J Biol Inorg Chem (2003) 8: 360–370
DOI 10.1007/s00775-002-0426-3

ORIGINAL ARTICLE

**Isabel A. Abreu · Alexandra I. Lourenço
António V. Xavier · Jean LeGall · Ana V. Coelho
Pedro M. Matias · David M. Pinto
Maria Arménia Carrondo · Miguel Teixeira
Lígia M. Saraiva**

A novel iron centre in the split-Soret cytochrome *c* from *Desulfovibrio desulfuricans* ATCC 27774

Received: 3 May 2002 / Accepted: 28 October 2002 / Published online: 19 December 2002

Abstract The facultative sulfate/nitrate-reducing bacterium *Desulfovibrio desulfuricans* ATCC 27774 harbours a split-Soret cytochrome *c*. This cytochrome is a homodimeric protein, having two bis-histidinyl *c*-type haems per monomer. It has an unique architecture at the haem domain: each haem has one of the coordinating histidines provided by the other monomer, and in each monomer the haems are parallel to each other, almost in van der Waals contact. This work reports the cloning and sequencing of the gene encoding for this cytochrome and shows, by transcriptional analysis, that it is more expressed in nitrate-grown cells than in sulfate-grown ones. In addition, the gene-deduced amino acid sequence revealed two new cysteine residues that could be involved in the binding of a non-haem iron centre. Indeed, the presence of a novel type of an iron-sulfur centre (possibly of the [2Fe-2S] type) was demonstrated by EPR spectroscopy, and putative models for its localization and structure in the cytochrome molecule are proposed on the basis of the so-far-known 3D crystallographic structure of the aerobically purified split-Soret cytochrome, which lacks this centre.

I.A. Abreu · A.I. Lourenço · A.V. Xavier · J. LeGall
A.V. Coelho · P.M. Matias · D.M. Pinto
M. Arménia Carrondo · M. Teixeira · L.M. Saraiva (✉)
Instituto de Tecnologia Química e Biológica,
Universidade Nova de Lisboa,
Rua da Quinta Grande 6,
2780-156 Oeiras, Portugal
E-mail: lst@itqb.unl.pt
Tel.: +351-214469834
Fax: +351-214468766

J. LeGall
Department of Biochemistry and Molecular Biology,
University of Georgia, Athens, GA 30622, USA

A.V. Coelho
Department of Chemistry,
Universidade de Évora,
Apartado 94, 7002-554 , Evora, Portugal

Keywords Cytochrome *c* · Iron-sulfur centre · *Desulfovibrio desulfuricans*

Introduction

The dimeric di-haem split-Soret cytochrome *c* was first isolated from the soluble fraction of the sulfate/nitrate reducer *Desulfovibrio desulfuricans* ATCC 27774 (Dd27k) by LeGall and co-workers [1]. They reported that although present in nitrate- or sulfate-grown cells, a lower yield of cytochrome was purified from cells cultured under sulfate-respiring conditions. So far, the split-Soret cytochrome *c* (Ssc) was only isolated from this organism, in spite of the large number of haem-containing proteins found in all the sulfate-reducing bacteria belonging to the *Desulfovibrio* genus [1, 2, 3, 4]. The name of this cytochrome derives from the unusual split observed in the Soret band (420 nm, with a shoulder at 415 nm) of the ferrocytochrome. The only other proteins known to contain a similar feature are the cytochrome c_{554} from *Nitrosomonas europaea* [5] and the cytochrome *bc* complex from *Rhodothermus marinus* [6]. The split-Soret cytochrome *c* is a homodimeric protein of 51.5 kDa containing two haems per monomer [1, 7] and the preliminary model for its three-dimensional structure revealed that the protein has unique structural features [7]: in each monomer the two haem groups are almost parallel and nearly stacked on top of each other; in the dimer molecule, the four haems are in close proximity, the shortest distance between the haem iron atoms of each monomer being only 9 Å, which results in a compact haem cluster arrangement. Furthermore, one of the axial ligands of each haem is provided by a histidine residue from the polypeptide chain of the other monomer.

The available amino acid sequence of Dd27k Ssc [8] showed that each monomer of the mature cytochrome is composed of a polypeptide chain of 247 amino acids

containing eight cysteine residues. Four cysteine residues are involved in two typical CxxCH haem binding motifs and it was suggested that the remaining cysteine residues (Cys29, Cys44, Cys147 and Cys155) were involved in the binding of a [2Fe-2S] cluster, although clear evidence for the presence of such a cluster has not been reported [8].

In the present work, the gene encoding the Dd27k Ssc was cloned and sequenced. Northern analyses demonstrated that the split-Soret cytochrome *c* gene is transcribed as a monocistronic operon and that the level of its mRNA is considerably lower in sulfate-grown Dd27k cells. When compared with the previously published sequence [8], the gene-derived amino acid sequence presented differences that were proved to be structurally important. Furthermore, the existence of a non-haem iron centre in this protein was demonstrated by EPR spectroscopy, and models for its structure are proposed on the basis of the available 3D crystallographic structure of this cytochrome.

Materials and methods

Cloning and sequencing of the *D. desulfuricans* ATCC 27774 split-Soret cytochrome *c* gene

Concert DNA Purification Systems from Gibco Invitrogen, the restriction enzymes and DIG Nonradioactive Nucleic Acid Labeling, Hybridization and Detection System from Roche Applied Science were used. General techniques for DNA manipulation were performed according to standard protocols [9].

Considering the *Desulfovibrio* codon usage [10], a pair of degenerated oligonucleotides was designed for the amino acid sequence near the N-terminal (YDGYWYKGF) and the C-terminal (KWCYENKIE) regions of the Dd27k Ssc [8]. A PCR product with 384 base pairs (bp) was then amplified from Dd27k genomic DNA extracted (as described in [9]) from cells grown under nitrate-respiring conditions [11]. For sequencing purposes the generated DNA fragment was purified, end repaired with Klenow polymerase, ligated to *Sma*I digested/dephosphorylated pUC19, yielding pUSsc1, and transformed in *Escherichia coli* JM109 cells. Plasmid pUSsc1 was then isolated, sequenced with the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems) and analysed on an Applied Biosystems 377A DNA Sequencer. Since the results confirmed the cloning of the partial Dd27k split-Soret cytochrome gene, this homologous fragment was random-primed DIG-labelled and used as a probe to screen a previously constructed λ-DASH genomic library of Dd27k. Plaque hybridization allowed the identification of two phages that contained the split-Soret cytochrome gene (named phages λ1.1.3 and λ1.2.2). Phage DNAs were isolated (Qiagen Lambda Maxi Kit) and sequenced, revealing the complete split-Soret cytochrome gene sequence, which was deposited in the GenBank sequence database under the accession number AF465622. Multiple sequence alignment with hierarchical clustering [12] and the Genetics Computer Group (Wisconsin) package provided by the Portuguese EMBnet Node (PEN) and Neural Networks for Promoter Prediction [13] were used to analyse the nucleotide sequence. The BLAST network service at NCBI was used for searching of homologous sequences and the Clustal X program [14] allowed the amino acid sequence alignments. Search on the sequence database of *D. vulgaris* Hildenborough, *D. desulfuricans* G20, *Desulfitobacterium hafniense* DCB-2, *Shewanella oneidensis* MR-1 ATCC 700550 and *Archaeoglobus fulgidus* genomes were done at the Institute for Genomic Research (http://www.tigr.org) or at GOLD, the Genomes Online Database (http://wit.integratedgenomics.com/GOLD/). Secondary structure predictions were done using PSIPRED v2.1 (http://bioinf.cs.ucl.ac.uk/psipred/). The three-dimensional model for the putative protein from the sulfate-reducing archaeon *A. fulgidus*, AF0016 [15], was obtained from Swiss Model (http://www.expasy.ch/swissmod/SWISS-MODEL.html) [16, 17, 18] using as template the 3D structure of the aerobic form of Ssc [7]. In order to avoid errors related to deviations between the model and its template structure [18, 19, 20, 21], other independent programs were also utilized for the prediction of the secondary and tertiary structures, yielding identical results (data not shown).

