X-RAY CRYSTALLOGRAPHIC ANALYSIS OF CYTOCHROME C' FROM TWO BACTERIAL SPECIES

This thesis is submitted as partial fulfilment of the requirements for the degree of Doctor of Philosophy in Chemistry at Massey University.

AARON JOHN DOBBS

March, 1995

ABSTRACT

The structure of cytochrome c' from two closely related bacterial species, *Alcaligenes* sp and *Alcaligenes denitrificans*, have been determined from X-ray diffraction data to 1.80 Å and 2.15 Å resolution respectively using the anomalous scattering of the single iron atom in each to identify and refine a weak molecular replacement solution. Molecular replacement studies, with the program AMORE, used two isomorphous data sets (from the two species), two independent search models (the cytochromes c' from *Rhodospirillum molischianum* and *Rhodospirillum rubrum*), both with and without sidechains, and two different resolution ranges (10.0-4.0 Å and 15.0-3.5 Å) to generate a large number of potential solutions. No single solution stood out and none appeared consistently. The iron position in each structure was then determined from its anomalous scattering contribution and all molecular replacement solutions were discarded which did not (i) place the iron correctly and (ii) orient the molecule such that a crystallographic 2-fold axis generated a dimer like those of the two search models. Finally, electron density maps phased solely by the iron anomalous scattering were calculated. As these were combined and subjected to solvent flattening and histogram matching (program SQUASH), correlation with the remaining molecular replacement solutions identified one as correct and enabled it to be improved and subjected to preliminary refinement. The correctness of the solution was confirmed by parallel isomorphous replacement studies. Both crystal structures have been refined using least-squares methods, to give final R-factors of 0.184 for the *Alcaligenes* sp cytochrome c' (953 protein atoms and 89 solvent molecules) and 0.167 for the *Alcaligenes denitrificans* cytochrome c' (950 protein atoms and 75 solvent molecules). Analysis and comparisons of these structures with three cytochrome c' structures previously determined show that despite the low level of sequence conservation among the family members the gross overall structure is maintained (four-α-helix bundle). The configuration of the iron ligands and of the surrounding haem environment are very similar indicating that although residues are not conserved they are replaced by residues which fulfil a similar function. There are relatively few sidechain...sidechain interactions in cytochromes c' which probably reflects the fact that the majority of stabilising $i^{interactions}$ are hydrophobic in nature. This hydrophobic packing could also explain the lack of sequence identity between the different species of cytochrome c' as strict conservation of the residues is not required. All the structures do have a similar haem stabilisation framework, with an approximately ten residue piece of structure situated in the BC loop involved in both haem stabilisation and interhelix stabilisation in all of the cytochromes c'

determined. The overall four-helix structure is also discussed in terms of its parameters with respect to the current interest in the *de novo* design of proteins.

ACKNOWLEDGEMENTS

Many people have given me their time and advice during my thesis and if I do not mention your name please don't be offended as I do thank you.

The biggest acknowledgement goes to my supervisors Prof. Ted N. Baker and Dr. Bryan F. Anderson who gave endless support and guidance throughout a long and difficult Ph.D. Their contribution was immeasurable in both knowledge and encouragement.

Also Mrs Heather M. Baker for her adept crystallisation skills and kindly ear when required. Dr. Gill E. Norris for providing me the crystals for this study plus a place to stay when finishing off this thesis.

Dr. Hale. H. Nicholson and the University of Oregon for the collection of data sets involved in this thesis.

The post-docs such as Drs. David Thomas, Hale Nicholson, Colin Groom, Rick Faber and Maria Bewley who all provided me with different ideas and techniques in macromolecular crystallography .

The other members of the crystallography research group (past and present) should not get away without mention, so thank you..

Two very good friends, Michael "you should be humanitarian ambassador" Nairn, who unfortunately for the general population has a very similar sense of humour to mine and Dean "pirate" Severinsen both of whom got me through this thesis and life in general.

I am grateful to Massey University for the position of Graduate Assistant to make this study possible and to the Howard Hughes Medical Institute grant for further funding.

Lastly I would like to thank my family who still are not quite sure about what I do but gave me their support throughout my time at Massey University.

CONTENTS

Chapter One

INTRODUCTION

Chapter Two

CRYSTALLISATION, DATA COLLECTION AND PROCESSING

Chapter Three
MOLECULAR REPLACEMENT

Chapter Four
PHASE DETERMINATION BY ISOMORPHOUS REPLACEMENT AND ANOMALOUS SCATTERING

Chapter Five
STRUCTURE DETERMINATION

Chapter Six
STRUCTURE DESCRIPTION

Chapter Seven
STRUCTURAL COMPARISONS WITH OTHER CYTOCHROMES C'

Chapter Eight
GENERAL DISCUSSION

ABBREVIATIONS

NO_3^-	Nitrate ion
A_{280}	Absorbance at 280nm
CM-cellulose	Carboxymethyl cellulose
R. rubrum	*Rhodospirillum rubrum*
R. molischianum	*Rhodospirillum molischianum*
Cys	Cysteine
His	Histidine
a.a	amino acid
NH_3	ammonia
NADP/H	phosphorylated NAD
NO	Nitric oxide
N_2O	Nitrous oxide
NO_2	Nitrogen dioxide
NO_2^-	Nitrite
Rps. viridis	*Rhodopseudomonas viridis*
Rps. palustrus	*Rhodopseudomonas palustrus*
Rps. capsulata	*Rhodopseudomonas capsulata*
Rps. sphaeroides	*Rhodopseudomonas sphaeroides*
pI	Isoelectric point
Chr. Vinosum	*Chromatium vinosum*
M	mol L^{-1}
IP	Image plate
Asp	*Alcaligenes* sp
Ade	*Alcaligenes denitrificans*
Rms	Root-mean-square
MR	Molecular replacement
MIR	Multiple isomorphous replacement
MIRAS	Multiple isomorphous replacement with anomalous scattering
SIR	Single isomorphous replacement
SIRAS	Single isomorphous replacement with anomalous scattering
Asp	Aspartic acid
Gln	Glutamine
Glu	Glutamic acid
Ala	Alanine
Pro	Proline
Asn	Asparagine
Hem	Haem

Val	Valine
Lys	Lysine
Ile	Isoleucine
Leu	Leucine
Thr	Threonine
Met	Methionine
Tyr	Tyrosine
Phe	Phenylalanine
Ser	Serine
AB loop	Loop between helices A and B
BC loop	Loop between helices B and C
CD loop	Loop between helices C and D
DEAE	Diethyl aminoethyl
ALDH	Aldehyde dehydrogenase
FPLC	Fast protein liquid chromatography
NAD/H	Nicotinamide adenine dinucleotide
ATP	Adenosine triphosphate
SDS	Sodium dodecylsulphate
PEG	Polyethylene glycol
DMNB	3,4-dihydro-3-methyl-6-nitro*2H*-1,3-benoxazin-2-one
Da	Dalton
ASCC	*Alcaligenes* sp cytochrome c'
ADCC	*Alcaligenes denitrificans* cytochrome c'
CVCC	*Chr. vinosum* cytochrome c'
RMCC	*R. molischianum* cytochrome c'
RRCC	*R. rubrum* cytochrome c'
VDW	van der Waals
PCMB	*para*-chloromecuribenzoate
PHMB	*para*-hydroxymecuribenzoate

