Solution Structure of a Monoheme Ferrocytochrome *c* from *Shewanella putrefaciens* and Structural Analysis of Sequence-Similar Proteins: Functional Implications†

Ilaria Bartalesi,‡,§ Ivano Bertini,*,‡,§ Parvana Hajieva,‡ Antonio Rosato,‡,§ and Paul R. Vasos‡

Centro di Risonanze Magnetiche, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy, and Department of Chemistry, University of Florence, Via della Lastruccia 5, 50019 Sesto Fiorentino, Italy

Received November 27, 2001; Revised Manuscript Received February 18, 2002

ABSTRACT: Within the frame of the characterization of the structure and function of cytochromes *c*, an 81-amino acid cytochrome *c* was identified in the genome of *Shewanella putrefaciens*. Because of the scarce information about bacterial cytochromes of this type and the large variability in sequences and possibly function, we decided to proceed to its structural characterization. This protein was expressed in *Escherichia coli* and purified. The oxidized species is largely high spin, with a detached methionine, whereas the reduced species has the classical His/Met axial ligation to iron. The NMR solution structure of the reduced form was determined on a ^{15}N-labeled sample, for which 99% of all non-proline backbone ^{1}H and ^{15}N resonances have been assigned. One thousand three hundred two meaningful NOEs, out of 1775 NOEs, together with 66 dihedral angles provide a structure with rmsd values from the mean of 0.50 and 0.96 Å for backbone and all heavy atoms, respectively. A search of gene banks allowed us to locate 10 different cytochromes *c*, the sequences of which are more than 30% identical to that of the *S. putrefaciens* cytochrome. For two of them, the structures are known. The structures of the others have been modeled by using the available templates and internally validated. Structural similarities in terms of surface properties account for their biophysical features and provide hints about the function.

c-type cytochromes are widespread metalloproteins, involved in a number of different metabolic pathways throughout all kingdoms of life (*1*−*3*). In particular, class I (*4*) cytochromes, the typical monoheme soluble cytochromes of mitochondria and bacteria, have been extensively investigated, and structural information is available for all the various subgroups of this class. The largest amount of physiological and chemicophysical information is available for eukaryotic mitochondrial cytochromes [which belong to subclass IB (*4*)], whose physiological role is that of shuttling electrons to cytochrome *c* oxidase, as the final step of aerobic respiration. For these systems, extensive studies on redox-dependent structural (*5*, *6*) and dynamic (*7*−*9*) properties, on folding−unfolding mechanisms (*10*−*12*), and on the factors tuning redox potential (*13*−*16*) are available (*9*, *17*, *18*). Comparatively less is known for bacterial cytochromes belonging to class I, albeit with significant differences among the different subgroups. It is noteworthy that bacterial *c*-type cytochromes display a large variability in sequence, and thus in physicochemical properties. This variability is in sharp contrast with what has been observed for mitochondrial *c*-type cytochromes, which constitute one of the protein families most conserved through evolution (*19*, *20*). Consequently, to obtain a degree of understanding of the structure−function relationships for bacterial cytochromes analogous to that of mitochondrial *c*-type cytochromes, a significantly wider array of experimental data on several different systems is needed.

This work focuses on an uncharacterized cytochrome *c* from the metal-reducing bacterium *Shewanella putrefaciens* (also known as *Pseudomonas putrefaciens*). *S. putrefaciens* is a Gram-negative facultative anaerobic bacterium, which can use a variety of compounds as terminal electron acceptors for respiration, including insoluble metal oxides and oxidized anions (*21*−*26*). The high level of versatility of *S. putrefaciens* in respiration is linked also to its capability of producing a number of different mono- and multiheme cytochromes *c*, the relative levels of expression of which depend on the terminal electron acceptor(s) available to the bacterium (*27*). The cytochrome studied here, encoded by the ScyA gene of *S. putrefaciens* (*Sp*cytc[1] hereafter), is a monoheme small (∼9 kDa) soluble protein. It most likely localizes in the periplasmic space of *S. putrefaciens*, where bacterial respiration occurs, as suggested by the presence of a signal peptide at the N-terminus of the protein sequence. *Sp*cyt*c* is similar in sequence to other two cytochromes of known structure: cytochrome c_5 from *Azotobacter vinelandii*

† Financial support from MURST COFIN99, Italian CNR (Contracts 98.01789.CT03 and 01.0359.PF49), and the European Commission (Contract HPRI-CT-1999−0009) is gratefully acknowledged.

* To whom correspondence should be addressed: Centro di Risonanze Magnetiche, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy. Telephone: +39 055 4574272. Fax: +39 055 4574271. E-mail: bertini@cerm.unifi.it.

‡ Centro di Risonanze Magnetiche.

§ Department of Chemistry.

[1] Abbreviations: *Sp*cyt*c*, *S. putrefaciens* cytochrome *c*; PCR, polymerase chain reaction; IPTG, isopropyl *β*-D-thiogalactopyranoside; ESI-MS, electrospray ionization mass spectroscopy; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlation spectroscopy; HSQC, heteronuclear single-quantum correlation spectroscopy; REM, restrained energy minimization; rmsd, root-mean-square deviation.

10.1021/bi015984z CCC: $22.00
Published on Web 03/29/2002

(*28*) (40% level of sequence identity) and cytochrome c_{551} from *Ectothiorhodospira halophila* (*29*) (35% level of sequence identity). The structure of the reduced form of the latter protein was determined by NMR methods (*29*), while that of cytochrome c_5 from *A. vinelandii* was determined by X-ray methods at 2.5 Å in the oxidized state (*28*). *A. vinelandii* cytochrome c_5 is the archetypal representative of the cytochrome subclass to which *Sp*cyt*c* belongs [IE (*4*)], and is also the protein of this class for which most is known about its physiological role (*30*). It is worth noting that *S. putrefaciens* and *A. vinelandii* are characterized by widely different metabolic processes. *E. halophila* cytochrome c_{551} displays significant sequence variations from typical cytochromes c_5 (e.g., it lacks the conserved disulfide bridge which is characteristic of these cytochromes), while it bears similarities with cytochromes c_6, which lead to the proposition that *E. halophila* cytochrome c_{551} could be regarded as a link between these two groups of proteins (*31*).

It is thus apparent that further characterization of cytochromes c_5 would be useful in obtaining deeper insights into the structural and functional features of this class of proteins, which displays a number of peculiar sequence variations with respect to *c*-type cytochromes from eukaryotes and to other bacterial cytochromes. The work presented here reports the solution structure of *Sp*cyt*c* in its reduced form, and a partial investigation of the oxidized species. These data, together with information derived by browsing publicly accessible DNA and protein databases and by performing homology modeling studies on other related cytochromes, provide intriguing hints about function.

