Acta Cryst. (1992). B48, 42–59

Refinement of Rubredoxin from *Desulfovibrio vulgaris* at 1.0 Å with and without Restraints

BY ZBIGNIEW DAUTER

European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, D-2000 Hamburg 52, Germany

LARRY C. SIEKER

Department of Biological Structures, SM-20, University of Washington, Seattle, WA 98195, USA

AND KEITH S. WILSON

European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, D-2000 Hamburg 52, Germany

(*Received* 22 *July* 1991; *accepted* 5 *September* 1991)

Abstract

X-ray data have been recorded from crystals of rubredoxin derived from the bacterium *Desulfovibrio vulgaris* to a resolution of 1.0 Å using in part synchrotron radiation and in part X-rays from a sealed-tube Mo $K\alpha$ source. In both cases an imaging-plate scanner was used as detector. The space group of the crystals is $P2_1$ with cell dimensions $a = 19.97$, $b = 41.45$, $c = 24.41$ Å and $\beta = 108.3°$. The overall merging $R(I)$ factor between symmetry-related reflections was 5.8%. The model was refined by least-squares minimization initially with stereochemical restraints to an R factor of 16.4%. Only atomic positional parameters and isotropic temperature factors for non-H atoms were used in the refinement. There were 18532 independent X-ray observations for a total of 1916 atomic parameters. A round of unrestrained refinement gave an R factor of 16.0%, acceptable geometry for more than 90% of the protein atoms, but emphasized the disorder inherent in eight of the residues. A final round of restrained refinement gave an R factor of 14.7%. Three of the 389 protein atoms in the molecule, in the side chain of Lys2, have been assigned zero occupancy in the model. A total of eight atoms in three side chains have been assigned two conformations, giving 393 protein atomic sites in the model. In addition there is one Fe atom, a sulfate ion and 102 water sites. 339 H atoms were included at their calculated positions, which were not refined. There is clear evidence for anisotropic thermal motion. This has not been incorporated in the present model.

Introduction

Rubredoxin is a small bacterial protein with just over 50 amino acids. It contains a single Fe atom, approximately tetrahedrally coordinated by the S atoms of four cysteine residues. The detailed function of rubredoxin in the cell remains unknown, but it is assumed to be part of a redox chain, with the Fe atom changing oxidation state by one electron. The protein has been purified from sulfate-reducing bacteria and also from several aerobic bacteria.

The first rubredoxin to be studied crystallographically was that from *Clostridium pasteurianum*, and pioneering work on the refinement of protein structures by least-squares minimization was carried out on this protein using data to 1.5 Å resolution (Watenpaugh, Sieker, Herriott & Jensen, 1972, 1973). Highly ordered crystals of rubredoxin diffracting to a resolution approaching 1.0 Å were reported some years ago (Pierrot, Haser, Frey, Bruschi, Le Gall, Sieker & Jensen, 1976) for protein from *Desulfovibrio vulgaris* and *Desulfovibrio gigas*, both sulfate-reducing bacteria. Refined structures have now been reported for rubredoxins from *Clostridium pasteurianum* at 1.2 Å (Watenpaugh, Sieker & Jensen, 1979), *Desulfovibrio gigas* at 1.4 Å (Frey, Sieker, Payan, Haser, Bruschi, Pepe & Le Gall, 1987), *Desulfovibrio desulfuricans* at 1.5 Å (Stenkamp, Sieker & Jensen, 1990) and *Desulfovibrio vulgaris* at 1.5 Å resolution (Adman, Sieker & Jensen, 1991). Only the most recent reference is given for each of these: reports of earlier refinements at lower resolutions can be found therein.

We here describe the collection of X-ray diffraction data to a nominal resolution of 1.0 Å for the *Desulfovibrio vulgaris* rubredoxin. The structure was refined firstly using least-squares minimization with stereochemical restraints. The ratio of observations to parameters indicated that refinement without stereochemical restraints should be possible. The results of the unrestrained refinement are presented

0108-7681/92/010042-18$03.00

and compared with those obtained with restraints. In the final restrained refinement two alternative conformations were included for three of the residues, and the H-atom contributions at their calculated positions. The final model is closely similar to that obtained at 1.5 Å resolution by Adman *et al.* (1991) and readers are referred to that publication for a more detailed discussion of the structure.

Table 1. *Summary of the data collection*

	High-resolution data set	Medium-resolution data set	Low-resolution data set
Radiation	Synchrotron	Mo $K\alpha$	Mo $K\alpha$
Wavelength (Å)	0.65	0.71	0.71
Crystal-to-plate (mm)	120	180	280
Maximum resolution (Å)	1.0	1.5	2.3
Oscillation per image (°)	1.0	2.0	4.0
Number of images	88	90	90
Crystal mount	c^*	*ca* b^*	*ca* b^*
Angular spread (°)	>0.5	0.4	0.4

Crystallization, data collection and processing

Rubredoxin from *Desulfovibrio vulgaris* was crystallized as reported previously (Adman, Sieker, Jensen, Bruschi & Le Gall, 1977). In brief the conditions were as follows. A protein solution of concentration 0.5–1.0%, buffered with 0.1 *M* sodium citrate at pH 4.0, was slowly adjusted to be 2 *M* in ammonium sulfate. Dark red prisms developed at ambient temperature. The space group is $P2_1$ with cell dimensions $a = 19.97$, $b = 41.45$, $c = 24.41$ Å and $\beta = 108.3°$. The packing density, V_m, has the very low value of 1.68 Å^3 dalton^{-1}, indicating tight packing of the protein in the crystal.

Data were first collected from a single crystal using synchrotron radiation from the EMBL beamline X31 at the DORIS storage ring, DESY, Hamburg. The wavelength used was 0.65 Å. The ring was operating in main-user mode at an energy of 3.7 GeV, with approximately 4 h between successive particle injections. Details of the beamline have been reported previously (Wilson, 1989). The first optical element is a channel-cut Si(111) monochromator. The second element is a toroidal mirror providing 1:1 focusing of the source. The final collimation is provided by two pairs of vertical and horizontal slits separated by about 25 cm. The apertures were set to 0.4 mm in each direction. The crystal in a capillary was mounted on a φ-rotation axis.

The data were recorded using an imaging-plate scanner developed and built in the EMBL outstation by Jules Hendrix and colleagues. The system and the software used for data reduction were as given in Dauter, Terry, Witzel & Wilson (1990). In brief, the images are treated as films and analyzed with a modified version of the *MOSFLM* film processing package (Leslie, Brick & Wonacott, 1986). Later computations were carried out with the *CCP*4 suite of programs (SERC Daresbury Laboratory, 1979).

Most of the low-resolution (>2 Å) reflections saturated the scanner electronics during the recording of the 1.0 Å data, and were not measured. Therefore a second crystal was later used with radiation from a conventional Seifert X-ray source with an Mo $K\alpha$ sealed tube operated at 3 kW equipped with a graphite monochromator. Data were initially recorded to a resolution of 1.5 Å, and then with shorter exposures to 2.5 Å to allow the measurement of those few reflections which remained saturated in the 1.5 Å set. The data-collection parameters for all three sets are summarized in Table 1.

The 'angular spread' given in Table 1 includes the contribution of both the mosaic spread of the crystal and the divergence of the X-ray beam. The crystal used on the conventional source was of high quality, with the divergence of the sealed-tube beam being the dominant factor: it is usual for this to be of the order 0.3–0.4°. However, for synchrotron radiation the beam divergence is very low, of the order of 0.1° or less. Thus, the high angular spread of the data from the crystal used on the synchrotron indicates it was substantially imperfect. The high mosaicity was evident in the images recorded from the synchrotron crystal. This is in keeping with it being mounted in a stabilizing solution retrospectively found to be slightly higher in ammonium sulfate concentration than is proper for the *Desulfovibrio vulgaris* rubredoxin crystals. The data for the 2.5 and 1.5 Å sets show essentially no indication of radiation damage as reflected in the relative '*B* values' used in scaling together the successive images. Perhaps surprisingly given its high mosaic spread, the crystal used on the synchrotron gave 1.0 Å data which had negligible '*B* values' relative to the first crystal. In this context the relative '*B* value' as well as the scale factor for each image are refined during the merging of data from the different images: the '*B* value' allows for different fall-off of intensity with resolution on different images, and thus takes up effects including the relative degree of order of different crystals, radiation damage and relative absorption. The result suggested that the crystal had suffered little decay during the experiment. However, an attempt to subsequently record low-resolution data from this crystal the following day revealed a substantial increase in disorder and the data collection was terminated. Given the use of such a short wavelength for the data collection no absorption correction was applied. It was felt that the image-to-image scaling would reveal, and to some extent take up, any residual differences.

The three sets of data were merged together, rejecting the 2.5 Å resolution inner shell from the 1.0 Å synchrotron data. This left a total of 59492

observations, comprising 38172 fully and 21320 partially recorded reflections, resulting in 18545 unique observations. The overall merging $R(I)$ factor was 5.8% (rising to 18% for the highest resolution shell) and the data between 10.0 and 1.0 Å were 90.5% complete (99% to 1.5 Å). The merging $R(I)$ factor and the completeness are shown as a function of resolution in Figs. 1(*a*) and 1(*b*) respectively. The Wilson plot (Wilson, 1949) is shown in Fig. 2 and gives an estimate of the overall temperature factor of 5.9 Å^2.

Restrained refinement

Initial model

The starting model for the refinement was the coordinate set 3RXN from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977), based on the refinement of the structure at 2.0 Å resolution (Adman *et al.*, 1977). No solvent molecules were present in this initial model. The original authors emphasized that the coordinates represented an intermediate stage of their refinement. We chose this model to allow refinement to proceed somewhat independently of that described at 1.5 Å resolution (7RXN; Adman *et al.*, 1991), leading to an increased significance in the comparison of results at 1.5 Å, and the present results at 1.0 Å resolution.

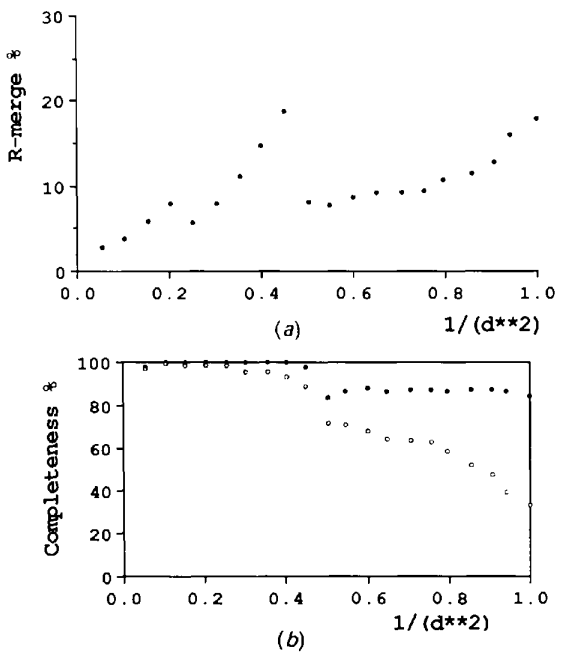


Fig. 1. (*a*) The merging $R(I)$ factor, defined as $\Sigma|I - \langle I\rangle|/\Sigma I$ and (*b*) the percentage completeness of the data as a function of resolution. In (*b*) the full circles refer to the total number of intensities recorded and the empty circles to those greater than $3\sigma(I)$.

