JBIC (1997) 2:680–689

ORIGINAL ARTICLE

Ana V. Coelho · Pedro Matias · Vilmos Fülöp
Andrew Thompson · Ana Gonzalez
Maria A. Carrondo

Desulfoferrodoxin structure determined by MAD phasing and refinement to 1.9-Å resolution reveals a unique combination of a tetrahedral FeS_4 centre with a square pyramidal $FeSN_4$ centre

Received: 16 April 1997 / Accepted: 31 July 1997

Abstract The structure of desulfoferrodoxin (DFX), a protein containing two mononuclear non-heme iron centres, has been solved by the MAD method using phases determined at 2.8 Å resolution. The iron atoms in the native protein were used as the anomalous scatterers. The model was built from an electron density map obtained after density modification and refined against data collected at 1.9 Å. Desulfoferrodoxin is a homodimer which can be described in terms of two domains, each with two crystallographically equivalent non-heme mononuclear iron centres. Domain I is similar to desulforedoxin with distorted rubredoxin-type centres, and domain II has iron centres with square pyramidal coordination to four nitrogens from histidines as the equatorial ligands and one sulfur from a cysteine as the axial ligand. Domain I in DFX shows a remarkable structural fit with the DX homodimer. Furthermore, three β-sheets extending from one monomer to another in DFX, two in domain I and one in domain II, strongly support the assumption of DFX as a functional dimer. A calcium ion, indispensable in the crystallisation process, was assumed at the dimer interface and appears to contribute to dimer stabilisation. The C-terminal domain in the monomer has a topology fold similar to that of fibronectin III.

Key words Desulfoferrodoxin · Non-heme iron centre · MAD method · Rubredoxin · Fibronectin III

Abbreviations *DFX* desulfoferrodoxin · *DX* desulforedoxin · *Rb* rubredoxin · *Rr* rubrerythrin · *Nl* Neelaredoxin · *D.d. Desulfovibrio desulfuricans* ATCC 27774 · *D.v. Desulfovibrio vulgaris* Hildenborough · *D.g. Desulfovibrio gigas* · *rbo* rubredoxin oxidoreductase

Introduction

The precise functional activity of several mononuclear non-heme iron proteins isolated from anaerobic sulfate-reducing bacteria is still unknown [1–5]. However, these proteins seem to play a role in the electron transfer chain of the sulfate reducing-bacteria, and enzymatic functions can be attributed to some of them [5]. Desulfoferrodoxin (DFX) has been isolated from the sulfate-reducing bacteria *Desulfovibrio* (*D.*) *desulfuricans* ATCC 27774 (*D.d.*), grown in nitrate medium [1] and from *D. vulgaris* Hildenborough (*D.v.*), grown in sulfate medium [1, 2]. Their molecular mass and subunit composition were determined by gel filtration and SDS-PAGE. Moura et al. found a 14 kDa monomer [1], while Verhagen et al. reported a dimer with a molecular mass of 2×14 kDa [2]. A value of 13881.3 Da was further determined by electrospray analysis for the protein isolated from *D. desulfuricans* [6]. An iron content of two atoms per monomer was determined for both proteins.

DFX was isolated from *D. desulfuricans* ATCC 27774 in two oxidation states; the oxidised, grey form and the semi-reduced, pink form. Both states can be interconverted by addition of appropriate oxidising or re-

A.V. Coelho
Chemistry Department, Universidade de Évora, P-7000 Évora, Portugal

A. Thompson
EMBL Grenoble Outstation, c/o ILL 20, BP-156, F-38042 Grenoble Cedex, France

A. Gonzalez[1]
ESRF, BP-220, F-38043 Grenoble Cedex, France

V. Fülöp
Laboratory of Molecular Biophysics and Oxford Centre for Molecular Sciences, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

A.V. Coelho · P. Matias · M.A. Carrondo (✉)
Instituto de Tecnologia Química e Biológica, Apartado 127, P-2780 Oeiras, Portugal
Tel.: +351-1-4418215; Fax: +351-1-4411277;
e-mail: carrondo@itqb.unl.pt

Present address:
[1] EMBL, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany

ducing reagents [7]. Spectroscopic studies of DFX, using optical spectroscopy, EPR and Mossbauer techniques [1, 8], suggest the existence of two different iron centers; centre I similar to the desulforedoxin (DX) centre, coordinated by four cysteines in a distorted tetrahedral geometry, and centre II with an octahedral coordination mainly by nitrogen and/or oxygen ligands. The possibility of pentagonal coordination for this centre was not, however, completely discounted [6]. Based on resonance Raman spectroscopy studies, a partial cysteinyl S coordination was proposed for centre II in its reduced form, and for the protein with this centre in the oxidised form the same studies indicated a probable coordination of two cysteines and at least, one histidine ligand [8]. The subsequent determination of the DFX primary structure showed five cysteine residues, suggesting that if four of them were involved in the coordination of centre I only one would be left to coordinate at centre II [6].

The midpoint redox potentials for both DFX iron-centres are 4 mV for centre I and 240 mV for centre II. The large difference between these values allows the separation of the protein into three oxidation states. Furthermore, the high redox potential value for centre II explains the stability of an aerobically stable semi-reduced form.

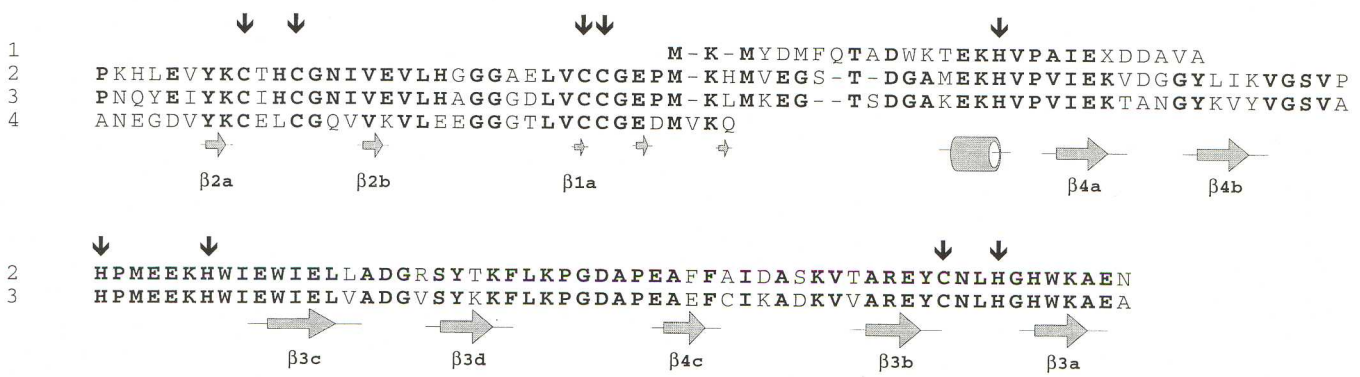
DFX was only found in *D.d.* and *D.v.* but not in *Desulfovibrio gigas* (*D.g*). On the other hand, desulforedoxin (DX), and neelaredoxin (Nl) were only found in *D.g.* but not in *D.d.* and *D.v.*[3, 5]. Some correlation between DFX molecules with DX and Nl has already been attempted based on spectroscopic data and sequence identity; this correlation will be completed when three-dimensional structures of the proteins become available.

