Atomic Resolution (0.94 Å) Structure of *Clostridium acidurici* Ferredoxin. Detailed Geometry of [4Fe-4S] Clusters in a Protein[†,‡]

Zbigniew Dauter,[§] Keith S. Wilson,[§] Larry C. Sieker,[∥] Jacques Meyer,[⊥] and Jean-Marc Moulis*[,⊥]

Department of Chemistry, University of York, Heslington, York, YO1 5DD, England, Department of Biological Structure, SM/20, University of Washington, Seattle, Washington 98195, and CEA, Département de Biologie Moléculaire et Structurale, Laboratoire Métalloprotéines, 17 avenue des Martyrs, 38054 Grenoble Cedex 9, France

Received August 29, 1997; Revised Manuscript Received October 16, 1997[⊗]

ABSTRACT: The crystal structure of the 2[4Fe-4S] ferredoxin from *Clostridium acidurici* has been solved using X-ray diffraction data extending to atomic resolution, 0.94 Å, recorded at 100 K. The model was refined with anisotropic representation of atomic displacement parameters for all non-hydrogen atoms and with hydrogens riding on their parent atoms. Stereochemical restraints were applied to the protein chain but not to the iron-sulfur clusters. The final *R* factor is 10.03 % for all data. Inversion of the final least-squares matrix allowed direct estimation of the errors of individual parameters. The estimated errors in positions for protein main chain atoms are below 0.02 Å and about 0.003 Å for the heavier [4Fe-4S] cluster atoms. Significant differences between the stereochemistry of the two clusters and distortion of both of them from ideal T_d tetrahedral symmetry can be defined in detail at this level of accuracy. Regions of alternative conformations include not only protein side chains but also two regions of the main chain. One such region is the loop of residues 25−29, which was highly disordered in the room temperature structure.

Metal active sites in proteins display an extremely wide diversity of structures and compositions. This diversity is amplified in many cases by the variety of different functions that can be assumed by proteins containing a given metal prosthetic group. This arises from the essential and multifarious effects of the polypeptide chains on the reactivities of their metal sites (*1*).

An example of such interactions is provided by proteins containing [4Fe-4S] clusters, which have been under intense scrutiny over the past three decades (*2*). Considering only proteins for which crystallographic models have been produced, these clusters are found in electron transfer proteins (*3*−*7*, *8* and references therein), in various enzymes (*9*−*14*), in DNA binding motifs (*15*) and in suspected oxygen-sensing sites (*16*). The intrinsic physicochemical properties of [4Fe-4S] clusters have been investigated in considerable detail through synthetic analogues (*17*). In addition, the role of the polypeptide chain as tuner of the cluster reactivity has been studied in a vast number of proteins (*18*, *19*).

Despite the powerful combination of these approaches, important gaps still limit our knowledge of the biological function of [4Fe-4S] clusters. For instance, the means by which the proteins exert their influence on these metal centers are not well understood. In particular, whereas the geometry of the clusters in synthetic compounds has been assessed in great detail (*17*, *20*−*24*), similar data are not available for proteins, due to the lack of atomic resolution X-ray structures [1.5 Å at best (*3*)]. Thus, the structural constraints imposed by the proteins on their [4Fe-4S] centers have so far remained undetermined, except in the very few cases in which changes in iron coordination occur (*9*). In recent years, significant progress has been made in the determination of protein structures at atomic resolution (1.2 Å or better). For instance, the geometries of FeS_4 and ZnS_4 active sites in rubredoxins have been determined with a precision approaching that obtained with synthetic analogues (*25*, *26*).

We report here the atomic resolution (0.94 Å) structure of a protein containing two iron-sulfur clusters, namely that of the 2[4Fe-4S] ferredoxin from *Clostridium acidurici* (henceforth CauFd).[1] This analysis at unprecedented resolution clarifies uncertainties in some features of the polypeptide chain as reported in a previous 1.84 Å resolution structure of the same protein (*28*). More importantly, for the first time in any protein, the geometries of the [4Fe-4S] clusters have been determined with a precision that allows comparisons with analogous features in synthetic model compounds and analysis of the strains imposed by the polypeptide chain.

MATERIALS AND METHODS

Crystallization and Data Collection. Crystals of CauFd were grown as described earlier (*28*). For data collection, a single crystal was immersed in mother liquor containing 25% glycerol for about 10 s and then transferred in a fiber loop to a stream of cold nitrogen gas at 100 K. Data were collected at the wiggler EMBL beam line BW7B at DESY,

[†] This work was supported in part by the EC BIOTECH programme (Contracts CT92-0524 and CT96-0189 to Z.D. and K.S.W.). CEA (France) and EMBL (Hamburg) awarded visiting fellowships to L.C.S.

[‡] The coordinates and structure factors have been deposited in the PDB and have received ID codes 2FDN and r2FDNSF, respectively.

* Corresponding author, CEA/Grenoble, DBMS-MEP, 17 rue des Martyrs, 38054 Grenoble Cedex 9, France. Tel: 33 476885623. Fax: 33 476885872. E-mail: moulis@ebron.ceng.cea.fr.

[§] University of York.

[∥] University of Washington.

[⊥] CEA.

[⊗] Abstract published in *Advance ACS Abstracts,* December 1, 1997.

[1] Abbreviations: CauFd, *Clostridium acidurici* [the former species name *Clostridium acidi-urici* has been changed to *C. acidurici* (27)] ferredoxin; NOE, nuclear Overhauser effect; ADP, atomic displacement parameters.

S0006-2960(97)02155-7 CCC: $14.00

Table 1: Crystallographic Data

space group	$P4_32_12$
cell (Å)	$a = 33.95$, $c = 74.82$
beam line	EMBL Hamburg, wiggler BW7B
wavelength (Å)	0.883
detector	MAR Research Image Plate
T (K)	100
resolution (Å)	20.0−0.94
reflections, unique	28 084
reflections, Friedels separate	47 452
completeness (%)	98.7
R (I) merge (%)	7.3
$I/\sigma(I)$ overall	15.0
$I/\sigma(I)$ outer shell	2.0

Hamburg, using a MAR Research 300 mm image plate scanner. To cover the whole range of intensities, three sets of images were recorded, differing in the exposure times and the resolution limit. The diffraction images were processed, and reflection intensities integrated with the program suite DENZO/SCALEPACK (*29*), keeping the Friedel related reflections separate. Table 1 shows a summary of data collection. Data complete at 98.7% extending to 0.94 Å resolution were obtained, consisting of 47 452 independent intensities. The cell dimensions in space group $P4_32_12$ estimated from the postrefinement procedure are $a = 33.95$ Å and $c = 74.82$ Å. The estimated accuracy of the dimensions was 0.1%. The crystals contain about 35% solvent and have a V_M of 1.85 Å^3/Da (*30*). The overall temperature factor estimated from the Wilson plot (*31*) based on the data from the frozen crystal is 5.1 Å^2.

