Acta Cryst. (1996). D**52**, 356–368

Three-Dimensional Structure of Cytochrome *c′* from Two *Alcaligenes* Species and the Implications for Four-Helix Bundle Structures

AARON J. DOBBS, BRYAN F. ANDERSON, H. RICK FABER AND EDWARD N. BAKER*

Department of Chemistry and Biochemistry, Massey University, Palmerston North, New Zealand

(*Received 13 April 1995; accepted 21 June 1995*)

Abstract

The three-dimensional structures of two cytochromes *c′* have been determined in order to analyse the common features of proteins of this family and their relationship with other four-helix bundle structures. The structure of cytochrome *c′* from *Alcaligenes sp* was determined by molecular replacement supplemented with the iron anomalous scattering and the use of a single isomorphous heavy-atom derivative, and was refined using synchrotron data to 1.8 Å resolution. The final model, comprising 956 protein atoms (one monomer) and 89 water molecules, has a final *R* value of 0.188 for all data in the range 20.0–1.8 Å resolution (14 673 reflections). The structure of the cytochrome *c′* from *Alcaligenes denitrificans* is isomorphous and essentially identical (r.m.s. deviation for all atoms 0.36 Å). Although its amino-acid sequence has not been determined chemically, only four differences from that of *Alcaligenes sp* cytochrome *c′* were identified by the X-ray analysis. The final model for *Alcaligenes denitrificans* cytochrome *c′*, comprising 953 protein atoms and 75 water molecules, gave a final *R* factor of 0.167 for all data in the range 20.0–2.15 Å (8220 reflections). The cytochrome *c′* monomer forms a classic four-helix bundle, determined by the packing of hydrophobic side chains around the enclosed haem group. There are very few cross-linking hydrogen bonds between the helices, the principal side-chain hydrogen bonding involving one of the haem propionates and a conserved Arg residue. The cytochrome *c′* dimer is created by a crystallographic twofold axis. Monomer–monomer contacts primarily involve the two *A* helices, with size complementarity of side chains in a central solvent-excluded portion of the interface and hydrogen bonding at the periphery. Both species have a pyroglutamic acid N-terminal residue. The haem iron is five-coordinate, 0.32 Å out of the haem plane towards the fifth ligand, His120. The unusual magnetic properties of the Fe atom may be linked to a conserved basic residue, Arg124, adjacent to His120.

1. Introduction

Cytochrome *c′* is a small *c*-type cytochrome which is found in a variety of photosynthetic and denitrifying bacteria (Bartsch, 1978). In several respects, the cytochromes *c′* bear a superficial resemblance to the proteins of the mitochondrial cytochrome *c* class; they have a polypeptide of 120–130 amino-acid residues (approximately 14 kDa) and the single protohaem IX group is attached through two Cys–thioether linkages provided by the common Cys-*X*-*Y*-Cys-His motif. Apart from these parallels, however, the cytochromes *c′* have many unique features, and their structural and physical properties mark them out as a distinct protein family.

The cytochromes *c′* are usually isolated as soluble homodimers (2 × 14 kDa), although several species also undergo an unusual ligand-controlled dimer dissociation (Cusanovich & Gibson, 1973; Kassner, 1991). The Cys-*X*-*Y*-Cys-His motif through which the haem is attached is close to the C-terminus of each polypeptide (in contrast to the mitochondrial cytochromes *c* where it is near the N-terminus). The His residue in this sequence motif also provides a fifth ligand for the Fe atom, but another unique feature of the cytochromes *c′* is the absence of any sixth ligand provided by the protein. Instead, they have the property of being able to bind certain small, neutral ligands such as CO and NO (but not O_2) at the vacant coordination site (Gibson & Kamen, 1966; Doyle, Weber & Gill, 1985). Their spectroscopic properties are also marked by an unusual mixed spin state of the Fe atom (Maltempo, Moss & Cusanovich, 1974) which has been attributed to the relatively weak ligand field provided by the five-coordinate site (Weber, 1982).

Although found in a wide variety of bacteria the actual functional role of cytochromes *c′* is not clear (Meyer & Kamen, 1982). Their physicochemical properties suggest a role in electron transfer, perhaps in denitrifying pathways, given their distribution. It is also not known whether their ability to bind species such as CO and NO has any physiological significance.

Structurally, the cytochromes *c′* are of considerable interest. Crystallographic analyses of the cytochromes *c′* from three different bacterial species have established that they have a classic four-helix bundle fold; the structures determined to date include the cytochromes *c′* from *Rhodospirillum molischianum*, at 1.67 Å resolution (Weber, Howard, Xuong & Salemme, 1981; Finzel, Weber, Hardman & Salemme, 1985); from *Chromatium vinosum*, at 1.8 Å resolution (Ren, Meyer & McRee, 1993); and from *Rhodospirillum rubrum*, at 2.8 Å resolution (Yasui, Harada, Kai, Kasai, Kusunoki &

©1996 International Union of Crystallography
Printed in Great Britain – all rights reserved

Matsuura, 1992). A striking feature of the cytochrome *c′* family is, however, the relatively low level of sequence conservation among its members (Ambler *et al.*, 1981). Pairwise sequence comparisons show only 25–30% identity, and very few residues are conserved through the family as a whole.

The structural constraints (or relative lack of them) within the cytochrome *c′* family are thus of considerable relevance to current interest in the *de novo* design of proteins. Many of these studies are directed towards the design of proteins with a four-helix bundle fold (Ho & De Grado, 1987; Regan & De Grado, 1988; Hecht, Richardson, Richardson & Ogden, 1990) and the ways in which this fold can be varied and stabilized are thus of much interest. Furthermore, the haem in cytochrome *c′* occupies a cleft at one end of the four-helix bundle, formed by the splaying outwards of the helices; this feature, too, has been utilized in attempts to incorporate a particular active site into a designed molecule of this type (Hahn, Klis & Stewart, 1990).

Here we describe high-resolution crystal structures for two closely related cytochromes *c′*, from *Alcaligenes sp* and *Alcaligenes denitrificans*. An earlier report described attempts to overcome the difficulties posed by the low similarity of available search models for the use of molecular replacement in the structure determination (Baker, Anderson, Dobbs & Dodson, 1995). Here the completion of the structure analysis and the details of the refined structure are presented, together with an analysis of the similarities and differences between these structures and the others determined to date; this is aimed at establishing the common structural framework of the cytochrome *c′* family, including the stabilization of haem binding, and the relevance to other four-helix bundle proteins.

2. Experimental

2.1. *Crystals and heavy-atom derivatives*

Cytochrome *c′* was isolated from cultures of *Alcaligenes sp* (NCIB 11015, formerly known as *Pseudomonas denitrificans* and now also known as *Achromobacter xylosoxidans*) and *Alcaligenes denitrificans* (NCTC 8582) and crystallized in its oxidized, Fe^{III}, form as previously described (Norris, Anderson, Baker & Rumball, 1979). The crystals are isomorphous, with cell dimensions $a = b = 54.4$, $c = 181.1$ Å, $\alpha = \beta = 90$, $\gamma = 120°$ (*Alcaligenes sp*) and $a = b = 54.6$, $c = 180.4$ Å, $\alpha = \beta = 90$, $\gamma = 120°$ (*Alcaligenes denitrificans*), space group $P6_122$ or $P6_522$, with one monomer in the asymmetric unit. The space group was subsequently determined to be $P6_522$ from the structure analysis by molecular replacement and anomalous scattering methods (Baker *et al.*, 1995).

Attempts to prepare heavy-atom derivatives were made by soaking the crystals in standard mother liquor (0.1 *M* Tris–HCl buffer, 90% saturated with ammonium sulfate, at pH 8.0) in which were dissolved small concentrations (1–10 m*M*) of various heavy-atom reagents. The small number of cytochrome *c′* crystals available limited these experiments, which were carried out on the *Alcaligenes sp* cytochrome *c′*, and only one derivative was found that gave suitable intensity changes without disrupting the crystals. This derivative was with K_2PtCl_4 at a concentration of 1 m*M* and a soaking time of 1 d.

Table 1. *Data-collection statistics*

Crystal	*Alcaligenes sp* Native*	*Alcaligenes sp* Native	*Alcaligenes denitrificans* Native*	*Alcaligenes sp* K_2PtCl_4
Resolution	1.80	2.82	2.15	2.32
Observations	75683	15232	60526	46282
Unique reflections	14814	4038	8220	7364
Completeness (%)	94	94	90	98
R_m†	0.075	0.016	0.032	0.035
I/σ (outer shell)	4.5	11.0	3.9	5.6

* Data sets used for structure refinement. † $R_m = \sum\|I - \bar{I}\| / \sum\bar{I}$ where the summation is over all redundant measurements.