Method

Restrained refinement was carried out in seven steps comprising 40 cycles using the stereochemically restrained least-squares minimization procedure of Konnert & Hendrickson (1980). Unit relative weights were applied to the X-ray terms throughout the refinement. In all cycles the complete X-ray data within the resolution range were used with no σ cutoff on amplitude. Fast Fourier algorithms were used for calculation of structure factors and gradients (Agarwal, 1978) as implemented by Dodson (Baker & Dodson, 1980). The restraints placed on the stereochemical parameters were as shown in Table 2. After this procedure had essentially converged at cycle 40, the stereochemical constraints were removed for a number of cycles, see later section.

Rebuilding of the structure was carried out between several of the steps, using the program *FRODO* (Jones, 1978), implemented on an Evans and Sutherland PS330 interactive graphics station. The rebuilding consisted mainly of the introduction of water molecules and a sulfate ion to the model, and some adjustments of side-chain conformations. Changes to the main chain were only needed at the N and C termini. There now follows a brief summary of the various steps. The course of the refinement is summarized in Table 2.

Step 1. Cycles 1–8. In the first five cycles only the positional parameters were refined, followed by the B values in the next three cycles. The original coordinate set, 3RXN, with no water molecules, was used as the model. Data were restricted to 1.5 Å resolution. The R factor fell from 30.8 to 23.3%.

Step 2. Cycles 9–15. Four cycles of *xyz* followed by two cycles of temperature-factor refinement. 48 water molecules and a sulfate ion were introduced to the model after inspection of the difference Fourier synthesis before this step. The R factor dropped to 18.4%.

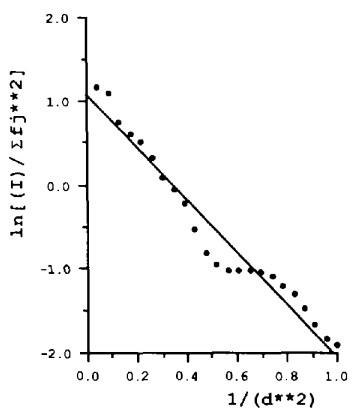


Fig. 2. The Wilson plot for the 1.0 Å rubredoxin data.

Table 2. *Course of the refinement*

Step		1	2	3	4	5	6	7	8*U*	9*U*		10	11	12	13
No. of cycles		1–8	9–15	16–20	21–25	26–30	31–35	36–40	41–45	46–50		51–55	56–60	61–65	66–80
No. of reflections		6127	6127	18532	18532	18532	18532	18532	18532	18532		18532	18532	18532	18532
No. of atom sites		388	441	444	444	462	472	479	479	435		481	485	485	502
No. of protein atom sites		387	387	370	383	383	388	389	389	345		391	393	393	394
No. of sulfate ions		0	1	1	1	1	1	1	1	1		1	1	1	1
No. of water molecules		0	48	68	68	73	78	84	84	84		84	86	86	102
Minimum resolution (Å)		10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0		10.0	10.0	10.0	10.0
Maximum resolution (Å)		1.5	1.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0		1.0	1.0	1.0	1.0
R factor at start (%)		30.8	21.8	23.4	19.3	18.2	16.8	16.4	16.4	19.1		17.9	16.5	15.7	18.9
R factor at end (%)		23.3	18.4	19.0	17.7	17.1	16.4	16.4	16.0	18.8		16.5	16.2	15.5	14.7
X-ray/geometry weighting		1.0	1.0	1.0	2.5	2.5	2.5	2.5	—	—		2.5	2.5	2.5	2.5
Mean temperature factor (Å^2)															
Main-chain atoms		8.9	8.3	8.1	7.7	7.7	7.9	7.9	8.0	7.6		8.0	8.0	7.6	7.8
Side-chain atoms		14.8	16.3	11.0	12.1	12.6	13.2	13.8	14.1	8.9		11.5	11.8	12.1	12.5
Overall protein		11.6	12.0	9.4	9.7	10.0	10.3	10.7	10.8	8.2		9.6	9.8	9.7	10.0
Overall water			29.0	34.9	30.8	33.3	34.9	21.6	20.8	20.1		19.6	19.2	18.9	38.8
Fe atom		7.0	5.8	5.5	5.3	5.2	5.2	5.2	5.3	5.4		5.2	5.2	5.1	5.1
Total (protein + solvent)		11.6	12.0	9.3	13.1	13.9	14.4	12.7	12.7	10.6		11.4	11.5	11.5	15.9
Mean deviations from ideality	Target										Target				
1–2 distances (Å)	0.020	0.030	0.030	0.014	0.027	0.022	0.022	0.021	0.106	0.080	0.060	0.036	0.033	0.031	0.035
1–3 distances (Å)	0.040	0.096	0.094	0.039	0.048	0.040	0.042	0.042	0.109	0.086	0.065	0.043	0.040	0.038	0.040
Planar groups (Å)	0.020	0.036	0.034	0.018	0.023	0.020	0.020	0.019	0.039	0.038	0.030	0.016	0.015	0.015	0.017
Chiral volumes (Å^3)	0.200	0.338	0.430	0.196	0.225	0.222	0.219	0.236	0.484	0.335	0.300	0.232	0.220	0.188	0.210
Torsion angle (ω) (°) of peptide plane	5.0	3.3	3.5	22.9	23.4	3.7	3.6	3.4	5.3	6.1	5.0	3.2	2.7	2.7	3.1

Step 3. Cycles 16–20. Positional parameters were refined in the first three cycles, then *B* values in the next two cycles. This (3 + 2) pattern was used in all of the following steps. The resolution was extended from 1.5 to 1.0 Å, increasing the number of X-ray observations from 6127 to 18532. The occupancies of a total of 17 side-chain atoms on residues Lys2, Lys3, Glu12, Glu17 and Thr21 were set to zero as they had no, or poor, density in the Fourier synthesis after step 2. The number of water molecules was increased to 68, and the *R* factor was 19.0% after this step. In this and the next cycle the r.m.s. deviation in the ω-peptide torsion angle increased dramatically from 3.5 to over 20°. This is totally explained by the behaviour of Pro26. This residue was described as being in a *cis* configuration in the coordinate set 3RXN, and was stereochemically restrained as such in steps 1 to 4. In spite of this restraint the introduction of the 1.0 Å data allowed the X-ray data to invert the residue to the correct *trans* configuration.

Step 4. Cycles 21–25. This step was similar to step 3, with the matrix parameter increased from 1.0 to 2.5 to relax the stereochemical restraints. The occupancies of all atoms, except N, of the carboxy terminal Ala52 were set to zero as there was very poor electron density for this residue. The *R* factor fell to 17.7%, and the deviation of the model from the target stereochemistry increased slightly from that with the tighter restraints, Table 2.

Step 5. Cycles 26–30. The model was rebuilt to include the protein atoms omitted in steps 3 and 4, with the exception of four atoms (C*A*, C, O and C*B*) of residue Ala52. The number of waters was raised to 73. The *R* factor fell to 17.1%.

Step 6. Cycles 31–35. The missing atoms of residue 52 were built into weak density in the difference Fourier synthesis. The number of waters rose to 78. The *R* factor fell to 16.4%.

Step 7. Cycles 36–40. Asp was substituted for Thr at position 21, as indicated by the sequence of the gene (Voordouw, 1988). Thus, there were now no protein atoms missing from the model. The positions of all water molecules were checked in a difference Fourier synthesis calculated with phases from which the waters had been omitted. It was decided to assign variable occupancies to the water molecules on the basis of their temperature factors. Those with *B* values less than 20 were left at full occupancy; those with *B* between 20 and 30 were given an occupancy of 0.75, between 30 and 40 a value of 0.5 and with *B* greater than 40 an occupancy of 0.25. The previous *B* values were multiplied by 1.0, 0.9, 0.75 and 0.5 for the four categories respectively. The number of water sites was now 84. The *R* factor remained at 16.4%.

The current phase of the restrained refinement was deemed to have converged at this point.

Restrained model

At this stage we assess the quality of a model refined with stereochemical restraints, and identify those regions of the structure which are problematical in terms of either high thermal mobility or disorder. The geometry is close to that required by the restraints, Table 2. The *R* factor is 16.4% for the 18532 X-ray data measured between 10.0 and 1.0 Å. The model contains all 389 protein atoms with unit occupancy, one Fe atom, one sulfate ion and 84 water sites. The σ_A plot defined by Read (1986) and

the plot of final R factor against resolution (Luzzati, 1952) are shown in Figs. 3(*a*) and 3(*b*) respectively. From the former, the upper limit for the error in the coordinates is estimated as 0.13 Å.

The Ramachandran plot (Ramakrishnan & Ramachandran, 1965) of the observed values of the conformational angles (φ, ψ) for the restrained model are shown in Fig. 4. Only glycine residues have values of the (φ, ψ) angles outside the allowed regions shown in the plot. The wide scatter of the (φ, ψ) values throughout the allowed regions reflects the lack of α and β secondary structure elements in rubredoxin.

In spite of the high quality of the rubredoxin crystals, with measurable diffraction to about 1.0 Å, a very low packing density V_m of 1.68 Å^3 dalton^{-1} and an R factor of 16.4%, several disordered areas can be identified in the model. As a function of residue number, the average main- and average side-chain B values are shown independently in Fig. 5(*a*), the average observed electron densities of these are in Fig. 5(*b*), and the surface area of each residue accessible to solvent, calculated according to Kabsch & Sander (1983), is in Fig. 5(*c*). In Fig. 5(*c*) the exposed surface area was first calculated for the molecule abstracted from its crystalline environment, and secondly including all the contacts with neighbouring molecules within the crystal. The results shown in Fig. 5(*b*) are very similar to those given in Fig. 5 of Adman *et al.* (1991). It is clear that eight residues (Met1, Lys2, Lys3, Glu12, Glu17, Asp21, Asp31 and Ala52) have substantial disorder, reflected in high B values and low electron density: all are located on the surface of rubredoxin, well away from the FeS_4 cluster. Only these residues contain atoms with B values greater than 20, with the exception of a single atom at the end of each of the long side chains of Tyr4 and Glu50. In the restrained model the average B value for main-chain atoms is 7.9. At first sight that of the side chains is much higher, 13.2. However, if the side chains of the eight residues listed above as disordered are excluded, the side-chain average drops to 9.4, close to that of the main chain. The larger average B value mainly reflects the contribution of a small number of disordered side chains.

These 'disordered' residues not only lie on the surface of the globular rubredoxin in the free molecule, but all have significant surface area exposed to the limited solvent region in the crystal, Fig. 5(*c*). The disordered parts are not directly involved in the tight packing of the protein molecule in the crystal lattice. In the isolated molecule, removed from its crystal environment, many more side chains are exposed to solvent, including some charged ones, and these might therefore also be expected to be mobile in solution.