LIST OF FIGURES

LIST OF TABLES

Chapter One

Introduction

1.1 Cytochromes in biology.

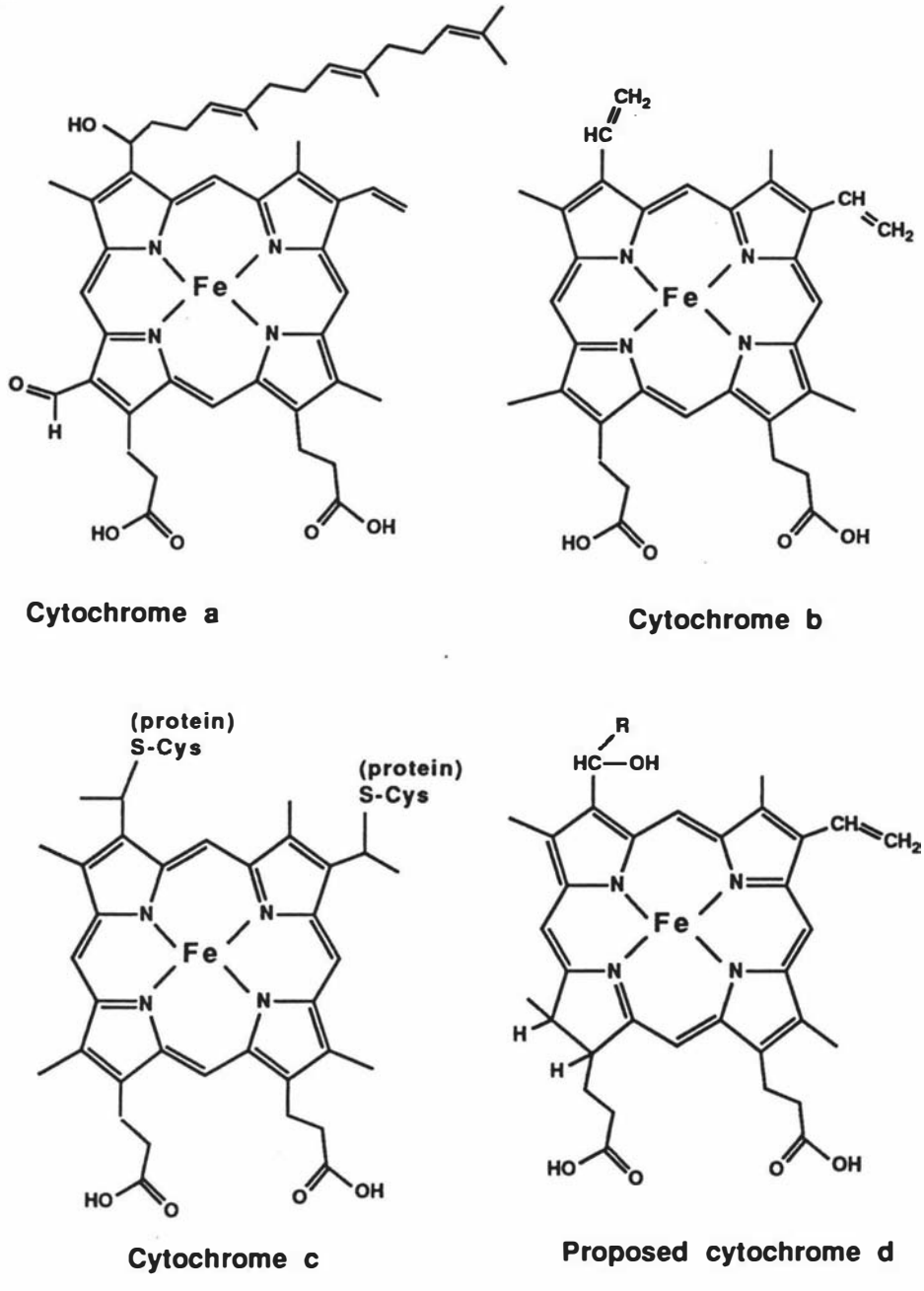
The 1965 Report of the Commission on Enzymes of the International Union of Biochemistry defined cytochromes as "haem proteins whose principal biological function is electron and/or hydrogen transport by virtue of a reversible valency change of their haem iron". Studies of cytochromes date back to the late 1880 s when the spectroscopic studies of MacMunn (1886, 1887) started to reveal the role of cytochromes in cellular respiration. However most of the knowledge of cytochromes is based on the research that was undertaken by David Keilin in the late 1920 s. Keilin recognised that the four banded spectrum of cytochromes was in fact due to several compounds which he called cytochromes a, b and c. He also proved that the change of iron valency was connected with their respiratory function as electron carriers between molecular oxygen, activated by an enzyme, and substrates activated by specific dehydrogenases. It was not until the mid 1950 s that the first cytochrome was crystallised, a cytochrome c from King penguin muscle (Bodo, 1955). Research into the purification and crystallisation of cytochromes then intensified, culminating with the first 3D X-ray diffraction structure of cytochrome c (Dickerson *et al.*, 1967).

The study of bacterial cytochromes began in 1928 with a study of Yaoi and Tamiya (1928) on cytochrome "a_2" (now classed as cytochrome d) in *Escherichia coli* and *Shigella dysenteriae*. The importance of these bacterial cytochromes lies not only in their greater variability, eg. the different oxidases a_1, a_2 and o as well as aa_3, and the unusual combinations of cytochromes found in them, but also in the role of these cytochromes beyond that of aerobic metabolism.

1.1.1 Classification on basis of optical spectra.

The classification of cytochromes into the four types a, b, c, and d is quite straightforward. The prosthetic haem groups to which the proteins are attached have sufficiently different characteristics to give rise to different optical absorption spectra. Classification in terms of the different haem groups is as follows (see also Figure 1.1.1.1).

1. Cytochromes a - Cytochromes in which the haem prosthetic group contains a formyl sidechain.
2. Cytochromes b - Cytochromes with protohaem as prosthetic groups.
3. Cytochromes c - Cytochromes with covalent "thioether" linkages between haem and protein. This group includes all cytochromes with prosthetic groups linked in this way.
4. Cytochromes d - Cytochromes with a dihydroporphyrin (chlorin) iron as prosthetic group.



Cytochrome a **Cytochrome b**

Cytochrome c **Proposed cytochrome d**

Figure 1.1.1.1 Haem groups of the cytochromes which give rise to their distinctive optical spectra.

The absorbance bands arising from the various attachments and structures of the prosthetic haem groups can be classified into three types. These are:

(i) π - π^* transitions. These involve the excitation of electrons from bonding (π) to antibonding (π^*) molecular orbitals that are located primarily on the porphyrin. They generally produce intense bands.

(ii) d-d transitions. These are transitions of electrons within the iron d-orbital set. They are formally forbidden by spectroscopic selection rules and therefore for cytochromes are generally too weak to be observed.

(iii) Charge transfer (CT) transitions. These result from the transfer of electrons between iron orbitals and orbitals of the axial ligands, and between iron orbitals and orbitals of the porphyrin. Intensities of the CT bands vary over a wide range but they are normally less intense than allowed π - π^* transitions.

For cytochrome spectra the most important molecular orbitals are the π-bonding and π^*-antibonding orbitals. The resulting π - π^* transitions give rise to the Soret band, and the α/β band. The absorption due to the haem group dominates the spectra, producing bands at ~ 410 nm (Soret band) and between 500 and 600 nm (the β and α bands). Bands at ~ 280 nm originate from aromatic residues and a weak CT band at 695 nm results from an interaction between the haem and its ligands. It is on the basis of the differences in the maxima of the α-band that the cytochromes can be classified. Cytochromes a, b , c and d have a maxima in their α-bands of ~ 600, ~ 560, ~ 550 and 620 nm respectively which allows for their characterisation. Figure 1.1.1.2 illustrates the distinctions in their optical spectra by showing three complex systems each of which has a c-type cytochrome associated with another haem (from Pettigrew and Moore, 1987).