Refinement. The starting model was that of entry 1FDN (*28*) from the Protein Data Bank (*32*). Only the protein atoms were accepted, without water molecules. The initial R factor was high, 48%, reflecting the use of a room temperature starting model with slightly different cell dimensions. Several cycles of isotropic refinement were carried out with the program REFMAC (*33*, *34*) coupled with automatic selection of solvent water atoms by ARP (*35*). At this stage, no attempt was made to model the multiple conformations of protein side chains. After the initial isotropic refinement converged, R factor 14.6%, refinement was continued with the program SHELXL-96 (*36*) using the F^2 as observables with no σ cutoff. Atomic displacement parameters (ADP) of all atoms of the model were converted to anisotropic. Hydrogen atoms were included at the calculated positions with their ADP isotropic and 20% (or 50% for methyl groups) higher than those of their parent atoms. The conjugate gradient option was used throughout anisotropic refinement, except at the end, when convergence had been achieved (see below). Restraints were applied to the stereochemical parameters (*38*) as well as to the ADP; their weights were default values within SHELXL-96. At this resolution with 10-fold excess of observables (47 452) over refined parameters (4980) the effect of geometrical restraints on the well-defined parts of the model is small (*25*); they are nevertheless necessary to hold the poorly defined parts of the model in a stereochemically acceptable geometry. The contribution of the solvent continuum was taken into account using the SHELXL SWAT instruction.

Rounds of anisotropic refinement with SHELXL-96 were interspersed by graphics sessions (QUANTA, Molecular Simulations Inc., San Diego, 1997) when alternative conformations of some protein regions were built and water molecules checked and rejected or appended to the model on the basis of $3F_o - F_c$ and $F_o - F_c$ electron density maps. The occupancies of alternative conformations were refined constraining their sum to unity. At this stage, all waters were assigned unit occupancies. When the refinement converged and no more significant features were identifiable in the maps, a round equivalent to three complete cycles of blocked-matrix refinement was performed with blocks containing positional parameters of about 30 residues or anisotropic displacement parameters of about 15 residues. The successive blocks were overlapped by two residues. Subsequently, a round of similar cycles was run with all restraints removed and simultaneously all shifts suppressed by damping them to zero. This allowed rigorous estimation of the errors associated with each independent refined parameter from the inverse least-squares matrix. This protocol has been established in the refinement of previous atomic resolution models (e.g., *26*, *37*, *39*, *40*) and therefore R-free cross validation was not used. Instead of global validation, the features in the electron density maps and their chemical interpretability were accepted as the ultimate validation. Indeed, since the atomic resolution of this analysis is similar to that of small crystal structures and the ratio of observations to parameters was about ten to one, it rendered the use of R-free unnecessary.

At this stage, it was realized that the positional errors of the cluster iron and sulfur atoms were as small as 0.002 Å. To confirm that the clusters were defined with such accuracy, the following procedure was employed. The positions of the cluster atoms (8 Fe and 8 S) were randomized ten times by 0.3 Å. Each of the ten models was independently refined to convergence by ten cycles of conjugate gradient minimization. The results concerning the geometry of the clusters were analyzed statistically (Table 2). The accuracy of the cluster stereochemistry estimated from the spread of positional parameters between the 10 "randomized" refinements is comparable to that obtained from the least-squares matrix inversion.

Table 2: Bond Lengths within the [4Fe-4S] Clusters[a]

	cluster I		cluster II	
Fe1−S2	2.295(3)	2.295(2)	2.293(3)	2.294(2)
S3	2.291(3)	2.290(2)	2.247(3)	2.247(3)
S4	2.257(3)	2.258(4)	2.298(3)	2.300(3)
Fe2−S1	2.310(3)	2.310(2)	2.325(3)	2.326(4)
S3	2.264(3)	2.261(3)	2.292(3)	2.294(3)
S4	2.307(3)	2.307(2)	2.278(3)	2.277(2)
Fe3−S1	2.274(3)	2.272(2)	2.242(3)	2.242(2)
S2	2.273(3)	2.275(4)	2.312(3)	2.311(2)
S4	2.302(3)	2.302(2)	2.304(3)	2.306(3)
Fe4−S1	2.262(3)	2.262(3)	2.283(3)	2.283(2)
S2	2.266(2)	2.267(2)	2.240(3)	2.239(4)
S3	2.284(3)	2.285(2)	2.299(3)	2.301(2)
Fe1−Fe2	2.760(2)	2.759(2)	2.730(2)	2.731(2)
Fe3	2.719(2)	2.719(3)	2.722(2)	2.724(2)
Fe4	2.698(2)	2.698(2)	2.690(2)	2.692(4)
Fe2−Fe3	2.779(2)	2.778(2)	2.752(2)	2.753(1)
Fe4	2.713(2)	2.713(3)	2.710(3)	2.712(2)
Fe3−Fe4	2.685(2)	2.686(2)	2.717(2)	2.718(2)
Fe1−SG8 (37)	2.265(2)	2.267(2)	2.274(3)	2.274(4)
Fe2−SG11 (40)	2.271(3)	2.272(3)	2.279(3)	2.278(3)
Fe3−SG14 (43)	2.233(3)	2.230(2)	2.259(3)	2.259(2)
Fe4−SG47 (18)	2.252(3)	2.251(2)	2.277(3)	2.275(4)

[a] Values from least-squares matrix inversion (columns 1 and 3) and those estimated from the average of 10 randomized refinements (columns 2 and 4), see text. Distances in angstroms are accompanied by their last digit standard uncertainties in brackets.

Table 3: Refinement Statistics

program	SHELXL-96
restraints	ref 38 for protein, none for [4Fe-4S] clusters
*wR*2 (%)	25.26
R factor, all reflections (%)	10.03 (47 452 reflections)
R factor, $F > 4\sigma(F)$ (%)	8.66 (37 200 reflections)
goodness of fit	1.00
Flack *x* parameter	0.055 (±0.009)

Table 4: Deviations from Target Stereochemistry

	su	contributors
bond distance (Å)	0.027	452
angle distance (°)	0.038	609
planar 1−4 distance (Å)	0.052	130
peptide planar groups (Å)	0.035	55
aromatic planar groups (Å)	0.017	2
chiral volumes (Å^3)	0.360	57
torsion angles, planar (°)	6.3	55
torsion angles, staggered (°)	15.6	58
torsion angles, orthonormal (°)	33.3	2
B factors, main chain, bond (Å^2)	0.78	274
B factors, main chain, angle (Å^2)	1.03	318
B factors, side chain, bond (Å^2)	1.97	178
B factors, side chain, angle (Å^2)	2.32	291

During inspection of the Fourier syntheses, it was clear that many water molecules lay too close together to have full occupancies and that they represent sites from overlapping and alternative solvent networks. Those waters generally had weaker associated electron density and some tended to drift away from the map peaks during refinement. It was decided to try to refine the occupancies of all water molecules. This refinement was stable, in spite of refining the occupancies simultaneously with the ADP. The tendency to move away from the density peaks also diminished. This stage was accepted as the end of refinement, and finally blocked matrix least-squares refinement without restraints or shifts (as described above) was performed to obtain standard uncertainties of refined parameters and derived quantities, such as bond lengths and angles and torsion angles. The refinement results are summarized in Table 3 and the details of the model stereochemistry are given in Table 4.

RESULTS AND DISCUSSION

Refined Model. The atomic model of CauFd was refined against 0.94 Å resolution data recorded at 100 K, consisting of 47 452 independent reflections (Friedel mates not merged), Table 1. The final *wR*2 factor (based on F^2) is 25.26%, final *R*1 factor (based on *F*) is 10.03% on all 47 452 reflections and 8.66% on 37 200 reflections with $F_o > 4\sigma(F_o)$, Table 3. Since there are 8 iron atoms/protein molecule and Friedel related reflections were not averaged, the data contained a significant anomalous signal and the Flack *x* parameter (*41*) refined to a value of 0.055 with $\sigma = 0.009$, providing clear enantiomorph distinction. *x* should refine to either 0.0 for the correct enantiomorph or 1.0 for the opposite one.

