NMR Solution Structure, Backbone Mobility, and Homology Modeling of *c*-Type Cytochromes from Gram-Positive Bacteria

Lucia Banci,[a, b] Ivano Bertini,*[a, b] Stefano Ciurli,[a, c] Alexander Dikiy,[a, c] Jens Dittmer,[a] Antonio Rosato,[a, b] Giuliano Sciara,[c] and Andrew R. Thompsett[a]

The solution structure of oxidized cytochrome c_{553} (71 amino acid residues) from the Gram-positive bacterium Bacillus pasteurii *is here reported and compared with the available crystal structure. The solution structure is obtained from 1609 meaningful NOE data (22.7 per residue), 76 dihedral angles, and 59 pseudocontact shifts. The root mean square deviations from the average structure are 0.25 ± 0.07 and 0.59 ± 0.13 Å for the backbone and all heavy atoms, respectively, and the quality assessment of the structure is satisfactory. The solution structure closely reproduces the fold observed in the crystal structure. The backbone mobility was then investigated through amide ^{15}N relaxation rate and $^{15}N-^{1}H$ NOE measurements. The protein is rigid in both the sub-nanosecond and millisecond time scales, probably due to the relatively large heme:number of amino acids ratio. Modeling of eight* c*-type cytochromes from other Gram-positive bacteria with a high sequence identity (> 30%) to the present cytochrome c_{553} was performed. Analysis of consensus features accounts for the relatively low reduction potential as being due to extensive heme hydration and indicates residues 34 – 35, 44 – 46, 69 – 72, and 75 as a conserved hydrophobic patch for the interaction with a protein partner. At variance with mitochondrial* c*-type cytochrome, this protein does not experience pH-dependent coordination equilibria. The reasons for this difference are analyzed.*

KEYWORDS:

backbone dynamics · cytochromes · electron transfer · molecular modeling · solution structure

Introduction

Gram-positive microorganisms lack a true periplasmic space and are therefore obliged to store cytochromes as membrane-bound proteins. The difficulties associated with the study of such water-insoluble proteins may explain why knowledge of the properties and physiological role of *c*-type cytochromes in Gram-positive bacteria is very limited as compared to the information available for Gram-negative bacteria and eukaryotes.[1]

The present work is focused on a small (9.6 KDa), acidic (pI = 3.3) cytochrome c_{553} present in the soluble fraction of the Gram-positive, alkaliphilic, and highly ureolytic soil bacterium *Bacillus pasteurii*[2] and believed to be involved in electron-transfer processes to the terminal oxidase.[3] This microorganism grows optimally at pH 9.2 in the presence of ammonium salts[4, 5] or urea.[6, 7] Large amounts of membrane-bound cytochromes of the *b*-and *c*-types, characterized by low redox potentials, are typically found in alkaliphilic bacteria. Cytochrome c_{553} from *B. pasteurii* (denoted *Bp*cytc hereafter) contains a *c*-type heme with a hexacoordinate low-spin Fe ion axially bound to His and Met residues, as determined by paramagnetic NMR spectroscopy.[8] The amino acid sequence of *Bp*cytc suggests that in vivo this protein is bound to the cytoplasmic membrane through a diacyl – glyceryl – cysteine anchor located at the N-terminus random coil tail and that the soluble form is obtained by proteolysis of the lipoprotein, which occurs during cell disruption.[3] Direct electrochemical measurements performed with a pyrolytic graphite edge electrode have shown that this protein is characterized by a low reduction potential ($E^{\circ\prime} = +47$ mV) as expected for cytochromes from alkaliphiles.[8] The thermodynamic parameters for the one-electron reduction process are consistent with a heme highly exposed to the solvent, as well as with solvent-reorganization effects upon reduction.[8] The *Bp*cytc structure was solved at 0.97 Å resolution by using synchrotron X-ray diffraction data on a single crystal obtained at pH 5.1 and frozen at 100 K.[9]

The present work aims at obtaining a thorough characterization of oxidized *Bp*cytc in solution, both from the structural and dynamic points of view, under experimental conditions more physiologically relevant than the solid-state structure. In

[a] *Prof. I. Bertini, L. Banci, S. Ciurli, A. Dikiy, J. Dittmer, A. Rosato, A. R. Thompsett*
Centro di Risonanze Magnetiche
University of Florence
Via Luigi Sacconi 6, 50019, Sesto Fiorentino (Italy)
Fax: (+39) 055-4574271
E-mail: bertini@cerm.unifi.it

[b] *Prof. I. Bertini, L. Banci, A. Rosato*
Department of Chemistry, University of Florence
Via Gino Capponi 9, 50121, Florence (Italy)

[c] *S. Ciurli, A. Dikiy, G. Sciara*
Department of Agro-Environmental Science and Technology
Via Filippo Re 8, 40127 University of Bologna (Italy)

Supporting information for this article is available on the WWW under http://www.chembiochem.com or from the author.

order to complement and extend this analysis, the structures of cytochromes from other Gram-positive bacteria displaying high sequence similarity to *Bpcytc* were modeled and comparatively analyzed, in order to establish a broader frame for the functional interpretation of the results. Additional hints on functionally important features of *Bpcytc* are provided by the detailed comparison of its dynamic properties with those of the prototype mitochondrial cytochrome *c*, that is, *Saccharomyces cerevisiae* iso-1 cytochrome *c*.

Results

Solution structure calculation

The present solution structure is determined for the oxidized state of the soluble fragment of *Bacillus pasteurii* cytochrome c_{553} (*Bpcytc*). A very large percentage of the total number of expected resonances (96% and 98% for ^{1}H and ^{15}N nuclei, respectively) was assigned. The ^{1}H and ^{15}N assignment achieved in this study, together with the stereospecific assignment, is reported in Table 1 of the Supporting Information (see also Figure 1), and has also been deposited in the BioMagResBank (entry number 5172).

Table 1 reports the details of the experimental constraints used for structure calculations; three different types of constraints were used: 1) proton–proton distance constraints, 2) pseudocontact shifts, and 3) dihedral angle constraints. The resulting average number of meaningful constraints per residue (22.7) allowed us to obtain a reliable solution structure. Figure 2A reports the distribution of the meaningful NOEs per residue. The pattern of sequential and medium-range NOEs involving NH, Hα, and Hβ protons features a series of strong sequential $d_{NH_i-NH_{i+1}}$ connectivities typical for proteins with a high content of α-helical secondary structure (see Supporting Information, Figure 1S). This is confirmed by the detection of $d_{H\alpha-NH_{(i,i+3)}}$ as well as $d_{H\alpha-H\beta_{(i,i+3)}}$ connectivities for a large number of amino acids. The presence of helical structure in these regions is also consistent with the small values of $^3J_{H\alpha-NH}$ constants observed, and the fact that their intra/interresidue Hα–NH NOESY cross-peak ratios are larger than unity.

All 30 conformers constituting the PSEUDYANA family have a total target function smaller than 0.7 Å^2 and show no consistent violations and no residual violation exceeding 0.17 Å. Their global root mean square deviations (RMSD) from the mean structure are 0.25 ± 0.07 and 0.59 ± 0.13 Å for the backbone and heavy atoms, respectively (for residues 24–89). The corresponding parameters for the PSEUDOREM family are 0.24 ± 0.04 and 0.60 ± 0.06 Å, with a total average penalty function of 5.2 ± 0.8 kJ mol^{-1}. The RMSD values relative to the PSEUDOREM family are plotted in Figure 2B. A "sausage" view of the backbone is shown in Figure 3A.

The structure quality parameters for the REM family and for the restrained energy minimized mean structure are reported in Table 1. The total number of meaningful constraints, the relatively small number and size of the violations, the distribution of backbone dihedral angles in the Ramachandran plot, and the overall G-factor (as defined in ref. [10]) indicate that the obtained structure is of very good quality. Coordinates for the final family of conformers are being deposited in the Protein Data Bank (PDB codes: 1K3H and 1K3G).

The magnetic anisotropy tensor

58 Pseudocontact shift (PCS) values were determined for ^{1}H backbone amide protons. In addition, the PCS value of the Hε1 proton of Trp87 was measured. These 59 input data were used in structure calculations as described in the Materials and Methods section. The final magnetic susceptibility anisotropy parameters were: $\Delta\chi_{ax} = (2.20 \pm 0.10) \times 10^{-32}$ m^3 and $\Delta\chi_{rh} = (-0.18 \pm 0.15) \times 10^{-32}$ m^3. The *z* axis of the tensor was closely aligned to the normal to the heme plane featuring a deviation of approximately 10°, while the *x* axis of the tensor formed an angle of roughly 30° with the Fe–pyrrole I nitrogen direction. The average deviation exceeding the tolerance between calculated (by fitting to the final family) and experimental PCS data was smaller than 0.03 ppm. The present tensor parameters, and particularly $\Delta\chi_{rh}$, are smaller than those previously reported for other *c*-type cytochromes, whose $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ values were around $2.2-2.8 \times 10^{-32}$ and $-1.0 - -1.4 \times 10^{-32}$ m^3, respectively.[11, 12] A different doming of the heme moiety and/or a different strength of the coordination

