

Structure of Ferricytochrome *c'* from *Rhodospirillum molischianum* at 1.67 Å Resolution

B. C. Finzel, P. C. Weber†, K. D. Hardman‡
and F. R. Salem†

Protein Engineering Division, Genex Corporation
16020 Industrial Drive
Gaithersburg, MD 20877, U.S.A.

(Received 16 April 1985, and in revised form 3 August 1985)

The structure of ferricytochrome *c'* from *Rhodospirillum molischianum* has been crystallographically refined to 1.67 Å resolution using a combination of reciprocal space and restrained least-squares refinement methods. The final crystallographic *R*-factor for 30,533 reflections measured with $I > \sigma(I)$ between infinity and 1.67 Å is 0.188. The final model incorporates 1944 unique protein atoms (of a total of 1972) together with 194 bound solvent molecules. The structure has been analysed with respect to its detailed conformational properties, secondary structural features, temperature factor behavior, bound solvent sites, and heme geometry. The asymmetric unit of the cytochrome *c'* crystal contains a dimer composed of chemically identical 128-residue polypeptide chains. Although the refined structure shows the monomers to be very similar, examination of the differences that do occur allows an evaluation of how different lattice contacts affect protein conformation and solvent binding. In particular, comparison of solvent binding sites in the two subunits allows identification of a common set that are not altered by lattice interactions. The preservation of these solvent interactions in different lattice environments suggests that they play a structural role in protein stabilization in solution. The refined structure additionally reveals some new features that relate to the ligand binding properties and unusual mixed-spin state character of cytochrome *c'*. Finally, comparison of the heme binding geometry in cytochrome *c'* and other structurally unrelated *c*-type cytochromes shows that two alternative, but sterically favorable, conformational variants occur among the seven examples examined.

1. Introduction

The cytochromes *c'* are a class of proteins found in a wide variety of photosynthetic and denitrifying bacteria (Bartsch, 1978). The native protein is usually isolated as a dimer of identical $M_r \sim 14,000$ polypeptide chains, each incorporating a covalently bound protoheme IX prosthetic group. Cytochromes *c'* are physically characterized by their unusual mixed-spin spectroscopic properties (Maltempo *et al.*, 1974; Weber, 1982) and ability to bind carbon monoxide, but not oxygen, in the reduced state (Gibson & Kamen, 1966; Doyle *et al.*, 1985). Although the amino acid sequences of

cytochromes *c'* from numerous species have been determined, extensive homology is generally restricted to a region near the carboxy terminus of the polypeptide that incorporates the heme binding sequence (Ambler *et al.*, 1981). Cytochrome *c'* differs from the class of mitochondrial *c*-type cytochromes in heme iron ligation and spin state (pentacoordinate high-spin *versus* hexacoordinate low-spin heme ligands), sequence situation of covalent polypeptide heme attachments (carboxy terminus *versus* amino terminus), and overall backbone fold (4- α -helical bundle *versus* "cytochrome *c*" fold). Here we describe the results of the crystallographic refinement of *Rhodospirillum molischianum* cytochrome *c'* at 1.67 Å resolution. The results obtained provide confirmation of structural features obtained from earlier interpretations at lower resolution (Weber *et al.*, 1980, 1981) and additionally provide new insights into both specific aspects of cytochrome *c'* physical properties, and protein architecture in general.

† Present address: E. I. DuPont Experimental Station, Bldg. 328, Rm B25, Central Research and Development Dept, Wilmington, DE 19898, U.S.A.

‡ Present address: Department of Medical Genetics, Medical Sciences, Bldg. Rm 4379, University of Toronto, Toronto, Canada M5S 1A8.

2. Methods

(a) Data collection

R. molischianum ferricytochrome *c'* was crystallized from ammonium sulfate by vapor diffusion (Weber & Salemme, 1977). Crystals belong to the orthorhombic space group $P2_12_12_1$ ($a = 56.58 \text{ \AA}$, $b = 72.38 \text{ \AA}$, $c = 75.44 \text{ \AA}$) and contain one cytochrome *c'* dimer per crystallographic asymmetric unit. An entire data set containing reflections to 1.5 \AA resolution was measured from a single crystal ($\sim 1.0 \text{ mm} \times 1.0 \text{ mm} \times 0.7 \text{ mm}$) in 9 days using a multiwire area detector (Xuong *et al.*, 1978). The sealed tube X-ray source was operated at 40 kV and 28 mA with a 3° take off angle. The higher resolution intensities were measured first with the chamber 381 mm from the crystal and at an angle of 40° to the incident beam. A chamber angle of 20° was used to measure the lower resolution intensities. The data collection strategy was similar to that used for the multiple isomorphous replacement structure determination (Weber *et al.*, 1980). A total of 213,471 observations of 46,290 unique reflections were made. Overall, 72% of the theoretically observable reflections to 1.5 \AA resolution exhibit $I > \sigma(I)$ (see Table 1). An unexplained anomaly in the data is the abrupt increase in mean intensity (and the associated $\langle I/\sigma(I) \rangle$) found for the 2 highest resolution shells. Since this feature is atypical and could not be duplicated in our molecular F_c , data between 1.5 \AA and 1.66 \AA were omitted during the final stages of refinement.

(b) The 2.5 Å starting model

The structure of *R. molischianum* cytochrome *c'* was solved by multiple isomorphous replacement methods using 2 heavy-atom derivatives (Weber *et al.*, 1980). Reassessment of the heavy-atom data and some model building in $2F_o - F_c$ maps resulted in a 2.5 \AA resolution structure with a crystallographic *R*-factor of 0.44 (Weber *et al.*, 1981). With the exception of the amino and carboxyl termini, the majority of the polypeptide chain was clearly delineated in this electron density map. In general, helical regions were better defined than the interhelical portions of extended chain. However, electron density in the vicinity of the heme prosthetic groups was

somewhat distorted due to residual phase errors caused by a nearby heavy-atom site.

(c) Refinement

The general strategy for refinement was to alternate between successive cycles of reciprocal space least-squares minimization, and manual map analysis and model rebuilding. All model building was performed on an Evans and Sutherland model PS300 color graphics system supporting a version of FRODO (Jones, 1978) to display molecular fragments and electron density and manipulate them interactively. Electron density maps were computed using fast Fourier transform algorithms written by Ten Eyck (1973) from Fourier coefficients ($F_o - F_c$ and $2F_o - F_c$) weighted as suggested by Sim (1960).

The refinement proceeded in 2 stages. The first stage consisted of 35 cycles of refinement by the Agarwal procedure (Agarwal, 1978) in conjunction with geometric regularization with the Modelfit program (Dodson *et al.*, 1976). The Agarwal method is a least-squares refinement that incorporates a rapid and accurate method for calculating the density grid from an atomic model and uses fast Fourier transform algorithms to compute structure factors and derivatives for the least-squares normal matrix. Regularization steps followed cycles 7, 14, 18, 21, 24, 28, 32, and 35. Used in this combined manner, atoms of the model which were poorly fit were easily detected by their erratic shifts in both co-ordinates and/or temperature factors (Hardman *et al.*, 1982; Baker, 1980). Extensive manual map analysis and rebuilding were undertaken after cycles 24 and 35. Individual isotropic temperature factors were refined starting with cycle 3. During this stage, the crystallographic *R*-factor was reduced from 0.44 to 0.27, while the resolution of the data included was extended from 2.0 to 1.5 \AA .

For the second stage of the refinement, we combined the fast computational methods used by the Agarwal procedures (Agarwal, 1978) and the restrained parameter least-squares procedure of Hendrickson & Konnert (1980). This allows for the specification of restraints on atomic positions and thermal parameters to maintain molecular geometry during the optimization of observed

Table 1
Summary of diffraction data

D_{\min}	Possible	Fraction			$\langle I \rangle$	$\langle \sigma(I) \rangle$
		Measured	Observed	Measured		
3.70	3578	0.99	0.98	3539	2316.84	38.17
3.00	3006	1.00	0.99	6541	987.90	17.58
2.60	3408	1.00	0.97	9945	387.70	15.96
2.35	3436	0.99	0.96	13,363	236.28	17.33
2.20	2853	0.99	0.92	16,183	152.11	19.30
2.10	2385	0.99	0.86	18,541	111.82	20.25
2.00	2875	0.99	0.82	21,375	81.51	21.67
1.92	2743	0.99	0.79	24,090	58.25	22.65
1.84	3241	0.98	0.73	27,282	45.89	23.65
1.76	3876	0.98	0.66	31,076	38.41	24.93
1.70	3381	0.97	0.61	34,364	34.75	25.73
1.65	3203	0.97	0.59	37,486	33.51	26.41
1.60	3589	0.97	0.54	40,955	32.07	27.19
1.55	4107	0.96	0.54	44,898	37.93	27.55
1.50	4639	0.30	0.18	46,290	92.64	30.22

Measured reflections are those for which an intensity measurement could be distinguished from background noise. Observed reflections are those with intensities (*I*) greater than $\sigma(I)$.

and calculated structure factors. The resulting normal matrix has nearly all the components of the usual restrained parameter refinement and is solved for shifts in the atomic parameters using the same conjugal gradient procedure applied by Konnert (1976). In this application, central processor unit time on a VAX 11/780 for a single cycle was reduced from 16.5 h to 20 min.

Bond lengths, angles and other parameters used to define standard geometry for restraints were from the set referenced by Sielecki *et al.* (1979). To begin this stage of refinement, data limits were returned to 2.0 Å resolution, and then gradually increased to 1.67 Å resolution. The model was manually adjusted every 4 or 5 cycles to fix poor geometric features or reinterpret improved electron density maps. Elements of the structure which could not be placed in density were omitted from the model until the next round of map analysis, when they could usually be rebuilt into interpretable difference density. Overall, 44 additional cycles of least-squares refinement were completed. The last 8 cycles were completed using analytical structure factor algorithms (rather than fast Fourier transforms) in order to incorporate anomalous scattering contributions arising from the 10 sulphur and 2 iron atoms in each asymmetric unit.

Ordered solvent molecules, each modeled as a single central oxygen atom, were included throughout the refinement process if density greater than 4σ † appeared in difference maps in positions consistent with hydrogen bonding either to protein atoms or to other water molecules already included. In addition, the final model includes a solvent plateau model as devised by Bolin *et al.* (1982) to correct for scattering from disordered solvent between protein molecules. An arbitrary electron density value ($0.34 \text{ e}/\text{\AA}^3$) was assigned to points on a grid not occupied by protein or ordered solvent atoms. This plateau was then subjected to Fourier inversion to yield scattering factor contributions for each reflection. Prior to combination with the molecular scattering factors, the solvent scattering contributions were multiplied by both a linear scale factor and a Gaussian smoothing function ($K_s \exp(-B_s s^2)$) to minimize the final crystallographic residual (R_2) defined by Bolin *et al.* (1982). Optimum values for K_s and B_s found were 0.35 \AA^2 and 300 \AA^2 , respectively. The large Gaussian factor was somewhat surprising, as it effectively eliminates any significant contribution to all but the very lowest resolution reflections. However, incorporation of the solvent model does improve the fit between F_o and F_c for the low resolution data and was included for this reason.