The model contains the following: 454 protein atoms (726 with hydrogens), 8 Fe and 8 inorganic S atoms, and 94 water sites. As many are partially occupied, the stoichiometric water content is 71 molecules. Several protein atoms are in addition split into two alternative sites as discussed below.

Overall Fold. As a reminder of the overall fold of this protein, the current model is shown in stereo in Figure 1. Briefly, the polypeptide chain is tightly wrapped around the

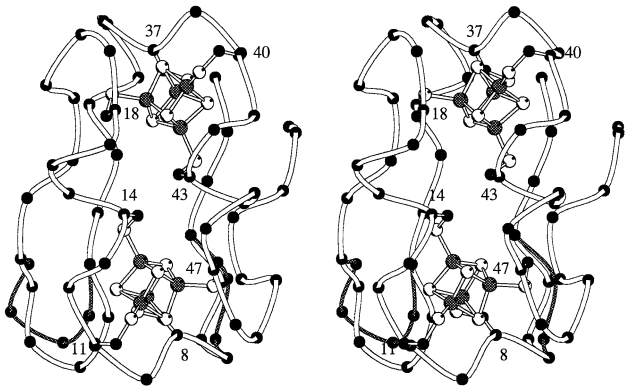

FIGURE 1: Protein fold, schematically shown as a CA coil with $[Fe_4S_4]$ clusters and coordinating cysteine side chains. The second conformation of the main chain is shown in dark gray. Figure drawn with MOLSCRIPT (*59*).

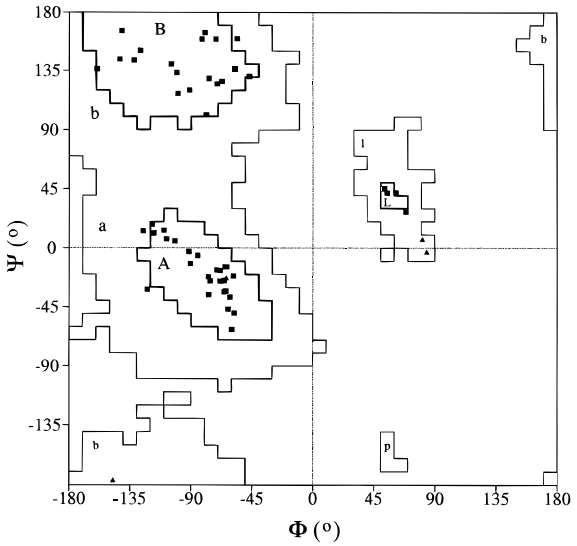

FIGURE 2: Ramachandran plot, calculated with PROCHECK (*60*).

two [4Fe-4S] clusters, which are each coordinated by four cysteines. The estimated errors in φ, ψ, and ω are typically smaller than 1° except in parts of the chain with two conformations (residues 5−6 and 25−29) where they rise to 3°. The Ramachandran plot (Figure 2) shows that only three residues (Asp28, Thr36, and Asp39) lie slightly outside the most favored regions of conformational space. In spite of CauFd containing essentially no *bona fide* α-helices or β-sheets, the residues nevertheless show a strong preference to lie within the corresponding regions of the plot.

Only the most salient features of the structure revealed by the present study are described in the following sections.

Accuracy of the Model and Comparison to 1FDN. A spectacular improvement in the resolution, from 1.84 Å for 1FDN (*28*) to 0.94 Å (this work), has been achieved by the combined use of cryogenic conditions and synchrotron radiation. This is the highest resolution achieved to date for a ferredoxin.

The unit cell volume is lower for the present cryogenic study at 100 K than that of 1FDN at room temperature (*28*) by more than 2%. This shrinkage from 88 700 to 86 240 Å^3 clearly reflects the effect of freezing but is not isotropic as the *c*-axis shows practically no difference. This is largely due to reduced thermal motion, as exemplified below by the freezing of the residue 25−29 loop which is only visible at low temperature. In addition, only 46 water molecules were

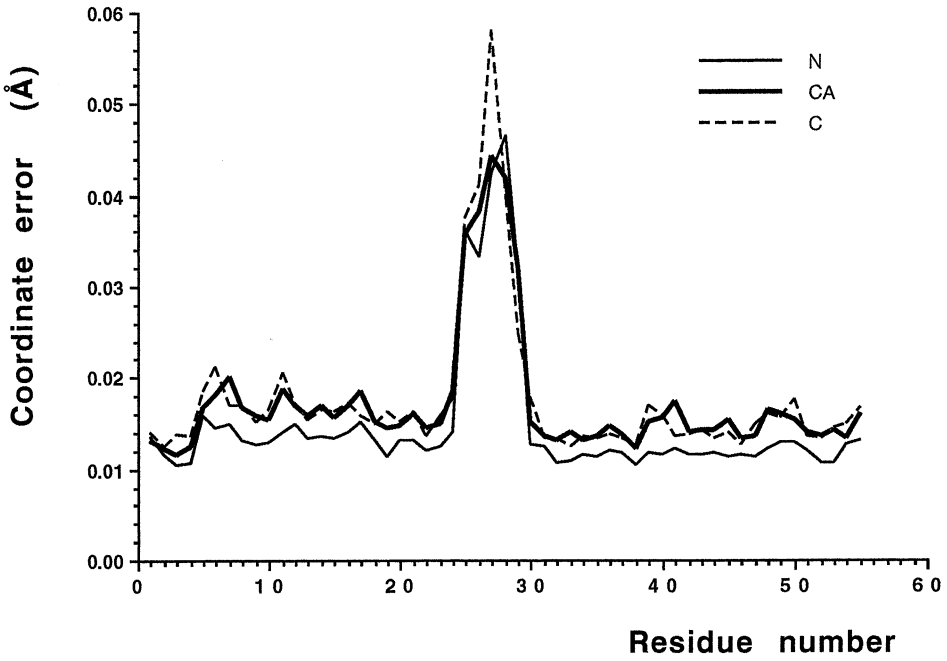

FIGURE 3: Individual atomic coordinate errors for main chain atoms estimated by least-squares matrix inversion. For regions 5−6 and 25−29, the values for only one conformation (A) are shown; much larger errors reflect the flexibility and partial occupancy of these regions.

located in the 1FDN model at 1.84 Å resolution, compared with the 94 water sites associated with the present model. Freezing the crystal clearly has the effect of not only reducing the average temperature factors of the protein atoms but also ordering the solvent. This should be borne in mind when comparing the detailed structure at 0.94 Å against the 1.84 Å model.

The accuracy of the present model can be judged by several indicators. The Ramachandran plot was already mentioned above. Errors in refined parameters were estimated by two methods, first rigorously from inversion of the least-squares matrix and second by a set of 10 parallel refinements of the atoms in the clusters after randomization. The latter is discussed in more detail in the cluster section below. The estimated coordinate errors of the main atoms are plotted as a function of residue number in Figure 3. Over most of the sequence, the coordinate errors of the main chain atoms are in the range 0.01−0.02 Å, only in the disordered loop 25−29 they rise up to 0.10 Å. Note that these error estimates are made truly independently for each parameter. The errors in the distances involving the heavy atoms of the clusters are spectacularly low, around 0.002−0.003 Å (Table 2). These are 1 order of magnitude smaller than the errors expected from a restrained-type refinement at more typical resolution for a protein (above 1.5 Å). Such accurate values for protein parameters might be expected to allow the creation of a stereochemical target library from protein data alone as previously suggested (*40*).