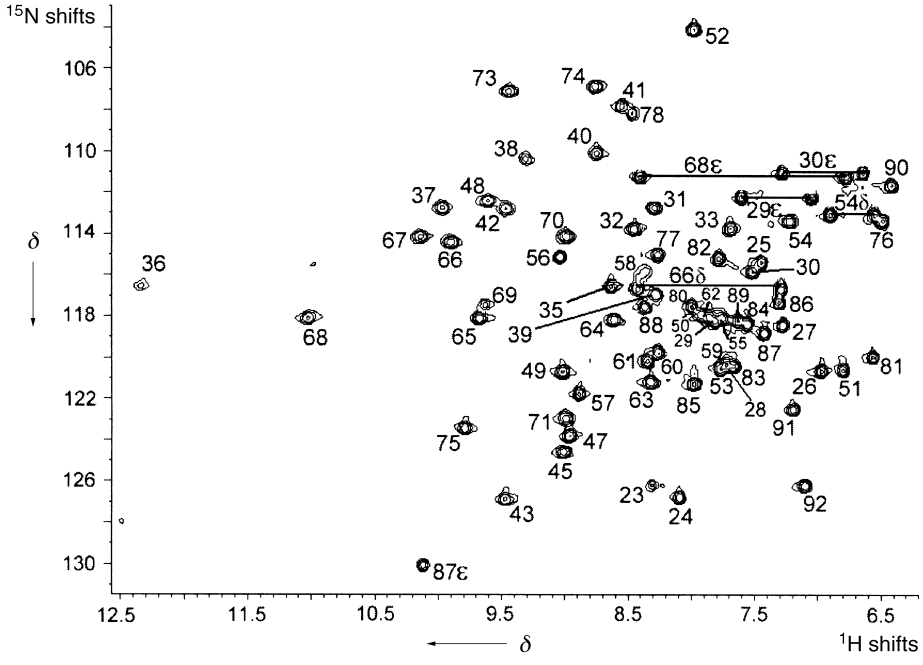

Figure 1. *^{1}H-^{15}N HSQC spectrum of oxidized recombinant cytochrome c_{553} from* Bacillus pasteurii, *at 500 MHz proton frequency, 298 K, and pH 7.5. The labels show the assignment of the peaks.*

Table 1. *Summary of NMR constraints used for structure calculation, restraint violations, structural statistics, and energetics for the restrained energy minimized solution structure of oxidized cytochrome c_{553} from* B. pasteurii.

Structural constraint	Total no.	Pseudorem		< Pseudorem >	
		Average no. of violations per conformer	RMS violation per constraint	No. of violations	RMS violation per constraint
meaningful (total) NOESY	1564 (2010)	12.1 ± 2.5	0.009 ± 0.001 Å	13	0.010 Å
intraresidue[a]	263	2.2 ± 0.8	0.008 ± 0.003 Å	3	0.008 Å
sequential[a]	382	4.6 ± 1.4	0.010 ± 0.002 Å	5	0.010 Å
medium range[a],[b]	426	3.0 ± 1.7	0.006 ± 0.003 Å	3	0.008 Å
long range[a]	493	2.3 ± 1.3	0.010 ± 0.002 Å	2	0.012 Å
1D NOE	45	0.93 ± 0.25	0.002 ± 0.001 Å	1	0.015 Å
ϕ	38	0.2 ± 0.4	$0.19 \pm 0.35^\circ$	1	0.67°
ψ	38	1.0 ± 0.6	$0.67 \pm 0.43^\circ$	1	0.58°
Fe – heme ligand distances	4	0	0 Å	0	0 Å
pseudocontact shifts[c]	59	4.8 ± 0.9	0.007 ± 0.001 ppm	4	0.006 ± 0.001 ppm
violations between 0.1 – 0.3 Å		3.3 ± 1.2		5	
violations larger than 0.3 Å		0		0	
largest distance violation		0.26 Å		0.23 Å	
largest ϕ violation		7°		4°	
largest ψ violation		11°		4°	
largest pseudocontact shift violation		0.42 ppm		0.19 ppm	

Energetics	Average value	Value
total target function	0.38 ± 0.04	0.47
AMBER average total energy	-896.1 ± 38.7 kJ mol^{-1}	-909.8 kJ mol^{-1}

Structure analysis	Value	Average value	Value
completeness of ^{15}N backbone assignment	98.5%		
completeness of ^{1}H backbone assignment	99.2%		
completeness of NMR-observable proton contacts within 4 Å	81.4%		
structural constraints per residue	24.6		
residues in most favored regions[d]		83.3%	80.7%
residues in allowed regions[d]		14.1%	15.8%
residues in generously allowed regions[d]		2.6%	3.5%
residues in disallowed regions[d]		0%	0%
overall G-factor[e]		−0.1	−0.09

[a] Only violations larger than 0.05 Å are taken into account. [b] Medium-range distance constraints are those between residues (i, i + 2), (i, i + 3), (i, i + 4), and (i, i + 5). [c] Only violations larger than 0.1 ppm are taken into account. [d] According to the Ramachandran plot. [e] The program PROCHECK was used to check the overall quality of the structure. An overall G-factor larger than −0.5 is expected for a good quality structure.[10]

bond of the axial methionine may be plausible reasons for this difference. It may be worthwhile to note here that the high-resolution X-ray crystal structure shows a very planar heme group, with the heavy-atom RMSD from the average plane being 0.08 Å.[11]

Protein mobility from ^{15}N relaxation and amide exchange data

Reliable ^{15}N relaxation data to be used in the model-free analysis were obtained for 65 out of 68 non-proline residues (excluding also the N-terminal residue Val 22) and for the Nε1 of Trp 87. The average R_1, R_2, and heteronuclear NOE were (3.06 ± 0.21) s^{-1}, (5.50 ± 0.84) s^{-1}, and 0.58 ± 0.06, respectively, at 400 MHz and (2.37 ± 0.30) s^{-1}, (6.11 ± 1.59) s^{-1}, and 0.79 ± 0.06 at 600 MHz. For the large majority of residues, six data were used as input for the Modelfree program, that is R_1, R_2, and the heteronuclear NOE at 400 and 600 MHz. In a few cases only five data were used, due to bad peak shape or low peak intensity in one of the data sets. The data used for Modelfree calculations, together with their uncertainty, are reported in the Supporting Information. The reorientation of the molecule in solution was taken to be isotropic, as the principal values of the tensor of inertia calculated from either the crystal or the solution structure are in a ratio of 1.00:0.98:0.92 or 1.00:0.98:0.95, respectively.

The results of the Modelfree fittings are shown in Figure 4. 63 out of the 66 input data sets could be fitted satisfactorily using the Modelfree approach. Relaxation data for residues 48, 71, and 80 were not fitted well by any model; consequently in the final calculations they were fitted to the simplest model. This is commonly observed behavior for a small fraction of residues (in the present case, approximately 4.5%).[13] In summary, all fittings were performed with a maximum of two adjustable parameters, by using the models involving only S^2, or S^2 and τ_e, or S^2 and R_{ex} of the standard Modelfree analysis (Figure 4). The average S^2 value over all residues was 0.86 ± 0.05. The large majority of residues were adequately fitted with the simplest model. Only four residues were fitted by using a correlation time for internal motions, τ_e (ranging from 50 – 100 ps), as an adjustable parameter: Asp 23, which is the second residue in the sequence, Lys 31, and the last two C-terminal residues, Lys 91 and Lys 92 (Figure 4). Inclusion of an R_{ex} term (up to 5 s^{-1}) in the fittings was required for Ile 33, residues 39 – 41, and Ser 44 (Figure 4).

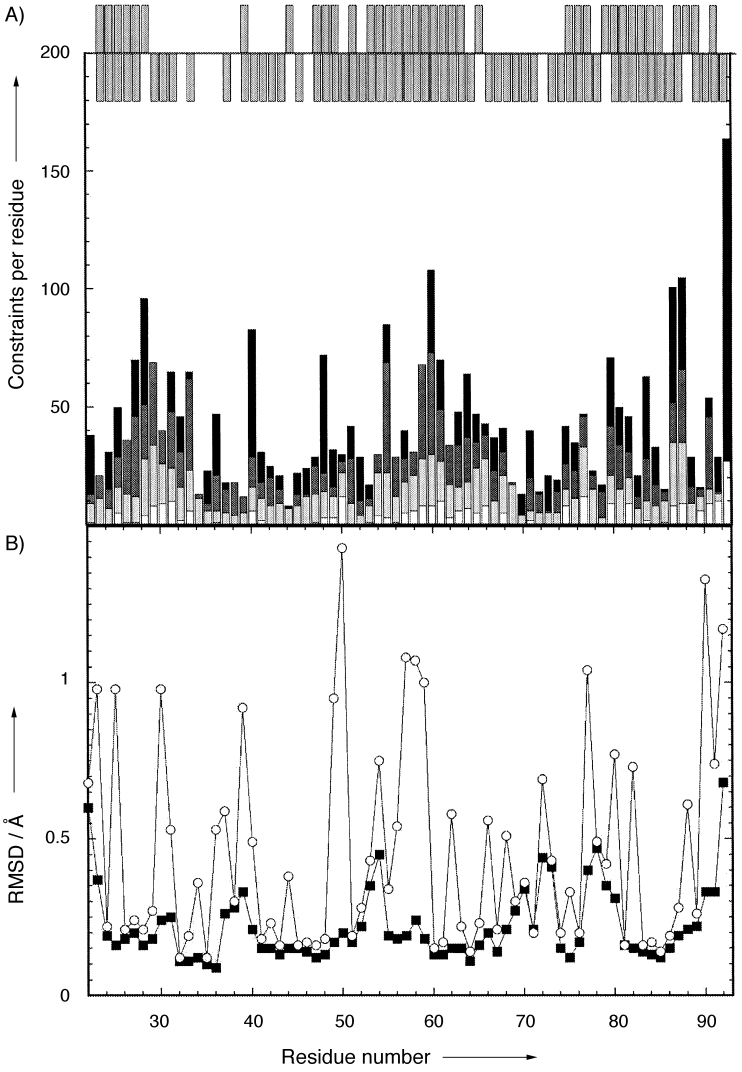

Figure 2. *A) Distribution of meaningful NOEs per residue, used for structure calculation of oxidized cytochrome c_{553} from* Bacillus pasteurii. *Intraresidual, sequential, medium-range, and long-range NOE constraints are in white, light gray, dark gray, and black, respectively. The NOE constraints for the heme moiety are reported as belonging to residue 93. At the top, the residue-by-residue constraints used for the calculation, other than NOEs, are also reported (upward bars: ϕ and Ψ dihedral angle constraints; downward bars: pseudocontact shift constraints). In addition, a pseudocontact shift constraint for the Trp 87 NHε proton was also used for structure calculations. B) Diagram of global backbone (filled squares) and heavy atom (open circles) RMSD values per residue with respect to the mean structure for the 30 PSEUDOREM structures of oxidized cytochrome c_{553} from* Bacillus pasteurii.