3. Results of Refinement

(a) The final model

The molecular model obtained through refinement is nearly complete. Only the N-terminal glutamine and the side-chain of Gln2 of each of the peptide chains have been excluded, since these atoms are disordered and cannot be resolved in the final electron density map. One side-chain, Gln42, is resolved in two conformations with approximately equal occupancy, differing by 120° rotation about the $C\alpha-C\beta$ bond. This duality is conserved between the two monomers. Several other side-chains are

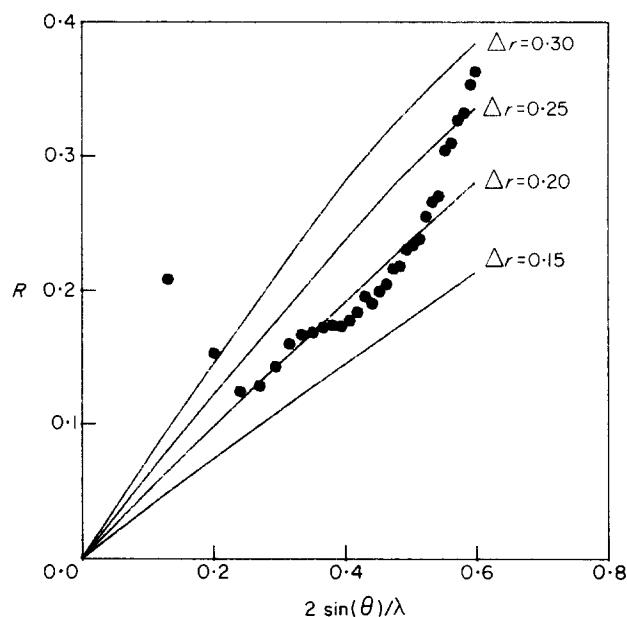


Figure 1. Mean discrepancy between F_o and F_c (filled circles) is plotted *versus* resolution and compared to discrepancies predicted on the basis of assumed specific mean co-ordinate errors (Δr). The best correlation between observed and predicted distributions is 0.20 Å.

poorly resolved in a single conformation, and have been assigned partial occupancy to reflect the reduced scattering capacity of these atoms. The resulting model contains 1944 unique protein atoms (of a total of 1972). Also included are 194 solvent molecules bound to the protein surface.

The final crystallographic R -factor, computed over all 30,533 reflections with $I > \sigma(I)$ between ∞ and 1.67 Å, is 0.188. For purposes of comparison with other refined proteins, we also report an R -factor for a best subset of the diffraction data with $I > 3\sigma(I)$. The R -factor for this subset of the cytochrome *c'* data (18,481 reflections) is 0.160, a value that could probably be lowered by further refinement incorporating only this data subset. Here we prefer a structural interpretation based on refinement against a more complete subset of the observed data (Table 2).

(b) Co-ordinate accuracy

To estimate the mean error in atomic positions, we have applied the method of Luzzati (1952) in which the discrepancy between F_o and F_c is plotted as a function of resolution, and compared to distributions anticipated from specific mean positional co-ordinate errors. As can be seen from Figure 1, there is poor correlation with any predicted distribution, particularly at high resolution. This discrepancy illustrates the failure of the Luzzati treatment to account for errors from other sources, such as data collection, and emphasizes that it provides an upper-bound estimate of the true mean co-ordinate error. The best correlation found by least-squares suggests a mean co-ordinate error of 0.20 Å using this method.

† σ is estimated as the standard deviation from the mean density value of the map.

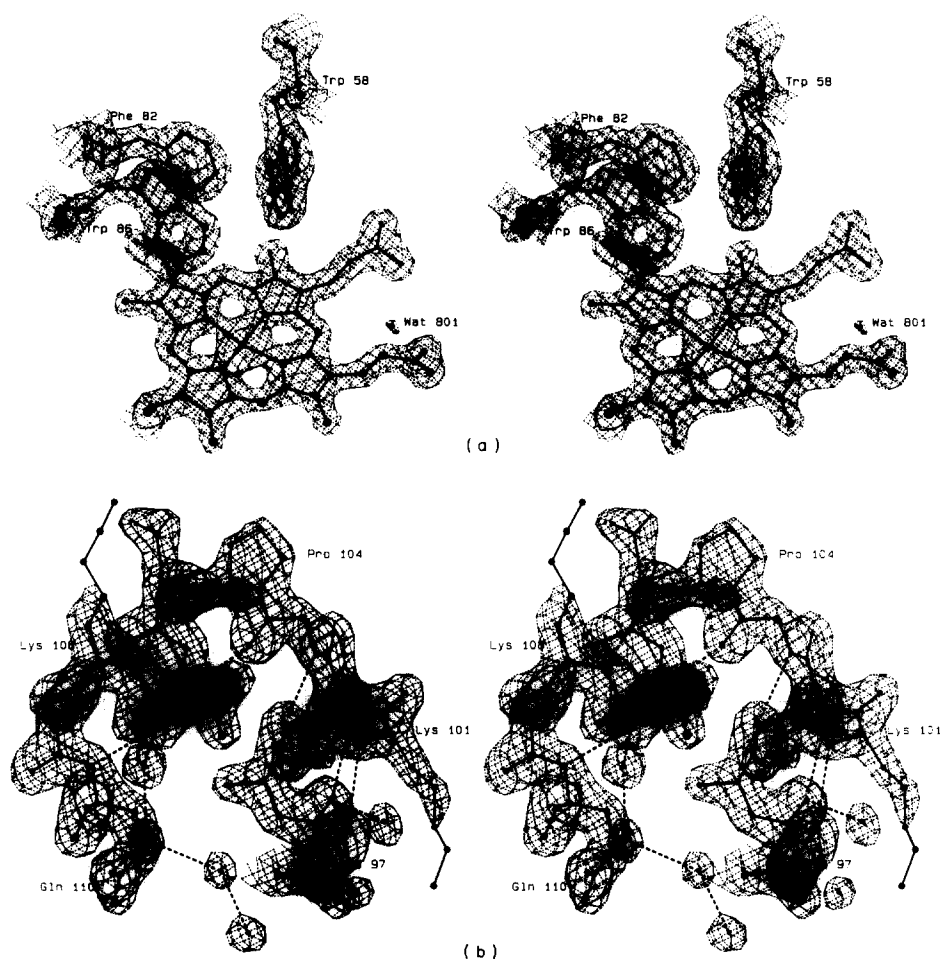


Figure 2. Representative sections of the refined 1.67 Å electron density map of cytochrome *c'*. (a) Monomer I heme density together with some adjacently packed aromatic residues. (b) The single residue chain reversal occurring between the *C* and *D* helices in monomer II, together with bound solvent localized at a lattice contact with an adjacent molecule in the crystal cell. Maps are contoured at 2.2σ and 1.8σ levels, respectively.

Table 2
Summary of least-squares refinement parameters

Reflections to 1.67 Å	36,396	
Reflections with $I > \sigma(I)$	30,533	
<i>R</i>	0.188	
	Target sigma	Final model
Distances		
1-2	0.030	0.030
1-3	0.050	0.051
1-4	0.050	0.059
Planes		
Peptides	0.030	0.027
Other	0.030	0.008
Chiral volumes	0.400	0.330
Non-bonded contacts		
1-4	0.500	0.192
Other	0.500	0.201

Final model value is the r.m.s. deviation from expected geometry. Sigmas are the inverse square root of the weight applied to this type of geometry restraint. Target sigma and r.m.s. deviations vary during the course of the refinement. All parameters given are those found at the end of refinement.

The notation 1-2, 1-3, 1-4 refers to atom pairs related through a bond, a bond angle, or dihedral angle, respectively.

As an alternative, an estimate of radial errors in atomic position was made from the diagonal components of an inverse, unrestrained least-squares normal matrix (Cruickshank, 1960). This method yields error estimates for individual atoms which vary regularly with increasing temperature factor (as expected), but gives an unreasonably small mean over all atoms of 0.03 Å. The discrepancy between this value and the 0.20 Å obtained by the Luzzati treatment illustrates the need for improved methods to estimate co-ordinate error from refined protein data. The true co-ordinate error is probably intermediate between the estimates produced by these alternative methods.

A summary of refinement statistics is presented in Table 2 as a further aid in evaluating the final molecular model. These statistics illustrate the extent to which the molecular model conforms to geometric requirements imposed in the form of restraints. Deviations between ideal and model geometrical features are generally small, and can be attributed to a random distribution of atomic displacements from idealized geometry whose average magnitude is consistent with the estimates of mean co-ordinate error described above.

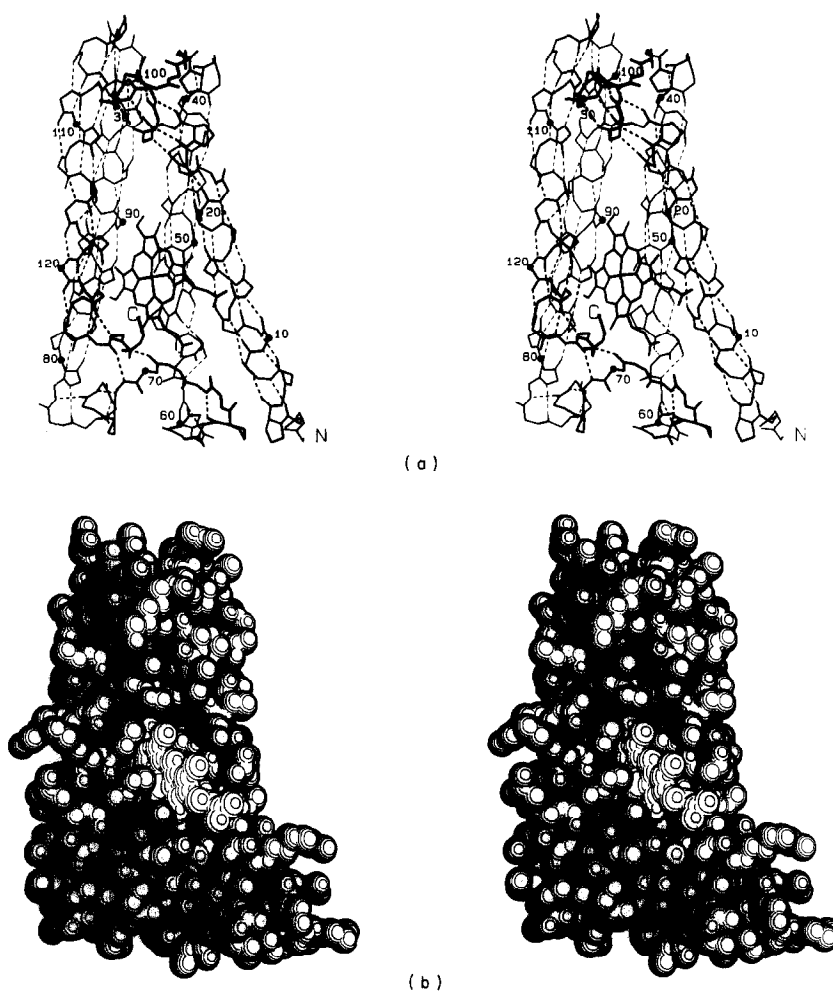


Figure 3. Stereoscopic views of cytochrome *c'* monomer geometry. (a) The backbone structural organization and hydrogen bond interactions common to both cytochrome *c'* monomers. (b) A corresponding space-filling representation that illustrates the overall packing and extent of heme exposure. Individual spheres represent joined hetero and attached hydrogen atoms (e.g. NH, CH₂, etc.) contoured with effective radii described by Shrake & Rupley (1973).

