Primary structure of desulfoferrodoxin from *Desulfovibrio desulfuricans* ATCC 27774, a new class of non-heme iron proteins

Bart Devreese[a], Pedro Tavares[b], Jorge Lampreia[b], Nancy Van Damme[a], Jean Le Gall[c], José J.G. Moura[b], Jozef Van Beeumen[a,*], Isabel Moura[b,**]

[a] *Vakgroep Biochemie, Fysiologie en Microbiologie, Universiteit Gent, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium*
[b] *Departamento de Química and Centro de Química Fina e Biotecnologia, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2825 Monte da Caparica, Portugal*
[c] *Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA*

Received 26 March 1996

Abstract The primary structure of desulfoferrodoxin from *Desulfovibrio desulfuricans* ATCC 27774, a redox protein with two mononuclear iron sites, was determined by automatic Edman degradation and mass spectrometry of the composing peptides. It contains 125 amino acid residues of which five are cysteines. The first four, Cys-9, Cys-12, Cys-28 and Cys-29, are responsible for the binding of Center I which has a distorted tetrahedral sulfur coordination similar to that found in desulforedoxin from *D. gigas*. The remaining Cys-115 is proposed to be involved in the coordination of Center II, which is probably octahedrally coordinated with predominantly nitrogen/oxygen containing ligands as previously suggested by Mössbauer and Raman spectroscopy.

Key words: Amino acid sequence; Desulfoferrodoxin; Rubredoxin; Desulforedoxin; *Desulfovibrio*; Non-heme iron

1. Introduction

Brumlik and Voordouw [1] showed the existence of an open reading frame upstream from the gene encoding rubredoxin from *Desulfovibrio vulgaris*, strain Hildenborough. The N-terminal sequence appeared to be homologous to the amino acid sequence of desulforedoxin (Dx) from *D. gigas* [2], the smallest redox protein isolated so far (a dimer of a 36 amino acid containing polypeptide chain). The authors regarded the protein encoded by this gene as a potential redox partner for rubredoxin and therefore designated it as rubredoxin oxidoreductase (*rbo*). Meanwhile, Moura et al. [3] isolated a new non-heme iron protein from *D. desulfuricans* ATCC 27774, named desulfoferrodoxin (Dfx). N-terminal sequence analysis showed that it was very similar to the putative *rbo* gene product. Mössbauer and EPR spectroscopic measurements showed that Dfx contains two mononuclear iron sites: an iron site I (Center I) with a distorted tetrahedral sulfur (cysteinic) coordination similar to the one found in Dx, and an iron site II (Center II) which appears to be octahedrally coordinated to nitrogen/oxygen/sulfur containing ligands [4]. This protein can exist in three different oxidation states. In the fully oxidized state (grey form) both irons are in the ferric state. Upon one-electron reduction, a semi-reduced form (pink form) is obtained corresponding to the reduction of Center II. Finally, the fully reduced state (in which both irons are in the ferrous state) is obtained by two-electron reduction. Aerobic purification procedures demonstrated that it was possible to isolate this protein in the first two of these redox states.

The structure of Center I is reminiscent of the one existing in the widespread small $FeCys_4$-containing proteins such as rubredoxin, in which the iron atom has a tetrahedral coordination to four cysteines. The distortion of the tetrahedral configuration in Dfx and Dx, in comparison to rubredoxin, seems to be caused by the primary structure that imposes a different relative position of the ligands of the Fe site: while rubredoxin has two Cys-X-Y-Cys motifs, Dfx and Dx have one Cys-X-Y-Cys and one Cys-Cys motif that produce significant differences in terms of metal environment and water structure as shown by X-ray crystallographic studies on *D. gigas* desulforedoxin [5]. In the latter protein, all four Fe-S bonds are nearly identical, while the S^{γ}-Fe-S^{γ} angles range from 101.9 to 125.9°. The S^{γ}-Fe-S^{γ} angle involving the adjacent cysteinyl residues Cys-28 and Cys-29 is substantially widened when compared with those found in rubredoxin [5], where a coordination closer to the tetrahedral geometry occurs.