The analysis of the dependence of the R_2 relaxation rates of the various residues on the Carr – Purcell – Meiboom – Gill (CPMG) frequency was largely consistent with the above findings from the model-free fittings. Indeed, the R_2 relaxation rate of residues 33 and 39 – 41 showed a distinct dependence on ν_{CPMG}, which could be satisfactorily fitted with Equation (1) as defined in the Materials and Methods section. The R_2 relaxation rate of residue Ser 44 was not dependent on ν_{CPMG}, and its value is significantly higher than the mean R_2 for all other residues (excluding 33 and 39 – 41); this implies that Ser 44 experiences conformational exchange on a time scale shorter than that presently sampled here (approximately 50 – 600 μs).

The protein mobility in the longer time range was probed by using H/D exchange measurements and comparing the results with amide NH solvent accessibility calculated through the program WHATIF.[58] The ^1H-^{15}N HSQC signals for 31 residues belonging to the N terminus, helix 2 and the following loop, the N-terminal sections of helices 4 and 5, and the region 69 – 73, which comprises the heme axial ligand Met 71, were found to totally disappear in a spectrum acquired 10 minutes after dissolving a lyophilized sample of *Bp*cytc in D_2O. Among the remaining 40 residues, the exchange rates were found to span a range of $10^{-3} - 10^{-5}\ s^{-1}$, with the residues with the smallest exchange rates found in helices 4 and 5. Overall, only 18 cross peaks were still observable in the ^1H-^{15}N HSQC spectrum after two days of contact with a large excess of D_2O, a fact indicating that the corresponding protons are not exchangeable. This is consistent with these protons forming strong hydrogen bonds typical of α helices and being inaccessible to the solvent. Some of the exchanging protons (17 in total) are inaccessible to solvent, and therefore must undergo conformational changes in order to interact with bulk water. These latter residues are mainly located in the hydrophobic region comprising residues 33 – 47, that is the long loop following the N-terminal helix, and it is significant that several residues in this region also experience mobility in the shorter time scales, as described above. The remaining regions are characterized

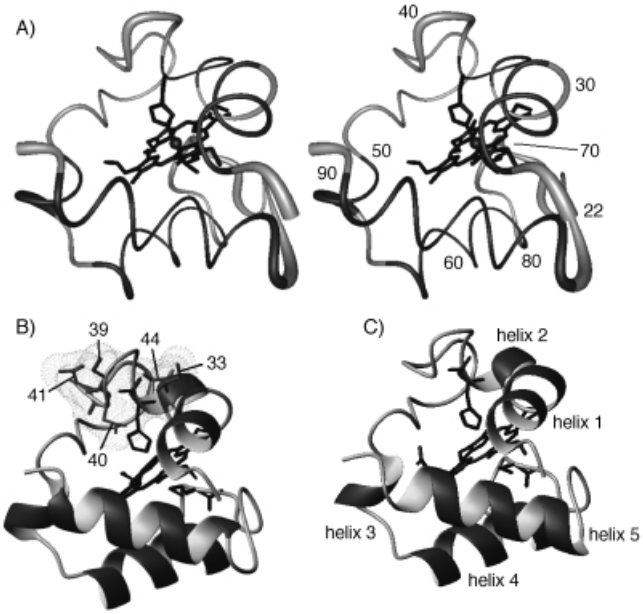

Figure 3. *A) Stereoview of the backbone of the final PSEUDOREM family of oxidized cytochrome c_{553} from* Bacillus pasteurii *as a tube with variable radius, proportional to the backbone RMSD of each residue. Residues in α helices are in dark gray. The heme moiety and the side chains of the axial ligands are also shown. B) Ribbon display of the average solution structure. Residues displaying conformational exchange as probed by ^{15}N-relaxation measurements are shown as sticks (only heavy atoms). The Van der Waals surface of the same residues is also shown. C) Ribbon display of the crystal structure (PDB entry: 1C75).[9] This figure was prepared with the MOLMOL program.[56]*

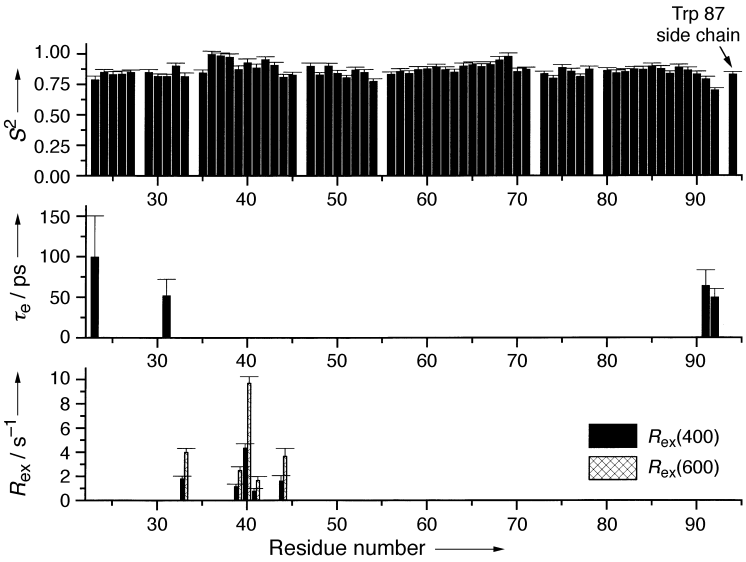

Figure 4. *Per-residue plots of the best-fit model-free parameters obtained from ^{15}N-relaxation data for oxidized cytochrome c_{553} from* Bacillus pasteurii.

by both solvent-exchange phenomena and by a significant solvent accessibility, the latter precluding a definitive conclusion about their mobility in the longer time scales.

Sequence analysis and structural modeling

From the analysis of the results of the BLAST[14] search and of the corresponding phylogenetic tree (not shown), it clearly appears that *Bpcytc* is closely related only to other *c*-type cytochromes from Gram-positive organisms of the *Bacillus*/*Clostridium* group. The sequence identity values between *Bpcytc* and these cytochromes are in the range 30 – 54 % (Table 2). The two sequences closest to *Bpcytc* are those of *B. licheniformis* and of the thermophilic *Bacillus* sp. PS3 cytochrome c_{551} (54 % and 52 % sequence identity, respectively). Identity percentages within the present family of sequences are in the range of 20 – 60 %, which indicates a high degree of variability (Table 2). All the above mentioned sequences, except that of *B. licheniformis*, have a 40 – 50 residue long hydrophobic N-terminal extension, whose role is likely to be that of anchoring the cytochrome to the membrane. This holds also for *Bpcytc*.[3] As the present work is focused on the soluble portion of *Bpcytc*, these N-terminal tails have not been taken into account. Sequence identity values between *Bpcytc* and other bacterial cytochromes are lower than 35 %, the one with highest similarity being a cytochrome from *Paracoccus denitrificans* (32 % sequence identity).

Interestingly, another distinct class of proteins to which *Bpcytc* bears some similarity is the *b*/*c* subunit of menaquinol-cytochrome *c* reductase, from three different *Bacillus* species. These proteins are approximately 250 amino acids long and are very similar to one another, with sequence identity values of 60 – 75 %. Their cytochrome *c* like portion has a sequence identity to *Bpcytc* of approximately 40 %. Finally, some distant sequence similarity is also found between *Bpcytc* and subunit II of bacterial cytochrome *c* oxidases (15 – 25 %).

Structural models have been built for the various *Bacillus* and *Heliobacterium* cytochromes mentioned above, based on the structure of *Bpcytc* and sequence alignments (Figure 5). The backbone RMSD values with respect to *Bpcytc* for the models of *Bacillus* cytochromes are lower than 0.5 Å, while those for *Heliobacterium* cytochromes are of the order of 0.7 Å. All structures display the same secondary elements, namely five helices involving residues 24 – 31, 33 – 36, 51 – 54, 57 – 66, and 78 – 89. With the exception of helix 33 – 36, which is a 3_{10} helix in one structure, they are all consistently α helices. The orientation of the iron axial ligands is strictly maintained in all the structural models. An interesting feature is the conservation of the hydrogen bond between the NH moiety of the imidazole ring of the axial His 36 and the backbone carbonyl oxygen atom of Pro 46, as also found in *c*-type cytochromes in eukaryotes.[15]

Discussion

Description of the structure and comparison with the X-ray crystal structure

The mean solution structure of *Bpcytc* is characterized by five α-helix regions: Ala 24 – Gln 29 (helix 1), Ile 33 – His 36 (helix 2),

Table 2. *Percentage sequence identity values (upper part) and backbone RMSD values (Å) for the structural models (lower part) of c-type cytochromes from different Gram-positive bacteria.*[a]

	B.p. c_{553}	*B.l.* c_{551}	*B.*PS3 c_{551}	*B.h.* c_{551}	*B.h.* c_{551} prec.	*B.s.* c_{550} (cccA)[b]	*B.s.* c_{551} (cccB)[b]	*H.m.* c_{553}	*H.g.* c_{553}
B.p. c_{553}	–	53.6	52.1	43.7	33.8	29.6	45.1	36.6	42.2
B.l. c_{551}	0.28	–	50.0	45.8	39.7	45.1	63.9	33.3	43.1
*B.*PS3 c_{551}	0.15	0.27	–	32.4	32.4	34.2	41.4	35.1	31.5
B.h. c_{551}	0.16	0.29	0.13	–	27.0	33.3	33.9	28.7	22.8
B.h. c_{551} prec.	0.19	0.25	0.16	0.17	–	27.9	34.9	25.2	21.8
B.s. c_{550} (cccA)	0.14	0.28	0.18	0.18	0.19	–	33.9	24.6	22.7
B.s. c_{551} (cccB)	0.15	0.28	0.17	0.21	0.21	0.15	–	35.7	33
H.m. c_{553}	0.70	0.71	0.71	0.69	0.69	0.71	0.67	–	60.6
H.g. c_{553}	0.71	0.70	0.72	0.71	0.67	0.66	0.72	0.35	–