Transcriptional analysis

Two synthetic oligonucleotides complementary to Dd27k *ssc* were used (5′-GGAGGGAATTCATGAACATCGG3′ and 5′-CGGGCAAAGGATCCTATCACA-3′) in a PCR amplification using *Pfu* polymerase and λ1.1.3 phagic DNA. The PCR product (877 bp) was cloned into previously digested *Eco*RV-pZErO-1, yielding plasmid pZSsc2 and transformed in *E. coli* XL2-Blue cells. After isolation, pZSsc2 was linearized with *Eco*RI and used as template to generate a homologous anti-sense DIG-RNA probe, following the DIG-RNA Labeling Kit (SP6/T7) protocol (Roche Applied Science).

Total RNA was extracted (Qiagen RNeasy extraction kit) from the mid-log phase of Dd27k cells grown anaerobically with lactate/nitrate or lactate/sulfate. RNAs were dissolved in diethyl pyrocarbonate (DEPC)-treated water and quantity and purity determined by UV spectrophotometry at 260 and 280 nm. Size fractionation of the total RNAs (2 μg) was carried out by electrophoresis on a 1.2% (w/v) agarose (Seakem LE from FMC) gel in 0.7% (v/v) formaldehyde, using a RNA ladder (0.24–9.5 kb) from Gibco Invitrogen. The 23S and 16S ribosomal bands confirmed the integrity and the amount of RNA loaded (Fig. 1B). RNAs were then transferred to positively charged nylon membranes (Roche Applied Science), and hybridization, using the previously prepared DIG-RNA *ssc* probe, was carried out according to the DIG System protocols (Roche Applied Science). In order to quantify the effect of the growth media on the relative amount of the split-Soret mRNA, the autoradiographs were subjected to scanning densitometry (Molecular Dynamics-Computing Densitometer) and bands intensities were quantified using the ImageQuant v3.1 program.

Purification of the *D. desulfuricans* ATCC 27774 split-Soret cytochrome *c*

Cell growth

The Dd27k cells were grown on a lactate/nitrate medium as previously described [1, 11]. Cells were broken in a French Press at 9000 psi and centrifuged 30 min at 10,000x*g* to separate the cell debris, yielding the crude extract. The crude extract was ultracentrifuged at 160,000x*g*, for 1 h, and the supernatant (soluble extract) decanted. For the preparation of the anaerobic cell-free extract, all the previous steps were done under an argon atmosphere [22].

Protein purification

The purification procedures were performed at 4 °C and pH 7.6:

1. Anaerobic purification was done in a Coy anaerobic chamber model A-2463. The anaerobically prepared soluble extract (800 mL) was loaded at 1 mL/min on a DEAE-52 column (6×39 cm) previously equilibrated with 10 mM Tris-HCl, and a two-step (2 L each) linear gradient, 10–250 mM Tris-HCl and 250–400 mM Tris-HCl, was applied. A fraction containing

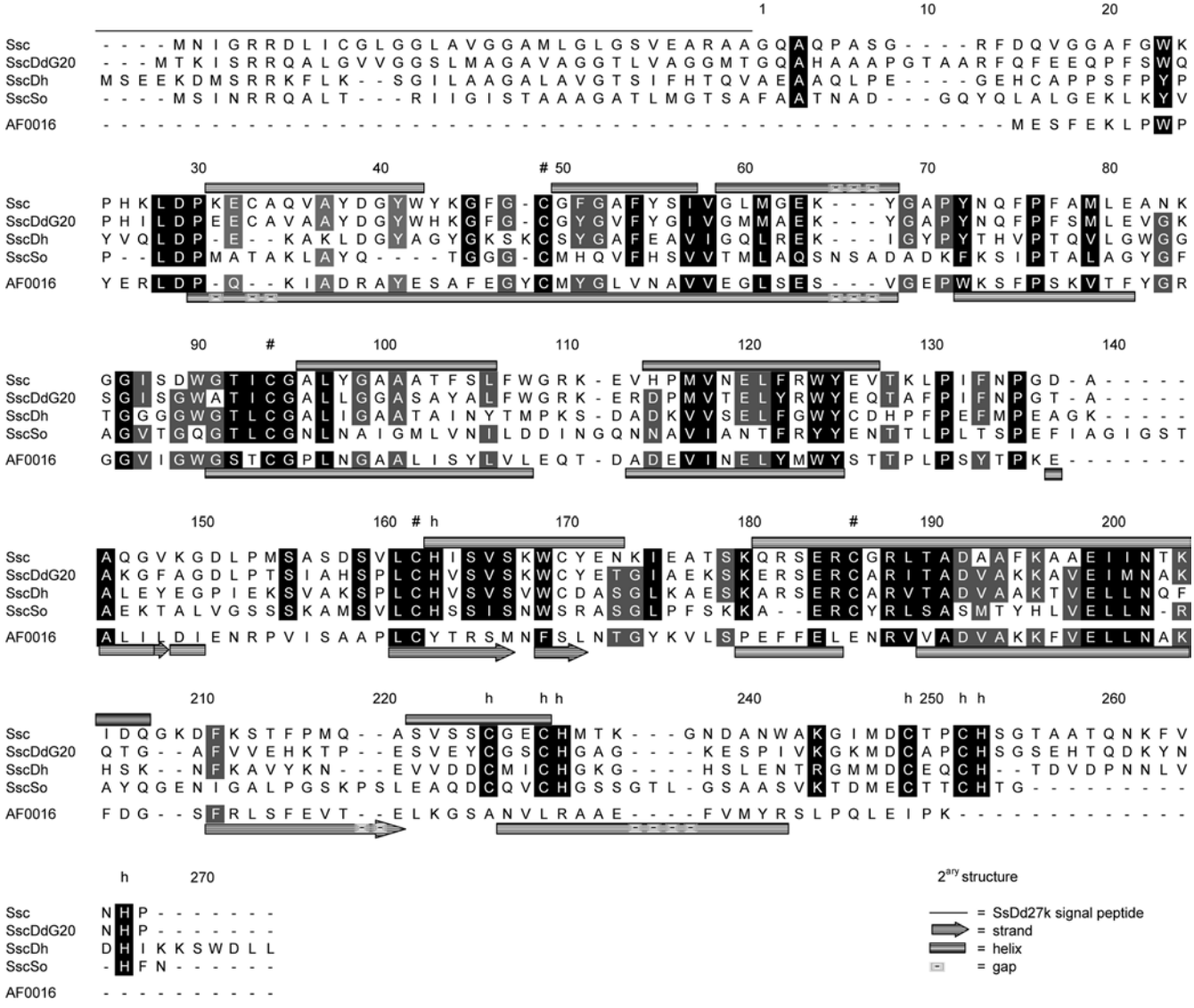

Fig. 1 Comparison between the primary structures of Dd27k Ssc and the putative gene products of *D. desulfuricans* G20 (Ssc DdG20), *Desulfitobacterium hafniense* (SscDh), *Shewanella oneidensis* (SscSo) and *A. fulgidus* AF0016. Amino acid sequence alignments produced in Clustal W 1.60 [14]; protein weight matrix: BLOSUM series. The symbols h and # represent the amino acids binding the haems and the [Fe-S] cluster, respectively. The signal peptide in the Dd27k split-Soret cytochrome is highlighted with a *full* line. The figure also compares the secondary structure of Dd27k, obtained from the 3D structure, with the PSIPRED 2.1 predicted secondary structure of *A. fulgidus* AF0016

cytochromes was eluted at ~250 mM. This fraction was then applied on a hydroxyapatite column (3×30 cm) (flow rate of 0.6 mL/min), equilibrated with 300 mM Tris-HCl. The column was eluted using a 0.5 L descending linear gradient (250 mM to 1 mM) of Tris-HCl, followed by a 1 L ascending linear gradient of 1–300 mM potassium phosphate buffer. The fraction containing the Ssc was eluted at ca. 150 mM phosphate and concentrated in a Diaflo apparatus (Amicon, Danvers, Mass., USA) with a YM3 membrane. This fraction was further purified through two passages on a Superdex-75 column (XK16/60; flow rate 0.5 mL/min).