We report here the results of the complete protein sequence analysis on Dfx from *D. desulfuricans* ATCC 2774 using a combination of classical Edman degradation and mass spectrometric techniques.

2. Materials and methods

2.1. Sequence and mass analyses

N-terminal sequence analysis of the protein was performed on an ABI 477A protein sequencer. Peptides were sequenced using either an ABI 475A, 476A or 477A instrument (Applied Biosystems, Foster City, CA). A few peptides were sequenced on a Porton LF3000 protein sequencer (Beckman, Palo Alto, CA).

Protein mass spectra were taken by electrospray mass spectrometry [6] on a VG BIO-Q mass spectrometer (Fisons Instruments, VG Biotech, Altrincham, UK). Typically, 50 pmol of protein were dissolved in 10 μl 50% acetonitrile/1% formic acid in water and injected via a Rheodyne injector 7125 (Rheodyne, Cotati, CA) in the electrospray source at a flow rate of 6 μl/min delivered by a syringe pump 11 (Harvard Apparatus, South Natick, MA). Calibration was performed separately by injection of 50 pmol horse myoglobin (Sigma, St. Louis, MO). Scans of 9 s covering the 600–1500 Da range were accumulated over 2.5 min. Spectra were collected via the Masslynx software delivered with the instrument.

Peptide masses were generally measured by matrix-assisted laser desorption mass spectrometry (Maldi-TOF) [7] on a VG Tofspec

*Corresponding author. Fax: (32) (9) 264 53 38;
E-mail: Jozef.VanBeeumen@rug.ac.be

**Corresponding author. Fax: (351) (1) 294 85 50;
E-mail: isa@dq.fct.unl.pt

Abbreviations: Dfx, desulfoferrodoxin; Dx, desulforedoxin; DTPA, diethylenetriaminepentaacetic acid

PII S0014-5793(96)00364-X

Table 1. Molecular weight determination of peptides used for determination of the Dfx protein sequence determined by electrospray or Maldi-TOF MS

Peptide	Sequence position	Calculated MW (Da)	Determined MW (Da)	Error(Da)
C12	P1-M33	3476.2	3172.3	−3.9[a]
S1	P1-E5	622.7	622.6	−0.1
Sc5a	V6-E17	1365.6	1364.4	−1.2
K2	H35-K47	1391.5	1407.5	+16.0[b]
S3	K47-E54	821.0	820.7	−0.3
S9	K54-E72	2055.4	2055.9	+0.5
S12	K73-E77	1140.3	1140.0	−0.3
Sc1	W78-E80	446.5	446.6	+0.1
S8	L81-E99	2078.3	2078.1	−0.2
S10	A100-E113	2055.4	2055.9	+0.5
Sc5b	Y114-E124	1471.5	1472.7	+1.2
R2	E113-N125	1598.3	1676.5	+78.2[c]
C28A1	A106-E124	2200.5	2198.7	−1.8

[a]Molecular weight of the homoserine lactone form of the CNBr fraction. The relatively large error of 3.9 mass units may be due to disulfide bridge formation between the four cysteines present in this peptide.
[b]Mass difference of 16.0 probably due to oxidation of Met-36.
[c]Mass difference due of 78.2 Da due to disulfide linkage of Cys-115 to a β-mercaptoethanol molecule.

mass spectrometer (Fisons Instruments, VG Analytical, Wythenshawe, UK) equipped with an N_2 laser (307 nm). Typically, 1 pmol of sample was mixed on the target with 1 μl of a 50 mM solution of α-cyanohydroxycinnamic acid (Aldrich, St. Louis, MO) in 40% acetonitrile/0.1% trifluoroacetic acid (TFA) in water. Spectra from 5–20 laser shots were accumulated and processed with the Opus software delivered with the instrument. In some cases, the 'Fast Evaporation' sample application was used as described by Vorm and Mann [8].

Some peptide masses were also measured by electrospray mass spectrometry as described above.

