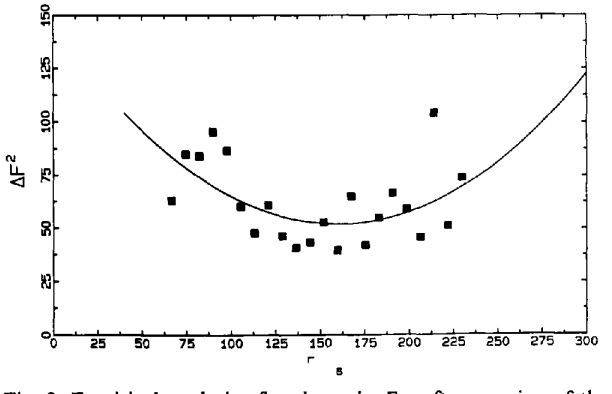

Fig. 2. Empirical analysis of variance in F_{obs} after merging of the scaled film (F_{film}) and diffractometer (F_{diff}) data sets. For reflections common to the film and counter data sets, the mean-square difference $\langle||F_{film}| - |F_{diff}||^2\rangle$ is plotted *versus* $|F_{obs}|$. The variance $\sigma^2(F)$ of an observation is given by the function $(144 - 1.15|F_{obs}| + 0.0036|F_{obs}|^2)$ fitted to the mean-square differences by least squares. The maximum and minimum variances cited in the text were read from the curve.

Table 2. *Summary of the refinement procedure*

Round	Cycles	Residual R	Resolution (Å)	N_{obs}	No. of atoms* Protein	Dummy	H's	Solvent	Comments
1	1–3	0.336–0.316	6.0–2.0	4901	739	77	0	0	Film data only. Fixed $B = 15$ Å^2 Weights $\sigma(F_{obs}) = 20$
	4–6	0.327–0.267	6.0–1.8	6710	739	77	0	0	B's refined Model rebuilt in ΔF maps
2	7–8	0.276–0.258	6.0–1.8	6710	739	51	0	16	
	9–11	0.259–0.251	6.0–1.6	9476	739	51	0	16	
	12–14	0.261–0.254	6.0–1.3	15 374	739	51	0	16	Diffractometer data merged with film data
3	15–18	0.234–0.213	6.0–1.3	16 073	739	51	0	16	ΔF maps. H atoms added to model
4	19–22	0.213–0.210	6.0–1.3	16 073	739	12	704	44	
5	23–27	0.221–0.194	7.0–1.3	16 156	739	12	704	44	
	28–32	0.194–0.188	7.0–1.3	16 156	739	12	704	44	Weights $\sigma(F_{obs}) = 16.0$ ΔF maps
6	33–40	0.199–0.171	7.0–1.3	16 156	739	4	704	66	Variable occupancies introduced for solvent ΔF maps
7	41–50	0.173–0.164	7.0–1.33	16 111†	739	4	704	80	Outermost shell of data removed† ΔF maps
8	51–57	0.166–0.159	7.0–1.33	16 111	739	4	704	99	
		0.152	7.0–1.33	14 251	739	4	704	99	Film data reprocessed, partials added if reflections observed on adjacent films Merged with diffractometer data
9	58–60	0.152–0.149	7.0–1.33	14 251	739	4	704	99	
10	60–62	0.149–0.148	7.0–1.33	14 251	739	4	704	99	Individual weights on structure factors based on standard deviations $w(F) = 1/\sigma_A^2(F)$, $\sigma_A(F) = 1.72\sigma(F)$
11	63–72	0.154–0.144	7.0–1.33	14 251	756	4	724	99	ΔF maps. Refinements and omit maps for potentially disordered groups Residues 54, 56, 61, 64 modelled as disordered
									Further partial refinements and omit maps Parts of the following residues removed: 43, 60, 66, 71. Some H_2O deleted
									Omit maps with intermediate refinement for all solvent sites
12	73–83	0.156–0.149	7.0–1.33	14 251	775	14	747	96	7 residues in two conformations: 45, 54, 56, 59, 61, 64, 75
									Further partial refinements and omit maps
13	84–92	0.150–0.144	7.0–1.33	14 251	784	0	760	105	9 residues in two conformations: 30, 45, 54, 56, 61, 64, 66, 71, 75
									ΔF maps
14	93–102	0.148–0.146	8.0–1.33	14 303	784	0	760	103	Omit maps with intermediate refinement for all solvent positions
15	103–112	0.146–0.145	8.0–1.33	14 303	784	0	760	104	
16	113–118	0.145–0.145	8.0–1.33	14 303	784	0	760	104	$w(F) = 1/\sigma_A^2(F)$, $\sigma_A(F) = 1.75\sigma(F)$ ΔF maps
17	119–131	0.145–0.146	8.0–1.33	14 303	784	0	760	120	$w(F) = 1/\sigma_A^2(F)$, $\sigma_A(F) = 2.3\sigma(F)$ ΔF maps
18	132–141	0.152–0.149	8.0–1.33	14 303	785	0	762	110	C^β included in second conformer of Asn64
									ΔF maps. Omit maps for all solvent sites in groups and for all disordered residues
19	142–151	0.150–0.149	8.0–1.33	14 303	785	0	762	112	
20	152–161	0.150–0.149	8.0–1.33	14 307	785	0	762	112	ΔF maps Anomalous-scattering corrections for Cu and S included. 4 reflections incorrectly omitted earlier included
21	162–171	0.149–0.149	8.0–1.33	14 307	785	0	762	112	
22	172–181	0.150–0.150	8.0–1.33	14 307	785	0	762	111	ΔF maps. All H_2O contacts checked
23	182–186	0.150–0.150	8.0–1.33	14 307	785	0	762	110	ΔF maps. All H_2O contacts checked

* Protein: Number of protein atoms (non-H), including Cu and second conformer of disordered groups. Dummy: Number of non-H atoms refined with zero (0.01) occupancy. H's: Number of H atoms included in the model (see text). Solvent: Number of solvent positions included in the refinement.

† 45 reflections, being the only observed data in the range $1.33 \geq d \geq 1.30$ Å, were deleted. The data in this shell were incomplete due to cut-off at the edges of the films, and their residual R was significantly higher than that for the adjacent shell.

new solvent atoms were added to the model, the temperature factors were kept constant for several cycles while the positional parameters of all atoms were refined. The refinement of the positional parameters then continued, temperature factors (of all atoms) and occupancies (of the solvent sites alone) being included in alternate cycles.

The refinement of solvent temperature factors and occupancies in alternate cycles followed a frequently used protocol (Hendrickson, 1985). In practice, the

protocol is justified by the fact that it converges to give plausible results; and it can be argued that the occupancies of solvent sites *must* be treated as variables if the model is to represent physical reality (Jensen, 1990). In principle, the protocol is open to the objection that the temperature factor and occupancy of an atom are so highly correlated that it may be meaningless to refine both of them. Model calculations have recently been used both to support the objection (Kundrot & Richards, 1987) and to challenge it (with qualifications) (Bhat, 1989). In the present work we have assessed the results of the refinement by means of an empirical test devised by Sielecki & James (1981). In Fig. 3, the refined solvent occupancies q_i are plotted *versus* the refined temperature parameters B_i. The scatter of values in this plot, as well as the correlation coefficient (0.33), indicate that the refined variables are only weakly correlated.

To assess the extent to which the protein structure might be affected by the treatment of solvent occupancies, three additional rounds of refinement not included in Table 2 were carried out after round 18. The temperature factors and occupancies of all solvent atoms were initially reset to 25.0 Å^2 and 1.0, respectively. The occupancies were not refined. After the three rounds, the only parameters which had values significantly different from those obtained after the three comparable rounds of refinement with variable occupancies were the temperature factors of the solvent atoms. These had increased ($8 \le B \le 73$ Å^2 compared with $7 \le B \le 57$ Å^2). Increases in the unweighted and weighted residuals ($R = 0.152$ and $R_w = 0.168$ compared with 0.149 and 0.165) and r.m.s. differences between the respective sets of atomic positions (0.027 Å for the C^α atoms, 0.051 Å for all atoms) were not significant.

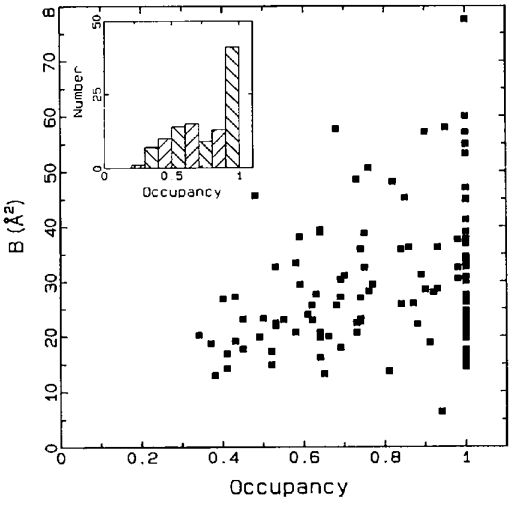

Fig. 3. Plot of refined occupancies q_i against the refined temperature factors B_i of the solvent molecules. The linear correlation coefficient is 0.33. (Inset) Histogram of occupancies.

Table 3. *Weighting parameters and r.m.s. deviations from target values in restrained least-squares refinement*

	σ	R.m.s. Δ	Worst Δ	$\Sigma w\Delta^2$
Bonding distances (Å)				
Bond length	0.030	0.016	0.052	
Angle distance	0.040	0.032	0.168	
Interplanar distance	0.050	0.036	0.162	0.13×10^4
X—H bond distance	0.100	0.028	0.153	
X—H angle distance	0.080	0.022	0.113	
Planar groups (Å)	0.020	0.011	0.042	0.28×10^3
Chiral volumes (Å^3)	0.150	0.118	0.571	0.75×10^2
Non-bonded contacts (Å)				
Single torsion	0.500	0.143	0.483	
Multiple torsion	0.500	0.199	0.657	0.69×10^2
Hydrogen bond (*X*···*Y*)	0.500	0.195	0.521	
Hydrogen bond (*X*—H···*Y*)	0.500	0.093	0.233	
Torsion angles (°)				
Peptide plane (ω) 0, 180	5.0	2.6	9.5	
Staggered (± 60°, 180°)	15.0	14.1	51.3	0.15×10^3
Orthonormal (± 90°)	22.0	12.4	24.7	
Thermal factors (Å^2)				
Main chain (1–2)	5.0	2.2	7.3	
Main chain (1–3)	7.5	3.0	8.0	
Side chain (1–2)	5.0	3.9	11.4	0.10×10^4
Side chain (1–3)	7.5	6.0	14.3	
X—H bond	10.0	2.5	5.3	
X—H angle	10.0	3.3	7.5	
Restraints against excessive shift				
Positional parameters (Å)	0.1	–	–	
Thermal parameters (Å^2)	3.0	–	–	
Diffraction data				
Structure-factor modulus	$2.3\sigma(F)$	–	–	0.13×10^5

The r.m.s. shifts in the positional and thermal parameters in the final refinement cycle of round 22 were 0.002 Å and 0.07 Å^2, respectively. The statistics of the final model are given in Table 3.*

Results and discussion

Improvements in accuracy of the structure

The most significant improvement produced by the new refinement is that the conformations of all side chains on the surface of the molecule can now be described with confidence. The conformations of some of these side chains in the original refinement were dictated by the geometrical constraints rather than by the diffraction data. The 14 residues where 'omit' maps previously revealed little or no electron density for all or part of the side chain now lie in

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1PLC, RIPLCSF), and are available in machine-readable form from the Protein Data Bank at Brookhaven. These data, together with a table of water molecule environments (hard copy), have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37064. Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

significant electron density (Glu18, Lys26, Lys30, Ser45, Lys54, Glu59, Glu60, Asp61, Asn64, Lys66, Ser75, Glu79, Lys95 and Asn99). The electron density features at seven of these residues are sufficiently well resolved to permit the modelling of the side chains as two conformers (Lys30, Ser45, Lys54, Asp61, Asn64, Lys66 and Ser75). Two other residues, which did lie in density at 1.6 Å resolution but were refined as a single conformer, are also seen as two conformers at 1.33 Å resolution (Ser56 and Glu71).† The latter two residues are included among six where significantly different side-chain orientations have been found: Glu43 (rotation of $-COO^-$ about $C^\gamma-C^\delta$), Ser56* (new orientation of $C^\beta-O^\gamma H$), Glu71* (new orientation of $C^\gamma-C^\delta OO^-$), Lys77 (new side-chain conformation), Asn76 and Gln88 (rotation of $-CONH_2$ by 180° consistent with

† The alternate conformer of a disordered side chain is identified by an asterisk elsewhere in the text and in the tables.

