

Refinement of the three-dimensional structures of cytochrome c_3 from *Desulfovibrio vulgaris* Hildenborough at 1.67 Å resolution and from *Desulfovibrio desulfuricans* ATCC 27774 at 1.6 Å resolution

P. Simões^a, P.M. Matias^a, J. Morais^a, K. Wilson^c, Z. Dauter^c, M.A. Carrondo^{a,b,*}

^a Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Apartado 127, 2780 Oeiras, Portugal

^b Instituto Superior Técnico, Universidade Técnica de Lisboa, Av. Rovisco Pais, 1000 Lisbon, Portugal

^c EMBL Hamburg, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany

Received 28 April 1997; revised 26 June 1997; accepted 31 July 1997

Abstract

The three-dimensional X-ray structures of cytochrome c_3 from *Desulfovibrio vulgaris* Hildenborough and from *Desulfovibrio desulfuricans* ATCC 27774 were previously determined at 1.9 and 1.75 Å resolution, respectively. More recently, higher resolution data were collected (at 1.67 and 1.6 Å respectively) using synchrotron radiation. The refinement of the previously determined three-dimensional structures using the new data resulted in more accurate structural models, with no significant changes of the initial structures. In cytochrome c_3 from *D. vulgaris* Hildenborough, the R -factor was lowered from 19.6 to 15.3% using SHELXL-93, complemented with inspection and correction of the model relative to the electron density. This cytochrome crystallises in space group $P6_1$ with $a=77.3$, $b=77.3$, $c=77.1$ Å, $Z=12$. Cytochrome c_3 from *D. desulfuricans* ATCC 27774 crystallises in space group $P6_122$ with $a=62.71$, $b=62.71$, $c=111.09$ Å, $Z=12$. The R -factor was lowered from 17.8 to 16.6% using the same refinement procedure. In this cytochrome the 71–74 loop region was rearranged with no evidence of the previously found disorder, and disorder models were introduced in the terminal regions of residues serine 84 and lysine 90. The resulting higher-resolution structural models for both cytochromes are analyzed and compared with those previously obtained. © 1998 Elsevier Science S.A. All rights reserved.

Keywords: Three-dimensional structure; Cytochrome c_3

1. Introduction

Sulfate- and sulfur-reducing bacteria contain a rich array of cytochromes c ranging from molecules with a single heme to others having a total of sixteen heme groups [1]. The understanding of the function and mechanism of action of these unusual cytochromes is becoming increasingly important because of the unique implications of these bacteria in biological energy conservation and environmental problems linked to sulfate and nitrate reduction. The best studied multiheme is the tetraheme cytochrome c_3 which is considered to be a co-factor of hydrogenase. A role in energy conversion of the bacteria which are capable of using di-hydrogen as an energy source has been proposed for this cytochrome [2].

Cytochromes c_3 have unusually low and pH dependent redox potentials, in the range -120 to -400 mV, and there is evidence of fast intramolecular electron transfer between

the four heme irons [3]. This small protein is found in all *Desulfovibrio* species with some common structural features: four heme prosthetic groups covalently bound to a single polypeptide chain through thioether bridges to cysteine residues; all axial ligands are histidine residues, rather than histidine and methionine as found in other c -type cytochromes.

Crystal structures of several cytochrome c_3 molecules have been determined. These include the cytochromes from *Desulfovibrio vulgaris* strain Miyazaki F at 1.8 Å [4], *Desulfo-microbium baculatus* strain Norway 4 at 1.7 Å [5], *D. vulgaris* strain Hildenborough for orthorhombic crystal forms at 2.0 Å [6,7] and hexagonal at 1.9 Å [8], *D. gigas* at 1.8 Å [9], *Desulfovibrio desulfuricans* ATCC 27774 at 1.75 Å [10] and finally *D. baculatus* ATCC 9974 at 1.8 Å [11]. All show the same general fold of the polypeptide chain except where deletions and insertions occur.

The sequence homology between *D. vulgaris* Hildenborough and *D. desulfuricans* ATCC 27774 cytochromes c_3 is 42% and the sequence alignment is shown in Fig. 1. Both

* Corresponding author. Tel.: +351 1 441 7823; fax: +351 1 441 1277; e-mail: carrondo@itqb.unl.pt

the individual B -values were reset to a value close to that of the overall B -factor obtained in the previous B -factor refinement stage. After each refinement step, new $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ electron density maps were calculated and inspected on a graphics workstation to carry out model corrections and addition of further water molecules. The R -factor was thus reduced to 17.8%.

The final refinement cycles were performed with SHELXL-93 [16] using a restrained conjugated gradient weighted least-squares procedure on F^2 . Each of the two SHELXL-93 runs consisted of 10 cycles of restrained least-squares refinement, and during each cycle of refinement an R_{free} coefficient [16] was calculated from 10% (2857) of the input 28 861 reflections. This R_{free} coefficient was not used in the X-PLOR refinement because it was carried out just before the version of X-PLOR which allowed the use of R_{free} became available in our laboratory. Following each refinement run, $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ electron density maps were calculated and inspected on a graphics workstation as above to adjust incorrect side-chain positions to the electron density, check for previously unseen side-chains and to add further water molecules to the model. The criteria followed throughout for including a water molecule were that the corresponding peak should be visible on both the $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ Fourier maps and that there should be at least one hydrogen bond (between 2.5 and 3.2 Å). As a result of this procedure, the R -factor was lowered from 17.8 to 15.3% and the final R_{free} was 19.0%. The refinement statistics are presented in Table 2. It should be mentioned that an attempt was made to refine the non-hydrogen atoms anisotropically with SHELXL-93 (while restraining the anisotropic thermal motion parameters of the solvent molecules to be nearly isotropic). However, although both the conventional R -factor and R_{free} dropped (to 12.1 and 17.0%, respectively) the drop in R_{free} was deemed insufficient due to the large increase in the number of parameters (from 8226 to 18 596) with a consequent substantial decrease in the ratio between the number of observations (including geometrical restraints) and the number of parameters from 6.8 to 3.8. Therefore, the final 1.67 Å model was taken as that resulting from the last isotropic SHELXL-93 refinement. The final atomic coordinates have been deposited in the Protein Data Bank [18,19] with id-code 2CTH.