The disorder is best illustrated by the electron density itself. Fig. 6 shows six regions of the density in the ($2F_o - F_c$) synthesis after cycle 40, contoured at the same two levels for each: 1.0 e Å^{-3} (1σ above the

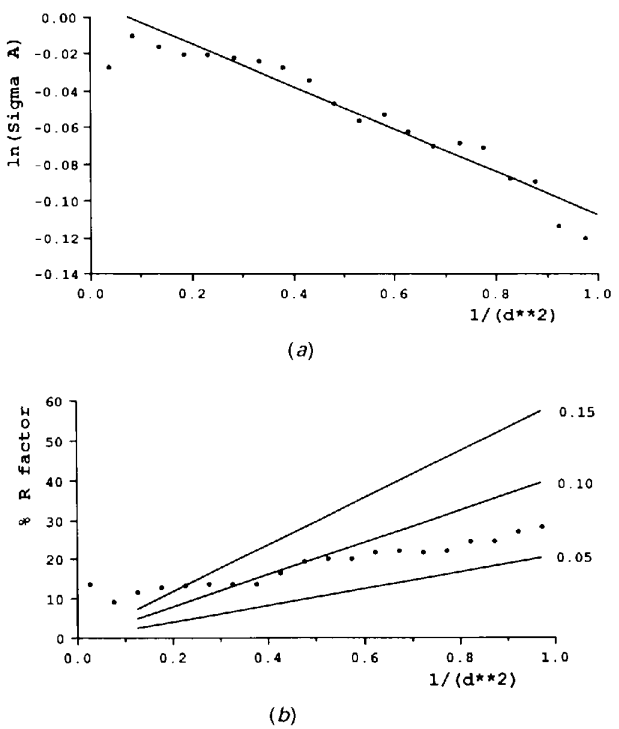


Fig. 3. (*a*) The Read σ_A plot and (*b*) the Luzzati plot of final R factor against resolution. The lines estimated for a series of average coordinate errors are shown in (*b*).

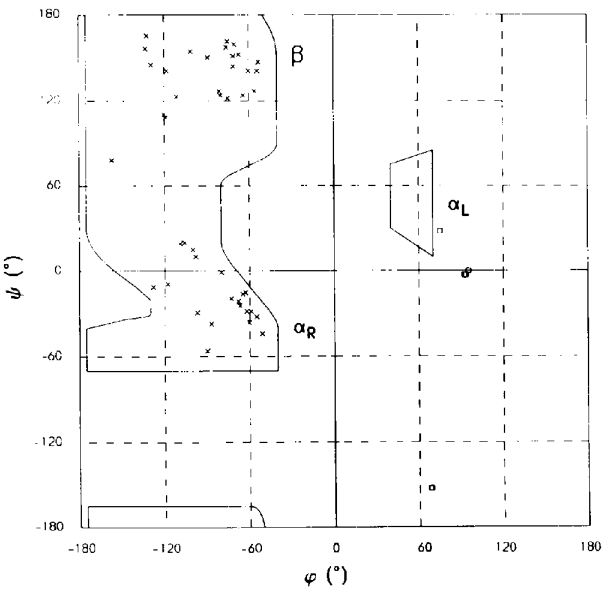


Fig. 4. The Ramachandran (φ, ψ) plot for the model after the first phase of the restrained refinement, cycle 40. Squares correspond to glycines, crosses to all other residues.

mean) and 2.3 e Å^{-3} (3σ). In this context we define σ as the root-mean-square deviation of the electron density from the mean value in the whole unit cell. $F(000)$ was included in the calculation giving a mean density of 0.35 e Å^{-3}. Figs. 6(*a*) to 6(*c*) show the density in three well defined regions: the FeS_4 cluster, Phe30 and Glu48 respectively. In Fig. 6(*a*) all atoms in the FeS_4 cluster are well resolved at both contour levels. In Fig. 6(*b*) the main chain and C*B* atoms are only resolved at the 3σ level. The atoms of the aromatic ring are not fully resolved even at the 3σ level, reflecting substantial concerted in-plane thermal vibration of these atoms. Thus, C*E* and C*Z* 'overlap' significantly more than C*G* and C*D*. Fig. 6(*c*) shows that for Glu48, one of the well ordered charged side chains, the atoms are both visible and resolved at the 3σ level right up to the carboxyl group.

Figs. 6(*d*) to 6(*f*) show the $(2F_o - F_c)$ density around three of the 'disordered' residues: Glu12, Glu17 and Asp31. For each there is no density at the 3σ level for atoms towards the end of the side chain, although all have density, albeit rather broad, at the 1σ level, indicating substantial disorder/thermal vibration in these side chains. Some of these regions are discussed in more detail after the results of the unrestrained refinement.

At the end of this restained refinement the overall impression is of a well ordered model with some disordered side chains, but including a single conformation for all protein atoms.

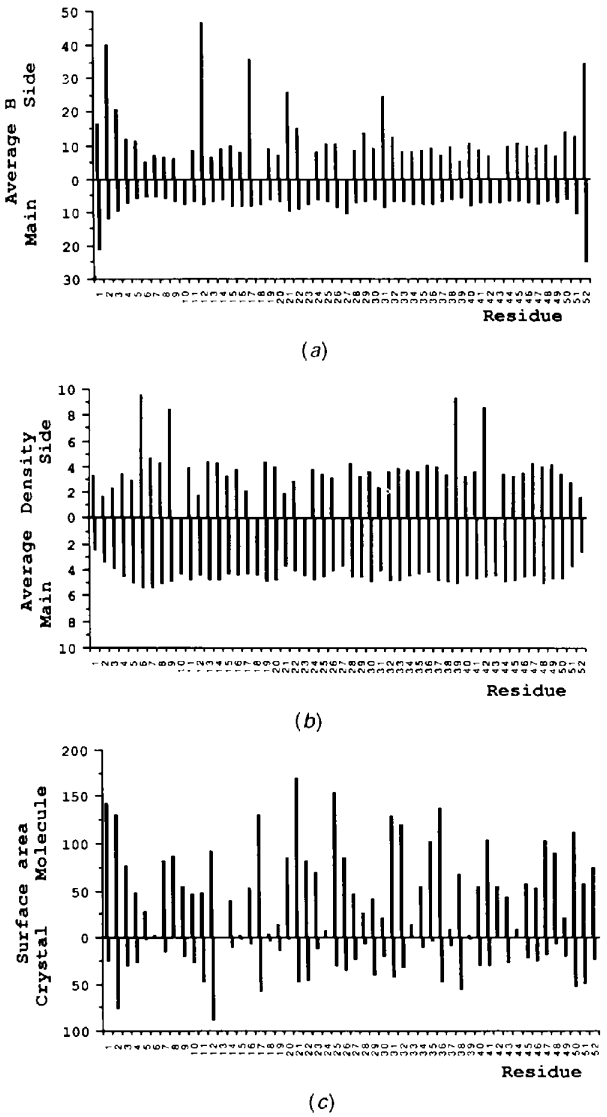


Fig. 5. (*a*) The *B* factors (Å^2) averaged independently for the main-chain and for the side-chain atoms; (*b*) the observed electron density at the atomic centres averaged over all atoms in units of e Å^{-3} in the $(2F_o - F_c)$ electron density synthesis, also for the main-chain and side-chain atoms separately; and (*c*) the surface accessible area (Å^2) for each residue in the isolated molecule and taking into account the contacts within the crystal lattice.

Features in the difference Fourier synthesis

In the difference Fourier synthesis after cycle 40, the maximum and minimum electron densities are 1.13 and -1.62 e Å^{-3}. The r.m.s. density in the map is 0.11 e Å^{-3}. The highest and lowest features are grouped around the FeS_4 cluster. In the $(2F_o - F_c)$ synthesis the maximum feature is the Fe atom with density 27.4 e Å^{-3} and the r.m.s. deviation in density was 0.67 e Å^{-3} from a mean value of 0.35 e Å^{-3}. Well ordered C atoms typically have peak electron densities of 3.5–4.0 e Å^{-3} in the map. The difference synthesis gives some indications of two sets of features absent from the model: H atoms and anisotropic thermal motion of the atoms.

In addition to the problem of the disordered side chains, the situation is not completely satisfactory with regard to the water sites. There are residual features at a low electron density close to the noise level of the difference Fourier synthesis after cycle 40. Many of these are in appropriate positions to take part in hydrogen-bonding networks. However, we did not feel confident in the significance of introducing these into the model as the sites would possess very low occupancy and/or high temperature factors. The problem with these waters may reflect the use of X-ray data from two crystals, one stored in a slightly different solution. We will return to these points in the final discussion.

Unrestrained refinement

Method

In the restrained model after cycle 40 there were 479 atoms giving a total of 1916 atomic parameters, with *xyz* positional and isotropic *B* values. There

were 18532 X-ray intensities giving a ratio of observations to parameters of almost 10:1. The ratio is 7.5:1 if observations less than $3\sigma(F)$ are excluded. This suggested that unrestrained refinement should be possible and we proceeded to perform this using the same program as above for calculating amplitudes, gradients and shifts (Baker & Dodson, 1980). The model was checked at the end of each unrestrained refinement step, Table 2, using the program *PROLSQ* (Konnert & Hendrickson, 1980), in order to obtain statistics on the stereochemistry for comparison with the restrained refinement. The refinement, although carried out without stereochemical restraints, uses only a diagonal approximation in place of the full least-squares matrix. The results obtained from these steps are now discussed.

Step 8. Cycles 41–45. The model after restrained refinement, cycle 40, was refined without stereochemical restraints, otherwise as in step 7, Table 2. As for preceding steps, three cycles of positional parameter refinement, followed by two of B values were performed. The R factor fell from 16.4 to 16.0%. The geometry, not surprisingly, diverged somewhat from the original target values.

Step 9. Cycles 46–50. The conditions were as for step 8. 44 protein atoms with B values greater than 15 in the model after cycle 45 were assigned zero occupancy, *i.e.* omitted from the current model. Atoms were thus omitted (always from the end of the side chain) from the following residues: Met1 (8), Lys2 (5), Lys3 (4), Tyr4 (1), Glu12 (4), Glu17 (4), Asp21 (4), Asn22 (3), Ser29 (1), Asp31 (3), Asp32 (1), Glu50 (2) and Ala52 (4). The number in brackets is the number of atoms left out of each residue. The final R factor was 18.8%. The omission of over 10% of the atoms of the protein increased the R factor by only 2.4%.

Stereochemistry of the unrestrained model

Those parameters monitored as deviating considerably from the target values by *PROLSQ* after step 8 (cycle 45) were inspected in detail. Excessive deviations were found for bond lengths, and other stereochemical parameters such as chiral volumes, almost exclusively in the eight residues in which there was indication of disorder from the restrained refinement, namely: Met1, Lys2, Lys3, Glu12, Glu17, Asp21, Asp31 and Ala52. In extreme cases, such as the side chain of Lys2, the deviation from stereochemical target values in some bond lengths was more than 1 Å: such side chains had become completely ill-defined in the unrestrained model and their geometry showed no relation to chemical reality (Pauling, 1960).