2.2. *Data collection*

The earlier molecular-replacement analysis of the two cytochromes *c′* used two sets of intensity data, a 2.8 Å native data set for *Alcaligenes sp* cytochrome *c′*, collected at room temperature on a Xuong–Hamlin multiwire area detector (Cu $K\alpha$, Rigaku RU200 rotating-anode generator), and a 1.86 Å native data set for *Alcaligenes denitrificans* cytochrome *c′* collected at room temperature, from a single crystal, on a Rigaku R-AXIS IIC image-plate detector (Cu $K\alpha$, Rigaku RU200 rotating-anode generator). The full details of data collection are as reported earlier (Baker *et al.*, 1995), but the statistics are also summarized in Table 1. In the case of the *Alcaligenes denitrificans* data set the completeness at the highest resolution was low, and for refinement the effective resolution was judged to be 2.15 Å at which resolution the overall completeness was 90%. For the completion of the structure analysis, described herein, two additional data sets were collected; a native data set to 1.8 Å resolution for *Alcaligenes sp* cytochrome *c′* and a data set for the K_2PtCl_4 derivative of the same protein (see also Table 1).

The 1.8 Å native data set for *Alcaligenes sp* cytochrome *c′* was collected at 283 K using the synchrotron radiation source of the Photon Factory (Tsukuba, Japan). Intensity data were measured by screenless Weissenberg photography, with image plates (Sakabe, 1991), using an X-ray source of wavelength 1.0 Å. Two crystals were used; one oriented along the unique [*c*] axis was used to collect data over a rotation range of 60°, and the second, oriented along the [*a*] axis gave data over a rotation range of 90°. Total exposure time was 65 min for crystal 1 and 84 min for crystal 2. Data were processed using the

program *WEIS* (Higashi, 1989) and after scaling and merging, the complete data set comprised 14 814 unique reflections, with an overall R_{merge} of 0.075. The overall data set was 94% complete and in the outer 1.8–2.0 Å shell (87% complete) the average $I/\sigma(I)$ was 4.5

For the K_2PtCl_4 derivative of *Alcaligenes sp* cytochrome c' intensity data were collected at room temperature using a Rigaku R-AXIS IIC image-plate detector on a Rigaku RU 200 rotating-anode generator, with graphite-monochromated Cu $K\alpha$ radiation. One crystal was used to collect data to 2.35 Å resolution; the oscillation range (about [c]) was 60°, comprising 30 frames of 2.0° oscillation each, and a total exposure of 30 h. The final data set comprised 7364 unique reflections, with a merging R value of 0.035.

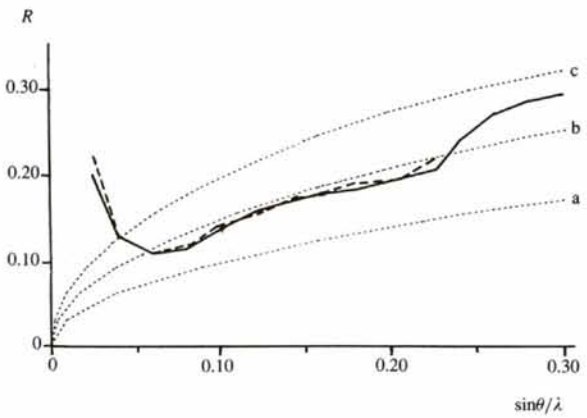

Fig. 1. Variation of the crystallographic R factor with resolution for the cytochromes c' from *Alcaligenes sp* (full line) and *Alcaligenes denitrificans* (broken line). The theoretical variation for coordinate errors of (*a*) 0.10, (*b*) 0.15 and (*c*) 0.20 Å (Luzzati, 1952) is shown.

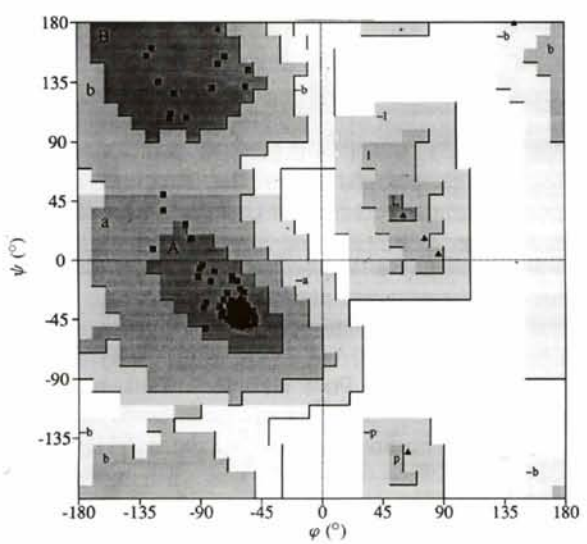

Fig. 2. Ramachandran plot (Ramakrishnan & Ramachandran, 1965) of the main-chain torsion angles (φ and ψ) for the final refined model of *Alcaligenes sp* cytochrome c'.

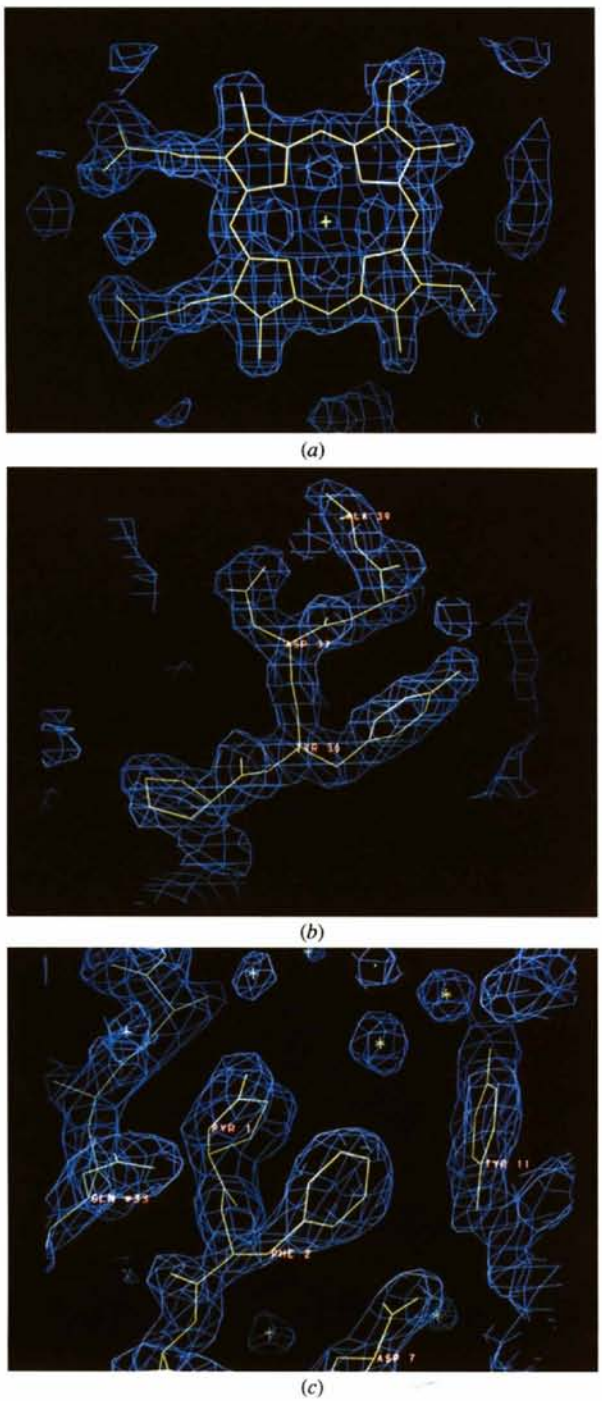

Fig. 3. Electron density for representative features of the *Alcaligenes sp* cytochrome c' structure, from the final $2F_o - F_c$ map, contoured at a level of 1.5σ. (*a*) The haem group. (*b*) A section of the *A–B* loop, residues 35–39. (*c*) The N-terminal pyroglutamic acid residue.

Table 2. *Derivative and phasing details*

Cytochrome c'	*Alcaligenes sp*
Heavy atom-reagent	K_2PtCl_4
Concentration (m*M*) (soaking time)	1 (1 d)
Crystal cell dimensions (Å)	$a = b = 54.4$, $c = 181.3$
Isomorphous intensity change (%)	15.1
Number of sites	2
Occupancies*	2.6, 1.3
R_{cullis}	0.71, 0.69 (acentric, centric)
Phasing power	1.2, 0.9 (acentric, centric)

* On an arbitrary scale.

2.3. *Structure determination*

Earlier attempts to solve the structures of either *Alcaligenes sp* cytochrome *c'* or *Alcaligenes denitrifcans* cytochrome *c'* by molecular replacement had been made difficult by the low degree of similarity of the available search models. In the end, the phase information from the iron anomalous scattering, combined from both structures, together with considerations of monomer association into dimers, proved sufficient to identify the correct molecular-replacement solution (Baker *et al.*, 1995). An electron-density map phased solely by the iron anomalous scattering (with solvent flattening) allowed some rebuilding of the structure and initial refinement, but the inter-helix connecting loops could not be traced satisfactorily and refinement was held up at $R = 0.28$–0.30. The use of calculated phases from the model, or calculated phases combined with anomalous phases, also failed to reveal the missing structure. It was for this reason that the single isomorphous derivative was used, to provide additional phasing.