2.2. Chemical and enzymatic digestions

Cyanogen bromide cleavage was carried out on 1 mg of Dfx. The protein was dissolved in 200 μl of 85% formic acid. The reagent (66 mg) was then added and the volume was adjusted to 500 μl with the formic acid solution. The reaction mixture was left overnight under argon and dried thereafter using a SpeedVac (Savant, Farmingdale, NY). Finally, the protein was resuspended in 200 μl of 0.1% TFA.

78 μg of native Dfx were digested with 2 μg of Glu-C endoproteinase (*Staphylococcus aureus* V8 protease, Miles, Naperville, IL) at 37°C in 50 mM ammonium bicarbonate in water (pH 7.8). A similar digestion was performed on the same amount of carboxymethylated protein.

Partial acid hydrolysis was performed on CNBr fraction 28. About 5 nmol of this peptide was dissolved in 20 μl 5% aqueous formic acid in a glass tube placed in a hydrolysis vessel to which 200 μl of 5% formic acid was added to saturate the atmosphere. The vessel was sealed under vacuum and incubated at 106°C for 2 h.

60 μg of dithionite-treated protein were digested with 2 μg Lys-C endoproteinase (Wako, Osaka, Japan) at 37°C in 50 mM Tris-HCl, pH 8. Another 60 μg of native protein were digested with 2 μg Arg-C endoproteinase (Boehringer, Mannheim, Germany) dissolved in the incubation buffer delivered with the enzyme and using the conditions indicated by the company.

2.3. Purification of the peptides

All digestion mixtures were dried in a SpeedVac and redissolved in 100 μl 0.1% TFA in water prior to injection on the HPLC column. The CNBr digestions were separated on a Gold HPLC system (Beckman, Palo Alto, CA) using a C18 reversed-phase column from Pharmacia (Uppsala, Sweden). The Glu-C and Lys-C digestions of the native protein were separated on an Alltima C18 column (Alltech, Deerfield, USA) using a 140 A solvent delivery system, and a 1000 S diode array detector (Applied Biosystems). Part of the column effluent was split and used to determine the masses of the eluted peptides on-line by electrospray mass spectrometry. The configuration was very similar to that described by Klarskov et al [9]. The major

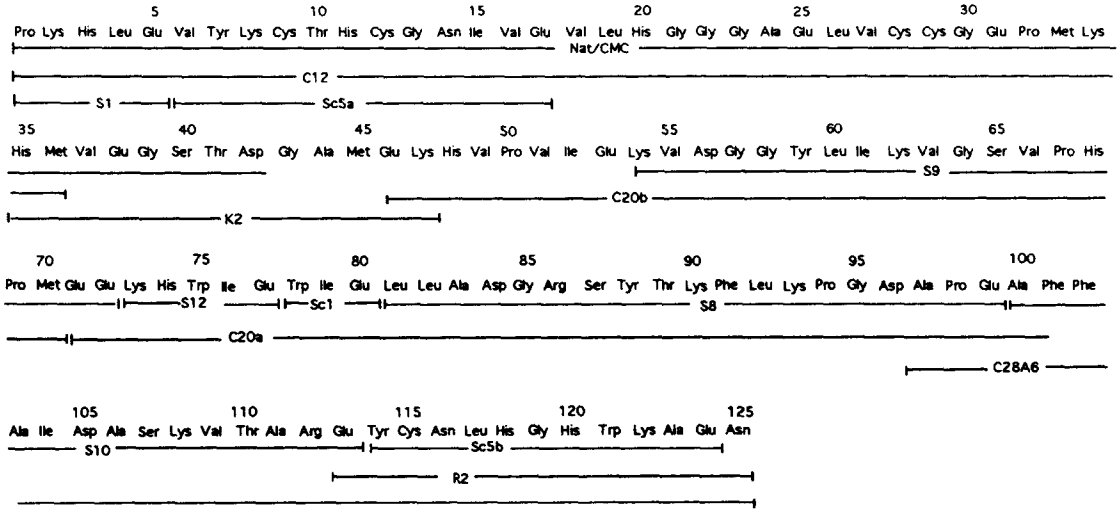

Fig. 1. Reconstruction of the sequence of Dfx. S, *Staphyloccoccus aureus* V8 endoproteinase derived peptides; Sc, idem, but derived from carboxymethylated protein; C, CNBr-derived peptides; K, endoproteinase Lys-C derived peptides; R, Arg-C endoproteinase derived peptides. C28A peptides were derived by partial acid hydrolysis from CNBR fraction 28. Nat and CMC are results of native and carboxymethylated protein sequence analysis, respectively. All cysteines where at least once determined as carboxymethylcysteine. Fraction numbering is according to the elution order of the HPLC chromatograms (not shown).