[a] *B.p.* = *Bacillus pasteurii*; *B.l.* = *Bacillus licheniformis*; *B.*PS3 = *Bacillus PS3*; *B.h.* = *Bacillus halodurans*; *B.s.* = *Bacillus subtilis*; *H.m.* = *Heliobacillus mobilis*; *H.g.* = *Heliobacterium gestii*. [b] The name with which the corresponding gene is labeled in the genome sequence is given in parentheses.

```
B.pasteurii     ------------------------------------------------------------
B.licheniformis ------------------------------------------------------------
B.sp. PS3       ------------MKWKLAAMFLGVSLALAACGGGGDNAGEKNGGSNGGGDTAA------
B.halodurans    MKGRPLLPFAIIAVVGILLMISISIIGLNQRAAMDEEGEGADTEEVVEFDDPVE------
B.halodurans(II)--------MKKCLFALSGLMVLGSLVACGGSGENVEGTPADTEEVVGDFDATMA------
B.subtilis cccA MKWNPLIPFLLIAVLGIGLTFFLSVKGLDDSREIASGGE---SKSAEKKDANAS------
B.subtilis cccB -----------MKSKLSILMIGFALSVLLAACGSNDAKE-EKTDTGSKTEATAS------
H.mobilis       -----MKNFKLIGVAAVLGLSLVTLTACNGGSQAPAEKKETPAVKTEQAPAPAPKVETKT
H.gestii        GGENTMKNLKLIGIAAVMGLSMVALTACNSGSKAPDAAKPAPAPSSAPAPAPAP-----A

                                    30            40        50        60
B.pasteurii     -----------VDAEAVVQQK--CISCH--GGDLTGASAPAIDKAGANYSEEEILDIILN
B.licheniformis --------ATATDGEEIYQQN--CTGCH--GTDLAGGSAPSLKEVGGKYKESEIKDIVVN
B.sp. PS3       ------------AAEQIFKQN--CASCH--GQDLSGGVGPNLQKVGSKYSKDEIKNIIAN
B.halodurans    ------------AGEQIVNQS--CIGCH--GGDLSGGAGPALTGLEGKYSAEEIADIVQN
B.halodurans(II)--------------RETYEAA--CIACH--GGNLEGGAGPQLTDG--AYSYEEIIHAIEH
B.subtilis cccA -------------PEEIYKAN--CIACH--GENYEGVSGPSLKGVGDKKDVAEIKTKIEK
B.subtilis cccB ------------EGEELYQQS--CVGCH--GKDLEGVSGPNLQEVGGKYDEHKIESIIKN
H.mobilis       EAKTETKTADAGKGKELFNGKGACNSCHKIGSEGTAAIGPDLEKVGAKYDADKIAKILAN
H.gestii        DKPTAAPAAAGADAKALFTGKGACITCHKLGTEGALEVGPNLGEVGKKYNEEKIYKILVN
                                       *  **  *        *            *

                        70         80        90
B.pasteurii     --G---QGGMP-GGIAKGAEAEAVAAWLAEKK
B.licheniformis --G---RGGMP-GNLVDEK-AEAVAKWLSEK-
B.sp. PS3       --G---RGAMP-AGIIKGEDADKVAEWLAAKK
B.halodurans    --G---KGAMP-AMPHDDAEADAIAQYLLSIE
B.halodurans(II)--G---KGAMP-PQNVDQEEAENLAKWIEAQ-
B.subtilis cccA --G---GNGMP-SGLVPADKLDDMAEWVSKIK
B.subtilis cccB --G---RGNMP-KGLVDDNEAAVIAKWLSEKK
H.mobilis       PTGEGLQATMPPAAALSDDEKKEVAKFLASQK
H.gestii        PSGEGLQATMPAATTLSDDEKKAVAKFLAEKK
                  *      **             *
```

Figure 5. *Sequence alignment of c-type cytochromes from different Gram-positive bacteria (see also Table 2). Stars indicate conserved residues. The numbering used is that of the complete sequence of cytochrome* c_{553} *from* Bacillus pasteurii.

Ala 51 – Asn 54 (helix 3), Glu 57 – Leu 65 (helix 4), and Gly 76 – Glu 90 (helix 5) (Figure 3B). This secondary structure compares very well with that observed in the crystal structure (Figure 3C). The RMSD values between these two structures are 0.61 and 1.15 Å, for the backbone and heavy atoms, respectively. The largest deviations occur for resides 68 – 70, where a displacement of this loop region with respect to the solid-state structure is observed. The backbone amide protons of residues 69 – 70 do not show significant mobility in the model-free analysis, and their large exchangeability in D_2O cannot be taken as an indication of the presence of possible slow conformational rearrangements in solution because of their concomitantly large solvent accessibility. A possible contribution to this displacement in solution may be simply due to the lower than average number of structural constraints in this region, which is in turn due to the location of these residues on top of a long loop. Other regions featuring relatively high backbone RMSD values with respect to the crystal structure are Ser 34, Gly 74, Ala 78, and Glu 79. The latter two residues correspond to a maximum of the RMSD within the present family itself (Figure 2B). Some of the differences observed may also be due to the relative paucity of long-range NOEs for these residues (Figure 2A). Thus, it may be concluded that the backbone of the protein in solution has a conformation very similar to that observed in the solid state (Figure 3). As far as the metal site is concerned, the only significant rearrangement is that of propionate-7. In the solid-state structure, solved at pH 5.1, its conformation suggests that it is protonated and hydrogen-bonded to the carbonyl oxygen atom of Ala 47.[9] The pH dependence of the reduction potential and NMR chemical shifts of *Bpcytc* suggested the presence of a pK_a value of ≈ 5.5, ascribed to the protonation equilibrium of propionate-7.[8] Consistently, in the calculated solution structure at pH 7.5 the propionate appears to protrude into the solvent, not forming any consistent hydrogen bonds (see also Figure 2S in the Supporting Information). This rearrangement is consistent with the lack of NOE data. In spite of the conformational rearrangement of propionate-7, the solvent accessibility of the heme moiety in the present structure and in the solid-state structure is essentially identical.

Homology modeling

The complete lack of structural information (except for the crystal structure of the present protein[9]) on cytochromes from Gram-positive bacteria prompted us to obtain the structures of several cytochromes belonging to this bacterial class through a building-by-homology molecular-modeling approach. The modeled structures of six *Bacillus* cytochromes as well as those of two

cytochromes from *Heliobacteria* are extremely similar to that of *Bp*cytc, as shown by their low backbone RMSD (Table 2). This similarity can be partially overestimated by the present modeling approach, which was based on the use of a single template structure. However, PROSA[16] analyses indicate that the predicted fold of these proteins corresponds to a minimum in energy, thereby indicating that they are substantially correct. The *Heliobacterium* and *Heliobacillus* models display somewhat higher backbone RMSD values because they have several insertions in the sequence. In particular, modeling of the N-terminal tail and of the region aligning with residues 45 – 47 of *Bp*cytc turns out to be quite difficult. This leads to less-satisfactory, but still acceptable, results in the analysis with PROSA.[16]

In agreement with the low RMSD between the various structures, all elements of secondary structure are well maintained within this family. The hydrophobic core of the protein is also well conserved, consistently involving residues 27 – 28, 32, 35 – 37, 40, 45, 48, 51, 60, 64, 67, 71, 81, 84 – 85, and 88 (*Bp*cytc numbering). Additional residues which are part of the hydrophobic core only in some of the structures are 24, 39, 52, 55, 61 – 63, 75 – 76, 83, and 89. The residues which are consistently in the hydrophobic core are, as expected, always hydrophobic and are either strictly conserved or conservatively substituted (with the exception of residue 51 which is deleted in one of the two *B. halodurans* cytochromes). On the other hand, the residues belonging to the core in only some structures show non-conservative substitutions in some sequences. With only a few exceptions, for the cytochromes in which there are hydrophilic residues at these positions, these residues are solvent accessible, which constitutes an indication of the reliability of the results of the modeling approach followed in this work. The residues that are absolutely conserved, or conservatively substituted, throughout the sequence are hydrophobic core residues or other hydrophobic residues contacting them. Residues found in positions 59, 66, and 80 always possess hydrophilic side chains (Glu or Lys, Asn, His, or Lys, Glu, Asp or Lys, respectively), and are consistently solvent exposed, a fact which suggests that they may play a role in molecular recognition processes. This could be particularly the case for residues 59 and 66 lying on the protein patch which is also the face surrounding the heme propionates. The observed reversals in the charge of the side chain may be correlated to variations of the amino acid type in the interaction region of the partner.