The data for the K_2PtCl_4 derivative were scaled to the native *Alcaligenes sp* data, giving a mean isomorphous difference of 15.1%, and a difference Patterson map was interpretable in terms of one major site of substitution and one minor site. Details are given in Table 2. The major binding site of the heavy atom later proved to be the side chain of His120. Heavy-atom parameters (x, y, z, B and occupancy) were refined with *MLPHARE* (Otwinowski, 1991), and a single isomorphous replacement (SIR) phase determination gave a mean overall figure of merit of 0.44 for all data in the resolution range 20.0–3.0 Å. A subsequent SIR-phased electron-density map correlated well with the electron-density map phased by the iron anomalous scattering (correlation coefficient 0.32) and was of similar quality; it confirmed the original molecular-replacement solution but failed to give better definition to the inter-helix connecting loops. Use of solvent flattening, with *SQUASH* (Cowtan & Main, 1993) or Wang's method (Wang, 1985) increased the figure of merit, but did not improve the electron-density map sufficiently to trace the missing structure. Attempts to combine the SIR phases with the iron anomalous phases were made difficult by the lack of suitable weights for the latter, because of the way they were obtained (Baker *et al.*, 1995).

An electron-density map from which the missing structure (primarily the inter-helix loops, but also some side chains) could be located and built was finally obtained by carrying out point-by-point additions of the independently phased maps, using the program *OVERLAPMAP* (*CCP*4, Collaborative Computational Project, Number 4, 1994). The SIR and iron anomalous maps were combined (correlation coefficient 0.20), as were the Wang phase-extended map and the iron anomalous map (correlation coefficient 0.17); these two combined maps were then themselves combined. The electron-density map that resulted from this combination of maps, at a resolution of 3.0 Å, finally allowed the inter-helix loops to be unequivocally traced, and several incorrectly built parts of the helices to be improved. Refinement then became straightforward.

2.4. *Refinement*

The initial model for *Alcaligenes sp* cytochrome *c'* was the molecular-replacement model, derived from the *Rhodospirillum molischianum* cytochrome *c'* model, with some rebuilding into the iron anomalous-phased electron-density map (Baker *et al.*, 1995). From this model, residues 33–38, 52–78 and 98–104 (the inter-helix connecting loops) were removed, together with a number of amino-acid side chains. These were re-included in the model by rebuilding into the combined maps described above, using the program *FRODO* (Jones, 1978), as the density became clearer during refinement.

The course of the refinement using the 1.8 Å Photon Factory data set throughout, is outlined in Table 3. Initial refinement was with *X-PLOR* 3.1 (Brünger, 1992), using the simulated-annealing protocols of Brünger, Kuriyan & Karplus (1987). The free R value was also monitored during this stage to check that real progress was being made. At the end of this phase of refinement the crystallographic R factor was 0.221 (free $R = 0.275$) for data in the resolution range 5.0–1.8 Å. Refinement was then continued with the restrained least-squares program *TNT* (Tronrud, Ten Eyck & Matthews, 1987). Restraints were placed on bond lengths, bond angles, planar groups, non-bonded contacts and temperature factors of neighbouring atoms; no correction for the solvent continuum was made. Omit maps were calculated at regular intervals to allow manual rebuilding of regions, with regions needing attention often being identified by their high B values. Solvent molecules (all regarded as water) were added conservatively with due regard for their environment including potential interactions with hydrogen-bond partners. The solvent model was comprehensively checked several times during refinement by omitting all water molecules that had high B values (>60 Å^2), or made either too-close contacts with each other or with protein atoms, or made no potential hydrogen-bonded contacts at all.

Table 3. *Course of refinement*

(*a*) *Alcaligenes sp* cytochrome c'

Round	R factor	Comments
1	0.304 → 0.297	Four rounds refinement with *X-PLOR*, model from molecular replacement, residues 2–122, all side chains included. Free $R = 0.345$ at end. Resolution range 5.0–1.8 Å.
2	0.297 → 0.255	Rebuilt inter-helix loops from combined maps (see text). Four rounds refinement with *X-PLOR*, residues 2–122. Free $R = 0.321$ at end. Resolution range 5.0–1.8 Å.
3	0.255 → 0.221	Rebuilt inter-helix loops from combined maps. Six rounds refinement with *X-PLOR*, residues 2–122. Free $R = 0.275$ at end. Resolution range 5.0–1.8 Å.
4	0.221 → 0.213	General rebuild of model from combined maps. Added residues 123–125. Refinement with *TNT* (*xyz* only), 15 cycles, resolution range 8.0–1.8 Å.
5	0.213 → 0.192	All side chains checked and 53 waters added from $2F_o - F_c$ and $F_o - F_c$ maps. Refinement with *TNT* (*xyzB*), 35 cycles, resolution range 8.0–1.8 Å.
6	0.189 → 0.183	Residue 1 and 66 waters added from $2F_o - F_c$ and $F_o - F_c$ maps. Refinement with *TNT* (*xyzB*), 40 cycles, resolution range 20.0–1.8 Å.
7	0.198 → 0.188	Check of all solvent structure, added 13 waters, removed 43 waters. Checked all side chains with $B > 50$ Å^2. Refinement with *TNT* (*xyzB*), 40 cycles, resolution range 20.0–1.8 Å.

(*b*) *Alcaligenes denitrificans* cytochrome c'

1		Rigid-body refinement with *TNT*, eight groups (i) Resolution 20.0–4.0 Å, 5 cycles, $R = 0.188$ (ii) Resolution 20.0–3.0 Å, 5 cycles, $R = 0.187$ (iii) Resolution 20.0–2.0 Å, 5 cycles, $R = 0.217$
2	0.217 → 0.185	Checked all side chains, corrected sequence (four residue changes), built residue 1 as pyroglutamic acid, added 103 waters. Refinement with *TNT* (*xyzB*), 30 cycles, resolution range 20.0–2.0 Å.
3	0.169 → 0.167	Revised water structure (removed 28 waters), checked whole model. Refinement with *TNT* (*xyzB*), 30 cycles, resolution range 20.0–2.15 Å.

Once the *Alcaligenes sp* cytochrome c', for which the amino-acid sequence was known (Ambler, 1973), had been refined, this refined model was used as starting model for refinement of the *Alcaligenes denitrificans* structure. This was expected to be appropriate because of the high degree of isomorphism between the two. Since no sequence is yet available for *Alcaligenes denitrificans* cytochrome c' electron-density maps were carefully scrutinized for indications of sequence differences, and where there was any doubt the side chain was cut back or removed completely. Refinement began with three rounds of rigid-body refinement using *TNT*. The molecule was split into eight rigid-body groups (four helices, three connecting loops and the haem) and the maximum resolution was extended from 3.0 to 2.0 Å in steps during the three rounds of refinement. Little movement occurred, and the groups were easily re-joined by building into an omit map. Conventional restrained least-squares refinement, with *TNT*, was then continued, and solvent molecules added as before.

The final refined model for *Alcaligenes sp* cytochrome c' comprises 956 protein atoms (one monomer; 125 residues and a haem group) and 89 water molecules. That for *Alcaligenes denitrificans* comprises 953 protein atoms and 75 water molecules. Atomic coordinates and structure factors have been deposited with the Brookhaven Protein Data Bank.*

3. Results and discussion

3.1. *Model quality*

Both refined models give extremely good agreement with the X-ray data, the final values for the crystallographic R factor being, for *Alcaligenes sp*, 0.188 for all data in the resolution range 20.0–1.8 Å (14 673 reflections, with no σ cutoff) and, for *Alcaligenes denitrificans*, 0.167 for all data in the range 20.0–2.15 Å (8220 reflections, with no σ cutoff). Luzzati plots (Luzzati, 1952) of the R factor as a function of resolution (Fig. 1) are consistent with average maximum coordinate errors of approximately 0.18 Å for both structures; a *SIGMAA* analysis (Read, 1986) gives similar values of about 0.16 Å. Both models conform with expected geometry; root-mean-square (r.m.s.) deviations from standard bond lengths and angles are 0.017 Å and 2.4°, respectively, for *Alcaligenes sp*, and 0.013 Å and 2.1° for *Alcaligenes denitrificans*. A Ramachandran plot (Ramakrishnan & Ramachandran, 1965) of main-chain torsion angles (φ, ψ) for *Alcaligenes sp* cytochrome c' is shown in Fig. 2, and is representative of both structures. There are no outliers, and 96% of residues fall into the 'most favoured' regions defined in the program *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993).