2. For aerobic purification the aerobically prepared soluble extract was loaded on a DEAE fast-flow column, previously equilibrated with 5 mM Tris-HCl. A linear NaCl gradient (0.05–1 M) was then applied and the fraction containing the Ssc was eluted at 0.3 M. This fraction was then dialysed overnight in a 3 kDa cut-off tube against 5 mM Tris-HCl buffer, and loaded on a Q-Sepharose column, equilibrated with 10 mM Tris-HCl. This column was eluted with a 0–1 M NaCl linear gradient. The fraction containing the protein (eluted at 0.25 M) was subsequently loaded on a HTP column equilibrated with 5 mM potassium phosphate buffer and eluted with a 0–1 M potassium phosphate linear gradient. The desired fraction was eluted at 0.2 M, concentrated in a Diaflo apparatus with a YM3 membrane and finally loaded on a Superdex-75 column (XK16/60; flow rate 0.5 mL/min) in 50 mM Tris-HCl, 0.1 M NaCl. The purity of the anaerobic and aerobic Ssc fractions was confirmed by SDS-PAGE.

Chemical analysis and pyridine-hemochrome

Haem content was determined by pyridine-haemochrome using the molar absorptivity $\varepsilon_{R\text{-}O,550-535} = 23.97\ mM^{-1}\ cm^{-1}$ for the *c*-type pyridine-haemochrome [23]. Non-haem iron content was chemically determined by the 2,4,6-tripyridyl-*s*-triazine (TPTZ) method [24].

Spectroscopic analysis

EPR spectra were measured on a Bruker ESP 380 spectrometer equipped with an Oxford Instruments continuous-flow helium cryostat. UV-visible spectra were recorded on a Shimadzu UV-1603 spectrophotometer. EPR spectra were simulated as described [25], using spectra obtained under non-saturating conditions. Spectra of the reduced cytochrome and of appropriate standards (copper nitrate and myoglobin azide solutions) were recorded under non-saturating conditions and quantitated by double integration of the experimental spectra and of the respective theoretical simulations.

Refinement of the 3D structure of the *D. desulfuricans* ATCC 27774 split-Soret cytochrome *c*

The preliminary model of the three-dimensional structure obtained previously [7] was edited according to the first published sequence data [8]. Refinement at 2.5 Å resolution using diffraction data obtained earlier from the aerobically purified and crystallized protein [7] was carried out with program X-PLOR [26]. Two-fold non-crystallographic symmetry (NCS) restraints were used between the four independent molecules in the asymmetric unit: the two molecules in each dimer were treated as different but were NCS-related to the corresponding molecule in the other dimer. In all refinements and map calculations with X-PLOR a bulk solvent correction was used and all low-resolution data were included. Prior to the refinement calculations, a random 5% sample of the reflection data was flagged for *R*-free calculation [27]. All model interactive visualization and editing was carried out using TURBO [28]. An initial simulated annealing refinement was carried out at 3000 K, followed by refinement of positional parameters and individual restrained *B*-factors were refined for non-hydrogen atoms. The model was then improved by manual adjustments to the electron density, using $2|F_o|-|F_c|$, $|F_o|-|F_c|$ maps and simulated annealed omit maps. In the final refinement stages the model was further corrected according to the newly determined Ssc sequence (see above) and a total of 188 water molecules was included. The original residue numbering scheme was kept, according to the mature protein, and ignoring the 32 residues in the signal peptide. This refinement converged to values of *R* and R_{free} of 19.9% and 26.4%, respectively. The final refinement results and statistics are shown in Table 1. The final model coordinates as well as the structure factors have been deposited with the Protein Data Bank [30], with accession codes 1h21 and r1h21sf, respectively.

Results and discussion

Dd27k split-Soret cytochrome: features of the nucleotide gene sequence and analysis of the complete amino acid sequence

The sequencing of part of the phage λ1.1.3 isolated from the Dd27k genomic library (GenBank AF465622) was shown to contain the *ssc* coding unit (840-nt). Located 6-bp upstream of the *ssc* start codon a ribosome binding site-like sequence is observed, which is preceded by a nucleotide sequence that has a high probability of acting as the promoter region (between nt-9 and -54). Downstream of the *ssc* stop codon, a predicted stem-loop structure that may serve as a *rho*-independent transcription terminator signal is also detected.

The Dd27k split-Soret cytochrome *c* gene encodes a 279 amino acid precursor polypeptide, carrying a signal peptide of 32 amino acid residues. The gene-deduced amino acid sequence differs by three residues from the one previously reported [8], namely Glu2 (previously Gln), Cys86 (previously Tyr) and Cys171 (previously Ala). The fact that positions 86 and 171 harbour cysteine residues had strong implications on the analysis of the 3D structure (see below). A BLAST search in the nucleotide/protein databases revealed that Dd27k Ssc potential homologues are present in *D. desulfuricans* G20 (54% identity and 64% similarity), *Desulfitobacterium hafniense* DCB-2 (31% identity and 46% similarity) and *Shewanella oneidensis* MR-1 ATCC 700550 (23% identity and 40% similarity) (Fig. 1). Dd27k Ssc also shares some degree of similarity with a hypothetical protein (AF0016) present in the sulfate-reducing thermophilic archaeon *A. fulgidus* [15], as previously noted [8]. This is a particularly interesting case, since the *A. fulgidus* protein is

Table 1 Final refinement statistics for Ssc[a]

Resolution limits (Å)				20.0–2.50
Final *R*-factor (%) [32,158 reflections, $F_o > 2\sigma(F_o)$]				19.9
Final free *R*-factor (%) [1681 reflections, $F_o > 2\sigma(F_o)$]				26.4
Number of non-hydrogen protein atoms				7728
Number of solvent molecules				188
Estimated overall r.m.s. coordinate error (Å) from a σ_A plot [29]				0.35
	Molecule A	Molecule B	Molecule C	Molecule D
Model r.m.s. deviations from ideality				
Bond lengths (Å)	0.014	0.017	0.015	0.014
Bond angles (°)	1.73	1.87	1.72	1.70
Model completeness				
Zero occupancy non-hydrogen atoms	6	10	3	6
Whole residues omitted from model	1–7	1–7	1–7	1–7
Average *B* values (Å^2)				
Main-chain protein	38.7	34.4	35.0	36.7
Side-chain protein	39.1	35.4	35.8	36.6
Haem groups	36.0	35.1	38.6	35.8
Solvent molecules	33.8			

[a]The two independent dimers in the asymmetric unit were made up of molecules A,B and C,D. The NCS restraints used in the refinement made equivalent molecules A,D and B,C

the only one that lacks the two *c*-type haem binding motifs (CxxCH) present in the C-terminal region of Dd27k Ssc, but contains some of the cysteine residues that in the split-Soret cytochrome *c* are predicted to bind the [2Fe-2S] centre. In addition, the 3D structure of the aerobic form of Dd27k Ssc and the one predicted for the *A. fulgidus* protein share relevant similarities (see below).

Expression of the split-Soret cytochrome *c* gene

A unique transcript of ~900 bp was observed in all RNAs analysed (Fig. 2A), confirming that *ssc* is transcribed as a monocistronic operon. Although *ssc* was expressed during both nitrate and sulfate growth, the densitometric analyses of the Northern blots revealed that the level of its mRNA was significantly lower in sulfate-grown cells of Dd27k. Considering as 100% the level of *ssc* mRNA in Dd27k nitrate-grown cells, sulfate-grown cells contained only 9% of *ssc* mRNA (Fig. 2C). This result contrasts, for example, with that obtained for the Dd27k nine-haem operon, for which the level of the mRNA was ~70% lower in nitrate-grown cells [31]. It should also be noted that the present result agrees with previous data reporting that the yield of the split-Soret cytochrome *c* isolated from sulfate-grown cells was lower when compared with the yield obtained from nitrate-grown cells [1].