2.3.2. *D. desulfuricans* ATCC 27774 cytochrome c_3

The initial R -factor calculated using the previously determined model with the new data set and cell parameters was 49.2%. All the water molecules included in the initial model were deleted and rigid body refinement was undertaken using data to 2.5 Å resolution only, lowering the R -factor to 30.5%. The data set was then split into a test set (1025 reflections, i.e., a random sample containing approximately 6% of the 17 593 measured independent reflections), used to calculate R_{free} [17], and a working set (the remaining 16 568 reflections), used in the structure refinement. Overall B refinement using 1.6 Å resolution data was followed by an X-PLOR

Table 2
Summary of the refinement statistics for $c_3\text{DvH}$ and $c_3\text{Dd}$

	$c_3\text{DvH}$	$c_3\text{Dd}$	
Resolution limits (Å)	8.0–1.67	8.0–1.6	
Final R_{free} (%) (test set) ^a	19.0	21.0	
Final R -factor (%) (working set, $F_o > \sigma F_o$)	15.3	16.6	
Final R -factor (%) (all 28 861 reflections)	16.1		
No. refinement cycles			
X-PLOR	~400	~200	
SHELXL-93	20	120	
No. reflections used			
X-PLOR	27 065	16 996	
SHELXL-93 (working set, $F_o > \sigma F_o$)	25 741	16 390	
SHELXL-93 (test set, $F_o > \sigma F_o$)	2857	1018	
No. non-hydrogen protein atoms	1964	981	
No. solvent molecules	124	79	
Model r.m.s. deviations from ideal ^b			
Bond lengths (Å) ^c	0.030 (0.033)	0.031	
Bond angles (°) ^c	2.700 (2.944)	2.874	
Numbering scheme	$c_3\text{DvH A}$	$c_3\text{DvH B}^d$	$c_3\text{Dd}$
Protein chain	1...107	113...219	1...107
Heme I	111	223	201
Heme II	110	222	202
Heme III	112	224	203
Heme IV	109	221	204
Solvent molecules	225...348		205...288

^a $F_o > \sigma F_o$ for $c_3\text{Dd}$.

^b Calculations with X-PLOR using the final SHELXL-93 model.

^c Values in parentheses refer to $c_3\text{DvH}$ molecule B.

^d In the text, residues 113...219 are referred to as 1B...107B and residues 223...221 as hemes IB...IVB.

[14] simulated annealing refinement at 3000 K using the standard protocol, another overall B refinement run, 15 cycles of grouped B -factors refinement (as described above for the $c_3\text{DvH}$ refinement) and 15 cycles of individual restrained individual B -factors. At this point, the R -factor was 24.7% and the R_{free} was 29.8%.

Fourier $2|F_o|-|F_c|$ and $|F_o|-|F_c|$ maps were calculated and inspected on a workstation using TOM [15] and TURBO [20]. Some side-chain positions were corrected but no water molecules were included at this stage. The alternate conformations for residues 71–75 present in the 1.75 Å model were not observed in these maps, and a single conformation was adopted for this region. The refinement was continued with SHELXL-93 [16] using a restrained conjugated gradient weighted least-squares procedure on F^2 . All the subsequent SHELXL-93 runs consisted of 20 cycles of restrained least-squares refinement. In order to ensure that the same reflection test set was used in X-PLOR and SHELXL-93, and because the two programs have incompatible test set generation algorithms, the working set and the test set obtained from X-PLOR were kept in separate files. After each SHELXL-93 run, the test set data were used to calculate the R_{free} based on the final refined coordinates, new $2|F_o|-|F_c|$ and $|F_o|-|F_c|$

maps were calculated, water molecules were included in the model, according to the rules mentioned above, and model adjustments were made taking into account the best fit between the model and the electron density. Disorder was modelled in residues Ser 84 and Lys 90. On the basis of the features observed in the electron density maps, Ser 84 O γ was modelled in three alternate positions with occupation factors set to 1/3 and not refined, due to a limitation in SHELXL-93 which does not allow the refinement of site occupation factors for more than two alternate atomic positions. Lys 90 N ζ was also modelled in two alternate positions with occupation factors v and $1 - v$, which refined to $v = 0.41$. The main chain positions of loop 70–73 were not clear, so an omit map was made and the loop was rebuilt (using O [21]) with a new conformation, using the electron density which resulted from the loop omission. This refinement lowered the R -factor to 16.6% and the final R_{free} was 21.0%. The refinement statistics are included in Table 2. As for $c_3\text{DvH}$ above, an anisotropic refinement of the thermal motion parameters of the non-hydrogen protein and solvent atoms in $c_3\text{Dd}$ did not prove to be statistically significant at the data resolution available. Both R and R_{free} were reduced to 13.8 and 20.1%, respectively. However, the number of parameters more than doubled, from 4263 to 9586, leading to a drop in the ratio between the number of observations (including restraints) from 6.2 to 3.6. Therefore, the final 1.6 Å model was that resulting from the last isotropic SHELXL-93 refinement. The final atomic coordinates have been deposited in the Protein Data Bank [18,19] with id-code 3CYR.

3. Structure analysis and discussion

3.1. General features of $c_3\text{DvH}$ and $c_3\text{Dd}$ cytochromes

MOLSCRIPT [22] stereo diagrams of $c_3\text{DvH}$ molecule A and $c_3\text{Dd}$ molecule are represented in Figs. 2 and 3, respectively. The two cytochromes have similar tertiary structures, keeping the overall architecture found for the previous studies of these proteins [8,10] and for all other c_3 cytochromes already determined [4–7,9,11]. In each of these molecules, the polypeptide chain wraps around four heme groups which are held in their positions through thioether bridges to cysteine residues. The octahedral environment of the iron atoms is completed via coordination to histidine residues on both sides of the hemes. The relative position of the four hemes is maintained in all c_3 cytochrome structures with similar intramolecular heme iron–iron distances, edge-to-edge ring distances and interplanar angles. Table 3 lists those values for $c_3\text{DvH}$ and $c_3\text{Dd}$ molecules, respectively. As noticed before [4–11], these values are distributed between two groups. Shorter iron–iron distances correspond to interplanar angles close to orthogonality, while longer distances correspond to those hemes that form more acute interplanar angles.

The vicinity of the axial ligands to the hemes is also conserved through hydrogen bonding patterns involving these residues. Tables 4 and 5 list, for the present study, hydrogen bonds and interplanar angles involving axial histidines in $c_3\text{DvH}$ and $c_3\text{Dd}$ molecules. In particular, three waters appear co-crystallised in similar positions in all known cytochromes, and through hydrogen bonds in both sides of heme I and in

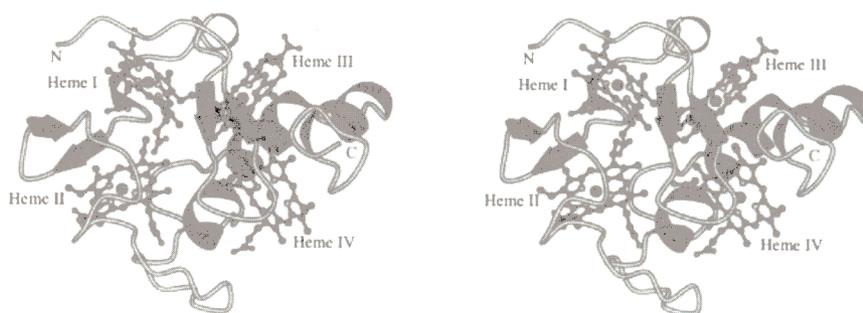


Fig. 2. MOLSCRIPT stereo diagram of $c_3\text{DvH}$ molecule A showing the secondary structure elements of the polypeptide chain and the heme groups.