MATERIALS AND METHODS

Sample Preparation. The ScyA gene of *S. putrefaciens* (kindly provided by D. A. Saffarini, Department of Microbiology, University of Massachusetts, Amherst, MA), which encodes a monoheme *c*-type cytochrome, was amplified by standard PCR methods. Only the portion of the gene corresponding to the mature protein (residues 22−99 of the amino acid sequence deposited in TrEMBL, entry O52685) was amplified. The gene was digested with *Bgl*II and *Xba*I and cloned into the pPB10 plasmid (*32*), obtaining the pPB10ScyA plasmid. This fuses the cytochrome *c* gene to the sequence encoding the signal peptide of *Escherichia coli* cytochrome b_{562} to ensure correct export into the periplasm and processing. After processing, three amino acids (Ala, Asp, and Leu, numbered from −3 to −1 in this work) of the signal peptide are left at the N-terminus of the protein, thus yielding an 81-residue polypeptide. The NMR analysis shows that these three residues are unstructured and protrude into the solvent, thus not interacting with the remainder of the protein. The gene sequence was confirmed by standard DNA sequencing methods. The pPB10ScyA plasmid was cotransformed in *E. coli* strain BL21(DE3)C41 with the pEC86 plasmid, which expresses the ccmABCDEFGH genes, which are needed for heme incorporation in cytochrome *c* (*33*). Cells harboring both plasmids were grown on 2×YT plates, selected with 100 μM/mL ampicillin and 40 μM/mL chloramphenicol. For protein preparation, cells were grown in 2 L flasks, containing 1.5 L of 2×YT with antibiotics at 37 °C, at 100 rpm and induced with 0.8 mM IPTG. The maximum level of protein expression was obtained 18−20 h after induction. Cells were harvested by centrifugation and lysed by sonication. After centrifugation, $(NH_4)_2SO_4$ was added to the supernatant up to a concentration of 326 g/L. The supernatant was dialyzed overnight in 50 mM Tris (pH 8), then applied to a CM52-Sepharose column equilibrated in the same buffer, and eluted with a 0 to 250 mM NaCl gradient. Red fractions were pooled, concentrated by ultrafiltration with stirred cells (Amicon), applied to a G75 Sephadex column, and eluted with 50 mM Tris (pH 8). Fractions with an *R* value of ≈5 ($R = A_{416}/A_{280}$) were pooled together. The mass of the protein was checked by electrospray ionization mass spectrometry (ESI-MS) using an API 365 ESI-MS instrument (Applied Biosystem) equipped with a Perkin-Elmer HPLC system and a series 200 autosampler. The experimental mass was 9205.3 Da against an expected value of 9206.7 Da.

The ^{15}N-labeled protein was obtained from cultures in M9 minimal medium containing 1.24 g/L of $(^{15}NH_4)_2SO_4$, supplemented with a solution of trace elements, a vitamin mix, 20 μL/mL chloramphenicol, 50 μL/mL ampicillin, δ-aminolevunilic acid (0.1 mM), and 1-mercaptoethanesulfonic acid (1 mM). The carbon source was glucose (2 g/L). Growth conditions, protein isolation, and purification were essentially as described above. Ten milligrams per liter of purified protein was obtained from rich medium and 2 mg/L from minimal medium.

Protein absorption spectra were acquired on a Varian Cary 3 UV−visible scanning spectrophotometer at room temperature using fused quartz cuvettes with a path length of 1 cm (Merck). Measurements were performed at room temperature in 100 mM phosphate buffer (pH 7.0).

Reduced *Sp*cyt*c* samples were prepared by adding 2 equiv of sodium dithionite, under anaerobic conditions, to a protein solution in 100 mM phosphate buffer (pH 7). Under these conditions, disulfide bridges are not reduced. Oxidized *Sp*cyt*c* samples were prepared in the same buffer, either by washing the protein with a slight excess of $K_3Fe(CN)_6$ or by bubbling O_2 through the solution. NMR samples had protein concentrations of 1−1.5 mM.

NMR Spectroscopy and Solution Structure Calculations. NMR spectra were acquired on reduced *Sp*cyt*c* on Bruker Avance 800, 700, and 600 spectrometers operating at proton nominal frequencies of 800.13, 700.13, and 600.13 MHz, respectively. The temperature for all experiments was set to 298 K, unless otherwise stated.

Three-dimensional (3D) ^{15}N-HSQC-TOCSY (*34*) and two-dimensional (2D) TOCSY (*35*) spectra were recorded at 600 MHz, with a spin lock time of 80 ms. 3D ^{15}N-HSQC-NOESY (*36*) and 2D NOESY (*37*) spectra were acquired at 800 MHz at two temperatures (298 and 288 K), with mixing times between 90 and 110 ms. An HNHA experiment (*38*) was performed at 700 MHz to determine $^3J_{HNH\alpha}$ coupling constants. These were used to obtain constraints for the ϕ torsion angles. An HNHB (*39*) spectrum was acquired at 800 MHz to obtain stereospecific assignments for Hβ protons and constraints for the χ_1 torsion angle.

The assignment of the ^{1}H and ^{15}N NMR signals of *Sp*cyt*c* was achieved using standard procedures (*40*) through the analysis of 2D and 3D homo- and heteronuclear spectra. The 2D and 3D NOESY spectra acquired at 298 K were used to obtain upper distance limits for structure calculations. Intensities of dipolar connectivities were converted into upper

distance limits through the program CALIBA, using the standard procedure (*41*). Stereospecific assignments of diastereotopic pairs were obtained using the program GLOMSA (*41*), as well as from relative cross-peak intensities in the HNHB spectrum (*39*). The latter were also used to extract constraints for the χ_1 torsion angle. $^3J_{HNH\alpha}$ coupling constants were correlated to the backbone torsion angle ϕ by means of the appropriate Karplus curve (*38*). The ratios between intraresidue and sequential HN−Hα NOESY cross-peak intensities were used to extract constraints for the ϕ and ψ torsion angles (*42*).

All the above constraints were used as input for the program DYANA (*43*). The heme special residue was introduced in calculations as previously described (*17*, *44*). Two hundred fifty random structures were generated and annealed using the standard DYANA protocol in 15 000 steps. The 30 structures with the lowest target function values were selected as the final DYANA family and subjected to a restrained energy minimization (REM) refinement with the program AMBER (*45*). The energy-minimized structures constituted the final REM family. The average structure of this family was built by superimposing residues 1−77 and again subjected to REM.