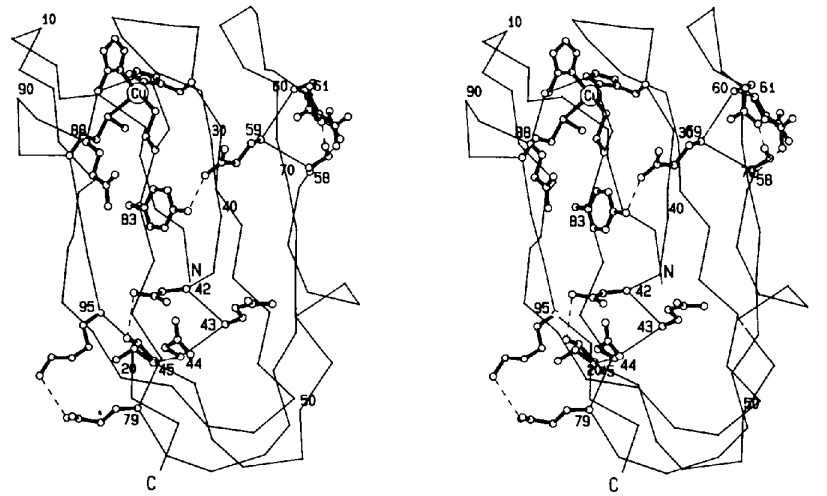

Fig. 4. The acidic patch of plastocyanin revisited. The side chains of Asp42, Glu43 (new orientation), Asp44, Ser45 (disordered), Glu59, Glu60, Asp61 (disordered) and Tyr83 are shown in relation to the molecular C^α backbone and the Cu site. The side chains of Ser58, Glu79, Gln88 and Lys95 (new orientation) are also shown. Dashed lines indicate hydrogen bonds between the side chains of Asp42 and Ser45 (new conformer), Ser58 and Glu60, Glu59 and Tyr83; a hydrogen bond between the side chain of Asp61 and the polypeptide backbone at Ser58; and a salt bridge (3.4 Å) between Glu79 and Lys95.

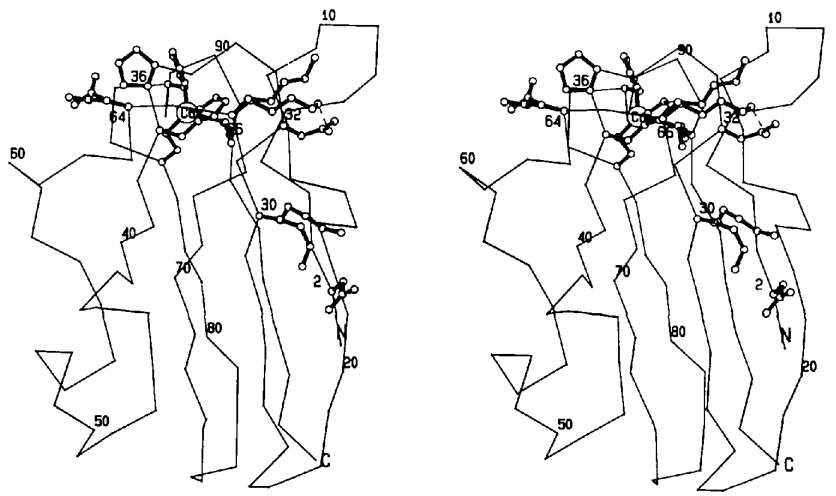

Fig. 5. New orientations of disordered side chains at Lys30, Asn64 and Lys66. The side chains of Asp2, Lys30 (disordered), Asn32, Pro36, Asn64 (disordered) and Lys66 (disordered) are shown in relation to the molecular C^α backbone and the Cu site. The orientations of Asp2 and Lys30 suggest the presence of a salt bridge, but the distance of closest approach (4.1 Å) corresponds to only a weak interaction. The dashed line indicates a hydrogen bond between the side chains of Asn32 and one conformer of Lys66. There is a tight contact (3.0 Å) between the side-chain amide group of Asn64 and the C^γ atom of Pro36.

hydrogen bonding). The new side-chain orientations are illustrated in Figs. 4–7.

The more detailed description of the surface side chains has led to a better identification of intramolecular hydrogen bonds, intermolecular hydrogen bonds and protein–solvent contacts. These are discussed in later sections. Side chains which were previously excluded from the discussion because their position and orientation were not known with certainty can now be included.

The present work has also confirmed unequivocally that residue 39 in the protein crystallized by us is correctly characterized as Ile. An inconsistency with the NMR assignment of residue 39 as Val (King & Wright, 1986) may be due to microheterogeneity; in a second form of poplar plastocyanin isolated from some trees, Ile39 is indeed replaced by Val39 (Dimitrov, Egorov, Donchev & Atanasov, 1987).

In contrast with the greatly improved description of the side chains, the polypeptide backbone has been left relatively unchanged by the refinement at 1.33 Å resolution (Fig. 8). The r.m.s. difference between the positions of the C^{α} atoms at 1.6 and 1.33 Å resolutions is 0.11 Å (without least-squares

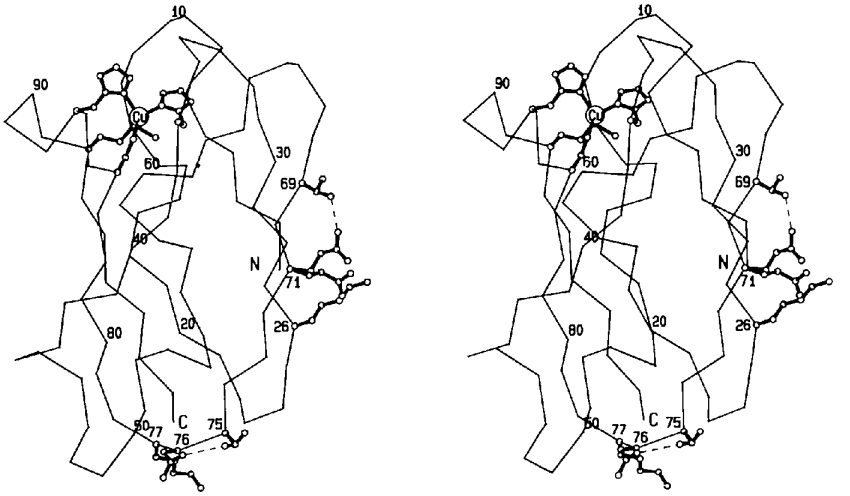

Fig. 6. New orientations of side chains at Lys26, Glu71, Ser75, Asn76 and Lys77. The side chains of Lys26 (new orientation), Thr69, Glu71 (two new conformers), Ser75 (disordered), Asn76 (side-chain amide group rotated by 180° from earlier structure) and Lys77 (new orientation) are shown in relation to the molecular C^{α} backbone and the Cu site. The dashed lines indicate hydrogen bonds between the side chains of Thr69 and one conformer of Glu71, and between the side chains of Asn76 and one conformer of Ser75.

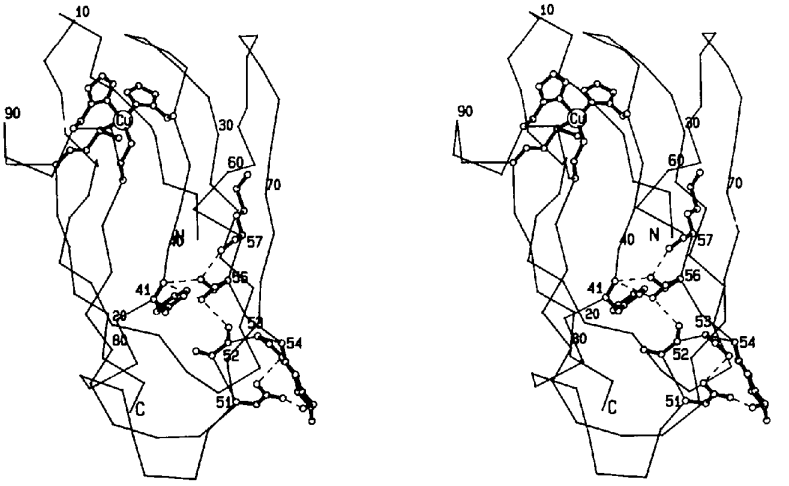

Fig. 7. New orientations of side chains at Lys54 and Ser56. The side chains of Phe41, Asp51, Ala52, Ser53, Lys54 (disordered), Ser56 (disordered) and Met57 are shown in relation to the molecular backbone and the Cu site. Dashed lines show hydrogen bonds between the side chain of Asp51 and the side chains of Ser53 and Lys54, and between the two side-chain conformers of Ser56 and the backbone peptide groups at Phe41, Ala52 and Met57.

minimization); the largest difference is 0.28 Å (at C^{α} Asn64). When all the polypeptide backbone atoms are included in the calculation, the r.m.s. difference is only 0.10 Å and the largest difference is 0.36 Å (at O Ser22). Changes greater than 0.20 Å (*i.e.* twice the r.m.s. change) have occurred at residues 13, 22, 24, 49, 51, 58, 61–64 and 71. Six of these residues (from 51 to 64) lie along the irregular fifth (non-β) strand of the polypeptide backbone. Least-squares fitting of the superposed atomic coordinates does not reduce the r.m.s. differences significantly.

Marked changes in the atomic temperature parameters B have occurred almost exclusively at the side chains (Fig. 9). Changes in the B's of polypeptide backbone atoms are small and not correlated with structural features. The majority of the B's of side-chain atoms have decreased, some by as much as 40 Å^2. The residues with the largest decreases include most of those with exposed side chains which were previously not located in density, plus several which were mentioned above as being modelled significantly better at 1.33 Å than at 1.6 Å resolution (Glu43, Ser56, Glu71). Only one very large decrease in B ($\geq$ 10 Å^2) has occurred at a residue which was previously well defined (Phe19). The B's of the Cu and side-chain atoms at the Cu site are discussed later.

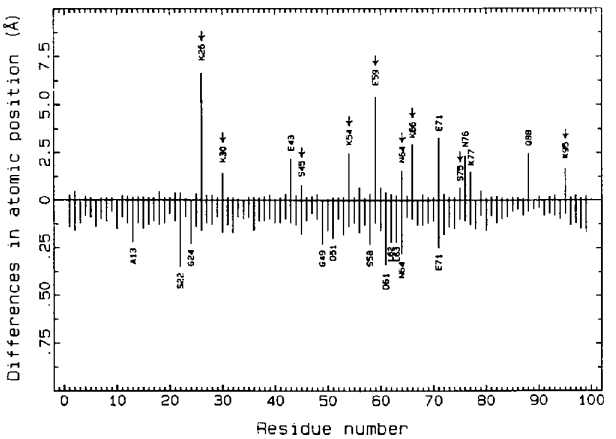

Fig. 8. Differences between corresponding atomic positions in the 1.3 and 1.6 Å models, plotted as a function of residue number. For each residue, the line *above* the horizontal axis represents the *maximum* shift of any side-chain atom; the line *below* the horizontal axis represents the *maximum* shift of any backbone atom (N, C^{α}, C, O). Arrows mark residues where the positions of side-chain atoms were previously poorly determined but are now unequivocally located in electron density. For disordered side chains, the shift is plotted for that conformer which differs least from the side chain in the 1.6 Å model. Substantial shifts of side-chain atoms have occurred at only five residues (see text: 43, 71, 76, 77 and 88). The large shifts at eight other residues (26, 30, 45, 54, 59, 64, 66, 95) are apparent rather than real, because the side chains are among the 14 for which there was previously no electron density. The changes at the remaining five side chains for which there was previously no electron density are small [18, 60, 61 (one conformer), 79, 99].

Hydrogen bonds and salt bridges

A revised assignment of the intra- and intermolecular hydrogen bonds and salt bridges is given in Tables 4 and 5.

There is no clear-cut definition of a hydrogen bond, and the decision whether a particular contact should be labelled as a hydrogen bond usually involves the application of subjectively chosen criteria. Both geometrical (Baker & Hubbard, 1984) and energetic (Kabsch & Sander, 1983) criteria have been proposed. In the present work we have adopted an even more conservative geometrical approach than in our earlier papers. While the cut-off for donor–acceptor distances has been slightly increased (from 3.2 to 3.3 Å), a number of additional criteria have been introduced: plausible values are now required for bond angles at donor and acceptor atoms, and for the H···acceptor distance and the N—H···O angle in all contacts involving amide groups. Details are given in the footnotes to Table 4. A small number of interactions which have acceptable donor–acceptor distances but fail to satisfy one or another of the more restrictive criteria are tagged with superscripts in Tables 4 and 5. No distinction has been made between hydrogen bonds and salt

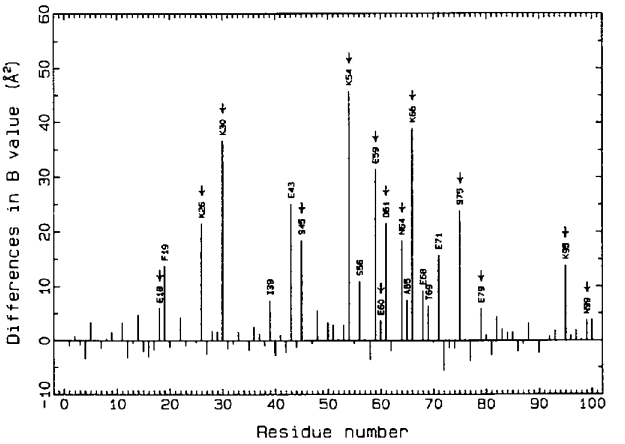

Fig. 9. Changes in atomic temperature parameters B plotted as a function of residue number. At each residue, the line above/below the horizontal axis represents the *maximum* difference ($B_{1.6} - B_{1.3}$) between the temperature parameters B of a side-chain atom in the models at 1.3 and 1.6 Å resolution. Labels indicate the residues with the largest decreases in B. They include (i) residues where the positions of exposed side-chain atoms were previously poorly determined but are now unequivocally located in electron density, (ii) residues where the description of the side-chain conformation has changed substantially (Glu43, Ser56, Glu71), and (iii) one residue which was previously well defined and where there is no obvious reason for a decrease $\geq$ 10 Å^2 in B (Phe19).