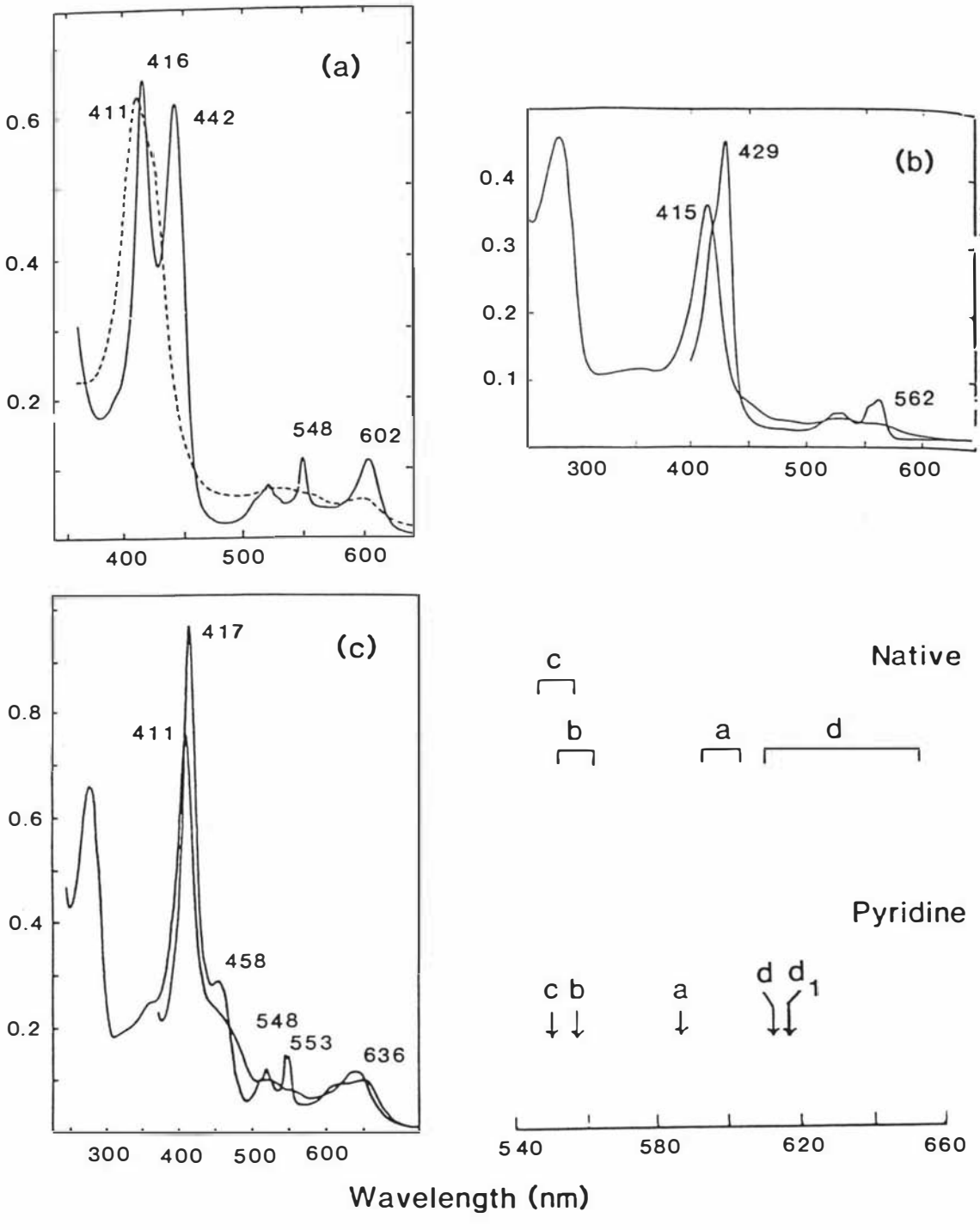


Figure 1.1.1.2 UV-visible absorption spectra of complex cytochromes containing haem c. The ferrocytochrome spectra are those showing a band near 550 nm. (**a**) The caa_3 complex of *Thermus thermophilus* (Hon-nami and Oshima 1980); (**b**) the cytochrome bc_1 complex of mitochondria (Yu *et al.*, 1974); (**c**) cytochrome cd_1 of *Ps. aeruginosa* (Meyer and Kamen, 1982). The *lower right* part of the figure shows the range of the α-peak maxima for native ferrocytochromes (top) and for the alkaline pyridine ferrohaemochromes (bottom) (Yamanaka and Okunuki, 1974; Falk, 1964; Lemberg and Barrett, 1973).

1.1.2 Distribution and roles.

Cytochromes can be generally found either in the inner membrane of mitochondria or in the bacterial cytoplasmic membrane. The major role of cytochromes is said to be that of electron carriers to molecular oxygen in the respiratory chain. This is closely interwoven with the conversion of adenosine diphosphate plus inorganic phosphate to adenosine triphosphate in the process of oxidative phosphorylation. The discovery of cytochromes in chloroplasts of green plants (cytochromes f and b_6) by Hill and co-workers (Hill and Scarisback, 1951; Davenport and Hill, 1952; Hill, 1954) led to the discovery of the role of cytochromes in photosynthetic reactions, not only in green plants and algae, but also in anaerobic photosynthetic bacteria. The role of cytochromes in anaerobic reactions is not restricted to photosynthetic reactions, however, as they have been found to be essential in nitrate, nitrite and sulphate reduction (Sato and Egami, 1949; Egami *et al.*, 1961; Postgate, 1959, 1961). Thus cytochromes must have been essential catalysts at a time before photosynthetic processes gave the earth's atmosphere sufficient molecular oxygen to make cell respiration possible.

1.2 c-Type cytochromes.

1.2.1 Common features.

Cytochromes c can be distinguished from the general family of cytochromes by their method of haem attachment to the protein via thioether linkages. This method of binding in turn gives rise to specific changes in their electronic spectra. The haem attachment is conserved throughout the cytochromes c in the characteristic sequence motif Cys-X-Y-Cys-His. Here the two Cys residues are covalently linked to the haem and the His residue serves as an axial ligand for the haem iron. Cytochromes c are small proteins of between 82 - 134 amino acids with molecular weights ranging from approximately 10,000 Da to 14,000 Da for each monomeric unit.

1.2.2 Subclassification of cytochromes c.

c-Type cytochromes have been classified into three groups, I, II and III, based on the sequences of the bacterial cytochromes (Ambler, 1980). This classification is based primarily on the position of covalent attachment of the haem group (either at the N-terminus or C-terminus) and whether the haem is hexacoordinate or pentacoordinate.

The validity of this classification has been confirmed by X-ray crystallography (Almassy and Dickerson, 1978; Korszun and Salemme, 1977; Weber *et al.*, 1980 and Haser *et al.*, 1979) as three-dimensional structures have been determined from each sub-class and correlated with the sequence classification. Apart from the Cys-X-Y-Cys-His haem attachment motif, there is no sequence similarity between the three classes.

Within each subclass, the 3D structures of all members appear to be homologous (at least for those those whose structures have been determined to date). For example mitochondrial cytochrome c (11,500 Da, 104 residues) and bacterial cytochrome c-551 (8,500 Da, 82 residues) have shown to have essentially the same fold, despite very little sequence identity and major size differences (Pettigrew and Moore, 1990). On the other hand, proteins from different subclasses appear to have no 3D structural relationship. Representative structures of each subclass are given in Figure 1.2.2.1 (reproduced from Dickerson, 1980). The shaded portion of the chain of mitochrondrial cytochrome c, consis ting of 16 residues, is apparently missing from the smaller cytochromes. The shaded portions of the chain of *R. rubrum* cytochrome c_2 correspond to three and eight residue insertions relative to the mitochrondrial cytochrome c. *Paracoccus* cytochrome c-550 is the largest in its class: the shaded areas correspond to two and six residue insertions plus a ten residue C-terminal tail. *Pseudomonas* cytochrome c-551 has a single residue insertion relative to the others, indicated by shading.

Table 1.2.2.1 gives the distinguishing properties of each class of c-type cytochrome. In Figure 1.2.2.2 the UV-visible absorption spectra for the three different types of cytochrome c are shown (from Pettigrew and Moore, 1987). In each case the the ferri- and ferrocytochrome spectra were obtained by the addition of potassium ferricyanide and sodium dithionite respectively to the solutions of the cytochrome in 0.1 mol L^{-1} phosphate pH 7.0 (Meyer and Kamen, 1982). The ferrocytochrome spectra are those showing a band near 550 nm which is often used to designate a particular cytochrome.

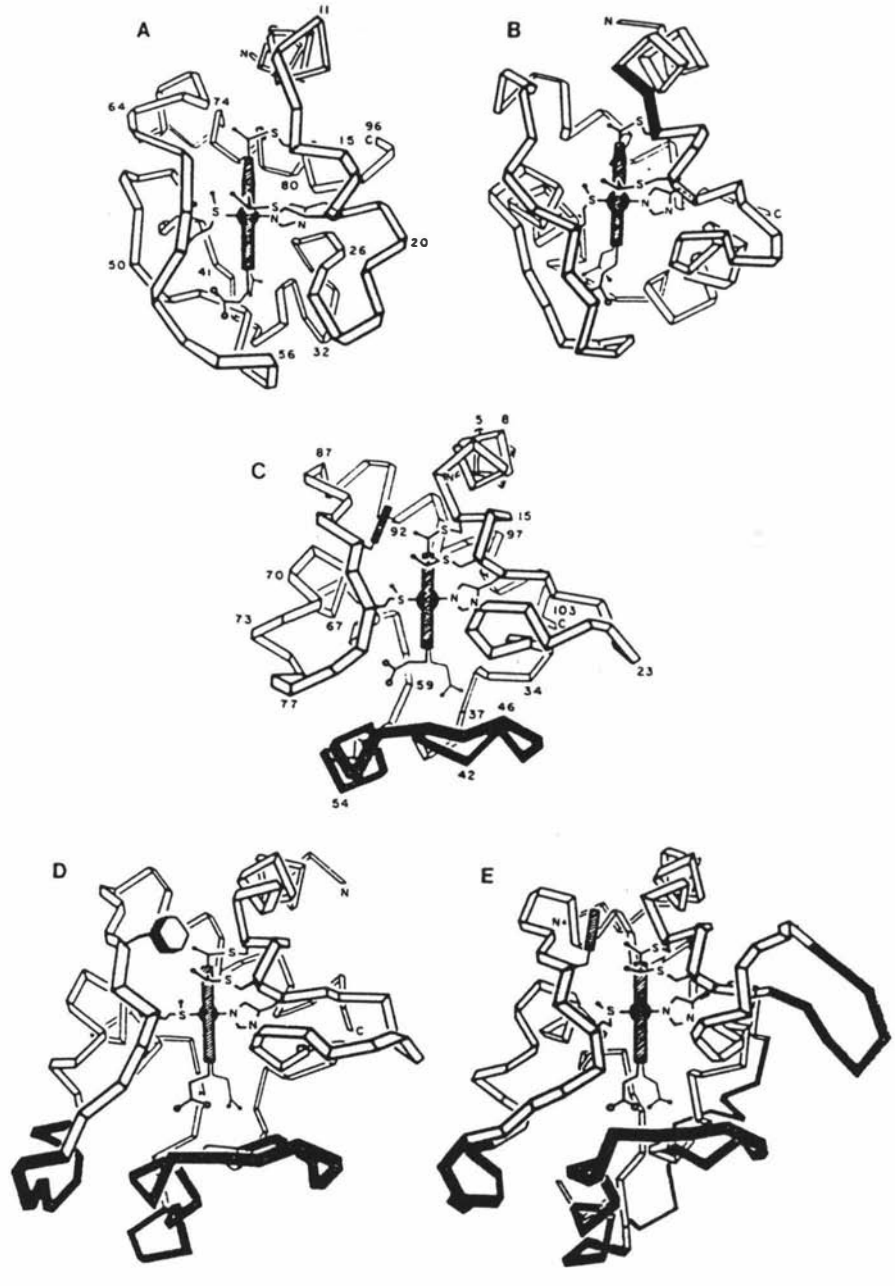


Figure 1.2.2.1 Ribbon diagrams of various cytochromes c. **A** Cytochrome c-555 from *Chlorobium limicola*; **B** Cytochrome c-551 from *Pseudomonas aeruginosa*; **C** Cytochrome c from Tuna; **D** Cytochrome c_2 from *Rhodospirillum rubrum*; **E** Cytochrome c-550 from *Paracoccus denitrificans*.