The entire polypeptide chain of *Alcaligenes sp* is represented by well defined electron density except for the two C-terminal residues, 126 and 127, which are missing from our final model. The amino-acid sequence for *Alcaligenes denitrificans* cytochrome c' has not been determined chemically, but in our model comprises residues 1–125, as for *Alcaligenes sp*. Several representative views of the final electron density are shown in Fig. 3. The N-terminal residue is in both cases pyroglutamic acid, with the side-chain carboxyl group covalently linked to the N-terminal α-amino group to give a cyclic structure, which is extremely clear in the electron density (Fig. 3*c*). Even the inter-helix connecting loops, which had proved so difficult to model originally, are now well defined (Fig. 3*b*); evidently the difficulties arose from inadequacies in the original phasing and from the structural differences in these regions from the molecular-replacement search models, rather than from a lack of definition in the structure. Very few of the 13 Lys residues are involved in stabilizing interactions (see later) and many of these side chains are

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1CGN, R1CGNSF for *Alcaligenes sp*, and 1CGO, R1CGOSF for *Alcaligenes denitrificans*). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: LI0204).

poorly ordered; in the final model for *Alcaligenes sp* cytochrome *c'* the side chains of Lys4, Lys31, Lys41, Lys49, Lys104 and Lys117 are incomplete.

3.2. *Overall structure*

The cytochrome *c'* monomers for both *Alcaligenes sp* and *Alcaligenes denitrificans* have the expected four-helix bundle structures, described in detail below. In both cases the functional dimer is created by a crystallographic twofold axis; the symmetry operation which relates the monomer coordinates to the other half of the dimer is $y-x,\ y,\ \frac{1}{2}-z$).

In keeping with the close isomorphism of the crystals and the apparently very high level of identity of the two sequences (see below), the three-dimensional structures of the two species of cytochrome *c'* are extremely similar. When the *Alcaligenes sp* and *Alcaligenes denitrificans* structures are superimposed, the r.m.s. deviation is 0.17 Å for C_α atoms and 0.36 Å for all atoms. This is in spite of the fact that the starting model for refinement with the *Alcaligenes denitrificans* data (which was the final refined model for *Alcaligenes sp*) was first subjected to rigid-body refinement in which the structure was broken up into eight groups, allowing ample scope for readjustment.

3.3. *Amino-acid sequence*

The crystallographic refinement has confirmed the published amino-acid sequence of *Alcaligenes sp* cytochrome *c'* (Ambler, 1973), including the blocked N-terminus (Fig. 3*c*).

In the *Alcaligenes denitrificans* structure only four sequence differences have been detected; Thr27 is modelled as Ala, Pro61 as Ala, Ser78 as Gly and Ala122 as Ser. All of these positions are highly variable in other cytochrome *c'* sequences. Of course, it is not possible to exclude the possibility that some of these side chains could be longer, but disordered; residues 4, 31, 41, 49, 104 and 117 are again modelled as incomplete Lys residues. Moreover, even at the relatively high resolution of the present structure analysis some side chains simply cannot be distinguished from the X-ray data alone. Nevertheless, it seems certain that the sequence identity must be greater than 90%, consistent with the high isomorphism of the two crystal structures.

3.4. *The four-helix bundle*

The four helices, labelled *A* to *D*, form a classic four-helix bundle, and account for almost all of the secondary structure in the cytochrome *c'* monomer (Table 4). At the top of the molecule, in the orientation shown in Fig. 4, the helices are joined by very short loops; the *A–B* loop comprises only five residues (33–37), and the *C–D* connection can hardly be described as a loop at all, since

Table 4. *Secondary structure*

Structural element	Residues	Comment
Helix *A*	4–33	α_N configuration (N-terminus) α_{C2} configuration (C-terminus) Kink around Pro28
Helix *B*	37–54	α_{C1} configuration (C-terminus)
Helix *C*	75–100	α_{C1} configuration (C-terminus)
Helix *D*	101–124	
B–C loop	55–75	55–59 3_{10}-helix 60–63 type II turn 65–68 type II' turn 69–75 $3_{10}/\alpha$-helix

Gly100 is the last residue of helix *C* and Asp101 the first of helix *D*. At the bottom of the molecule, however, the *B–C* loop is long (21 residues) and helps to provide some of the packing round the haem (see below). This combination of short loops at one end, and a long loop at the other end, is associated with the splaying out of the helices to create a haem-binding crevice at the wider end.

The four helices contain a number of features that are commonly found in globular proteins (Baker & Hubbard, 1984). Helix *A* begins with an α_N configuration, with two 1–4 hydrogen bonds, 4O···7N and 5O···8N, and terminates with an α_{C2} configuration, marked by 1–6, 2–5 hydrogen bonding (28O···33N, 29O···32N) and a residue (Gly32) with the left-handed α configuration. It also has a distinct kink, caused by the presence of Pro28, which bends the helix through an angle of 30° and results in the loss of four hydrogen bonds (the C=O groups of residues 22–25 are not involved in intrahelical hydrogen bonding). Both helix *B* and helix *C* tighten towards 1–4 hydrogen bonding near their C-termini, this being another characteristic termination mode, labelled α_{C1} by Baker & Hubbard (1984).

Apart from the changed hydrogen-bonding patterns at their N- and C-termini, and the Pro-associated kink in helix *A*, the four helices are remarkably regular. The 1–5 hydrogen-bond lengths are extremely consistent. The plot in Fig. 5 suggests that there are slight bends in helix

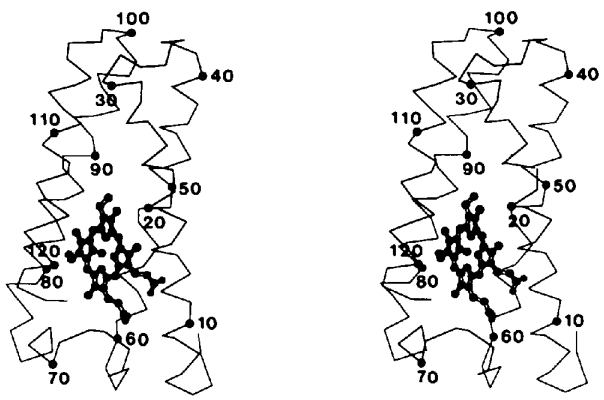

Fig. 4. Stereo C_α plot of the *Alcaligenes sp* cytochrome *c'* monomer, with the haem group added in a ball-and-stick representation.

B, at around residue 45, and in helix *D* around residues 117 and 120, the latter two associated with haem binding; apart from these the average hydrogen-bond length is 2.97 (0.13) Å. Similar regularity, with low standard deviations is found, in the (φ, ψ) angles, which average [63.2 (6.2)°, 40.2 (8.1)°] over the four helices (78 residues, omitting only the termini and the kink in helix *A*).

One other feature of the helices deserves comment. Each of them has a negatively charged residue at its N-terminus, presumably adding stabilization by interaction with the positive charge of the helix N-terminus (Hol, van Duijnen & Berendsen, 1978; Nicholson, Becktel & Matthews, 1988). Helix *A* has Glu6, which hydrogen bonds with its own peptide NH, and Asp7, which hydrogen bonds with the peptide NH of residue 4, helix *B* has Asp37, which hydrogen bonds with the peptide NH of residue 40, and helix *D* has Asp101 which hydrogen bonds with the peptide NH of residue 104; these are all typical 'N-cap' hydrogen-bonding interactions, side chain $n\cdots NH(n-3)$ or $n\cdots NH(n+3)$ (Baker & Hubbard, 1984; Richardson & Richardson, 1988; Presta & Rose, 1988). Helix *C* has Asp75 at its N-terminus and although the carboxyl group does not make a geometrically favourable hydrogen bond, mutagenesis of other systems has shown that the charge interaction is the more important effect (Nicholson *et al.*, 1988).

Apart from the four helices the only other secondary-structural elements are found in the long *B–C* loop. This loop is quite tightly organized, beginning with a short 3_{10}-helix (55–59), followed by two β-turns both of which contribute to interactions with the haem group (see later) and finishing with another short helix (69–75) which is intermediate in character between 3_{10} and α.

3.5. *Side-chain hydrogen bonding*

Very few of the side chains in the cytochrome *c'* structure are involved in hydrogen bonds or ion pairs. Of those that are, most form local interactions, with residues close in the amino-acid sequence. This can be attributed to the high helical content, which favours the formation of hydrogen bonds in which a side chain interacts with a carbonyl O atom three or four residues earlier in the sequence (Baker & Hubbard, 1984; Gray & Matthews, 1984); Ser14, Thr17, Arg25, Thr27, Thr50, Thr52, Trp56, Thr63, Ser74 and Asn89 all form local hydrogen bonds of this type.