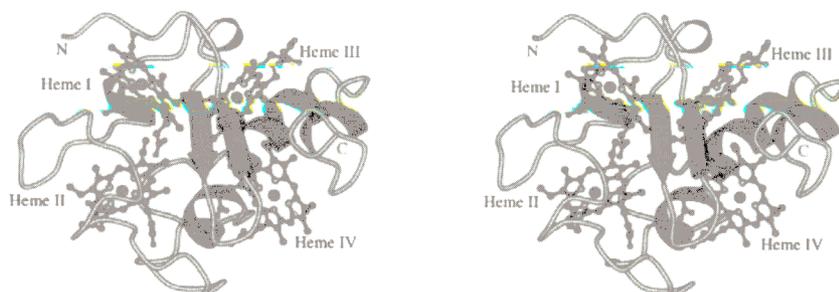


Fig. 3. MOLSCRIPT stereo diagram of the $c_3\text{Dd}$ molecule showing the secondary structure elements of the polypeptide chain and the heme groups. The molecule is drawn in the same orientation as $c_3\text{DvH}$ in Fig. 2.

Table 3
Intramolecular iron–iron distances and heme–heme angles

Heme ^b	Angles (°)/Distances (Å) ^a			
	I	II	III	IV
I	c ₃ DvH A	12.4 (6.6)	11.1 (5.9)	17.8 (9.8)
	c ₃ DvH B	12.4 (6.6)	11.1 (5.9)	17.8 (9.7)
	c ₃ Dd	12.5 (6.5)	11.2 (5.9)	17.8 (9.8)
II	c ₃ DvH A	89.7	16.1 (8.7)	16.7 (10.3)
	c ₃ DvH B	89.2	16.2 (8.8)	16.6 (10.3)
	c ₃ Dd	86.4	16.3 (8.7)	16.9 (10.7)
III	c ₃ DvH A	82.6	60.2	12.0 (5.4)
	c ₃ DvH B	82.1	60.2	12.0 (5.4)
	c ₃ Dd	85.0	59.7	12.3 (5.8)
IV	c ₃ DvH A	21.3	73.0	81.3
	c ₃ DvH B	20.6	72.7	80.6
	c ₃ Dd	20.9	69.8	83.6

^a Distances in parentheses refer to the nearest distance between porphyrin atoms.

^b Heme planes are calculated for all atoms in the porphyrin ring.

one side of heme IV, help to maintain the internal environment of the four hemes and eventually provide pathways for intra- and intermolecular electron transfer events.

The angles between the histidine planes on both sides of the hemes are considered significant structural parameters. The angle relating histidines coordinated to heme II is in all cases ~30 to 40° away from orthogonality, while the other three angles in all cytochromes are in a range close to zero degrees.

3.2. *D. vulgaris* Hildenborough cytochrome c₃

Ramachandran [23] plots for the refined model of the polypeptide chains of each of the two independent molecules in the crystal structure are shown in Fig. 4. Cys 51 in molecule A lies slightly outside the normally allowed regions for non-glycine residues, but the same residue in molecule B has a

Table 5
Hydrogen bonds and interplanar angles involving ligated histidines in c₃Dd

	Donor atom	Hydrogen bond distance N ^{δ1} ...donor ^a (Å)	His–His angle ^{a,b} (°)
Heme I			
His 22	O (Wat 210)	2.74 (2.78)	7.7 (9.8)
His 34	O (Wat 206)	2.88 (2.73)	
Heme II			
His 35	O (Leu 36)	2.92 (2.77)	69.1 (60.2)
His 52	O (Glu 61)	2.74 (2.65)	
Heme III			
His 25	O (Pro 21)	2.87 (2.74)	4.8 (5.3)
His 83	O (Leu 97)	2.83 (2.75)	
Heme IV			
His 69 ^c	O (Tyr 65) ^d	2.76 (2.77)	4.3 (1.4)
His 106	O (Wat 232)	2.88 (2.87)	

^a Values in parentheses are for the 1.75 Å model (Ref. [10]).

^b Acute angle between the normals to the least-squares planes of the imidazole rings.

^c Corresponds to His 70 in c₃DvH.

^d Corresponds to Tyr 66 in c₃DvH.

very similar conformation. All known cytochrome c₃ structures have this feature. Cys 51 binds heme II and is involved in a hydrogen bond that holds a β-turn, as can be seen in Fig. 5. The geometrical constraints imposed by these two bonds distort the conformation of Cys 51, pushing it outside the normally allowed regions for non-glycine residues. This is further supported by the observation that no other heme-binding cysteine is involved in a β-turn, although all of them have hydrogen bonds to other residues.

Asn 38 in molecule B is also slightly outside the most favoured regions for non-glycine residues. This residue is located in a surface loop (its counterpart in molecule A is sitting in the middle of the 'L' region in Fig. 4(a)) and, furthermore, this loop has a poor electron density, hence a slight conformational error in the model is likely, which is

Table 4
Hydrogen bonds and interplanar angles involving ligated histidines in c₃DvH

	Donor atom/hydrogen bond distance N ^{δ1} ...donor (Å) ^a		His–His angle (°) ^{a,b}	
	c ₃ DvH A	c ₃ DvH B	c ₃ DvH A	c ₃ DvH B
Heme I				
His 22	O (Wat 228)	2.82 (2.77)	4.8 (3.7)	2.42 (6.3)
His 34	O (Wat 232)	2.80 (2.61)		
Heme II				
His 35	O (Pro 36)	3.04 (3.12)	55.4 (53.5)	72.1 (72.2)
His 52	O (Ala 62)	2.62 (2.64)		
Heme III				
His 25	O (Asn 21)	2.88 (2.74)	11.2 (16.0)	13.5 (6.1)
His 83	O (Leu 97)	2.83 (2.69)		
Heme IV				
His 70	O (Tyr 66)	2.76 (2.79)	8.2 (9.0)	7.4 (9.2)
His 106	O (Wat 238)	2.83 (2.92)		

^a Values in parentheses are for the 1.9 Å model (Ref. [8]).

^b Acute angle between the normal to the least-squares planes of the imidazole rings.