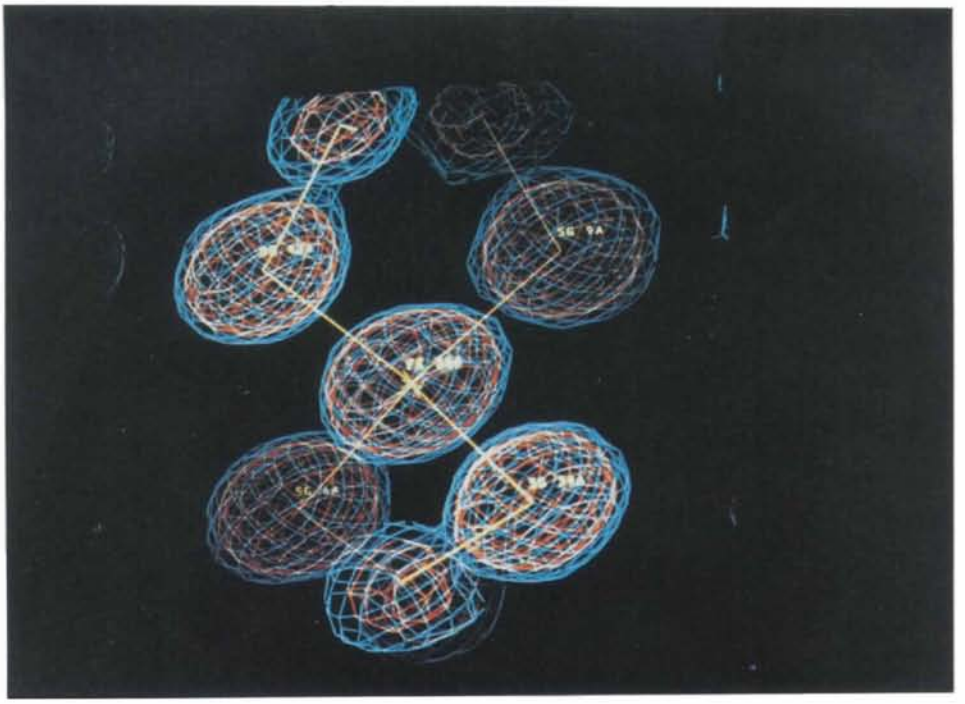
At this point two sets of statistics were calculated for the stereochemistry, Table 3. In the first all the protein atoms were included. In the second those 44 protein atoms with B values greater than 15.0 (roughly twice the average for the protein main chain) were excluded. It is clear that most of the increased deviation from the target values in comparison with the restrained-model values is due to the 44 'disordered' atoms with the high B values: these compose about 10% of the atoms in the protein. Target σ values for evaluating the stereochemistry of the coordinates of the unrestrained model were selected from the observed geometry of the remaining 345 protein atoms, and then applied to both the complete set of 389 atoms and the well ordered 345 atoms independently. The target σ values for the bond lengths was thus 0.06 Å. Fig. 7 is a histogram of the deviation in observed bond length from the library value divided by these target σ values, a statistical t plot. The histogram is plotted separately for the 389 and for the 345 atoms. If the target deviations were correct estimates of the errors and the average observed lengths equal to the library bond lengths then the standard deviation and mean value of the distribution in the t plot should be 1.0 and 0.0 respectively. For the 345 atoms the standard deviation is 0.98 and the mean value is 0.14. The value of 0.98 for the standard deviation suggests that the target value of 0.06 Å is a good estimate of an essentially normal distribution of errors in the bond lengths. The small (0.14) positive value of the mean indicates that the bond lengths refine to slightly longer than the target values in the stereochemical library used in the program by 0.009 Å on average.

There are several possible explanations for this small tendency towards apparently elongated bonds. Firstly, the library values based on small-molecule

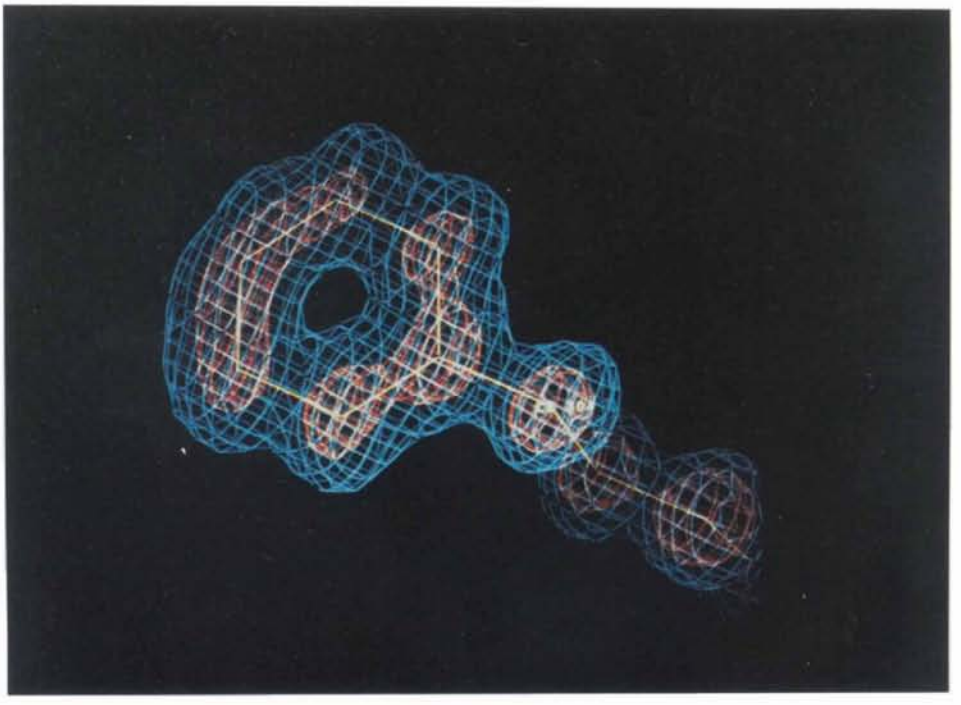
Table 3. *Observed deviations of the stereochemistry of the model after the unrestrained refinement step* 8 *(cycle* 45*) from the ideal values computed by PROLSQ*

The weight in each class of target values corresponds to $1/\sigma^2$. Two sets of statistics are given. The first is for all 389 protein atoms. The second is for those 345 well ordered protein atoms with temperature factors less than 15.0

	σ	Observed σ (389 atoms)	Observed σ (345 atoms)
Distances (Å)			
Bond lengths (1 2 neighbours)	0.060	0.106	0.059
Bond angles (1 3 neighbours)	0.065	0.109	0.063
Dihedral angles (1 4 neighbours)	0.090	0.095	0.087
Planar groups	0.030	0.039	0.032
Chiral volumes (Å^3)	0.300	0.484	0.316
Non-bonded contacts (Å)			
Single torsion contacts	0.500	0.166	0.161
Multiple torsion contacts	0.500	0.175	0.183
Torsion angles (°)			
Peptide plane (ω)	5.0	5.2	5.3
Staggered	15.0	15.5	9.9
Orthonormal	20.0	14.5	14.6

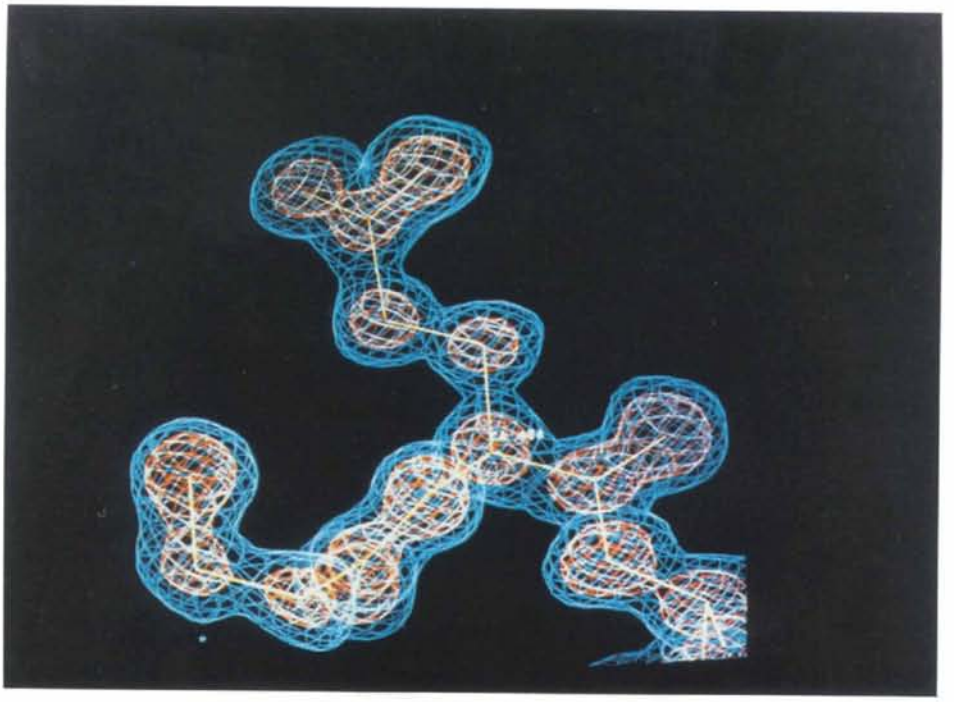


(*a*)

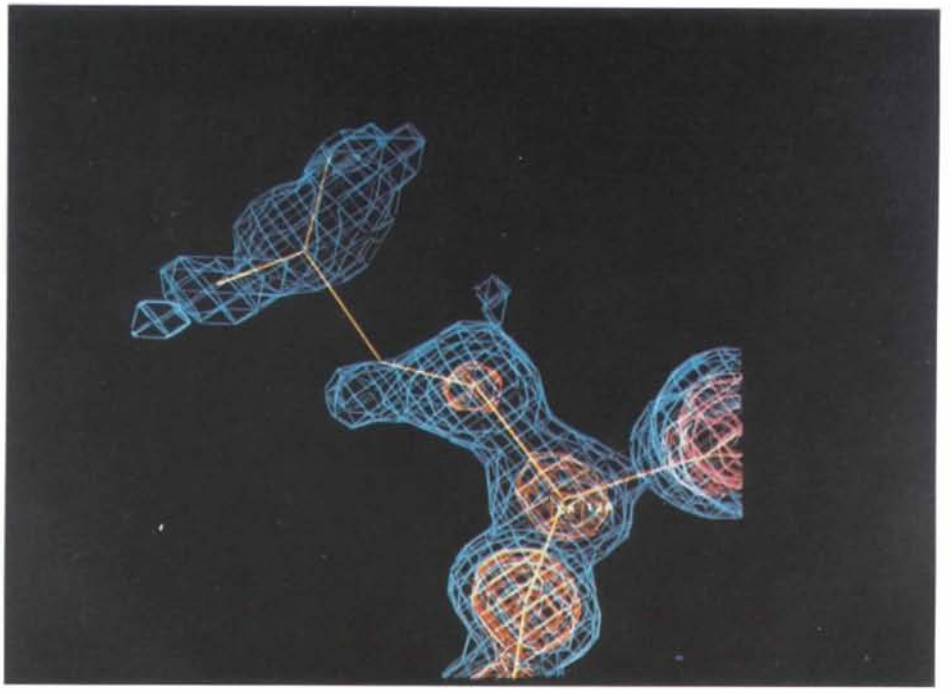


(*b*)

Fig. 6. The density in the $(2F_o - F_c)$ Fourier synthesis after the first phase of the restrained refinement, cycle 40. The blue contour corresponds to 1.0 e Å^{-3} (1σ above the mean density) and the red contour to 2.3 e Å^{-3} (3σ above the mean): (*a*) the region around the FeS_4 cluster, (*b*) the side chain of Phe30.