difference is in the commercial flow splitter which is replaced by a home-built device. A 1/20 fraction of the column effluent (5 μl/min) was split using a Valco tee to a liquid sheath probe. The sheath liquid was 2-methoxyethanol/isopropanol (2:1) pumped at a flow rate of 3 μl/min by a 140 A solvent delivery system (Applied Biosystems). The electrospray instrument was a VG BIO-Q triple quadrupole unit (Fisons Instruments, Altrincham, UK).

The Glu-C digestion mixture from carboxymethylated protein and the CNBr fraction C20 mixture were separated on a C2/C18 mRPC column. The Arg-C digestion and the partial acid hydrolysis subcleavage of CNBr fraction E28 were separated on a Sephasil C18 column. These separations were all performed on a Smart System (Pharmacia Biotech, Uppsala, Sweden).

3. Results and discussion

The N-terminal sequence analysis of the native protein, already described in [3] covered the first 44 amino acids of which a few were corrected during this work. It showed a high similarity with the sequence of Dfx in *D. vulgaris*, which simplified the alignment of the peptides obtained and the interpretation of the sequencing results. The result of the sequence analysis is summarized in Fig. 1. After each digestion, the masses of the obtained fractions were determined, and only those of which the mass did not fit within the sequence, known at that stage, were further analysed. A first digestion, using cyanogen bromide, yielded 3 main fractions which were all analysed by mass spectrometry and sequence analysis. One fraction labelled C12 was the N-terminal peptide of which the mass fitted well with the calculated mass of the peptide Pro-1–Met-35. The two other fractions (C12 and C28) turned out to be identical mixtures and contained two main fragments which could be aligned with Glu-46–Met-70 and Glu-71–Ala-100 of the *D. vulgaris* Dfx sequence. Digestion of the protein with Glu-C endoproteinase yielded some 15 fractions. Sequence analysis of fraction S9 demonstrated an overlap between C20b and C20a; analysis of S8 confirmed the C-terminal region of C20a which had not been unambiguously determined before. Sequence analysis of fraction S10 extended the sequence of C20a to residue 113. A peptide K12 obtained by Lys-C digestion of the protein yielded an overlap between C12 and C20b. Subdigestion of a fraction of C28 by partial acid hydrolysis yielded a C-terminal fragment Ala-97–Asn-125. The C-terminal sequence was confirmed by sequence analysis of peptide R2 obtained after Arg-C endoproteinase digestion. Finally, cysteines were confirmed as carboxymethylcysteine in sequence analyses on the carboxymethylated protein and peptides obtained after Glu-C digestion. The molecular weight of some relevant peptides is summarized in Table 1.

The mass obtained by electrospray analysis of Dfx was 13 881.3 ± 0.1 Da (Fig. 2), which corresponds well with the molecular mass calculated from the sequence (13 882.0 Da). The difference is smaller than 0.01%, the generally accepted accuracy level of the method. The sequence derived amino acid composition is also in good agreement with that experimentally determined [3].

The sequence of *D. desulfuricans* (ATCC 27774) Dfx is the first of its class to be fully characterised at the protein level. The other known sequence of a Dfx, the one from *D. vulgaris* Hildenborough, was deduced from the nucleotide sequence [1]. The gene product from this strain has also been isolated and physically characterised [10]. Originally this protein was called rubredoxin oxidoreductase [1]. There is 79% similarity between the two Dfxs (Fig. 3). This is not surprising since the two species have a 91.3% similarity based on their 16S rRNA sequences [11].