(c) *Comparison of multiple isomorphous replacement and refined model*

The r.m.s.† shift in atomic co-ordinates during refinement was 1.3 Å for all atoms and 0.9 Å for Cαs only, confirming that the initial m.i.r. model was essentially correct. As a result of refinement, individual monomers appear more structurally similar, so that the molecular dimer is more symmetric than inferred previously from the m.i.r. determination. The largest individual atomic shifts are to sections of extended chain connecting helices *A* and *B* in monomer I, and helices *B* and *C* in monomer II. Local conformational differences in these sections of chain have converged on refinement to produce a single peptide chain conformation that is nearly identical in both monomers. Figure 2 shows two sections of the final $2F_o - F_c$ map that are generally illustrative of the quality of the refined electron density map.

4. The Structure

Figure 3(a) shows a stereoscopic representation of the cytochrome *c'* monomer polypeptide backbone.

Each identical 128-residue monomer is comprised of four α -helices (*A* to *D*) that fold with sequentially adjacent helix axes oriented antiparallel to form a left-twisted four- α -helical bundle (Weber & Salemme, 1980). Helical bundles of this type are known to occur in a diverse group of proteins including the coat protein of tobacco mosaic virus (Bloomer *et al.*, 1978), the non-heme oxygen transport proteins hemerythrin (Stenkamp *et al.*, 1978) and myohemerythrin (Ward *et al.*, 1975), and the heme-containing electron transport cytochrome *b*₅₆₂ (Matthews *et al.*, 1979). Such bundles are well-suited for incorporation of prosthetic groups because of packing requirements peculiar to parallel arrays of α -helices (Weber & Salemme, 1980). In the cytochrome *c'* bundle, the helices spatially diverge from their point of closest approach at the bundle end incorporating the *A*-*B* and *C*-*D* helix connectors to create a sizable hydrophobic cavity for heme binding. The *B*-*C* connector appears to cap the divergent end of the bundle and shield one of

† Abbreviations used: r.m.s., root-mean-square; m.i.r., multiple isomorphous replacement.

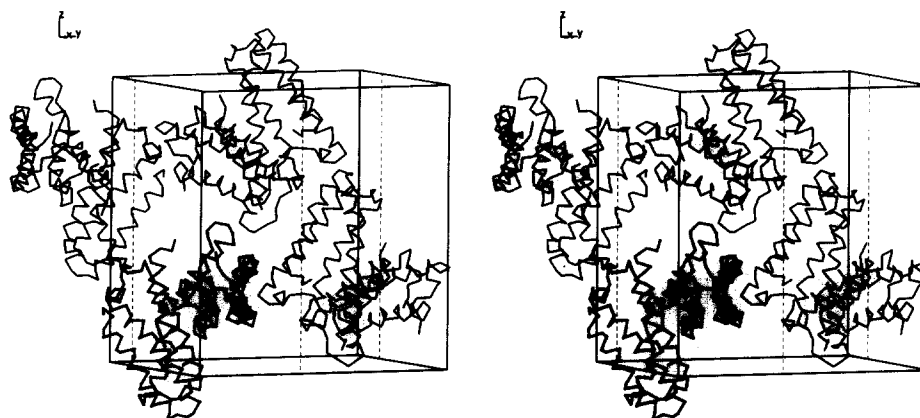


Figure 4. Cytochrome *c'* lattice packing. The illustration shows the contents of a unit cell (4 dimeric molecules) and their situations relative to alternate 2-fold crystallographic screw axes (broken lines) parallel to the *c*-unit cell edge. Monomers II form a zig-zag pattern of interactions along the *c*-axis screws. Each monomer I makes a native dimer interaction with a monomer II, and crystal contacts with monomers II aligned along adjacent screws.

the heme propionate groups from solvent (Fig. 3(b)).

Ferri-cytochrome *c'* monomers associate to form a stable dimer. The subunit interface is characterized by extensive hydrophobic interactions formed between side-chains of the first two helices (*A* and *B*) of each monomer (Weber *et al.*, 1981). These interactions produce a symmetrical "V-shaped" dimer whose subunits are related by a 2-fold rotation axis of symmetry.

(a) Crystal packing

R. molischianum cytochrome *c'* crystallizes with a functional dimer in the crystallographic asymmetric unit. Consequently, the two monomer subunits are situated in different crystalline environments in the $P2_12_12_1$ unit cell. As illustrated in Figure 4, one set of monomers (II) forms tightly packed head-to-tail arrangements aligned along alternate 2_1 screw axes parallel to the *c* unit cell direction. In contrast, monomers I are oriented roughly perpendicular to the *c*-cell screw axes owing to the "V" shape of the cytochrome *c'* dimer. This results in the appearance of large gaps between successive monomers I along the *c*-axis direction, so that they do not directly contact each other. Instead, the principal crystal contacts involving monomers I occur with monomers II on adjacent occupied 2_1 screw axes aligned parallel with the *c*-cell axis. The lattice is thus organized as an array of 2_1 -helices composed of cytochrome *c'* dimers, with monomers II forming the backbone of the 2_1 -helices aligned parallel to the *c*-cell axis, and monomers I forming the connections between the dimer helices in the *ab* crystal planes. The differences in subunit environment that result from this packing arrangement provide an opportunity to assess effects of lattice interactions on protein conformation and solvent organization.

(b) Dimer symmetry

Despite differences in crystalline environment, the cytochrome *c'* subunits are structurally very

similar. Least-squares superposition of the two monomers produces an r.m.s. difference in C α position of 0.4 Å (0.7 Å for all atoms) with only three C α s (Gln2, Ser3 and Asp128) differing by more than 0.6 Å. As noted above, the precise conformation of both N and C-terminal residues remains tenuous even after refinement, so that the observed differences between these groups are comparable to the estimated co-ordinate errors. Neglecting these residues, the r.m.s. difference in C α positions for the remaining 124 residues is reduced to 0.2 Å (0.5 Å, all atoms). Only eight side-chains have markedly different conformations (>1.0 Å r.m.s.) in the two subunits. In three cases (Lys60, Lys116 and Lys119), the side-chains in both monomers are disordered in solvent so that positional differences are again comparable to estimated co-ordinate errors. In the other five cases (Glu81, Glu84, Lys87, Thr91, Glu124), the electron density for the side-chains is well-resolved, and the observed differences in conformation are the result of alternative crystal contacts formed by the monomers in their different lattice environments.

The similarity between the two monomers is also reflected in the behavior of the individual isotropic temperature factors obtained during the refinement. Temperature factors are incorporated during refinement of the molecular model to account for the reduced scattering power of atoms that undergo thermal motion or are statically disordered in the crystal. The temperature factor behavior of refined proteins is of interest because it can serve as a probe of molecular dynamics in solution. Figure 5 plots the mean isotropic temperature factor (*B*-value), averaged over all main-chain atoms, versus amino acid sequence position in both subunits. There is very good correlation (0.82) in the overall temperature factor behavior between the two monomers. As anticipated, *B*-values tend to be lower in extended sections of ordered secondary structure and higher in the less-ordered loops connecting them. Also notable is the periodic variation in *B*-value along the helical segments (e.g.

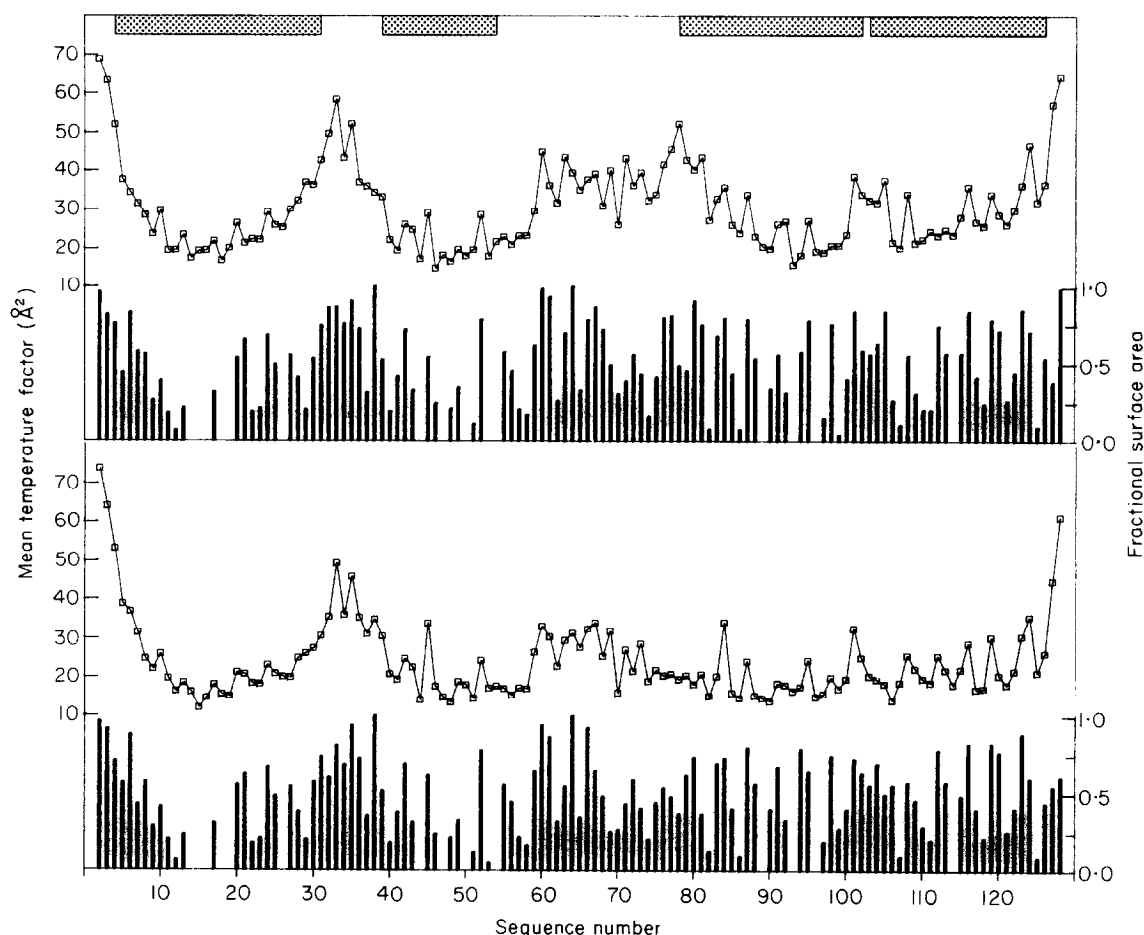


Figure 5. Temperature factor and lattice accessible surface area variation in cytochrome *c'* monomers I (upper) and II (lower). The mean temperature factor behavior is indicated by the connected curve, while vertical bars indicate fractional accessible surface areas of each residue in their different lattice environments. The shaded bars along the top of the plot indicate α -helical regions of the structure.

helix A) where residues exposed to the molecular surface exhibit higher *B*-values than their solvent inaccessible neighbors. Following the lead of others (Sheriff *et al.*, 1985; Higuchi *et al.*, 1984; Kamphuis *et al.*, 1984), we have plotted the fractional solvent-accessible surface area (the fraction of the surface area accessible relative to a Gly-X-Gly tripeptide reference (Shrake & Rupley, 1973)) for each residue, taking into account the presence of molecules related by crystal symmetry which might prevent access by a 1.4 Å probe sphere. We find good correlation between lattice-accessible surface area (0.67 and 0.61 for monomers I and II, respectively) with *B*-value variation.