Only seven side chains make longer range, cross-linking interactions within the cytochrome *c'* monomer. Probably the most important of these are the network of hydrogen bonds linking Arg12, on helix *A*, to the haem *D* ring propionate and to the *B–C* loop. The full hydrogen-bonding potential of the guanidinium group is used, with NE hydrogen bonded to the propionate O atom, O1D, NH1 hydrogen bonded to the carbonyl O atoms of residues 58 and 63 and Thr63 OG1, all on the *B–C* loop, and NH2 hydrogen bonded to 63 O and 65 O, both also from the *B–C* loop (see Fig. 6). Gln13, adjacent to Arg12 on helix *A*, also hydrogen bonds to the *D* ring propionate. The only other cross-linking hydrogen bonds involve Asn44 on helix *B*, which interacts with His22 on helix *A* (OD1···22 ND1; ND2···22 O); Thr63 on the *B–C* loop, which interacts with Arg12 on helix *A* (OG1···12 NH2); Trp73, which helps stabilize the conformation of the *B–C* loop (NE1···60 O); and Lys92 on helix *C* and Asp111 on helix *D* which form a salt bridge linking the two helices.

3.6. *The haem environment*

The principal determinant of the cytochrome *c'* structure appears to be the packing of the helix bundle around the haem; the helices, together with the long *B–C* loop, provide more than 20 side chains that pack against, or interact with, the haem group (see Table 5). Only three of the helices, *A*, *C* and *D* are directly involved, however, and the lack of involvement of helix *B* may correlate with its lesser spatial conservation among cytochrome *c'* structures (see later). Thus helix *A* provides, from its first five turns, Val9, Arg12 and Gln13, Leu16 and Thr17, Met19 and Ala20, and Phe23; helix *C* provides, from its first three turns, Phe79, Lys82 and Gln83, and Phe86; and helix *D* provides, from its last three turns, the

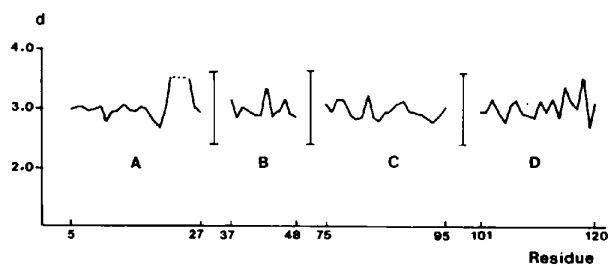

Fig. 5. Plot of potential hydrogen bond lengths (*d*) along each of the four helices. Distances are given in Å for $O(n)\cdots N(n+4)$ separations. Residue numbers refer to those of the C=O groups.

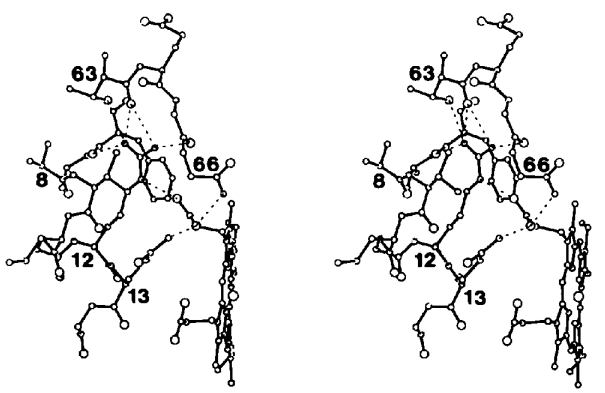

Fig. 6. Stereoview showing the environment of the conserved Arg12 side chain and the inner of the two haem propionate groups. Hydrogen bonds are shown with broken lines.

Table 5. *Haem packing contacts*

Residue	No. of contacts*	Closest contact (Distance, Å)	Residue	No. of contacts*	Closest contact (Distance, Å)
Val9	1	CG1···O1D (4.31)	Asp67	7	N···O2D (2.79)†
Arg12	7	NE···O1D (2.76)†	Ala68	7	CB···O1D (3.76)
Gln13	18	NE2···O2D (2.66)†	Ile72	1	CD1···CMD (3.59)
Leu16	27	CD1···C4D (3.42)	Phe79	4	CE1···CAD (3.76)
Thr17	3	OG1···O2A (3.21)	Lys82	1	CB···CBC (4.22)
Met19	2	CG···CBB (4.20)	Gln83	1	CB···CBC (3.97)
Ala20	3	CB···CMA (3.75)	Phe86	11	CG···CMC (3.48)
Phe23	5	CG···CBB (4.14)	Cys116	13	SG–CAB (1.81)‡
Trp56	16	CH2···C3C (3.59)	Cys119	13	SG–CAC (1.82)‡
Phe59	6	CE2···CMD (3.90)	His120	37	NE2–Fe (2.01)‡
Gly66	3	CA···O2D (3.84)	Arg124	24	NH1···C2A (3.48)

* Side-chain atom–haem atom contacts < 4.5 Å. † Hydrogen bonds. ‡ Covalent bonds.

Table 6. *Haem geometry*

	Alicagenes sp	*Alcaligenes denitrificans*
(*a*) Distances (Å)		
Fe–NA	1.97	1.94
Fe–NB	1.95	2.00
Fe–NC	1.99	2.01
Fe–ND	2.02	2.01
Fe–$N_{\varepsilon 2}$(120)	1.94	2.02
Fe···haem plane	0.32	0.32
Fe···pyrrole N plane	0.33	0.33
His120 $N_{\delta 1}$···water	2.92	2.93
(*b*) Other parameters		
Angle between imidazole plane and NA–NC vector (°)	56.8	55.2
R.m.s. deviation of atoms from haem plane (Å)	0.10	0.10

covalently bound Cys116, Cys119 and His120, as well as Arg124.

Many of the side-chain–haem contacts are hydrophobic, with the most extensive interactions made by Leu16, which covers the vacant sixth iron coordination site, and the aromatic side chains of Trp56, Phe59 (both from a short 3_{10}-helix of the *B–C* loop), Phe79 and Phe86. In addition the guanidinium group of Arg124 packs plane-to-plane with the *A* pyrrole ring, with presumed overlap of the positive charge of the arginine with the negative charge of the haem π system; the distance apart is approximately 3.7 Å.

As in the other cytochrome *c′* structures, the two propionate groups are buried to different degrees. The *D* ring propionate is the more buried and is stabilized by the hydrogen bonds from Arg12 and Gln13 described above and shown in Fig. 6. It also receives a hydrogen bond from the peptide NH of residue 67; this is made possible by the II′ configuration of the β-turn 65–68, which orients the 66–67 peptide so that 67 NH points towards the haem. The *A* ring propionate is much more solvent exposed and is hydrogen bonded primarily to water molecules, as well as to the side chain of Thr17; the latter interaction is probably weak, however, as the hydroxyl is also in a position to hydrogen bond to the main-chain carbonyl O atom of residue 13, suggesting a bifurcated interaction.

3.7. *Haem geometry*

The independent refinements of the two cytochrome *c′* structures have resulted in very similar haem geometries (Table 6); the atoms of the two haem groups, together with the axial His ligand, superimpose with an r.m.s. deviation of 0.09 Å. The Fe atom is in each case five-coordinate, as expected. Bonds to the four pyrrole N atoms average 1.98 and 1.99 Å in the two structures (range 1.94–2.02 Å), and the bond to the fifth (axial) ligand, His120, is essentially the same, averaging 1.98 Å. The Fe atom is displaced by 0.3 Å from the haem plane towards the His ligand. This displacement is remarkably constant through all the cytochrome *c′* structures, 0.32 Å in the two *Alcaligenes* proteins, 0.23 Å in *R. molischianum* cytochrome *c′* and 0.29 Å in *C. vinosum* cytochrome *c′*.

The His ligand coordinates iron through $N_{\varepsilon 2}$ in each case, as in other haem proteins. In contrast to the latter, however, the other imidazole N atom, $N_{\delta 1}$, does not hydrogen bond to a carbonyl O atom. Instead of hydrogen bonding to the peptide O atom of Cys116, as might have been anticipated [since side-chain $n\cdots O{=}C$ $(n-4)$ interactions are common in helices], $N_{\delta 1}$ is instead hydrogen bonded to a solvent molecule. The histidine orientation is such that the imidazole plane approximately bisects the angle between *cis* pyrrole N atoms; the angle with the NA—Fe—NC vector is close to 45° in both cytochrome *c′* structures. A significant influence on the orientation of the His120 imidazole ring may be the presence of Arg124 adjacent to it. The guanidinium group of Arg124 lies parallel to the haem plane, 3.7 Å from it, and is oriented such that the protons on NE and NH2 are directed towards the plane of the His120 imidazole ring in a manner suggestive of amino–aromatic hydrogen bonding (Burley & Petsko, 1986; Levitt & Perutz, 1988) (Fig. 7). The distances of 124 NE and NH2 from the imidazole ring are 3.5 and 3.4 Å, respectively.