Structure analyses were performed with the programs PROCHECK (*46*), PROCHECK-NMR (*47*), and AQUA (*47*). Structure visualization was done with the program MOLMOL (*48*).

Sequence Analysis and Structural Modeling. We searched for sequences similar to that of *Sp*cyt*c* in the SwissProt and TrEMBL (for protein sequences translated from genomic codes) databases using sequence similarity criteria using the program BLAST (*49*). The Phi-Blast version of the program (Pattern-Hit Initiated Blast) (*49*) was also used, to identify sequences more distantly related, but still showing the CXXCH heme-binding pattern.

The identified sequences were selected and organized by performing alignments through the ClustalX program (*50*), installed on local machines. The service at http://bioweb.pasteur.fr (*51*) was used to draw phylogenetic trees. Modeling was performed with the model-default option of the program MODELLER version 4.0 (*52*). Models were analyzed with the program PROSA (*53*). Sequences which were more than 30% identical with that of *Sp*cyt*c* and less than 95% identical among each other have been retained. Several sequences had a long N-terminal extension with respect to the sequence of *Sp*cyt*c*, corresponding to the signal peptide for periplasmic export (which is normally cleaved in the mature protein). As mentioned, this is also present in the ScyA gene, but was removed at the step of gene cloning. The boundaries of the signal peptide in the ScyA gene were identified through available sequence analysis software tools (see www.expasy.ch). The N-terminal extension, where present, was not considered for either the alignment or the modeling (see also the legend of Figure 5).

RESULTS AND DISCUSSION

Spectroscopic Properties. Under the experimental conditions described in Materials and Methods, the oxidized *Sp*cyt*c* features a Soret band at 411 nm, α and β bands at 525 and 552 nm, respectively, and a charge-transfer band (CT1) band at 620 nm, indicating the presence of both a high-spin and a low-spin Fe^{3+} species. In contrast, the UV−vis spectrum of reduced *Sp*cyt*c* is consistent with the presence of a single low-spin Fe^{2+} species. Figure 1 reports a comparison of the ^{1}H NMR spectra of oxidized and reduced *Sp*cyt*c*. It clearly appears that the pattern of hyperfine-shifted signals observed for oxidized *Sp*cyt*c* signals is that typical of a high-spin species, while reduced *Sp*cyt*c* displays a spectrum typical of a low-spin (thus diamagnetic) species. This behavior is extremely uncommon for *c*-type cytochromes, which usually are in a low-spin state in all oxidation states. The signal dispersion in the ^{15}N-HSQC spectrum and the NOESY patterns of the polypeptide chain clearly indicate that the oxidized protein is folded. The 2D NOESY pattern of the hyperfine-shifted signals of oxidized *Sp*cyt*c* (not shown) is similar to that of oxidized cytochromes *c*′, in which the iron ion is pentacoordinated, with only one His axial ligand (*54*, *55*). However, the chemical shifts of the hyperfine-shifted proton signals are somewhat smaller than those of cytochromes *c*′, which feature shifts as large as >80 ppm (see Figure 1 and refs *54* and *55*). This can indicate that, in the present case, equilibria fast on the NMR time scale between high- and low-spin species are operative. Similar observations had been reported for a monoheme cytochrome *c* from *Wolinella succinogenes* (*56*), but no further experimental data are available. The temperature dependence of 1D ^{1}H NMR spectra of oxidized *Sp*cyt*c* displays a significant broadening of the hyperfine-shifted resonances with an increase in temperature, indicating the presence of equilibria between multiple species. The characterization of the oxidized form of *Sp*cyt*c* is in progress in our lab as part of an effort to ascertain the reasons for its peculiar spectral properties. The observations presented here provide another new piece of information about the variability of the strength of the bond between the iron ion and the axial Met in *c*-type cytochromes.

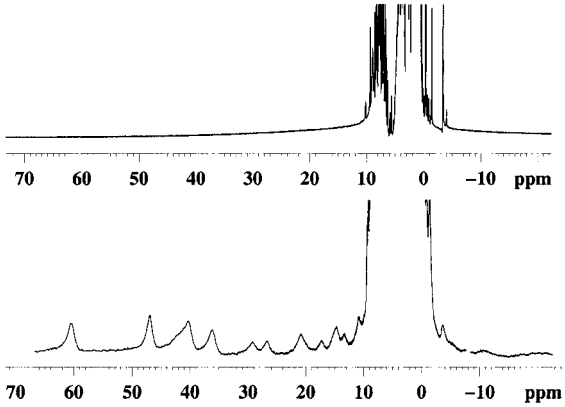

FIGURE 1: Comparison of the ^{1}H NMR spectra (acquired at 16.5 T and 285 K) of reduced (top) and oxidized (bottom) *S. putrefaciens* cytochrome *c*.

Solution Structure. A high percentage of the total number of expected resonances (97 and 94% for ^{1}H and ^{15}N nuclei, respectively) was assigned. No assignment was obtained for residue 1. The full ^{1}H and ^{15}N assignment achieved in this study, together with the stereospecific assignment, is reported in Table S1 of the Supporting Information.

Table 1 reports the details of the experimental constraints used for structure calculation, while Figure 2 displays sequential and medium-range NOE connectivities involving the HN, Hα, and Hβ protons. The pattern of sequential $d_{NH-NH(i,i+1)}$, of $d_{H\alpha-NH(i,i+3)}$, and of $d_{H\alpha-H\beta(i,i+3)}$ connectivities

Table 1: Summary of NMR Constraints Used for Structure Calculation, Restraint Violations, Structural Statistics, and Energetics for the Restrained Energy-Minimized Solution Structure of Reduced *S. putrefaciens* Cytochrome *c*