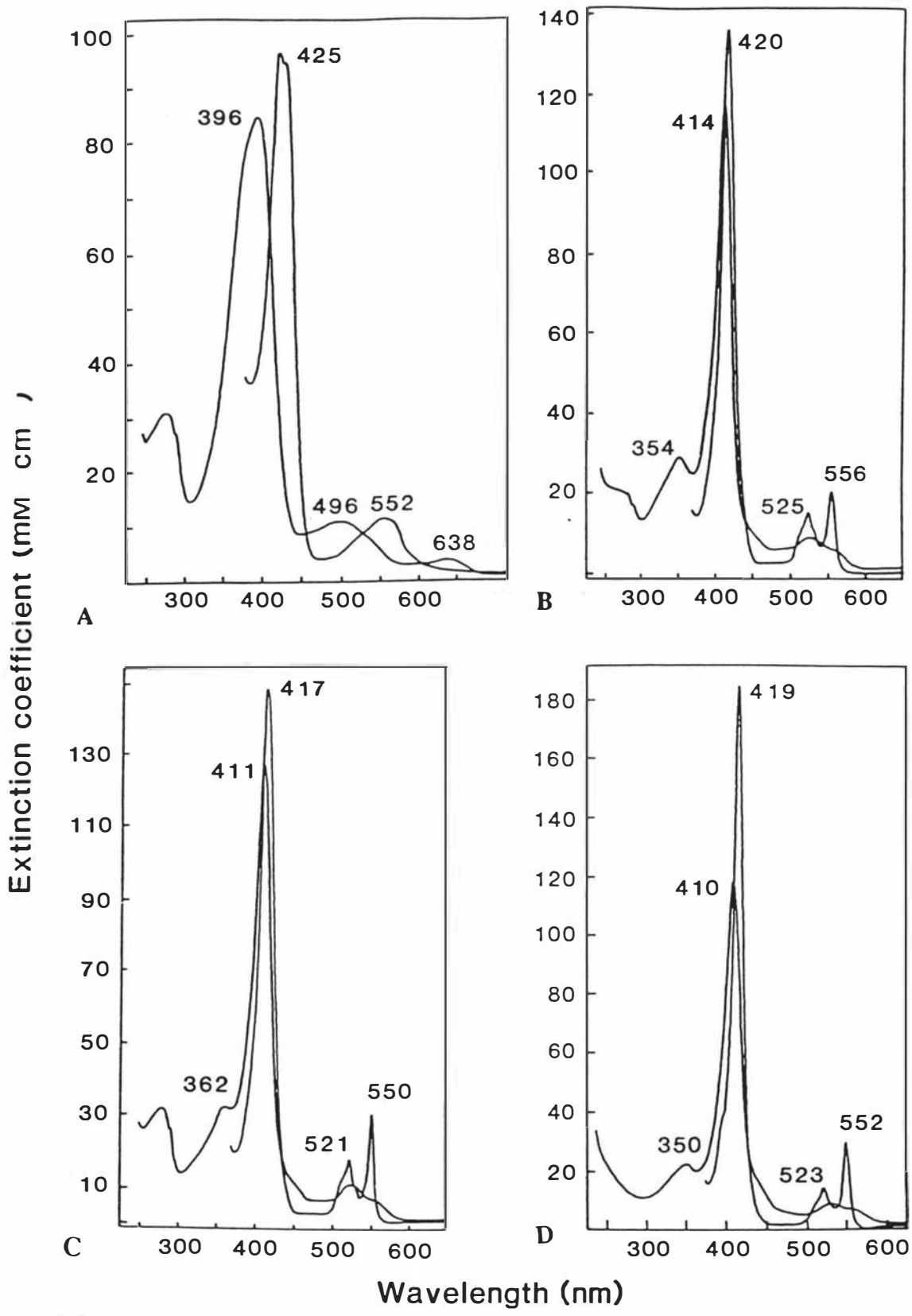


Figure 1.2.2.2 A-D. UV-visible spectra of selected c-type cytochromes. **A** cytochrome c' from halotolerant *Paracoccus* sp; **B** cytochrome c-556 from *Rps. palustrus*; **C** cytochrome c_2 from *Rps. sphaeroides*; **D** cytochrome c_3 from *Desulfovibrio vulgaris*.

Table 1.2.2.1 Distinguishing properties of c-type cytochromes.

Class	Features	Subdivision	Features	Examples
I	Low spin His+Met haem coordination Haem attachment near N-terminus 80-130 aa	(a) large (>100aa)	Loop of residues closes bottom of haem crevice	mitochondrial cyt.c (ref a)
		(b) small (<100aa)	Lacks loop of (a). Left side folds downward to close bottom of haem crevice	*Pseudomonas* cyt.c-551 (ref b) *Chlorobium* cyt.c-551 (ref c)
II	Haem attachment near C-terminus	(a) High-spin	His-only haem coordination	cyt.c'(ref d)
		(b) Low-spin	His + Met haem coordination	cyt.c-556 (ref e)
III	Multi-haem, one per 30-40 aa bis-Histidyl coordination	On basis of haem content		cyt.c_3 (3 haem) (ref f) cyt.c_3 (4 haem) (ref g)

a - Dickerson *et al.* (1971)
b - Almassy and Dickerson (1978)
c - Matsuura *et al.* (1982)
d - Weber *et al.* (1980)
e - no structure yet
f - no structure yet
g - Czjek *et al.* (1994)

1.2.3 c-Type cytochromes in bacteria.

A very large group of the bacterial cytochromes c (class I) probably act as diffusible links between cytochrome c reductase systems and terminal reactions (as shown in Figure 1.2.3.1), while other c-type cytochromes are involved as part of these reductases and terminal enzymes. Figure 1.2.3.1 illustrates the central importance of the class I cytochromes c in the modular construction of the oxidising end of electron transport systems which allows flexibility in the use of electron donors and acceptors (from Pettigrew and Moore, 1987). The redox potentials of cytochromes c in electron transport systems generally occupy the positive end of the biological redox potential scale (with the exception of cytochrome c_3) with redox potentials of between 0 and +400 mV. Only the redox centres associated with the terminal enzymes themselves are more oxidising, while flavins and quinones have more negative redox potentials.