It is known that surface electrostatic potentials are often important in guiding the interaction between cytochromes and their partners (see ref. [15], and references cited therein). The analysis of the surface potentials in the present family of sequences reveals a significant variability from one sequence to another, in agreement with the high variability of amino acid type for the solvent-exposed residues. In addition, very marked differences are observed between the *Heliobacterium* and *Heliobacillus* sequences on one side and the *Bacillus* sequences on the other side, probably due to the various insertions occurring in the *Heliobacteria* sequences. As a consequence, we restrict the following analysis of surface electrostatic potentials to the family of *Bacillus* cytochromes. Figure 6 shows two views

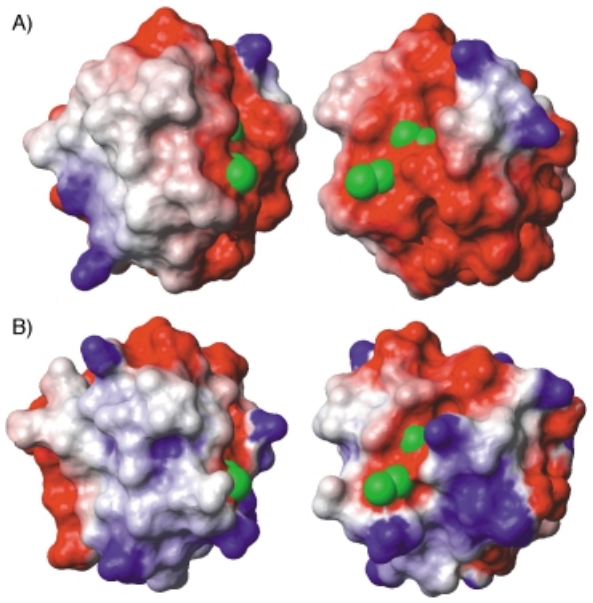

Figure 6. *Surface electrostatic potentials of cytochrome* c_{553} *from* Bacillus pasteurii *(A) and cytochrome* c_{551} *(cccB) from* Bacillus subtilis *(B). The orientation of the protein is such that the heme plane is normal to the page, and the Fe–Nε2(His) bond is oriented towards the top of the page. In the left column the exposed side of the heme where the thiomethyl 4 and methyl 5 moieties lie faces the observer. In the right column the protein is rotated by 90° around the Fe–Nε2(His) bond, so that the exposed side of the heme where the propionates lie faces the observer. The oxygen atoms of the heme carboxylates are shown as spheres of 2 Å radius (in green). This figure was prepared by using the default values for electrostatic potential calculations of the program MOLMOL [56] and by imposing a charge of − 0.5 to each oxygen of the propionates.*

of the two most diverse potential surfaces, that is, that of *Bp*cytc and that of *B. subtilis* c_{551} (cccB). It can be seen (Figure 6 A) that the protein patch surrounding the heme propionates always shows a negative surface potential, while the regions further from the heme display a large charge variability (from very negative to very positive). The other *Bacillus* sequences display intermediate properties. Rotation of the protein around the Fe–Nε2(His) bond by approximately 90° reveals that a large nonpolar patch is conservatively present, surrounding the other consistently exposed heme edge (Figure 6 B). This patch comprises residues 34 – 35, 44 – 46, 69 – 72, and 75. Some minor variations of the potential in the regions surrounding the patch in the two structures are apparent, but overall the gross surface features of this side of protein are conserved. This suggests that such a nonpolar patch, which also features backbone mobility in the long time scales (as discussed below), essential to protein recognition processes, is a likely site for the physiological electron-transfer processes involving *Bacillus c*-type cytochromes. The present suggestion is also consistent with the fact that *Bp*cytc is thought to be anchored to the cell membrane of this Gram-positive bacterium, with the soluble globular domain protruding into the extracellular space[3] and therefore exposed to the typical high ionic strength medium used for the growth of *B. pasteurii*.[2] Such high dielectric medium would reduce the efficiency of protein recognition through electrostatic interactions. In addition, it should be mentioned that *Bp*cytc, although isolated from a mesophilic organism, is

thermostable[8] and features a high sequence homology with *Bacillus* PS3, a thermophile. Hydrophobic protein – protein interactions are favored at high temperatures as compared to electrostatic interactions, and this concept has been utilized to rationalize the large hydrophobic patch present on the surface of the terminal cytochrome *c* oxidase in *Thermus thermophilus*[17] as well as the partner cytochrome c_{552}.[18] These observations further strengthen the hypothesis that *Bp*cyt*c* is the electron donor to the terminal oxidase also in *B. pasteurii*[3] and support the present proposition for the interaction surface involved in the process.

All modeled cytochrome structures display a hydrogen bond between the Hδ1 proton of the side chain of axial His 36 and the backbone carbonyl oxygen atom of Pro 46. This structural feature corresponds to the conserved hydrogen bond between the Hδ1 proton of His 18 and the carbonyl oxygen atom of Pro 30 observed in all mitochondrial cytochrome *c* proteins.[15] Its role is believed to be that of ensuring the proper orientation of the plane of the axial His in solution. Finally, the solvent accessibility of the metal site is also a conserved feature within the present protein family. The solvent accessible area of the heme indeed varies from 110 – 190 Å^2, which is only slightly larger than the range observed for the experimental family of conformers representing the solution structure of *Bp*cyt*c* (135 – 160 Å^2); this suggests that the low redox potential observed for *Bp*cyt*c* may be a conserved feature of this class of *c*-type cytochromes from Gram-positive bacteria.[8, 9]

Analysis of the mobility of *Bacillus pasteurii* and comparison with other oxidized *c*-type cytochromes

*Bp*cyt*c* shows relatively constant S^2 values along the sequence (Figure 4), with two more rigid regions being the stretches 36 – 38 and 67 – 69, which are close to the heme binding site. Enhanced mobility is instead observed for the C-terminal tail, and, to a smaller extent, for the N terminus, plus residues 54 and 73 – 74. The two termini can be fitted satisfactorily only if a τ_e value for their internal motions is explicitly included in the fittings. This observation suggests that the N- and C-terminal residues are involved in motions at the slow end of the range accessible through the model-free analysis. Interestingly the S^2 value of the NH vector of the side chain of Trp 87 is as low as 0.83, which suggests that it is somewhat mobile in solution. Indeed, this side chain is quite solvent-exposed, and the Hε1 proton is readily exchanged in D_2O. The solvent exchangeability of backbone amide protons varies quite significantly along the polypeptide chain, showing a good degree of correlation with the presence of secondary structure elements (besides helix 2, where all backbone amide protons are exchanged in D_2O in less than 10 minutes). Residues at the N termini of helices have a tendency to exchange more readily than those at the C termini, possibly because of the insertion of the C-terminal helix 5 between helices 1 and 4.

Overall, the most striking dynamic feature of *Bp*cyt*c* is the existence of conformational exchange processes in the μs – ms time scale, as independently highlighted by both the Modelfree fittings and the ν_{CPMG} dependencies for residues 33, 39 – 41, and 44. Residue 33 immediately follows the first heme-bound Cys, and is the N-terminal residue of the short α helix 2, which extends up to the axial His. Residues 39 – 41 are at the N-terminal region of a loop following the axial His and are in contact with Ile 33 (through Leu 40 and the backbone of Asp 39). Ser 44, on the other hand, is in the center of the same loop and not in direct contact with any of the above-mentioned residues. Interestingly, only Ser 44 displays a deviation between the conformation observed in solution and in the crystal larger than the average for the rest of the protein. The conformational equilibrium which involves residues 33 and 39 – 41 could play a role in modulating the interactions of the polypeptide chain with the heme cofactor, for example, by affecting its doming. Indeed, the residues located between the two heme-bound Cys residues (Ile 33 and Ser 34 in the present case) are essential in this respect,[19] a fact that might be relevant for biological function. The protein region formed by residues 33 and 39 – 41 lies above the heme, extending in the direction opposite to the propionates, and is in close contact with residues 47 – 48, which lie in the vicinity of the heme propionates. It is noteworthy that residue 44 is located in the central region of the exposed hydrophobic patch involving the side of the heme where thiomethyl 4 and methyl 5 lie (Figure 6). Ile 33 is also close to this interface. This suggests a possible influence of the conformational equilibria detected here in the recognition of and/or interaction with the partner, as discussed above. It is also possible that residues 33 and 39 – 41 play a role in the interaction with the partner by making contacts in the region close to the axial His. This would be particularly important in the case of *Bp*cyt*c* interacting with its partner through binding in a cleft.

Mobility studies on oxidized, paramagnetic, *c*-type cytochromes are available only for yeast iso-1 cytochrome *c*, for which ^{15}N relaxation data have been used to probe both motions in the sub-ns time scale[20] and in the μs – ms time scale.[21] The Modelfree analysis for oxidized yeast iso-1 cytochrome *c* shows that fast internal motions (as described by the order parameter S^2) vary significantly along the sequence, the residues in loops between helices being significantly more flexible than those in an element of regular secondary structure.[20] This mobility pattern is different from that observed in *Bp*cyt*c*, for which the pattern of S^2 values is quite uniform throughout the sequence. Additionally, several residues in oxidized yeast iso-1 cytochrome *c*, belonging mainly to helical regions, experience motions in the μs – ms time scale (that is, conformational exchange processes),[21] while none of the corresponding residues in *Bp*cyt*c* (aligned from the point of view of the sequence, by using the CLUSTAL program,[22] or of the structure, by using the DALI program[23]) shows any exchange contribution to R_2. Moreover, the 39 – 41 loop region displays conformational exchange in *Bp*cyt*c*, while no conformational exchange processes are detected for the corresponding region 21 – 25 in yeast iso-1 cytochrome *c*. The same holds for Ile 33 in *Bp*cyt*c* and the corresponding Leu 15 in the yeast cytochrome. In summary, no correlation exists for the mobility patterns determined by the model-free analysis in the two cytochromes. On the other hand, in both *Bp*cyt*c* and yeast iso-1 cytochrome *c*[24] the pattern of backbone amide exchange rates (or protection factors), which

sample time scales for protein flexibility of the order of a minute or longer, correlates broadly with the pattern of secondary structure elements. This confirms that hydrogen exchange rates are mainly influenced by local structure.[25, 26]

The above mobility studies may be significant with respect to the observation that eukaryotic cytochrome *c* proteins, such as yeast iso-1, give rise to a so-called alkaline form at high pH values.[27] The formation of the alkaline form involves the detachment of the axial Met and its replacement by another side-chain ligand (most likely a Lys).[27–30] Indeed, several residues in the proximity of Met 80 in the yeast-oxidized cytochrome experience conformational averaging,[21] which indicates that different conformers with comparable energy are in equilibrium. This observation suggests that there may be some structural strain in this region, which prevents the polypeptide from folding into a single well-defined minimum. In *Bp*cyt*c* the region around axial Met 71 does not display any conformational exchange. Residues in this region have order parameter values ranging from above to below the average of the entire protein. The absence of a global trend suggests that local mobility is dominated only by residue-specific interactions, analogously to the other parts of the sequence, and thus that there are no structural strains in this protein region. It should also be considered that the absence of Lys residues in *Bp*cyt*c* in positions corresponding to those of the most likely candidates for binding to the iron in the alkaline form of yeast iso-1 cytochrome *c* proteins certainly contributes to the relative stability of the axial Met ligation in *Bp*cyt*c* at high pH value, as evidenced by the pH dependence (up to pH 10) of the signals in the ^{1}H NMR spectrum.[8] However, Lys 77 is within reach of the iron center, and in the presence of protein flexibility could have, in principle, also given rise to the alkaline form for *Bp*cyt*c*. This stability of *Bp*cyt*c* at high pH values is consistent with the fact that the extracytoplasmatic region of cell, where *Bp*cyt*c* is likely to be localized, is exposed to the high pH ($>$ 9.2) medium in which *Bacillus pasteurii* optimally grows.