Table 4. *Intramolecular hydrogen bonds and salt bridges in poplar* Cu^{II}*-plastocyanin: differences between the models at* 1.33 *and* 1.6 Å *resolution*

(*a*) Hydrogen bonds between polypeptide backbone groups

Peptide N	Peptide O	Distance (Å) in 1.33 Å structure	Distance (Å) in 1.6 Å structure
1	26	2.7	2.8
3	28	2.9	2.9
4	15	2.9	2.8
5	30	2.8	2.8
6	13	2.9	2.9
7	11	2.9	2.8
10	7	2.9	2.9
13	11	3.1[a,c]	–[e,f]
15	4	3.0	2.9
19	95	3.3[a]	3.3
21	97	2.9	3.0
24	74	2.8	2.8
25	22	3.0	3.3
27	72	2.9	2.7
28	1	2.7	2.7
29	70	2.8	2.9
30	3	2.7	2.8
31	68	2.8	2.7
32	5	2.8	2.8
33	5	3.1	3.0
37	63	3.0	3.0
40	83	2.9	2.9
42	81	3.0	3.1
45	42	2.9	2.9
50	47	3.2	3.1
54	51	3.1	3.2
55	52	3.3[a]	3.1
56	52	3.0[a]	3.0
57	39	3.0	2.9
61	58	3.1	3.2
63	37	2.8	2.8
67	31	2.9	3.0
68	65	3.2	3.1
70	29	3.0	3.0
72	27	3.0	3.0
78	98	3.0	2.9
80	96	2.8	2.8
81	45	3.0	2.9
82	94	3.0	3.0
83	40	2.8	2.8
84	92	3.0	3.1
87	84	3.2[a]	3.1
88	84	2.9[a,b]	3.0
90	87	3.2[a]	3.2
91	88	2.9	2.9
92	87	3.1	3.2
93	91	3.0	–[e,f]
94	82	3.0	2.9
95	17	2.9	2.9
96	80	2.9	3.0
97	19	3.0	3.0
98	78	2.9	2.9
99	21	3.0	3.1

(*b*) Hydrogen bonds involving side-chain atoms

Atom *X* (donor)	Atom *Y* (acceptor)	Distance (Å) in 1.33 Å structure	Distance (Å) in 1.6 Å structure
11 N	9 $O^{\delta1}$	3.1	2.9
11 O^{γ}	9 $O^{\delta1}$	2.8	3.0
22 N	25 $O^{\epsilon2}$	3.2	3.4[e]
31 $N^{\delta2}$	63 O	3.0[c]	3.0
31 $N^{\delta2}$	65 O	3.0	3.1
32 $N^{\delta2}$	6 O	2.8	2.9
32 $N^{\delta2}$	8 $O^{\delta1}$	2.8	2.7
37 $N^{\epsilon2}$	33 O	2.7	2.6
38 N	84 S^{γ}	3.5	3.4
38 $N^{\delta2}$	61 O	2.9	2.9
38 $N^{\delta2}$	85 O^{γ}	3.0	3.0

Table 4 (*cont.*)

Atom *X* (donor)	Atom *Y* (acceptor)	Distance (Å) in 1.33 Å structure	Distance (Å) in 1.6 Å structure
41 N	56 O^{γ}	3.0	3.3[e]
41 N	56* O^{γ}	2.9	–
44 N	42 $O^{\delta1}$	3.0	2.9
45* O^{γ}	42 $O^{\delta2}$	3.0[c]	–
53 N	51 $O^{\delta2}$	2.9	3.1
53 O^{γ}	51 $O^{\delta2}$	2.7	2.8
54* N^{ζ}	51 $O^{\delta1}$	2.2	–
56 O^{γ}	52 O	2.9[c]	2.8
56* O^{γ}	57 O	3.1	–
58 N	61 $O^{\delta2}$	2.5	2.8
58 O^{γ}	60 $O^{\epsilon2}$	2.4	2.7
60 N	58 O^{γ}	3.3[a]	3.2
64 N	68 $O^{\epsilon2}$	2.5	2.8
65 N	68 $O^{\epsilon2}$	2.7	2.8
66 N^{ζ}	32 $O^{\delta1}$	3.0	> 3.5[e]
66 N^{ζ}	32 O	3.0	> 3.5[e]
69 $O^{\gamma1}$	71 $O^{\epsilon2}$	2.7	> 3.5[e]
75* O^{γ}	76 $O^{\delta1}$	3.3	–
80 O^{η}	76 O	2.6	2.6
83 O^{η}	59 $O^{\epsilon2}$	2.8	> 3.5[e]
85 N	38 $O^{\delta1}$	2.8	2.9
85 O^{γ}	59 O	3.2	3.3[e]
99 $O^{\delta1}$	21 O	3.1[d]	3.0[d]

The following criteria have been used to define hydrogen bonds. Backbone and side-chain amide groups: $d_{N\cdots O} \le 3.3$ Å, $d_{H\cdots O} \le 2.3$ Å, $\angle N—H\cdots O \ge 120°$, $\angle C—N\cdots O$ and $\angle C—O\cdots N \ge 90°$. Other functional groups *X*, *Y*: $d_{X\cdots Y} \le 3.3$ Å, $\angle C—X\cdots Y$ and $\angle C—Y\cdots X \ge 90°$.

Superscripts *a–c* denote contacts determined at 1.33 Å resolution which fail to meet one of the above criteria: (*a*) $d_{H\cdots O} > 2.3$ Å, (*b*) $\angle N—H\cdots O < 120°$, (*c*) $85 < \angle C—X\cdots Y < 90°$.

Superscripts *d–f* have the following significance: (*d*) Rotation of the Asn99 side-chain amide group by 180° about $C^{\beta}—C^{\gamma}$ permits the formation of a hydrogen bond Asn99 $N^{\delta2}\cdots$Ile21 O. If the amide group is *not* rotated, it can form an intermolecular hydrogen bond Asn99 $N^{\delta2}\cdots$Glu60^{xii} O (Table 5). The two bonds cannot be present simultaneously. (*e*) Contact was not identified as a hydrogen bond at 1.6 Å resolution (Guss & Freeman, 1983). (*f*) Contact was identified as a hydrogen bond in a γ-bend (1–3 bend) by Baker & Hubbard (1984).

* Alternate conformer of a disordered residue.

bridges in the case of side-chain $—NH_3^+\cdots{}^-OOC—$ contacts. A cut-off distance of 3.3 Å for salt bridges has the virtue of simplicity but lies well within the probable range of electrostatic interactions (Rashin & Honig, 1984).

Intramolecular hydrogen bonds

There have been no chemically significant changes in the polypeptide backbone conformation as a result of the refinement at 1.33 Å resolution. Table 4(*a*) lists the same peptide–peptide hydrogen bonds as Fig. 5 of Guss & Freeman (1983), with the addition of two contacts identified as '1–3' or 'γ' hydrogen bonds by Baker & Hubbard (1984). However, six of the peptide–peptide interactions violate at least one of the more rigorous geometrical criteria mentioned above. Three of these six interactions involve residues 84, 87, 88 and 90 on the important double loop which forms one side of the Cu site. The *positions* of

Table 5. *Intermolecular hydrogen bonds in poplar* Cu^{II}*-plastocyanin crystals at* 1.33 Å *resolution*

All intermolecular N—H···O, O—H···O and $N^+ \cdots O^-$ contacts $d_{X\cdots Y} \leq 3.3$ Å are listed. There are four non-bonded intermolecular contacts ≤ 3.0 Å: Ser58 $C^\beta \cdots$Asp61 $O^{\delta 2}$, 2.9; Glu71 $C^\delta \cdots$Asp44^{vi} $O^{\delta 2}$, 3.0; Lys*54 $C^\epsilon \cdots$Gln88^{vii} $O^{\epsilon 1}$, 2.9; 188···Lys26^x C^δ, 2.7 Å.

Contact			Reverse contact			$d_{X\cdots Y}$ (Å)
Asp8	O	···Ser17^v O^γ	Ser17	O^γ	···Asp8^{ii} O	2.9
Ser20	O^γ	···Asp61^{xi} $O^{\delta 1}$	Asp61	$O^{\delta 1}$	···Ser20^{xi} O^γ	2.6
Gly24	O	···Asn64^i $N^{\delta 2}$	Asn64	$N^{\delta 2}$	···Gly24^x O	3.2
Asp44	$O^{\delta 1}$	···Thr69^{viii} $O^{\gamma 1}$	Thr69	$O^{\gamma 1}$	···Asp44^{vi} $O^{\delta 1}$	3.1
Asp44	$O^{\delta 2}$	···Glu71^{viii} $O^{\epsilon 1}$	Glu71	$O^{\epsilon 1}$	···Asp44^{vi} $O^{\delta 2}$	3.2
Asp44	$O^{\delta 2}$	···Glu71^{viii} $O^{\epsilon 2}$	Glu71	$O^{\epsilon 2}$	···Asp44^{vi} $O^{\delta 2}$	3.0
Asp44	$O^{\delta 2}$	···Glu*71^{viii} $O^{\epsilon 2}$	Glu*71	$O^{\epsilon 2}$	···Asp44^{vi} $O^{\delta 2}$	3.2
Ser48	O^γ	···Asn99^{iv} $O\gamma$	Asn99	$O\gamma$	···Ser48^{iii} O^γ	3.1
Gly49	O	···Ser85^{vi} O^γ	Ser85	O^γ	···Gly49^{ix} O	2.7
Asp51	$O^{\delta 1}$	···Gly89^{vii} N	Gly89	N	···Asp51^{ix} $O^{\delta 1}$	2.9
Lys*54	N^ζ	···Gln88^{vii} $O^{\epsilon 1}$	Gln88	$O^{\epsilon 1}$	···Lys*54^{ix} N^ζ	2.9
Glu60	O	···Asn99^{xi} $N^{\delta 2}$	Asn99	$N^{\delta 2}$	···Glu60^{xii} O	2.7

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

* Alternate conformer of a disordered residue.

the peptide groups are not in doubt but the network of peptide–peptide hydrogen bonds stabilizing this part of the molecule may not be as extensive as in our original description. On the other hand, the same interactions have been found in a structure analysis of French bean plastocyanin by two-dimensional NMR spectroscopy (Moore, Lepre, Gippert, Chazin, Case & Wright, 1991). The NMR analysis is in fact in striking overall agreement with the present work. Only at Gln88 is there a small difference in interpretation due to the use of different distance criteria. We identify the interaction N Gln88···O Cys84 as a hydrogen bond, with qualifications (Table 4). In the NMR structure analysis, the interaction described as a hydrogen bond is N Gln88···O Ser85. The relevant interatomic distances in the crystal structure of poplar plastocyanin ($N_{88}\cdots O_{85} = 3.6$ Å, $H\cdots O_{85} = 2.7$ Å) lie significantly outside the range which we have adopted for hydrogen bonds.

Since there have been substantial changes in the positions and orientations of side-chain functional groups, it is not surprising that the intramolecular backbone/side-chain and side-chain/side-chain hydrogen bonds listed in Table 4(*b*) include 12 more contacts than were shown in Fig. 6 of Guss & Freeman (1983). [Four contacts in Table 4(*b*) violate one of the proposed additional geometrical criteria.] Two of the newly identified hydrogen bonds involve residues at bends linking strands of the protein backbone (N Ser22···$O^{\epsilon 2}$ Glu25, O^γ Ser75*···$O^{\delta 1}$ Asn76); one lies along β-strand 6 (O^γ Thr69···$O^{\epsilon 2}$ Glu71); two lie along the non-β strand 5 of the backbone (N^ζ Lys54*···$O^{\delta 1}$ Asp51, O^γ Ser56*···O Met57); two involve side chains of residues which contribute to the acidic patch (O^γ Ser45···$O^{\delta 2}$ Asp42, O^η Tyr83···$O^{\epsilon 2}$ Glu59); the last of these and four others are interactions between residues in adjacent strands of the backbone (N Phe41···O^γ Ser56*, N^ζ Lys66···$O^{\delta 1}$ Asn32, N^ζ Lys66···O Asn32, O^γ Ser85···O Glu59).