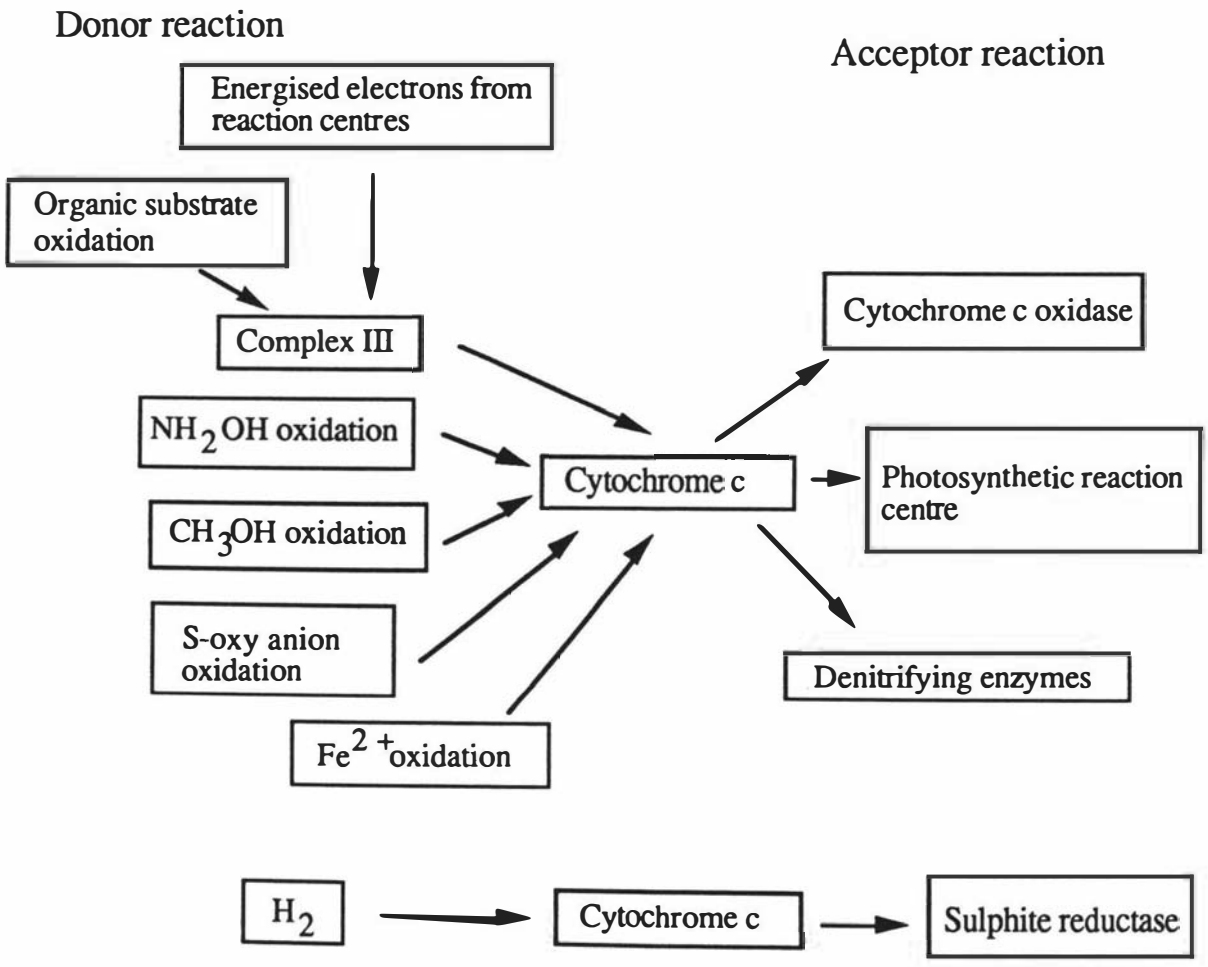


Figure 1.2.3.1 Terminal reactions in bacterial electron transport.

The c-type cytochromes display a wide range of functional roles in bacterial respiration systems. In nitrification, cytochromes c play a role in electron transfer and are a part of the enzyme hydroxylamine oxidase which oxidises one of the intermediates in the aerobic oxidation of NH_3 (Tsang and Suzuki, 1982). The oxidation of methane involves a cytochrome c as a mediator in the recycling of electrons between methanol dehydrogenase (Colby and Dalton, 1978; Tonge *et al.*, 1977) and methane mono-oxygenase (Duine and Frank, 1980). Their ability to use sulphide, sulphur and the oxides of sulphur is widespread. The phototropic bacteria (such as *Chromatiaceae* and *Rhodospirallaceae*) generally oxidise sulphur compounds anaerobically using the electrons to reduce $NAD(P)^+$. The reduction of molecular oxygen to water involves several types of bacterial cytochrome oxidases including aa_3, a_1, o or d. The cytochrome c oxidases are transmembrane enzymes which oxidise cytochrome c in the periplasmic space (Wood, 1983), a compartment topographically equivalent to the intermembrane space of the mitochrondrion. They reduce oxygen to water using protons taken up from the cytoplasmic side of the membrane. c-Type cytochromes are also involved in nitrate respiration and denitrification, one of the best characterised of these proteins being cytochrome cd_1 (Newton, 1969; Kuronen *et al.*, 1975; Henry and Bessieres, 1984). Denitrification is defined as the respiratory (dissimilatory) reduction of nitrite to gaseous products (N_2, NO and N_2O) which are lost to the atmosphere (Delwiche and Bryan, 1976). This process is of ecological importance because some of the products may contribute to the destruction of the ozone layer. Most denitrifiers are non-fermentative facultative gram-negative chemheterotrophic bacteria which use NO_3^- as a less preferred respiratory substitute for oxygen (Payne, 1981). Usually nitrate is the initial respiratory acceptor for denitrification but intermediates in the reduction pathway can often be used and in some cases (e.g. NO_2^- in *Alcaligenes*) are required.

c-Type cytochromes are found in large numbers in denitrifying bacteria. Their roles are generally poorly defined but several probably act as redox mediators between the electron transport systems and the terminal enzymes of denitrification. In addition, c-type cytochromes are involved as components of some of these terminal enzymes.

1.3 Cytochromes c'.

1.3.1 Distribution.

Cytochrome c' was first isolated by Vernon and Kamen (1954), who named it pseudohaemoglobin, and it underwent a succession of name changes until the present

one was adopted. A full description of its discovery and characterisation is given by Lemberg and Barrett (1973) and Bartsch (1978). Cytochrome c' is found in the photosynthetic bacteria - *Chromatiaceae* and *Rhodospirillaceae* - the aerobe, *Azotobacter vinelandii* and the facultative anaerobes, *Alcaligenes* sp and the halotolerant *Paracoccus* (Meyer and Kamen, 1982). Although widespread in the Rhodospirillaceae it is not found in *Rhodomicrotobium vanniellii* (Morita and Conti, 1963), *Rps. viridis* (Olson and Nadler, 1965) and the one strain of *Rps. palustrus* that contains the low spin cytochrome c-556 (Deklerk and Kamen, 1966). In *R. rubrum* and *Rps. capsulata* it is present in both photosynthetic and respiratory cells (Taniguchi and Kamen, 1965; Prince *et al.*, 1975), while in *Rps. sphaeroides* it is only present in the former (Kikuchi *et al.*, 1965).

The erratic distribution of cytochrome c' makes definition of its function difficult. Further difficulty arises from the probability that cytochrome c' is altered upon solubilisation from the periplasmic membrane (Pettigrew and Moore, 1987). In *Chromatium*, although purified cytochrome c' binds CO, intact cells bind much less CO than expected from their c' content (Cusanovich *et al.*, 1968). Taniguchi and Kamen (1965) could not detect cytochrome c' by immunological methods in membrane particles of *R. rubrum*, yet Kakuno *et al.* (1971) were able to solubilise cytochrome c' from such membranes using detergent.

1.3.2 Functional role.

A role for cytochrome c' in photosynthetic electron transport has been proposed on the basis of light-induced absorption changes (Olson and Chance, 1960; Morita, 1968; Cusanovich *et al.*, 1968), but the problem of identifying the cytochrome *in situ* cast a doubt on these interpretations (Pettigrew and Moore, 1987). A different possibility is that the cytochrome c' is involved as part of the denitrification process. Thus cytochrome c' was isolated from the denitrifying bacterium *Alcaligenes* sp (Suzuki and Iwasaki, 1962) and Shidara (1980) also purified a cytochrome c-556 from the same organism.

Denitrification as a complete sequence of reactions is rare amongst the *Rhodospirillaceae* but the ability to reduce nitrous oxide may be more common (McEwan *et al.*, 1985). McEwan *et al.* (1985) suggested that cytochrome c' may mediate the transfer of electrons to N_2O reductase at the level of reduced quinone (this is shown in Figure 1.3.2.1). Reduction of N_2O is an antimycin-insensitive process and in *Rps. capsulata* it is proposed this occurs via cytochrome c' (McEwan *et al.*, 1985).