Conclusion

The solution structure of *Bp*cyt*c* was solved by NMR spectroscopy at pH 7.5 and compared with the crystal structure at pH 5.1. It was shown that the overall folding of the protein is extremely well maintained when passing from the crystal to solution. A possible pH-dependent conformational change of propionate-7 was suggested. The structures of homologous *c*-type cytochromes from another six *Bacillus* and two *Heliobacterium* species were modeled and analyzed in order to single out conserved features which might have a functional role or a role in stabilizing the protein fold. It was found that the hydrophobic core is well defined and conserved within this family. In addition, two conserved solvent-accessible patches located in the surroundings of the heme moiety were identified. These regions may play a role in the biological processes of recognition and interaction with the partners (for example, the aa_3 terminal oxidase).

The mobility of the protein backbone was characterized through ^{15}N relaxation measurements. The protein displays uniform mobility, with the exception of a few residues and of the N- and C-terminal tails, which have S^2 values lower than the average, and of two more rigid stretches close in sequence to the axial ligands. The most prominent dynamic feature of this protein is the presence of conformational exchange processes for residues 33, 39 – 41, and 44, both in the μs – ms and longer (minutes) time scales. These residues are close to, or part of, conserved nonpolar exposed patches around the heme moiety, identified from the comparative modeling. This suggests that the conformational process observed may have some influence on *Bp*cyt*c* physiological interactions. A comparison of the dynamic behavior of *Bp*cyt*c* and of oxidized mitochondrial yeast iso-1 cytochrome *c* revealed that these two proteins have largely different characteristics in the regions around the axial Met ligand. This difference may be indicative of the presence of some structural strain in yeast iso-1 cytochrome *c*, which, in turn, would at least partly explain the fact that in the yeast protein the axial Met gets displaced at high pH value, while this does not happen for *Bp*cyt*c*.

Materials and Methods

Preparation of recombinant *Bacillus pasteurii* cytochrome *c* samples: The gene coding *Bp*cyt*c* has been amplified by polymerase chain reaction (PCR) from genomic DNA and sequenced by standard dideoxy sequencing methods (Genbank entry AJ318066). The sequence coding the soluble portion (residues 22 – 92) of *Bp*cyt*c* was cloned between the NcoI and XhoI sites of the expression vector pET20b (Novagen), resulting in pAT1. In this way, the pelB leader sequence is fused to the *Bp*cyt*c*, directing the newly synthesized polypeptide to the periplasmic space, where heme attachment occurs.[31] The pelB leader peptide is removed in the process of heme attachment. *E. coli* strain DH5α was used for subcloning and the BL21(DE3) derivative strain C41[32] was used for expression of *Bp*cyt*c*.

For protein expression, C41 competent cells were cotransformed with pAT1 and pEC86, kindly provided by Dr. Thöny Meyer. Plasmid pEC86 encodes the complete *ccm*A-H gene cluster from *E. coli*, which is needed for incorporation of the heme moiety in the expressed cytochrome, under the control of the *tet* promoter,[33] and carries a marker for chloramphenicol. Cells harboring both plasmids were selected for their ability to grow on 2xYT plates containing ampicillin (100 μg mL^{-1}) and chloramphenicol (50 μg mL^{-1}). Rich media cultures were performed with the 2xYT culture medium supplemented with antibiotics. Cultures were usually incubated overnight with shaking at 37 °C, then induced with isopropyl-β-D-thiogalactopyranoside (IPTG; 0.5 mM) and finally harvested after 24 h by centrifugation at 4 °C. For preparation of ^{15}N-enriched samples, cultures were grown on minimal medium consisting of M9 salts supplemented with $MgSO_4$, trace metal, and vitamin solutions, and also containing δ-aminolevulinic acid (0.1 mM) and 2-mercaptoethanesulfonic acid (1 mM). The nitrogen source was $(^{15}NH_4)_2SO_4$ (1.2 g L^{-1}) and the carbon source was glucose (4 g L^{-1}).

The protein was extracted from *E. coli* cells by isolating the periplasmic contents, which were loaded directly onto a DE52 column equilibrated with 2 mM sodium ascorbate and 0.1 M tris(hydroxymethyl)aminomethane (Tris)-HCl at pH 8.0 and eluted with a 0 – 300 mM NaCl gradient in the same buffer. Red fractions were concentrated, applied to a gel filtration column, and then eluted with 15 mM sodium phosphate buffer (pH 7.5).

Samples for NMR spectroscopy were at a concentration of 1.0 – 1.5 mM protein in the elution buffer of the gel filtration column. Immediately prior to NMR experiments on oxidized *Bp*cytc, O_2 was bubbled into the sample solution for 15 – 30 minutes. Reduction of *Bp*cytc was achieved by addition of an equimolar amount of sodium dithionite to the solution under anaerobic conditions.

NMR structural data acquisition and processing: NMR experiments were performed on Bruker Avance spectrometers operating at 400, 600, and 800 MHz proton Larmor frequencies. Data acquisition and processing was performed by using a standard Bruker software package (XWINNMR).

1D NOE spectra on hyperfine-shifted signals were recorded as previously described[34] by using repetition and irradiation times of 75 ms. Two-dimensional (2D) ^{1}H homonuclear TOCSY[35] and NOESY[36] experiments were acquired at 288, 293, and 298 K, in phase-sensitive TPPI mode[37] by using unlabeled *Bp*cytc. Water suppression was achieved with the WATERGATE pulse sequence[38] or, alternatively, through solvent presaturation in spectra tuned to detect connectivities involving hyperfine-shifted signals. The values for the spin-lock time in TOCSY experiments varied from 15 – 50 ms, and the mixing time in NOESY spectra from 55 – 100 ms. 2D ^{1}H-^{15}N HSQC spectra[39] were performed at 278, 288, 298, and 308 K. A three-dimensional (3D) WATERGATE NOESY – HSQC spectrum[40] was recorded with a mixing time of 100 ms. A 3D HNHA experiment[41] was collected in order to determine the $^3J_{H\alpha-NH}$ coupling constants. For reduced *Bp*cytc, a 2D ^{1}H-^{15}N HSQC spectrum[39] was recorded at 298 K at 600 MHz. For all heteronuclear spectra, ^{15}N decoupling was performed during acquisition. Other experimental details are given in Table 3.

Peak assignment and structural constraints: The program XEASY[42] was used for spectral analysis and for NOESY cross-peak integration. The assignment of the ^{1}H and ^{15}N NMR signals of protein residues not directly coordinated to Fe^{III} was achieved with standard procedures[43] through the analysis of 2D and 3D homo- and heteronuclear spectra. The assignment of the signals of protons of the heme moiety or of the coordinating iron residues was obtained from 1D NOE spectra, acquired upon saturation of the hyperfine-shifted signals, and from paramagnetic 2D TOCSY and NOESY experiments, according to previously published procedures.[44] The latter assignment is in complete agreement with that already reported.[8]

The volumes of the NOESY cross-peaks assigned for the oxidized form were transformed into ^{1}H – ^{1}H upper distance limits by using the program CALIBA.[45] Standard calibration was performed by using backbone, side-chain, and methyl classes, with a $(1/r)^6$ distance dependence. 1D NOEs were calibrated separately by classifying them as either weak or strong. In the first case, upper distance limits of 5.5 and 6.5 Å were used for single protons and methyl groups, respectively. In the case of strong NOEs, an upper distance limit of 3.5 Å was used for both proton and methyl classes. Stereospecific assignments (33 in total) were obtained by using the program GLOMSA.[45]

38 ϕ dihedral angle constraints were obtained, either from the analysis of $^3J_{HN-H\alpha}$ coupling constants obtained from the HNHA spectrum[41] or from the relative intensities of the intra- and interresidue Hα – NH NOESY cross-peaks measured in the NOESY spectra.[46] 38 ψ dihedral angle constraints were obtained from the relative intensities of the intra- and interresidue Hα – NH NOESY cross-peaks.[46]

For the reduced protein, resonance assignments of the ^{1}H – ^{15}N HSQC peaks were performed following a procedure reported previously.[47], [48] This is based on the temperature dependence of the observed shifts of the signals in the already assigned oxidized species (which, for residues not bound to the heme moiety, is entirely due to the pseudocontact contribution). At infinite temperature, the hyperfine (paramagnetic) contribution to the shifts extrapolate nothing, and the observed chemical shift is thus the diamagnetic shift. By raising the temperature, the peaks thus shift towards the corresponding position in the diamagnetic reduced form. This permits, by comparison with the ^{1}H – ^{15}N HSQC spectrum of the reduced protein, the unambiguous assignment of an initial set of resonances for the reduced cytochrome. These chemical shifts are used to calculate the PCS values by taking the difference between the corresponding shifts in the oxidized and reduced form.[11] A magnetic susceptibility anisotropy tensor is then calculated based on these data. The assignment is then extended to other nuclei, whose shifts extrapolated to crowded regions of the spectrum. The corresponding diamagnetic shifts were discriminated by comparing the experimental and predicted hyperfine shifts.