(c) Hydrogen bonding

Hydrogen bonds between protein atoms are summarized schematically in Figure 6. Interactions between potential partners were initially identified as hydrogen bonds if donor and acceptor atoms were less than 3.4 Å apart, and if their interaction exhibited qualitatively acceptable linearity. These preliminary assignments were compared (in the case of main-chain-main-chain hydrogen bonds) to those obtained using the algorithm of Kabsch & Sander

(1983) obtained from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977), which computes an interaction energy based on the distance between potential hydrogen bond partners and their angle of interaction. Several additional interactions were added to our original list as a result of these computations, though none was eliminated. A quantitative assessment of hydrogen bond strengths between side-chain atoms was not possible, since hydrogen atom positions have not been included in the molecular model. Some hydrogen bond distances of up to 3.6 Å have been included in the final list if there is good hydrogen bond geometry and if the donor and acceptor participate in no other hydrogen bonds. While these longer interactions could arguably be excluded from the hydrogen bond list, they may favorably contribute to the stability of the protein and have been included for this reason.

Table 3 summarizes the assignment of secondary structural features of cytochrome *c'*, as determined from analysis of hydrogen bonding patterns and backbone torsion angles. Designated α -helical segments extend from the residue contributing the first carbonyl oxygen $O_{(n)}$ for hydrogen bonding with $NH_{(n+4)}$ through the residue bearing the last amide

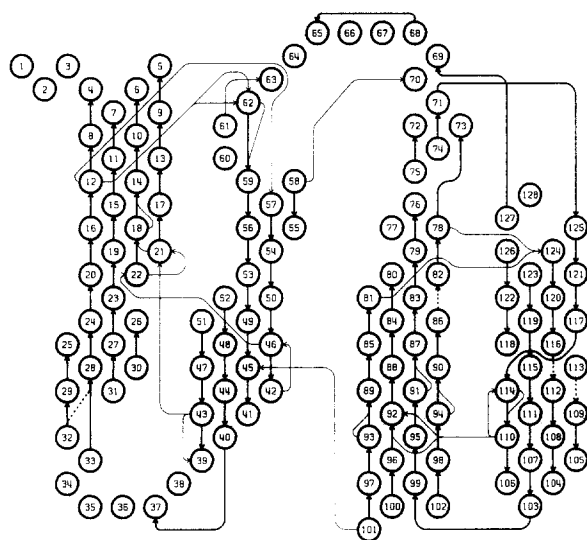


Figure 6. Schematic diagram of cytochrome *c'* hydrogen bonds. Hydrogen bonds are represented by arrows from the donor (e.g. peptide NH) to the acceptor (e.g. peptide carbonyl oxygen). Arrows originating or terminating in vertical line segments involve participation of backbone amide groups. Similarly, horizontal segments designate hydrogen bonding through side-chains. Broken lines correspond to long hydrogen bonds as described in the text. Heme interactions are not included. There are no intermonomer hydrogen bonds in cytochrome *c'*.

nitrogen in sequence. Four helical segments result; the longest being 26 residues. The average main-chain torsion angles (ϕ , ψ) computed over residues in α -helices are -65° and -41° , consistent with values found in other proteins refined at high resolution (Baker, 1980; Finzel *et al.*, 1984; James & Sielecki, 1983). We suggest that the database of refined structures is now sufficiently large to justify abandoning the "ideal" helix conformation (-57° , -47°) suggested by Arnott & Dover (1967) in favor of one (perhaps -64° , -40°) more consistent with observation.

Two abnormalities in the α -helices should be noted. In our earlier description of the structure (Weber *et al.*, 1981), helix *A* was described as terminating with proline at position 25. However, the refined structure clearly shows that helix *A* extends another turn to residue 30, but is bent at proline 25 by approximately 25° (Fig. 3). Since the bend incurs little irregularity in expected α -helical torsion angles (see below), the region comprising residues 5 through 30 seems better described as a single bent α -helix. A second irregularity occurs in the middle of helix *D* where an almost imperceptible change in helix pitch allows for a 3_{10} -helical hydrogen bond between Thr114 and Val117, thereby disrupting the normal interactions of the intervening residues in the chain. The basis for this disruption is unclear, but it may be a distortion required to optimize protein-heme interactions.

Two short 3_{10} -helical segments have also been identified. The longest provides a gradual termination to α -helix *B*, while the other represents a single

3_{10} -helical turn which reverses chain direction prior to the initiation of helix *C*.

Refinement at high resolution also permits the confident assignment of non-repetitive secondary structure (hairpin turns) summarized in Table 3. These are classified according to the nomenclature of Lewis *et al.* (1973) whose expansion upon the original set of six explored by Venkatachalam (1968) enables the categorization of all turns for which the distance between $C\alpha_{(n)}$ and $C\alpha_{(n+3)}$ is less than 7.0 \AA . Two such turns fall into the miscellaneous category IV used to include all turns not otherwise classifiable. These turns are found in a contiguous section of chain which makes considerable contact with the heme propionates and other protein side-chains *via* Thr61 and may be stabilized in this unusual conformation as a result. Turns of type II prefer glycine at position 3, but one turn reported here (Leu-Pro-Asn-Gly, residues 65 to 68) violates this restriction despite near perfect correlation to ϕ , ψ angles that define this turn type.

(d) Conformational torsion angles

Main-chain conformational torsion angles are summarized independently for both monomers in Figure 7 using a representation devised by Balasubramanian (1977). The presentation of torsion angles (ϕ and ψ) as directional vectors allows easy identification of secondary structural features. Residues involved in α -helices exhibit short ϕ to ψ vectors centered around -50° while extended sections of chain are shown as much longer lines, occasionally spanning the length of the ordinate. One feature that stands out in this representation is the very short and smooth transition between helices *C* and *D* (Fig. 2(b)). Only a single amino acid (Gly103) is required to connect

Table 3
Secondary structure in cytochrome *c'*

A. Helices			
Helix	Type	Residues	Length (residues)
A	3-6/13	5-30	26
B	3-6/13	40-53	14
C	3-6/13	79-102	24
D	3-6/13	104-125	22
B+	3-0/10	54-58	5
C-	3-0/10	72-74	3
B. Hairpin turns			
Type	Residues	Sequence	Hydrogen bond
I	29-32	Phe-Ala-Ala-Gly	29 O-N 32
I	37-40	Pro-Asp-Asp-Ala	37 O-N 40
I	56-59	Ile-Gly-Trp-Ala	56 O-N 59
II	59-62	Ala-Lys-Gly-Thr	59 O-N 62
IV	60-63	Lys-Gly-Thr-Glu	
IV	62-65	Thr-Glu-Ala-Leu	
II	65-68	Leu-Pro-Asn-Gly	65 O-N 68
I	73-76	Glu-Ala-Phe-Gly	
I	76-79	Gly-Ser-Lys-Ser	76 O-N 79

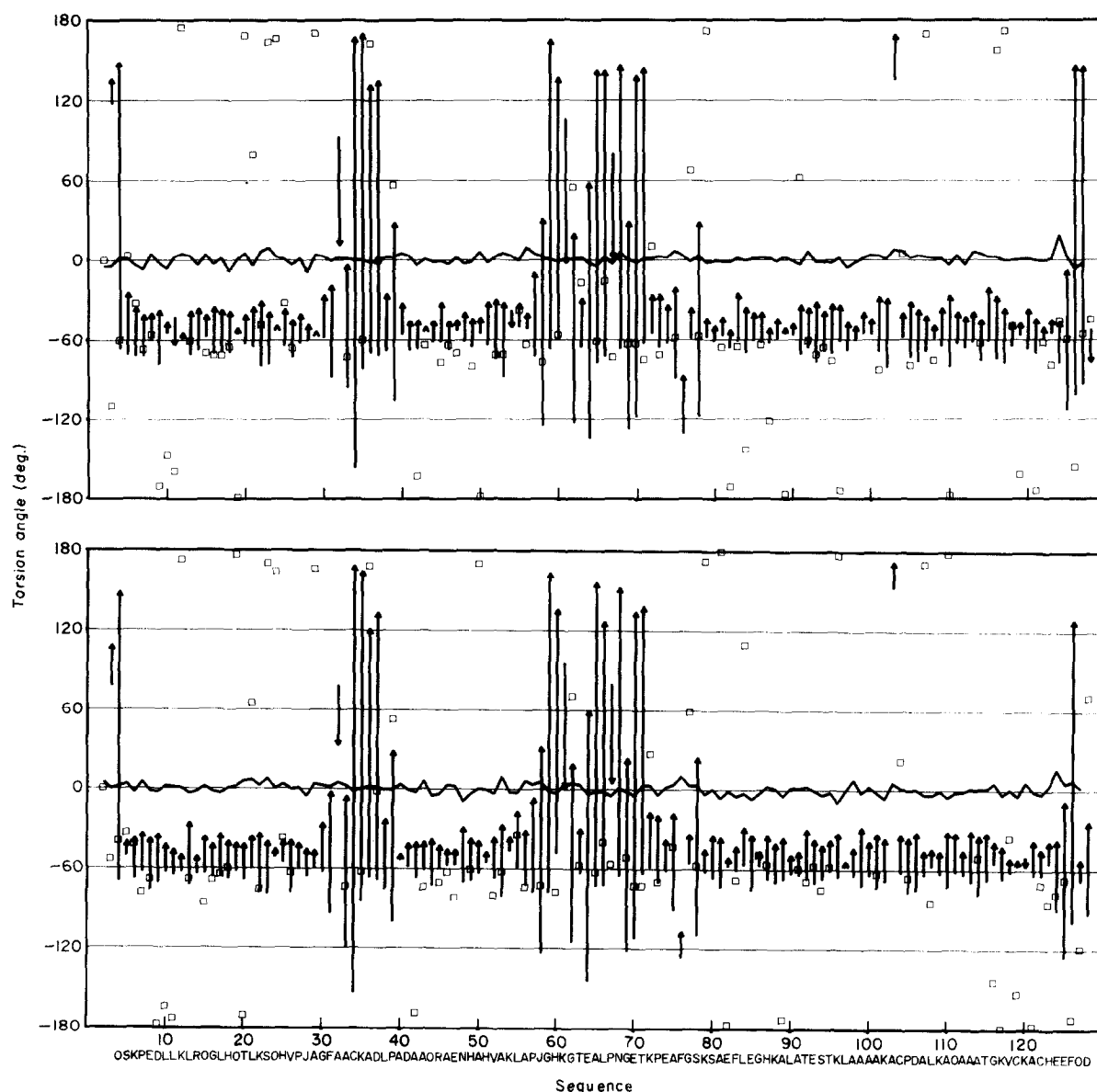


Figure 7. Summary of refined conformational parameters for monomers I (upper) and II (lower) of cytochrome *c'* plotted as a function of sequence position. The line oscillating about 0° in each plot indicates the refined deviations in peptide bond planarity ω . The vertical arrows plot ϕ - ω backbone conformational values for each residue. Squares indicate side-chain χ_1 torsional angle values. Torsional angle values or differences are given in degrees.

these two helices, despite the fact that there is a 180° reversal in the direction of the chain. Our search of the Brookhaven Protein Data Bank indicates that this is the only example of a turn of this type connecting two helices in such compact fashion. The turn may be possible because of the high percentage of alanine residues involved in the adjacent helix contacts that allow an unusually short interhelix connection. In this context, it is interesting to note that while all known sequences of cytochromes *c'* have a glycine at position 103, only *R. molischianum* has alanine residues at all the key interhelix contact points (positions 99, 102 and 106) necessary to complete this turn. In other sequences, the turn is accomplished by insertion of an additional amino acid following Gly103.