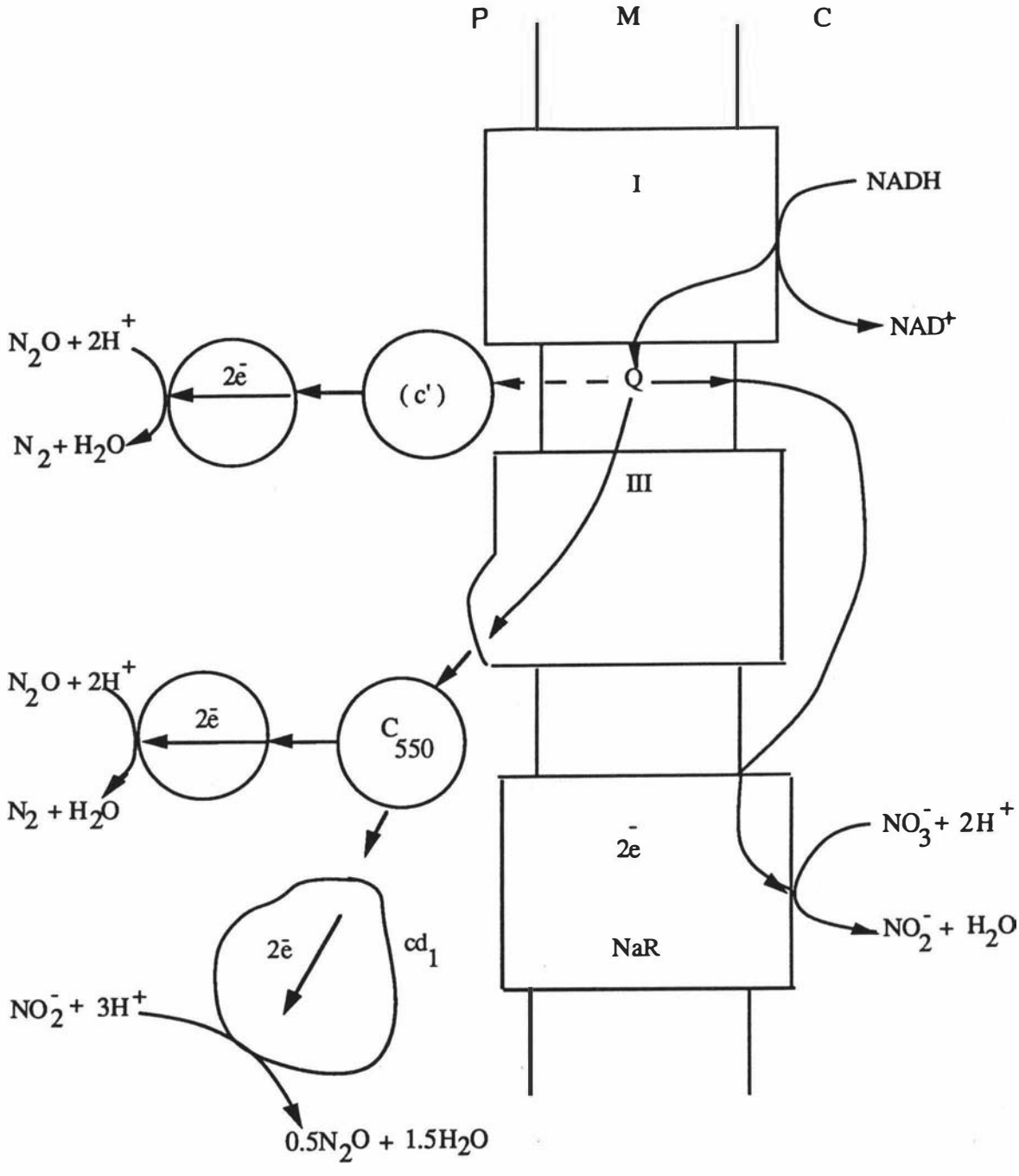


Figure 1.3.2.1 The denitrifying electron transport system of *Pa. denitrificans*. Respiratory complexes are inserted in a membrane (M) separating the cytoplasm (C) from the periplasm (P) (from Pettigrew and Moore, 1987).

1.3.3 Physicochemical properties.

All cytochromes c' form a homologous class with respect to optical spectra, mid-point potential and molecular weight (Table 1.3.3.1 gives the properties for some selected cytochromes c' - adapted from Pettigrew and Moore, 1987).

Table 1.3.3.1 Properties of some cytochromes c'.

Species	m/d[a]	pI	E° (mV)	pK_a [b]	
Alcaligenes denitrificans [c]	d		132	7.1	(optical)
R. rubrum [d]	d	5.6	-8	8.2	(optical)
			10	8.1	(E_m)
				8.5 - 9.0	(EPR, NMR, MCD, RR, Mössbauer)
R. molischianum [e]	d			8.8	(optical, NMR)
Rps. palustrus [f]	m	9.4	102	7.8	(optical, MCD)
			100	8.0	(E_m)

a - m = monomer and d = dimer. Meyer and Kamen (1982) point out that there may be a rapid equilibrium between monomer and dimer in the case of *Rps. palustrus* cytochrome c'. This is consistent with the crystallisation of both monomeric and dimeric forms (Salemme, 1974).

b - The pK_a is for the ionisation of His, the axial histidine ligand to a histidinate. Methods used to determine these values are in parentheses.

c - Cusanovich *et al.* (1970).

d - Bartsch and Kamen (1960); Bartsch (1978); Kitigawa *et al.* (1977); Barakat and Strekas (1982); Rawlings *et al.* (1977); Emptage *et al.* (1977) and La Mar *et al.* (1981).

e - Moore *et al.* (1982).

f - Bartsch (1978).

In haem proteins iron can exist in two common oxidation states, the ferrous state Fe(II):[Ar]$3d^6 4s^0$ or the ferric state Fe(III):[Ar]$3d^5 4s^0$. There are five 3d orbitals each capable of being occupied by two electrons, and the distribution of electrons between these governs the electronic properties of iron. Using Crystal Field Theory (CFT) the ligands are treated as point charges and it is assumed that the bonding between the iron and its ligands is entirely electrostatic in nature. This is depicted in Figure 1.3.3.1 which shows the effect of an octahedral crystal field on d-orbital splitting.

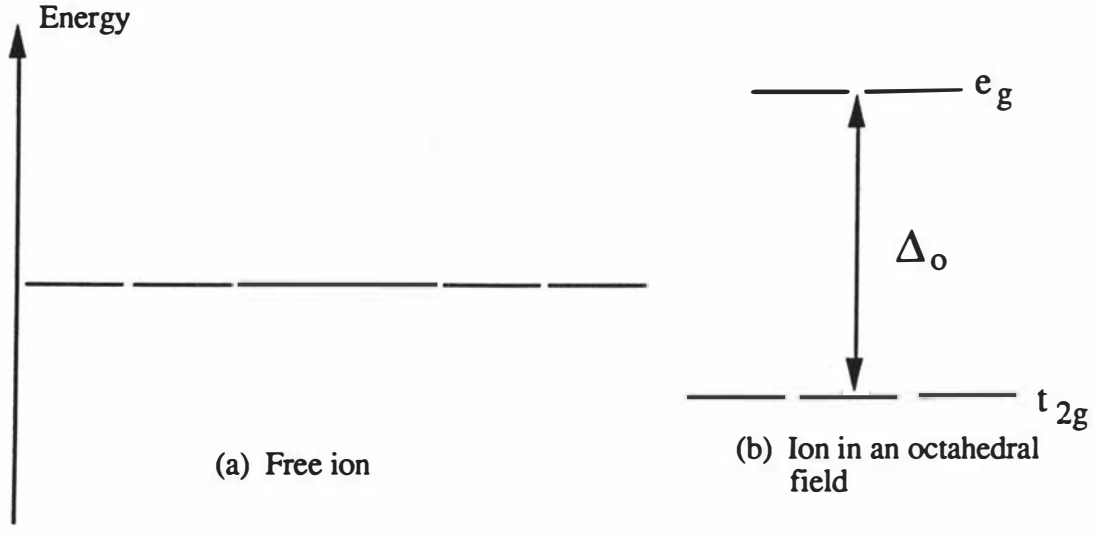


Figure 1.3.3.1 Effect of an octahedral crystal field on the d-orbital splitting. Δo is the crystal field stabilisation energy and e_g and t_{2g} are the symmetry labels. The $d_{x^2-y^2}$ and d_{z^2} orbitals comprise the e_g set and the d_{xz}, d_{yz} and d_{xy} orbitals comprise the t_{2g} set.

In a six-coordinate octahedral complex the orbitals along the Cartesian axes ($d_{x^2-y^2}$ and d_{z^2}) are destablised by the incoming ligands more than the orbitals lying between the axes. These former are called the e_g set, with the latter forming the t_{2g} set. Electrons in the t_{2g} have lower energy.

There are two major energy terms that govern the distribution of electrons in d orbitals: the strength of the d-orbital splitting (given the symbol Δ_o), and the energy required to place electrons in the same orbital (P). The pairing energy is always unfavourable but what is important is whether it requires less energy to pair electrons in t_{2g} orbitals than it does to keep them unpaired by occupying the e_g orbitals. Where there is a weak crystal field the d-orbital splitting is smaller than the pairing energy and the electrons remain unpaired and enter separate orbitals. This distribution is the high-spin configuration. When there is a strong crystal field, the electrons enter the t_{2g}

orbitals and pair up to produce the low-spin configuration. These electronic configurations are shown in Figure 1.3.3.2.

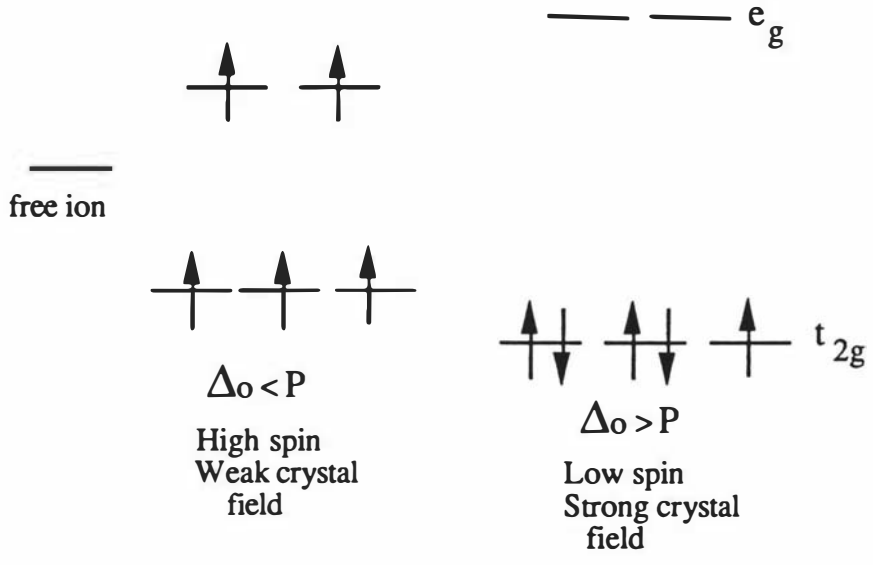


Figure 1.3.3.2 Spin states of octahedral Fe(III). The arrows represent electrons occupying the d-orbitals. In the low spin state the unpaired electron is shared by the three t_{2g} orbitals.