While all the hydrogen bonds listed in Table 4(*b*) must contribute to the stabilization of the conformation of the poplar plastocyanin molecule, only those which involve the side chains of invariant residues are likely to be of global significance in describing plastocycin molecules in general. Conversely, the fact that these residues contribute to the stability of the molecular conformation may be one reason why they are invariant. We summarize here the interactions where side chains of invariant residues appear to stabilize important features of the molecular structure. Some of these interactions were previously described by Guss & Freeman (1983); they are repeated here in the light of more recent evidence concerning the residues which are to be regarded as invariant (Sykes, 1990).

The residues in Table 4(*b*) which are currently regarded as being invariant in the plastocyanins of higher plants are Glu25, Asn31, Asn32, His37, Asn38, Asp42, Lys54, Ser56, Glu68, Tyr80, Tyr83 and Cys84. Several of these residues are part of a network of hydrogen bonds which define not only the geometry of the Cu site itself but also the relationship of the Cu site to a probable electron-transfer pathway. The orientation of the imidazole ring of His37, one of the Cu-binding residues, is stabilized by a hydrogen bond to the backbone amide group of residue 33. The residue adjacent to His37 is Asn38 which appears to play a very important role in defining the geometry of the Cu site. The side-chain amide group of Asn38 forms a N—H···O hydrogen bond with the backbone O(amide) atom of residue 61, and an O···H—N hydrogen bond with the backbone N(amide) atom of residue 85. The significance of the first interaction is that residue 61 is located at the northern kink in the non-β-strand 5, is acidic in all higher plant plastocyanins, and is part of the northern lobe of the acidic patch which is likely to provide electrostatic guidance for the interaction with biological redox partners. The significance of the second interaction is that residue 85 is adjacent to Cys84, the Cu-binding residue which is almost certainly part of an electron-transfer pathway to and from the Cu site. Thus the side chain of Asn38 is involved in interactions which help to define the relationship between two of the Cu-binding residues as well as the relationship between the Cu site and part of the putative electrostatic recognition patch. Further stabilization of the position of Cys84 with respect to His37 is provided by an N—H···S hydrogen bond from the backbone N(amide) atom of

Asn38 to the Cu-binding S(thiolate) atom of Cys84. An important feature of this N—H···S bond, which has also been found in all other structurally characterized cupredoxins, is that the donor N(peptide) atom is remote from Cys84 in the primary structure and belongs to a different strand of the polypeptide in the tertiary structure. It is to be noted that in ferredoxins, where N(peptide)—H···S(Cys) hydrogen bonds likewise play a prominent role, the donor N(peptide) is characteristically only two or three residues from the acceptor Cys (Adman, Watenpaugh & Jensen, 1975; Backes, Mino, Loehr, Meyer, Cusanovich, Sweeney, Adman & Sanders-Loehr, 1991).

The finding that there is a hydrogen bond between the side chains of Tyr83 and Glu59 is interesting in view of evidence that the invariant residue Tyr83 is directly involved in electron transfer to plastocyanin from its biological redox partner, cytochrome *f* (Beoku-Betts, Chapman, Knox & Sykes, 1985). The hypothesis that electron transfer takes place in a collision complex in which the redox partner makes contact with plastocyanin at Tyr83 is supported by electrostatic interaction energy calculations using cytochrome *c* to represent cytochrome *f* (Roberts, Freeman, Olson, Tainer & Getzoff, 1991). It is possible that the hydrogen bond to Glu59, which is highly (but not totally) conserved in the plastocyanins of higher plants, stabilizes the side chain of Tyr83 in a conformation which leads to a favourable contact with the biological redox partner.

Hydrogen bonds at side chains of other conserved residues can be assigned functions not directly related to the Cu site. The newly found interaction between the side chain of Glu25 and the backbone at residue 22 clearly stabilizes the loop between β-strands 2 and 3. Three hydrogen bonds involving the side chains of Asn31 and Asn32 link β-strand 3 to β-strands 1 and 6. The side chain of Ser56 in the non-β-strand 5 interacts not only with two backbone O(amide) atoms in the same strand but also with the backbone N(amide) atom of the invariant residue Phe41 in the adjacent β-strand 4. Glu68 at the northern end of β-strand 6 forms hydrogen bonds to the backbone at residues 64 and 65, stabilizing the loop which links strand 6 to strand 5. The phenolic group of Tyr80 is hydrogen bonded to the backbone at residue 76, an interaction whose significance is still not understood any better than when it was first described by Guss & Freeman (1983).

Dimensions and thermal parameters of the Cu *site*

The new dimensions of the Cu site in Cu^{II}-plastocyanin are shown in Table 6. The standard deviations of the Cu–ligand bond lengths and the ligand–Cu–ligand bond angles are conservatively estimated as 0.04 Å and 2.5°, respectively (see *Estimates of accuracy and precision*, below).

Table 6. *Dimensions of the copper site in poplar* Cu^{II}*-plastocyanin*

	Refined at 1.6 Å resolution*	Refined at 1.33 Å resolution*
Cu–ligand bond lengths (Å)		
Cu—N(His37)	2.04	1.91
Cu—N(His87)	2.10	2.06
Cu—S(Cys84)	2.13	2.07
Cu—S(Met92)	2.90	2.82
Bond angles at Cu atom (°)		
N(His37)—Cu—N(His87)	97	97
—S(Cys84)	132	132
—S(Met92)	85	88
N(His87)—Cu—S(Cys84)	123	121
—S(Met92)	103	101
S(Cys84)—Cu—S(Met92)	108	110
Temperature factors *B* of Cu atom and Cu-binding side chains (averaged) (Å^2)		
Cu	13	9
His37	9	9
His87	8	11
Cys84	9	8
Met92	7	7
Torsion angles involving Cu-binding side chains (°)		
His37		
S^{γ}(Cys84)—Cu—N^{δ}(His37)—C^{γ}	−5	4
N^{δ}(His87)—Cu—N^{δ}(His37)—C^{γ}	143	148
S^{δ}(Met92)—Cu—N^{δ}(His37)—C^{γ}	−115	−112
Cu—N^{δ}(His37)—C^{γ}—C^{β}	−6	−15
N^{δ}(His37)—C^{γ}—C^{β}—C^{α} (χ_2)	−62	−60
C^{γ}(His37)—C^{β}—C^{α}—N (χ_1)	−67	−65
Cys84		
N^{δ}(His37)—Cu—S^{γ}(Cys84)—C^{β}	−105	−110
S^{δ}(Met92)—Cu—S^{γ}(Cys84)—C^{β}	−5	−3
N^{δ}(His87)—Cu—S^{γ}(Cys84)—C^{β}	115	113
Cu—S^{γ}(Cys84)—C^{β}—C^{α}	−166	−168
S^{γ}(Cys84)—C^{β}—C^{α}—N (χ_1)	170	169
His87		
N^{δ}(His37)—Cu—N^{δ}(His87)—C^{γ}	159	167
S^{γ}(Cys84)—Cu—N^{δ}(His87)—C^{γ}	−49	−44
S^{δ}(Met92)—Cu—N^{δ}(His87)—C^{γ}	72	77
Cu—N^{δ}(His87)—C^{γ}—C^{β}	−21	−25
N^{δ}(His87)—C^{γ}—C^{β}—C^{α} (χ_2)	138	143
C^{γ}(His87)—C^{β}—C^{α}—N (χ_1)	−62	−61
Met92		
N^{δ}(His37)—Cu—S^{δ}(Met92)—C^{γ}	157	161
S^{γ}(Cys84)—Cu—S^{δ}(Met92)—C^{γ}	24	27
N^{δ}(His87)—Cu—S^{δ}(Met92)—C^{γ}	−107	−102
Cu—S^{δ}(Met92)—C^{γ}—C^{β}	72	69
C^{ε}—S^{δ}(Met92)—C^{γ}—C^{β} (χ_3)	172	−168
S^{δ}(Met92)—C^{γ}—C^{β}—C^{α} (χ_2)	165	165
C^{γ}(Met92)—C^{β}—C^{α}—N (χ_1)	−168	173

* E.s.d.'s of bond lengths and bond angles, respectively, at the Cu atom: 0.05 Å, 3° at 1.6 Å resolution (Guss & Freeman, 1983); 0.04 Å, 2.5° at 1.33 Å resolution (see text).

From a chemical viewpoint, the most important findings are (i) that the two Cu—N(His) bond lengths, 1.91 (4) and 2.06 (4) Å, are significantly different, and (ii) that the revised Cu—S(Met92) bond length, 2.82 (4) Å, still falls between the values in

A. denitrificans azurin, 3.11 Å (Baker, 1988), and in the cucumber basic protein 'CBP', 2.62 Å (Guss, Merritt, Phizackerley, Hedman, Hodgson & Freeman, in preparation). The Cu–ligand bond lengths are in excellent agreement with the results of EXAFS measurements on oriented single crystals of plastocyanin at pH 6.0 (Scott, Hahn, Doniach, Freeman & Hodgson, 1982) and on solutions of plastocyanin at pH 5.5 (Murphy, Hasnain, Strange, Harvey & Ingledew, 1990) (Table 7). Agreement with the results of EXAFS measurements at pH 8.0 (Murphy *et al.*, 1990) is less convincing (Table 7). The coordination geometry is as previously reported, the bond angles at the Cu atom having changed by less than 3° (the e.s.d. at 1.6 Å resolution).

The B of the Cu atom has decreased from 13 to 9 Å^2 (Table 6 and Fig. 9). The change in B_{Cu} is consistent with the inclusion of an anomalous-scattering correction in the scattering curve for Cu: in a trial calculation without the anomalous scattering correction for Cu, B_{Cu} decreased only to 11.5 Å^2. The use of an anomalous-scattering correction for S, on the other hand, had no significant effect on B_S. The B's of the side chains at the Cu site have changed neither dramatically nor systematically, and trial calculations have revealed no correlation between the changes and the use of anomalous-scattering corrections for Cu or S.

The inclusion of side-chain torsion angles in Table 6 reflects recent interest in the detailed mechanism of electron transfer to and from the Cu site. The three torsion angles along the sequence of bonds from the Cu atom to N Cys84 are all close to 180°. Similar conformations occur in other blue copper proteins, and it has been suggested that an approximately coplanar Cu—S^{γ}—C^{β}—C^{α}—N bond sequence is required for a favourable electron-transfer pathway (Han, Loehr, Freeman, Codd, Huq, Beppu, Adman & Sanders-Loehr, 1991). It is to be noted that several other torsion angles in Table 6 are close to 0 to 180°.

Solvent structure

Generally improved detail is found in the solvent structure, a total of 110 positions being assigned to solvent atoms. There is appreciable disorder: only 40 (36%) of the solvent sites have occupancies in the range 0.95–1.0, while 15 (14%) have occupancies ≤0.5 (see histogram, Fig. 3).

If the 110 solvent molecules identified in this analysis are assumed to be water, then the solvent structure comprises 18 water molecules which are hydrogen bonded to protein side chains and/or amide groups, 76 water molecules which are hydrogen bonded to protein groups as well as to other water molecules, and 16 water molecules which are hydrogen bonded only to other water molecules.

Table 7. *Comparison between* Cu–*ligand bond lengths* (Å) *determined by X-ray diffraction at* 1.33 Å *resolution and by extended X-ray absorption fine structure* (*EXAFS*)

	XRD pH 6 (*a*)	EXAFS Crystal pH 6 (*b*)	EXAFS Solution pH 5.5 (*c*)	EXAFS Solution pH 8.0 (*c*)
Cu—N	1.91 (4)	–	1.940	1.872
Cu—N	2.06 (4)	–	2.024	1.999
Cu—N (mean)	1.98	1.98	1.98	1.94
Cu—S	2.07 (4)	2.08	2.090	2.117
Cu—S′	2.82 (4)	–	–	2.857

References: (*a*) Present work. (*b*) Scott, Hahn, Doniach, Freeman & Hodgson (1982). (*c*) Murphy, Hasnain, Strange, Harvey & Ingledew (1991).

Among the 94 water molecules which interact with the protein, 22 are hydrogen bonded only to side chains, 40 only to backbone amide groups, and 32 both to side-chain and amide groups. The distribution of water molecules hydrogen bonded to surface atoms of the protein is uneven, since solvent is obviously excluded from those regions of the molecular surface which make direct contact with the surface of another molecule. Details of the environments of the water molecules have been deposited as supplementary material.*

The solvent structure necessarily reflects the criteria used to identify an electron density peak as a solvent atom during the refinement, *i.e.* (i) electron density above a threshold value, and (ii) contacts ≤3.3 Å with one or more plausible hydrogen-bonding partners. No attempt was made to distinguish between molecules of solvent and ions of the ammonium sulfate electrolyte or the sodium phosphate buffer. Thirty-four of the 110 solvent positions are within 0.5 Å, and seven others within 1.0 Å, of a solvent atom in the 1.6 Å structure. There are no equivalents for three of the original 44 solvent atoms.