None of the non-glycine amino acids represented (excluding poorly located N-terminal residues) are

found in conformations far outside of the allowed regions originally defined by Ramakrishnan & Ramachandran (1965). Only one residue (Asn67), with $\phi = 80^\circ$ and $\psi = 0^\circ$, lies near the shallow potential minimum allowed for a left-handed α -helix. This residue falls in the midst of a relatively common type II reverse turn (see above).

Deviations from peptide bond planarity obtained on refinement are shown as the jagged line oscillating about 0° in Figure 7. The variations in peptide bond torsion angle ω from the expected value (180°) range in magnitude from 0 to 18° and produce a nearly Gaussian distribution of ω values centered about -178° . The magnitudes of the refined ω values depend strongly upon the weight given peptide bond planarity restraints during the refinement process. However, the direction of displacement may reflect stress placed upon the

bond by contiguous chain segments. This suggestion is supported by the significant correlation (0.4) obtained on comparing the variations in ω versus sequence position for the two independently refined monomers. Presumably, the deviations would be uncorrelated if they resulted only from random co-ordinate error. Although the amplitude of the ω angle deviations can be damped by weighting planarity restraints more heavily during the least-squares refinement, relaxing these restraints further did not result in a significant increase in the amplitude of the variations, or reduction of the *R*-factor. This suggests that the magnitude of the ω variations are realistic as modeled.

Figure 7 also illustrates the variation in values of the first side-chain torsion angle (χ_1), as defined for side-chains extending beyond $C\beta$. Previous work has shown that there is a pronounced tendency for amino acid side-chains to assume staggered conformations with $\chi_1 \pm 60^\circ$ or 180° (Bhat *et al.*, 1978) that reduce unfavorable steric interactions (Janin *et al.*, 1978). The exception is proline whose side-chain is geometrically restricted in an eclipsed conformation required for ring closure. Examination of Figure 7 confirms a general adherence to these expected trends in cytochrome *c'*. The distribution of observed χ_1 angles is consistent with the results of a survey conducted by James & Sielecki (1983) who noted that most amino acid side-chains prefer the $\chi_1 = -60^\circ$ or $\chi_1 = 180^\circ$ conformations over $\chi_1 = +60^\circ$. Serine and threonine favor $\chi_1 = \pm 60^\circ$, particularly in α -helical conformations, where the side-chain hydroxyl can form a hydrogen bond to the backbone carbonyl oxygen of the preceeding turn of helix (Gray & Matthews, 1984). All of the threonine and serine residues found in helices in our model are involved in such hydrogen bonds (see Fig. 6).

(e) Solvent structure

Refinement has enabled the identification of 194 partially ordered water sites in the asymmetric unit. These represent 20% of the approximately 900 water molecules estimated to fill the crystal lattice volume unoccupied by protein. The small fraction of the solvent molecules observed crystallographically reflects the disordered character of the great majority of water in the protein crystal lattice. Nevertheless, careful examination of the ordered water sites in this and other proteins can help evaluate the role of solvent interactions in stabilizing protein structures.

No water is found buried in the interior of the cytochrome *c'* dimer, a situation that may simply reflect the packing efficiency of hydrophobic side-chains in the interior of α -helical bundles. Instead, ordered solvent appears unevenly distributed over the protein surface, occupying sheltered niches where favorable hydrogen bonding can occur (Fig. 8).

Some tightly bound water molecules that exhibit high occupancies are found in narrow channels at or

near intermolecular crystal contacts (e.g. see Fig. 2(b)). These water molecules typically form one or two-dimensional hydrogen-bonded arrays with each other, as well as with protein atoms. However, water is more commonly found in partially ordered clusters, where the combination of low individual site occupancies and high overall site densities suggests that the solvent structure model represents only the most probable arrangement in a dynamic ensemble. There appears to be no extensive preference for the arrangement of water into pentagonal rings as observed in the more highly ordered crystals of crambin (Teeter, 1984).

Because the cytochrome *c'* crystals incorporate chemically identical subunits in different lattice environments, it becomes possible to assess both the role of bound water in stabilizing the structure and how the lattice environment affects protein solvent binding. The refined water sites in cytochrome *c'* were accordingly divided into two populations to facilitate this investigation. The first set contains all water molecules found in analogous positions in both of the crystallographically independent monomers, subject to the requirements that two sites exhibit hydrogen bonds to equivalent protein hetero-atoms in each monomer, and that they be less than 1.5 Å apart following least-squares superposition of the monomers (Bolin *et al.*, 1982). Of the total of 194 refined solvent sites, 92 (46/monomer) have been assigned to this "symmetric" class. The remaining 102 sites have been assigned to the second "asymmetric" class of solvent sites.

The most striking structural difference between the two populations is the extent to which they interact with main-chain atoms. Some 77% of the symmetric water subset are positioned to form hydrogen bonds with main-chain carbonyl oxygen or amide NH groups, compared with 24% of the asymmetric class. In addition, 16 of the 92 symmetric water molecules form more than one, and often as many as three, hydrogen bonds with protein atoms. In contrast, none of the asymmetric waters exhibits more than one protein interaction. The symmetric waters binding to the molecule thus appear to be stabilized either by multiple hydrogen bonding interactions or, in the case of water molecules singly bound to main-chain helix carbonyls, reflect local polar interactions arising as a result of polypeptide chain conformation. In either case, these symmetric water molecules make interactions that may appreciably contribute to protein stability in solution and might therefore be regarded as integral parts of the structure. The asymmetric solvent sites are primarily localized at or near monomer surfaces making alternative lattice contacts in the crystal. Some of these water molecules appear to extend hydrogen-bonded solvent arrays involving symmetric waters, or reflect specific interactions formed at crystal contacts. In other cases, differences occur whose cause is difficult to identify, although they presumably must reflect some aspect of the

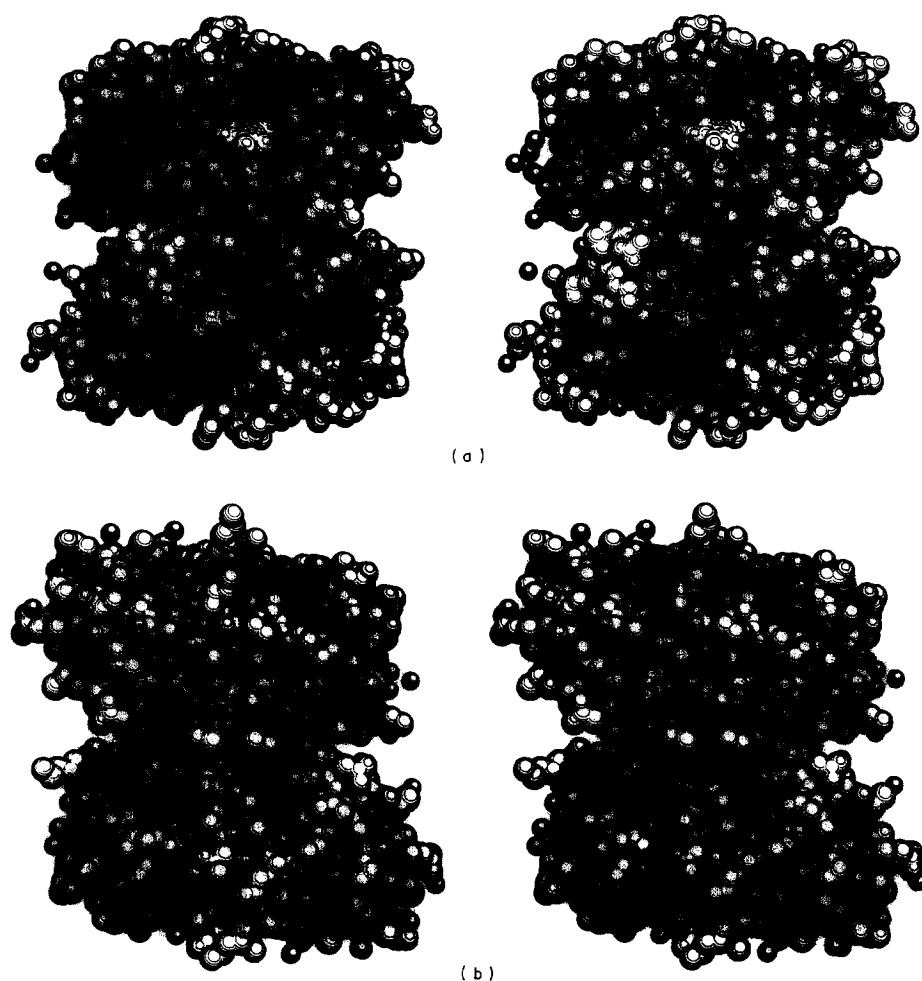


Figure 8. Bound solvent sites in cytochrome *c'*. The positions of bound solvent sites are illustrated by darkly shaded spheres on front (a) and rear (b) stereoscopic images of the cytochrome *c'* molecule viewed along the dimer symmetry axis.

asymmetric crystal lattice environment. It would be anticipated that the asymmetric sites would have varying stabilities in different crystal or solution environments.

Symmetric and asymmetric water sets were also compared with respect to their refined occupancies and temperature factors. Although it is difficult to quantify relative occupancies and temperature factors for partially occupied sites because these parameters interact during refinement, it was surprising to find that little obvious distinction existed between the two classes in terms of refined occupancies, temperature factors, or their ratios. Moreover, sites from both the symmetric and asymmetric classes exhibited a wide range of occupancy/*B*-value ratios depending on their local environment. Although some structural water sites (such as those that form bridging interactions with the heme propionic acid groups: Figs 2 and 10) refine with unit occupancy in both subunits, other pairs of symmetry-related solvent molecules show more variability. Figure 2(b) illustrates an example where members of the symmetric water subset exhibit near-unit occupancy in the subunit shown, but have substantially lower occupancies in the

dyad-related monomer. In the monomer illustrated, the bound water molecules are stabilized by crystal contacts with a symmetry-related molecule (omitted from Fig. 2(b) for clarity of the electron density). Although the different geometric and packing effects that result in the increased stability of these solvent sites relative to their symmetry mates are apparent in this case, the general situation is clearly much more complicated.