The cytochromes c' have a pentacoordinate haem which gives rise to an essentially high-spin spectrum that distinguishes it from the other cytochromes c. There is a band at 600 - 640 nm which is characteristic of the high-spin ferrihaemproteins. This "640 band" may arise from a porphyrin -> Fe(III) CT transition (Makinen and Churg, 1983). The distinctive EPR signals of purified cytochrome c' are not observed in intact cell or membrane fractions (Prince *et al.*, 1974; Corker and Sharpe, 1975). Kakuno *et al.* (1971) have suggested that cytochrome c' in situ may have a b-type optical spectrum and in this context it is noteworthy that the low spin sequence homologues of cytochrome c' (such as cytochrome c-556) have red-shifted absorption bands (Pettigrew and Moore, 1987). Although cytochromes c' are normally regarded as high-spin iron proteins they do in fact have somewhat anomalous properties which distinguish them from other such proteins. The spectroscopic behaviour of cytochromes c' has been suggested to reflect a quantum mechanical admixture of intermediate (S = 3/2) and high (S = 5/2) spin states (Maltempo, 1974; Maltempo *et al.*, 1974; Maltempo, 1976; Maltempo and Moss, 1976).

Weber (1982) hypothesised that the environment of the haem in cytochromes c' supported the existence of the admixed spin state at neutral pH, and suggested that the pH-dependent transition to a pure high-spin state at alkaline pH involved deprotonation of the axial histidine ligand to the haem iron. In other high-spin haem

proteins, the $N_{\delta 1}$ nitrogen of the axial histidine is hydrogen bonded to a protein oxygen (Salemme *et al.*, 1973: Valentine *et al.*, 1979; Poulos and Kraut, 1980). This hydrogen bonding is thought to increase the imidazole ligand field strength (Nappa *et al.*, 1977; Valentine *et al.*, 1979). The lack of this hydrogen bond in cytochromes c' suggests that the histidine acts as a weaker field ligand than in other high-spin haem proteins (Weber, 1982). Combined with the absence of a sixth axial ligand to the haem iron (Emptage *et al.*, 1981; Weber *et al.*, 1981; Yasui *et al.*, 1992; McRee *et al.*, 1993) this indicates an overall weak field axial coordination in cytochromes c' thus correlating with the observed spectroscopic properties.

Cytochromes are also able to bind CO, alkylisocyanides and CN^- to the sixth ligand position (Kassner, 1991; Wood, 1984). The rate and equilibrium constants are 10^2 - 10^6 smaller than other high spin haem proteins which is thought to be a result of steric interactions at the haem coordination site (Kassner, 1991).

1.3.4 Amino acid sequences.

The amino acid sequences of 13 cytochromes c' have been determined (Ambler *et al.*, 1981; Pettigrew and Moore, 1990). Cytochromes c' have polypeptide chains that are approximately 130 amino acids in length with the haem binding region situated at the C-terminus. A striking feature of cytochromes c', unlike the various class I cytochromes, is that there are few invariant residues present in the sequences. Pairwise sequence comparisons show only 25-30 % identity and taking the family as a whole ~ 10 % of residues are absolutely invariant (Ambler *et al.*, 1981). This low level of identity makes sequence alignment of cytochromes c' difficult. In chapter 7 of this study an alignment has been derived from the cytochromes c' whose 3D structures have been solved, by aligning the conserved haem attachment region and then using least-squares superposition methods to determine the rest of the alignment. The Cys-X-Y-Cys-His haem attachment motif, however, is present throughout the cytochromes c' and other invariant residues include an arginine close to the N-terminus, a glycine-threonine pair at approximately residues 60 and 61, and lastly a conserved glycine at residue 100.

1.3.5 X-ray crystallographic structures of cytochrome c'.

Three 3D structures of cytochrome c' have already been determined. Two high resolution structures, a 1.67 Å structure of cytochrome c' from *R. molischianum* (Weber *et al.*, 1981; Finzel *et al.*, 1985), and a 1.8 Å resolution structure from *Chr. vinosum* (McRee *et al.*, 1993) have been deposited in the Brookhaven Data Bank (Bernstein *et al.*, 1977; entries 2CCY and 1BBH respectively). The third structure

reported is a 2.8 Å resolution structure of cytochrome c' from *R. rubrum* (Yasui *et al.*, 1992). The crystallographic analyses of these three cytochromes c' have established that they form a classic four-helix fold. The four α-helices are packed sequentially into a left-twisted anti-parallel bundle; ie. with the N-terminus of one helix packed against the C-terminus of another helix. All of the cytochromes c' form dimers in solution. The dimer interface is an anti-parallel four-α-helical AA'BB' bundle, where A and B come from one monomer and A' and B' from the other. Most of the residues forming contacts between the monomers are hydrophobic. The haem environments of the structures are identical, with the iron atoms being pentacoordinate and having one axial histidine ligand.

1.4 Aims of this study.

The aims of this study were to determine the X-ray crystallographic structures of the cytochromes c' from two closely related *Alcaligenes* species (*Alcaligenes* sp and *Alcaligenes denitrificans*) and analyse these structures in order to establish the common structural framework of the cytochrome c' family. This included analysing the common bonding motifs in the stabilisation of haem binding, and in the formation of the cytochrome c' dimer. The structural basis of the proposed admixed spin state and its pH dependence in these cytochromes is not really understood so analyses of the *Alcaligenes* structures was expected to help clarify these questions. The presence of a modified N-terminal residue in these cytochromes c' is of interest as such a feature has not been found in any of the other cytochrome c' structures previously solved.

The structural constraints (or lack of them) within the cytochrome c' family are also of considerable relevance to current interest in the *de novo* design of proteins. Many synthetic studies have been directed towards the design of proteins with a four-helix fold (Ho and DeGrado, 1987; Regan and DeGrado, 1988; Hecht *et al.*, 1990) and the ways in which this fold can be varied and stabilised 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 utilised in attempts to incorporate a particular active site into a designed molecule of this type (Hahn *et al.*, 1990).

Chapter Two

Crystallisation, Data Collection and Processing

2.1 Purification of cytochrome c'.

The crystals of cytochrome c' from *Alcaligenes* sp and *Alcaligenes denitrificans* used in these studies were grown by Dr. G. Norris (Norris *et al.*, 1979). The protein was purified concurrently with the copper protein azurin from the same two species. *Alcaligenes denitrificans* was grown under aerobic conditions, with stirring, while *Alcaligenes* sp was grown anaerobically. Azurin and cytochrome c' were purified using a method based on those of Parr *et al.* (1976), Cusanovich *et al.* (1970) and Ambler (1973). The cells were suspended in 0.02 mol L^{-1} phosphate buffer (pH 7.0) then broken in a French pressure cell and treated with deoxyribonuclease, and the cell debris removed by centrifugation. The dark green supernatant was brought to 45 % saturation with ammonium sulphate and the precipitate discarded. The supernatant was then brought to 100 % saturation, the resulting precipitate collected by centrifugation, dissolved in distilled water and dialysed against distilled water until its conductivity reached that of distilled water. The protein solution was acidified to pH 4.1 with acetic acid, centrifuged to remove a white precipitate and the green solution loaded onto a CM-cellulose column (Whatman CM-52) at pH 4.1. Cytochrome c' was eluted as already described (Ambler, 1973; Parr *et al.*, 1976) at pH 7.0. It was then further purified, after concentration by ultrafiltration, on a DEAE-cellulose column (Ugurbil and Bersohn, 1977; Ambler and Brown, 1967). The cytochrome c' solution was rechromatographed on CM-cellulose, precipitated with ammonium sulphate and subjected to gel filtration. The final pure material had a spectral ratio Soret/A_{280nm} of 0.33 to 0.35.

2.2 Crystallisation of cytochrome c' from *Alcaligenes* sp and *Alcaligenes denitrificans*.

The best cytochrome c' crystals were obtained by vapour diffusion. Finely divided ammonium sulphate was added slowly, with stirring, to the protein solution (2 to 10 mg ml^{-1} in 0.1 mol L^{-1} phosphate buffer, pH 8.0) until it showed slight turbidity (approximately 50 % saturated). Buffer was then added drop by drop until the turbidity had vanished. The solution was then centrifuged at high speed for two hours, and left to equilibrate with 95 % saturated ammonium sulphate at room temperature. Crystals were

also obtained by microdialysis against 95 % saturated ammonium sulphate in 0.1 mol L^{-1} phosphate (pH 8.0), with 1.0 mol L^{-1} sodium chloride, or by free interface diffusion.