		REM		⟨REM⟩	
structural constraint	no.	average no. of violations per conformer	rms violation per restraint	no. of violations	rms violation per restraint
meaningful NOESY (total NOESY)	1302 (1775)	35.2 ± 3.7	0.017 ± 0.001 Å	40	0.020 Å
intraresidue	205	2.0 ± 1.5	0.010 ± 0.006 Å	2	0.020 Å
sequential	409	9.1 ± 2.6	0.014 ± 0.003 Å	10	0.015 Å
medium-range	454	13.7 ± 2.6	0.018 ± 0.002 Å	14	0.020 Å
long-range	234	10.4 ± 2.6	0.021 ± 0.003 Å	14	0.025 Å
Φ	32	1.7 ± 0.9	0.9 ± 0.22°	3	1.08°
Ψ	20	0.9 ± 0.8	0.5 ± 0.38°	1	0.64°
χ_1	14	0.6 ± 0.7	0.6 ± 0.63°	0	0°
violations between 0.1 and 0.3 Å		11.7 ± 2.8		14	
violations larger than 0.3 Å		0		0	
largest distance violation (Å)			0.23		0.27
largest Φ violation (deg)			6.3		3.9
largest Ψ violation (deg)			3.6		2.9
largest χ_1 violation (deg)			8.9		0

energetics	average value	value
total target function	0.45 ± 0.04	0.57
Amber average energy (kJ/mol)	−606 ± 204	−518

structure analysis		average value	value
completeness of ^{15}N backbone assignment	98.8%		
completeness of ^{1}H backbone assignment	98.8%		
completeness of backbone NMR-observable proton contacts within 4 Å		87.4%	85.8%
structural constraints per residue	16.9		
residues in most favored regions[a]		69.4%	72.5%
residues in allowed regions[a]		26.2%	27.5%
residues in generously allowed regions[a]		3.6%	0.0%
residues in disallowed regions[a]		0.8%	0.0%
overall *G* factor[b]		−0.37	−0.40

[a] As defined by the Ramachandran plot. [b] As defined in the program PROCHECK (*46*).

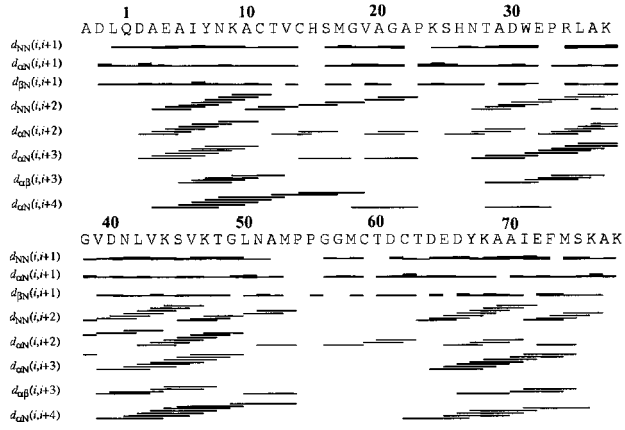

FIGURE 2: Schematic representation of the sequential and medium-range NOE connectivities involving NH, Hα, and Hβ protons of reduced *S. putrefaciens* cytochrome *c*. The thickness of the bar indicates the NOE intensity.

observed are characteristic for a protein with a quite high content of helical secondary structure, involving residues 3−10, 31−36, 41−48, and 65−74. The presence of helical structure in these regions is also consistent with the small values of the $^{3}J_{H\alpha-NH}$ constants observed, and the fact that their intraresidue:interresidue Hα−NH cross-peak ratios are larger than unity. In total, 1302 meaningful upper distance limits (of 1775) could be obtained from NOESY spectra (Table 1). Figure 3A reports the distribution of these constraints on a per-residue basis.

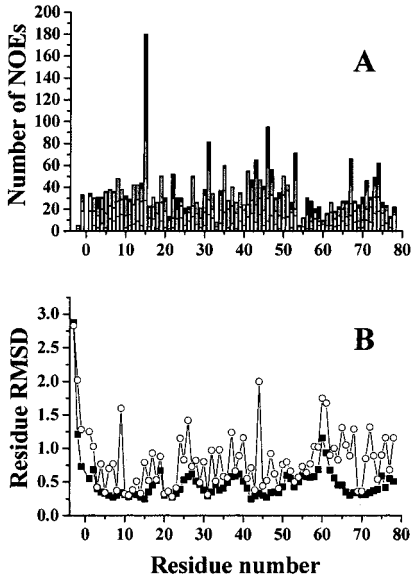

FIGURE 3: (A) Number of NOEs per residue for reduced *S. putrefaciens* cytochrome *c*. White, light gray, dark gray, and black bars indicate intraresidue, sequential, medium-range, and long-range connectivities, respectively. (B) Per-residue plot of backbone (■) and heavy atom (○) rmsds with respect to the mean structure for the final family of conformers.

The average target function of the 30 conformers constituting the DYANA family is 0.63 ± 0.08 Å^2. Their global root-mean-square deviations (rmsd) from the mean structure

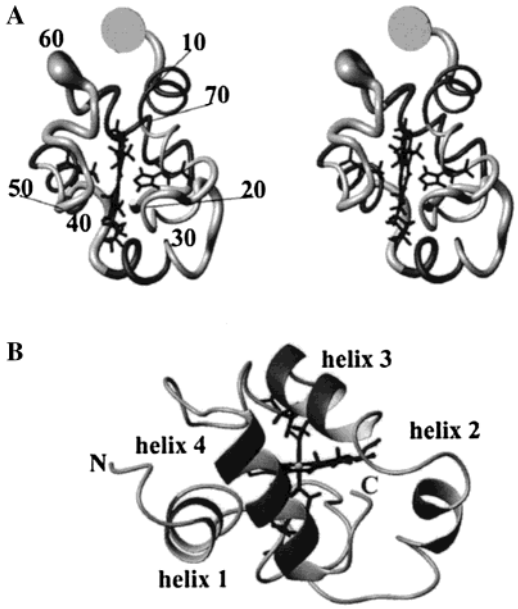

FIGURE 4: (A) Stereoview of the backbone of the final family of conformers of reduced *S. putrefaciens* cytochrome *c* represented as a tube with a variable radius, proportional to the backbone rmsd of each residue. Helices are shown in dark gray. (B) Ribbon display of the energy-minimized average structure of reduced *Sp*cyt*c*.

are 0.51 ± 0.08 and 0.96 ± 0.09 Å for the backbone and heavy atoms, respectively (for residues 1−78). These values drop to 0.46 ± 0.08 and 0.91 ± 0.09 Å, respectively, if the less defined loop region involving residues 58−64 is not included in the superposition. The same parameters over residues 1−78 for the REM family are 0.50 ± 0.08 and 0.96 ± 0.09 Å, respectively. The rmsd values for the REM family are plotted in Figure 3B. An overview of the quality parameters calculated for the REM family as well as for the restrained energy-minimized mean structure is given in Table 1. They indicate that the obtained structure has a very good quality. Coordinates for the final family of conformers and for the energy-minimized average structure have been deposited in the PDB (PDB entries 1KX7 and 1KX2).