A more detailed description of the solvent structure and solvent–protein interactions would be inappropriate since a recent study of the same crystals at 173 K and 1.6 Å resolution has revealed many additional solvent sites (Fields, Bartsch, Bartunik,

* See deposition footnote. We draw attention to a movement of O_w 187 from its position in the structure at 1.6 Å resolution. In the earlier refinement the solvent atom equivalent to O_w 187 (labelled O_w 130) was within hydrogen-bonding distances of O Gly34 and $N^{\varepsilon 2}$ His87. The latter contact is now too long to be classified as a hydrogen bond (3.58 Å) (see supplementary material). Since this $O_w \cdots N^{\varepsilon 2}$ His87 hydrogen bond was used as part of the evidence for an imidazole 'ring flip' in *apo*-plastocyanin (Garrett, Clingeleffer, Guss, Rogers & Freeman, 1984), it is important to note that the description of the *apo*-protein structure remains intact. In the *apo*-plastocyanin structure a solvent atom is within hydrogen-bonding distance of the exposed δ atom of the imidazole ring, implying that the δ atom is not $C^{\delta 1}$ His87 (as in Cu^{II}-plastocyanin) but $N^{\delta 1}$ His87. In Cu^{II}-plastocyanin the orientation of the His87 imidazole ring, with the $C^{\delta 1}$—$N^{\varepsilon 2}$ ring edge exposed to the solvent, remains as before since $N^{\delta 1}$ is still unequivocally bonded to the Cu atom.

Guss & Freeman, in preparation). The 158 unambiguously assigned solvent sites in that analysis include 70 of the present 110 solvent sites. The differences almost certainly reflect rearrangements that produce a more ordered solvent structure at low temperature.

An unexpected feature of the low-temperature structure which was not recognized in the present analysis is the presence of two internal (*i.e.* buried) water molecules. Significant peaks [$> 3\sigma(\rho)$] did appear at the relevant positions in the final room-temperature electron density maps, but were rejected as possible solvent atoms on the grounds that their contacts with one or more neighbouring atoms were uncomfortably short (peak 1: $C^{\delta 1}$ 21 at 3.16 Å, $C^{\gamma 1}$ 27 at 3.28 Å; peak 2, $C^{\delta 1}$ 21 at 3.03 Å, O 25 at 2.21 Å). In the light of the low-temperature structure it is likely that the internal water sites are at least partly occupied at room temperature. The two internal water molecules fill the gap in a classic β-bulge (Richardson, Getzoff & Richardson, 1978) between two anti-parallel β-strands of the polypeptide backbone (β-strand 3 at residues 25–27 and β-strand 6 at residues 72–74). A detailed description will be given elsewhere (Fields, Bartsch, Bartunik, Guss & Freeman, in preparation).

Intermolecular contacts

As a result of the more accurate description of the surface side chains, the list of direct intermolecular interactions (Table 5) differs significantly from the list at 1.6 Å resolution [Table 9 of Guss & Freeman (1983)]. Twelve pairs of intermolecular hydrogen bonds between the 'origin' molecule at x, y, z and six pairs of symmetry-related molecules can be assigned with confidence. One pair of hydrogen bonds identified at 1.6 Å resolution (between Glu60 and Asn76) is deleted. One pair of hydrogen bonds previously marked as doubtful due to uncertainties in the orientation of a side chain (between Ser20 and Glu60) is confirmed. A pair of close contacts between the carboxylate groups of Asp44 and Glu71, previously regarded as uncertain due to possible errors in position, is confirmed and shown to be more extensive than in the original description (see below). Two new pairs of hydrogen bonds are identified (between Ser48 and Asn99, and between Lys54* and Gln88, respectively). The C-terminal residue Asn99 is unique in making contacts with two different symmetry-related molecules, one at the side chain and the other at the free carboxylate group.

Two of the interactions included in Table 5 and not described earlier involve the side chain of Asp44. In fact, Asp44 $O^{\delta 2}$ has close contacts not only with Glu71 $O^{\epsilon 1}$ but also with Glu71 $O^{\epsilon 2}$ and Glu71* $O^{\epsilon 2}$ in a symmetry-related molecule. These contacts clearly imply intermolecular carboxyl–carboxylate hydrogen bonds (—COOH$\cdots^{-}$OOC—) of the type discussed by Sawyer & James (1982). Since the crystals were maintained in a medium at pH 6, the pK_a of the hydrogen-bonded carboxyl groups must be significantly higher than the normal value. A similar but *intra*molecular —COOH$\cdots^{-}$COO— interaction occurs in *E. prolifera* plastocyanin (Collyer, Guss, Sugimura, Yoshizaki & Freeman, 1990).

In summary, the residues involved in direct contacts with symmetry-related molecules are Asp8, Ser17, Ser20, Gly24, Asp44, Ser48, Gly49, Asp51, Lys54*, Glu60, Asp61, Asn64, Thr69, Glu71, Glu71*, Ser85, Gln88, Gly89 and Asn99. At residues 8, 24, 49, 60 and 89 a polypeptide backbone atom makes the intermolecular contact. At the remaining residues the side chain is involved.

In addition to the direct intermolecular contacts there are 60 intermolecular contacts *via* bridging solvent molecules, *i.e.* contacts in which a solvent molecule is hydrogen bonded to functional groups on the origin protein molecule at x, y, z and on one of the 12 adjacent protein molecules in the crystal (Table 8). The number of hydrogen bonds between bridging solvent molecules and functional groups on the 'origin' protein molecule exceeds 60 since some solvent molecules interact with more than two functional groups. The number of hydrogen bonds involving each type of functional group (and, in parentheses, the number of participating residues) are: O(peptide), 19 (18); O(carboxyl), 15 (9); N(peptide), 12 (12); O(hydroxyl), 9 (7); $N(NH_3^+)$, 6 (3); O/N(side-chain amide), 7 (4). The prominent role played by carboxylate groups is not surprising in view of the acidic nature of plastocyanin.

Estimates of accuracy and precision

If the crystallographic residual R is used as a guide, the structure at 1.33 Å resolution is significantly 'better' than that at 1.6 Å resolution. The residual R is 0.149 for 14 307 reflections ($8.0 \geq d \geq 1.33$ Å), and 0.141 for 10 093 reflections ($8.0 \geq d \geq 1.6$ Å), compared with 0.168 for 8285 reflections ($7.0 \geq d \geq 1.6$ Å) in the previous refinement. The significantly improved residual in the range covered by the original 1.6 Å data is noteworthy: the model refined using the higher-resolution data agrees better with the lower-resolution data than a model refined against the lower-resolution data alone. The comparison is not a simple one since the original intensities in the range $2.8 \geq d \geq 1.6$ Å were modified by merging the counter and film data. Improvements in the data are, however, not the sole (or even the main) cause of the improved residual R. If the values of F_{calc} from the 1.33 Å model are compared with the *original* values of F_{obs} for the 8285 reflections ($7.0 \geq d \geq 1.6$ Å) used in the 1.6 Å refinement, the residual R is 0.145.

Table 8. *Intermolecular contacts via bridging water molecules*

The table includes all intermolecular interactions in which a solvent site lies within hydrogen-bonding distance (≤ 3.3 Å) of functional groups on two adjacent protein molecules. Contacts involving more than one bridging solvent molecule, which were included in our earlier description of the structure (Guss & Freeman, 1983), are not listed. Superscripts denoting symmetry operations are explained in the footnote to Table 5. The symbol # denotes a solvent site which is too close to another to be fully occupied. The symbol * denotes an atom in the alternate conformer of a disordered side chain.

Links				Reverse links			
1	Ile N	174	77^{iii} Lys N^{ζ}	77	Lys N^{ζ}	174	1^{iv} Ile N
25	Glu $O^{\epsilon 1}$					25^{iv}	Glu $O^{\epsilon 1}$
8	Asp N	194^{v}	91^{v} Gly O	91	Gly O	194	8^{ii} Asp N
8	Asp O	147	18^{v} Glu $O^{\epsilon 1}$	18	Glu $O^{\epsilon 1}$	147^{ii}	8^{ii} Asp O
8	Asp $O^{\delta 1}$	169	17^{v} Ser O^{γ}	17	Ser O^{γ}	169^{ii}	8^{ii} Asp $O^{\delta 1}$
8	Asp $O^{\delta 2}$	104^{v}	14^{v} Phe N	14	Phe N	104	8^{ii} Asp $O^{\delta 2}$
			93^{v} Val O	93	Val O		
8	Asp $O^{\delta 2}$	$115^{\#}$	12^{v} Leu O	12	Leu O	$115^{\#ii}$	8^{ii} Asp $O^{\delta 2}$
9	Asp $O^{\delta 2}$	196	81^{v} Ser O^{γ}	81	Ser O^{γ}	196^{ii}	9^{ii} Asp $O^{\delta 2}$
9	Asp $O^{\delta 2}$	109^{v}	82^{v} Phe O	82	Phe O	109	9^{ii} Asp $O^{\delta 2}$
19	Phe O	199	61^{xii} Asp $O^{\delta 1}$	61	Asp $O^{\delta 1}$	199^{xi}	19^{xi} Phe O
20	Ser O^{γ}						20^{xi} Ser O^{γ}
20	Ser N	$\#185^{xii}$	58^{xii} Ser O^{γ}	58	Ser O^{γ}	$\#185$	20^{xi} Ser N
22	Ser O^{γ}	133^{ii}	48^{iii} Ser N	48	Ser N	133	22^{iv} Ser O^{γ}
99	Asn O						99^{iv} Asn O
22	Ser O^{γ}	203	48^{iii} Ser O	48	Ser O	203^{iv}	22^{iv} Ser O^{γ}
99	Asn $O^{\delta 1}$						99^{iv} Asn $O^{\delta 1}$
22	Ser O^{γ}	143	77^{iii} Lys O	77	Lys O	143^{iv}	22^{iv} Ser O^{γ}
25	Glu $O^{\epsilon 2}$						25^{iv} Glu $O^{\epsilon 2}$
23	Pro O	114	79^{iii} Glu N	79	Glu N	114^{iv}	23^{iv} Pro O
25	Glu $O^{\epsilon 1}$	135^{ii}	77^{iii} Lys N	77	Lys N	135	25^{iv} Glu $O^{\epsilon 1}$
48	Ser O^{γ}	157^{iv}	99^{iv} Asn O^{δ}	99	Asn O^{δ}	157	48^{iii} Ser O^{γ}
49	Gly O	129	59^{vii} Glu $O^{\epsilon 1}$	59	Glu $O^{\epsilon 1}$	129^{ix}	49^{ix} Gly O
51	Asp $O^{\delta 2}$	107^{vii}	86^{vii} Pro O	86	Pro O	107	51^{ix} Asp $O^{\delta 2}$
			90^{vii} Ala N	90	Ala N		
52	Ala N	101	86^{vii} Pro O	86	Pro O	101^{ix}	52^{ix} Ala N
53	Ser O	145	66^{*i} Lys N^{ζ}	66*	Lys N^{ζ}	145^{x}	53^{x} Ser O
53	Ser O^{γ}	117	66^{*i} Lys N^{ζ}	66*	Lys N^{ζ}	117^{x}	53^{x} Ser O^{γ}
54	Lys O	106^{i}	66^{i} Lys N	66	Lys N	106	54^{x} Lys O
54	Lys N^{ζ}	146^{i}	66^{i} Lys O	66	Lys O	146	54^{x} Lys N^{ζ}
54	Lys N^{ζ}	209^{vii}	88^{vii} Gln O	88	Gln O	209	54^{ix} Lys N^{ζ}
54*	Lys N^{ζ}	126^{vii}	85^{vii} Ser O	85	Ser O	126	54^{*ix} Lys N^{ζ}
			88^{vii} Gln $O^{\epsilon 1}$	88	Gln $O^{\epsilon 1}$		
59	Glu $O^{\epsilon 1}$	184^{ix}	76^{ix} Asn $O^{\delta 1}$	76	Asn $O^{\delta 1}$	184	59^{vii} Glu $O^{\epsilon 1}$
62	Leu N	167^{xi}	99^{xi} Asn $N^{\delta 2}$	99	Asn $N^{\delta 2}$	167	62^{xii} Leu N
64	Asn O	120	73^{x} Ala N	73	Ala N	120^{i}	64^{i} Asn O
64	Asn $O^{\delta 1}$	188	71^{x} Glu $O^{\epsilon 1}$	71	Glu $O^{\epsilon 1}$	188^{i}	64^{i} Asn $O^{\delta 1}$
64	Asn $N^{\delta 2}$	179^{xi}	99^{xi} Asn $N^{\delta 2}$	99	Asn $N^{\delta 2}$	179	64^{xii} Asn $N^{\delta 2}$

It is well known that the accuracy and precision of a macromolecular structure determination are difficult to estimate on the basis of a single experiment. Although the problem of distinguishing a correct macromolecular structure from an incorrect one has recently received attention (Brändén & Jones, 1990), there has been relatively little progress in estimating the precision which can be claimed for the atomic positions in a correctly determined and competently refined structure. The problem has its roots in the present methodology of protein structure refinement: firstly, the least-squares refinement matrices are usually diagonalized so that the terms required for the meaningful calculation of variances in the final parameters (see *e.g.* Dunitz, 1979) are absent; secondly, the final parameters are affected to an unknown extent by the inclusion of geometrical restraints and by the weighting of these restraints in relation to the observed structure amplitudes. The effectiveness of the restraints is shown by the r.m.s. and maximum deviations of various types of dimension from their ideal or target values (Table 3), but this information does not help us to assess the precision of individual atomic positions, bond lengths or even bond types.