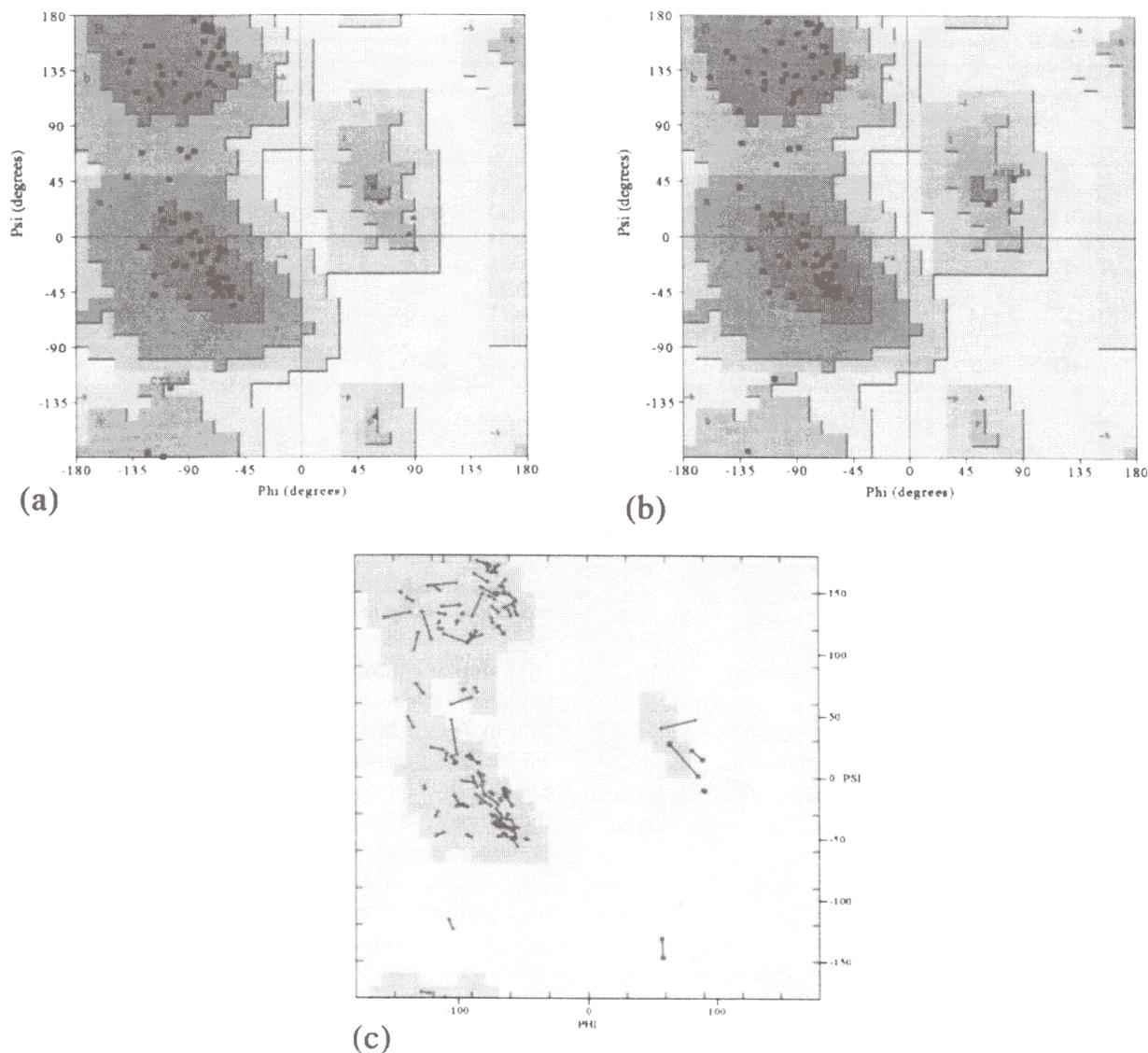


Fig. 4. Ramachandran plots for c_3 DvH molecules A and B. (a) and (b) are standard Ramachandran plots for molecules A and B, respectively, where glycine residues are represented by triangles and other residues by squares; (c) is a multi-model Ramachandran plot of molecules A and B where the symbols for equivalent residues in each molecule are connected. In this plot, glycine residues are represented by squares, other residues by crosses.

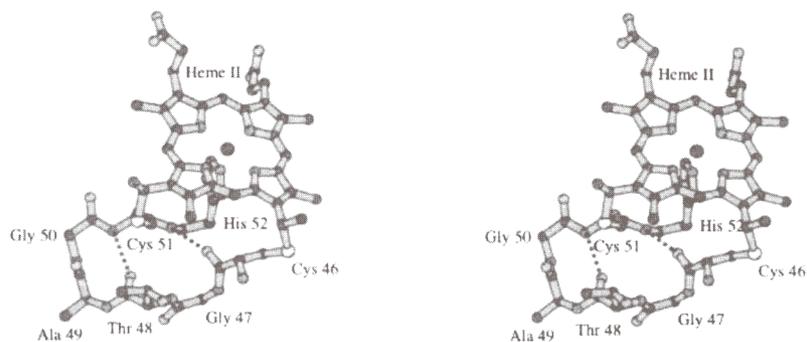


Fig. 5. MOLSCRIPT stereo diagram showing the CXXXXCH binding loop to heme II in c_3 DvH molecule A. Cys 51 is part of this loop, and is involved in a β -turn that distorts its φ, ψ conformation.

revealed by the Ramachandran plot. Since the structural model contains two copies of the same molecule, it is possible to compute a Multiple-Model Ramachandran plot [24] where the symbols for the same residue in each of the two molecules

are connected. This plot is represented in Fig. 4(c) and it can be seen that the equivalent residues in molecules A and B have reasonably similar conformations. Indeed, analysis of the (ϕ, ψ) data shows that the maximum variation in ϕ is

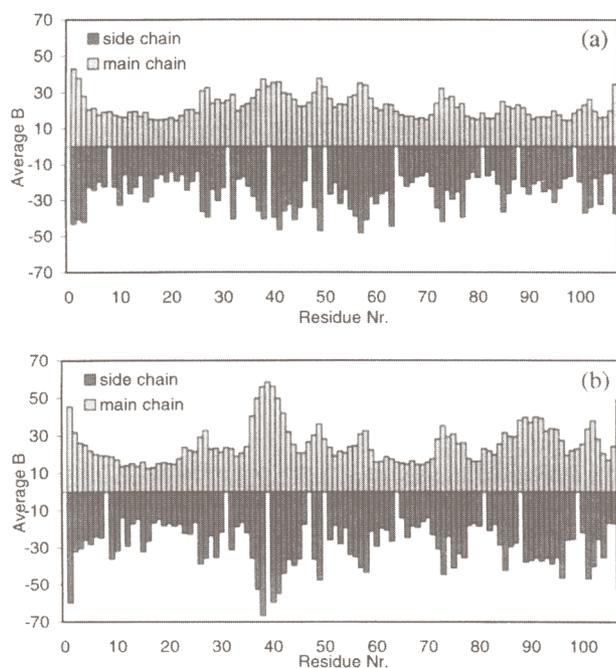


Fig. 6. Variation in average main-chain (above x -axis) and side-chain (below x -axis) B -factors along the protein chains of c_3 DvH molecules A and B.

26.6° for Asn 38, and 89% of the residues have ϕ variations of less than 10°, while the maximum variation in φ is 28.4° for Ala 89, and 94% of the residues have φ variations which are below 10°. Also, most of the residues with higher (ϕ, φ) variations are located in surface loops, where conformational changes due to crystal packing contacts are more likely to occur.

The stereochemical quality of the final model was assessed with program package PROCHECK [25], and all parameters were found to be within their confidence intervals. Although some of the side-chain atoms, which could not be located

from the final 1.9 Å Fourier maps (see Table 8 in [8]), were located during the refinement against 1.67 Å data, in the final Fourier maps it was still not possible to locate the side-chain atoms of some residues, a few of which had been fully determined from the 1.9 Å refinement. All these atoms were included in the model, but given zero occupancy. With a few exceptions, all the residues with side-chain atomic positions not completely known either have long side-chains and/or are located at the protein surface. These are then likely to be mobile, leading to disorder which prevents their location by X-ray crystallography.