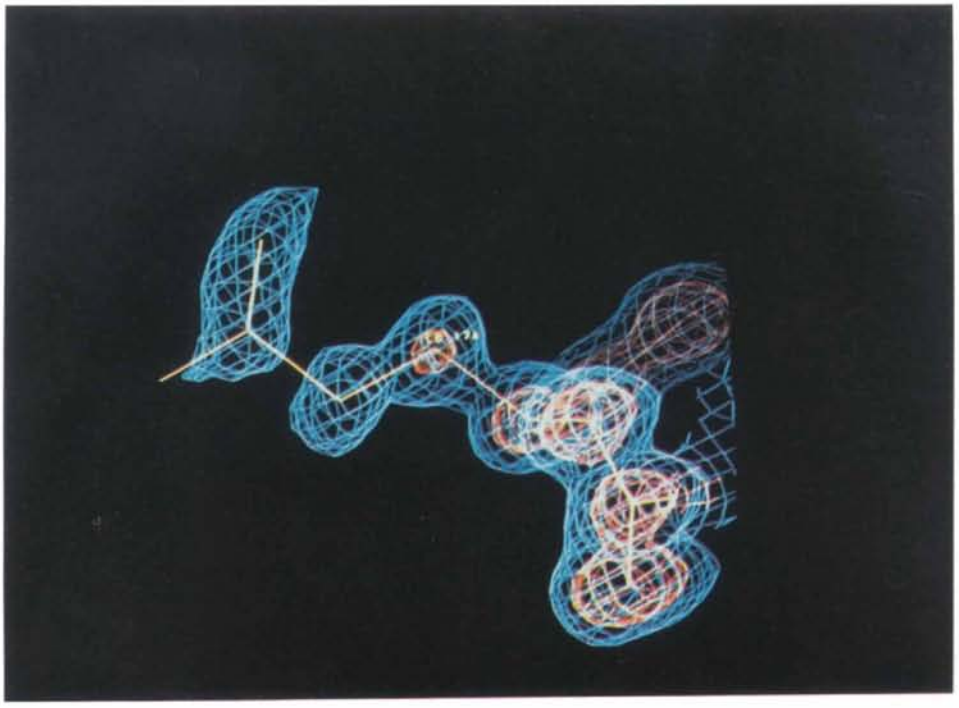


(*c*)

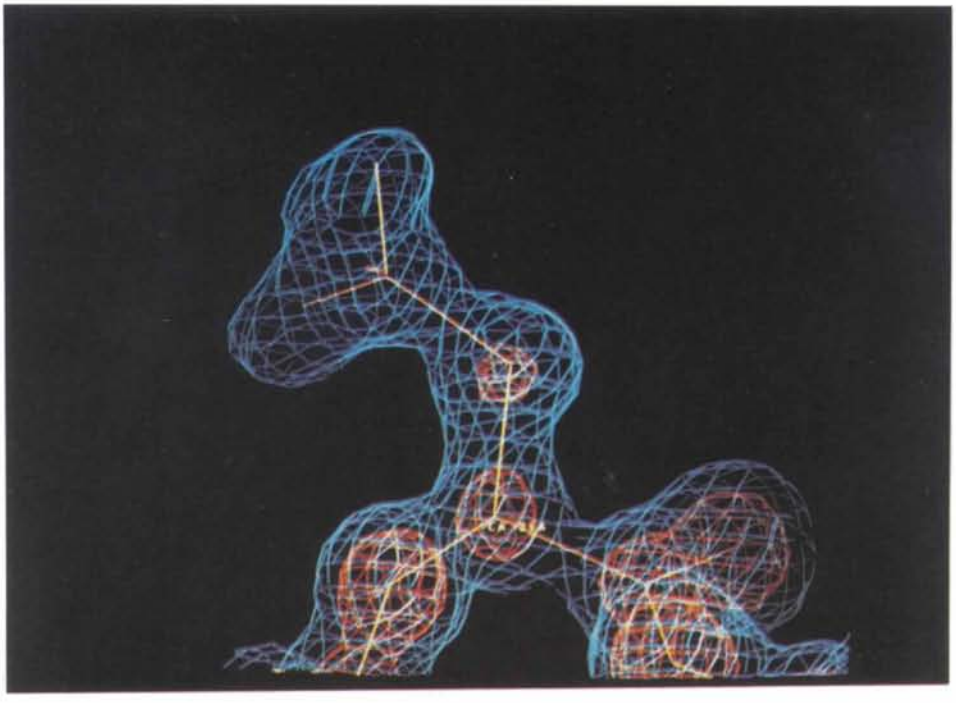


(*d*)

Fig. 6 (*cont.*) The density in the $(2F_o - F_c)$ Fourier synthesis after the first phase of the restrained refinement, cycle 40. The blue contour corresponds to 1.0 e Å^{-3} (1σ above the mean density) and the red contour to 2.3 e Å^{-3} (3σ above the mean): (*c*) Glu48, (*d*) Glu12.



(*e*)



(*f*)

Fig. 6 (*cont.*) The density in the $(2F_o - F_c)$ Fourier synthesis after the first phase of the restrained refinement, cycle 40. The blue contour corresponds to 1.0 e Å^{-3} (1σ above the mean density) and the red contour to 2.3 e Å^{-3} (3σ above the mean): (*e*) Glu17, (*f*) Asp31.

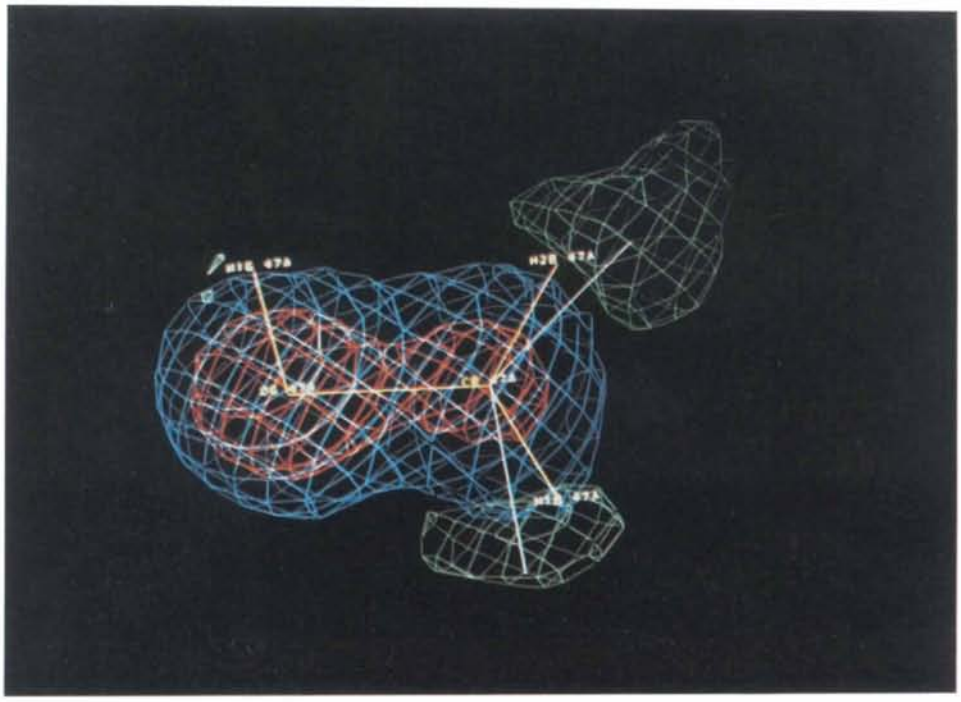


Fig. 11. The side chain of Ser47 viewed down the C*B*—C*A* bond, with the $(2F_o - F_c)$ and $(F_o - F_c)$ syntheses superimposed. The contour levels are 1.0 e Å^{-3} (blue, 1σ) and 2.3 e Å^{-3} (red, 3σ) in the $(2F_o - F_c)$ map and 0.25 e Å^{-3} (green, 2.5σ) in the difference map. The two alternative positions of the O*G* atom are shown in white. The calculated positions of the H atoms, not included in the refinement, from the single C*B*—O*G* conformation in the model are marked.

amino-acid and peptide X-ray analyses could be imperfect for macromolecules. Secondly, the cell dimensions could be overestimated by up to 1.5%. During data processing the relative values of the cell dimensions were refined, but there is not an absolute measure of their magnitude. However, the error is unlikely to reach such a level. In addition the cell dimensions recorded here are in reasonable agreement with those reported by the Seattle group on the basis of diffractometer measurements (Adman *et al.*, 1991). Thirdly, the omission of H atoms from the model could cause a small shift in the coordinates to compensate for their contribution to the electron density. This is referred to later after inclusion of H atoms at their calculated positions in the model. Fourthly, any remaining systematic errors in the model, particularly in surface residues, are likely to move atoms further apart than their true positions. Evidence for a contribution from the last point comes from the *t* plot for the complete 389 protein atoms where about 10% of the atoms are substantially disordered. After the unrestrained cycle 45 the standard deviation and mean from the *t* plot are 1.75 and 0.30 (+0.02 Å) respectively, again using 0.06 Å as a target σ value.

Disordered residues

From the results of the above restrained and unrestrained refinements it is clear that there are residues which are considerably disordered or highly mobile in the crystal structure. We therefore decided to perform one step (step 9) of unrestrained refinement discarding those 44 atoms described above which had *B* values higher than 15 after step 8. Detailed inspection of the difference Fourier synthesis after step 9, down to a level of 0.25 e $Å^{-3}$, allowed us to investigate the possible conformation(s) of these atoms without bias from their presence in poorly defined positions.

We now describe these poorly defined regions in some detail. Unless otherwise stated there is no clear evidence for more than one conformation.

(1) Met1. There was significant density for all of the atoms, but this was both broad and low suggesting considerable mobility. The density for all atoms (all atoms of the residue had been omitted) was greater than 0.5 e $Å^{-3}$. The residue was put back into the model.

(2) Lys2. There was no density at the 0.25 e $Å^{-3}$ level for any atom of the side chain. This side chain is the most disordered in the structure and all side-chain atoms were given zero occupancy.

(3) Lys3. There was low but significant density for the conformation of this side chain as it was after step 8. The side chain was reintroduced into the model.

(4) Tyr4. The OH oxygen was clearly visible and thus reintroduced.