Previously, Bruschi et al. [2] sequenced desulforedoxin, a dimer (2×4 kDa) isolated from *D. gigas*. The monomer shows

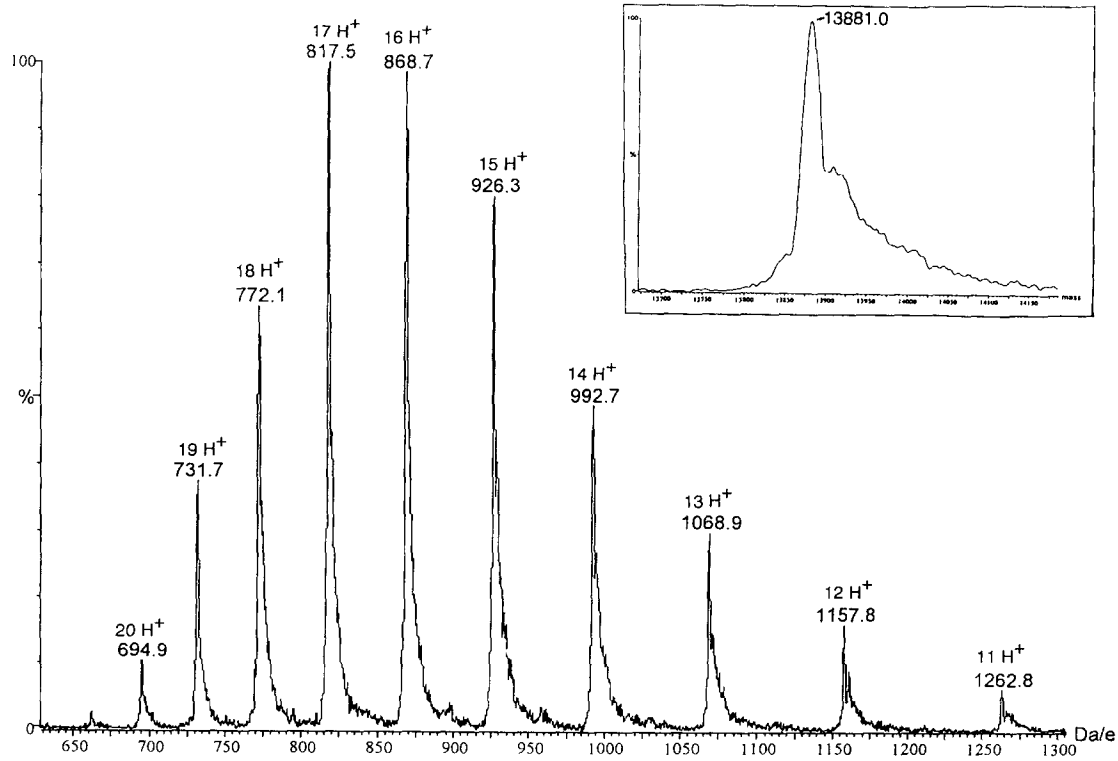

Fig. 2. Electrospray mass spectrum of Dfx. The inset shows the transformed spectrum. Peak broadening at the higher mass is due to sodium and potassium adduct formation.