The DX structure determined by X-ray crystallography [9] shows that each monomer contains an $FeCys_4$ rubredoxin(Rb)-like centre [4] in a distorted tetrahedral arrangement. A non-crystallographic twofold axis relates the two monomers, which are firmly hydrogen bonded and folded as an incomplete β-barrel, with the two iron centres placed at opposite poles of the molecule. The structure of Nl has not yet been determined. However, it is known that this protein contains two iron atoms, and its N-terminal sequence has some homology with the second domain of DFX (50% identity with DFX from *D.d.*), as can be seen in Fig. 1. This result, together with optical spectroscopy studies and EPR measurements performed at neutral pH, suggest structural similarity between the neelaredoxin iron centre and DFX centre II [5].

The polypeptide sequence of DFX from *D.d.* was determined by Devreese et al. [6], while that from *D.v.* was deduced from its nucleotide sequence [10]. The latter protein was previously called rubredoxin oxidoreductase (*rbo*) by Brumlik et al., who regarded the protein encoded by this gene as a potential redox partner for rubredoxin, since its open reading frame is upstream from the gene encoding rubredoxin. The two DFX proteins have 77% sequence identity. DX, however, has 53% similarity with the first 36 residues of DFX from *D.d.* (domain I) and 56% with the same region in DFX from *D.v.* The sequence alignment of these three proteins is presented in Fig. 1. The similarity between DX and the corresponding counterpart of the *rbo* gene prompted Brumlik et al. [11] to suggest that *rbo* may be the product of a gene fusion between DX and another redox protein. The evidence presented above indicates neelaredoxin as the natural candidate [5].

Mononuclear iron centres combining nitrogen and oxygen ligands are already well described, for example, in the non-heme iron centre of the photosynthetic reaction centre [12] or the centres in transferrins [13–15], dioxygenases [16–20], superoxide dismutase [21], and galactose-1-phosphate uridylyltransferase [22]. In these cases the more frequent ligands are histidines, aspartic or glutamic acids, and solvent water molecules, with tyrosine and C-terminal residues being less common. The preferred geometries for these metal centres are tetrahedral and octahedral.

Fig. 1 Sequence alignment of DFX and related proteins: *1* N-terminal fragment of neelaredoxin from *D.g.* [5], *2* DFX from *D.d.* ATCC 27774 [51], *3* DFX from *D.v.* [10], *4* DX from *D.g.* [52]. Residues in *bold* are common to both DFX sequences. Residues from DX or neelaredoxin common to those are also displayed in *bold*. Residues bound to the iron in DFX *D.d.* are marked with a *pointer*. The secondary structure of DFX *D.d.* is identified along the sequence. This figure was adapted from fig. 3 in [6]



Structures of iron proteins with a combination of sulfur and nitrogen as iron ligands have only recently been characterised. Mononuclear centres of this type were described in the structure of nitrile hydratase, where the iron shows an incomplete octahedral coordination to three sulfurs from cysteines and two main-chain amide nitrogens [23] and in the complex of isopenicillin *N* synthase with the substrate where the iron atom is in a square pyramidal coordination to two nitrogens from one histidine, one aspartic carboxylate oxygen, one water molecule and to a sulfur atom from the substrate [24]. Multinuclear iron centres with sulfur and nitrogen ligands can also be found in clusters such as the "Meatball" found in a protein isolated from *D.v.* Hildenborough, where one of the four iron atoms in the cluster has a coordination of the $FeSNO_2X$ type [25] and in the "Rieske" iron-sulfur protein [26], where one of the two iron atoms has a tetrahedral arrangement to two bridging sulfur atoms and to two nitrogens from histidines.

In the present paper we describe the structure determination of the grey, oxidised form of DFX from *D.d.* by the MAD method [27]. In this structure, a unique combination has been found of a rubredoxin type centre and a second, square pyramidal centre, with nitrogen and sulfur ligands from histidine and cysteine residues respectively.

Experimental

Crystallisation and crystal flash-cooling

Rhombohedral crystals of DFX were grown as previously reported [28]. A stoichiometric amount of potassium ferricyanide was added to the protein solution to induce the crystallisation conditions of the oxidised form. The addition of $CaCl_2$ is indispensable to crystal growth. To assess which type of cation was required to help crystallisation, both $Ca(NO_3)_2$ and NaCl were used; crystals only appeared when the Ca^{2+} ion was present, and the crystals used for this study were those obtained with the addition of $CaCl_2$.

In order to preserve the diffraction quality during the MAD experiments the freezing conditions were optimised, and the best results were achieved by transferring the crystals to crystallisation solutions containing gradually increasing glucose concentrations up to a maximum of 15% (w/v). Crystals were flash-cooled at 100 K. Unit cell dimensions for the frozen crystals are a = 113.1 Å and c = 63.0 Å.

Data collection and crystal characterisation

Room temperature diffraction data were collected at station 9.5 at the SRS, CCLRC Daresbury Laboratory using a wavelength of 0.99 Å and processed to 1.9 Å. The statistics for these data, although summarised here, have already been reported in a previous paper [28], which includes the crystal characterisation. This data set (DAR data set) was used for the final refinement steps.

Data for MAD phasing were collected at four different wavelengths using a MAR Research detector at the ESRF beam line BM-14 (Grenoble, France) from a single frozen crystal. A preliminary X-ray fluorescence spectrum of a DFX crystal near the K-shell edge of the iron was measured. This was used to select the monochromator settings for the peak (λ_1 = 1.7417 Å, maximum $\Delta f''$), the inflection point (λ_2 = 1.7430 Å, the minimum of $\Delta f'$), and two remote points (λ_3 = 1.7514 Å and λ_4 = 1.0908 Å). A crystal mounted with the c axis 90° away from the spindle allowed Bijvoet pairs to be recorded on the same or adjacent images, thus minimising systematic errors. The scaling between the four data sets was then straightforward. The diffraction images were processed and scaled by the programs DENZO and SCALEPACK [29], and merged using the CCP4 [30] program suite (ROTAPREP/AGROVATA/TRUNCATE). Data processing statistics are summarised in Table 1.

DFX crystals belong to space group R32, as previously determined, and the unit cell for the frozen crystals is 0.5% longer in the a and b dimensions, while in the c direction a 0.3% decrease is observed. These differences appear to have no significant structural importance, since the DAR data set was used to continue the refinement of the structure without difficulty.