A "sausage" view of the backbone is shown in Figure 4, which also depicts the secondary structure elements resulting from calculations. The protein displays an all-helical fold, as it is commonly found for *c*-type cytochromes, with two long helices at the N- and C-termini (α_1 and α_4, spanning residues 3−9 and 64−74, respectively), and two additional helices: a short helix involving residues 32−35 (α_2) and a longer helix of residues 42−48 (α_3). The two terminal helices are packed one against the other, and also have contacts with the heme substituents. The first ligand to the iron ion is His15, and the second ligand is Met53; both are located in loops between helices. The heme moiety is quite solvent accessible, with two sides of the porphyrin ring being exposed: the edge where the propionates lie and the edge where methyl 5 and vinyl 4 are. In addition, methyl 3 (on the side of the porphyrin ring opposite from that of the propionates) is also exposed. Helix α_2 in reduced *Sp*cyt*c* is shorter than the corresponding helix in the structure of *A. vinelandii* cytochrome c_5 [PDB entry 1CC5 (*28*)], which spans residues 28−37. The reason helix α_2 is shorter in *Sp*cyt*c* is most likely the presence of a Pro in position 33 in the present protein. From the point of view of statistics, prolines show a strong preference for the N-terminal position in helices (i.e., a proline in a helix most likely occurs as the first amino acid of the helix) (*57*, *58*). However, the probability of finding a Pro as the second (as in the present case) or third residue in a helix is still significant, while the probability of having a Pro past the third position in a helix or even as the first residue immediately after the helix is essentially zero (*57*, *58*). This is due to the fact that a proline at position *i* in the protein sequence cannot form a hydrogen bond with the carbonyl oxygen of residue $i - 4$, thus destabilizing the helix. The presence of Pro33 thus prevents residues 28−31 from adopting a helical conformation.

Sequence Alignment Analysis and Structural Modeling. The analysis of the results of the BLAST (*49*) search in protein sequence databases and of the corresponding phylogenetic tree shows that *Sp*cyt*c* displays a level of sequence similarity >30% when it is compared to only 10 cytochromes from other proteobacteria (purple bacteria) or from green-sulfur bacteria. The most similar cytochromes belong to the γ-proteobacteria group (which includes also *S. putrefaciens*). Among the proteins in this group, a 2.5 Å X-ray structure is available for *A. vinelandii* cytochrome c_5 [PDB entry 1CC5 (*28*), level of sequence identity to *Sp*cyt*c* of 40%], whereas an NMR structure (with a backbone rmsd from the mean of 0.48 Å) is available for reduced *E. halophila* cytochrome c_{551} [PDB entry 1GKS (*29*), level of sequence identity to *Sp*cyt*c* of 35%]. A number of other cytochromes display level of sequence identity values to *Sp*cyt*c* of ∼20%, including proteins from eukaryotes, algae, and other bacteria. By drawing a phylogenetic tree including all proteins with a level of sequence identity to *Sp*cyt*c* as low as 20%, we observe that sequences that are more than 30% identical to that of *Sp*cyt*c* essentially all cluster on a single branch of the tree (not shown). This implies that the subclass of bacterial cytochromes *c* containing *Sp*cyt*c* is well-defined and that it does not bear close sequence relationships with other cytochrome *c* subclasses. The alignment of the sequences that the above considerations indicate as being homologous to *Sp*cyt*c* is shown in Figure 5. As expected, the heme attachment signature CXX′CH is absolutely conserved in all the family, as well as the second iron ligand, Met53. Other absolutely conserved residues (*Sp*cyt*c* numbering) are Ala22, Gly49, and Pro54. Pro23 and Trp31 are conserved in all sequences but the one from *Cyanophora paradoxa*. It is noteworthy that the two residues between the heme-binding cysteines display a considerable variability, in terms of both the chemicophysical properties and the volume of the side chain. Indeed, Thr, Asn, Ser, His, Gly, Lys, and Ala can all be observed at position 12, while position 13 can be Val, Ala, Ser, Ile, Leu, Thr, or Met. As a general rule, it appears that polar amino acids are favored at position 12, while position 13 favors hydrophobic amino acids (Figure 5). There is no clear-cut restriction on side chain volume. In mitochondrial cytochromes *c*, less variability is observed, especially for position 13, where, in sharp contrast with the present case, there is a clear preference for polar amino acids (*20*). An intriguing feature highlighted by the sequence alignment is the presence of two consecutive prolines after the axial Met, which is observed in all proteins of the family except the two with the lowest level of sequence identity to *Sp*cyt*c* (last two in Figure 5), and the cytochrome from *E. halophila*. These three sequences also have in common a

```
                    1        10        20        30        40        50        60        70
Spcytc      -ADLQDAEAIYNKACTVCHSMG--VAGAPKSHNTADWEPRLAK--GVDNLVKSVKTGLN----AMPPGGMCTDCTDEDYKAAIEFMSKAK--
1CC5        GGGARSGDDVVAKYCNACHGTG--LLNAPKVGDSAAWKTRADAKGGLDGLLAQSLSGLN----AMPPKGTCADCSDDELKAAIGKMSGL---
1GKS        -DGESIYINGTAPTCSSCHDRG--VAGAPELNAPEDWADRPSS---VDELVESTLAGKG----AMPAYDGRADRED--LVKAIEYMLSTL--
cytc2       -VSAQEGKAVYDKACHICHSMG--VAGAPKAHDAAAWEPRIAQ--GLDTLVSTVKTGKG----AMPPGGMFYRLFR----------------
cytc3       -QAAQDPEAVFNRTCGACHDGQ--LPMAPKKGDHAAWEPRLAK--GMPTLVQHVTNGFN----AMPPRGLCTDCSAEDYQATIQWMVKK---
cytc4       -TGPRDGATVYGTFCTACHSAG--ISGAPKTGNAADWGPRIAQ--GKDVLKNHALNGFN----AMPPKGTCMDCSDDEIVAAIEHMIAGL--
cytc5       -PAEPALAQVYSASCKLCHADP--ASGAPLTGDAAAWAPRLAQ--GIDTLLDHSINGYK----GMPPMGMCMQCSEEQFRALIDFMSAAQDR
cytc6       -GGARSGEDIVGKTCNTCHGTG--LLGAPKVGDKAEWGKRAKEQGGLDGLLAKAISGIN----AMPPKGTCADCSDDELKAAIKHMSGL---
cytc7       -GGARSADDIIAKHCNACHGAG--VLGAPKIGDTAAWKERADHQGGLDGILAKAISGIN----AMPPKGTCADCSDDELREAIQKMSGL---
cytc8       -YDAAAGKATYDASCAMCHKTG--MMGAPKVGDKAAWAPHIAK--GMNVMVANSIKGYKGTKGMMPAKGGNPKLTDAQVGNAVAYMVGQSK-
cytc9       -VFAADGAAIFTNNCAACHAGGNNVIAAEKTLKKAALEQYLDGGYNVDAIKKQVTGGKN----AMPAFGGRLAEDEIAAVAEYVYSQAGNGW
                     *  **        *                                 *        **
```