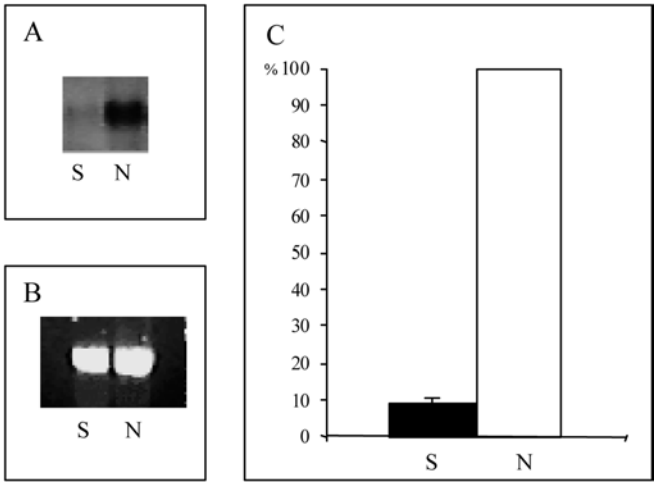

Fig. 2A–C Analysis of *ssc* transcription in Dd27k cells. **A** Total RNA (2 μg) was isolated at early-log phase of anaerobically nitrate- (N) or sulfate- (S) grown cells. Hybridization was carried out with the dioxigenin-labeled RNA probe, as described in Material and methods. **B** The 23S rRNA was used as loading control. **C** Densitometric analysis of Northern blots of **A**, considering the *ssc* mRNA level for nitrate-grown cells as 100%. Values correspond to the average of five independent experiments

EPR studies of the aerobically and anaerobically purified Dd27k split-Soret cytochrome *c*

The Dd27K Ssc was purified to homogeneity, both aerobically and anaerobically, yielding a single band on SDS-PAGE at 27 kDa. The non-haem iron content, determined in relation to the haem content obtained by the pyridine-haemochromogen method, shows that the aerobically purified protein contains a lower non-haem iron content (0.8 atoms Fe/per monomer) when compared with the one determined for the anaerobically purified protein (1.4 Fe atoms/per monomer).

The two preparations exhibit identical visible spectra in both the oxidized and reduced states (data not shown). The EPR spectra of the oxidized states are also identical, with *g*-values identical to the ones previously reported [1]. However, upon reduction with sodium dithionite, resonances at $g \approx 2.01$, 1.95 and 1.89 are observed (Fig. 3A). These signals are detected

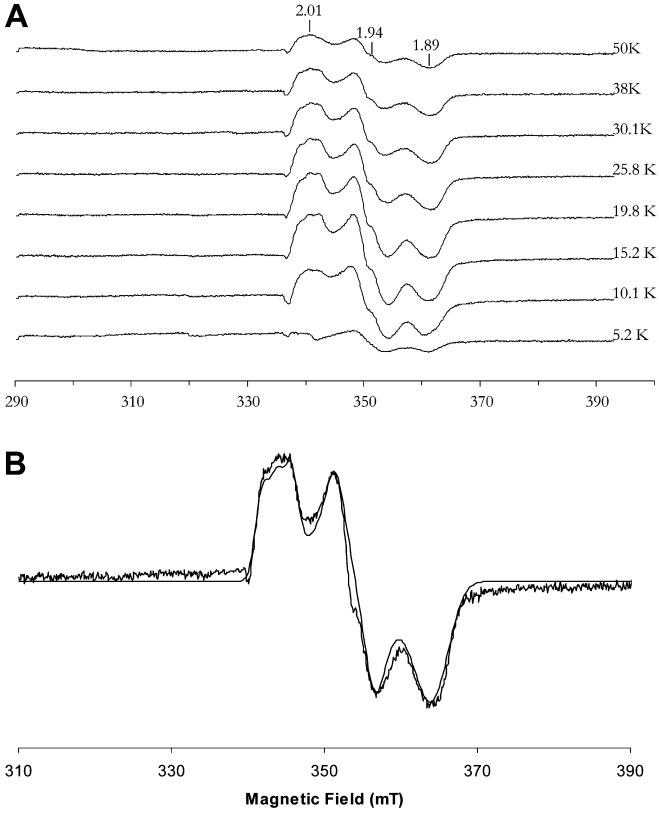

Fig. 3 **A** EPR spectra of Dd27k split-Soret cytochrome reduced with sodium dithionite, at the indicated temperatures. Microwave power, 2.4 mW; microwave frequency, 9.44 GHz; modulation amplitude, 1 mT. **B** Spectrum of Dd27k split-Soret cytochrome, reduced with sodium dithionite, at 21 K and a microwave power of 24 μW, superimposed with a spectral simulation (*dotted* line). The following parameters were used for the simulation: $g = 2.004$, 1.946, 1.892; linewidth of 2.1, 2.6 and 4.0 mT; hyperfine constant ($I = 1$) of 1.4 mT for the g_{max} and g_{med} resonances; at g_{min} no resolved hyperfine splitting is observed

Table 2 Structural analysis statistics for Dd27k Ssc

Ramachandran plot	Molecule A		Molecule B		Molecule C		Molecule D	
	%	Nr.	%	Nr.	%	Nr.	%	Nr.
Core regions	89.1	180	90.1	182	88.1	178	89.6	181
Additionally allowed regions	9.4	19	9.4	19	10.9	22	8.9	18
Generously allowed regions	1.5	3	0.5	1	1.0	2	1.5	3
Disallowed regions	0	0	0	0	0	0	0	0
ProCheck statistics								
Morris et al. class	1 2 2		1 2 2		1 2 2		1 2 2	
Overall *G*-factor	0.26		0.26		0.26		0.27	
cis -Proline residues	63, 122		63, 122, 247		63, 122, 247		63, 122	

at least up to 50 K, with only negligible line broadening, and at 21 K the power for half saturation is 3.5 mW. These relaxation properties are characteristic of reduced iron-sulfur centres of the [2Fe-2S] type. The g_{max} resonance shows an apparent hyperfine splitting of approximately 20 mT, and indeed the EPR spectrum could be simulated assuming a hyperfine interaction with an $I=1$ nucleus, leading to a well-resolved triplet at $g=2.004$ (Fig. 3B). However, at this stage, this interpretation of the EPR spectrum is presented just as a hypothesis, since, mainly due to the cluster instability, it is not possible to exclude that the apparent splitting may result from the superposition of multiple slightly different conformations of the iron centre; nevertheless, it should be noted that the spectra do not change their lineshape, apart from saturation at very low temperatures or very high microwave power. Also, since this putative splitting was not observed in all samples, it most probably would not be due to an intrinsic protein ligand (in agreement with the structural prediction, see below).

The reduction potential of this centre appears to be quite low. In fact, redox titrations up to pH 9 (monitored by EPR spectroscopy and performed anaerobically in the presence of redox mediators and using sodium dithionite as reductant) did not lead to the complete reduction of the centre, showing that at, for example, pH 8 its reduction potential is lower than ~–500 mV. In agreement with this result, EPR quantification of the cluster spectrum does not yield a stoichiometric value of the amount of reduced cluster in relation to the non-haem iron content, in either the anaerobic or aerobic preparations (values lower than 0.5 spin per monomer were obtained for the anaerobically purified sample, which has the highest iron content).

The low amount of iron observed for the aerobically purified protein strongly indicates that this cluster is quite sensitive to oxygen exposure, thus explaining why so far it has been so difficult to prove its presence. The centre now observed probably corresponds to the one proposed earlier on the basis of Mössbauer spectra of the aerobically purified cytochrome, which displayed in the oxidized form a diamagnetic component, substoichiometric in relation to the dominant haem components, with average isomeric shifts and quadrupole splittings typical of oxidized [2Fe-2S] centres [8].