Structure determination using MAD phasing

Data sets collected at wavelengths λ_1, λ_2 and λ_3 were scaled to data measured at λ_4 using the program SCALEIT of the CCP4 program suite [30]. The heavy-atom positions were determined

Table 1 X-ray data collection

MAD data set ESRF BM-14	Resolution range (Å)	*N* measured [*N* uniques]	Completeness (%)[a] [Multiplicity]	*N* Bijvoet pairs [Completeness (%)]	$I/\sigma(I)$[a]	R_{sym} (%)[ab] R_{anom} (%)[ac]	R_{disp} (%)[ad]
1.7514 Å (remote λ_3)	40.0–2.80	18405 [3856]	98.9 (96.1) [4.8]	3448 [96.0]	14.3 (11.4)	3.5 (5.9) 2.0 (3.2)	4.4 (5.5)
1.7430 Å (inflection λ_2)	40.0–2.80	16774 [3876]	99.4 (100) [4.3]	3413 [85.7]	14.6 (11.2)	3.5 (6.0) 3.6 (4.6)	5.3 (6.5)
1.7417 Å (Peak λ_1)	40.0–2.75	19163 [4059]	98.7 (95.1) [4.7]	3614 [74.9]	14.1 (7.7)	3.6 (8.2) 4.1 (6.3)	3.9 (5.1)
1.0908 Å (remote λ_4)	40.0–2.50	24873 [5230]	96.3 (87.5) [4.8]	4608 [91.3]	22.9 (10.1)	2.6 (7.4) 2.1 (4.8)	–
DAR data set, Daresbury, station 9.5							
0.994 Å	20.1–1.9	25678 [11867]	97.7 (96.3) [2.2]	–	8.8 (2.9)	4.5 (25.7) –	–

[a] Numbers between brackets refer to the last resolution shell; between square brackets are presented the values for the magnitude indicated

[b] $R_{sym} = \Sigma_h \Sigma_i |I_i(h) - \langle I(h) \rangle| / \Sigma_h \Sigma_i I_i(h)$ where I_i(h) is the ith measurement

[c] $R_{anom} = \Sigma |\langle I+ \rangle - \langle I- \rangle| / \Sigma(\langle I+ \rangle + \langle I- \rangle)$

[d] $R_{disp} = \Sigma ||F_p(\lambda_a)| - |F_p(\lambda_4)|| / \Sigma |F_p(\lambda_4)|$ where $F_p(\lambda_a)$ and $F_p(\lambda_4)$ are the protein structure factors amplitudes at $\lambda_a (a = 1,2,3)$

from a Patterson map calculated using the anomalous differences from the data set with highest anomalous signal (λ_1) as coefficients and confirmed from additional Patterson maps calculated using both the anomalous differences from the λ_2 data set and the maximal dispersive differences ($||F(\lambda_2)|-|F(\lambda_4)||^2$) as coefficients. The positions of the two iron atoms were determined with program RSPS [30] and confirmed by SHELXS-86 [31].

Phase refinement of these iron positions and MAD phasing were performed using the program MLPHARE [30]. For the phasing using MLPHARE, the data set λ_2 was treated as a native data set with intrinsic anomalous scattering, and the other three were regarded as derivatives, following a procedure described by Ramakrishnan et al. [32] and Glover et al. [33]. The preliminary phases were calculated and refined using the data in the range from 10-Å to 2.8-Å resolution. This process converged to an overall figure of merit of 0.69. The statistics for the phase calculation and refinement are presented in Table 2. Density modification (solvent flattening and histogram matching) and phase extension to 2.5 Å using the program DM [30] was then performed, and as a result the figure of merit increased to 0.78. A strong contrast between the protein and the solvent regions was revealed in the calculated electron density map. Its quality was high enough to build the protein model with the exception of a solvent-exposed loop of eight residues.

Model building and refinement

The atomic model was constructed using the program O [34]. A first refinement step including rigid-body minimisation and a simulated annealing round was performed with XPLOR [35] using the λ_4 data set (maximum resolution 2.5 Å). Since a significant improvement in the electron density map was not achieved after more refinement steps, further refinement of the model proceeded with the highest resolution data set available (maximum resolution 1.9 Å). The R factor (based on 11017 reflections in the range of 8.0-Å to 1.9-Å resolution) and the free R factor (based on 573 (5%) randomly selected reflections in the same range) [36] for the initial model including 117 amino acid residues (residues 40 to 47 were missing) were 40.4% and 44.2%, respectively. These values were reduced to 24.5% and 30.1%, respectively, after rigid-body minimisation, simulated annealing and a gradual refinement of the B factors with XPLOR [35]. The improvement of the electron density map allowed the introdution of the eight missing residues and 56 water molecules. SHELXL-93 [37] was used to finish the refinement of the complete model using reflections covering the resolution range from 20.0-Å to 1.9-Å. Refinement based on intensities as obtained directly from SCALEPACK, rather than on structure factors amplitudes obtained from TRUNCATE, resulted in a significant decrease in the free R factor. Also, a significant decrease in R_{free} was observed when the anisotropic refinement of iron, sulfur and calcium atoms was included in the refinement protocol. No geometrical restraints between the iron atoms and their ligands were imposed during the refinement procedure. The R factor and the free R factor converged to 19.6% and 23.1%, respectively. In the final refinement all the reflections were used and no significant changes in either the model or the R factor were observed. The same set of reflections were used during all the refinement procedures performed with XPLOR and SHELXL-93 and in the calculation of the free R factor.

A peak of 0.32 eÅ^{-3} (6.4σ) density persisted in the $2F_o$–F_c and in the F_o–F_c maps near the iron atom from centre II, in the direction of a possible sixth ligand, but at a distance from the iron of 4.9 Å. This peak is also remote from other atoms in the structure, and has not yet been fully explained.

A sphere of density positioned on the crystallographic twofold axis between the two monomers observed during the refinement process suggested the presence of an ion which could help to establish and maintain the interaction between monomers. This density is surrounded by eight oxygen atoms from neighbouring residues, within 2.7–3.1 Å. The height of this peak is almost twice that attributed to a water molecule. Since the addition of Ca^{2+} is indispensable to crystallisation and as oxygens are the local ligands to this ion, this density was attributed to and refined as a calcium ion.

The final model includes the complete sequence of amino acids, two iron atoms, 71 water molecules and a calcium ion sitting on the crystallographic twofold axis. The refinement parameters indicating the quality of the final model are presented in Table 3. The stereochemical quality of the model was assessed with XPLOR [36] and PROCHECK [38], which gives a G factor of –0.2 for the refined DFX model (the G factor value for structures with the same nominal resolution is -0.3 ± 0.3). The cavity volume calculations were performed with VOIDOO [39].