FIGURE 5: Alignment of sequences similar to that of *S. putrefaciens* cytochrome *c*. The codes to the sequences are as follows: *Sp*cytc, *S. putrefaciens*; 1CC5, *A. vinelandii* (labeled with the PDB entry number); 1GKS, *E. halophila* cytochrome (labeled with the PDB entry number); cytc2, *S. violacea* (gi 6683508) (residues 19−85); cytc3, *P. aeruginosa* (gi 11351223) (residues 18−87); cytc4, *V. cholerae* (gi 11354851) (residues 55−135); cytc5, *P. aeruginosa* (gi 11348388) (residues 47−129); cytc6, *P. aeruginosa* (gi 11348388) (residues 55−136); cytc7, *P. mendocina* (gi 117915) (residues 6−87); cytc8, *C. limicola* (gi 115250); and cytc9, *C. paradoxa* (gi 9369269) (residues 58−144). Asterisks denote absolutely conserved residues. The numbering is that of *S. putrefaciens* cytochrome *c*.

lack of the disulfide bond between residues 59 and 62. Finally, in *C. paradoxa* cytochrome, which is most distant from the other sequences, the highly conserved Met74 is substituted with serine.

Structural models have been built for the sequences that are more than 30% homologous to that of *Sp*cytc based on the present energy-minimized average solution structure, and on the X-ray structure of *A. vinelandii* cytochrome c_5. The backbone rmsd values for this group of structures are on the order of 1 Å, which is in keeping with previous reports for the family of mitochondrial *c*-type cytochromes (*20*). The PROSA analysis indicates that the energetics of the models that have been built is comparable to that of experimental structures, thereby giving us confidence in the theoretical structures that were generated. A somewhat less favorable PROSA score is obtained for *C. paradoxa* cytochrome, consistent with the fact that it is the protein in the family most distant from *Sp*cytc. It is interesting to analyze the network of interresidue interactions in which the conserved residues mentioned above are involved. Ala22 and Pro23 are in close contact with the imidazole ring of the axial His15, which is actually H-bonded to the carbonyl oxygen of Pro23. This is a common feature of *c*-type cytochromes (*20*). Trp31 and Met74 are also in contact with the imidazole ring of the axial His15. Residues 22 and 23 contact the ring from the edge containing Nδ1 and are approximately above propionate 6, while residues 31 and 74 are in front of one of the two faces of the ring and lie above heme methyls 1 and 8. The side chain of Met74 is in contact with the ring of Trp31. On the basis of the observations described above, one can postulate a concerted role for these four residues, where Ala22 and Pro23 are involved in maintaining the proper orientation of the imidazole ring of the axial His15, thereby permitting its interaction with the conserved pair (Trp31 and Met74). This interaction can (i) provide a part of the ET pathway or (ii) play a role in modulating the electronic properties of the iron ion. It is to be noted that all these residues are consistently buried in the present set of related structures.

The hydrophobic core of the proteins comprises residues 6, 11, 14, 15, 22, 31, 42, 46, 53, 57, 67, 70, and 71 (with solvent accessibility of less than 10% in more than 90% of the modeled structures). Residues 7, 10, 23, 43, 49, 59, 62, 74, and 77 have a slightly higher (but still below 20%) solvent accessibility. The heme moiety is quite solvent exposed in the present family of related structures, two edges of the porphyrin ring being completely accessible to water molecules. As a consequence of the high solvent accessibility

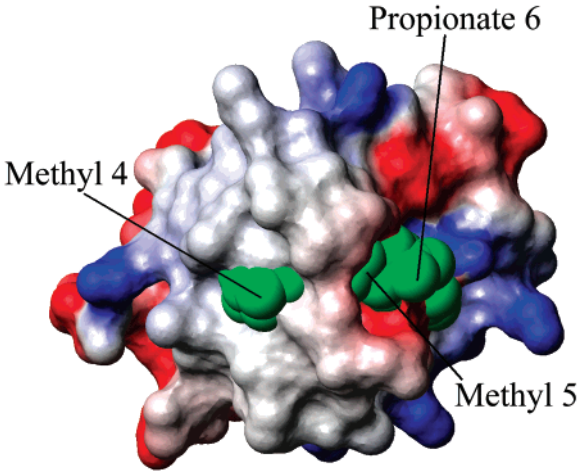

FIGURE 6: Surface electrostatic potential of *S. putrefaciens* cytochrome *c*. The orientation of the protein is such that the heme plane is normal to the page and the Fe−Nϵ2 bond is oriented toward the top of the page. The exposed side of the heme where thiomethyl 4 and methyl 5 lie faces the observer. The heme atoms are shown as spheres with a radius of 2 Å (in green). This figure was prepared using the default values for electrostatic potential calculations of the program MOLMOL (*48*), and by imposing a charge of −0.5 to each oxygen of the propionates.

of the heme, the cytochromes belonging to the present homologous family should have a quite low redox potential (*14*, *59*). The redox potential of *E. halophila* cytochrome c_{551} is only 58 mV (*60*), but the redox potential of *A. vinelandii* cytochrome c_5 is high [312 mV (*28*)]. It would be interesting to ascertain whether this difference is related to the presence of the disulfide bridge. All structures show a conserved solvent-exposed hydrophobic patch around the edge of the heme where methyl 5 and thioether 4 lie (as observed from the calculated electrostatic potential surface, Figure 6) and a second conserved negatively charged patch on the opposite side of the protein. The latter is next to the surface area closest to Trp31 and Met74, discussed above. The hydrophobic patch is a likely candidate for recognition of the partner and, consequently, protein−protein interaction. Interestingly, thioether 4 is involved in intermolecular contacts in the crystal structure of the complex between yeast cytochrome *c* peroxidase (CCP) and its substrate cytochrome *c* (*61*). Even though yeast cytochrome *c* is not a homologue of *Sp*cytc, the present observation suggests that thioether 4 may be a conserved key docking site for electron transfer. It is to be noted that the region of the surface of yeast cytochrome *c* involved in protein−protein interaction, as identified via both crystallography (*61*) and NMR (*62*), is highly charged. On the other hand, the presently proposed

interaction patch is essentially hydrophobic. A similar contention has been proposed for other bacterial cytochromes, based on their structural features (*63*−*65*). It is also relevant to observe that NMR studies of the interaction between plastocyanin and cytochrome *f* point out that hydrophobic interactions play an important role in determining the correct orientation of the complex (*66*, *67*). Interestingly, the electrostatic contribution was found to be quite relevant in the studies on proteins from plants (*66*), but essentially negligible for the complex of the proteins from the cyanobacterium *Phormidium laminosum* (*67*). The above-mentioned and present studies suggest that hydrophobic interactions may be more important for protein recognition in bacteria with respect to higher organisms.