The best that can at present be done in many cases is to rely on global or qualitative indicators. These include the familiar Luzatti (1952) plot, the σ_A plot (Read, 1986) and the σ_E plot (Read, 1990). In the present work, a Luzatti plot (not shown) yields an upper bound of 0.15 Å for the average error in the atomic positions at 1.33 Å resolution, compared with 0.20 Å at 1.6 Å resolution. The improvement is consistent with the decrease in the overall residual R noted earlier. It is also possible to estimate the radial errors in atomic positions from the diagonal elements of the inverse normal matrix in a final cycle of unrestrained least-squares refinement (Holland, Clancy, Muchmore, Ryde, Einspahr, Finzel, Heinrickson & Watenpaugh, 1990).

In order to obtain what we consider to be a more realistic estimate of the average uncertainty in the atomic positions at 1.33 Å resolution we adopt the fiction that the refinements at 1.33 and 1.6 Å resolution are completely independent determinations. An indication of the random errors can then be obtained from the r.m.s. difference between the positions of corresponding atoms. For C^{α} atoms the r.m.s. difference is 0.11 Å. For all polypeptide backbone atoms (N, C, O, C^{α}) it is 0.10 Å. (Side-chain atoms cannot be included in the comparison due to the disparities between the numbers of side chains and conformers located in the two refinements.) The r.m.s. differences between the two sets of refined atomic positions are accounted for almost completely by the estimated uncertainties in the atomic positions after the 1.6 Å refinement (0.09 Å for C^{α} atoms, 0.13 Å for all atoms with $B \leq 30$ Å^2) (Guss & Freeman, 1983). The ratio between the Luzatti estimates of the maximum average errors after the two refinements is 0.75 (see above). If we make the

further assumption that the average uncertainties are related in the same ratio as the maximum average errors, we obtain an estimate of $0.75 \times 0.09 = 0.07$ Å for the uncertainty in the backbone atom positions at 1.33 Å resolution. If the 'uncertainty' were an e.s.d. then, as Jensen (1985) has pointed out, 5% of the backbone atoms (20 atoms) would be at least 0.14 Å from their true positions, in surprisingly good agreement with the upper bound estimated from the Luzatti plot.

For metalloprotein structures, global indicators of precision have the disadvantage that they do not distinguish between the metal site and the rest of the molecule. Depending on the atomic numbers of the metal and ligand atoms and on the validity of restraints applied to the ligand groups, the atomic positions at the metal site may well be determined more or less precisely than those in the rest of structure. Yet it is just in relation to the geometry of the metal site that estimates of precision become very important since this is often the focus of attention for non-crystallographers. Further, as will be seen later, the crystallographically determined metal-site dimensions may be subject to systematic uncertainties in addition to random errors. Assessing the accuracy may be just as difficult as assessing the precision.

A strategy that has been adopted in a number of cases where the asymmetric unit of a crystal comprises two crystallographically independent metalloprotein molecules has been to treat the dimensions of the metal sites as though they have been determined independently. For example, an average 'maximum error' of 0.04 Å estimated in this way has been reported for the Cu–ligand bonds in *A. denitrificans* azurin (Baker, 1988), and similarly calculated 'measures of precision' ranging from 0.01 to 0.11 Å have been quoted for individual Cu–ligand bonds in two mutants of *Ps. aeruginosa* azurin (Nar, Messerschmidt, Huber, van de Kamp & Canters, 1991). It is of course a fact that the metal sites of independent molecules in a single crystal are *not* independently determined quantities in a statistical sense, but this objection may have to be put aside for reasons of practicality. A more serious objection is that such error estimates fail to take into account the systematic uncertainties which affect the metal-site dimensions, regardless of the number of molecules in the asymmetric unit (see *Trials of refinement strategy*, below). Error estimates which are based solely on internal self-consistency without allowance for systematic effects are likely to be too low. Our own strategy of treating the refinements of a structure at 1.6 and 1.33 Å as independent determinations would be subject to much the same criticism if the refinements were not based on independent data sets recorded from different crystal specimens.

In our previous studies of the plastocyanin structure we have taken advantage of the fact that the molecule is small enough to permit some experiments which in the case of larger proteins would be prohibitively expensive in terms of time and resources. In the refinement of the structure at 1.6 Å resolution, we carried out two independent sets of refinement calculations using independently recorded data sets (Guss & Freeman, 1983). Comparisons between the Cu-site geometries derived from the two refinements led to estimates of the standard deviations of the Cu–ligand bond lengths (0.05 Å) and ligand–Cu–ligand bond angles (3°). These values were shown to be consistent with estimates obtained from the r.m.s. difference between the atomic positions after superposing the two refined molecules and optimizing the fit between them (Guss & Freeman, 1983). Similar calculations led to comparable estimates for the best of the structures of reduced poplar plastocyanin where data sets with resolutions ranging from 1.8 to 2.1 Å were recorded at six pH values (Guss, Harrowell, Murata, Norris & Freeman, 1986). Larger e.s.d.'s (0.07 Å) were assigned to the Cu–ligand bond lengths in *Enteromorpha prolifera* plastocyanin after considering the quality of the data and the ratio of observations to refined variables, despite a low residual R (Collyer *et al.*, 1990).

We estimate the standard deviations of the Cu–ligand bond lengths and the ligand–Cu–ligand bond angles at 1.33 Å resolution as 0.04 Å and 2.5°, respectively. These values are derived subjectively from the corresponding e.s.d.'s in the refinement at 1.6 Å resolution, 0.05 Å and 3°, after considering the much higher ratio of diffraction measurements to refined variables (3.9 *versus* 2.7), the improved overall residual R (0.149 *versus* 0.168), and the improved Luzzati estimate of the maximum average error in the atomic positions (0.15 Å *versus* 0.20 Å, see above).

Trials of refinement strategy

A surprising result of the refinement at 1.33 Å resolution was that all four Cu–ligand bond lengths were *shorter* than at the end of the refinement at 1.6 Å resolution (Table 7). The largest decrease in bond length, −0.13 Å for Cu—N(His37), was 2.6 times our previous estimate for the standard deviation (0.05 Å). The remaining decreases, −0.04 to −0.08 Å, were not significant in terms of the e.s.d. but notable because they were *systematically* negative. Since there had been a number of changes in the refinement procedures and parameters apart from the increase in resolution, we investigated the extent to which each change had contributed to the systematic shortening of the Cu–ligand bonds.

Plots of the Cu–ligand bond lengths as functions of the number of 'rounds' of refinement (Fig. 10)

showed that the most dramatic changes had occurred during the early rounds of refinement. Thus changes in geometry were directly linked to the use of the new diffraction data, the use of a new *PROLSQ* dictionary, and the inclusion of H atoms in the model (Table 2). The subsequent effects of adding a substantial number of new solvent atoms to the model (rounds 6–8), adding the second conformer of disordered side chains (rounds 11–13), and making allowance for anomalous scattering were smaller. It was also apparent that abrupt excursions from general trends, such as the changes in Cu—N(His87) at round 2 and in Cu—N(His37) at round 11, had been rapidly reversed (Fig. 10); these temporary aberrations at the Cu site were presumably caused by the manual adjustment of other parts of the model. By the end of the refinement, the Cu–ligand bond lengths were fluctuating within limits (0.01–0.02 Å) which were well within the estimated standard deviation.

In order to obtain a quantitative estimate of the effect of change in the refinement conditions we proceeded as follows (see Fig. 11). The entire refinement was repeated, starting from the published 1.6 Å model, but introducing only one change of condition at a time. Each step in the calculations comprised ten least-squares cycles without operator intervention. As a test for possible correlations among the effects of the changes of condition, the entire sequence of calculations was carried out twice, the changes of condition being introduced in different orders. The factors tested were: (i) the use of a new set of merged counter and film data, replacing the original counter data at 1.6 Å; (ii) the increase in resolution from 1.6 to 1.33 Å; (iii) the use of different *PROLSQ* 'dictionary' parameters representing ideal geometry;* (iv) the inclusion of H atoms in the model; (v) the allowance for anomalous scattering by S and Cu; and (vi) the inclusion of solvent site occupancies as refinement variables.

Starting from the 1.6 Å model (*P*6 in Fig. 11), these calculations produced two closely similar models, *A* and *B*, which were refined under the same conditions as the fully refined 1.33 Å model *P*3 but differed from it by *excluding operator intervention* (manual adjustment to improve the fit with density, to model disordered side chains and to include additional solvent atoms). Operator intervention could not be reproduced in a *controlled* way because it had affected the original refinement during interactive computer-graphics sessions at intervals over an extended period. The effects of operator intervention could, however, be estimated from the lack of clos-

* The 'dictionary' parameters used in the present refinement were distributed with the 1985 release of *PROLSQ*. In the 1.6 Å refinement (Guss & Freeman, 1983) the 'dictionary' parameters were taken from the 1979 release. The standard values of the bond lengths and bond angles in histidine, cysteine and methionine residues in the two dictionaries differ by as much as 0.036 Å and 3.8°, respectively.

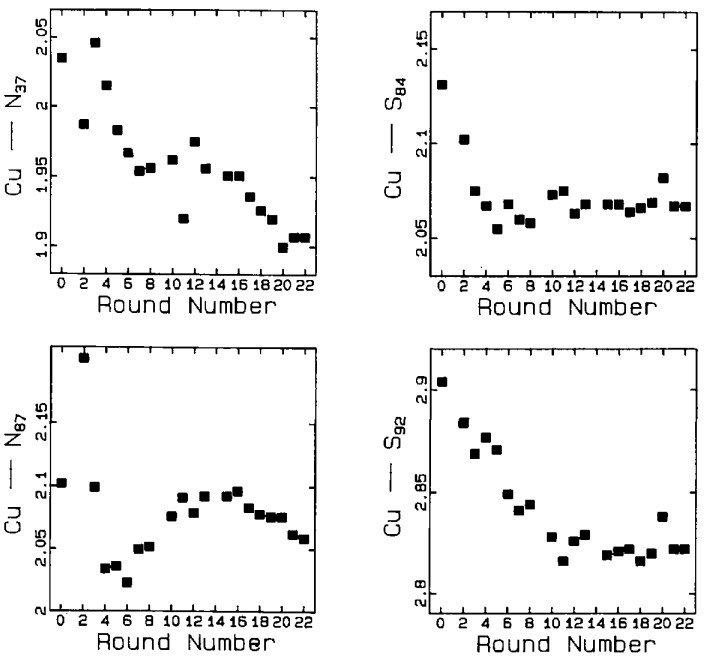

Fig. 10. Changes in the geometry of the Cu site during the refinement. The Cu–ligand bond lengths are plotted as functions of the number of 'rounds' of refinement.

ure between the calculations leading to models *A* and *B* on the one hand, and the full refinement leading to *P*3 on the other hand.

The estimates derived from this step-by-step refinement were checked by reversing the entire process: starting with the fully refined 1.33 Å model *P*3 and reversing every change of condition, a series of calculations produced model *C* which was refined under the same conditions as the original 1.6 Å structure *P*6 but differed from it by *including* the effects of operator intervention (since these were incorporated in *P*3). This sequence of calculations, too, was repeated with the changes of condition in a different order, yielding model *D* (see Fig. 11).

The calculations are summarized in Table 9 and Fig. 12. The final outcome comprised 25 sets of atomic coordinates and 25 sets of Cu–ligand bond lengths (including those from the actual 1.6 and 1.33 Å refinements).

Quantitative estimates of the contribution which each change in refinement conditions made to the final result were obtained by superposing the 25 coordinate sets (Table 9) in pairs. In each pair of coordinate sets, the r.m.s. difference between the atomic positions was calculated for (*a*) the C^{α} atoms, (*b*) the backbone N, C^{α}, C, O atoms, and (*c*) all atoms excluding solvent and the second conformer of disordered side chains. The r.m.s. differences for the C^{α} atoms and backbone atoms were in all cases approximately equal, and the values for the C^{α} atoms will not be cited or discussed here. The differences between the atomic positions and Cu–ligand bond lengths in selected pairs of coordinate sets are summarized in Tables 10 and 11.