Results and discussion

Refined structure

The refined model is complete except for a solvent-exposed loop (Ser40 to Lys47) for which continuous density could not be obtained. Main-chain B–factors of some residues in this region are 2.5 times higher than the average B value, reflecting the high flexibility or the statistical disorder of this loop. The B factors for the

Table 2 MLPHARE phase refinement statistics

Wavelength	Occ[ac]		Aocc[bc]		Phasing power		R_{cullis}		
	Fe 1	Fe 2	Fe 1	Fe 2	Centric	Acentric	Centric[d]	Acentric[d]	Anomalous[e]
λ_1 (peak)	0.892	0.435	1.532	1.265	0.90	1.16	0.69	0.78	0.62
λ_2 (inflection)	0.0	0.0	1.301	1.129	–	–	–	–	0.65
λ_3 (remote 1)	1.163	0.734	0.306	0.246	1.20	1.54	0.63	0.69	0.97
λ_4 (remote 2)	2.890	2.201	0.602	0.491	1.59	1.99	0.52	0.59	0.79

[a] Occupancy
[b] Anomalous occupancy
[c] The occupancies were refined using unitary form factors and reflect the number of electrons constituting the dispersive differences between data collected at different wavelengths and contributing to the anomalous differences using data on an approximately absolute scale

[d] $R_{cullis} = \Sigma||F_p(\lambda_a)| - |F_p(\lambda_2)|| - |\Delta F_{calc}||/\Sigma||F_p(\lambda_a)| - |F_p(\lambda_2)||$, where $F_p(\lambda_a)$ and $F_p(\lambda_2)$ are protein structure factors at λ_a (λ_1, λ_3, λ_4) and λ_2, respectively. ΔF_{calc} is the calculated dispersive difference between λ_a and λ_2. Summation was done using centric and acentric reflections

[e] $R_{cullis\_anom} = \Sigma||F_p^+(\lambda_a)| - |F_p^-(\lambda_a)|| - |\Delta F_{calc}||/\Sigma||F_p^+(\lambda_a)| - |F_p^-(\lambda_a)||$, where $|F_p^+(\lambda_a)| - |F_p^-(\lambda_a)|$ is the Bijvoet difference at λ_a (λ_1, λ_3, λ_4). ΔF_{calc} is the calculated anomalous difference at λ_a

Table 3 Refinement statistics and the quality of the final model

Parameter	
Residue range	1–125
Number of non-hydrogen atoms	975
Number of water molecules included (with 0.5 occupancy)	71 (27)
Resolution range (Å)	20.0–1.9
Average temperature factor (Å^2)	
Main-chain atoms	34.1
Side-chain atoms	37.8
Metal centre atoms (I; II)	32.6; 37.8
Water molecules	40.8
Calcium pocket ligands	27.2
R factor (F>4σ(F);F>0)[a] (%)	18.7; 19.6
Number of reflections	12875
Free *R* factor (F>4σ(F);F>0) (%)	21.3; 23.1
Number of reflections	689
RMS deviations from ideal values	
Bond lengths (Å)	0.02
Bond angles (°)	2.3
Dihedral angles (°) (not restrained)	30.7
Improper torsion angles (°)	2.5
Ramachandran plot (non-Gly, non-Pro residues)	
Residues in most favoured regions (%)	89.4
Residues in additional allowed regions (%)	10
Residues in generously allowed regions (%)	1
Residue in disallowed regions (%)	0
Other structural quality indicators[b]	
Number of bad contacts	0
Omega angle deviation (°)	3.2
Zeta angle deviation (°)	3.7
Hydrogen-bond energy standard deviation (kcal mol^{-1})	0.8
Overall *G* factor	−0.2

[a] R factor $=\Sigma_h \| F(\mathrm{h})_{\mathrm{obs}} \| - \| F(\mathrm{h})_{\mathrm{calc}} \| / \Sigma_h | F(\mathrm{h})_{\mathrm{obs}} |$
[b] Several parameters as in PROCHECK [38]

iron atoms (31.8 Å^2 and 39.1 Å^2), similar to the mean values of their ligands, are in accordance with a full occupancy for their positions. Asp84, in a 2:2 type I' hairpin loop [40], is the only residue in a less favourable region of the Ramachandran plot [41].

Overall structure

The molecule of DFX is a homodimer composed of two crystallographically equivalent monomers related by a twofold axis. A ribbon diagram of the DFX molecule with the twofold axis vertically in the plane of the figure and the trace of the α-carbon backbone with this axis perpendicular to the plane are shown in Figs. 2a and b, respectively. The molecule has approximate maximum dimensions of $50 \times 47 \times 25$ Å^3.

The monomer consists of two domains, each containing a non-heme iron centre. However, the homodimer can also be described with advantage as arranged in two domains, each with the respective contribution from the corresponding subunits. In this paper, we shall refer to domain I and II as in the DFX dimer. Domain I has an approximately spherical shape very similar to the structure of DX and will therefore be called the DX-like domain. Domain II has an elongated shape extended in the direction normal to the twofold axis. The two domains are linked by a region of polypeptide chain giving rise to an elongated van der Waals cavity in the dimer with an approximate volume of 430 Å^3. There are three levels of antiparallel β-sheets in the dimeric molecule.

A topology diagram for DFX monomer is shown in Fig. 3. The N-terminal domain is 34 amino acid residues long, while the C-terminal domain has 88 amino acid residues arranged in a "3+4 sheet structure" typical of the fibronectin III fold [42]. This motif has been described in several proteins of animal origin, predominantly extracellular proteins or their domains [43], but has been found in the structures of few proteins of bacterial and plant origin, namely in bacterial chitinase [44] and in chloroplast cytochrome *f* [45].

The iron atoms are close to the molecular surface and well exposed to the solvent (Fig. 2a); their positions, together with the distances between the four iron atoms in the molecule, are given in Fig. 2b. The solvent molecules are mainly distributed at the surface of the protein, but a few are buried near the iron centre I, at the dimer interface and in the inter-domain region.

Domain I, the DX-like domain

The superposition of DFX domain I and DX gives a root-mean-square (rms) deviation of 0.59 Å for main-chain atoms. This value is an indication of the strong similarity between the two structures.

Three β-strands from each monomer contribute to two almost parallel β-sheets in the dimer (β1 and β2), forming an incomplete β-barrel [9] with a predominantly hydrophobic core (see Fig. 2a).

The iron centres in Domain I have, as in DX, a distorted tetrahedral coordination to four sulfur atoms from cysteine residues, as can be seen in Fig. 4a. Table 4 lists bond lengths and bond angles around the iron atom in centre I for DFX and the equivalent values in DX and Rb. Iron-sulfur distances are similar in the three molecules, but the bond angles show the same distorted geometry in DFX as that in DX, with a slightly higher angle involving the vicinal cysteines. The coordinating cysteines in DFX, Cys9 and Cys12, contribute to a motif, Cys-X-X-Cys, usually called a rubredoxin "knuckle" [46] and also found in DX, with characteristic fold and hydrogen bonds.