The deviations from the idealized library target values (*38*) in Table 4 are greater than the weights often (incorrectly) applied in refinements at lower resolution. This supports the notion that many low- and medium-resolution refinements are actually overrestrained. In addition, inclusion of hydrogen atoms allows the other atoms to refine more properly into their own positions; they no longer refine to the center of gravity of, e.g., CH or CH_2 groups. Indeed, a majority of the hydrogen atoms in the well-defined parts of the protein chain were visible in the difference Fourier synthesis (Figure 4).

The atomic resolution of the data permitted the refinement of displacement parameters of the water atoms together with their occupancies. This proceeded stably and led to meaningful results. Several waters refined to values higher than 0.9; their occupancies were subsequently fixed at 1.0. Others acquired occupancies less than 0.9; they clearly reflect the alternative H-bonding networks. The statistics of water occupancies are summarized in Figure 5.

Internal 2-fold Symmetry of 2[4Fe-4S] Ferredoxins. CauFd displays a high degree of internal 2-fold symmetry (*28*), the result of an internal sequence repeat arising from an ancestral gene duplication (*42*). Since the presence of two very similar but slightly different halves of the molecule is almost certainly important for its biological role (*43*, *44*), it is worth asking to what extent such noncrystallographic symmetry is preserved in the high-resolution model. The main chain carbons of the two halves belonging to 52 out of the 55 amino acids of the sequence (only residues 27, 28, and 34 have no counterpart) were superimposed with an rms of about 0.9 Å. This average difference in position decreases to 0.4 Å when the 32 most closely related CA atoms are superimposed. One (A) of the two conformations of the main chain around residues 24−29 (see below) superimposes better over the symmetry-related half of the molecule than does the other (B) conformation.

Thus, the internal symmetry seen at lower resolution holds quite well at the higher accuracy reached in this study. The main chain fragments which do not correspond well between the two halves of the molecule are the segments linking the clusters (residues 15−17 and 44−46), residues 23 and 52, and a peripheral short segment (residues 24 and 25 and 53 and 54).

Alternate Conformations and Disordered Regions. In the 1FDN model, there was no clear electron density for residues 27 and 28, associated with the turn of a β-strand. Improved

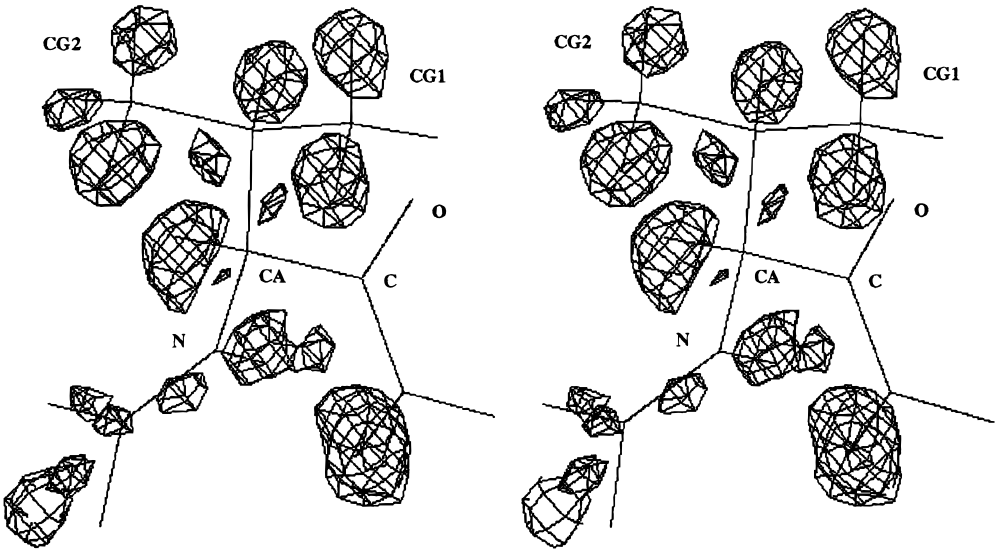

FIGURE 4: Residue Val46 with overlapped difference Fourier synthesis at 3σ level calculated for the model with all hydrogen atoms of this residue omitted.

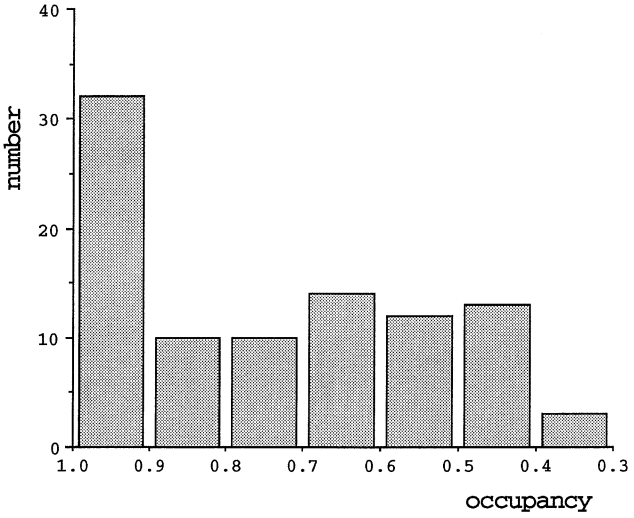

FIGURE 5: Histogram of site occupancies for solvent water molecules.

resolution and cryogenic conditions allow the definition of two conformations for this region (Figure 1). These two folding patterns of the main chain have essentially equal occupancy. They start to diverge by a roughly 40° difference in the Ser25 ψ angle and reconverge onto a single conformation at Tyr30. The gap between main chain atoms is up to 2.5 Å for residues 27 and 28. Such difference is obviously too large to escape detection at 1.84 Å resolution, and the lack of clear electron density in this region in 1FDN was probably due to dynamic disorder which is frozen out at 100 K. Although these two conformations are well defined at 0.94 Å, relatively few water molecules, all with partial occupancies and high B factors, could be associated with this exposed loop.

Due to insertions or deletions in the 25−29 loop in other comparable structures of ferredoxins, the only model available for comparison to CauFd was derived from NMR data on the homologous ferredoxin from *C. pasteurianum* where the β-turn composing this loop produces a relatively high number of NOE and appears well defined (*45*). The folding in solution approximately follows conformation B of the present crystal structure up to residue 25 and conformation A after residue 27. Gly26, which folds similarly in the two crystallographic conformations, appears to have a different ψ angle in the NMR solution study.

Table 5: Amino Acid Side Chains with Alternate Conformations[a]

	occupancy (%)	
	conformation A	conformation B
Ser 10	87	13
Glu 17	68	32
Asn 21	59	41
Ser 24	70	30
Val 31	85	15
Gln 54	86	14

[a] The estimated errors of the occupancies are less than 2%.

The main chain of residues 5 and 6 also assumes two slightly different conformations in a 3:1 ratio. In addition, five side chains involving external hydrophilic amino acids, as well as the exposed Val31, have been modeled with two alternative conformations with occupancies summed to unity (Table 5). Figure 6 shows the electron density for Asn21.

Hydrogen Bonds. The clear definition of residues 25−29 in the present high-resolution model defines additional hydrogen bonds compared to 1FDN (*28*). Some involve side chain atoms (Table 6). An interesting one, between Gly26 N and Arg29 O, reaches across the β-turn around residues 27 and 28 in conformation B.