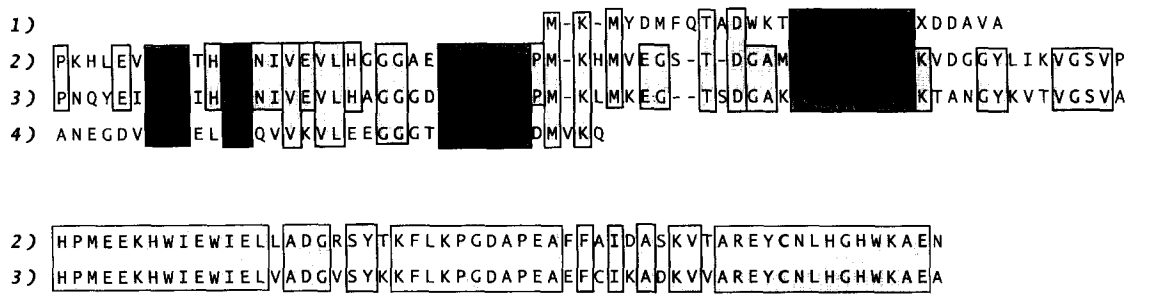

Fig. 3. Sequence alignment of Dfx and related proteins. (1) N-terminal fragment of *D. gigas* neelaredoxin [14]; (2) *D. desulfuricans* ATCC 27774 sequence of desulfoferrodoxin (this work); (3) *D. vulgaris* desulfoferrodoxin sequence [1]; (4) *D. gigas* desulforedoxin sequence [2]. The boxes show the homologous regions between the different proteins. Some of the ligands that can bind the iron are given in bold.

48.7% similarity with the amino-terminus (36 residues) of *D. desulfuricans* Dfx and 51.4% with Dfx from *D. vulgaris* (Fig. 3). Brumlik et al. [12] sequenced the gene for Dx and demonstrated a 67% nucleotide sequence similarity with the corresponding counterpart of the gene from *D. vulgaris* Dfx. These authors suggested that Dfx might have arisen from a gene fusion event of Dx with another redox protein.

Recently, a new blue non-heme iron protein named neelaredoxin was purified from *D. gigas*. It was described as a two identical iron atom containing protein with a molecular mass of 15 kDa [13]. Its visible spectroscopic data as well as its redox potentials show some resemblance with those attributed to Center II in *D. desulfuricans* Dfx [4]. Moreover, its N-terminal sequence has a fair degree of similarity with the initial part of the second domain (starting at amino acid 36) of the *D. desulfuricans* Dfx (Fig. 3). On the other hand, *D. gigas* appears to contain no Dfx, whereas *D. vulgaris* (Hildenborough) and *D. desulfuricans* are devoid of Dx and neelaredoxin. Thus, so far, the coexistence of both types of protein (Dfx and Dx) in the same organism has not been demonstrated.

On the basis of 16S rRNA similarity data *D. vulgaris* and *D. desulfuricans* are 91.3% related, while they are related to *D. gigas*, respectively, by 88 and 87%. The evolutionary tree drawn from these similarities shows that *D. gigas* is the one most related to their common ancestor [12]. Therefore, it is elegant to assume that Dx and neelaredoxin might be the ancestor proteins which in the more evolved *Desulfovibrio* strains have given rise to the new protein Dfx. In order to gain a better insight in this topic, several *Desulfovibrio* strains should be screened for the presence of Dx and neelaredoxin and/or Dfx.

The sequence data presented in this paper help to clarify some of the as yet unresolved aspects of Dfx active centers coordination. The initial proposition for the coordination of its Center I was based on Mössbauer and resonance Raman data [4,14], and indicated that this center was coordinated by four cysteines in a way similar to that found in rubredoxins. A second proposition was made, in which Center I would have a coordination similar to that found in pentagonal bipyramidal iron in ferric complexes of diethylenetriaminepentaacetic acid (DTPA) [11]. Recently, Archer et al. [5] determined the X-ray structure of *D. gigas* Dx and proved that the iron is found in a distorted tetrahedral sulfur coordination, as first proposed. Supporting evidence for this fact also came from the determination of the structure of the zinc derivative of Dx by 2D NMR techniques [15] and from EXAFS data of oxidized and reduced Dx (Scott, R.A., unpublished results). In light of these results and due to the strong homology found between Dx and the N-terminal sequence of Dfx, it is sensible to assign Cys-9, Cys-12, Cys-28 and Cys-29 as the residues responsible for the coordination of Center I.