(f) *Heme geometry and axial ligands*

The high degree of similarity observed between monomers is also reflected in the structural properties of the heme groups. The hemes and side-chains of covalently connected protein residues Cys118, Cys121 and His122 superimpose with an r.m.s. co-ordinate error of 0.12 Å. The averaged geometrical properties of the heme prosthetic groups are summarized in Table 4. As is typical for highly refined heme protein structures, the heme planes deviate from perfect planarity. The average displacement of the 23 core porphyrin atoms from the mean porphyrin plane is 0.1 Å (Table 4). Systematic deviations of the pyrrole rings from the

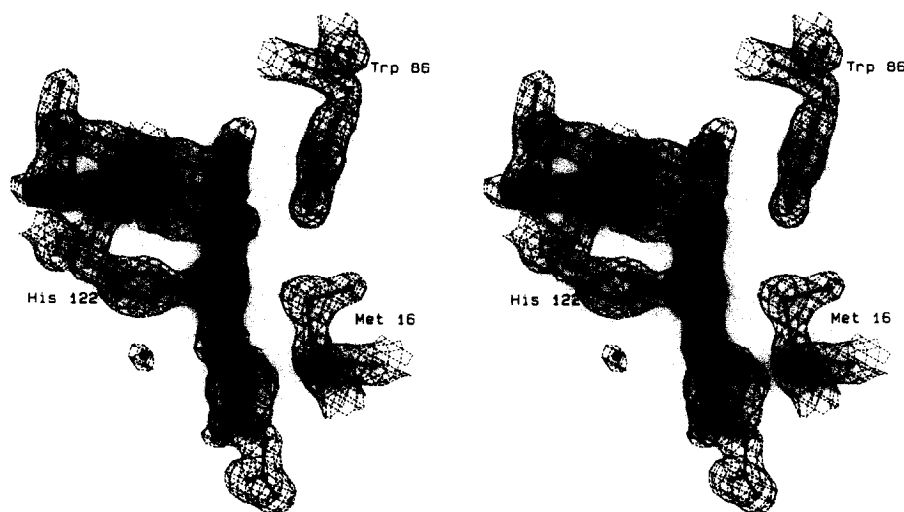


Figure 9. Stereoscopic view of monomer I heme electron density illustrating the situation of Met16 and the absence of a ligand at the sixth iron co-ordination site.

average heme plane result in a saddle-shaped prosthetic group with pyrrole rings B and D bent toward and pyrrole rings A and C bent away from the histidine axial ligand.

The heme iron is pentaco-ordinate with four in-plane pyrrole nitrogen ligands and a single axial ligand provided by the imidazole side-chain N ϵ 2 of His122. The heme iron is displaced 0.2 Å from the mean pyrrole nitrogen plane toward the co-ordinating histidine imidazole ligand. This displacement is somewhat larger than observed in other refined heme proteins, although cytochrome *c'* is the first pentaco-ordinate ferriheme protein to be refined at high resolution. The dihedral angle between the plane of the axial ligand histidine and the NA-NC pyrrole nitrogen vector is 45°, an orientation that minimizes steric overlap of the pyrrole nitrogens and histidine imidazole hydrogens (Collins *et al.*, 1972).

The unusual spectroscopic properties of the cytochromes *c'* has lead to speculation about the

existence and nature of a possible sixth axial ligand (Bartsch, 1978). However, inspection of refined electron density maps contoured at levels that clearly reveal bound solvent (Fig. 9) show no density corresponding to a sixth axial ligand. This observation is consistent both with an inspection of the heme environment van der Waals' surface, that shows close residue packing near the putative ligand site, and the absence of side-chains in the interior heme pocket oriented to form stabilizing hydrogen bonds with an axial ligand. A curious feature of the cytochrome *c'* heme environment is the situation of a methionine residue (Met16), whose side-chain sulfur is near the sixth axial heme co-ordinate position. Although methionine universally forms the sixth iron ligand in the mitochondrial family of low-spin cytochromes *c* (Salemme, 1977), Met16 is very well-defined in the cytochrome *c'* map (Fig. 9), and clearly oriented to disallow co-ordination with the heme iron. The observed iron-to-methionine sulfur distance of 3.75 Å in cytochrome *c'* is substantially longer than the 2.3 Å typically observed for iron-sulfur ligand distances in the mitochondrial cytochromes *c* and heme model compounds (Korszun *et al.*, 1982; Masiko *et al.*, 1979). Computer graphics modeling studies indicate that the lack of sixth-ligand co-ordinate bond formation in cytochrome *c'* is not a result of local conformational effects, as no alternative side-chain conformation could be found to situate the methionine sulfur within 2.3 Å of the heme iron in an axial orientation.

(g) Heme environment

The heme group in each cytochrome *c'* monomer is situated in an asymmetric environment, with one heme surface partially exposed to solvent and the other buried in a pocket lined with aromatic amino acid side-chains (Fig. 10). Although the heme-binding residues 118 to 122 of helix *D* shield most of the exterior heme face, atoms of the heme pyrrole

Table 4
Heme geometry

A. Distances (Å)	
Fe to pyrrole N plane†	0.2
Fe to porphyrin plane	0.2
Fe to His122 N ϵ 2	2.0
Fe to pyrrole N (average)	2.0
His122 N ϵ 2 to pyrrole N plane	2.2
B. Angles (°)	
Pyrrole A normal to porphyrin normal	2.3
pyrrole B normal to porphyrin normal	5.0
Pyrrole C normal to porphyrin normal	4.1
pyrrole D normal to porphyrin normal	5.3
pyrrole N normal to porphyrin normal	1.1
His122 imidazole plane to pyrrole N normal	3.4
His122 N ϵ 2-Fe bond to pyrrole N normal	6.3
His122 imidazole plane to NA-NC vector	42.9

† The mean porphyrin plane was determined by least-squares fits incorporating all 23 atoms in the porphyrin core. The pyrrole N plane is defined by the 4 pyrrole nitrogen positions.

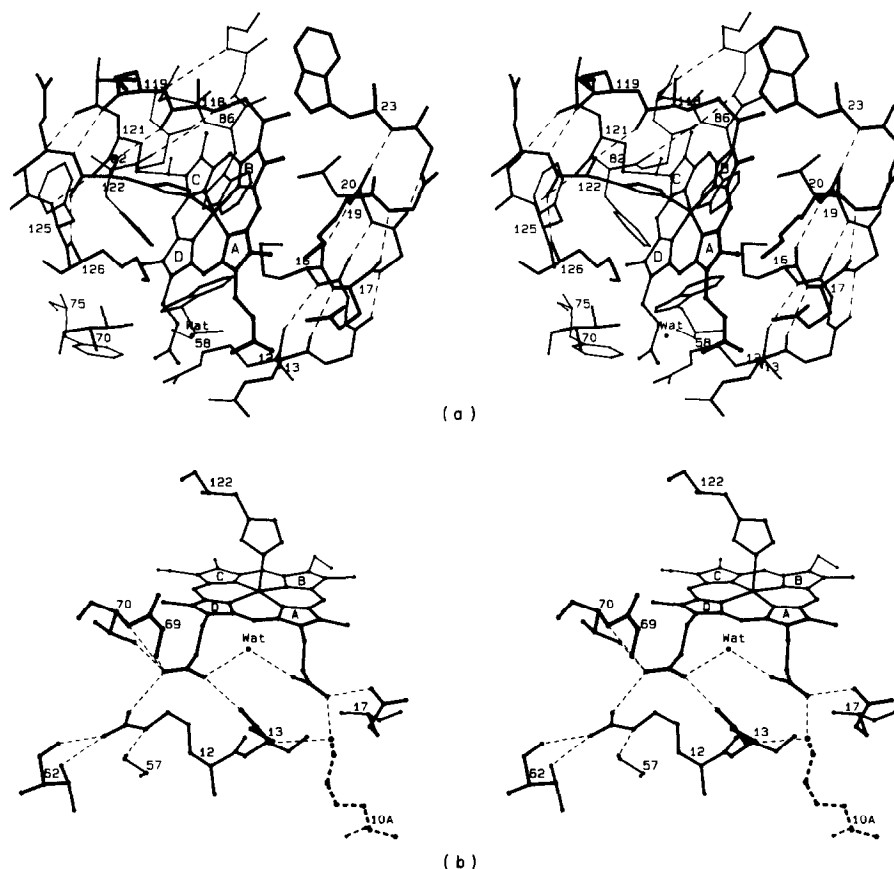


Figure 10. Stereoscopic views of the cytochrome *c'* heme environment. (a) The heme environment from a perspective approximately corresponding to Fig. 2. Broken lines show hydrogen bonds in helical segments incorporating residues that make packing interactions with the heme. (b) The pattern of ionic and hydrogen bond interactions stabilizing the heme propionates. These incorporate a salt-link between one of the heme propionates and a lysine residue (with heavy-broken side-chain) from the dyad-related monomer of the dimeric molecule. Individual residue assignments can be made by reference to Fig. 6.

rings A and B and the fifth axial ligand histidine are exposed to solvent. Residues of the *A*, *B*, and *C* helices otherwise make extensive hydrophobic and aromatic interactions with the buried heme surface. Hydrophobic residues of the *A* helix, including Trp23, form packing interactions with the interior faces of pyrrole rings A and B. The heme pyrroles C and D are more deeply buried in an interior pocket lined with the aromatic side-chains, Trp58, Phe75, Phe82, Trp86 and Phe126. The situation of the groups making heme interactions is essentially similar to that described by Weber, *et al.* (1981).

As is typical for the mitochondrial cytochromes *c* (Salemme, 1977) the heme propionate groups in cytochrome *c'* are partially buried in the protein interior. The refined structure of cytochrome *c'* reveals that the heme propionates are stabilized through participation in an extended hydrogen-bonding and salt-link network that extends between subunits in the dimer. As shown in Figure 10(b), this network includes an intersubunit salt-linkage between Lys10 of one monomer and the more solvent-exposed heme propionate, a structural water molecule that bridges the heme propionates, and a salt-link or hydrogen bond between a sequentially invariant arginine (Arg12) and the more buried propionate. In addition, a number of

other residues participate in the network through hydrogen bonding interactions involving either side-chain or backbone atoms.

(h) Structural basis for heme iron spin state

The heme iron spin state in ferricytochrome *c'* differs from the high-spin state usually found in other heme proteins. It has been postulated that the spin state of cytochrome *c'* reflects an admixture of high-spin ($S = 5/2$) and intermediate-spin ($S = 3/2$) character (Maltempo, 1974). As described, this proposal is in general agreement with both structural features of the cytochrome *c'* molecule and a variety of experimental data (Weber, 1982; and references cited therein). The refined structure reveals additional features consistent with the proposal for a mixed intermediate- and high-spin iron state.