As for all the stereochemical parameters, the lengths of all H bonds are more accurately defined in the 0.94 Å model. The distances between donor and acceptor atoms decrease on average when compared to those derived from lower resolution data (Table 6). This is partly related to the shrinkage of the molecule/crystal lattice in the cryogenic structure: the 2.8% shrinkage in volume corresponds roughly to a cell dimension change of 1.4%.

The H bonds involving the sulfur atoms of the active sites are presented in more detail below.

Geometry of the Clusters. The present work provides the first opportunity to examine the precise geometry of [4Fe-4S] clusters in a protein. The coordinates of the iron and inorganic sulfur atoms were refined throughout without any restraint applied to the clusters. These atoms were then moved by an rms of 0.3 Å from their refined positions, and

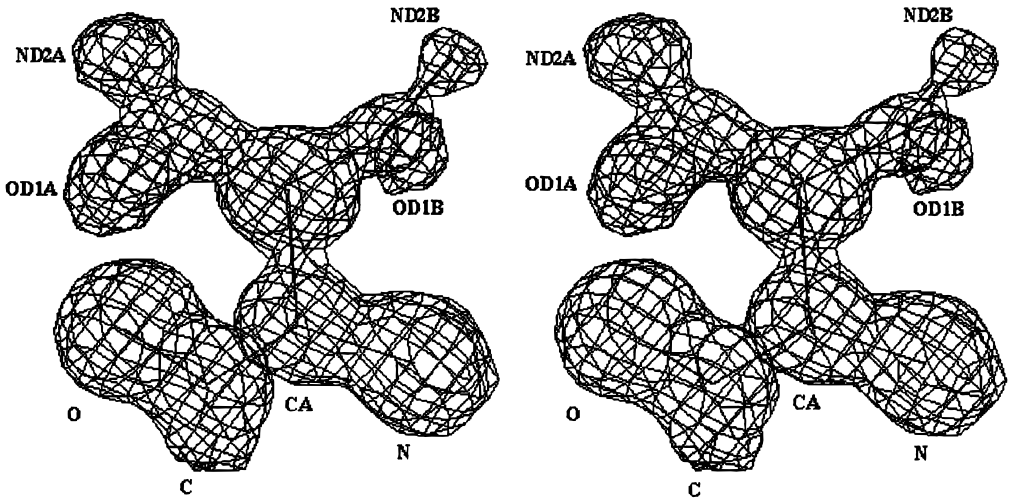

FIGURE 6: Asn21 with corresponding $3F_o - 2F_c$, α_c density at 1σ level showing alternate conformations of its side chain with occupancies ratio A/B of 60/40%.

the entire structure was again refined until convergence. This procedure was repeated 10 times and the data were then used to calculate mean values and empirical standard uncertainties for the intracluster distances (Table 2). The empirical standard deviations, between 0.002 and 0.003 Å, agree very well with those estimated from the matrix inversion, confirming that the latter are realistic. The spread among Fe−S distances in the clusters is up to 30−40 times these values. Thus, the geometry of the clusters is sufficiently accurate to allow meaningful comparisons with the structures of synthetic analogues.

Inorganic Cores. It has been documented that synthetic compounds of general formula $[Fe_4S_4(SR)_4]^{2-}$ often display a compressed tetragonal geometry along one pseudo-fourfold axis resulting in a local symmetry close to the D_{2d} point group (*20* and references therein). However, a few structures differing from this idealized description have been reported (*21*, *23*, *46*).

Cluster II, coordinated by cysteines 37, 40, 43, and 18, shows approximately tetragonally compressed geometry (Figure 7a). The compression axis passes through the (Fe2, Fe4, S1, S3) face of the cubane core containing the iron atoms coordinated by Cys40 and Cys18 on one side and the (Fe1, Fe3, S2, S4) face on the opposite side. Accordingly, the Fe−S distances distribute into four short ones (2.251 ± 0.015 Å) and eight long ones (2.302 ± 0.012 Å). This is borne out by the distances between inorganic sulfur atoms (average 3.615 Å) which clearly separate into two long distances of 3.676 Å (±0.004) and four short ones at 3.584 Å (±0.012). In contrast, the Fe−Fe distances follow a different trend: they are clustered around 2.722 Å (±0.019), with the Fe2−Fe3 distance being the longest (2.753 Å) and the Fe1−Fe4 distance being the shortest (2.692 Å). This indicates that the deviation from ideal T_d symmetry is actually a twist (compression and rotation) of the Fe_4 tetrahedron towards D_2 symmetry (Figure 7a).

In contrast to cluster II, cluster I does not display a clear-cut deviation from idealized T_d symmetry for its [4Fe-4S] core. Distances and angles are more tightly distributed around the mean than for cluster II and for most of the structurally characterized synthetic compounds. Only the Fe−Fe distances (2.726 ± 0.032 Å) are slightly more scattered. Cluster I constitutes another example, in addition

Table 6: Intramolecular Hydrogen Bonds[a]

(a) Main Chain to Main Chain

donor	acceptor	this work	1FDN	donor	acceptor	this work	1FDN
Val3 N	Val53 O	2.842	2.87	Val31 N	Ser24 O	2.881	2.91
Asn5 N(A)	Ala51 O	2.927	2.94	Asp33 N	Ala22 O	2.776	2.88
Cys8 N	Asn5 O(A)	2.969	3.32	Cys37 N	Asp33 O	2.965	3.08
Glu15 N	Gly12 O	3.292	3.42	Ala44 N	Gly41 O	3.236	3.23
Glu17 N	Cys14 O	2.980	2.89	Val46 N	Cys43 O	3.241	3.27
Cys18 N	Cys14 O	3.272	3.40	Cys47 N	Cys43 O	3.187	3.13
Cys18 N	Glu15 O	3.606	3.61	Cys47 N	Ala44 O	3.046	3.21
Asn21 N	Cys18 O	3.295	3.30	Asp50 N	Cys47 O	3.151	3.20
Ser24 N	Val31 O	3.026	3.09	Val53 N	Val3 O	2.909	2.88
Gly26 N(B)	Arg29 O(B)	2.670		Ala55 N	Ala1 O	2.975	3.07
Thr36 N	Asp33 O	3.064	3.00				
Gly45 N	Gly41 O	3.039	3.29				
Gly45 N	Ala42 O	3.105	3.02				
Val46 N	Ala42 O	3.187	3.24				

(b) Main Chain to Side Chain

donor	acceptor	this work	1FDN
Ala1 N	Asp39 OD1	2.699	2.88
Arg29 NH1	Ile4 O	2.799	2.78
Ala7 N	Asn5 OD1(B)	3.242	3.08
Asp35 N	Asp33 OD1	2.902	3.07

(c) N-H···SG

cluster I				cluster II			
donor	acceptor	this work	1FDN	donor	acceptor	this work	1FDN
Ser10 N	Cys8 SG	3.366	3.29	Asp39 N	Cys37 SG	3.483	3.44
Tyr30 N	Cys8 SG	3.563	3.65	Tyr2 N	Cys37 SG	3.337	3.37
Ala13 N	Cys11 SG	3.415	3.44	Ala42 N	Cys40 SG	3.326	3.37
Val49 N	Cys47 SG	3.567	3.65	Val20 N	Cys18 SG	3.554	3.59
Ala51 N	Cys47 SG	3.333	3.26	Ala22 N	Cys18 SG	3.306	3.16

(d) Side Chain to Side Chain

donor	acceptor	this work	1FDN
Thr36 OG1	Asp33 OD2	2.751	2.69

(e) N-H···S (Inorganic)

cluster I				cluster II			
donor	acceptor	this work	1FDN	donor	acceptor	this work	1FDN
Cys14 N	S1	3.621	3.60	Cys43 N	S1	3.405	3.50
Ile9 N	S3	3.363	3.30	Ile38 N	S3	3.546	3.47
Gly12 N	S4	3.273	3.19	Gly41 N	S4	3.316	3.29

[a] The distances in angstroms are given for the current model and for 1FDN (*28*).