(5) Glu12. There was evidence for at least two conformations of the side chain beyond C*B*. These were modelled in the broad density feature, each with half occupancy.

(6) Glu17. There was weak density for a single conformation and this was accepted.

(7) Asp21. As for Glu12, two conformations were identified in rather broad density, and added to the model.

(8) Asn22. The amide group was clearly visible in its previous conformation.

(9) Ser29. Density for the single omitted O*G* was evident.

(10) Asp31. There was very diffuse density for the carboxyl group. This was modelled as a single conformation, but in reality may well have more than this.

(11) Asp32. The single missing O atom reappeared in the difference density.

(12) Glu50. Both O atoms were in well defined density regions.

(13) Ala52. The C*B* density was very weak. There was broad density for the terminal carboxyl groups, extending into the solvent. One water was added at this point.

Neglecting the three residues (Tyr4, Ser29 and Asp32) which have only a single high *B* value atom, from the remaining ten 'disordered' residues, six in principle carry negative and three positive charge, the remaining Asn22 is neutral. However, the crystallization pH for the rubredoxin is 4.0. The p*K* values for the carboxyl groups of Asp and Glu amino acids in free solution are 3.9 and 4.2 respectively, which should be similar for the residues exposed to solvent on the surface of rubredoxin. This would result in a dynamic equilibrium between the charged and uncharged states of these side chains at pH 4.0, with

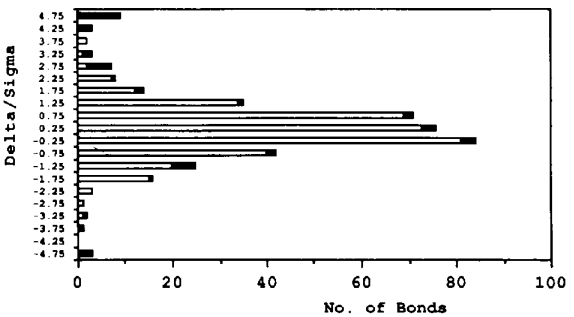


Fig. 7. The *t* plot of the distribution of the ratio between the actual deviations of bond lengths from the ideal values and the target values for deviations (σ) set in *PROLSQ*. The plot refers to the coordinates after step 8 (cycle 45) of the unrestrained refinement. The inner (empty) bars of the histogram correspond to the 345 protein atoms (with 361 bonds) with *B* values less than 15.0; the full length of the bars (empty plus full) to the complete protein model of 389 atoms (405 bonds).

multiple conformations. Omitting the relatively well ordered Asn22 and Glu50, there are seven disordered charged residues lying on the surface of rubredoxin on the opposite side of the molecule to the FeS_4 group, Fig. 8. They form a cluster where the intermolecular contacts are relatively loose in the crystal. The cluster is made up of residues Met1, Lys2, Lys3, Glu12, Glu17 and Ala52 from one molecule, and Asp21 and Asp31 from two adjacent molecules. Fig. 8 shows a view of four molecules down the *b* axis, related by translation of the whole unit cell along **a** and/or **c**. These molecules form part of a two-dimensional layer, with those generated by the twofold screw axes forming successive layers above and below. The diameter of the rubredoxin molecule is between 15 and 20 Å and corresponds to the full *a* and *c* cell dimensions, but only half of **b**. Thus, residues from the other layers do not contribute to the cluster of disordered side chains shown in the figure. In spite of a very low V_m such a mode of packing leaves some void within the layer which is filled with solvent.

The arrangement of rubredoxin molecules in the crystal resembles that of a hexagonal close packing of spheres. Each molecule is in contact with twelve symmetry-related molecules: six of them related by translations along *a* and *c* axes only [symmetry operations: (i) $x+1, y, z$; (ii) $x, y, z+1$; (iii) $x+1, y, z+1$; (iv) $x-1, y, z$; (v) $x, y, z-1$; (vi) $x-1, y, z-1$], six related by twofold screw axes, three with a positive translation along the *b* axis [symmetry operations: (vii) $-x, \frac{1}{2}+y, -z$; (viii) $-x, \frac{1}{2}+y, 1-z$; (ix) $1-x, \frac{1}{2}+y, 1-z$] and three analogous, with a negative translation along the *b* axis [symmetry operations: (x) $-x, -\frac{1}{2}+y, -z$; (xi) $-x, -\frac{1}{2}+y, 1-z$; (xii) $1-x, -\frac{1}{2}+y, 1-z$].

Summary of unrestrained refinement

The unrestrained least-squares minimization led to well defined stable positions for more than 90% of the protein atoms. The coordinates for these positions are amenable to refinement without stereochemical restraints with data of the present accuracy to 1.0 Å. The r.m.s. deviation in bond lengths of this part of the structure from the library values usually used for the stereochemistry is about 0.06 Å, suggesting positional errors for the C, N and O atoms of about 0.04 Å. These deviations are clearly substantially higher than those obtained after imposition of stereochemical restraints, by a factor of about three. However, they have the advantage of being only dependent on the experimental X-ray data.

In contrast, a number of the atoms with high *B* values, especially in the 'disordered' residues, did not behave well in unrestrained refinement. This may well reflect multiple conformations for these residues rather than merely high mobility, and it is unlikely that it would be possible to refine models for these without some form of local restraint even with higher resolution data. The pH value (4.0; close to the p*K* values of the side-chain carboxyl groups) of the rubredoxin crystallization medium probably exacerbates this problem.

Final refinement (restrained)

It was concluded that the complete model could not be presently refined without some stereochemical restraints. We chose to continue with restrained refinement but with the looser stereochemical restraints shown in Table 2 based on the analysis of the results of the refinement of the well behaved 90% of the atoms without restraints. A model was now assembled on the basis of the results of the restrained and unrestrained refinements. All protein atoms were included with unit occupancy with the following exceptions. Firstly, the side-chain atoms of Lys2 were given zero occupancy. Secondly, two conformations with half occupancy were used to model the C*G*, C*D*, O*E*1 and O*E*2 atoms of Glu12, and the C*G*,

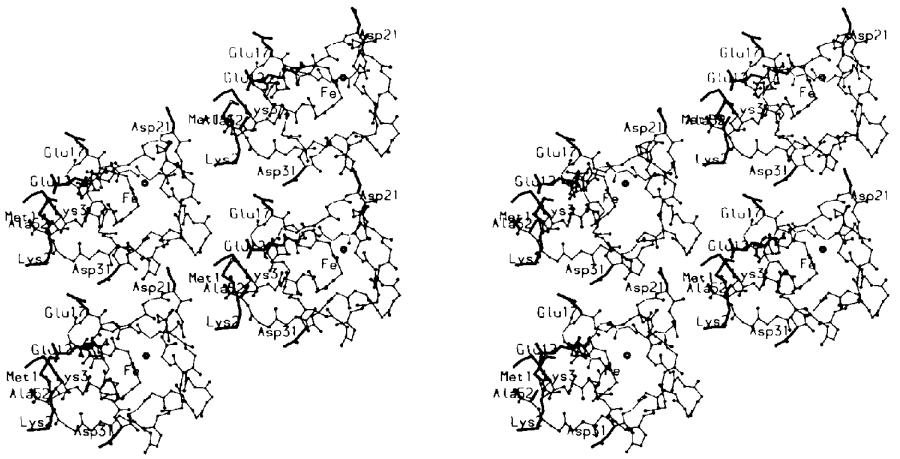


Fig. 8. Packing of molecules related by translation along **a** and **c**. Side chains are shown only for the disordered residues in thick lines. The disordered residues can be seen to form a cluster on the surface of the molecule opposite to the Fe atom.

*OD*1 and *OD*2 atoms of Asp21. The solvent structure was initially unchanged, with the assigned partial occupancies for the waters and the addition or deletion of one or two waters close to the disordered residues.

Step 10. Cycles 51–55. Three cycles of positional plus two of *B*-value refinement were performed as usual. The *R* factor was 16.5%.

Step 11. Cycles 56–60. Inspection of the difference Fourier synthesis after step 10 revealed defined positions for the C*B* and C*G* atoms of Lys2 at the 0.25 e $Å^{-3}$ level. These were added to the model. This left only the three atoms at the end of the Lys2 side chain with undefined positions and zero occupancy. Two waters were added. The *R* factor dropped to 16.2%.

Inspection of the difference Fourier synthesis after cycle 60 revealed four identifiable types of residual feature, all close to the noise level of the difference Fourier. The r.m.s. electron density was 0.1 e $Å^{-3}$ in this map. The maximum and minimum values were 1.1 and -1.6 e $Å^{-3}$. The four types of feature were as follows.

(1) Anisotropic thermal motion is most clearly illustrated by the difference electron density around the Fe atom and its chelating S atoms, Fig. 9(*a*). Here there is typical doughnut-like density around the Fe atom suggesting most of the vibration lies in the plane of the doughnut with much less in the direction normal to the plane. Around the Fe itself the peaks in the difference density are roughly 1.0 e $Å^{-3}$. These are the largest positive features in the map; the most negative minima also lie around the Fe. Similar, albeit lower, features are present around the S atoms. Evidence for anisotropic motion can also be seen around the lighter atoms. For example Fig. 9(*b*) shows the difference Fourier synthesis around the carbonyl O atom of Phe30 with typical bending rather than stretching motion predominating. Such features are especially evident around most of the carbonyl O atoms. In this and in Fig. 10 the calculated positions of the H atoms are shown, although at this point these were not included in the refinement. The presence of peaks corresponding to H atoms can be seen in this figure, leading on to the next point.

(2) Fig. 10 shows the difference density for Phe30 viewed in the same orientation as in Fig. 6(*b*). There is well defined electron density for the H atoms attached to the C*A*, C*B* and both C*D* atoms. The difference density at the further end of the ring suggests that here anisotropic motions of the C*Z* and the two C*E* atoms in the plane of the ring dominate: no density is evident for the C*E* or C*Z* H atoms, but density can rather be seen between the C*E* and C*Z* atoms and outside the C*E* positions, reflecting the in-plane vibration of the ring. This is in keeping with the density observed after cycle 40, Fig. 6(*b*). The peaks corresponding to many of the H atoms in the structure can be seen in this map. The contributions from the H atoms at their calculated positions were included in the structure-factor calculations in steps 12 onwards.