Tyrosine 7, the only aromatic residue present in DX, is conserved in Rb and in DFX in a similar position to that found in DX. However, DFX has four more aromatic residues in domain I. One of these residues, His

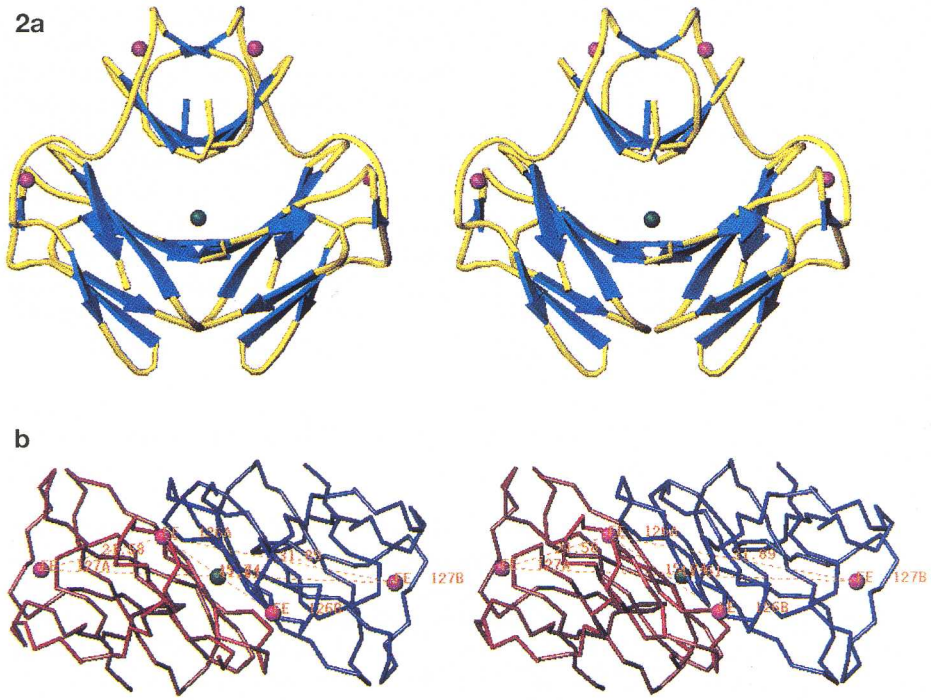
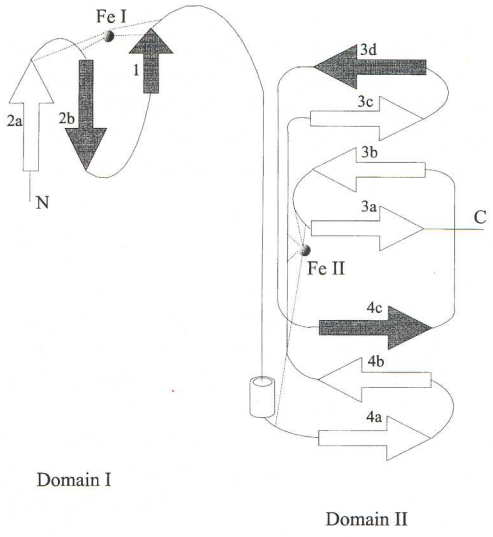


Fig. 2a,b DFX dimer with the iron atoms displayed in *magenta* and the calcium ion in the interdomain region in *green*. **a** Ribbon diagram showing the two DFX dimeric domains oriented with the twofold crystallographic axis along the plane of the figure: domain I is the top and domain II the bottom domain. The two domains are linked by a piece of randomly coiled polypeptide chain. The C (backbone is displayed in *blue* for β-strands and *yellow* for random coils. **b** Cα tracing of DFX dimer viewed along the twofold axis with the monomers represented in different colours (*red* and *blue*). Distances between iron atoms are 15.8 Å in domain I, 41.6 Å in domain II and 21.6 Å and 31.9 Å between domains. These and all the subsequent drawings were made with TURBO-FRODO [54]



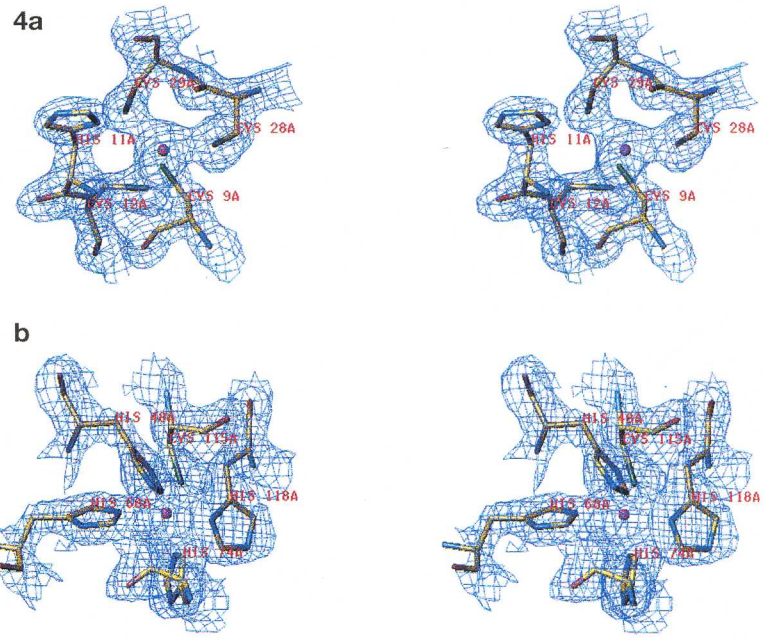


Fig. 3 The topology diagram of DFX monomer. The *shadowed arrows* represent the strands interacting with the symmetry-related monomer and the spheres represent the iron atoms. The residues coordinating the iron atom are referred to Fig. 1

Fig. 4a,b Stereo representations of electron density maps contoured at 1σ level are superimposed with the two different iron centres structures. **a** Centre I, an Rb-like centre, is coordinated by four sulfur atoms from cysteines 9, 12, 28 and 29. His 11, with its Nδ1 atom 5.8 Å from the iron atom, is also represented. **b** Centre II is coordinated in the equatorial positions by the Nε2 atoms from His 48,68,74, by the Nδ1 atom from His 118 and in the axial position by the sulfur atom from Cys115. This centre presents a square pyramidal coordination geometry

11, which replaces Leu11 in DX, has its Nδ1 atom 5.8 Å from the iron atom (see Fig. 4a).

Domain II

In domain II, strands β3a–d from each monomer contribute to the eight-stranded antiparallel β-sheet that extends through the molecule, showing a large twist between the first and the eighth strands. Below this region, four strands forming a β-sheet in each monomer, as represented in Fig. 2a, do not show interaction between the subunits.