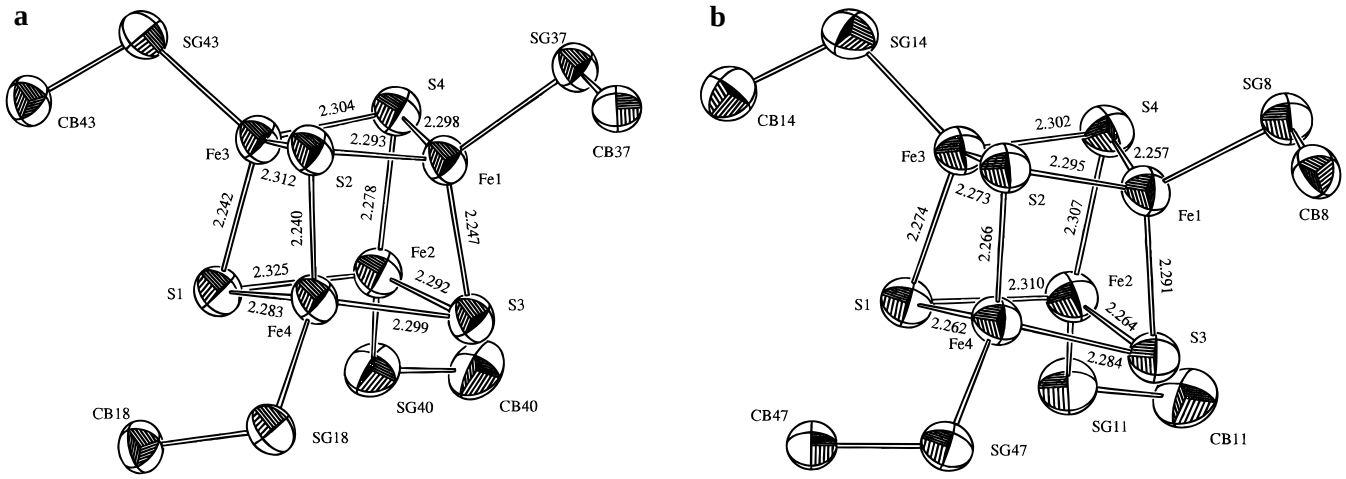

FIGURE 7: Probability (33%) thermal ellipsoid representation of (a) cluster II and (b) cluster I.

Table 7: Structural Comparisons of [4Fe-4S] Clusters in CauFd and in Model Compounds[a]

compd	r_{Fe} (Å)	β_{Fe} (°)	V_{Fe} (Å^3)	r_S (Å)	β_S (°)	V_S (Å^3)
CauFd, cluster I	1.668	55.1	2.38	2.210	55.0	5.59
CauFd, cluster II	1.666	54.7	2.37	2.214	56.1	5.56
$[Fe_4S_4(SCH_2Ph)_4]^{2-}$	1.682	55.6	2.44	2.208	55.6	5.52
$[Fe_4S_4(SPh)_4]^{2-}$	1.675	54.6	2.41	2.211	55.6	5.54
$[Fe_4S_4(SCH_2CH_2CO_2)_4]^{6-}$	1.687	55.4	2.46	2.206	55.0	5.50
$[Fe_4S_4(SCH_2CH_2OH)_4]^{2-}$	1.671		2.39	2.212	56.5	5.53
$[Fe_4S_4(S\text{-}t\text{-}Bu)_4]^{2-}$ (a)	1.688	54.1	2.45	2.214	56.9	5.55
$[Fe_4S_4(S\text{-}t\text{-}Bu)_4]^{2-}$ (b)	1.687	55.0	2.47	2.207	54.9	5.53
$[Fe_4S_4Cl_4]^{2-}$	1.694	54.4	2.49	2.197	55.9	5.43
$[Fe_4S_4(SPh)_4]^{3-}$	1.680	54.3	2.43	2.24	53.6	5.76

[a] The volumes were calculated as described in ref 17. r and β refer to the distances and angles from the mean coordinates of the eight atoms and $\beta = 54.74°$ for a perfect tetrahedron. Note that these calculations are approximate for the clusters of CauFd, as they apply to exact D_{2d} local symmetry. The volumes of the synthetic analogues are taken from refs 17 and 20.

Table 8: Cysteine Conformations in CauFd[a]

cysteine	8	11	14	47
χ_1	179.3	72.7	54.5	181.2
θ (Fe-S-C_β-C_α)	63	306.16	252.1	87.24
cysteine	37	40	43	18
χ_1	178	71.7	52.3	176.1
θ (Fe-S-C_β-C_α)	67.35	308.85	273.9	84.09

[a] The dihedral angle is in degrees.

to a few synthetic analogues (*21*, *23*), of a [4Fe-4S] cluster not following the common compression from T_d toward approximate D_{2d} symmetry (Figure 7b).

A noteworthy feature of both clusters is the relatively short distance of Fe4 to the other iron atoms (Table 2) which contributes to the local symmetry lowering from T_d. No obvious strain is induced by the polypeptide chain on this atom, but the effect must be due to the environment. Indeed, Fe4 is surrounded by hydrophobic side chains belonging to the β-turn flanking the remote cysteine of the coordinating motif (*43*, *47*). The same distortion has not been evidenced with model compounds (*17*, *20*−*24*, *46*), even with those displaying a differentiated subsite (*48*). Only in the case of some trialkylphosphine coordinated clusters, does a shift of one iron atom toward the plane formed by the three coordinating inorganic sulfurs occur (*49* and references therein).

Keeping in mind that the inorganic cores of CauFd are not ideal D_{2d} structures, their volumes have been calculated by a previously devised procedure (*17*). The volume of the tetrahedron defined by the four sulfur atoms is slightly larger, while the one defined by the iron atoms is smaller, in the case of the protein than for all reported synthetic analogues (Table 7). However, in both clusters of this protein as well as in model compounds, the iron tetrahedron assumes a more irregular shape than the sulfur tetrahedron. Despite these discrepancies, the values obtained for $[4Fe\text{-}4S]^{2+}$ clusters are similar, in contrast to those of $[4Fe\text{-}4S]^{+}$ clusters (Table 7, compare the volumes of the S4 tetrahedra). Although similarly accurate values are not available for $[4Fe\text{-}4S]^{3+}$ analogs, the corresponding volumes are also expected to be significantly different (*50*). These data strongly suggest that no redox transitions have been induced in the crystal by X-ray irradiation. Moreover, it is likely that photooxidation of this protein might have produced cluster conversion rather than mere oxidation, as documented in solution (*51*).

Coordination to the Polypeptide Chain. The structural discrepancies observed among otherwise similar synthetic analogues of [4Fe-4S] clusters have been ascribed to a subtle interplay between intrinsic chemical properties of these cores and lattice effects in the crystals (*20*, *21*, *23*). In proteins, the extrinsic influences on the properties of the clusters are contributed mainly by the polypeptide chain.