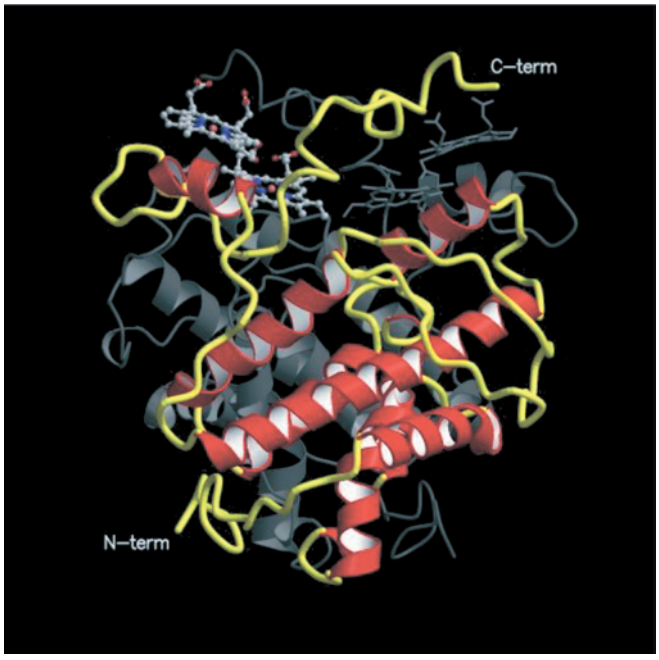

Fig. 4 Ribbon diagram of the 3D structure of the aerobic form of Dd27k Ssc. One monomer is drawn in colour: *yellow* represents loops, *red* denotes α-helices, and haem groups are represented in *ball-and-stick* using the standard atomic colouring scheme. The second monomer is behind the first and is drawn in *gray*. The non-crystallographic two-fold symmetry axis is vertical and in the plane of the page. Figure created with Molscript [34] and Raster3D [35]

Analysis of the 3D structure of the *D. desulfuricans* ATCC 27774 split-Soret cytochrome *c*

The refined model for the 3D structure was analysed with PROCHECK [32] and its stereochemical quality parameters were within their respective confidence intervals. The main Ramachandran [33] φ,ϕ plot and PROCHECK results are listed in Table 2.

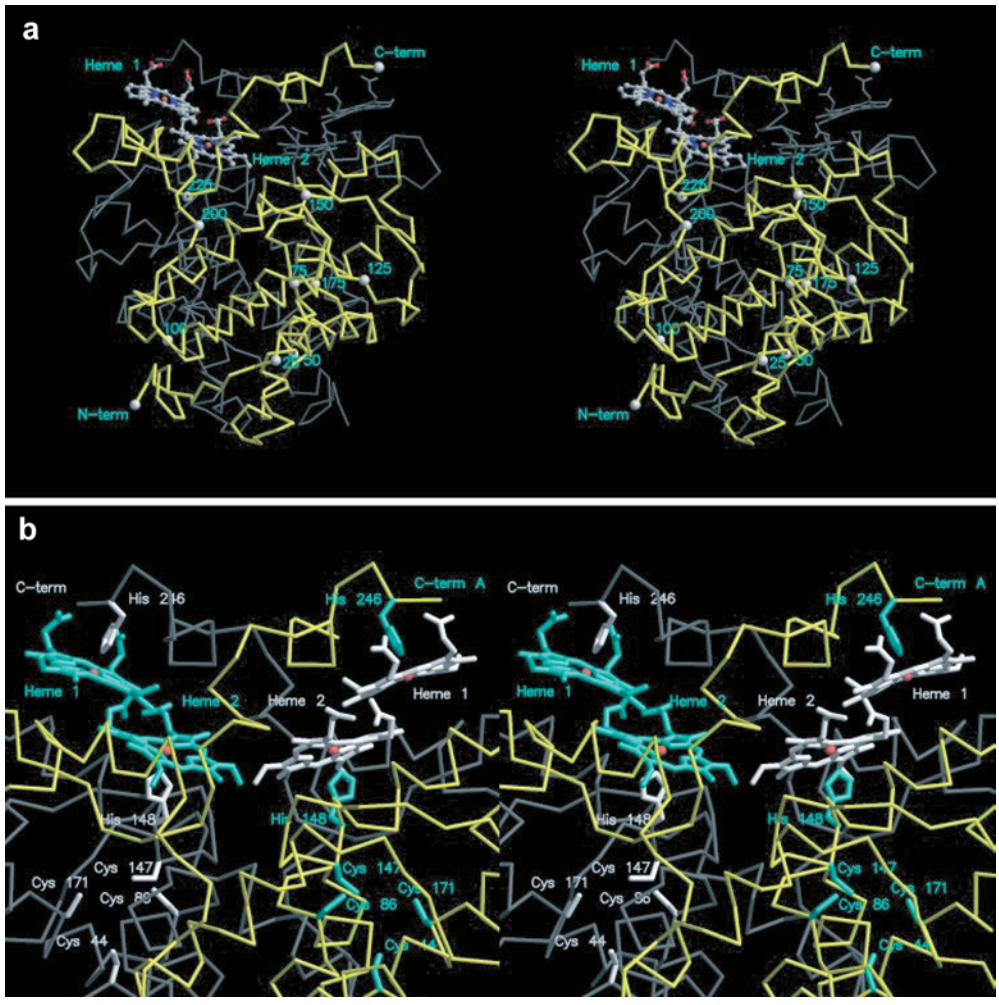

Fig. 5a, b Stereo views of the of the 3D structure of the aerobic form of Dd27k Ssc. **a** The C^{α} trace of one monomer is drawn in *yellow* and the C^{α} atoms in the protein chain are numbered about every 25 residues. Haem groups are represented in *ball-and-stick* using the standard atomic colouring scheme. The second monomer is behind the first and is drawn in *gray*. **b** Detailed view of the haem core showing the stacking arrangement in each monomer as well as the side chains of the histidine and cysteine residues discussed in the text. Residues coloured and labelled in *cyan* belong to the first monomer (C^{α} trace drawn in *yellow*). Residues coloured and labelled in *white* belong to the second monomer (C^{α} trace drawn in *gray*). In each monomer, His148 is directly attached to haem 2 and may provide a short electron transfer route between the haem core and the putative binding pockets of the two [2Fe-2S] centres in the molecule. The non-crystallographic two-fold symmetry axis is vertical and in the plane of the page. Figure created with Molscript [34] and Raster3D [35]

Overall, and excluding the obvious differences due to the lack of sequence information, the refined three-dimensional structure of Ssc is very similar to the preliminary model [7]. The main difference was that the initial model was three residues shorter, with two missing residues (Lys20 and Ser164) being located in loop regions for which the electron density maps were less clear and the terminal Pro 247 was not modelled, also due to unclear electron density. As shown in Fig. 4, Ssc is a cytochrome with a most unusual fold. First, the haem groups are located at one end of the molecule rather than being tightly wrapped by the polypeptide chain. Second, they display a stacking arrangement that was interpreted as the origin of the characteristic splitting of the Soret band in the UV-visible spectrum of the reduced protein. Ssc was the first reported example of such a haem stacking arrangement, which has since been found in other structures (see for example [5, 36]). Finally, the second axial histidine ligand of the iron atom in the haem groups of one monomer in the dimer is actually provided by a histidine residue from the other monomer, a clear evidence that the dimer is the functional unit of Ssc. Calculations with AREAIMOL [37] show that over 36% of the total molecular surface of each monomer, i.e. 4,942 $Å^2$ in 13,640 $Å^2$, are buried by the dimerization process. Figure 5a shows an overall view of the molecule, and includes the residue and haem numbering scheme, while Fig. 5b shows a close-up view of the haem core and surroundings, which further highlights the above discussion.