Table 4 Bond lengths and angles around iron centre I in DFX and equivalent values in DX and Rb

Bond length Fe-S_{cys} (Å)	DFX	DX[a]	Rb[b]
9	2.3	2.3–2.3	2.3–2.3
12	2.3	2.3–2.4	2.3–2.3
28	2.3	2.3–2.4	2.3–2.3
29	2.3	2.2–2.3	2.2–2.3
Bond angle S_{cys}-Fe-S_{cys} (°)			
9–12	110	109–110	111–115
9–28	106	107–108	111–113
9–29	106	108–110	105–106
12–28	104	102–104	101–104
12–29	105	104–105	109–114
28–29	126	120–126	112–113

[a] Values for the two monomers in DX molecule
[b] Range of values for the three rubredoxin structures from *Desulfovibrio* species

Iron centre II is the second example, after nitrile hydratase [23], where a mononuclear non-heme iron centre is coordinated only to sulfur and nitrogen atoms. However, centre II in DFX is the first case where these ligands are provided by histidine and cysteine residues. The iron coordination, shown in Fig. 4b, can be described as square pyramidal, with the four histidines, His48, His68, His74 and His118, occupying the equatorial positions and Cys115 the axial position. Table 5 gives details of the geometry around this centre. All histidines coordinate to the metal atom through their Nε2 atom except His118, which bonds through Nδ1.

The electron density in the iron vicinity does not suggest a sixth ligand. The residue closer to a possible sixth coordination, Lys47, is partially visible, but is at the end of a exposed loop that shows poor electron density. Its current position gives a distance of 7.5 Å between Lys47 Nζ and the iron atom. Therefore the sixth iron position is easily accessible to solvent.

Table 5 Geometrical parameters in the coordination sphere of iron II

Ligands		Bond length Fe-X (Å)	Distance Fe-CαX (Å)	Distance Cys115-X (Å)
His48 Nε2		2.2	6.3	3.7
His68 Nε2		2.1	6.5	3.2
His74 Nε2		2.2	6.2	3.2
His118 Nδ		2.1	4.5	3.3
Cys115 Sδ		2.3	–	–
X	Y	Bond angles X-Fe-Y(°)	Angles between planes from opposite histidines (°)	
His48	His68	84	His48-His74	130
His68	His74	94	His68-His118	69
His74	His118	86		
His118	His48	92		
Cys115	His48	112		
Cys115	His68	95		
Cys115	His74	90		
Cys115	His118	96		
Deviations from the plane (Å)			Angle between the plane and the Fe-Cys115 Sγ bond (°)	
His48 Nε2		−0.1	79	
His48 Nε2		0.1		
His48 Nε2		−0.1		
His118 Nδ		0.1		
Cys115 Sγ		2.6		
Fe		0.3		

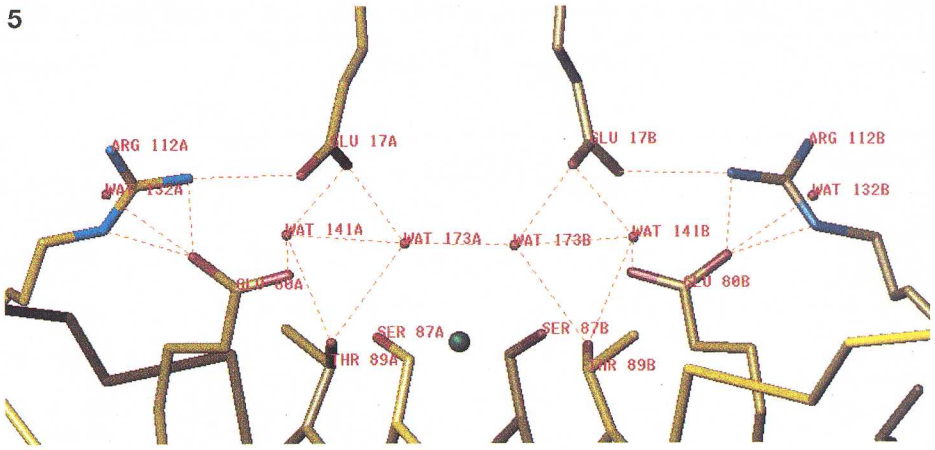
Distribution of aromatic residues and interdomain region

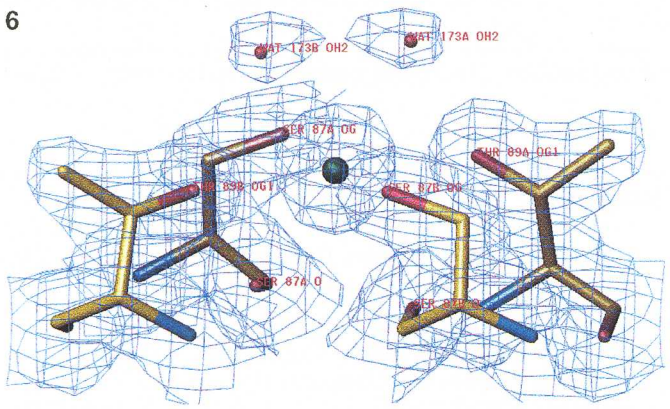
There are six structural water molecules in the interdomain region, three from each subunit. Of these six water molecules, a chain of four contribute to a network of hydrogen bonds with side-chain atoms from neighbouring residues. This network is extended, through Glu80A and Glu80B, to two further water molecules, one on each side of this chain. This arrangement can be seen in Fig. 5.

The calcium ion lies on the twofold axis between β-strand 3d and its symmetry-related strand, and is coordinated by eight oxygen atoms, two of which are the central waters in the chain referred to above. The electron density for this region is shown in Fig. 6, and its coordination geometry is listed in Table 6. The unusually long bond lengths between the calcium ion and its oxygen ligands may be explained by a considerable thermal motion of this ion (refined B factor of 52.6 Å^2 compared to 27.2 Å^2, the average B factor of the neighbouring residues), which may be due to the size of the cavity where it is located, leading to some uncertainty regarding its fixed location on the twofold axis. The major donor groups in proteins to calcium ions are carbonyl or carboxylates, but alcohol groups and water molecules can also be found [47]. The usual coordina-

Table 6 Calcium-binding pocket bond lengths and angles

Ligands		Bond Ca-X (Å)
Ser87 Oγ		3.1
Ser87 O		2.9
Thr89 Oγ1		3.1
W173		2.7
X	Y	Angle X-Ca-Y (°)
Ser87A Oγ	Thr89A Oγ1	73
Ser87A Oγ	W173 A	85
Ser87A Oγ	Ser87A O	62
Ser87A Oγ	Thr89B Oγ1	107
Thr89A Oγ1	Ser87B O	100
Thr89A Oγ1	W173A	59
W173A	W173B	64
Ser87A O	Ser87B O	76