We first consider how the atomic positions behaved during the stepwise refinement. To assist the visualization of the results, the movements of the Cu and individual ligand atoms are shown in Fig. 13. In every case the two refinements in each direction

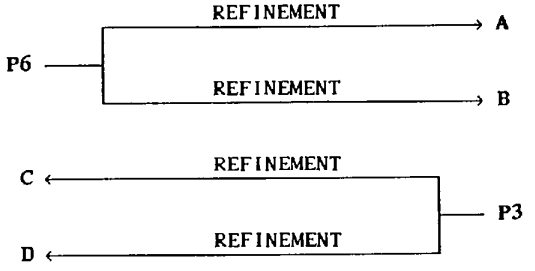

Fig. 11. (Top) *P*6, the Cu^{II}-plastocyanin structure refined at 1.6 Å resolution (Guss & Freeman, 1983); *A* and *B*, models derived from *P*6 by stepwise refinement without operator intervention. (Bottom) *P*3, the Cu^{II}-plastocyanin structure refined at 1.33 Å resolution (present work); and *C* and *D*, models derived from *P*3 by stepwise refinement without operator intervention, the changes of condition introduced in the top half of the figure being reversed.

produced closely similar results even though they followed different paths. We conclude that the effect of new refinement conditions is independent of the order in which the changes of condition are introduced. Further, the cumulative effect of altering the conditions of refinement sequentially is to move each atom towards its position in the model obtained by a complete refinement using the same final conditions but including operator intervention. As already noted, the gap between the position of an atom at the end of the step-by-step refinement and its position in the fully refined model represents the effect of operator intervention. The results of the refinement sequences starting from *P*6 (the refined 1.6 Å model) always resemble *P*3 (the refined 1.33 Å model) more closely than the results of the sequences starting from *P*3 resemble *P*6. This is to be expected, since *P*3 was originally derived from *P*6 by a refinement which corresponded to the steps in Fig. 13 *plus* operator intervention, whereas *P*6 was not derived from *P*3 but from the 2.7 Å MIR model.

Relative effects of refinement conditions on atomic positions. If the r.m.s. changes in the positions of the backbone atoms are compared (Table 10), the largest effects – as judged by the *maximum* r.m.s. change for each group of examples – are generated by the change from counter to counter/film data (0.06 Å), by the inclusion of H atoms in the model (0.03 Å), and by the increase in resolution (0.03 Å). Smaller but detectable effects are caused by the changes in the *PROLSQ* dictionary, and by the treatment of the solvent site occupancies as refinement variables (both ~0.02 Å). The inclusion of anomalous-scattering terms has a negligible effect (0.01 Å).

Relative effects of refinement conditions on Cu–*ligand bonds.* We now return to the question why the Cu–ligand bond lengths became systematically shorter during the refinement at higher resolution. On the whole, the two factors which most influenced the Cu–ligand bond lengths were also those that had the largest effects on the positions of the backbone atoms (Table 10). The largest r.m.s. changes in the Cu–ligand bond lengths were caused by the change from counter to counter/film data, and by the inclusion of H atoms in the model (both ~0.03 Å). More moderate but still significant effects were caused by the increase in resolution from 1.6 to 1.33 Å, and by the changes in the *PROLSQ* dictionary (both ~0.02 Å). Among the procedures and parameters tested, the smallest contributions to the final values were made by the allowance for anomalous scattering in the atomic scattering factors of Cu and S, and by the treatment of the solvent site occupancies as refinement variables (both ~0.01 Å).

Effect of refinement conditions on individual Cu–*ligand bonds.* Only the use of a new *PROLSQ* dictionary caused changes that were limited to one type

Table 9. *Trials of refinement strategy*

Explanation of column headings: Code, identifier for refinement trial; Resolution, '1.60' = 7.0–1.6 Å, '1.33' = 8.0–1.33 Å; Data, 'Old' (diffractometer) or 'New' (diffractometer plus film) data set; Dict., *PROLSQ* dictionary of target values for molecular geometry; H, inclusion of H atoms at calculated positions; A/S, inclusion of anomalous scattering terms in Cu and S scattering factors; Occupancy, treatment of occupancies of solvent sites ('Yes' = variable, 'No' = fixed).

Code	Resolution (Å)	Data	Dict.	H	A/S	Occupancy	Bond lengths (Å) from Cu to N_{37}	S_{84}	N_{87}	S_{92}
*P6**	1.60	Old	Old	No	No	No	2.035	2.131	2.102	2.904
a1/b1	1.60	New	Old	No	No	No	2.011	2.113	2.158	2.864
a2	1.60	New	New	No	No	No	1.984	2.113	2.132	2.856
a3	1.60	New	New	Yes	No	No	1.965	2.117	2.086	2.858
a4	1.60	New	New	Yes	Yes	No	1.954	2.109	2.076	2.860
a5	1.33	New	New	Yes	Yes	No	1.955	2.095	2.076	2.838
A	1.33	New	New	Yes	Yes	Yes	1.950	2.096	2.075	2.839
b2	1.33	New	Old	No	No	No	1.987	2.094	2.120	2.836
b3	1.33	New	New	No	No	No	1.986	2.096	2.124	2.836
b4	1.33	New	New	Yes	No	No	1.966	2.093	2.090	2.836
b5	1.33	New	New	Yes	Yes	No	1.963	2.093	2.090	2.835
B	1.33	New	New	Yes	Yes	Yes	1.947	2.088	2.081	2.840
P3†	1.33	New	New	Yes	Yes	Yes	1.907	2.068	2.059	2.822
c1	1.60	New	New	Yes	Yes	Yes	1.898	2.083	2.065	2.840
c2	1.60	New	New	Yes	No	Yes	1.908	2.091	2.078	2.839
c3	1.60	New	New	No	No	Yes	1.916	2.091	2.135	2.832
c4	1.60	New	Old	No	No	Yes	1.943	2.092	2.166	2.833
c5	1.60	Old	Old	No	No	Yes	1.943	2.072	2.120	2.854
C	1.60	Old	Old	No	No	No	1.964	2.081	2.132	2.835
d1	1.33	New	New	Yes	No	Yes	1.915	2.073	2.065	2.822
d2	1.33	New	New	No	No	Yes	1.923	2.075	2.125	2.809
d3	1.33	New	Old	No	No	Yes	1.949	2.078	2.156	2.810
d4	1.60	New	Old	No	No	Yes	1.944	2.094	2.172	2.828
d5	1.60	Old	Old	No	No	Yes	1.945	2.073	2.129	2.845
D	1.60	Old	Old	No	No	No	1.964	2.081	2.118	2.833

* These are the deposited coordinates (1PCY) resulting from 1.60 Å refinement.
† These are the final deposited coordinates resulting from the 1.33 Å refinement.

of Cu–ligand bond: the shortening of Cu—N_{37} and Cu—N_{87} accounted for nearly all of the average shortening of ~0.02 Å mentioned above. The inclusion of H atoms, on the other hand, (Table 10) causes little change in Cu—S_{84} and Cu—S_{92}, a −0.01 Å change in Cu—N_{37}, and a −0.05 Å change in Cu—N_{87}. Thus the introduction of H atoms has different effects not only on different types of Cu–ligand bond, but even on two Cu–ligand bonds of the same type. A similar disparity is found if the effects of changing from counter data to merged counter/film data at 1.6 Å resolution are examined. In this case the Cu—N_{37} and Cu—N_{87} bond lengths actually change in opposite directions (−0.01 and

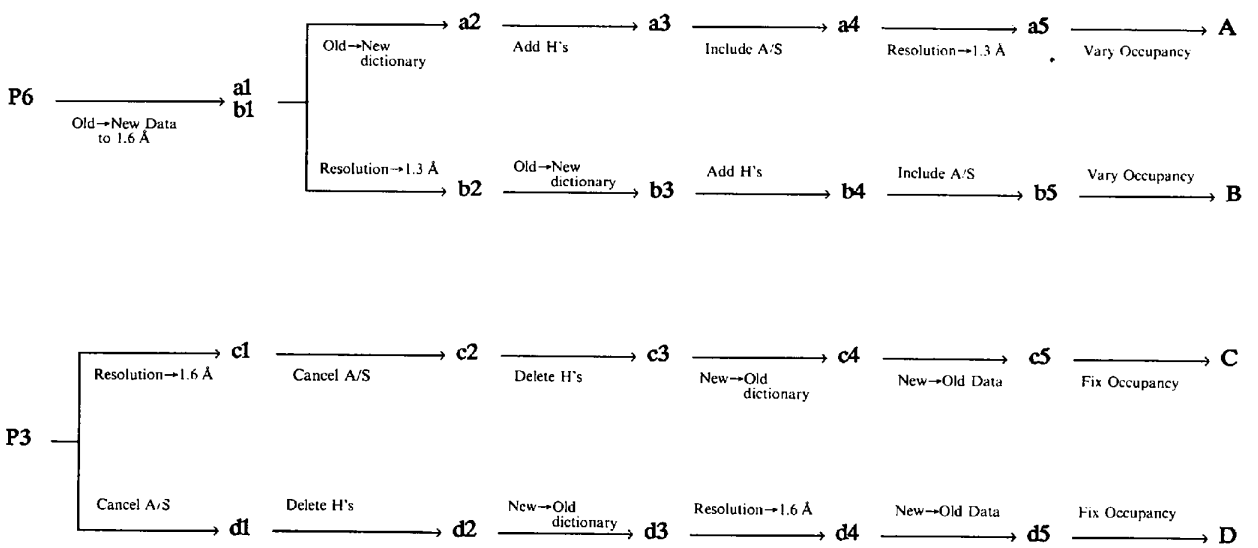

Fig. 12. Trials of refinement strategy. The calculations listed in Table 8 are arranged to show the relationships between the sets of refined coordinates.

Table 10. *Changes in atomic positions and* Cu-*site geometry as a result of changes in refinement conditions*

Change in refinement conditions	Examples*	R.m.s. changes† (× 10^3 Å) in positions NC°CO atoms	All atoms‡	Mean change (× 10^3 Å) in Cu–ligand bond lengths N_{37}	S_{84}	N_{87}	S_{92}	All	R.m.s. change (× 10^3 Å) in Cu–ligand bond lengths
Increase in resolution from 1.6 to 1.33 Å	(1)	13–25	22–69	−2	−16	−15	−21	−14	15
Change from counter to merged counter/film data at 1.6 Å resolution	(2)	31–58	30–100	−8	8	48	−26	5	31
Change to new dictionary of standard amino-acid geometries	(3)	9–17	13–46	−20	−1	−21	−2	−11	15
Inclusion of H atoms	(4)	31–34	20–46	−14	0	−49	6	−14	26
Allowance for anomalous scattering by S and Cu	(5)	3–8	7–28	−8	−5	−7	0	−5	6
Inclusion of variable occupancies for solvent sites	(6)	5–18	16–37	−15	−5	−3	9	−4	9

* Examples: (1) *a*4→*a*5, *b*1→*b*2, *c*1→*P*3, *d*4→*d*3; (2) *P*6→*a*1, *c*5→*c*4, *d*5→*d*4; (3) *a*1→*a*2, *b*2→*b*3, *c*4→*c*3, *d*3→*d*2; (4) *a*2→*a*3, *b*3→*b*4, *c*3→*c*2, *d*2→*d*1; (5) *a*3→*a*4, *b*4→*b*5, *c*2→*c*1, *d*1→*P*3; (6) *a*5→*A*, *b*5→*B*, *C*→*c*5, *D*→*d*5.

† Only the smallest and largest r.m.s. changes in each group of examples are cited.

‡ For disordered side chains, the atoms of only one conformer were included in the calculation of the r.m.s. changes.

Table 11. *Comparison between refinements including/excluding operator intervention*

Superimposed models	Atomic positions NC°CO atoms R.m.s. Δ (× 10^3 Å)	All atoms R.m.s. Δ (× 10^3 Å)	Cu–ligand bond lengths Mean Δ* (× 10^3 Å)	R.m.s. Δ (× 10^3 Å)
Complete refinement including operator intervention				
*P*6→P3	103	606	−70	85
Effect of operator intervention				
Only one model of a pair includes contribution of operator intervention				
*P*6→*C*	98	609	−40	57
*P*6→*D*	97	611	−44	57
A→*P*3	67	617	−26	28
B→*P*3	67	622	−25	26
Effect of all other changes in refinement conditions				
Neither model or both models include contribution of operator intervention				
*P*6→*A*	81	197	−53	58
*P*6→*B*	83	192	−54	60
C→*P*3	55	95	−39	47
D→*P*3	56	99	−35	42

* Changes in bond lengths are calculated by subtracting the values in the first model from those in the second model of each pair.

+0.05 Å). Similar observations can be made about the other changes in condition listed in Table 10.

Effect of operator intervention on the Cu-*site geometry*. The factor which made the largest single contribution to the positions of the atoms at the Cu site – and therefore to the systematic shrinkage of the Cu–ligand bonds – during the high-resolution refinement was 'operator intervention', as defined above. The magnitude of the effect can be seen by comparing models *A* and *B* with *P*3, and models *C* and *D* with *P*6 (see Fig. 11 and Table 11). In each example, the models compared differ only by the inclusion of operator intervention in the refinement of one but not the other. The total contribution of *all other* changes in refinement condition can be seen by comparing models *A* and *B* with *P*6, and models *C* and *D* with *P*3 (see Fig. 11 and Table 11). The mean changes in the Cu–ligand bond lengths due to operator intervention (−0.03 to −0.04 Å) are 75–80% as large as those attributable to all other refinement conditions (−0.04 to −0.05 Å). The r.m.s. changes in the Cu—ligand bond lengths are similarly related (0.03–0.06 Å *versus* 0.04–0.06 Å). If, instead of concentrating on the Cu—ligand bond lengths, we consider changes in atomic position averaged over the whole molecule (Table 11), the effect of operator intervention is seen to be somewhat larger for the polypeptide backbone atoms (0.07–0.10 Å *versus* 0.06–0.08 Å), and very much larger for 'all' atoms (0.6 Å *versus* 0.1–0.2 Å). The latter result is to be expected since the averages over 'all' atoms include the side chains which are usually more frequent targets for operator intervention than the backbone atoms.