The nature of Center II as well as the coordination of its Fe atom are still unknown. The Mössbauer and resonance Raman data for this site are indicative of an iron center with octahedral coordination and predominantly nitrogen/oxygen/sulfur-containing ligands (although the possibility of a pentagonal coordination cannot be ruled out) [4]. As shown in Fig. 3, the alignment of the two Dfxs and the N-terminal region of neelaredoxin shows a strongly conserved motif **EKHPVAIE** with four residues that could provide some of the postulated nitrogen/oxygen coordination. However, it should be stressed that, at least for the two DFxs, there are up to 22 conserved amino acids (Glu, Asp, Lys and His) that could provide the nitrogen/oxygen ligands required for the binding of the second iron. A resonance Raman study has already provided evidence for one or two cysteinyl-S ligands at Center II [4]. However, and in contrast with the sequence derived for the *D. vulgaris* protein, only one cysteinyl residue is present in the C-terminal region of *D. desulfuricans* Dfx (Cys-115). Therefore, only one cysteine is available to be involved in the coordination of Center II, ruling out the possibility of a double sulfur coordination for this center.

Further structural and spectroscopic studies are necessary to fully describe and understand the nature of Center II in desulfoferrodoxin as well as its evolutionary origin in *Desulfovibrio* species.

Acknowledgements: We thank Manuela Regalla for the N-terminal sequence analysis and Maria João Lima for electrophoresis and homogeneity studies on Dfx. We also thank Dr. R. Toci for growth of biomass at the Fermentation Plant of the CNRS, Marseille, France. This work was supported by Junta Nacional de Investigação Científica e Tecnológica (PBIC/QUI/1646/93 (to I.M.) and PBIC/BIO/1289/92 (to J.L.)), the Flemish (Concerted Research Action Grant no. 12052293) and by the Belgian National Fund for Joint Basic Research (nr 32001891) (to J.V.B.).

References

[1] Brumlik, M.J. and Voordouw, G. (1989) J. Bacteriol. 171, 4996–5004.
[2] Bruschi, M., Moura, I., LeGall, J., Xavier, A.V. and Sieker, L.C. (1979) Biochem. Biophys. Res. Commun. 90, 596–605.
[3] Moura, I, Tavares, P., Moura, J.J.G., Ravi, N., Huynh, B.H., Liu, M.-Y. and LeGall J. (1990) J. Biol. Chem. 265, 21596–21602.
[4] Tavares, P., Razvi, N., Moura, J.J.G., LeGall, J., Huang, Y.-H., Crouse, B.R., Johnson, M.K., Huynh, B.H. and Moura, I. (1994) J. Biol. Chem. 269, 10504–10510.

[5] Archer, M., Huber, R., Tavares, P., Moura., I., Moura, J.J.G., Carrondo, M.A., Sieker, L.C., LeGall and Romão, M.J. (1995), J. Mol. Biol. 215, 690–702.
[6] Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. and Whitehouse C.M. (1990) Mass Spectrom. Rev. 9, 37–70.
[7] Karas, M., Bachman, D., Bahr, U. and Hillenkamp, F. (1987) Int. J. Mass Spectrom. Ion Processes 78, 53–68.
[8] Vorm, O. and Mann, M. (1994) J. Am. Soc. Mass Spectrom. 5, 955–958.
[9] Klarskov, K., Roecklin, D., Bouchon, B., Sabatié, J., Van Dorsselaer, A. and Bisschoff R. (1994) Anal. Biochem. 216, 127–134.
[10] Verhagen, M.F., Voorhorst, W.G., Kolkman, J.A., Wolbert, R.B. and Hagen, W.R. (1993) FEBS Lett. 336, 13–18.
[11] Devereux, R., Delaney, M., Widdel, F. and Stahl, D.A. (1989) J. Bacteriol. 171, 6689–6695.
[12] Brumlik, M.J., Leroy, G., Bruschi, M. and Voordouw, G. (1990) J. Bacteriol. 172, 7289–7292
[13] Chen, L., Sharma, P., LeGall J., Mariano, A.M., Teixeira, M. and Xavier, A.V. (1994) Eur. J. Biochem. 226, 613–618.
[14] Moura, I., Huynh, B.H., Hausinger, R.P., LeGall, J., Xavier, A.V. and Munck, E. (1980) J. Biol. Chem. 255, 2493–2498.
[15] Goodfellow, B.J., Tavares, P., Romão, M.J., Czaja, C., Rusnak, F., LeGall, J., Moura, I. and Moura, J.J.G. (1995) J. Biol. Inorg. Chem., submitted for publication.