Correlation of the three-dimensional structure of the aerobic form of the *D. desulfuricans* ATCC 27774 split-Soret cytochrome *c* with the spectroscopic EPR data

In the refined Ssc structure, no evidence of a non-haem iron-sulfur centre was observed, probably because this centre is labile under aerobic conditions (see previous sections). In fact, the crystals that were used to determine the structure were obtained from preparations for which no special precautions for anaerobicity were taken, either in the purification or the crystallization steps. Further crystallization trials have, so far, systematically failed to produce good diffracting crystals. If this non-haem iron-sulfur centre was present in

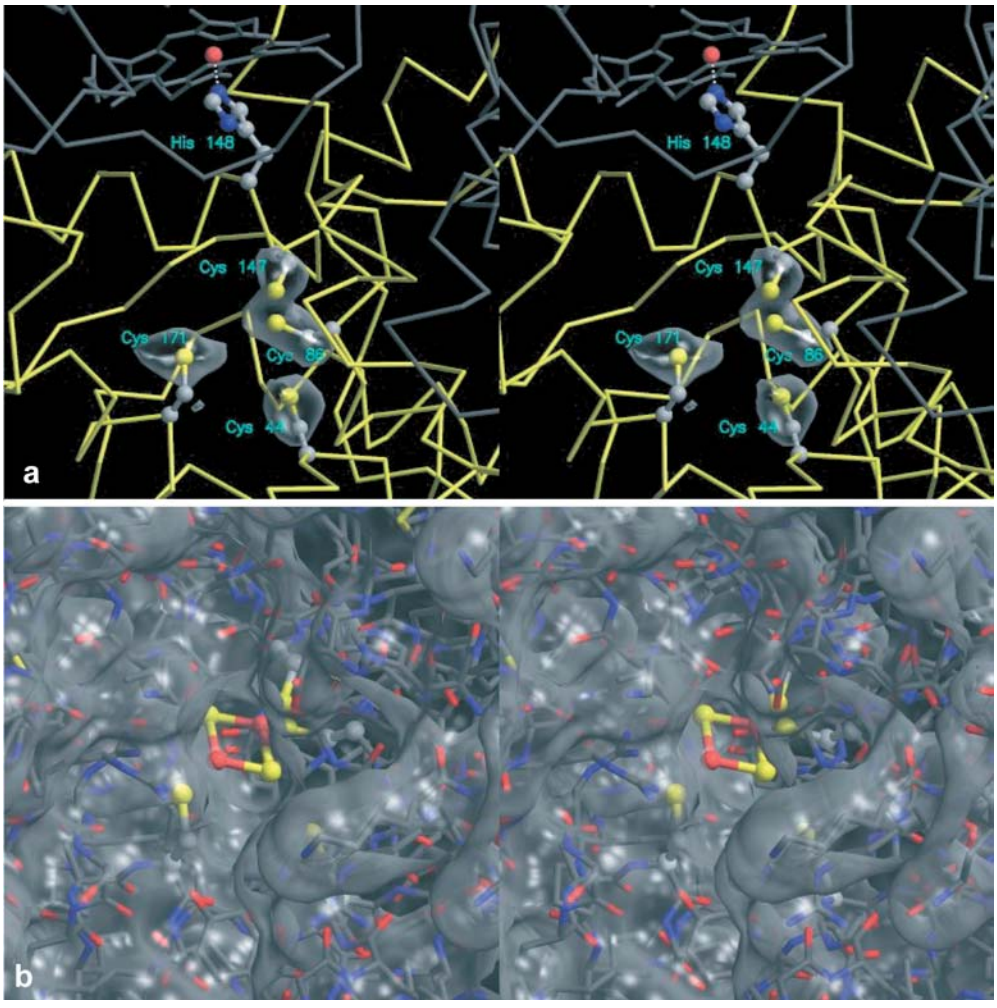

Fig. 6a, b Stereo views of the putative binding pocket of the [2Fe-2S] centre in the 3D structure of the aerobic form of Ssc Dd27k. A modelled [2Fe-2S] centre and the side chains of Cys44, Cys86, Cys147, His148 and Cys171 are represented in *ball-and-stick* mode, with atom colours by atom type, the rest of the molecule being represented by its C^{α} trace. **a** $|F_o - F_c|$ electron density map contoured at the 3σ level, calculated after a refinement cycle where the residues at positions 44, 86, 147 and 171 were truncated to Gly, showing electron density consistent with the presence of four Cys residues at those positions. His148 coordinates the iron atom from a haem group in the other monomer. **b** Molecular surface calculated with a 1.4 Å probe radius, showing the existence of a surface pocket in the 3D structure of Ssc, which could accommodate a [2Fe-2S] centre. Figures are drawn in the same orientation and were prepared with Molscript [34], Raster3D [35], Bobscript [38] and MSMS [39]

the aerobic structure, even sub-stoichiometrically, the Fe atoms should have been clearly observed in a Fourier map using the anomalous differences as coefficients.

The most remarkable implication of the three-dimensional structure of this cytochrome resulted from the corrections to the primary structure reported above. While the Tyr86Cys error had been realized early in the refinement, that was not the case for the Ala171Cys error. Figure 6a shows the $|F_o|-|F_c|$ electron density used to validate the presence of the four Cys residues at positions 44, 86, 147 and 171. In the three-dimensional structure of Ssc, Cys86 and Cys171 occur near two other Cys residues (Cys44 and Cys147). The obvious consequence of this fact is that these four Cys residues could constitute a binding site for a tetra-coordinated non-haem iron-sulfur centre. The presence of a non-haem iron centre was previously proposed [8]; however, the cysteine residues considered by those authors as being involved in the binding of the centre (Cys29, Cys44, Cys147 and Cys155) are not, with the exception of Cys44 and Cys147, located near one another in the 3D structure of Ssc, which invalidated that hypothesis. Besides being located in close proximity in the 3D structure of Ssc, as shown in Figs. 5b and 6a, the four Cys residues (Cys44, Cys86, Cys147 and Cys171) which are now proposed to form the binding site of a non-haem iron-sulfur centre are located at an open surface pocket in the protein which is accessible from the outside, as illustrated in Fig. 6b. The pocket volume calculated with CASTp [40] of the basis on its molecular surface is 680 $Å^3$. Thus, solely on the basis of the pocket size, the non-haem iron-sulfur centre could be of the [2Fe-2S] or [4Fe-4S] type, which occupy volumes of 264 and 416 $Å^3$, respectively, as calculated by ASC [41] also on the basis of their molecular surfaces. The S^{γ} atoms of the four cysteine residues for each of the four independent molecules in the asymmetric unit are located at the vertices of a distorted tetrahedron with S-S distances ranging between 3.7 and 7.4 Å. This arrangement might suggest a [4Fe-4S] centre. However, the available spectroscopic data (EPR (see above) and Mössbauer [8]) favour a [2Fe-2S] centre instead. The coordinating S^{γ} atoms of these clusters are usually co-planar and located at the corners of a rectangle. Therefore, the tetrahedral arrangement of the coordinating S^{γ} atoms in Ssc, unless it turns out to be an artefact of the structure relaxation following the aerobic degradation of the iron-sulfur centre, will originate an

unusual coordination of the putative [2Fe-2S]. Also, since the molecule is a homodimer, there are two such possible [2Fe-2S] binding sites in the molecule.

A simple modelling of an idealized [2Fe-2S] centre into this four-Cys binding site showed that under some circumstances it could be fit into the binding pocket with only a minor structural rearrangement of the protein. Two binding schemes were considered and are represented in Fig. 7: (a) a fully tetra-coordinated [2Fe-2S] centre and (b) a tri-coordinated [2Fe-2S] centre, where Cys44 is not ligated and one of the Fe coordinating positions would remain vacant. As shown in Fig. 7, because of the tetrahedral arrangement of the S^{γ} atoms in the observed crystal structure of Ssc, the typical [2Fe-2S] coordination mode in scheme (a) introduces a far greater strain in the molecular structure of Ssc, requiring a significantly larger readjustment of the structure in order to accommodate the [2Fe-2S] centre with an acceptable coordination geometry. On the other hand, scheme (b) only requires small (< 0.6 Å) changes in the positions of the Cys S^{γ} atoms which could be mainly accommodated by changes in their C-C^{α}-C^{β}-S^{γ} dihedral angles. The presence of a [2Fe-2S] centre with only three cysteine ligands, although unusual, is not unprecedented, as in NifU from *Azotobacter vinelandii* one of the iron-sulfur centres is bound by only three cysteines, resulting in a labile centre. The lability of this cluster is functionally important for the role proposed for NifU in the assembly of the FeS centres [42]. Interestingly, the vacant Fe coordinating position in the hypothetical tri-coordinate binding mode of the [2Fe-2S] centre would be accessible to any possible substrate species that might interact with the exposed [2Fe-2S] centre, as illustrated in Fig. 6b. Finally, as illustrated in Figs. 5a and 6a, one of the cysteine residues in the putative binding site (Cys147) is followed in the sequence by one histidine residue (His148) which is one of the axial ligands of the innermost haem group from the other monomer. This could constitute a simple electron transfer path between the [2Fe-2S] centres (one from each monomer) and the haem core.