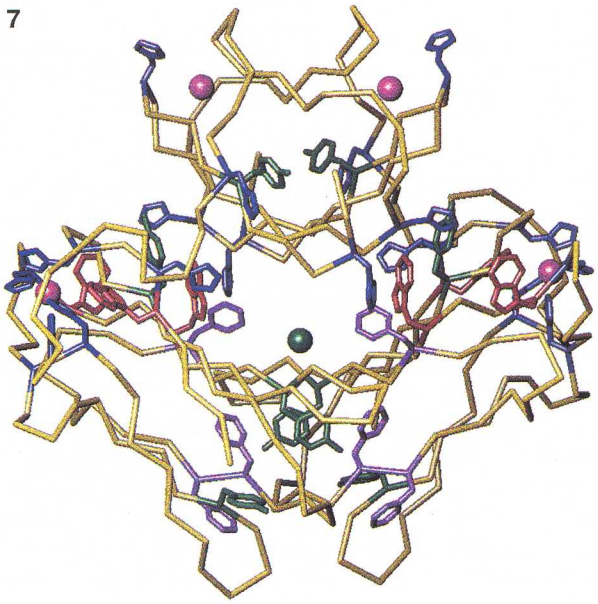


Fig. 5 Zoom of the interdomain region showing the core of polar and charged residues and the chain of buried water molecules. *Dashed lines* indicate all possible hydrogen bonds. Atoms are *coloured* by atom type and the calcium ion is *coloured in green*

Fig. 6 The calcium pocket is represented with its eight ligands and electron density contoured at 1σ level

Fig. 7 Cα backbone of DFX molecule is represented with the side chains of the aromatic residues displayed in different colours according to the residue type: histidine in *blue*, tryptophan in *red*, tyrosines in *green* and phenylalanine in *violet*. The iron atoms are displayed in *magenta* and the calcium ion in *green*. A core of aromatic residues surrounding centres II extends through the inner part of the molecule, and another aromatic core can be seen underneath the calcium position

tion number for calcium in proteins is six or seven, but a larger number of ligands is also possible [48].

Figure 7 shows the Cα-backbone of the DFX molecule, with the side chains of aromatic residues also represented. These are mainly distributed along the interdomain region, from iron centre II in one subunit to the equivalent centre in the other subunit. This arrangement is interrupted, at its centre, by a polar core of side chains (as represented in Fig. 5), which includes the calcium ion and the six water molecules previously mentioned. This core could eventually provide an adequate binding pocket for a suitable substrate.

Below this polar core, another region with a high density of aromatic side chains can be observed, in domain II, close to the subunit interface. Two tyrosines, one from each subunit, have their phenyl rings almost parallel, within 3.6 Å from each other (Fig. 7).

Biological and structural characteristics of proteins with mononuclear non-heme iron centres isolated from sulfate-reducing bacteria

The strong interaction between the two crystallographically related DFX monomers and the structural similarity of DFX domain I and DX suggest, as in the case of DX structure, that these two proteins are functionally active as dimers. However, DX must be a more stable dimer than the DFX *D.d.* dimer, because electrospray mass spectroscopy and gel filtration results are only consistent with a dimeric molecule for DX [49], while in the same experimental conditions a monomer was assigned for DFX *D.d.* [1, 6]. Nevertheless, SDS-PAGE and gel filtration experiments indicate that DFX from *D.v.* is also a homodimer in solution [2]. Interestingly, DFX seems to need a calcium ion to help the stabilization of the dimer in the crystal, a result suggesting fur-

Table 7 Biochemical and iron centre characterization of non-heme proteins isolated from sulfate-reducing bacteria

Protein	Fe/monomer (monomers/ molecule)	E'_0 (mV)	Iron ligands/geometry
Desulfoferrodoxin[b]			
D.d. [1]	2 (2)	+4, +240	Rb-like/tetrahedral distorted; 4 his, 1 cys/square pyramidal
D.v. [2]	2 (2)	+2, +90	
Desulforedoxin[b]			
D.g. [3]	1 (2)	−35	Rb-like/tetrahedral distorted
Rubredoxin[b]			
D.d., *D.v.*, *D.g.* [4]	1 (1)	between −5 and 0	4 cys/tetrahedral
Neelaredoxin[a]			
D.g. [5]	2 (1)	+190, +190	proposed in [5] similar to DFX centre II
Rubrerythrin[b]			
D.v. [51]	3 (2)	+230 centre Rb-like [51],	Rb-like/tetrahedral; 5 glu, 1 his/diiron oxo-bridge octahedral per Fe [50]
D.d. [7]		+281, +339, +246 [53]	
Nigerrythrin[a]			
D.v. [53]	3–4 (2)	+213, +300, +209	Rb-like; proposed in [53] similar to dinuclear Fe Rr centre

The E'_0 values are quoted in the same order as the iron-centre ligands and geometry. Technique used for the metal centre characterisation: [a] spectroscopy, [b] X-ray crystallography

ther enzymatic studies and, in particular, the reassessment of the non-observed pyridine-linked rubredoxin oxidoreductase activity [7]. The amino acid sequence identity (79%) between DFX from *D.d.* and *D.v.* suggests a very similar structure for these proteins. The large difference between centres II redox potentials of these two proteins (see Table 7) still lacks an explanation.

A comparison between biochemical and structural characteristics of several mononuclear non-heme proteins isolated from sulfate-reducing bacteria is presented in Table 7. Although the biological function for these proteins is not yet known, a role in electron transfer has been proposed for all of them, while enzymatic activity has been suggested for some [5]. X-ray structures are now known for four of these proteins, namely DFX from *D.d.*, DX [9], Rb [4] and Rr [50]. It is interesting to note that all these proteins, except neelaredoxin, contain an Rb-like centre. As previously suggested [5], this protein, in combination with DX, may have in *D. gigas* an equivalent biological function to DFX in *D.d.*

DFX and rubrerythrin are the only known structures where an Rb-like centre is combined with another iron centre, a mononuclear centre of type $FeSN_4$ in the case of DFX and a diiron-oxo centre in the case of rubrerythrin.

Acknowledgements The authors would like to thank Profs. J. Le Gall and A.V. Xavier for most useful discussions, Profs. J.J.G. Moura and I. Moura and Dr. P. Tavares for the purified protein sample used for crystallisation and for supplying the DFX polypeptide sequence before it was published, Dr. C. Frazão for helping with synchrotron data collection, and the fermentation plant at University of Georgia, Athens, Ga., USA for growing the bacteria. Vilmos Fülöp is a Royal Society University Research Fellow. This work was funded by EU-HCM-CHRX-CT93-0143 and supported by EMBL and ESRF in Grenoble and CCLRC in Daresbury.