Studies of model heme compounds show that the intermediate-spin state depends critically on the presence of a weak field axial ligand (Scheidt & Reed, 1981). In high-spin heme proteins having a histidine fifth axial ligand, the ligand field strength can be modulated by hydrogen-bonding to the imidazole side-chain. In general, increased ligand field strength is correlated with the extent of

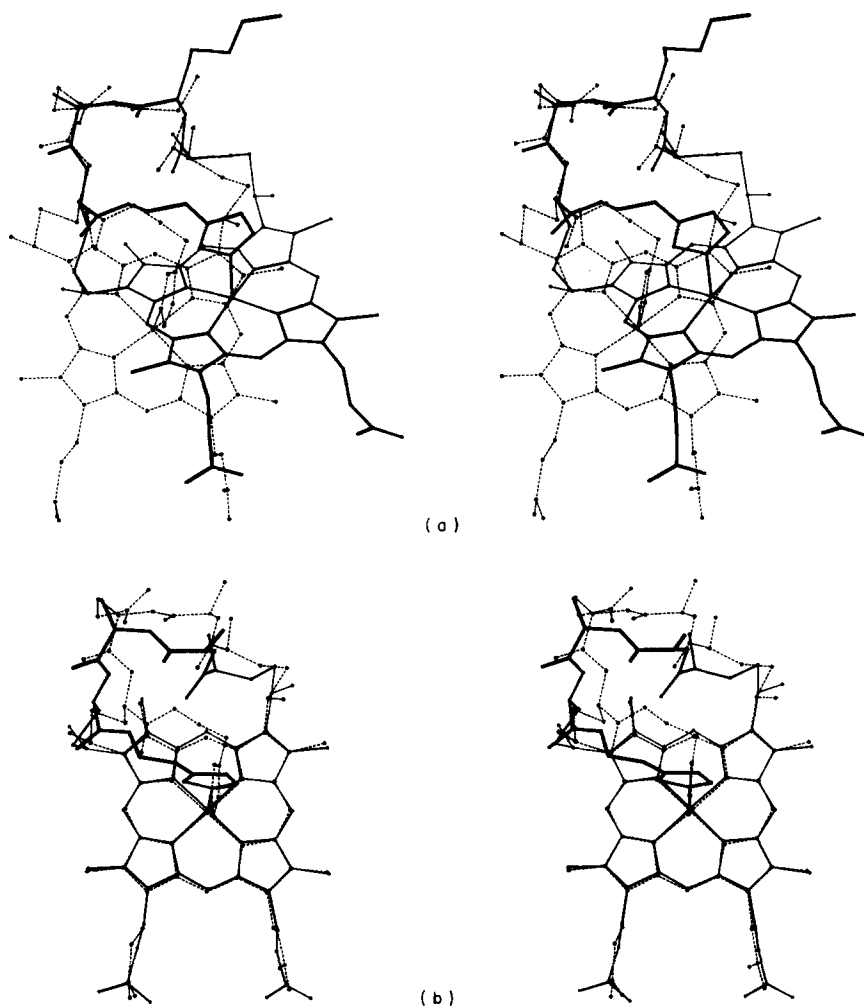


Figure 11. Stereoscopic views of comparative heme binding geometries in cytochromes *c*. Least-squares comparisons of the heme binding geometries in refined cytochromes *c* revealed the occurrence of 2 classes of heme binding conformation, exemplified here by *R. molischianum* cytochrome *c'* (continuous lines) and tuna cytochrome *c* (broken lines). (a) A superposition based on the Cys-X-X-Cys polypeptide backbones common to both structures. This produces a relative displacement in the heme iron atoms of approximately 3.7 Å and an interplanar heme angle of 32°. (b) Results of a heme group superposition. This clearly shows that the orientation of the axial ligand histidine plane differs by a 90° rotation about the imidazole nitrogen–iron bond in the 2 conformations. Structures included in the comparison are tuna cytochrome *c*, *Pseudomonas aeruginosa* cytochrome *c*₅₅₁, rice cytochrome *c*, *R. rubrum* cytochrome *c*₂, cytochrome *c*₃ (containing 2 of 4 total heme groups with C-X-X-C-H linkages), and *R. molischianum* cytochrome *c'*.

deprotonation of the imidazole Nδ1 nitrogen as would result, for example, from the formation of a strong hydrogen bond between the imidazole Nδ1 and a bound solvent molecule (Landrum *et al.*, 1980; Weber, 1982). Conversely, the behavior of His122 as a very weak field ligand in cytochrome *c'* would imply the absence of solvent strongly bound to the histidine imidazole. In fact, electron density maps contoured to reveal solvent sites show five low occupancy sites in the vicinity of the exposed histidine residues. However, this diffuse solvent density near the histidine axial ligand differs in the two subunits, and in neither case appears to correspond to an ordered solvent molecule bound to Nδ1 of the histidine imidazole. This suggests that, at least in the crystal, the cytochrome *c'* histidine axial ligand is essentially fully protonated at the Nδ1 position, and consequently in its weakest ligand field state.

In addition to ligand electronic effects, steric interactions that constrain the ligand plane in an eclipsed conformation relative to the porphyrin pyrrole nitrogens have been associated also with intermediate-spin state in crystalline model heme compounds (Scheidt *et al.*, 1983). However, steric factors do not appear to play a role in modulating the heme-spin state in cytochrome *c'*. The histidine ligand is relatively unconstrained in cytochrome *c'* and adopts the energetically preferred conformation where the dihedral angle between the plane of the axial ligand and the iron–pyrrole nitrogen bonds is 45° (Fig. 10). Structural studies of the intermediate-spin model compound (Scheidt *et al.*, 1983) also showed shorter iron-to-pyrrole nitrogen distances and a longer axial ligand-to-iron distance than observed in high-spin model compounds. However, the small differences observed do not provide a structural basis for distinguishing the heme geo-

metrical features in cytochrome *c'* from other heme proteins at comparable resolution, because the expected differences are comparable to estimated co-ordinate errors. Curiously, ferriheme model compound studies provide no clear correlation between spin state and the extent of iron displacement from the heme plane. Although the observed heme iron displacement in cytochrome *c'* is relatively large at 0.2 Å, and probably a consequence of heme pentaco-ordination by a very weak field ligand, the functional correlation of this displacement remains unclear. Further discussion of this and other detailed heme geometrical features is consequently deferred until the completion of high resolution studies at low temperature.

(i) Covalent bonds between heme and polypeptide

Heme proteins are classified as *c*-type cytochromes if the protoporphyrin IX prosthetic group is covalently attached to the polypeptide by thioether linkages formed through condensation of the heme vinyl groups and cysteine side-chains. With very few exceptions, the cysteine residues occur within the Cys-X-X-Cys-His sequence where the histidine residue immediately following the C-terminal cysteine provides the fifth axial ligand to the heme iron. The geometry of heme attachment is similar in *c*-type cytochromes of known structure, with the amino- and carboxy-terminal cysteine sulfur atoms bonded to the CAB atom of pyrrole B and the CAC atom of pyrrole C, respectively. In addition to this similarity in the pattern of covalent connections, the linkages also form identical stereoisomers at the CAB and CAC carbons which form chiral centers on condensation of the vinyl groups with the cysteine sulfur atoms. These observations of the similarities in heme attachment between otherwise structurally unrelated molecules suggested that a common enzymatic mechanism was responsible for heme attachment in both mitochondrial *c*-type cytochromes and cytochromes *c'* (Weber *et al.*, 1981). The structural refinement of cytochrome *c'* motivated a closer examination of the conformation of the covalent heme linkages. While generally confirming previous observations, it was found that covalent heme linkages in fact fall into two conformationally distinct classes.

The backbone atoms of heme-binding residues in *R. molischianum* cytochrome *c'* and six other examples in refined structures of *c*-type cytochromes were superimposed by least-squares procedures. In all cases, the Cys-X-X-Cys main-chain atoms adopt a quite regular α -helical conformation (average ϕ of -72° and ψ of -31°). Despite this local structural similarity, the *c*-type cytochromes clearly partition into two groups that differ in heme group orientation relative to the Cys-X-X-Cys α -helix. Figure 11 abstracts representatives from each group and makes clear that the differences in heme orientation principally result from alternate side-chain conformations of the C-terminal cysteine

and histidine residues. In the more common case that includes the majority of *c*-type cytochromes, both χ_1 of the cysteine and χ_2 of the histidine approximate preferred staggered conformations at -60° (see section (d), above). The second arrangement, that characterizes the heme binding in cytochrome *c'* and one of the cytochrome *c*₃ hemes (Higuchi *et al.*, 1984) results from rotation of both side-chain bonds to an alternative sterically preferred conformation Cys $\chi_1 = 180^\circ$ and His $\chi_2 = 180^\circ$. This alteration is accompanied by an $\sim 30^\circ$ relative rotation of the heme plane and a 90° rotation of the histidine plane around the imidazole nitrogen-iron bond (Fig. 11(b)). Both arrangements consequently orient the histidine plane so that the imidazole hydrogen atoms are staggered relative to the heme pyrrole nitrogens to minimize unfavorable steric interactions (Collins *et al.*, 1972). Thus, the two heme binding configurations appear to reflect alternative conformations that maintain both preferred side-chain torsional angles and an energetically favorable histidine-heme interaction.

5. Conclusions

The high-resolution structural refinement of *R. molischianum* cytochrome *c'* has revealed numerous structural features of the molecule that were unresolved in the initial 2.5 Å resolution electron density map. These include features that relate both to specific aspects of cytochrome *c'* physical properties, and more general aspects of protein molecular architecture and stabilization. Here we briefly summarize and comment upon some major results.

(1) As is typical of many proteins diffracting to high resolution, crystallographic refinement produces a structure that is extensively organized through backbone hydrogen bonds. In the case of cytochrome *c'*, some 88 of the total 127 backbone carbonyl groups, and 93 of the total backbone amide NH groups, are involved in hydrogen bond formation with other protein side-chain or backbone atoms. These interactions are reflected both in the organization of the extended α -helical regions of the structure, and more localized 3_{10} -helices and hairpin loops. The helices in cytochrome *c'* manifest a number of local structural perturbations. These include a pronounced bend at Pro25 in helix A, and kinks associated with the local formation of 3_{10} hydrogen bonds in the midst (Thr114 of the D helix) or at the ends of α -helices. The helices are otherwise reasonably regular, and characterized by average backbone torsional angles quite similar to those observed in other refined globular proteins (Baker, 1980; Finzel *et al.*, 1984; James & Sielecki, 1983). The observed geometry corresponds to a slight rotation of the peptide planes that orients the backbone carbonyl oxygens away from the helix axis. Previous observations (Bolin *et al.*, 1982) and the results presented here suggest that this reflects the frequent participation of helix carbonyl oxygens as additional hydrogen bond acceptors from

hydroxylic amino acid side-chains (Ser and Thr) or solvent water molecules.

A unique structural feature of the molecule is the single residue interconnection between the anti-parallel aligned *C* and *D* helices. This short connection occurs with no apparent disruptions in either the conformation or hydrogen bonding of the adjacent helix residues. The arrangement apparently depends upon the occurrence of alanine residues at the contact points between adjacent helix segments. These short contacts result in a tight interhelix packing interaction that allows a single residue to span the distance between the helix ends.

(2) Lattice organization. The cytochrome *c'* lattice contains a dimer in the crystallographic unit. As a result, the two monomers are situated in different lattice environments and make different lattice contacts with adjacent molecules. As no evidence exists for functional asymmetry in the molecule (see below), it is presently unclear why this situation should occur. Other species of dimeric cytochrome *c'* have been crystallized where the monomers are related by a crystallographic 2-fold symmetry axis (Weber & Salemme, 1977).