Functional Considerations. The functional role of *Sp*cyt*c* is completely uncharacterized from the biochemical point of view. However, on the basis of the analysis of the solution structure, and the sequence alignments and structural modeling reported above, as well as on the basis of the availability of the complete genome sequence of *Pseudomonas aeruginosa*, it is possible to infer some hypotheses in this respect. Indeed, here it is proposed that *Sp*cyt*c* acts as an electron carrier involved in the metabolism of nitrogen in *S. putrefaciens*. Below we give the lines of evidence supporting this hypothesis. It should be noted that the observed behavior of oxidized *Sp*cyt*c* may also be related to this protein playing a role in sensing the environmental redox potential.

(1) The sequence of *Sp*cyt*c* is 53% identical to that of the protein product of *P. aeruginosa* open reading frame PA5491, which encodes a putative monoheme *c*-type cytochrome also very similar in length to the present protein. The ORF in the *P. aeruginosa* genome next to this ORF encodes an uncharacterized putative diheme *c*-type cytochrome. Diheme cytochromes similar in sequence to the latter are commonly present in organisms which also contain monoheme cytochromes similar to *Sp*cyt*c*. An example is cytochrome c_4 from *A. vinelandii*, which is implicated in the N_2 fixation pathway (*30*). The sequences of other *c*-type cytochromes from *Pseudomonas* species are less identical to that of *Sp*cyt*c*, and have longer N-terminal extensions.

(2) *A. vinelandii* cytochrome c_5 (also called cycB), which is similar in sequence to *Sp*cyt*c* (level of sequence identity of ∼40%), has been shown to be significantly upregulated (ca. 10-fold) under conditions of nitrogen starvation (*30*). The high expression levels of *A. vinelandii* cytochrome c_5 induced by nitrogen starvation *in vivo* correlated with high nitrogenase activity, indicating that the experimental conditions prompted conversion to N_2 fixation physiology (*30*).

(3) *Sp*cyt*c* is similar in sequence to a cytochrome from *Pseudomonas mendocina*, which has been implicated in nitrate respiration (as indicated in the annotation of the SwissProt entry for this protein).

(4) The ORF encoding *Sp*cyt*c* is between two ORFs, one encoding a protein similar to the product of cytochrome *c* maturation gene A (an ATP-binding protein part of an ABC-type transporter) and the other a protein similar both to another protein proposed to be involved in cytochrome *c* maturation (called NrfG or cycH) and to an *E. coli* protein part of a seven-gene operon involved in nitrite reduction (*68*). Interestingly, some proteins of the latter operon have also been associated with cytochrome *c* maturation. The operon is expressed under anaerobic conditions (*68*), i.e., when O_2 respiration is inactivated.

ACKNOWLEDGMENT

We thank Dr. L. Thöny Meyer for providing us with pEC86 and Dr. D. A. Saffarini for providing us with the ScyA gene.

SUPPORTING INFORMATION AVAILABLE

Three tables listing ^{15}N and ^{1}H NMR chemical shifts in reduced *Sp*cyt*c*, upper distance limits, and dihedral angle constraints used in solution structure calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Pettigrew, G. W., and Moore, G. R. (1987) in *Cytochromes c; Biological Aspects*, Springer-Verlag, Berlin.
2. Moore, G. R., and Pettigrew, G. W. (1990) in *Cytochromes c; Evolutionary, Structural and Physicochemical Aspects*, Springer-Verlag, Berlin.
3. Scott, R. A., and Mauk, A. G. (1996) in *Cytochrome c. A multidisciplinary approach*, University Science Books, Sausalito, CA.
4. Ambler, R. P. (1991) *Biochim. Biophys. Acta 1058*, 42−47.
5. Mathews, F. S. (1985) *Prog. Biophys. Mol. Biol. 45*, 1−56.
6. Banci, L., Bertini, I., Luchinat, C., and Turano, P. (2000) in *The Porphyrin Handbook* (Kadish, K. M., Smith, K. M., and Guilard, R., Eds.) pp 323−350, Academic Press, San Diego.
7. Baxter, S. M., and Fetrow, J. S. (1999) *Biochemistry 38*, 4493−4503.
8. Fetrow, J. S., and Baxter, S. M. (1999) *Biochemistry 38*, 4480−4492.
9. Barker, P. B., Bertini, I., Del Conte, R., Ferguson, S. J., Hajieva, P., Tomlinson, P., Turano, P., and Viezzoli, M. S. (2001) *Eur. J. Biochem. 268*, 4468−4476.
10. Bai, Y. W., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) *Science 269*, 192−197.
11. Pascher, T., Chesick, J. P., Winkler, J. R., and Gray, H. B. (1996) *Science 271*, 1558−1560.
12. Englander, S. W., Sosnick, T. R., Mayne, L. C., Shtilerman, M., Qi, P. X., and Bai, Y. W. (1998) *Acc. Chem. Res. 31*, 737−744.
13. Churg, A. K., and Warshel, A. (1986) *Biochemistry 25*, 1675−1681.
14. Bertrand, P., Mbarki, O., Asso, M., Blanchard, L., Guerlesquin, F., and Tegoni, M. (1995) *Biochemistry 34*, 11071−11079.
15. Battistuzzi, G., Borsari, M., Sola, M., and Francia, F. (1997) *Biochemistry 36*, 16247−16258.
16. Battistuzzi, G., Borsari, M., Cowan, J. A., Eicken, C., Loschi, L., and Sola, M. (1999) *Biochemistry 38*, 5553−5562.
17. Banci, L., Bertini, I., Gray, H. B., Luchinat, C., Reddig, T., Rosato, A., and Turano, P. (1997) *Biochemistry 36*, 9867−9877.
18. Banci, L., Gori Savellini, G., and Turano, P. (1997) *Eur. J. Biochem. 249*, 716−723.
19. Fitch, W. M. (1976) *J. Mol. Evol. 8*, 13−40.
20. Banci, L., Bertini, I., Rosato, A., and Varani, G. (1999) *JBIC, J. Biol. Inorg. Chem. 4*, 824−837.
21. Myers, C. R., and Nealson, K. H. (1988) *Science 240*, 1319−1321.
22. Myers, C. R., and Nealson, K. H. (1990) *J. Bacteriol. 172*, 6232−6238.
23. Myers, C. R., and Myers, J. M. (1992) *J. Bacteriol. 174*, 3429−3438.
24. Myers, C. R., and Myers, J. M. (1993) *FEMS Microbiol. Lett. 114*, 215−222.
25. Myers, C. R., and Myers, J. M. (1997) *J. Bacteriol. 179*, 1143−1152.