References

1. Moura I, Tavares P, Moura JJG, Ravi N, Huynh BH, Liu M-Y, LeGall J (1990) J Biol Chem 265:21596–21602
2. Verhagen MFJM, Voohorst, WGB, Kolkman JA, Wolbert RBG, Hagen WR (1993) FEBS Lett 336:13–18
3. Moura I, Bruschi M, LeGall J, Moura JJG, Xavier AV (1977) Biochem Biophys Res Commun 75:1037–1044
4. Sieker LC, Stenkamp RE, LeGall J (1994) In: Wyckhoff HW, Hirs CHW, Timasheff SN (eds) Methods in Enzymology 243:203–216. Academic Press, London
5. Chen L, Sharma P, LeGall J, Mariano, AM, Teixeira M, Xavier AV (1994) Eur J Biochem 226:613–618
6. Devreese B, Tavares P, Lampreia J, Van Damme N, LeGall J, Moura JJG, Van Beeumen J, Moura I (1996) FEBS Lett 385:138–142
7. Moura I, Tavares P, Ravi N (1994) In: Wyckhoff HW, Hirs CHW, Timasheff SN (eds) Methods in Enzymology 243:216–240, Academic Press, London
8. Tavares P, Ravi N, Moura JJG, LeGall J, Huang Y-H, Crouse BR, Jonhson MK, Huynh BH, Moura I (1994) J Biol Chem 269:10504–10510
9. Archer M, Huber R, Tavares P, Moura I, Moura JJG, Carrondo MA, Sieker LC, LeGall J, Romão MJ (1995) J Mol Biol 251:690–702
10. Brumlik MJ, Voordouw G (1989) J Bacteriol 171:4996–5004
11. Brumlik MJ, Leroy G, Bruschi M, Voordouw G (1990) J Bacteriol 172:7289–7292
12. Ermler U, Fritzsch G, Buchanan SK, Michel H (1994) Structure 2:925–936
13. Anderson BF, Baker HM, Norris GE, Rice DW, Baker EN (1989) J Mol Biol 209:711–734
14. Dewan JC, Mikami B, Hirose M, Sacchettini JC (1993) Biochemistry 32:11963–11968
15. Bailey S, Evans R, Garratt RC, Gorinsky B, Hasnain S, Horsburgh C, Jhoti H, Lindley PF, Mydin A, Sarra R, Watson J (1988) Biochemistry 27:5804–5812
16. Minor, W, Steczko J, Bolin JT, Otwinowski Z, Axelrod B (1993) Biochemistry 32:6320–6323
17. Boyington JC, Gaffney BJ, Amzel LM (1993) Science 260:1482–1486
18. Han S, Eltis LD, Timmis KN, Muchmore SW, Bolin JT (1995) Science 270:976–980
19. Ohlendorf DH, Lipscomb JD, Weber PC (1988) Nature 336:403–405

20. Roach PL, Clifton IJ, Fülöp V, Harlos K, Barton GJ, Hajdu J, Andersson I, Schofield CJ, Baldwin JE (1995) Nature 375:700–704
21. Stoddard BL, Howell PL, Ringe D, Petsko GA (1990) Biochemistry 29:8885–8893
22. Wedekind JE, Frey PA, Rayment I (1995) Biochemistry 34:11049–11061
23. Huang W, Jia J, Cummings J, Nelson M, Schneider G, Lindqvist Y (1997) Structure 5:691–699
24. Roach Pl, Clifton IJ, Hensgens CMH, Shibata N, Schofield CJ, Hajdu J, Baldwin JE (1997) Nature 387:827–830
25. Lindley P, Hadden J, Card G, McAlpine A, Bailey S, Zaitsev V, Duke E, Arendsen A, Bulsink Y, Hagen W (1997) European Research Conference "Chemistry of metals in biological systems", Tomar, Portugal
26. Iwata S, Saynovits M, Link TA, Michel H (1996) Structure 4:567–579
27. Hendrickson WA (1991) Science 254:51–58
28. Coelho AV, Matias PM, Carrondo MA, Tavares P, Moura JJG, Moura I, Fülöp V, Hajdu J, LeGall J (1996) Protein Science 5:1189–1191
29. Otwinowski, Z (1993) In: Sawyer L, Isaacs N, Bailey S (eds) Data collection and processing. SERC Daresbury Laboratory, Warrington, UK, pp 80–86
30. Collaborative Computational Project Number 4 (1994) Acta Cryst D50:760–763
31. Sheldrick GM (1990) Acta Cryst A46:467–473
32. Ramakrishanan V, Finch JT, Graziano V, Lee PL, Sweet RM (1993) Nature 362:219–223
33. Glover ID, Denny RC, Nguti ND, McSweeney SM, Kinder SH, Thompson AW, Dodson EJ, Wilkinson AJ, Tame JRH (1995) Acta Cryst D51:39–47
34. Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Acta Cryst A47:110–119
35. Brünger AT (1992) X-plor: Version 3.1. A system for protein crystallography and NMR. Yale University Press, New Haven, Conn
36. Brünger AT (1992) Nature 355:472–475
37. Sheldrick GM, Schneider TR (1997) In: Sweet RM, Carter JR CW (eds) Methods Enzymol. Vol 277. Academic Press, Orlando, Florida, pp 219–243
38. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) J Appl Cryst 26:283–291
39. Kleywegt GJ, Jones TA (1994) Acta Cryst D50:178–185
40. Sibanda BL, Blundell TL, Thornton JM (1989) J Mol Biol 206:759–777
41. Ramachandran GN, Sassiekharan V (1968) Adv Prot Chem 28:283–437
42. Baron M, Main AL, Driscoll PC, Mardon HJ, Boyd J, Campbell ID (1992) Biochemistry 31:2068–2073
43. Bork P, Doolittle RF (1992) Proc Natl Acad Sci USA 89:8990–8994
44. Perrakis A, Tews I, Dauter Z, Oppenheim AB, Chet I, Wilson KS, Vorgias CE (1994) Structure 2:1169–1180
45. Martinez SE, Huang D, Szczepaniak A, Cramer WA, Smith JL (1994) Structure 2:95–105
46. Schwabe JWR, Klug A (1994) Nature Struct Biol 1:345–349
47. Declercq J-P, Tinant B, Parello J, Etienne G, Huber R (1988) J Mol Biol 202:349–353
48. Silva JJRF da, Williams RJP (1991) The biological chemistry of the elements – the inorganic chemistry of life. Clarendon Press, Oxford
49. Moura I, Bruschi M, Le Gall J, Moura JJG, Xavier AV (1977) Biochem Biophys Res Commun 75:1037–1044
50. DeMaré F, Kurtz DM, Nordlund P (1996) Nature Structural Biology 3:539–546
51. LeGall J, Prickril BC, Moura I, Xavier AV, Moura JJG, Huynh B-H (1988) Biochemistry 27:1636–1642
52. Bruschi M, Moura I, LeGall J, Xavier AV, Sieker LC (1979) Biochem Biophys Res Commun 90:596–605
53. Pierik AJ, Wolbert RBG, Portier GL, Verhagen MFJM, Hagen WR (1993) Eur J Biochem 212:237–245
54. Roussel A, Cambillau C (1989) In: Silicon Graphics (eds) Silicon Graphics Geometry Partner Directory. Silicon Graphics, Mountain View, Calif., pp 77–78