(3) Dimer symmetry. The independently refined monomers are remarkably similar as judged by a number of structural and conformational criteria. Significant differences that do occur are almost wholly confined to alternate side-chain conformations that reflect differing lattice contacts made by the two subunits. These observations suggest that structures organized as regular arrays of α -helices are relatively rigid and not readily deformed by lattice packing forces.

(4) Temperature factor behavior. The temperature factor behavior in cytochrome *c'* is typical in the sense that the extended regions of secondary structure appear to be more highly constrained than the less-organized sections of chain connecting them. Moreover, the refined temperature factor behavior is very similar for the two monomers despite their differing crystal environments. Taken together, these observations are consistent with the results of molecular dynamic simulations suggesting that refined *B*-values provide an indication of the relationship between structural organization and flexibility (Karplus & McCammon, 1983). While the overall similarity in *B*-value behavior in the two monomers must certainly reflect intrinsic dynamical properties of the structure, it is additionally clear that these can be modulated by lattice packing interactions (Finzel & Salemme, 1985). In the present case, lattice effects are principally manifest in differing orientations and extents of side-chain ordering for surface residues forming structured interactions at crystal contacts.

(5) Bound solvent organization. Some previous structural comparisons have noted the nearly equivalent situation of some bound water molecules in different crystal forms of the same or similar proteins (Bolin *et al.*, 1982; Moulton *et al.*, 1976). An analogous case occurs in cytochrome *c'*, where 92

(46/monomer) of the total 194 water molecules located in the crystal asymmetric unit are equivalently situated in independently refined monomers. These water molecules might consequently be regarded as parts of the molecular structure serving to stabilize the protein in solution. Conversely, the binding of the remaining 102 water molecules reflect lattice environmental effects which differ for the two monomers, and so are likely to exhibit a wider range of statistical occupancies in solution. It is interesting to note that while a substantial fraction of the asymmetrically situated water is localized near crystal contacts, many of the highly occupied sites at crystal contacts are populated by members of the symmetrical water set. In these cases, analogous sites on the symmetry-related subunit typically exhibit lower occupancies. These observations suggest that structural water molecules are important participants in the formation of lattice interactions, and may serve to nucleate the formation of the more extended arrays that incorporate both symmetric and asymmetric water sites at the cytochrome *c'* crystal contacts.

(6) Heme geometry and ligand environment. The refined electron density of the cytochrome *c'* heme environment shows no evidence of either a sixth axial ligand co-ordinating the heme iron, or a water molecule forming a strong hydrogen bond to the free imidazole nitrogen of the histidine axial ligand. These observations are consistent with spectroscopic and structural studies of model compounds. These indicate that the occurrence of a mixed intermediate- and high-spin state is associated with pentaco-ordination of the heme iron by a very weak field ligand (Scheidt & Reed, 1981).

As is the case for other *c*-type cytochromes (Salemme, 1977), the heme propionates in cytochrome *c'* are stabilized by an extensive salt-link and hydrogen bond network. The network incorporates a well-defined intersubunit salt-link between a heme propionate and lysine side-chain on the dyad-related monomer. This interaction constitutes the only salt-link or hydrogen bond that extends between subunits. Despite the interesting nature of this interaction, recent studies give no evidence of co-operative ligand binding to *R. molischianum* cytochrome *c'* (Doyle *et al.*, 1985) although this may be a property of cytochromes *c'* derived from other species.

(7) Geometry of covalent heme connections. Comparison of cytochrome *c'* with other highly refined *c*-type cytochromes shows the occurrence of two alternative conformations for the covalent connections formed between the heme and the Cys-X-X-Cys-His heme binding sequence. These differ mainly in the torsional orientations of the side-chain bonds of the ligating histidine and the adjacent cysteine that forms a heme thioether linkage. Both arrangements situate the amino acid side-chains in conformationally preferred orientations and additionally orient the ligand imidazole to minimize steric interference with the heme pyrrole nitrogen atoms. The latter effect results from a 90°

rotation of the histidine plane about the ligand-iron bond axis, that preserves the staggered orientation of the ligand imidazole relative to the heme pyrrole nitrogens in the two conformations.

The authors acknowledge the aid of Dr Robert C. Ladner for development of interactive computer graphics software. Earlier steps of refinement were done at the IBM Thomas J. Watson Research Center, Yorktown Hts., NY. This work was supported in part by N.I.H. research grants GM30393 and GM33325.

References

- Agarwal, R. C. (1978). *Acta Crystallogr. sect. A*, **34**, 791–809.
- Ambler, R. P., Bartsch, R. G., Daniel, M., Kamen, M. D., McLellan, L., Meyer, T. E. & Van Beeuman, J. (1981). *Proc. Nat. Acad. Sci., U.S.A.* **78**, 6854–6857.
- Arnott, S. & Dover, S. D. (1967). *J. Mol. Biol.* **30**, 209–212.
- Baker, E. N. (1980). *J. Mol. Biol.* **141**, 441–484.
- Balasubramanian, R. (1977). *Nature (London)*, **266**, 856–857.
- Bartsch, R. G. (1978). In *The Photosynthetic Bacteria* (Clayton, R. K. & Sistrom, W. R., eds), pp. 249–279, Plenum Press, New York.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
- Bhat, T. N., Sasisekharan, V. & Vijayan, M. (1978). *Int. J. Pept. Protein Res.* **13**, 170–184.
- Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. (1978). *Nature (London)*, **276**, 362–368.
- Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C. & Kraut, J. (1982). *J. Biol. Chem.* **257**, 13650–13662.
- Collins, D. M., Countryman, R. & Hoard, J. L. (1972). *J. Amer. Chem. Soc.* **94**, 2066–2072.
- Cruickshank, D. W. J. (1960). *Acta Crystallogr.* **13**, 774–777.
- Dodson, E. J., Isaacs, N. W. & Rollett, J. S. (1976). *Acta Crystallogr. sect. A*, **32**, 311–315.
- Doyle, M., Weber, P. C. & Gill, S. J. (1985). *Biochemistry*, **24**, 1987–1991.
- Finkel, B. C. & Salemme, F. R. (1985). *Nature (London)*, **315**, 686–688.
- Finkel, B. C., Poulos, T. L. & Kraut, J. (1984). *J. Biol. Chem.* **259**, 13027–13036.
- 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.
- Hardman, K. D., Agarwal, R. C. & Freiser, M. J. (1982). *J. Mol. Biol.* **157**, 69–86.
- Hendrickson, W. A. & Konnert, J. H. (1980). In *Biomolecular Structure, Function, Conformation and Evolution* (Srinivasan, R., ed.), vol. 1, pp. 43–57, Pergamon, Oxford.
- Higuchi, Y., Kusunoki, M., Matsuura, Y., Yasuoka, N. & Kakudo, M. (1984). *J. Mol. Biol.* **172**, 109–139.
- James, M. N. G. & Sielecki, A. R. (1983). *J. Mol. Biol.* **163**, 299–361.
- Janin, J., Wodak, S., Levitt, M. & Maignet, B. (1978). *J. Mol. Biol.* **125**, 357–386.
- Jones, T. A. (1978). *J. Appl. Crystallogr.* **11**, 268–272.
- Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
- Kamphuis, I. G., Kalk, K. H., Swarte, M. B. A. & Drenth, J. (1984). *J. Mol. Biol.* **179**, 233–256.
- Karplus, M. & McCammon, J. A. (1983). *Annu. Rev. Biochem.* **53**, 263–300.
- Konnert, J. (1976). *Acta Crystallogr. sect. A*, **32**, 614–627.
- Korszun, Z. R., Moffat, K., Frank, K. & Cusanovich, M. A. (1982). *Biochemistry*, **21**, 2253–2258.
- Landrum, J. T., Hatano, K., Scheidt, W. R. & Reed, C. R. (1980). *J. Amer. Chem. Soc.* **102**, 6729–6735.
- Lewis, P. N., Momany, F. A. & Scheraga, H. A. (1973). *Biochim. Biophys. Acta*, **303**, 211–229.
- Luzzati, V. (1952). *Acta Crystallogr.* **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.
- Masiko, T., Marchon, J.-C., Musser, D. T., Reed, C. A., Kastner, M. E. & Scheidt, W. R. (1979). *J. Amer. Chem. Soc.* **101**, 3653–3655.
- Matthews, F. S., Bethge, P. H. & Czerwinski, E. W. (1979). *J. Biol. Chem.* **254**, 1699–1706.
- Moult, J., Yonath, A., Traub, W., Smilansky, A., Podjarny, A., Rabinovich, D. & Sayer, A. (1976). *J. Mol. Biol.* **100**, 179–195.
- Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.
- Salemme, F. R. (1977). *Annu. Rev. Biochem.* **46**, 299–329.
- Scheidt, W. R. & Reed, C. A. (1981). *Chem. Rev.* **81**, 543–555.
- Scheidt, W. R., Geiger, D. K., Hayes, R. G. & Lang, G. (1983). *J. Amer. Chem. Soc.* **105**, 2625–2632.
- Sheriff, S., Hendrickson, W. A., Stenkamp, R. E., Sieker, L. C. & Jensen, L. H. (1985). *Proc. Nat. Acad. Sci., U.S.A.* **82**, 1104–1107.
- Shrake, A. & Rupley, J. A. (1973). *J. Mol. Biol.* **79**, 351–371.
- Sielecki, A. R., Hendrickson, W. A., Broughton, C. G., Delbaere, L. T. J., Brayer, G. D. & James, M. N. G. (1979). *J. Mol. Biol.* **134**, 781–804.
- Sim, G. A. (1960). *Acta Crystallogr. sect. A*, **34**, 791–809.
- Stenkamp, R. E., Sieker, L. C., Jensen, L. H. & McQueen, J. E. Jr (1978). *Biochemistry*, **17**, 2499–2504.
- Teeter, M. M. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 6014–6018.
- Ten Eyck, L. F. (1973). *Acta Crystallogr. sect. A*, **29**, 183–191.
- Venkatachalam, C. M. (1968). *Biopolymers*, **6**, 1425–1436.
- Ward, K. B., Hendrickson, W. A. & Klippenstein, G. L. (1975). *Nature (London)*, **257**, 818–821.
- Weber, P. C. (1982). *Biochemistry*, **21**, 5116–5119.
- Weber, P. C. & Salemme, F. R. (1977). *J. Mol. Biol.* **117**, 815–820.
- Weber, P. C. & Salemme, F. R. (1980). *Nature (London)*, **287**, 82–84.
- Weber, P. C., Bartsch, R. G., Cusanovich, M. A., Hamlin, R. C., Howard, A., Jordan, S. R., Kamen, M. D., Meyer, T. E., Weatherford, D. W., Xuong, Ng.-h. & Salemme, F. R. (1980). *Nature (London)*, **286**, 302–304.
- Weber, P. C., Howard, A., Xuong, Ng.-h. & Salemme, F. R. (1981). *J. Mol. Biol.* **153**, 399–424.
- Xuong, Ng.-h., Freer, S. T., Hamlin, R., Nielsen, C. & Vernon, W. (1978). *Acta Crystallogr. sect. A*, **34**, 289–296.