PCS values were used in structure calculations following standard protocols.[49] The ^{15}N PCS values were not used for calculations, because it is known that the variation of ^{15}N chemical shifts between the oxidized and reduced forms of cytochromes depends not only on the pseudocontact contribution, but also on other, still uncharacterized, factors.[50, 51]

Structure calculation and refinement: 400 random structures were generated by using the PSEUDYANA module[52] of the program DYANA[53] and annealed in 8000 steps by using the standard DYANA annealing protocol. All constraints were used simultaneously from the beginning of the annealing procedure. The porphyrin cofactor was included in the structure calculations as previously described.[52, 54] The axial and rhombic magnetic susceptibility anisotropies were given as input values and kept constant during the minimization. The 30 conformers with the lowest target function were then used to determine again the anisotropy parameters of the magnetic susceptibility tensors with the program FANTASIA,[49] for the next calculation round. Few cycles of 1) structure calculation,

Table 3. *Acquisition parameters for multidimensional NMR experiments performed on oxidized recombinant cytochrome* c_{553} *from* Bacillus pasteurii.

Experiment	Acquired data matrix (nucleus)			Spectral width (ppm)			Ref.
	T_1	T_2	T_3	F_1	F_2	F_3	
[^{1}H – ^{1}H] NOESY[a]	1024(^{1}H)	4096(^{1}H)		20	20		[32]
[^{1}H – ^{1}H] NOESY[a]	1024(^{1}H)	2048(^{1}H)		70	70		[32]
[^{1}H – ^{1}H] TOCSY[b]	1024(^{1}H)	4096(^{1}H)		20	20		[31]
[^{1}H – ^{1}H] TOCSY[b]	1024(^{1}H)	2048(^{1}H)		70	70		[31]
^{1}H – ^{15}N HSQC[a],[b]	256(^{15}N)	1024(^{1}H)		38	10		[35]
^{1}H – ^{15}N NOESY – HSQC[a]	320(^{1}H)	64(^{15}N)	1024(^{1}H)	10	40	7	[36]
HNHA[a]	256(^{1}H)	48(^{15}N)	1024(^{1}H)	15	38	7	[37]

[a] Data acquired on a 800 MHz spectrometer. [b] Data acquired on a 600 MHz spectrometer.

2) recalibration by using CALIBA, and 3) FANTASIA fittings yielded an ensemble of structures satisfying all experimental constraints. The 30 conformers with the lowest total target function were selected as the PSEUDYANA family.

The PSEUDYANA family was then subjected to restrained energy minimization with the SANDERPC[49] module of the AMBER program package,[55] to yield the PSEUDOREM family. Calculations were performed as previously described.[11, 49] In addition, the mean structure of the PSEUDOREM family was calculated by using the program MOLMOL,[56] and the restrained energy was minimized.

The quality of the structure was checked with the PROCHECK,[10, 57] WHATIF,[58, 59] and AQUA programs.[57] The coordinates of the PSEUDOREM family and of the minimized mean structure have been deposited in the PDB (PDB codes: 1 K3H and 1 K3G).

NMR mobility data acquisition and processing: The NMR experiments for determination of ^{15}N longitudinal and transversal relaxation rates and $^{1}H-^{15}N$ NOE were recorded at 298 K at 400 MHz and 600 MHz. The ^{15}N R_1 and R_2 relaxation rates and $^{1}H-^{15}N$ NOEs were measured by using, respectively, inversion-recovery,[60] CPMG,[61] and saturation[62][63] experiments. In all experiments, solvent suppression was achieved with the water flip-back scheme, which avoids water saturation.[64]

The dependence of the ^{15}N R_2 relaxation rates on the CPMG repetition rate was probed by measuring these rates at 400 MHz with six different CPMG frequencies (ν_{CPMG}): 435, 500, 588, 714, 909, and 1111 Hz. The data measured with the latter CPMG frequency were used in Modelfree calculations. The present experimental conditions, together with the choice of magnetic field and the small size of the protein ($\tau_r \approx 3.6$ ns), are such that off-resonance effects on the present measurements are negligible.[65]

Backbone amide exchange data were obtained by dissolving a lyophilized ^{15}N-labeled sample in buffered D_2O and acquiring HSQC spectra at regular time intervals over a period of three days. The first point was acquired ten minutes after sample preparation.

Analysis of ^{15}N relaxation data: R_1 and R_2 relaxation rates were obtained by fitting the cross-peak volumes (I), measured as a function of the relaxation delay, to a single exponential decay by using the Levenberg – Marquardt algorithm,[66, 67] as described in the literature.[68] Uncertainties had been evaluated by using a Monte Carlo approach.[68–70] Heteronuclear NOE values were calculated as the ratio of peak volumes in spectra recorded with and without ^{1}H saturation. The heteronuclear NOE values and their errors were estimated by calculating the mean and the standard error of the values obtained from the various data sets available.

^{15}N relaxation data were analyzed in terms of the model-free formalism[71] through the Modelfree program, version 4.0, following the reported protocol.[13] The amide proton – nitrogen backbone distance was taken to be 1.02 Å, and the ^{15}N chemical shift anisotropy (CSA) was taken to be $\delta = -172$.[72] For the tryptophan side chain NH moiety, the CSA was taken to be 89 ppm.[73] Model selection for the Modelfree calculations was performed according to the procedure previously described.[13] In the last stage of calculations τ_m was optimized together with all other Modelfree parameters.

The contribution from conformational exchange to ^{15}N R_2 shows a dependence on ν_{CPMG}, which can be approximated by Equation (1), where p_A and p_B are the populations of the two states a and b between which the exchange process occurs, $\delta\Omega$ is the difference in

$$R_{ex} = p_A p_B \delta\Omega^2 \tau_{ex}\left[1 - 4\nu_{CPMG}\tau_{ex}\tanh\left(\frac{1}{4\nu_{CPMG}\tau_{ex}}\right)\right] \quad (1)$$

the chemical shift of the ^{15}N in the two states, and τ_{ex} is the time constant for the exchange process.[74, 75]

R_2 relaxation rate values were obtained by adding the R_{ex} contribution of Equation (1) to the R_2 value in the absence of conformational exchange; these values were calculated for each residue through the model-free formalism from experimental R_1 and heteronuclear NOE data.[75, 76] This calculated sum was used to least square fit the experimental dependence of the R_2 relaxation rates, by using τ_{ex} and the product $p_A p_B \delta\Omega^2$ as adjustable parameters. Backbone amide exchange data were fitted to an exponential decay.

Sequence analysis and structural modeling: Sequences similar to that of *Bp*cyt*c* were searched for in the SwissProt and TrEMBL (for protein sequences translated from genomic codes) databases by using sequence similarity criteria. The sequence of *B. licheniformis* cytochrome *c* was taken from the literature.[3] The program used to perform the search was the "Basic Local Alignment Search Tool" (BLAST) program.[14] The pattern-hit initiated BLAST (Phi-BLAST) version of the program[14] was also used, in order to identify sequences more distantly related to *Bp*cyt*c*, but still showing the CXXCH heme-binding pattern.

The sequences identified were selected and organized by performing alignments through the ClustalX program,[22] installed on local machines. The service at: http://bioweb.pasteur.fr[77] was used to draw phylogenetic trees. Modeling was performed with the standard parameters of the model-default option of the program MODELLER v. 4.0,[78] based on the alignment obtained for all sequences from ClustalX. No further refinement was carried out. Models were analyzed with the program PROSA.[16]

We thank Dr. L. Thöny Meyer for providing us with pEC86. Financial support from the European Union (contracts ERB-FMRX-CT98 – 0218 and HPRI-1999-CT-50006) is gratefully acknowledged. This work has also been partly supported by MURST (COFIN99).

[1] R. A. Scott, A. G. Mauk, *Cytochrome c. A multidisciplinary approach*, University Science Books, Sausalito, CA, **1996**.
[2] S. Benini, S. Ciurli, W. Rypniewski, K. S. Wilson, *Proteins Struct. Funct. Genet.* **1997**, *28*, 280 – 285.
[3] I. H. Vanderberghe, S. Ciurli, S. Benini, J. Van Beeumen, *Biochem. Biophys. Res. Commun.* **1999**, *264*, 380 – 387.
[4] W. R. Wiley, J. L. Stokes, *J. Bacteriol.* **1962**, *84*, 730 – 734.
[5] W. R. Wiley, J. L. Stokes, *J. Bacteriol.* **1963**, *86*, 1152 – 1156.
[6] A. D. Larson, R. E. Kallio, *J. Bacteriol.* **1954**, *68*, 67 – 73.
[7] G. H. Bornside, R. E. Kallio, *J. Bacteriol.* **1956**, *71*, 627 – 634.
[8] S. Benini, M. Borsari, S. Ciurli, A. Dikiy, M. Lamborghini, *J. Biol. Inorg. Chem.* **1998**, *3*, 371 – 382.
[9] S. Benini, W. Rypniewski, K. S. Wilson, J. Van Beeumen, S. Ciurli, *Biochemistry* **2000**, *39*, 13 115 – 13 126.
[10] R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **1993**, *26*, 283 – 291.
[11] L. Banci, I. Bertini, H. B. Gray, C. Luchinat, T. Reddig, A. Rosato, P. Turano, *Biochemistry* **1997**, *36*, 9867 – 9877.
[12] L. Banci, I. Bertini, K. L. Bren, H. B. Gray, P. Sompornpisut, P. Turano, *Biochemistry* **1997**, *36*, 8992 – 9001.
[13] M. A. Mandel, M. Akke, A. G. Palmer III, *J. Mol. Biol.* **1995**, *246*, 144 – 163.
[14] F. S. Altschul, T. L. Madden, A. Schaeffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, *Nucleic Acids Res.* **1997**, *25*, 3389 – 3402.
[15] L. Banci, I. Bertini, A. Rosato, G. Varani, *J. Biol. Inorg. Chem.* **1999**, *4*, 824 – 837.
[16] M. J. Sippl, *Proteins Struct. Funct. Genet.* **1993**, *17*, 355 – 362.
[17] T. Soulimane, G. Buse, G. P. Bourenkov, H. D. Bartunik, R. Huber, M. E. Than, *EMBO J.* **2000**, *19*, 1766 – 1776.
[18] M. E. Than, P. Hof, R. Huber, G. P. Bourenkov, H. D. Bartunik, G. Buse, T. Soulimane, *J. Mol. Biol.* **1997**, *271*, 629 – 644.
[19] W. Jentzen, J.-G. Ma, J. A. Shelnutt, *Biophys. J.* **1998**, *74*, 753 – 763.