Why are the Cu–*ligand bond lengths systematically shorter at higher resolution?* The calculations described in the preceding sections provide a new insight to systematic effects in the refinement of macromolecular structures. However, while our observations explain why refinements at 1.6 and 1.33 Å resolution have produced *different* values for the Cu–ligand bond lengths, we still have no explanation for the fact that the bond lengths obtained at the higher resolution are systematically *shorter*.

Concluding remarks

There are two objective indicators that the present work has produced a more accurate model – that is, a set of atomic positional coordinates which more closely represent the true structure – than the previous refinement at 1.6 Å resolution. Firstly, the

electron density maps contain more interpretable detail and less noise; and secondly, the agreement between the observed and calculated structure amplitudes is better. The trial calculations described in the preceding section demonstrate unequivocally that the accuracy of the structure was strongly influenced by a number of refinement procedures and parameters. The following conclusions appear to be of general validity.

(i) The accuracy of a structure is strongly dependent on the decisions which the operator makes during those refinement procedures which involve the interpretation of electron density maps. Such decisions (for example, distinguishing solvent from noise, and recognizing disorder in side chains) involve human pattern recognition, and therefore depend on the operator's skill and experience.

(ii) The experimental factors which most strongly influence the accuracy of a structure are the resolution and quality of the data.

(iii) Other experimental factors which significantly influence the accuracy of a structure are the inclusion of H atoms, the values of the parameters representing ideal amino-acid geometry, and the use of atomic scattering factors including anomalous-scattering contributions.

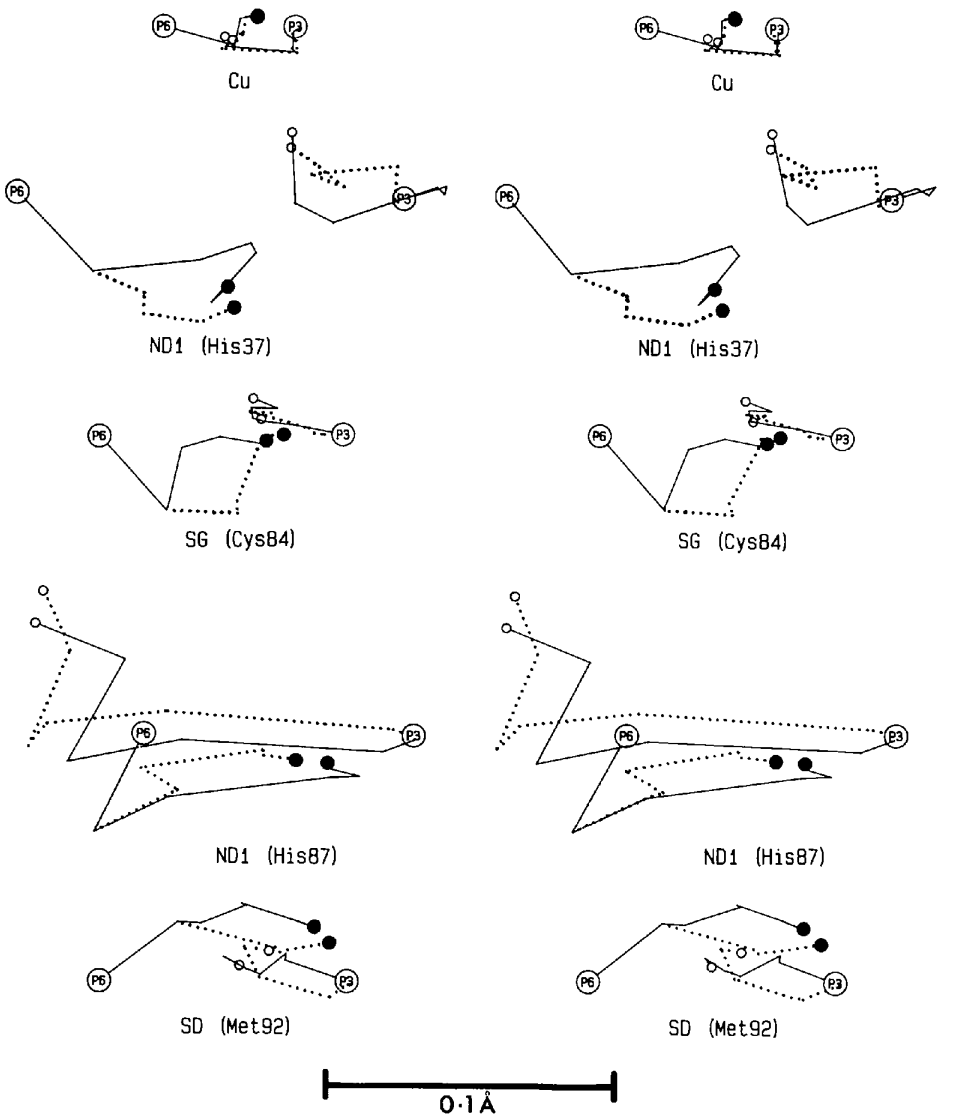

Fig. 13. Stereo drawings showing the movement of the Cu and ligand atoms during the trials of refinement strategy. In each drawing, the two large circles indicate the positions of an atom in the structures *P6* and *P3* (see Fig. 11). The continous lines —— and dotted lines ..... trace the movements of the atom in the two refinement series starting from each structure. The small circles indicate the final positions of the atom. Filled circles refer to the refinements starting from *P6*. Empty circles refer to the refinements starting from *P3*.

(iv) Treating solvent site occupancies as refinement variables, on the other hand, affects the accuracy of a protein molecular structure only to a minor extent.

Finally, given the movements of atoms observed in our trial calculations, we conclude that the uncertainties (quasi-e.s.d.'s) in the metal–ligand bond lengths in the well refined structure of a ~ 10 kDa metalloprotein are 0.04 Å at 1.3 Å resolution, 0.05 Å at 1.6 Å resolution, and 0.07 Å at 1.8 Å resolution.

Access to the facilities at the EMBL outstation at Hamburg was provided by the European Molecular Biology Organisation. Financial support from the Australian Research Grants Scheme (grants A2860032P and A28930307) and the Dr Joan R. Clark Research Fund is gratefully acknowledged.

References

ADMAN, E. T. (1985). *Topics in Molecular and Structural Biology. Metalloproteins*, Vol. 1, edited by P. M. HARRISON, pp. 1–42. London: Macmillan.

ADMAN, E. T. & JENSEN, L. H. (1981). *Isr. J. Chem.* **21**, 8–12.

ADMAN, E. T., TURLEY, S., BRAMSON, R., PETRATOS, K., BANNER, D., TSERNOGLOU, D., BEPPU, T. & WATANABE, H. (1989). *J. Biol. Chem.* **264**, 87–99.

ADMAN, E. T., WATENPAUGH, K. D. & JENSEN, L. H. (1975). *Proc. Natl Acad. Sci. USA*, **72**, 4854–4858.

AMBLER, R. P. (1983). Unpublished work cited by Guss & Freeman (1983).

ARNDT, U. W. (1977). *The Rotation Method in Crystallography*, edited by U. W. ARNDT & A. J. WONACOTT, pp. 19–31. Amsterdam: North-Holland.

BACKES, G., MINO,Y., LOEHR, T. M., MEYER, T. E., CUSANOVICH, M. A., SWEENEY, W. V., ADMAN, E. T. & SANDERS-LOEHR, J. (1991). *J. Am. Chem. Soc.* **113**, 2055–2064.

BAKER, E. N. (1988). *J. Mol. Biol.* **203**, 1071–1095.

BAKER, E. N. & HUBBARD, R. (1984). *Prog. Biophys. Mol. Biol.* **44**, 97–179.

BARTUNIK, H. D., GEHRMANN, T. & ROBRAHN, B. (1984). *J. Appl. Cryst.* **17**, 120.

BEOKU-BETTS, D., CHAPMAN, S. K., KNOX, C. V. & SYKES, A. G. (1985). *Inorg. Chem.* **24**, 1677–1681.

BHAT, T. N. (1989). *Acta Cryst.* A**45**, 145–146.

BOULTER, D., HASLETT, B. G., PEACOCK, D., RAMSHAW, J. A. M. & SCAWEN, M. D. (1977). *International Review of Biochemistry*, Vol. 13, *Plant Biochemistry* II, edited by D. H. NORTHCOTE, pp. 1–40. Baltimore: Univ. Park Press.

BRÄNDÉN, C.-I. & JONES, T. A. (1990). *Nature (London)*, **343**, 687–689.

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.

DIMITROV, M. I., EGOROV, A. A., DONCHEV, T. A. & ATANASOV, B. P. (1987). *FEBS Lett.* **226**, 17–22.

DUNITZ, J. D. (1979). *X-ray Analysis and the Structure of Organic Molecules*, pp. 206–207. Ithaca: Cornell Univ. Press.

FRASER, R. D. B., MACRAE, T. P. & SUZUKI, E. (1978). *J. Appl. Cryst.* **11**, 693–694.

GARRETT, T. P. J., CLINGELEFFER, D. J., GUSS, J. M., ROGERS, S. J. & FREEMAN, H. C. (1984). *J. Biol. Chem.* **259**, 2822–2825.

GETZOFF, E. D., TAINER, J. A., SIMPSON, M. M., BELL, G. I. & HALLEWELL, R. A. (1989). *Proteins*, **5**, 322–336.

GUSS, J. M. & FREEMAN, H. C. (1972). *Acta Cryst.* B**28**, 2090–2096.

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.

GUSS, J. M., MERRITT, E. A., PHIZACKERLEY, R. P., HEDMAN, B., MURATA, M., HODGSON, K. O. & FREEMAN, H. C. (1988). *Science*, **241**, 806–811.

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. (1985). *Methods in Enzymology*, Vol. 115, *Diffraction Methods for Biological Molecules, Part B*, edited by H. W. WYCKOFF, C. H. W. HIRS & S. N. TIMASHEFF, p. 267. Orlando: Academic Press.

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

HOLLAND, D. R., CLANCY, L. L., MUCHMORE, S. W., RYDE, T. J., EINSPAHR, H. M., FINZEL, B. C., HEINRICKSON, R. L. & WATENPAUGH, K. D. (1990). *J. Biol. Chem.* **265**, 17649–17656.

JENSEN, L. H. (1985). *Methods in Enzymology*, Vol. 115, *Diffraction Methods for Biological Molecules, Part B*, edited by H. W. WYCKOFF, C. H. W. HIRS & S. N. TIMASHEFF, pp. 227–234. Orlando: Academic Press.

JENSEN, L. H. (1990). *Acta Cryst.* B**46**, 650–653.

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

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

KATOH, S. (1960). *Nature (London)*, **186**, 533–534.

KING, G. & WRIGHT, P. E. (1986). *Biochemistry*, **25**, 2364–2374.

KUNDROT, C. E. & RICHARDS, F. M. (1987). *Acta Cryst.* B**43**, 544–547.

LANGRIDGE, R., MARVIN, D. A., SEEDS, W. E., WILSON, H. R., HOOPER, C. W., WILKINS, M. H. F. & HAMILTON, L. D. (1960). *J. Mol. Biol.* **2**, 38–64.

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

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). *X-ray Absorption Fine Structure*, edited by S. S. HASNAIN, pp. 152–154. Chichester: Ellis Horwood.

NAR, H., MESSERSCHMIDT, A., HUBER, R., VAN DE KAMP, M. & CANTERS, G. W. (1991). *J. Mol. Biol.* **218**, 427–477.

PETRATOS, K., DAUTER, Z. & WILSON, K. S. (1988). *Acta Cryst.* B**44**, 628–636.

RASHIN, A. A. & HONIG, B. (1984). *J. Mol. Biol.* **173**, 515–521.

READ, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

READ, R. J. (1990). *Acta Cryst.* A**46**, 900–912.

RICHARDSON, J. S., GETZOFF, E. D. & RICHARDSON, D. C. (1978). *Proc. Natl Acad. Sci. USA*, **75**, 2574–2578.

ROBERTS, V. A., FREEMAN, H. C., OLSON, A. J., TAINER, J. A. & GETZOFF, E. D. (1991). *J. Biol. Chem.* **266**, 13431–13441.

SAWYER, L. & JAMES, M. N. G. (1982). *Nature (London)*, **295**, 79–80.

SCHWAGER, P., BARTELS, K. & JONES, A. (1975). *J. Appl. Cryst.* **8**, 275–280.

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

SIELECKI, A. R. & JAMES, M. N. G. (1981). *Refinement of Protein Structures*, edited by P. A. MACHIN, J. W. CAMPBELL & M. ELDER, pp. 78–87. Science and Engineering Research Council, Daresbury Laboratory, Warrington, England.

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

SYKES, A. G. (1990). *Struct. Bonding (Berlin)*, **75**, 177–224.