The r.m.s. error in the final atomic coordinates was estimated by means of a SIGMA plot [26] to be 0.08 Å. A similar calculation was carried out for the 1.9 Å model, giving 0.32 Å, larger but close to the value (between 0.2 and 0.3 Å) which had been previously estimated (Fig. 4 in Ref. [8]) by means of a Luzatti [27] plot. The variation in average main-chain and side-chain B -factors along the protein chains of molecules A and B are represented in Fig. 6, and show a similar pattern to that observed for the 1.9 Å model (Fig. 5 in Ref. [8]). Also, as was the case in the 1.9 Å model, molecule B has a higher average B -factor (26.4 Å²) than molecule A (23.9 Å²). The average values for the heme groups are given in Table 6. The corresponding heme groups in both molecules have similar values, and the average B variation is similar to that observed in the previous model (Table 9 in Ref. [8]). The B -factor values for the solvent molecules ranged between 15.6 and 81.1 Å², with an average value of 43.5 Å².

The calculation of r.m.s. main-chain and side-chain coordinate deviations was performed with X-PLOR [14] after superimposing the 1.67 Å and 1.9 Å models and the results are illustrated in Fig. 7. Molecules A and B were compared at 1.67 Å, and each molecule at 1.67 Å was compared with the corresponding 1.9 Å model. These results show that the

Table 6
Average B values and r.m.s. differences for the heme groups

Molecule A		R.m.s. difference (Å)					
		c_3 DvH B (1.67 Å, this work)			c_3 DvH A (1.9 Å, Ref. [8])		
Heme	Average B (Å ²)	Overall	Heme ring ^a	Propionate ^b	Overall	Heme ring ^a	Propionate ^b
I	20.5	0.33	0.33	0.32	0.17	0.17	0.17
II	22.9	0.58	0.17	1.15	0.47	0.12	0.97
III	19.9	0.32	0.19	0.56	0.37	0.12	0.72
IV	16.0	0.22	0.16	0.34	0.16	0.15	0.18
Molecule B		R.m.s. difference (Å) toc_3 DvH B (1.9 Å, Ref. [8])					
I	19.2	0.21	0.13	0.37			
II	23.8	0.48	0.15	0.97			
III	19.6	0.15	0.12	0.21			
IV	16.2	0.15	0.15	0.16			

^a The heme ring includes the 4 pyrrole ring atoms and all atoms directly attached to them.

^b The propionate side chain atoms except C^{AA} and C^{AD}.

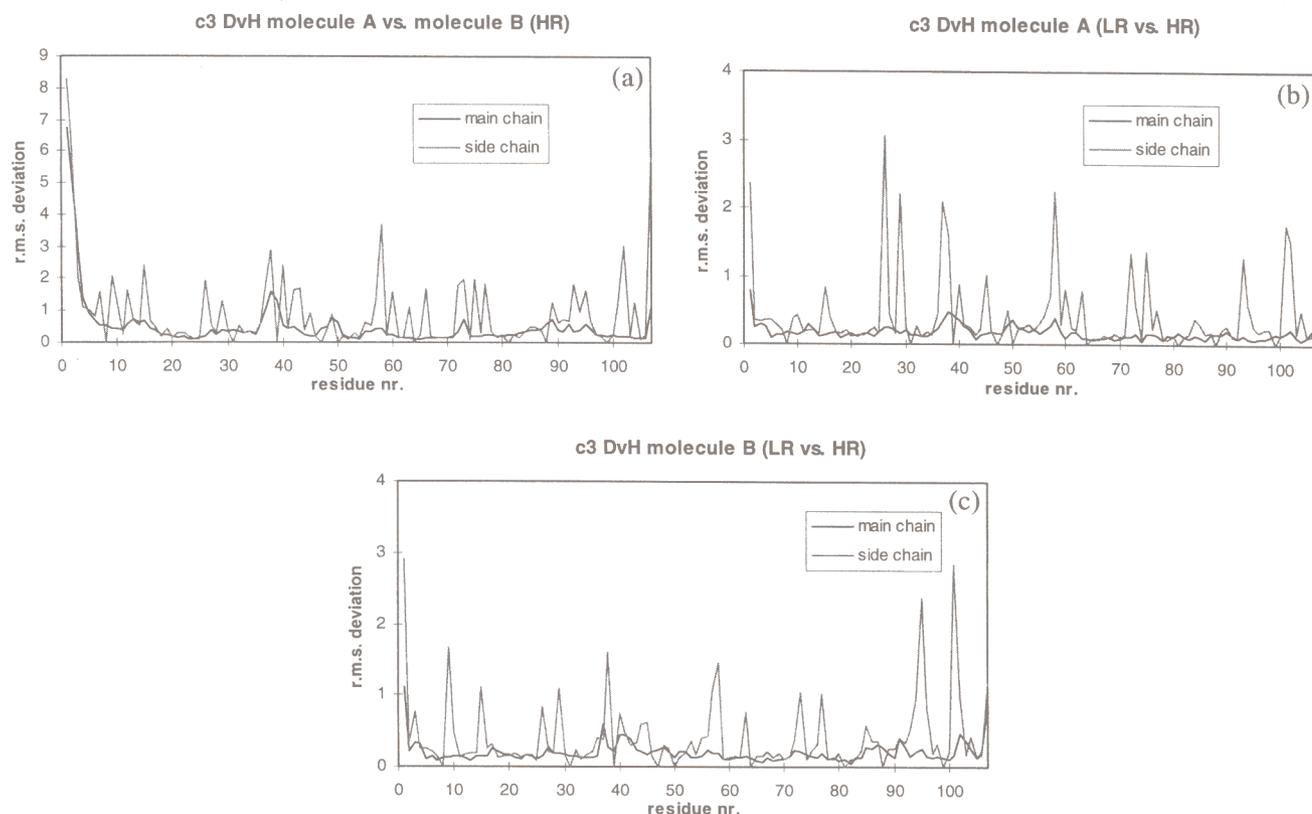


Fig. 7. R.m.s. main-chain and side-chain coordinate deviations for c_3 DvH. (a) Between molecules A and B at 1.67 Å; (b) between the 1.9 Å (lower resolution) and the 1.67 Å (higher resolution) models of molecule A; (c) between the 1.9 Å (lower resolution) and the 1.67 Å (higher resolution) models of molecule B.

difference pattern between molecules A and B is similar at 1.67 and 1.9 Å (Fig. 9(a) in Ref. [8]). The large r.m.s. deviation observed between the N-terminal regions of both molecules arises from a different position of that region in each molecule, due to different crystal packing interactions: in molecule A, the N-terminal interacts with a loop in a symmetry-related B-molecule containing the residues 71–78 (closest main-chain interaction, Lys 3A C α ...Lys 75B C α , 5.3 Å), while in molecule B the N-terminal interacts with a loop in a symmetry-related A-molecule containing the residues 87–93 (closest main-chain interaction, Pro 2B C α ...Asp 90A C α , 4.2 Å). Most of the differences greater than 1.0 Å that are observed between the models for each molecule at the two different resolutions occur for side-chains with atomic positions not completely known, with a few exceptions: in molecule A, Ala 1 has a slightly different conformation for the N-terminal, Val 37 side-chain is in a different rotamer conformation, and Lys 45 has a different side-chain conformation beyond C δ ; in molecule B, the N-terminal is flipped by about 180°, Leu 9 side-chain is in a different rotamer conformation, and Lys 15 has a different side-chain conformation. These differences are probably not significant in terms of the biological properties of this molecule, as they occur in surface regions of the protein, and are likely to be caused by crystal packing interactions. The results of this comparison for the heme groups are included in Table 6, and show that the main differences between the corresponding

heme groups in the two independent molecules occur for the propionate groups in heme II, as was already true in the previous model (results not shown). This is also the case when the models at the two different resolutions for the same molecule are compared.