The haem group does not deviate greatly from planarity, as a plane calculated through the entire haem group, excluding only the Fe atom and the propionate and thioether groups, shows an r.m.s. atomic displacement of only 0.10 Å. Consideration of the individual atomic displacements (Fig. 8), however, shows that it is slightly saddle shaped, with the *A* and *C* rings tilted down

(away from the His ligand), and the *B* and *D* rings tilted up.

The amino acid that occupies the site distal to the Fe atom, Leu16, makes a closest approach of 3.5 Å to the centre of the haem group, too close to allow a solvent molecule to bind in the axial sixth coordination position. The possible significance of this is discussed below.

3.8. *The dimer interface*

As in the other cytochrome *c'* structures, the dimer is created by the association of helices *A* and *B* of one molecule with the corresponding helices *A'* and *B'* of a second molecule. The twofold axis relating the two monomers then gives rise to an antiparallel four-helix bundle of helices *A*, *A'*, *B*, *B'* (Weber *et al.*, 1981). In the two *Alcaligenes* cytochromes *c'* the major part of the dimer interface lies between Tyr11 and His22 of helix *A* and Val47 and Leu54 of helix*B*. At the two ends of this region are strong hydrogen bonds between the side chains of Tyr11 and His22 (11OH···22'NE2 and 22NE2···11'OH). In between these residues solvent is totally excluded and the monomer–monomer interactions are primarily hydrophobic and marked by shape complementarity of the two surfaces. The clearest example of this involves Leu18, which packs between Tyr11' and Leu18' of the opposing monomer and fits into a hole created by the presence of a small side chain (Ala) at position 15' (*i.e.* on the intervening turn of helix). This is shown schematically in Fig. 9. Outside this central, solvent-excluded region the N-terminus of the polypeptide chain folds back such that it interacts with helix *A* of its own monomer (Phe2 packs against Ala8 and the aromatic ring of Tyr11) and helix *A'* of the opposing monomer (the N-terminal pyroglutamic acid side chain packs against the side chain of Arg25'). A number of well defined water molecules are found in this region of the interface. In general, most of the monomer–monomer contacts involve helix *A*, as is shown in Table 7, although Val47, Leu51 and Leu54 of helix *B* do contribute significantly to the central hydrophobic core of the interface.

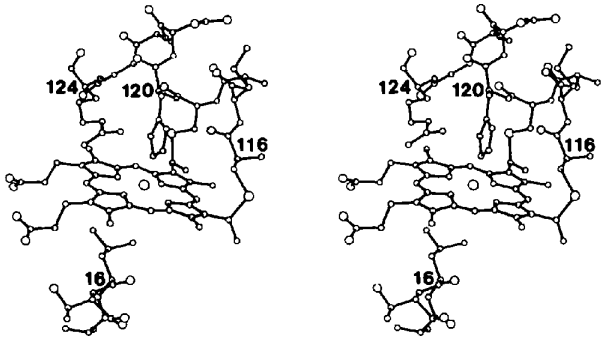

Fig. 7. Stereoview showing the arrangement of His120 and Arg124 with respect to each other and the haem group.

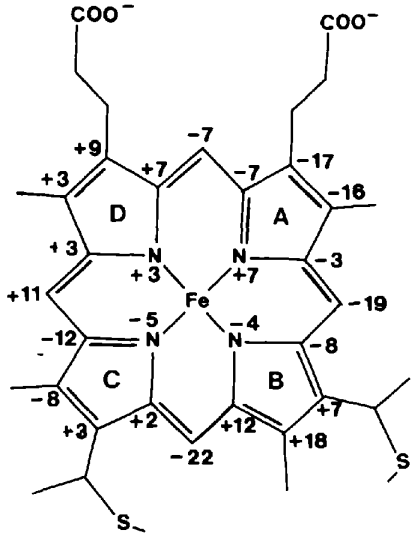

Fig. 8. Diagram of the haem group showing the deviations of individual atoms from the least-squares plane of best fit. The numbers give the deviations in Å × 100. A negative sign indicates a displacement on the same side as the axial His ligand. The pyrrole rings are labelled *A*, *B*, *C*, *D*.

3.9. *Temperature factors*

The cytochrome *c'* molecule as a whole is tightly organized and well ordered and this is reflected in the atomic *B* factors. The average *B* factor for main-chain atoms is 16.7 Å^2 for the *Alcaligenes sp* structure and a plot of the variation of *B* factors along the polypeptide chain (Fig. 10) shows that even the connecting loops between helices have relatively low values; only at positions 60–64 in the *B–C* loop, 74–80 at the start of helix *C*, 97–100 at the end of helix *C* and at the polypeptide chain N- and C-termini do the main-chain *B* factors rise above 20 Å^2.

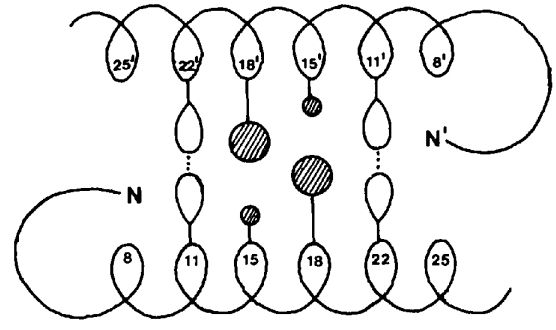

Fig. 9. Schematic representation of the interlocking packing of side chains between helices *A* and *A'* of the two monomers of the cytochrome *c'* dimer.

Table 7. *Monomer–monomer contacts*

Residue	No. of contacts*	Contacts with	Closest contact [Distance (Å)]
Pyr1	6	Arg25	CG···Arg25 NH1 (3.70)
Phe2	3	Ser21, Arg25	CE2···Arg25 NH1 (4.03)
Lys10	2	Thr17	CG···Thr17 CG2 (4.11)
Tyr11	25	Leu18, Ser21, His22, Arg25	OH···His22 NE2 (2.67)
Ser14	22	Ser14, Thr17, Leu18	OG···Ser14 OG (3.72)
Ala15	5	Leu18	CB···Leu18 CD1 (4.08)
Thr17	6	Lys10, Ser14	CG2···Ser14 OG (3.75)
Leu18	18	Tyr11, Ser14, Ala15	CD2···Tyr11 CD1 (3.42)
Ser21	6	Phe2, Tyr11	OG···Tyr11 CD1 (3.58)
His22	8	Tyr11, Leu54	NE2···Tyr11 OH (2.67)
Arg25	15	Pyr1, Phe2, Tyr11	NH2···Tyr11 CE1 (3.56)
Val47	2	Leu54	CG1···Leu54 CD1 (4.21)
Thr50	2	Thr50	CB···Thr50 CG2 (4.33)
Leu51	2	Leu51, Leu54	CD2···Leu51 CD2 (3.92)
Leu54	4	His22, Val47, Leu51	CD1···Val47 CG1 (4.21)

* Side chain···side chain contacts < 4.5 Å.

3.10. *Comparisons between cytochromes c′*

All of the cytochrome *c′* structures determined to date share the same basic fold. In addition to the four long α-helices that comprise the four-helix bundle this also includes a short 3_{10}-helix immediately following helix *B* and a second short 3_{10}-helix immediately prior to helix *C*, as well as two turns in the middle of the *B–C* loop where it makes contacts with the haem.

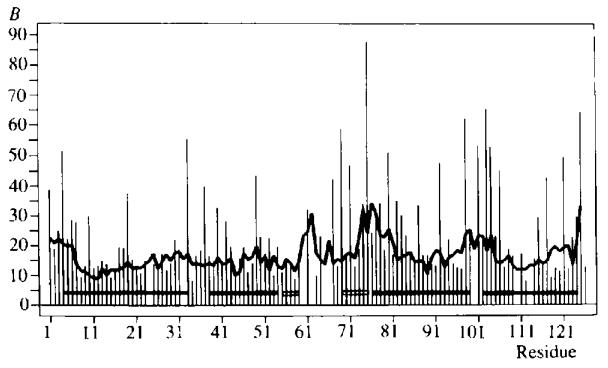

Fig. 10. Plot of crystallographic *B* values (Å^2) along the polypeptide chain of *Alcaligenes sp* cytochrome *c′*. Main-chain *B* values are shown by the solid line trace (indicating the average value for N, C_α, C, O for each residue) and average side-chain values are indicated by the vertical lines. The extent of α-helices and 3_{10}-helices are shown with solid and open horizontal bars respectively.

Table 8. *Superposition of helices*

Superpositions are based on C_α positions of all the helices plus the haem group. Individual r.m.s. deviations for atoms of each helix dissected out. *Alcaligenes sp* cytochrome *c′* is compared with three other species.