While the conformations of the identical ligands are generally similar in $[Fe_4S_4(SR)_4]^{2-}$ analog compounds, the orientation of the cysteine side chains varies in CauFd (Table 8). Despite these differences, the lengths of the Fe-SCys bonds linking cluster II to the polypeptide chain are all closely similar, with a mean value of 2.272 Å (±0.007). In cluster I, the mean value of the Fe−SCys distances is smaller and the spread is larger (2.255 ± 0.016) than for cluster II (Table 2). For both clusters, the shortest distance is observed for the third cysteine (Cys14 and Cys43) of each of the cysteine triads. The Fe−SCys distances are slightly larger (cluster II), or smaller (cluster I), than the Fe−SR distances found in synthetic analogues (ca. 2.26 Å on average), except for one having sterically hindered ligands (*24*; 2.274 Å on average). The significance of these rather small differences is unclear. However, in all cases the narrow distributions of these values agree with a similar oxidation state for all iron ions at this redox level (formally 2.5+), with electron delocalization over each cluster. For comparison, the spread of Fe−SCys distances in oxidized (Fe^{3+}) rubredoxin (*25*, *26*)

is larger than in CauFd. The main difference between model compounds and the clusters in CauFd concerns the tilt of the Fe2−SCys (Cys11 and 40 for clusters I and II, respectively) bonds from the cube diagonal, as previously noted (*28*). This deviation from any ideal symmetry is obviously absent from model compounds with four identical ligands. In addition, it seems larger here than in 1FDN, but whether this increased distortion is due to freezing or is revealed by the higher resolution remains unknown. The tilt of the Fe2−SCys bond is larger for cluster II (18.1°) than for cluster I (15.4°), which must be related to the forces imposed by the polypeptide chain around Cys11 and Cys40.

Environment of the Clusters. One of the most scrutinized parameters thought to tune the properties of iron-sulfur clusters in proteins is the number and distribution of hydrogen bonds involving their sulfur atoms (*52*, *53*). As shown in Table 6, the same number of hydrogen bonds surrounds each cluster, five of them involving cysteinyl sulfurs and the remaining three inorganic sulfurs. Moreover, the lengths of symmetry-related hydrogen bonds are very similar around the two clusters, with only a few exceptions. The main difference is the shorter (stronger) NH-S bond between Tyr2 N and Cys37 SG of cluster II compared to the longer (weaker) bond between Tyr30 N and Cys8 SG of cluster I. Other minor differences between clusters I and II concern the hydrogen bonds involving inorganic sulfurs and backbone atoms Ile9 NH and Ile38 NH, and Cys14 NH and Cys43 NH (Table 6).

One may try to correlate these features with the distortions of the clusters discussed above. The differences in length (Table 6) between the hydrogen bonds to SG37 and SG8, as well as between those involving S*1, may explain at least in part the large difference (5°) between the S*1−Fe−SCys37 and S*1−Fe−SCys8 angles, as compared to those within the three other pairs of such angles (mean 2.5°). Also, the absence of hydrogen bonds to SG14 and SG43 may be one of the reasons why the corresponding Fe−SCys bonds are shorter than the other three in each cluster. In contrast to the Fe−SCys bonds, the structural differences between the two [4Fe-4S] inorganic cores cannot be straightforwardly rationalized on the basis of hydrogen-bond-mediated interactions with the polypeptide chain. It seems indeed that the Raman active modes of the Fe−SCys bonds differ in various [4Fe-4S] proteins while the Fe−S* modes remain relatively invariant (*53*-*56*). However, even for synthetic analogues for which many high precision crystal structures are available, the factors that determine the detailed geometry of the [4Fe-4S] core remain somewhat obscure.

Other structural factors may explain some of the properties of [4Fe-4S] clusters. The importance of the dipoles surrounding the metal site, and not only the hydrogen bonds, has recently been emphasized as a main determinant of the value of the reduction potential (*57*, *58*). Here, comparison with analogue model compounds is of very little relevance, as the latter are generally solvated by organic solvents and are not surrounded by peptide ligands. Nevertheless, no well-defined water molecules have been identified in the present model at less than 6 Å from any of the inorganic cores. NMR studies have clearly established that the difference between the redox potentials of the two clusters in clostridial 2[4Fe-4S] ferredoxins is only 10 mV on average (*47* and references therein). Such differences probably arise from slight variations (a few tenths of an angstrom) in the relative positions of some NH, CO, and other dipoles around the cluster, as found in the superposition of the two halves of CauFd.

CONCLUSIONS

The 0.94 Å model of CauFd has confirmed the previously reported fold of the polypeptide chain and, in particular, the exceptional 2-fold noncrystallographic symmetry of the molecule. The new model reveals slight deviations from this symmetry, which have bearings on the geometry of the [4Fe-4S] clusters. An important gain afforded by atomic resolution and low temperature is the positioning, in two conformations, of the previously poorly defined loop (residues 25−29). The mobility of this loop is substantiated by NMR data showing that it predominantly assumes at room temperature a conformation intermediate between the two crystallographic ones. An additional bonus of the 0.94 Å resolution model is the possibility of locating with high precision those atoms which, in the neighborhood of the clusters, are likely to create dipoles that may be important effectors of the redox potential.

Most importantly, the atomic resolution attained here has allowed the accurate inspection of the geometry of [4Fe-4S] clusters, for the first time in any protein. The two clusters do not assume exactly identical structures, in keeping with the fact that the 2-fold symmetry of the ferredoxin molecule is not perfect. This suggests that the geometry of [4Fe-4S] clusters is finely tuned by the polypeptide chain, as most obviously shown by the tilt of one Fe−SCys bond imposed by a constraint of the protein. The Fe_4S_4 core of cluster II approaches the most frequently occurring D_{2d} symmetry, whereas that of cluster I does not display any clear-cut symmetry.

We have recently refined the structures of Zn- and Fe-containing rubredoxins at resolutions better than 1.2 Å (*25*, *26*). The structure reported here, at an even higher resolution and for a protein containing metal sites of much higher nuclearity, shows that the characterization of an extensive set of metal active site structures at atomic resolution is not a remote goal. This should allow a generalization of structural comparisons with synthetic analogues and the unveiling of detailed structural features that are associated with spectroscopic and functional properties of protein active sites.

REFERENCES

1. Lippard, S. J. and Berg, J. M. (1994) *Principles of Bioinorganic Chemistry*, University Science Books, Mill Valley, CA.
2. Johnson, M. K. (1994) in *Encyclopedia of Inorganic Chemistry* (King, R. B., Ed.) Vol. 4, pp 1896−1915, John Wiley, London.
3. Rayment, I., Wesenberg, G., Meyer, T. E., Cusanovich, M. A. and Holden, H. M. (1992) *J. Mol. Biol. 228*, 672−686.
4. Georgiadis, M. M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J. J. and Rees, D. C. (1992) *Science 257*, 1653−1659.
5. Krauss, N., Schubert, W.-D., Klukas, O., Fromme, P., Witt, H. T. and Saenger, W. (1996) *Nat. Struct. Biol. 3*, 965−973.
6. Fujii, T., Hata, Y., Wakagi, T., Tanaka, N. and Oshima, T. (1996) *Nat. Struct. Biol. 3*, 834−837.
7. Macedo-Ribeiro, S., Darimont, B., Sterner, R. and Huber, R. (1996) *Structure 4*, 1291−1301.
8. Moulis, J.-M., Sieker, L. C., Wilson, K. S. and Dauter, Z. (1996) *Protein Sci. 5*, 1765−1775.
9. Lauble, H. and Stout, C. D. (1995) *Proteins: Struct., Funct., Genet. 22*, 1−11.
10. Lim, L. W., Shamala, N., Mathews, F. S., Steenkamp, D. J., Hamlin, R. and Xuong, N. H. (1986) *J. Biol. Chem. 261*, 15140−15146.