26. Myers, C. R., and Myers, J. M. (2001) *J. Appl. Bacteriol. 76*, 253−258.
27. Morris, C. J., Gibson, D. M., and Ward, F. B. (1990) *FEMS Microbiol. Lett. 69*, 259−262.
28. Carter, D. C., Melis, K. A., O'Donnell, S. E., Burgess, B. K., Furey, W. F., and Wang, B. C. (1985) *J. Mol. Biol. 184*, 279−295.
29. Bersch, B., Blackledge, M. J., Meyer, T., and Marion, D. (1996) *J. Mol. Biol. 264*, 567−584.
30. Rey, L., and Maier, R. J. (1997) *J. Bacteriol. 179*, 7191−7196.
31. Ambler, R. P., Meyer, T. E., and Kamen, M. (1993) *Arch. Biochem. Biophys. 306*, 83−93.
32. Barker, P. D., Nerou, E. P., Freund, S. M. V., and Fearnley, I. M. (1995) *Biochemistry 34*, 15191−15203.
33. Arslan, E., Schulz, H., Zufferey, R., Kuenzler, P., and Thoeny-Meyer, L. (1998) *Biochem. Biophys. Res. Commun. 251*, 744−747.
34. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A. (1989) *J. Am. Chem. Soc. 111*, 1515−1517.
35. Bax, A., and Davis, D. G. (1985) *J. Magn. Reson. 63*, 207−213.
36. Wider, G., Neri, D., Otting, G., and Wüthrich, K. (1989) *J. Magn. Reson. 85*, 426−431.
37. Wider, G., Macura, S., Kumar, A., Ernst, R. R., and Wüthrich, K. (1984) *J. Magn. Reson. 56*, 207−234.
38. Vuister, G. W., and Bax, A. (1993) *J. Am. Chem. Soc. 115*, 7772−7777.
39. Archer, S. J., Ikura, M., Torchia, D. A., and Bax, A. (1991) *J. Magn. Reson. 95*, 636−641.
40. Wüthrich, K. (1986) in *NMR of Proteins and Nucleic Acids*, Wiley, New York.
41. Güntert, P., Braun, W., and Wüthrich, K. (1991) *J. Mol. Biol. 217*, 517−530.
42. Gagné, R. R., Tsuda, S., Li, M. X., Chandra, M., Smillie, L. B., and Sykes, B. D. (1994) *Protein Sci. 3*, 1961−1974.
43. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) *J. Mol. Biol. 273*, 283−298.
44. Banci, L., Bertini, I., Cremonini, M. A., Gori Savellini, G., Luchinat, C., Wüthrich, K., and Güntert, P. (1998) *J. Biomol. NMR 12*, 553−557.
45. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (1999) *AMBER 6*, University of California, San Francisco.
46. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
47. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR 8*, 477−486.
48. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graphics 14*, 51−55.
49. Altschul, F. S., Madden, T. L., Schaeffer, A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res. 25*, 3389−3402.
50. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res. 24*, 4876−4882.
51. Felsenstein, J. (1989) *Cladistics 5*, 164−166.
52. Sali, A., and Blundell, T. L. (1993) *J. Mol. Biol. 234*, 779−815.
53. Sippl, M. J. (1993) *Proteins: Struct., Funct., Genet. 17*, 355−362.
54. Banci, L., Bertini, I., Turano, P., and Vicens Oliver, M. (1992) *Eur. J. Biochem. 204*, 107−112.
55. Bertini, I., Gori, G., Luchinat, C., and Vila, A. J. (1993) *Biochemistry 32*, 776−783.
56. Moura, I., Liu, M. Y., Costa, C., Liu, M. C., Pai, G., Xavier, A. V., LeGall, J., Payne, W. J., and Moura, J. J. (1988) *Eur. J. Biochem. 177*, 673−682.
57. Kumar, S., and Bansal, M. (1998) *Proteins: Struct., Funct., Genet. 31* (4), 460−476.
58. Richardson, J. S., and Richardson, D. C. (1988) *Science 240*, 1648−1652.
59. Tezcan, F. A., Winkler, J. R., and Gray, B. H. (1998) *J. Am. Chem. Soc. 120*, 13383−13388.
60. Meyer, T. E. (1985) *Biochim. Biophys. Acta 806*, 175−183.
61. Pelletier, H., and Kraut, J. (1992) *Science 258*, 1748−1755.
62. Worrall, J. A. R., Kolczak, U., Canters, G. W., and Ubbink, M. (2001) *Biochemistry 40*, 7069−7076.
63. Than, M. E., Hof, P., Huber, R., Bourenkov, G. P., Bartunik, H. D., Buse, G., and Soulimane, T. (1997) *J. Mol. Biol. 271*, 629−644.
64. Soulimane, T., Buse, G., Bourenkov, G. P., Bartunik, H. D., Huber, R., and Than, M. E. (2000) *EMBO J. 19*, 1766−1776.
65. Banci, L., Bertini, I., Ciurli, S., Dikiy, A., Dittmer, J., Rosato, A., Sciara, G., and Thompsett, A. (2002) *ChemBioChem* (in press).
66. Ejdeback, M., Bergkvist, A., Karlsson, B. G., and Ubbink, M. (2000) *Biochemistry 39*, 5022−5027.
67. Crowley, P. B., Otting, G., Schlarb-Ridley, B., Canters, G. W., and Ubbink, M. (2001) *J. Am. Chem. Soc. 123*, 10444−10453.
68. Iobbi-Nivol, C., Crooke, H., Griffiths, L., Grove, J., Hussain, H., Pommier, J., Mejean, V., and Cole, J. A. (1994) *FEMS Microbiol. Lett. 119* (1−2), 89−94.

BI015984Z