[20] J. S. Fetrow, S. M. Baxter, *Biochemistry* **1999**, *38*, 4480 – 4492.
[21] P. B. Barker, I. Bertini, R. Del Conte, S. J. Ferguson, P. Hajieva, P. Tomlinson, P. Turano, M. S. Viezzoli, *Eur. J. Biochem.* **2001**, *268*, 4468 – 4476.
[22] J. D. Thompson, T. J. Gibson, F. Plewniak, F. Jeanmougin, D. G. Higgins, *Nucleic Acids Res.* **1997**, *24*, 4876 – 4882.
[23] L. Holm, C. Sander, *J. Mol. Biol.* **1993**, *233*, 123 – 138.
[24] S. M. Baxter, J. S. Fetrow, *Biochemistry* **1999**, *38*, 4493 – 4503.
[25] Y. W. Bai, T. R. Sosnick, L. Mayne, S. W. Englander, *Science* **1995**, *269*, 192 – 197.
[26] J. S. Milne, L. Mayne, H. Roder, A. J. Wand, S. W. Englander, *Protein Sci.* **1998**, *7*, 739 – 745.
[27] X. Hong, D. W. Dixon, *FEBS Lett.* **1989**, *246*, 105 – 108.
[28] P. B. Barker, A. G. Mauk, *J. Am. Chem. Soc.* **1992**, *114*, 3619 – 3624.
[29] J. C. Ferrer, J. G. Guillemette, R. Bogumil, S. C. Inglis, M. Smith, A. G. Mauk, *J. Am. Chem. Soc.* **1993**, *115*, 7507 – 7508.
[30] D. L. Brautigan, B. A. Feinberg, B. M. Hoffman, E. Margoliash, J. Peisach, W. E. Blumberg, *J. Biol. Chem.* **1977**, *252*, 574 – 582.
[31] M. Dudley Page, Y. Sambongi, S. J. Ferguson, *Trends Biochem. Sci.* **1998**, *23*, 103 – 108.
[32] B. Miroux, J. E. Walker, *J. Mol. Biol.* **1996**, *260*, 289 – 298.
[33] E. Arslan, H. Schulz, R. Zufferey, P. Kuenzler, L. Thoeny-Meyer, *Biochem. Biophys. Res. Commun.* **1998**, *251*, 744 – 747.
[34] L. Banci, I. Bertini, C. Luchinat, M. Piccioli, A. Scozzafava, P. Turano, *Inorg. Chem.* **1989**, *28*, 4650 – 4656.
[35] A. Bax, D. G. Davis, *J. Magn. Reson.* **1985**, *65*, 355 – 360.
[36] G. Wider, S. Macura, A. Kumar, R. R. Ernst, K. Wüthrich, *J. Magn. Reson.* **1984**, *56*, 207 – 234.
[37] D. Marion, K. Wüthrich, *Biochem. Biophys. Res. Commun.* **1983**, *113*, 967 – 974.
[38] V. Sklenar, M. Piotto, R. Leppik, V. Saudek, *J. Magn. Reson. Ser. A* **1993**, *102*, 241 – 245.
[39] G. Bodenhausen, D. J. Ruben, *Chem. Phys. Lett.* **1980**, *69*, 185 – 188.
[40] L. E. Kay, D. Marion, A. Bax, *J. Magn. Reson.* **1989**, *84*, 72 – 84.
[41] G. W. Vuister, A. Bax, *J. Am. Chem. Soc.* **1993**, *115*, 7772 – 7777.
[42] C. Eccles, P. Güntert, M. Billeter, K. Wüthrich, *J. Biomol. NMR* **1991**, *1*, 111 – 130.
[43] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, **1986**.
[44] L. Banci, I. Bertini, C. Luchinat, P. Turano in *The Porphyrin Handbook* (Eds.: K. M. Kadish, K. M. Smith, R. Guilard), Academic Press, San Diego, CA, **2000**, pp. 323 – 350.
[45] P. Güntert, W. Braun, K. Wüthrich, *J. Mol. Biol.* **1991**, *217*, 517 – 530.
[46] R. R. Gagné, S. Tsuda, M. X. Li, M. Chandra, L. B. Smillie, B. D. Sykes, *Protein Sci.* **1994**, *3*, 1961 – 1974.
[47] M. Allegrozzi, I. Bertini, M. B. L. Janik, Y.-M. Lee, G. Liu, C. Luchinat, *J. Am. Chem. Soc.* **2000**, *122*, 4154 – 4161.
[48] Z. Xia, B. D. Nguyen, G. N. La Mar, *J. Biomol. NMR* **2000**, *17*, 167 – 174.
[49] L. Banci, I. Bertini, G. Gori Savellini, A. Romagnoli, P. Turano, M. A. Cremonini, C. Luchinat, H. B. Gray, *Proteins Struct. Funct. Genet.* **1997**, *29*, 68 – 76.
[50] J. Boyd, C. M. Dobson, A. S. Morar, R. J. P. Williams, G. J. Pielak, *J. Am. Chem. Soc.* **1999**, *121*, 9247 – 9248.
[51] I. Bertini, C. Luchinat, P. Turano, *J. Biol. Inorg. Chem.* **2000**, *5*, 761 – 764.
[52] L. Banci, I. Bertini, M. A. Cremonini, G. Gori Savellini, C. Luchinat, K. Wüthrich, P. Güntert, *J. Biomol. NMR* **1998**, *12*, 553 – 557.
[53] P. Güntert, C. Mumenthaler, K. Wüthrich, *J. Mol. Biol.* **1997**, *273*, 283 – 298.
[54] L. Banci, I. Bertini, F. Ferroni, A. Rosato, *Eur. J. Biochem.* **1997**, *249*, 270 – 279.
[55] D. A. Case, D. A. Pearlman, J. W. Caldwell, T. E. Cheatham, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsui, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. K. Weiner, P. A. Kollman, *AMBER 6*, University of California, San Francisco, **1999**.
[56] R. Koradi, M. Billeter, K. Wüthrich, *J. Mol. Graphics* **1996**, *14*, 51 – 55.
[57] R. A. Laskowski, J. A. C. Rullmann, M. W. MacArthur, R. Kaptein, J. M. Thornton, *J. Biomol. NMR* **1996**, *8*, 477 – 486.
[58] G. Vriend, *J. Mol. Graphics* **1990**, *8*, 52 – 56.
[59] R. W. W. Hooft, G. Vriend, C. Sander, E. E. Abola, *Nature* **1996**, *381*, 272 – 272.
[60] J. W. Peng, G. Wagner, *J. Magn. Reson.* **1992**, *98*, 308 – 332.
[61] L. E. Kay, D. A. Torchia, A. Bax, *Biochemistry* **1989**, *28*, 8972 – 8979.
[62] L. E. Kay, L. K. Nicholson, F. Delaglio, A. Bax, D. A. Torchia, *J. Magn. Reson.* **1992**, *97*, 359 – 375.
[63] G. Barbato, M. Ikura, L. E. Kay, R. W. Pastor, A. Bax, *Biochemistry* **1992**, *31*, 5269 – 5278.
[64] S. Grzesiek, A. Bax, *J. Am. Chem. Soc.* **1993**, *115*, 12 593 – 12 594.
[65] D. M. Korzhnev, E. V. Tischenko, A. S. Arseniev, *J. Biomol. NMR* **2000**, *17*, 231 – 237.
[66] D. W. Marquardt, *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431 – 441.
[67] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, *Numerical Recipes in C—The Art of Scientific Computing*, Cambridge University Press, New York, **1988**.
[68] A. G. Palmer III, M. Rance, P. E. Wright, *J. Am. Chem. Soc.* **1991**, *113*, 4371 – 4380.
[69] J. W. Peng, G. Wagner, *Biochemistry* **1992**, *31*, 8571 – 8586.
[70] S. Zinn-Justin, P. Berthault, M. Guenneugues, H. Desvaux, *J. Biomol. NMR* **1997**, *10*, 363 – 372.
[71] G. Lipari, A. Szabo, *J. Am. Chem. Soc.* **1982**, *104*, 4546 – 4559.
[72] R. Ishima, D. A. Torchia, *Nat. Struct. Biol.* **2000**, *7*, 740 – 743.
[73] T. A. Cross, S. J. Opella, *J. Am. Chem. Soc.* **1983**, *105*, 306 – 308.
[74] Z. Luz, S. Meiboom, *J. Chem. Phys.* **1963**, *39*, 370.
[75] F. A. Mulder, P. J. Van Tilborg, R. Kaptein, R. Boelens, *J. Biomol. NMR* **1999**, *13*, 275 – 288.
[76] M. Akke, A. G. Palmer III, *J. Am. Chem. Soc.* **1996**, *118*, 911 – 912.
[77] J. Felsenstein, *Cladistics* **1989**, *5*, 164 – 166.
[78] A. Sali, T. L. Blundell, *J. Mol. Biol.* **1993**, *234*, 779 – 815.

Received: July 12, 2001 [F 264]