The unusual type of coordination of a [2Fe-2S] centre to the Dd27k split-Soret cytochrome still needs to be confirmed by the X-ray structure of the intact protein.

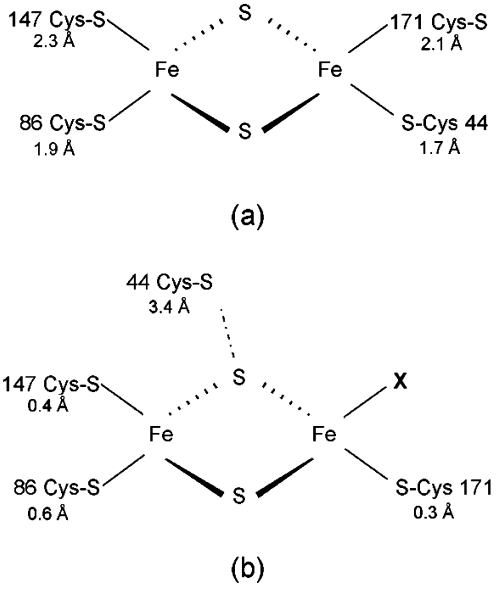

Fig. 7a, b Schemes showing two possible binding modes of the [2Fe-2S] centre in the putative binding pocket of Dd27k Ssc. **a** Idealized [2Fe-2S] centre with four Fe binding positions fit to Cys44 S^{γ}, Cys86 S^{γ}, Cys147 S^{γ} and Cys171 S^{γ}. **b** Idealized [2Fe-2S] centre with three Fe binding positions fit to Cys86 S^{γ}, Cys147 S^{γ} and Cys171 S^{γ}. The numbers below the residues indicate the calculated distances between the Cys S^{γ} atoms in the idealized centre as modelled and the actual Cys S^{γ} positions in the 3D structure of the aerobic form of Dd27k Ssc. Similar values are observed for all the four independent monomers in the asymmetric unit

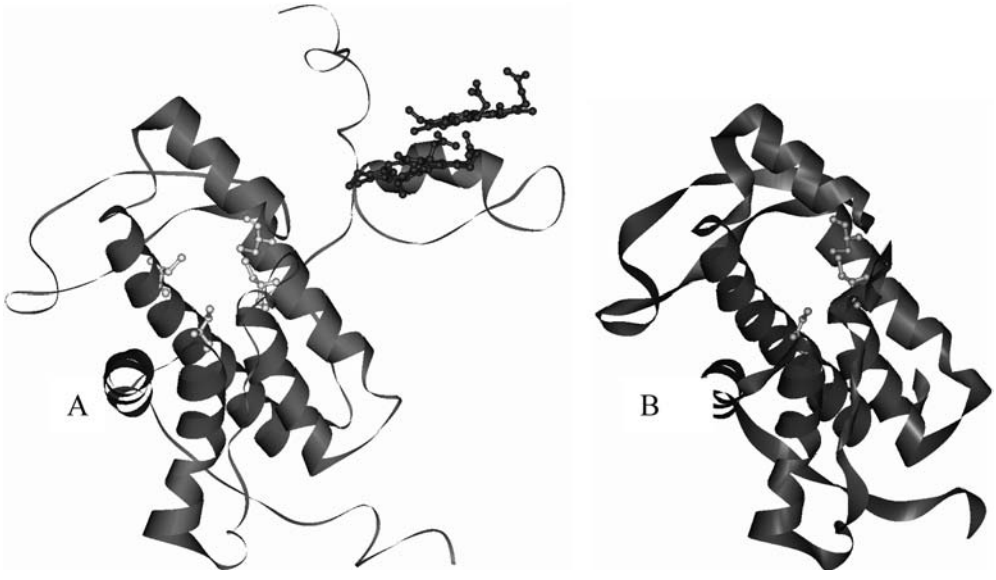

Fig. 8A, B Comparison of *A. fulgidus* AF0016 structural model with the X-ray structure of Dd27k Ssc. **A** The 3D structure of Ssc, with the N-terminal α domain and the haem-binding domain clearly separated, is represented in *red*. **B** The model for the structure of *A. fulgidus* AF0016 (in *blue*) was predicted using its sequence homology to the Ssc α domain

Since the present work revealed that the stability of this cluster depends on the degree of oxygen exposure, future work will proceed with samples prepared under anaerobic conditions, in order to confirm the structure of this novel iron-sulfur centre.

Comparison between the structure of Dd27k Ssc and the model for *A. fulgidus* AF0016

As expected, considering the significant amino acid similarity between Dd27k Ssc and the putative homologues from *Desulfovibrio desulfuricans* G20, *Desulfitobacterium hafniense* DCB-2 and *Shewanella oneidensis* MR-1, prediction of their 3D structures (data not shown) indicate that all the proteins may have similar structures.

In the structure of the Ssc monomer, two domains can be considered: an N-terminal α domain, containing nine helices and including the putative ligands for the iron-sulfur cluster, and a C-terminal haem-binding domain. To our knowledge, none of these domains have structural similarity with any other known 3D structure (results from a search made using the DALI server at EBI-EMBL, on the PDB database, http://www.ebi.ac.uk/dali). As mentioned above, the sequence database search also detected a putative protein, *A. fulgidus* AF0016, with significant primary structure similarity with Dd27k Ssc (23% identity and 38% similarity) (Fig. 1), that does not contain the haem binding motifs. The comparison between the primary structure of just the α domain of Dd27k Ssc (from amino acid residues 11 to 199) with part of AF0016 sequence (from amino acid residues 1 to 188) leads to an increase in the identity and similarity values to 27% and 44%, respectively. Thus, it was interesting to extend this comparison at the level of the secondary and tertiary structures of the two proteins. In fact, the secondary structure of the Dd27k Ssc and that predicted for the AF0016 protein were found to be very similar (Fig. 1), and a model for the 3D structure of AF0016 could be obtained (Fig. 8) using the N-terminal α domain of the Ssc 3D structure as a template. These results suggest that the 3D structure of the non-haem binding domain of the split-Soret cytochrome shares relevant common features with the structure of the AF0016 putative protein. Hence, the *A. fulgidus* protein could also accommodate an iron-sulfur centre, but it only contains three cysteine residues, lacking that equivalent to Dd27k Ssc Cys171, which is substituted by a glutamate (Glu155).