	C. vinosum R.m.s. (Å)	*R. molischianum* R.m.s. (Å)	*R. rubrum* R.m.s. (Å)
Helix *A*	0.90	1.34	1.06
Helix *B*	1.82	2.26	2.56
Helix *C*	0.68	1.45	1.35
Helix *D*	0.54	0.77	0.56
Overall	0.97	1.22	1.27

In spite of their folding similarities the cytochrome *c′* structures do not superimpose particularly well in three dimensions. The overall structural superposition of *Alcaligenes sp* cytochrome *c′* on to the other two high-resolution structures, those from *R. molischianum* (Finzel *et al.*, 1985) and *C. vinosum* (Ren *et al.*, 1993), is summarized in Fig. 11. In constructing the structural alignment, helices *A* and *D*, which show moderate sequence identity between species (Fig. 12), were used together with the haem to give an initial superposition, which could then be extended to the other two helices and finally to the connecting loops. The closest similarity is between *Alcaligenes sp* and *C. vinosum* structures, which match with an r.m.s. deviation of 1.37 Å for 113 C_α atoms (1.07 Å for 75 C_α if only the four helices, without their C-terminal residues, are included). Agreement is worse between *Alcaligenes sp* cytochrome *c′* and the two structures used in our molecular-replacement calculations (Baker *et al.*, 1995), *i.e.* those from *R. molischianum* and *R. rubrum*; r.m.s. deviations of 1.94 Å (109 C_α) and 2.07 Å (108 C_α), respectively, or 1.44 Å (75 C_α) and 1.56 Å (75 C_α) if only the α-helices are used. This undoubtedly accounted for some of the difficulties encountered in the molecular-replacement analysis.

The individual helices vary in their structural correspondence. If the molecules are superimposed as a whole, and the contribution of each helix is then dissected out, it can be seen that helices *A* and *D* match the best (Table 8). These are the two helices most closely associated with the haem, *D* through its three covalent linkages and *A* through its predominance in non-bonded contacts (Table 5); this makes the point that it is the packing of these helices against the haem that defines the core of the structure. Helix *B* matches least well and this is the helix which makes the fewest contacts with the haem. It is also involved in the generation of the dimer interface and it seems probable that the variations in the orientation of helix *B* are a major factor in the variations in the mode of association of the two monomers (Yasui *et al.*, 1992).

The sequence alignment based on structural correspondence (Fig. 12) shows few identical residues, only 12 (~10%) being totally invariant across the four species. Of these Cys116, Cys119 and His120 are involved in covalent links to the haem, Arg12 provides a network of internal hydrogen bonds, to the haem *D*

propionate and to the *B–C* loop, and Gly62 and Thr63 are on a conserved turn that interacts with Arg12. The others are two glycines, 100 and 113, Pro5 at the N-terminus of helix *A*, two internal residues on helix *C*, Leu93 and Ala97, and Lys117. Although within the haem-binding C*xx*CH sequence motif it is difficult to see why Lys117 should be conserved as it projects into the external solution and does not interact with any other protein atoms.

The lack of sequence identity between cytochromes c' can be traced to the four-helix bundle fold. This is defined by the interactions of hydrophobic residues, with each other and with the haem; almost all remain hydrophobic but there are evidently many combinations of such side chains that give similar packing and the same fold. Thus, almost none of the internal hydrophobic side chains are conserved because their interactions are not geometrically specific. A second feature of the four-helix bundle fold, at least in cytochromes c', is the lack of cross-linking hydrogen bonds. Hydrogen-bonded side chains, because of their more specific geometrical requirements, tend to be more highly conserved in homologous proteins. In cytochrome c' the vast majority of polar side chains either project into the external solvent or make local interactions, and there is thus little pressure for absolute conservation of their identity.

3.11. *Stabilization of the four-helix bundle*

In many respects the cytochrome c' structure fits the design principles advanced for the stabilization of four-helix bundles (Ho & DeGrado, 1987; Cohen & Parry, 1990). The polypeptide N-terminus is blocked and, therefore, does not introduce repulsion with the N-terminus of the first helix. All four helices have N-caps, and in each case the residue is negatively charged, for maximum stabilization. The first three helices all terminate with Gly or Pro residues. Most importantly, the distribution of hydrophobic residues in the helical wheel representation (Fig. 13) shows the role these play. All of the residues in positions *a* and *d* of the four helices are hydrophobic, with the exception of Arg12 and Gln83. The hydrophobic residues, which include the two Cys residues that provide covalent linkages to the haem group, are used in packing between the haem and the helices, and between the helices themselves. At the top of the molecule (as in Fig. 4) residues from all four helices, Phe23 and Met26 (helix *A*), Val41, Val45 and Leu48 (helix *B*), Leu93 and Ala97 (helix *C*) and Leu105 and Phe109 (helix *D*) pack together. Lower down the bundle, where the haem slots in as the helices splay out, the packing is about the haem and between pairs of helices. Of the two non-hydrophobic residues, Arg12 forms important internal hydrogen bonds with one haem propionate group and with the *B–C* loop, but Gln83 is oriented such that its side-chain amide projects into solution.

The distribution of residues at the remaining positions *b*, *c*, *e*, *f*, *g* shows that alanine residues, which presumably help to favour helix formation, given their propensity to appear in helices, are scattered liberally through all these sites. As regards larger hydrophobic residues, however, five are found in positions *c* and *g* of helices *A* and *B* (Leu18, Val29, Val47, Leu51, Leu54) and all except Val29 are involved in the dimer interface; only one other such residue (Val91) is found anywhere else round the four helices. Thus, the distribution of hydrophobic residues outside the core *a* and *d* positions creates a surface patch that strongly favours dimerization through helices *A* and *B* (see Fig. 13). It is not known whether dimerization has any functional significance, but it seems unlikely that it plays any role in stabilization of the basic four-helix bundle structure.

Although there are numerous charged residues in the external positions, *b*, *c*, *e*, *f*, *g* (five Asp, one Glu, eight Lys, two Arg) ion pairs are almost non-existent; there is just a single inter-helical one, Lys92–Asp111 (*c–c*), and no intrahelical ion pairs. Thus, specific ionic interactions do not seem to play any part in assembling the four-helix bundle. As suggested by Cohen & Parry (1990) the internal hydrophobic interactions and the influence of the dipolar nature of α-helices must be sufficient to drive the antiparallel packing of the helices and the formation of the four-helix bundle. The almost complete absence of cross-linking hydrogen bonds further strengthens this conclusion.

3.12. *Structural basis of iron spin state*

The present structures suggest a reason for the unusual electronic properties of proteins of the cytochrome c' family. These proteins, in their oxidized, Fe^{III}, state, have EPR and magnetic properties that are best explained by an admixture of intermediate-spin ($S = \frac{3}{2}$) and high-spin ($S = \frac{5}{2}$) states (Maltempo, 1974; Maltempo *et al.*, 1974). Theoretical considerations associate the intermediate-spin state with a critical weakening of the axial ligand field strength (Weber, 1982), but there has been no clear structural explanation for this effect in cytochromes c'. The Fe—N bond lengths are normal, and the displacement of the Fe atom from the haem plane is not significantly less than in other five-coordinate haem proteins. Weber (1982) has suggested that the weaker hydrogen bonding of the His ligand, to a water molecule rather than to a carbonyl O atom, could in turn weaken its bonding to the Fe atom.

In the two *Alcaligenes* cytochromes c' it is true that the His-bound water molecule is in each case poorly defined, with an elongated peak suggestive of positional disorder. The environment of the imidazole ring suggests another perturbing influence, however, in the presence of a positively charged residue, Arg124 in the present case, four residues (one turn of helix) removed from the His ligand. The guanidinium group of Arg124 interacts

intimately with both the haem group and the His120 imidazole ring (Fig. 7), the latter interactions resembling amino–aromatic hydrogen bonding (Burley & Petsko, 1986; Levitt & Perutz, 1988). The close interaction of this positively charged group with His120 must stabilize negative charge on the imidazole group substantially, and reduce its ability to denote electrons to the Fe atom. All cytochromes c' have a basic residue (Lys or Arg) in this position. Moreover the intermediate-spin state is pH dependent, with a change to pure high spin occurring at alkaline pH, the pK for this transition being between 7.1 and 9.0 (Bartsch, 1978). Thus we suggest that this pK may be associated with protonation of the residue (Arg or Lys) at position 124; the positively charged form, at

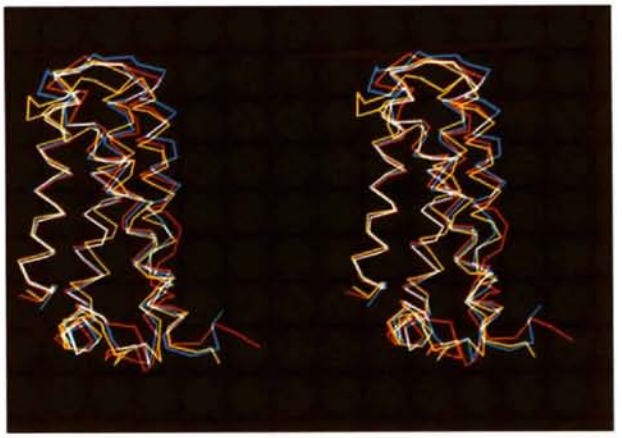

Fig. 11. Stereo C_α plot showing superposition of the cytochrome c' structures from *Alcaligenes sp* (blue), *Chromatium vinosum* (orange), and *Rhodospirillum molischianum* (yellow). The N-terminus of each polypeptide is on the right of the picture.