The relative solvent accessibility was calculated with X-PLOR [14] using the Lee and Richards algorithm [28] with an H₂O probe radius of 1.6 Å. First, the solvent accessible surface was calculated for each residue in the structure and next the same calculation was performed for the same residue, in the same conformation as in the molecule but isolated from the rest of the structure. The ratio of these two numbers for each residue is then its relative solvent accessibility. Comparing these results with those previously obtained for the 1.9 Å model, the differences in relative solvent accessibility for the polypeptide chain (not shown) are all less than 10%, and only a few residues have variations by more than 5%. Some of these variations occur for residues for which there are undetermined side-chain atomic positions (Lys 29 A, Lys 75 A, Lys 77 B, Lys 102 B). Of the remainder variations, Asp 32 A (–5.5%) is due to neighbouring Lys 29 A, which has undetermined side-chain atoms but the others are genuine variations (Ala 1 B, +9.0%; Ala 92 B, +6.3%) arising from a residue conformational change (Ala 1 B) or from a positional shift in a neighbouring residue (Lys 95 B, near Ala 92 B, which also has undetermined side-chain atoms but bearing no impact on the solvent accessibility of Ala 92 B). The

Table 7
Relative percentage of solvent accessibility for the heme groups

Heme	c_3 DvH A ^a	c_3 DvH B ^a	c_3 Dd ^b
I	12.9 (12.6)	11.7 (12.2)	10.9 (10.3)
II	14.8 (14.7)	14.6 (14.5)	13.1 (11.7)
III	14.8 (14.4)	13.9 (13.5)	18.0 (17.6)
IV	13.8 (14.0)	12.6 (12.9)	12.8 (11.6)

^a Values in parentheses are for the 1.9 Å model (Ref. [8]).

^b Values in parentheses are for the 1.75 Å model (Ref. [10]).

results for the heme groups are listed in Table 7 and are very similar to those obtained previously for the 1.9 Å model.

3.3. *D. desulfuricans* ATCC 27774 cytochrome c_3

A MOLSCRIPT [22] stereo diagram of the c_3 Dd molecule is represented in Fig. 3 and the Ramachandran [23] plot for the refined model of the polypeptide is shown in Fig. 8. No residue lies outside the normally allowed regions for non-glycine residues, although Cys 51 is only marginally inside

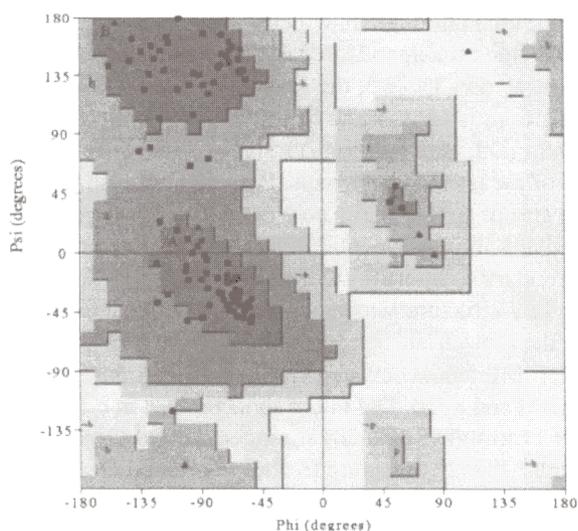


Fig. 8. Ramachandran plot for the c_3 Dd molecule. Glycine residues are represented by triangles, other residues by squares. Drawing made with PROCHECK.

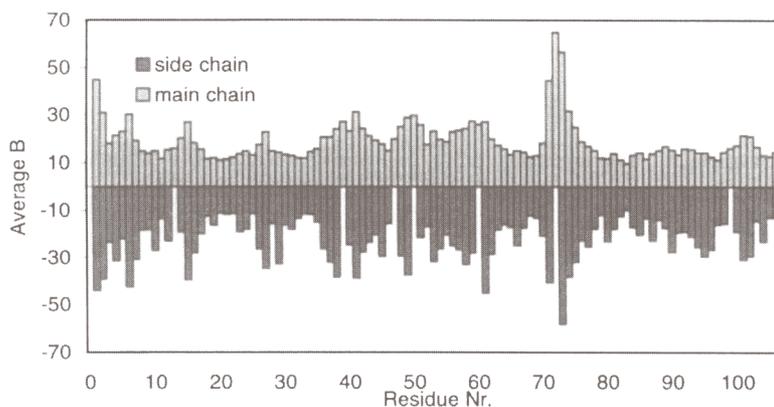


Fig. 9. Variation in average main-chain (above x -axis) and side-chain (below x -axis) B -factors along the protein chain of c_3 Dd molecule.

those regions. This situation is analogous to what was discussed above about the same residue in c_3 DvH.

The stereochemical quality of the final model was analysed with program package PROCHECK [25], and all parameters were found to be within their confidence intervals. Although some of the side-chain atoms which could not be located from the final 1.75 Å Fourier maps (see Table 4 in [10]) were located during the refinement against 1.6 Å data, in the final Fourier maps it was not possible to locate the side-chain atoms of some residues, a few of which had been fully determined from the 1.75 Å refinement. As in the c_3 DvH refinement, all these atoms were included in the model, but given zero occupancy. Like for c_3 DvH, with a few exceptions, all the residues with side-chain atomic positions not completely known either have long side-chains and/or are located at the protein surface. The surface loop containing Gly 72 shows some missing electron density, especially for Glu 73, possibly because of loop flexibility. As a consequence, this loop conformation is poorly defined.