(3) Fig. 11 shows the difference density around the side chain of Ser47. Two small peaks, about 0.35 e $Å^{-3}$ in height, are present in the positions for the O*G* atom corresponding to the two *gauche* con-

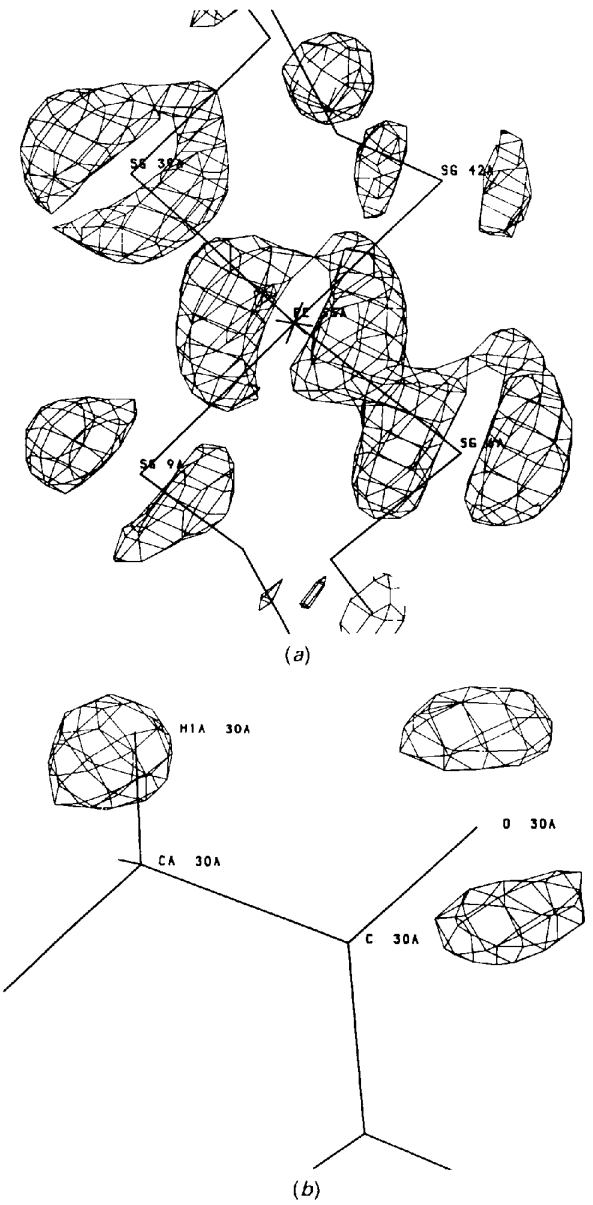


Fig. 9. Two regions of the difference Fourier synthesis after cycle 60 showing evidence for anisotropic thermal motion of the atoms. (*a*) The region around the FeS_4 cluster and (*b*) around the Phe30 carbonyl oxygen. In (*a*) the contour level is 0.3 e $Å^{-3}$ (3σ) and in (*b*) it is 0.2 e $Å^{-3}$ (2σ).

formations of the side chain. As can be seen from the predicted H*B* positions in the figure, the distance to the C*B* atom for both peaks is rather too long for a C—H, but fits very well as a C—O bond. Allowing for the usual 50% underestimation of density in the difference Fourier synthesis, the density for the additional O*G* positions should be 0.7 e Å^{-3}. The peak for the major conformation of the O*G* atom in the $(2F_o - F_c)$ map was 5.0 e Å^{-3}. This suggests a ratio of the occupancy of the major to that of the other two conformations of about 9:1:1. Evidence for similar low level multiple conformations can be seen at other positions in the chain. However, the densities for these minor conformations lie close to the r.m.s. level of the difference Fourier synthesis, presenting a cumbersome modelling problem, and have not been included in the final model.

(4) In the solvent regions many peaks are present at about the r.m.s. level of the map. Some were included in subsequent steps, provided they were in a position to take part in hydrogen-bonding networks.

Step 12. Cycles 61–65. (3+2) cycles of refinement were performed as before. H atoms were included at their calculated positions in this and the next step. The *R* factor dropped to 15.5%.

Step 13. Cycles 66–80. Three sets of (3+2) cycles of refinement were performed in this, the final step. A second position for the C*G* atom of Pro15 was introduced, representing the inverted pucker of the ring, with 1/4 occupancy. An additional 16 waters were added on the basis of the peaks in the difference Fourier synthesis. The *R* factor for the lowest resolution range up to 4.5 Å after cycle 65 was felt to be somewhat high (17.5%), and the occupancies of all water molecules were reset to unity. The results of this step confirmed that this improved the solvent model, as the *R* factor in the lowest range fell to 11.1%. The *R* factor finally converged at 14.7% (the value drops to 13.7% for the 16175 reflections with observed amplitude greater than 3σ).

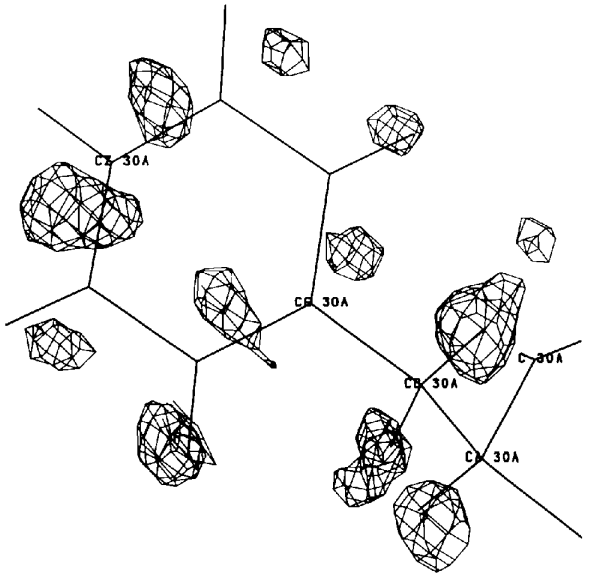


Fig. 10. The difference Fourier synthesis after cycle 60 around residue Phe30, contoured at a level of 0.2 e Å^{-3} (2σ). The computed positions of H atoms are shown, although these were not included in the refinement.

Table 4. *Weighting scheme and standard deviations at the end of the refinement, cycle* 80

The weight in each class of restraints corresponds to $1/\sigma^2$.

Distances (Å)		σ	Standard deviation	Number of parameters
Bond lengths	(1 2 neighbours)	0.060	0.035	414
Bond angles	(1–3 neighbours)	0.065	0.040	560
Dihedral angles	(1–4 neighbours)	0.090	0.062	157
Planar groups		0.030	0.017	70
Chiral volumes (Å^3)		0.300	0.210	55
Non-bonded contacts (Å)				
Single torsion contacts		0.500	0.165	145
Multiple torsion contacts		0.500	0.176	30
Torsion angles (°)				
Peptide plane (ω)		5.0	3.1	51
Staggered		15.0	11.6	55
Orthonormal		20.0	15.0	6

The statistics of the stereochemistry of the final model are summarized in Table 4. In the final $(2F_o - F_c)$ synthesis the electron density (in e Å^{-3}), estimated at the atomic positions, and averaged for particular atom types is for main-chain N 4.4, for C*A* and C 3.5, carbonyl O 4.6, C*B* 3.1, cysteine S*G* 13.1 and Fe 24.8. In the final difference map the maximum and minimum features were 1.09 and −1.62 e Å^{-3}, and the r.m.s. value 0.087 e Å^{-3}. There were no features above 0.4 e Å^{-3} except in the vicinity of the FeS_4 cluster.

Iron–sulfur cluster

As pointed out by previous authors (Stenkamp *et al.*, 1990; Adman *et al.*, 1991; and others referenced therein) the conformation of the iron–sulfur cluster, including the four cysteine residues, has a symmetry very close to C_2, with the local twofold axis relating Cys6 with Cys39 and Cys9 with Cys42. Adman *et al.* detail the unusual conformations of the residues making up the Fe-binding loops. Our results are in overall agreement with those reported previously. After cycle 80, least-squares superposition of Cys6 and Cys9 onto Cys39 and Cys42, using the twofold axis through the Fe atom, gives an average displacement of 0.10 Å and a maximum value of 0.18 Å for the N atom most distant from the Fe.

The geometry of the FeS_4 cluster, including the C*B* atoms of the cysteines, is shown in Fig. 12. This emphasizes the high degree of local C_2 symmetry. The pair of cysteines on the left-hand side of the

figure, Cys6 and Cys9, not related by the local twofold axis, differ in their conformational properties. For Cys6 the Fe—S bond length is 2.30 Å, the C*B*—S—Fe angle is 101° and the torsion angle C*A*—C*B*—S—Fe is about 180°. In contrast for Cys9 the values are 2.27 Å, 111 and about 270°. As can be seen from Fig. 12 the conformation of Cys39 (values 2.30 Å, 99 and about 180°) is closely related to that of Cys6; that of Cys42 (values 2.27 Å, 110 and about 270°) relates to that of Cys9. The torsion angles for the cysteines bonded to metals follow the observations of Chakrabarti (1989) as described by Adman *et al.* (1991). The Fe—S distances were loosely restrained to a target value of 2.30 Å in the final refinement. However, the values after unrestrained refinement cycle 50 were essentially identical; Cys6 2.30, Cys9 2.27, Cys39 2.30 and Cys42 2.26 Å. The bonding and torsion angles for the cysteines were also very similar.

The features seen in the final difference Fourier indicating significant anisotropy for the thermal vibrations of the FeS_4 cluster, Fig. 9(*a*), do not obey the local twofold symmetry. Instead the 'doughnuts' lie in the *ac* crystal plane, by chance perpendicular to the Fe—S39 bond. This may indicate that they are to some extent an effect of the packing in the crystal rather than an intrinsic property of the molecules.

Comparison with previous models

During the course of refinement the average change of C*A* positions between the starting model [3RXN data set of Adman *et al.* (1977) in the Protein Data Bank (Bernstein *et al.*, 1977)] and the final model was 0.20 Å with the highest values being 1.1 Å for Pro26 and more than 0.5 Å for the first two N-terminal amino acids. As mentioned above, Pro26 was present in the 3RXN coordinate set in the *cis* conformation and its peptide group flipped to *trans* during the restrained refinement without manual intervention in spite of being declared as *cis* in the restraints.

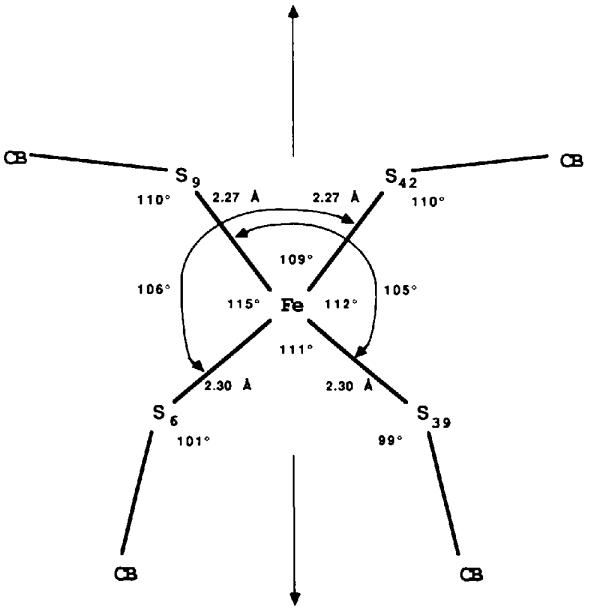


Fig. 12. The conformation around the iron–sulfur cluster. The local twofold axis is shown and is vertical in the plane of the figure.

Table 5. *Classification of the water molecules in the final model, after cycle* 80, *according to their B values* (Å^2)

The electron density (e Å^{-3}) in the corresponding $(2F_o - F_c)$ synthesis averaged over the waters within each range is also given.