11. Crane, B. R., Siegel, L. M. and Getzoff, E. D. (1995) *Science 270*, 59−67.
12. Chan, M. K., Mukund, S., Kletzin, A., Adams, M. W. W. and Rees, D. C. (1995) *Science 267*, 1463−1469.
13. Volbeda, A., Charon, M.-H., Piras, C., Hatchikian, C. E., Frey, M. and Fontecilla-Camps, J. C. (1995) *Nature 373*, 580−587.
14. Boyington, J. C., Gladyshev, V. N., Khangulov, S. V., Stadtman, T. C. and Sun, P. D. (1997) *Science 275*, 1305−1308.
15. Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P. and Tainer, J. A. (1995) *EMBO J. 16*, 4108−4120.
16. Smith J. L., Zaluzec, E. J., Wery, J.-P., Niu, L., Switzer, R. L., Zalkin, H. and Satow, Y. (1994) *Science 264*, 1427−1433.
17. Berg, J. M. and Holm, R. H. (1982) in *Iron-Sulfur Proteins* (Spiro, T. G., Ed.) pp 1−66, John Wiley, New York.
18. Cammack, R. (1992) *Adv. Inorg. Chem. 38*, 281−322.
19. Moulis, J.-M., Davasse, V., Golinelli, M.-P., Meyer, J. and Quinkal, I. (1996) *J. Bioinorg. Chem. 1*, 2−14.
20. Mascharak, P. K, Hagen, K. S., Spence, J. T. and Holm, R. H. (1983) *Inorg. Chim. Acta 80*, 157−170.
21. Kanatzidis, M. G., Baenziger, N. C., Coucouvanis, D., Simopoulos, A. and Kostikas, A (1984) *J. Am. Chem. Soc. 106*, 4500−4511.
22. Müller, A., Schladerbeck, N. H. and Bögge, H. (1987) *J. Chem. Soc., Chem. Commun.* 35−36.
23. Excoffon, P., Laugier, J. and Lamotte, B. (1991) *Inorg. Chem. 30*, 3075−3081.
24. Ueyama, N., Sugawara, T., Fuji, M., Nakamura, A. and Yasuoka, N. (1985) *Chem. Lett.* 175−178.
25. Dauter, Z., Sieker, L. C. and Wilson, K. S. (1992) *Acta Crystallogr., Sect. B 48*, 42−59.
26. Dauter, Z., Wilson, K. S., Sieker, L. C., Moulis, J.-M. and Meyer, J. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 8836−8840.
27. Cato, E. P. and Stackebrandt, E. (1989) in *Clostridia* (Minton, N. P. and Clarke, D. J., Eds.) pp 1−26, Plenum Press, New York.
28. Duée, E. D., Fanchon, E., Vicat, J., Sieker, L. C., Meyer, J. and Moulis, J.-M. (1994) *J. Mol. Biol. 243*, 683−695.
29. Otwinowski, Z. and Minor, W. (1997) *Methods Enzymol. 276*, 307−326.
30. Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491−497.
31. Wilson, A. J. C. (1942) *Nature 150*, 151−152.
32. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Simanouchi, T. and Tasumi, M. (1977) *J. Mol. Biol. 112*, 2055−2064.
33. Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) *Acta Crystallogr., Sect. D 53*, 240−255.
34. CCP4 (1994) *Acta Crystallogr., Sect. D 50*, 760−763.
35. Lamzin, V. S. and Wilson, K. S. (1993) *Acta Crystallogr., Sect. D 49* 129−147.
36. Sheldrick, G. M. (1996) *SHELXL-96: A program for crystal structure refinement*, University of Göttingen, Germany.
37. Sheldrick, G. M. and Schneider, T. R. (1997) *Methods Enzymol. 277*, 319−343.
38. Engh, R. A. and Huber, R. (1991) *Acta Crystallogr., Sect. A 47*, 392−400.
39. Sevcik, J., Dauter, Z., Lamzin, V. S. and Wilson, K. S. (1996) *Acta Crystallogr., Sect. D 52*, 327−344.
40. EU validation network (1997) *J. Mol. Biol.* (in press).
41. Flack, H. D. (1983) *Acta Crystallogr., Sect. A 39*, 876−881.
42. Otaka, E. and Ooi, T. (1987) *J. Mol. Evol. 26*, 257−267.
43. Quinkal, I., Davasse, V., Gaillard, J. and Moulis, J.-M. (1994) *Protein Eng. 7*, 681−687.
44. Moulis, J.-M. and Davasse, V. (1995) *Biochemistry 34*, 16781−16788.
45. Bertini, I., Donaire, A., Feinberg, B. A., Luchinat, C., Piccioli, M. and Yuan, H. (1995) *Eur. J. Biochem. 232*, 192−205.
46. Christou, G., Garner, C. D., Drew, M. G. B. and Cammack, R. (1981) *J. Chem. Soc., Dalton Trans.* 1550−1555.
47. Kyritsis, P., Huber, J. G., Quinkal, I., Gaillard, J. and Moulis, J.-M. (1997) *Biochemistry 36*, 7839−7846.
48. Stack, T. D. P. and Holm, R. H. (1988) *J. Am. Chem. Soc. 110*, 2484−2494.
49. Goh, C., Segal, B. M., Huang, J., Long, J. R. and Holm, R. H. (1996) *J. Am. Chem. Soc. 118*, 11844−11853.
50. O'Sullivan, T. and Millar, M. M. (1985) *J. Am. Chem. Soc. 107*, 4096−4097.
51. Thomson, A. J., Robinson, A. E., Johnson, M. K., Cammack, R., Rao, K. K. and Hall, D. O. (1981) *Biochim. Biophys. Acta 637*, 423−432.
52. Carter, C. W., Jr. (1977) in *Iron-Sulfur Proteins* (Lovenberg, W., Ed.) Vol. 3, pp 158−204, Academic Press, New York.
53. Backes, G., Mino, Y., Loehr, T. M., Meyer, T. E., Cusanovich, M. A., Sweeney, W. V., Adman, E. T. and Sanders-Loehr, J. (1991) *J. Am. Chem. Soc. 113*, 2055−2064.
54. Moulis, J.-M., Meyer, J. and Lutz, M. (1984) *Biochemistry 23*, 6605−6613.
55. Moulis, J.-M., Lutz, M., Gaillard, J. and Noodleman, L. (1988) *Biochemistry 27*, 8712−8719.
56. Staples, C. R., Ameyibor, E., Fu, W., Gardet-Salvi, L., Stritt-Etter, A.-L., Schürmann, P., Knaff, D. B. and Johnson, M. K. (1996) *Biochemistry 35*, 11425−11434.
57. Warshel, A., Papazyan, A. and Muegge, I. (1997) *J. Bioinorg. Chem. 2*, 143−152.
58. Stephens, P. J., Jollie, D. R. and Warshel, A. (1996) *Chem. Rev. 96*, 2491−2513.
59. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
60. Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.

BI972155Y