The r.m.s. error in the final atomic coordinates was estimated as 0.09 Å by means of a SIGMAA plot [26]. A similar calculation for the 1.75 Å model gave 0.13 Å as result, in agreement with the value (between 0.1 and 0.2 Å) which had been previously estimated (Fig. 5 in Ref. [10]) by means of a Luzatti [27] plot. The variation in average main-chain and side-chain B -factors along the protein chain is represented in Fig. 9, and shows a similar pattern to that observed for the 1.75 Å model (Fig. 7 in Ref. [10]), except for the loop region comprising residues 71 to 75. This region had been previously modelled as having two alternate conformations, and although there was no clear evidence of this kind of static disorder in the 1.6 Å model, it has nevertheless the highest B -factors, suggesting a high mobility for this loop within a single conformation. Also, the 1.6 Å model for the c_3 Dd molecule has an higher average B (20.5 versus 16.5 Å²). Like for c_3 DvH, all the higher B -factor regions are located at the protein surface. The average values for the heme groups are given in Table 8. The corresponding heme groups have higher B -values for the 1.6 Å model, and although the pattern is similar for hemes I, II and III (heme II has the highest B , and hemes I and III have similar, lower B -values) heme IV

Table 8
Average B values and r.m.s. differences for the heme groups in c_3Dd

Heme	Average B (\AA^2) ^a	R.m.s. difference (\AA) to c_3Dd (1.75 \AA , Ref. [9])		
		Overall	Heme ring ^b	Propionate ^c
I	15.9 (11.7)	0.18	0.15	0.27
II	20.4 (14.7)	0.54	0.16	1.08
III	14.1 (11.3)	0.44	0.13	0.87
IV	17.4 (10.7)	0.20	0.18	0.24

^a Values in parentheses refer to the 1.75 \AA model (Ref. [10]).

^b The heme ring includes the 4 pyrrole ring atoms and all atoms directly attached to them.

^c The propionate side-chain atoms except C^{AA} and C^{AD} .

now has the second highest B instead of the lowest as was the case in the 1.75 \AA model [10]. The B-factor values for the solvent molecules ranged between 14.4 and 78.9 \AA^2 , with an average value of 37.8 \AA^2 .

R.m.s. main-chain and side-chain coordinate deviations were calculated with X-PLOR [14] after superimposing the 1.6 and the 1.75 \AA models and the results are illustrated in Fig. 10. This calculation was performed for both alternate conformations of loop 71–75 in the 1.75 \AA model, and that which gave the lower deviations was chosen for inclusion in Fig. 10. Not surprisingly, the most significant main-chain r.m.s. deviations occur in the loop region 71–75. While most of the side-chain r.m.s. deviations greater than 1.0 \AA occur for residues with side-chain atomic positions not completely known, where the main contribution to the r.m.s. deviation arises mainly from the differences in the atomic positions which were given a zero occupation factor, there are a few exceptions: in Val 4, the side-chain is in a different rotamer conformation, in Asn 6, the main-chain has a slightly different conformation, in Ser 48, the side-chain is in a different rotamer conformation and the main-chain is also in a slightly different conformation, and finally Ser 49 has the side-chain in a different rotamer conformation. The results of this comparison for the heme groups are shown in Table 8, where it can be seen that the largest deviations occur in the propionate groups of hemes II and III.

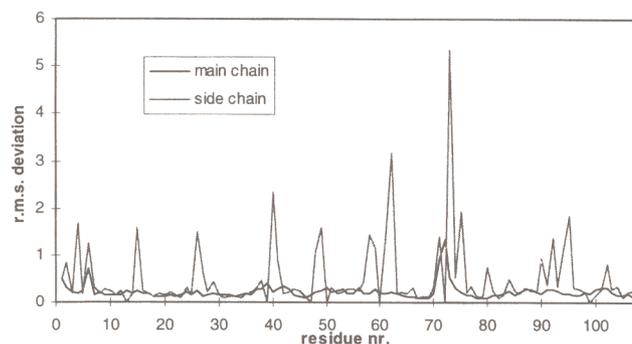


Fig. 10. R.m.s. main-chain and side-chain coordinate deviations for the c_3Dd molecule, between the 1.9 \AA and the 1.67 \AA models.

The relative solvent accessibility was calculated with X-PLOR [14] as described above for c_3DvH . Comparison of these results with a similar calculation on the 1.75 \AA model (not shown) reveals that only four residues have variations larger than 5%, and for the most part these variations are likely to be artifacts caused by residues with undetermined side-chain atomic positions (Lys 40, Lys 62 and Lys 75). The remaining variation (Ala 23, +6.1%) is mainly due to the side-chain conformation change in neighbouring residue Glu 26, which has the terminal carboxyl group with undetermined atomic positions. The results of this calculation for the heme groups are included in Table 7 and are comparable to those previously obtained for the 1.75 \AA model.

3.4. Comparison between the two cytochrome c_3 structures

The two refined cytochromes c_3 , despite the relatively low sequence analogy, have similar tertiary structures overall, as can be seen in Fig. 11, where the superimposed $C\alpha$ -traces and heme groups of c_3DvH A and c_3Dd are represented. This fold is characteristic of tetraheme cytochrome c_3 molecules, with the four heme groups covalently bound to a single polypeptide chain through thioether bridges to cysteine residues and all heme axial ligands are histidine residues. Furthermore, as illustrated in Table 3, the geometrical arrangement of the heme groups is very similar in c_3DvH (both molecules A and B) and c_3Dd , although Fig. 11 shows some differences for some of the propionate groups. Probably because these four heme groups form a bulky core around which the folding of a relatively short polypeptide chain takes place, the elements of secondary structure present are few, mainly α -helices. The secondary structure was analyzed with PROCHECK [25] using the Kabsch and Sander method [29] and revealed some detailed differences between the three-dimensional structures of c_3DvH and c_3Dd . The initial short β -sheet in c_3DvH (Leu 9-Met 11 to Val 18-Phe 20) is longer in c_3Dd (Val 9-Lys 12 to Thr 17-Phe 20), leading to a more ordered conformation of the initial loop in the structure. His 25, which coordinates heme III, is located in both molecules at the end of a very short 3_{10} helix (Ala 23-His 25). While the first heme-binding loop (Cys 30-X-X-Cys 33-His 34) in c_3Dd constitutes a short α -helix, in DvH it defines a shorter 3_{10} helix, which does not include His 34. On the other hand, the second short β -sheet

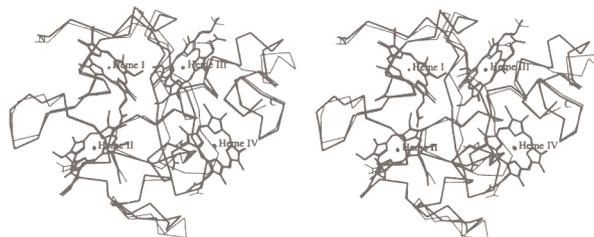


Fig. 11. Stereo view of the superimposed $C\alpha$ -trace and heme groups of c_3DvH molecule A (thin line) and c_3Dd (thick line). Drawing made with TURBO. The molecules are drawn in approximately the same orientation as in Figs. 2 and 3.