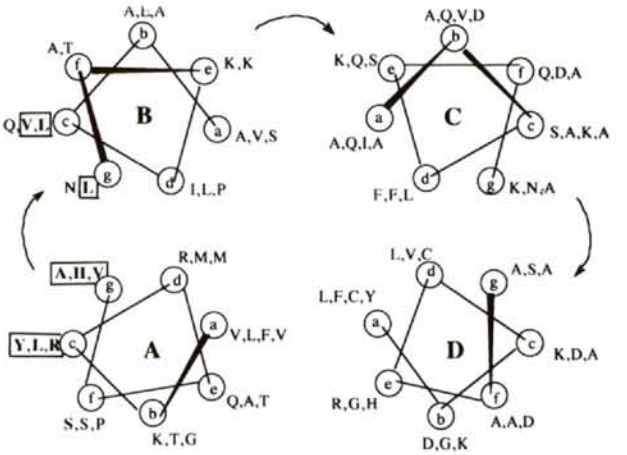

Fig. 13. Helical wheel representation of the four helices in *Alcaligenes sp* cytochrome c'. The external dimerization surface is indicated by boxed residues in bold face.

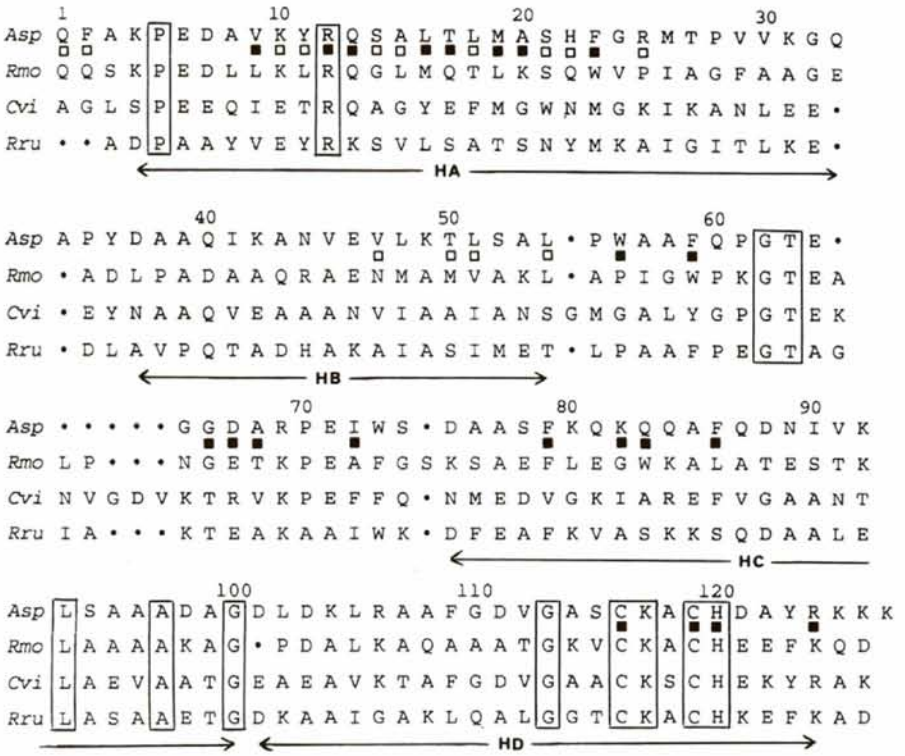

Fig. 12. Alignment of the amino-acid sequences of the cytochromes c' from *Alcaligenes sp* (*Asp*), *Rhodospirillum molischianum* (*Rmo*), *Chromatium vinosum* (*Cvi*) *and Rhodospirillum rubrum* (*Rru*). The extent of helices in *Asp* is indicated (HA, HB, HC, HD). Identical residues are boxed and those involved in haem or dimer packing are indicated with filled and empty squares, respectively.

neutral and lower pH, weakens iron binding by stabilizing negative charge on the adjacent His side chain.

Support is gratefully acknowledged from Massey University (award of a Graduate Assistantship to AJD), from the Howard Hughes Medical Institute through an International Research Scholar's award to ENB and from the New Zealand Lottery Grants Board. Thanks also to Dr Hale Nicholson for help in the early stages of the work.

References

Ambler, R. P. (1973). *Biochem. J.* **135**, 751–758.

Ambler, R. P., Bartsch, R. G., Daniel, M., Kamen, M. D., McLennan, L., Meyer, T. E. & van Beeuman, J. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 6854–6857.

Baker, E. N., Anderson, B. F., Dobbs, A. J. & Dodson, E. J. (1995). *Acta Cryst.* D**51**, 282–289.

Baker, E. N. & Hubbard, R. E. (1984). *Prog. Biophys. Mol. Biol.* **44**, 97–179.

Bartsch, R. G. (1978). *The Photosynthetic Bacteria*, edited by R. K. Clayton & W. R. Sistrom, pp. 249–279. New York: Plenum Press.

Brünger, A. T. (1992). *X-PLOR Version* 3.1, Yale University, New Haven, CT, USA.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.

Burley, S. K. & Petsko, G. A. (1986). *FEBS Lett.* **203**, 139–143.

Cohen, C. & Parry, D. A. D. (1990). *Proteins Struct. Funct. Genet.* **7**, 1–15.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cowtan, K. D. & Main, P. (1993). *Acta Cryst.* D**49**, 148–157.

Cusanovich, M. A. & Gibson, Q. H. (1973). *J. Biol. Chem.* **248**, 822–834.

Doyle, M. L., Weber, P. C. & Gill, S. J. (1985). *Biochemistry*, **24**, 1987–1991.

Finzel, B. C., Weber, P. C., Hardman, K. D. & Salemme, F. R. (1985). *J. Mol. Biol.* **186**, 627–643.

Gibson, Q. H. & Kamen, M. D. (1966). *J. Biol. Chem.* **241**, 1969–1976.

Gray, T. M. & Matthews, B. W. (1984). *J. Mol. Biol.* **175**, 75–81.

Hahn, K. W., Klis, W. A. & Stewart, J. M. (1990). *Science*, **248**, 1544–1547.

Hecht, M. H., Richardson, J. S., Richardson, D. C. & Ogden, R. C. (1990). *Science*, **249**, 884–891.

Higashi, T. (1989). *J. Appl. Cryst.* **22**, 9–18.

Ho, S. P. & De Grado, W. F. (1987). *J. Am. Chem. Soc.* **109**, 6751–6758.

Hol, W. G. J., van Duijnen, P. T. & Berendsen, H. J. C. (1978). *Nature (London)*, **273**, 443–446.

Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

Kassner, R. J. (1991). *Biochim. Biophys. Acta*, **1058**, 8–12.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Levitt, M. & Perutz, M. F. (1988). *J. Mol. Biol.* **201**, 751-754.

Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.

Maltempo, M. M. (1974). *J. Chem. Phys.* **61**, 2540–2547.

Maltempo, M. M., Moss, T. H. & Cusanovich, M. A. (1974). *Biochim. Biophys. Acta*, **342**, 290–305.

Meyer, T. E. & Kamen, M. D. (1982). *Adv. Protein Chem.* **35**, 105–212.

Nicholson, H., Becktel, W. J. & Matthews, B. W. (1988). *Nature (London)*, **336**, 651–656.

Norris, G. E., Anderson, B. F., Baker, E. N. & Rumball, S. V. (1979). *J. Mol. Biol.* **135**, 309–312.

Otwinowski, Z. O. (1991). In *Proceedings of the CCP4 Study Weekend*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington, England: Daresbury Laboratory.

Presta, L. G. & Rose, G. D. (1988). *Science*, **240**, 1632–1641.

Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Regan, L. & De Grado, W. F. (1988). *Science*, **241**, 976–978.

Ren, Z., Meyer, T. & McRee, D. E. (1993). *J. Mol. Biol.* **234**, 433–445.

Richardson, J. S. & Richardson, D. C. (1988). *Science*, **240**, 1648–1652.

Sakabe, N. (1991). *Nucl. Instrum. Methods Phys. Res. A*, **303**, 448–463.

Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Cryst.* A**43**, 489–501.

Wang, B. C. (1985). *Methods Enzymol.* **115**, 90–112.

Weber, P. C. (1982). *Biochemistry*, **21**, 5116–5119.

Weber, P. C., Howard, A., Xuong, N. H. & Salemme, F. R. (1981). *J. Mol. Biol.* **153**, 399–424.

Yasui, M., Harada, S., Kai, Y., Kasai, N., Kusunoki, M. & Matsuura, Y. (1992). *J. Biochem.* **111**, 317–324.