B-value range	Number	Electron density
10–20	12	3.5
20–30	21	2.2
30–40	24	1.6
40–50	21	1.4
50–60	14	1.2
60–70	9	1.1
70–80	1	1.0

The protein model resulting from the present refinement and that at 1.5 Å resolution (Adman *et al.*, 1991, coordinate set 7RXN) are essentially identical. Least-squares overlap of the C*A* atoms of the two models gave an average displacement of 0.061 Å with a maximum difference of 0.25 Å for the C-terminal Ala52 and a similar value of 0.22 Å for the N-terminal Met1. Omitting those two terminal residues from the superposition reduced the average displacement to 0.053 Å and left a maximum of 0.14 Å for Asp21, one of the rather poorly defined residues (see above). A similar analysis superimposing all main-chain atoms gave an average displacement of 0.069 Å; omitting the two terminal residues reduced this value to 0.060 Å.

In addition, the conformation of all side chains is practically identical. The single conformations of residues Glu12, Pro15 and Asp21 in the 7RXN model correspond closely to one of the two conformations of each of these residues present in the current model. Two atoms, C*G*1 and C*G*2, of Val24 are labelled in the opposite way in the two models. We preserve our labelling for this residue as it conforms with the chirality assumed by the *PROLSQ* program.

The close similarity of the protein model is not reflected in the modelling of the solvent structure. Adman *et al.* (1991) quote 180 water sites of varying occupancy and one sulfate ion: the total water content is equivalent to 98 molecules. In the model presented here we included the sulfate ion and 102 water sites, all assigned full occupancy. It is somewhat surprising that in such a small tightly packed protein crystal only 12 waters had *B* values less than 20 Å^2, Table 5.

The high degree of similarity of the two models, independently refined at 1.5 and 1.0 Å resolution,

enhances confidence in the accuracy of the resulting atomic parameters. The average displacement of about 0.06 Å between well defined atoms agrees well with the error estimated by the analysis of bond-length discrepancies in the model refined without restraints.

Concluding remarks

Diffraction data have been collected for the small metalloprotein, rubredoxin, to a nominal resolution of 1.0 Å, a comparable resolution to that often attained for 'large' small-molecule structures. The structure has been refined at this resolution, first with stereochemical restraints, then for a small number of cycles with the X-ray data only. These refinements showed that about 90% of the protein atoms are highly ordered, with B values not much higher than one might expect for a small molecule. However, 44 of the total of 389 protein atoms have B values greater than an (arbitrary) value of 15.0. These are clearly highly mobile or disordered. In only three cases was there clear evidence for two conformations and these were introduced into the final model. The 'disordered' residues form a cluster on the surface of the protein in the crystal. The unrestrained refinement gives a model for the 'ordered' 345 atoms which has r.m.s. deviations of the bond lengths from the library values used in the restrained refinement of 0.06 Å. This would suggest an r.m.s. error in coordinates for these atoms of about 0.04 Å. With the present data it is possible to refine the 'good' part of the protein structure free from stereochemical restraints, resulting in an acceptable geometry. However, the disordered or mobile regions require some form of local restraint. This is likely to remain valid even if the quality of the data is improved or the resolution extended.

The final model was thus refined with relatively loose chemical restraints. The latter were set on the basis of the deviations from ideality observed for the good part of the structure in the unrestrained refinement. The final model included the H atoms in their calculated positions, two alternative conformations for three residues, 102 water molecules and a sulfate ion. The water molecules were all assigned unit occupancy, although some had very high B values. As suggested by Adman *et al.* (1991) there is space for about 100 water molecules in the asymmetric unit. These coordinates have been deposited.*

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 8RXN, R8RXNSF), and are available in machine-readable form from the Protein Data Bank at Brookhaven. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37055 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

The errors in the final protein coordinates can be estimated in various ways. The σ_A plot gives an upper bound of 0.12 Å. The behaviour of the 'good' part of the protein in the unrestrained refinement suggests a value close to 0.04 Å, for these atoms at least. Superposition of the present model onto the 7RXN coordinates gives a discrepancy of 0.06 Å for the main-chain atoms, suggesting a coordinate error again around 0.04 Å. After the unstrained refinement the bond lengths were on average 0.009 Å longer than the library values. For the final model, after inclusion of the hydrogen contributions, the bonds were in contrast 0.005 Å shorter than the library values. This is substantially smaller than the estimated errors.

In the difference Fourier synthesis at the end of the refinement features are present suggesting three sources of residual error through omission of the following from the model: anisotropic thermal vibration, water molecules with low occupancy and low occupancy alternative conformations for some side chains. However, with the exception of the features around the Fe atom, these features are all of similar height, about $1–3\sigma$ (0.1–0.3 e $Å^{-3}$) of the density in the synthesis.

However, in the light of the possible differences in mother liquor between the two crystals and in particular the high mosaicity of the crystal used to collect the high-resolution data we have deferred such a detailed analysis until we have the opportunity to recollect the data on fresh crystals. We then plan to carry out a more extensive least-squares refinement of the model, including the above features. The need for more accurate data is evidenced by Fig. 1(*a*): the merging $R(I)$ factor rises to about 20% at the high 1.0 Å resolution limit. This could be considerably reduced by prolonging counting times during data collection. Together with the use of fresh stable crystals we expect to be able to extend the resolution beyond 1.0 Å.

Nevertheless the present data have not only provided a protein structure with most of the atoms resolved as discrete peaks in the electron density maps, but have been used in a direct-methods phasing of the protein with no chemical information being required (Sheldrick, Dauter, Sieker & Wilson, manuscript in preparation). The advantages of analyzing at least a small number of proteins at atomic resolution is self evident. The refinement of these structures without stereochemical restraints will provide us with increased understanding of their conformational behaviour and perhaps with better models for those which are limited to lower resolution.

We thank Dr I. Moura and Dr J. Le Gall at the Department of Biochemistry, University of Georgia, Athens, GA, for the sample of rubredoxin used in

this study. We thank Dr E. Adman for the 7RXN coordinates of the structure of *Desulfovibrio vulgaris* rubredoxin refined at 1.5 Å resolution prior to their release from the Brookhaven Protein Data Bank.

References

ADMAN, E. T., SIEKER, L. C. & JENSEN, L. H. (1991). *J. Mol. Biol.* **217**, 337–352.
ADMAN, E. T., SIEKER, L. C., JENSEN, L. H., BRUSCHI, M. & LE GALL, J. (1977). *J. Mol. Biol.* **112**, 113–120.
AGARWAL, R. C. (1978). *Acta Cryst.* A**34**, 791–809.
BAKER, E. N. & DODSON, E. J. (1980). *Acta Cryst.* A**36**, 559–572.
BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MAYER, E. F. JR, BRYCE, M. D., RODGERS, J. R., KENNARD, O., SIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.
CHAKRABARTI, P. (1989). *Biochemistry*, **28**, 6081–6085.
DAUTER, Z., TERRY, H., WITZEL, H. & WILSON, K. S. (1990). *Acta Cryst.* B**46**, 833–841.
FREY, M., SIEKER, L. C., PAYAN, F., HASER, R., BRUSCHI, M., PEPE, G. & LE GALL, J. (1987). *J. Mol. Biol.* **197**, 525–541.
JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
KABSCH, W. & SANDER, C. (1983). *Biopolymers*, **22**, 2577–2637.
KONNERT, J. H. & HENDRICKSON, W. A. (1980). *Acta Cryst.* A**36**, 344–350.
LESLIE, A. G. W., BRICK, P. & WONACOTT, A. J. (1986). *CCP4 Newsl.* **18**, 33–39.
LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.
PAULING, L. (1960). *The Nature of the Chemical Bond*, 3rd ed. New York: Cornell Univ. Press.
PIERROT, M., HASER, R., FREY, M., BRUSCHI, M., LE GALL, J., SIEKER, L. C. & JENSEN, L. H. (1976). *J. Mol. Biol.* **107**, 179–182.
RAMAKRISHNAN, C. & RAMACHANDRAN, G. N. (1965). *Biophys. J.* **5**, 909–933.
READ, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
SERC Daresbury Laboratory (1979). *CCP4. A Suite of Programs for Protein Crystallography*. Daresbury Laboratory, Warrington, England.
STENKAMP, R. E., SIEKER, L. C. & JENSEN, L. H. (1990). *Proteins Struct. Funct. Genet.* **8**, 252–264.
VOORDOUW, G. (1988). *Gene*, **67**, 75–83.
WATENPAUGH, K. D., SIEKER, L. C., HERRIOTT, J. R. & JENSEN, L. H. (1972). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 359–367.
WATENPAUGH, K. D., SIEKER, L. C., HERRIOTT, J. R. & JENSEN, L. H. (1973). *Acta Cryst.* B**29**, 943–956.
WATENPAUGH, K. D., SIEKER, L. C. & JENSEN, L. H. (1979). *J. Mol. Biol.* **131**, 509–522.
WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
WILSON, K. S. (1989). *Synchrotron Radiation in Structural Biology*, edited by R. M. SWEET & A. D. WOODHEAD, pp. 47–54. New York, London: Plenum Press.

Acta Cryst. (1992). B**48**, 59–67

Structure of Monoclinic Papain at 1·60 Å Resolution

BY R. W. PICKERSGILL,* G. W. HARRIS AND E. GARMAN†

Protein Engineering Department, Protein Crystallography and Biomolecular Computing Section, AFRC Institute of Food Research, Reading Laboratory, Shinfield, Reading RG2 9AT, England

(*Received* 26 *July* 1990; *accepted* 3 *June* 1991)

Abstract

X-ray diffraction data to 1·60 Å resolution have been collected from monoclinic crystals of papain. The monoclinic model was derived from the orthorhombic one by molecular replacement, X-ray restrained molecular-dynamics simulation and least-squares refinement. Refinement against 1·60 Å data produced a model with reasonable stereochemistry and an *R* factor of 16·0%. The X-ray structures of orthorhombic and monoclinic papain are compared. The two structures are similar, the r.m.s. deviation between the two structures is 1·049 Å (mean difference 0·531 Å). The monoclinic model is shown to shift considerably (r.m.s. 3·083 Å) during refinement which indicates that bias due to the starting model may reasonably be expected to be low; in addition, solvent structure is independently determined. Tightly bound solvent occupies the same position in both structures and weakly bound solvent structures (high temperature factors) are different. Differences in protein structure are attributable to different crystal contacts, different covalent modification of the active-site cysteine, or different interpretation of weak density. The temperature factors for both structures show similar trends.

Introduction

Papain is one of the cysteine proteases found in the latex of *Carica papaya*. The single polypeptide chain of 212 amino-acid residues is cross linked by three disulfide bridges. The monoclinic crystals reported here diffract to high angle and the structure of the orthorhombic form has been determined by multiple isomorphous replacement (Drenth, Jansonius, Koekoek, Swen & Wolthers, 1968). The orthorhombic

* To whom correspondence should be addressed.
† Laboratory of Molecular Biophysics, Rex Richards Building, University of Oxford, Oxford OX1 3QU, England.

0108-7681/92/010059-09$03.00