in c_3 DvH (Pro 36-Val 37 to Lys 40-Glu 41) is absent in c_3 Dd. As mentioned above, the second heme-binding loop (Cys 46-X-X-X-X-Cys 51-His 52) is in a region that contains an hydrogen-bonded β -turn in both molecules. Although both cytochromes have polypeptide chains which are 107 amino acid residues long, in c_3 Dd there is one deletion and one insertion with respect to c_3 DvH as shown in Fig. 1. This occurs between Ser 61 and Phe 76 in c_3 DvH, and as a result in this region the two sequences are out of register. Nevertheless, there is a conserved α -helix in this region (Tyr 65-His 70 in c_3 DvH, Leu 64-His 69 in c_3 Dd), which contains one of the histidine residues that coordinate heme IV. The final secondary structure elements are two α -helices (Cys 79-Ala 87 and Ala 91-Thr 98 in c_3 DvH, Cys 79-Glu 89 and Lys 94-Thr 98 in c_3 Dd) connected by a short loop. The first of these helices contains the third heme-binding region (Cys 79-X-X-Cys 82-His 83) and is longer in c_3 Dd than in c_3 DvH while the second is longer in c_3 DvH than in c_3 Dd. The final heme-binding region (Cys 100-X-X-X-X-Cys 105-His 106) is located in a loop at the end of the second helix near the C-terminal of the protein chain. The two molecules as shown in Fig. 11 were superimposed with TURBO [20], using all C α atoms less than 1.0 Å apart to derive a superposition rotation-translation operator. As a result of this distance criterion, 69 out of the 107 C α atoms were used in the calculation, with an r.m.s. C α distance between corresponding residues of 0.23 Å. Also, the matching between the two sequences obtained from this calculation is in agreement with the a priori sequence alignment represented in Fig. 1. The zones where the two structures are most different are mainly the loops which are the most variable regions between different tetraheme cytochrome c_3 molecules.

4. Conclusions

More accurate structural models were obtained for the three-dimensional structures of cytochromes c_3 from *D. vulgaris* Hildenborough and *D. desulfuricans* ATCC 27774 by refinement against higher resolution diffraction data measured with synchrotron radiation, while at the same time the crystallographic R -factors were improved. The comparison between the new and the previous structural models for each molecule showed that there were no significant differences between them. The comparison between the two different cytochrome c_3 molecules showed that the differences were consistent with the a priori sequence alignment and were mainly concentrated in the variable loop regions common to all the tetraheme cytochrome c_3 molecules.

Acknowledgements

This work was funded with European Union grants BIO2-CT942052, HCM-CHRX-CT93-0143 and PRAXIS/2/2.1/

QUI/17/94. Partial financial support with BJI 2746 to Paulo Simões from Junta Nacional de Investigação Científica e Tecnológica is gratefully acknowledged. The authors would like to thank Dr Cláudio Soares (ITQB-UNL) for his participation in the initial stages of the c_3 DvH refinement; the fermentation plant at University of Georgia, Athens, GA, USA, for growing the bacteria; Ms Isabel Pacheco (ITQB-UNL) for the purified c_3 DvH protein sample used for crystallization and Dr Cristina Costa, Professor José J.G. Moura, and Professor Isabel Moura (FCT-UNL) for the purified c_3 Dd protein sample used for crystallization.

References

- [1] H.D. Peck Jr. and J. LeGall (eds.), *Inorganic Microbial Sulfur Metabolism*. Methods in Enzymology, Vol. 243, Academic Press, New York, 1994.
- [2] R.O. Louro, T. Catarino, C.A. Salgueiro, J. LeGall and A.V. Xavier, *JBIC*, 1 (1996) 34–38.
- [3] H. Santos, J.J.G. Moura, I. Moura, J. LeGall and A.V. Xavier, *Eur. J. Biochem.*, 141 (1984) 283–296.
- [4] Y. Higuchi, M. Kusunoki, Y. Matsuura, N. Yasuoka and M. Kakudo, *J. Mol. Biol.*, 172 (1984) 109–139.
- [5] M. Czjzek, F. Payan, F. Guerlesquin, M. Bruschi and R. Haser, *J. Mol. Biol.*, 243 (1994) 653–667.
- [6] Y. Morimoto, T. Tani, H. Okumura, Y. Higuchi and N. Yasuoka, *J. Biochem.*, 110 (1991) 532–540.
- [7] Y. Higuchi, H. Akutsu and N. Yasuoka, *Biochimie*, 76 (1994) 537–545.
- [8] P.M. Matias, C. Frazão, J. Morais, M. Coll and M.A. Carrondo, *J. Mol. Biol.*, 234 (1993) 680–699.
- [9] P.M. Matias, J. Morais, R. Coelho, M.A. Carrondo, K. Wilson, Z. Dauter and L. Sieker, *Prot. Sci.*, 5 (1996) 1342–1354.
- [10] J. Morais, P. Nuno Palma, C. Frazão, J. Caldeira, J. LeGall, I. Moura, J.J.G. Moura and M.A. Carrondo, *Biochemistry*, 34 (1995) 12 830–12 841.
- [11] A.V. Coelho, P. Matias, C. Frazão and M.A. Carrondo, Crystal structure of c_3 cytochrome from *D. baculatus* ATCC 9974 at 1.8 Å resolution, in 4th Eur. Workshop on Crystallography of Biological Macromolecules, Como, Italy, 1995.
- [12] A.G.W. Leslie, CCP4 and ESF-EACMB Newsletter on Protein Crystallography, 26 (1992).
- [13] Collaborative Computational Project Number 4, *Acta Crystallogr.*, Sect. D, 50 (1994) 760–763.
- [14] A.T. Brünger, X-PLOR: a system for crystallography and NMR, The Howard Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, 1992.
- [15] C. Cambillau and E. Horjales, *J. Mol. Graph.*, 5 (1987) 175–177.
- [16] G.M. Sheldrick and T.M. Schneider, *Methods in Enzymology*, in press.
- [17] A.T. Brünger, *Nature*, 355 (1992) 472–474.
- [18] E.E. Abola, F.C. Bernstein, S.H. Bryant, T.F. Koetzle and J. Weng, in F.H. Allen, G. Bergeroff and R. Sievers (eds.), *Crystallographic Databases — Information Content, Software Systems, Scientific Applications*, Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester, 1987, pp. 107–132.
- [19] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, *J. Mol. Biol.*, 112 (1977) 535–542.
- [20] A. Roussel, J.C. Fontecilla-Camps and C. Cambillau, TURBO-FRODO: a new program for protein crystallography and modelling. XV IUCr Congress, Bordeaux, France, 1990.

- [21] T.A. Jones and M. Kjelgaard, *Acta Crystallogr., Sect. A*, 47 (1991) 110–119.
- [22] P.J. Kralis, *J. Appl. Cryst.*, 24 (1991) 946–950.
- [23] G.N. Ramachandran and V. Sasisekharan, *Adv. Prot. Chem.*, 23 (1968) 283–437.
- [24] G.J. Kleywegt, *Acta Crystallogr., Sect. D*, 52 (1996) 842–857.
- [25] R.A. Laskowski, M.W. MacArthur, D.S. Moss and J.M. Thornton, *J. Appl. Cryst.*, 26 (1993) 283–291.
- [26] R.J. Read, *Acta Crystallogr., Sect. A*, 42 (1986) 140–149.
- [27] V. Luzatti, *Acta Crystallogr.*, 5 (1952) 802–810.
- [28] B. Lee and F.M. Richards, *J. Mol. Biol.*, 55 (1971) 379–400.
- [29] W. Kabsch and C. Sander, *Biopolymers*, 22 (1983) 2577–2637.