Conclusions

The cloning and sequencing of the Dd27k Ssc confirmed its periplasmic localization and revealed its signal peptide sequence. The correct amino acid sequence had implications in the structure of the protein since it was realized that the two new cysteine residues (86 and 171) are located in a surface pocket in the protein, along with, and in close proximity to, two other cysteine residues (44 and 147). Altogether they may form the binding site for a [2Fe-2S] centre that, however, was not detected in the crystal structure obtained from an aerobically prepared sample. The instability of this centre was shown to be related to the degree of oxygen exposure during the purification procedures, since the anaerobically prepared sample of the split-Soret cytochrome contains a considerable higher amount of non-haem iron. A simple modelling of an idealized [2Fe-2S] centre into the putative binding pocket, located near the mouth of a surface pocket in each monomer, suggests a possibly unusual coordination, where only three of the four available Cys residues (86, 147 and 171) bind the two iron atoms, leaving one free coordination position which is accessible from the outside of the molecule. An easy electron transfer route between the [2Fe-2S] centre and the haem core can be achieved via His148, which coordinates the iron atom in one of the haem groups in the other monomer. The overall structure of the haem core and of the iron-sulfur centres therefore suggests that this cytochrome has a not yet found enzymatic activity. The fact that the levels of *ssc* mRNA in nitrate-grown cells of Dd27k are considerably higher than in sulfate-grown cells may indicate an involvement in the nitrate respiratory chain of *D. desulfuricans*. This hypothesis is consistent with the finding of genes encoding for very similar putative proteins in the genomes of other bacteria capable of respiring nitrate or nitrite: *D. desulfuricans* G20, *Shewanella oneidensis* and *Desulfitobacterium hafniense*. Unlike the *A. fulgidus* putative protein, the putative gene products from the other three organisms contain the necessary residues for the binding of two *c*-type haems, in addition to the iron-sulfur binding domain.

The reduction of nitrite by ammonia-forming organisms appears to be straightforward as it involves a single six-electron transferring step, performed by a similar multi-haem-containing enzyme in different bacterial species such as *E. coli* [43, 44] or *D. desulfuricans* [11, 45]. However, since the split-Soret cytochrome can be found in several other nitrate/nitrite respiring organisms and its synthesis in *D. desulfuricans* ATCC 27774 is clearly induced by these conditions, the reduction of nitrite to ammonia, an important reaction in the nitrogen cycle, is most probably more complicated than it appears from our present knowledge. The presence of a putative homologous protein in the sulfate-reducing archaeon *A. fulgidus*, whose genome also encodes a putative protein involved in the respiration of nitrogen oxides [15], further supports this conclusion. Interestingly, in the genome of the sulfate reducer *D. vulgaris*, neither a cytochrome equivalent to the split-Soret cytochrome nor proteins sharing similarities with the iron-sulfur domain of the Ssc or with the *A. fulgidus* protein could be detected.

Acknowledgements We would like to thank Isabel Pacheco and Célia V. Romão for their help in the purification of the Dd27k split-Soret cytochrome *c*. This work was supported by Fundação da Ciência e Tecnologia projects POCTI/1999/BME/36558 (to M.T.) and POCTI/1999/BME/35021 (to A.V.X.).

References

1. Liu MC, Costa C, Coutinho IB, Moura JJ, Moura I, Xavier AV, LeGall J (1988) J Bacteriol 170:5545–5551
2. Pereira IAC, Teixeira M, Xavier AV (1998) Struct Bonding 91:65–89
3. Coelho AV, Matias P, Sieker LC, Morais J, Carrondo MA, Lampreia J, Costa C, Moura JJ, Moura I, LeGall J (1996) Acta Crystallogr Sect D 52:1202–1208
4. Coutinho IB, Xavier AV (1994) Methods Enzymol 243:119–140
5. Arciero DM, Collins MJ, Haladjian J, Bianco P, Hooper AB (1991) Biochemistry 30:11459–11465
6. Pereira MM, Carita JN, Teixeira M (1999) Biochemistry 38:1268–1275
7. Matias PM, Morais J, Coelho AV, Meijers R, Gonzalez A, Thompson AW, Sieker L, LeGall J, Carrondo MA (1997) J Biol Inorg Chem 2:507–514
8. Devreese B, Costa C, Demol H, Papaefthymiou V, Moura I, Moura JJ, Van Beeumen J (1997) Eur J Biochem 248:445–451
9. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1995) Current protocols in molecular biology. Greene/Wiley, New York
10. Stokkermans JP, Pierik AJ, Wolbert RB, Hagen WR, Van Dongen WM, Veeger C (1992) Eur J Biochem 208:435–442
11. Liu MC, Peck HD Jr (1981) J Biol Chem 256:13159–13164
12. Corpet F (1988) Nucleic Acids Res 16:10881–10890
13. Reese MG, Harris NL, Eeckman FH (1996) In: Hunter L, Klein TE (eds) Biocomputing: proceedings of the 1996 Pacific symposium. World Scientific, Singapore, pp 737–738
14. Thompson JD, Higgins DG, Gibson TJ (1994) Nucleic Acids Res 22:4673–4680
15. Klenk HP, Clayton RA, Tomb J-F, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey EK, Peterson JD, Richardson DL, Kerlavage AR, Graham DE, Kyrpides NC, Fleischmann RD, Quackenbush J, Lee NH, Sutton GG, Gill S, Kirkness EF, Dougherty BA, McKenney K, Adams MD, Loftus B, Peterson S, Reich CI, McNeil LK, Badger JH, Glodek A, Zhou L, Overbeek R, Gocayne JD, Weidman JF, McDonald L, Utterback T, Cotton MD, Spriggs T, Artiach P, Kaine BP, Sykes SM, Sadow PW, D'Andrea KP, Bowman C, Fujii C, Garland SA, Mason TM, Olsen GJ, Fraser CM, Smith HO, Woese CR, Venter JC (1997) Nature 390:364–370
16. Peitsch MC (1995) Bio/Technology 13:658–660
17. Peitsch MC (1996) Biochem Soc Trans 24:274–279
18. Guex N, Peitsch MC (1997) Electrophoresis 18:2714–2723
19. Chothia C, Lesk AM (1986) EMBO J 5:823–826
20. Harrison RW, Chatterjee D, Weber IT (1995) Proteins Struct Funct Genet 2:463–471
21. Bajorath J, Stenkamp R, Aruffo A (1993) Protein Sci 2: 1798–1810
22. LeGall J, Liu MY, Gomes CM, Braga V, Pacheco I, Regalla M, Xavier AV, Teixeira M (1998) FEBS Lett 429:295–298
23. Berry EA, Trumpower BL (1987) Anal Biochem 161:1–15
24. Fischer DS, Price DC (1964) Clin Chem 10:21–25
25. Aasa R, Vanngard VT (1975) J Magn Reson 19:308–315
26. Brünger AT (1992) X-PLOR, version 3.1: a system for crystallography and NMR. Yale University Press, New Haven
27. Brünger AT (1992) Nature 355:472–474
28. Roussel A, Fontecilla-Camps JC, Cambillau C (1990) In: XV IUCR congress abstracts, Bordeaux, France, pp 66–67
29. Read RJ (1986) Acta Crystallogr Sect A 42:140–149
30. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235–242
31. Saraiva LM, da Costa PN, Conte C, Xavier AV, LeGall J (2001) Biochim Biophys Acta 1520:63–70
32. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) J Appl Crystallogr 26:283–291
33. Ramachandran GN, Sasisekharan V (1968) Adv Protein Chem 23:283–437
34. Kraulis PJ (1991) J Appl Crystallogr 24:946–950
35. Merritt EA, Murphy MEP (1994) Acta Crystallogr Sect D 50:869–873
36. Iverson TM, Arciero DM, Hooper AB, Rees DC (2001) J Biol Inorg Chem 6:390–397
37. Collaborative Computational Project Number 4 (1994) Acta Crystallogr Sect D 50:760–763
38. Esnouf RM (1999) Acta Crystallogr Sect D 55:938–940
39. Sanner MF, Spehner J-C, Olson AJ (1996) Biopolymers 38:305–320
40. Liang J, Edelsbrunner H, Woodward C (1998) Protein Sci 7:1884–1897
41. Eisenhaber F, Lijnzaad P, Argos P, Sander C, Scharf M (1995) J Comput Chem 16:273–284
42. Yuvaniyama P, Agar JN, Cash VL, Johnson MK, Dean DR (2000) Proc Natl Acad Sci USA 97:599–604
43. Fujita T (1966) J Biochem (Tokyo) 60:204–215
44. Kajie S-I, Anraku Y (1986) Eur J Biochem 154:457–463
45. Pereira IAC, Abreu IA, Xavier AV, LeGall J, Teixeira M (1996) Biochem Biophys Res Commun 224:611–618