Dark brown crystals from both species appeared in three to four weeks. Those from *Alcaligenes denitrificans* were chunky hexagonal bipyramids, up to 3 mm in length, while those from *Alcaligenes* sp formed as long needles of hexagonal cross-section up to 5mm long. Figure 2.2.1 shows crystals of cytochrome c' from *Alcaligenes* sp. Precession photographs taken indicated hexagonal symmetry (Norris *et al.*, 1979), corresponding to space group $P6_122$ (or its enantiomer $P6_522$), with preliminary cell dimensions:

a = b = 54.7 Å, c = 181.5 Å, γ = 120.0 ° (*Alcaligenes* sp)
a = b = 54.7 Å, c = 187.3 Å, γ = 120.0 ° (*Alcaligenes denitrificans*)

The crystal density was measured to be 1.22 g cm^{-3}, and assuming a monomer in the asymmetric unit, values of V_m = 2.80 $Å^3$ $Dalton^{-1}$ and 2.89 $Å^3$ $Dalton^{-1}$ were obtained (Matthews, 1968).

The crystals were stored in 0.1 mol L^{-1} phosphate buffer (pH 8.0) which was 95 % saturated with ammonium sulphate solution.

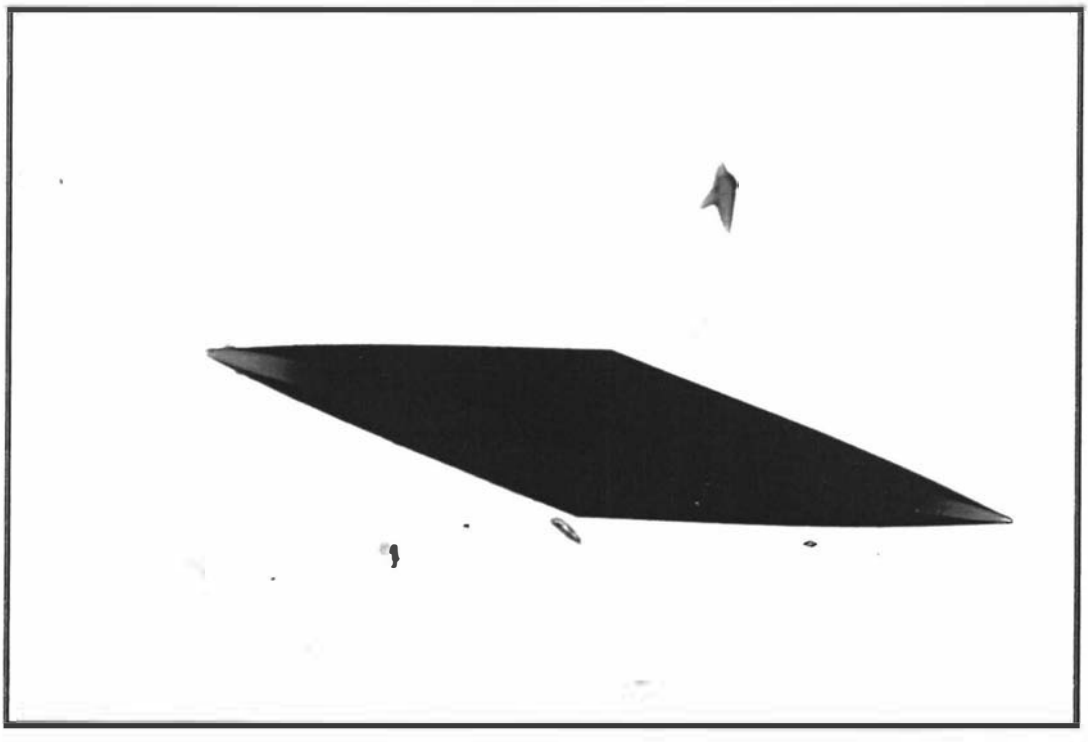


Figure 2.2.1 Crystals of cytochrome c' from *Alcaligenes* sp.
(x25 magnification)

2.3 Data Collection and Processing.

2.3.1 Photon factory data set - *Alcaligenes* sp native data.

Data Collection

A native data set from *Alcaligenes* sp cytochrome c' was collected using synchrotron radiation at the Photon Factory, Tsukuba, Japan. Two crystals were used, one with its a-axis along the capillary and the other with its c-axis along the capillary. The crystals were aligned along the desired axis using small angle oscillation photographs. Intensities were then recorded by screenless Weissenberg photography (Sakabe, 1991). The reflections were recorded on Fuji imaging plates and measured as laser stimulated luminescence with a BA100 image plate densitometer. Camera settings and exposure times used to collect the c-axis data are summarised in Table 2.3.1.1, while the settings and angles used for the a-axis data are given in Table 2.3.1.2.

Table 2.3.1.1 Image plate (IP) data collection statisics for collection along c-axis.

Parameter	Value
Source	Beam-line 6A2, Photon Factory
Method	Screenless Weissenberg photography
Wavelength	1.04 Å
Crystal to IP distance	430 mm
Coupling constant (mm/°)	2.0
Oscillation range (ω°/IP)	9.0
Number of osc^{n} / IP	4
ω scan speed (°/s)	4.0
Exposure per image (s)	36
Number of IP's	8
Total exposure (s)	288

Table 2.3.1.2 Image plate (IP) data collection statisics for collection along a-axis.

Parameter	Value
Source	Beam-line 6A2, Photon Factory
Method	Screenless Weissenberg photography
Wavelength	1.04 Å
Crystal to IP distance	430 mm
Coupling constant (mm/°)	1.0
Oscillation range (ω°/IP)	8.0
Number of osc^{n} / IP	6
ω scan speed (°/s)	2.0
Exposure per image (s)	48
Number of IP's	12
Total exposure (s)	576

Data processing

The image plate data were processed using the program WEIS (Higashi, 1991; see appendix I) using the positions of the fiducial spots that had been recorded for each image. Each image was first oriented using the fiducial marks, after which a number of medium-strong reflections were chosen and used to refine the initial orientation matrix parameters. Initially data to 10 Å were used for refinement and then once the missetting parameters had been determined and DR and DS (DR and DS are *rms* differences between observed and calculated spot positions) had reached a minimum the resolution of the data used for refinement was extended to 4.0 Å. Once the orientation matrix had been determined the intensity of each reflection was measured using a profile fitting method. All the data within a sphere of radius 1.8 Å were processed. Details of the c-axis data processsing are given in Table 2.3.1.3 and of the a-axis data processing are in Table 2.3.1.4.

Table 2.3.1.3 Data collection and processing statistics for the c-axis *Alcaligenes* sp cytochrome c' images.

ω range represents the oscillation range for which data were collected for each image. The coupling constant was kept the same thoughout the data collection. DR and DS give a measure of the missetting of the centre of each reflection, DR is the *rms* difference between the observed and calculated spots of the horizontal displacement and DS the *rms* difference of the observed and calculated spots of the vertical displacement. RMS/BGD is the ratio of the standard error to the background level of density.

Image plate	ω range	Coupling constant	% expected reflections observed	Number of whole reflections	DR	DS	RMS/ BGD
1	-3.6° - 5.4°	2.0	84.3	5269	0.20	0.23	0.8
2	4.9° - 13.9°	2.0	84.2	5187	0.22	0.17	0.8
3	13.4° - 22.4°	2.0	80.7	4823	0.18	0.19	1.2
4	21.9° - 30.9°	2.0	83.1	5032	0.24	0.21	0.7
5	30.4° - 39.4°	2.0	82.9	5099	0.23	0.18	0.8
6	38.9° - 47.9°	2.0	80.3	4795	0.19	0.17	0.9
7	47.4° - 56.4°	2.0	85.5	5287	0.20	0.20	0.9
8	55.9° - 64.9°	2.0	85.8	5256	0.18	0.18	0.8

-Total range of 68.5 ° (0.5 ° overlap per image).

Table 2.3.1.4 Data collection and processing statistics for the a-axis *Alcaligenes* sp cytochrome c' images.

Image plate	ω range	Coupling constant	% expected reflections observed	Number of whole reflections	DR	DS	RMS/ BGD
1	-21.0° - -14.0°	1.0	71.0	4213	0.27	0.29	4.8
2	-14.5° - -6.5°	1.0	72.4	4250	0.27	0.29	2.9
3	-7.0° - 1.0°	1.0	75.8	4460	0.27	0.31	2.6
4	0.5° - 8.5°	1.0	69.1	4093	0.33	0.40	1.8
6	15.5° - 23.5°	1.0	74.1	4325	0.26	0.31	1.5
7	23.0° - 31.0°	1.0	75.6	4458	0.22	0.27	1.6
8	30.5° - 38.5°	1.0	75.6	4441	0.23	0.26	1.5
9	38.0° - 46.0°	1.0	77.1	4562	0.24	0.29	1.5
10	45.5° - 53.5°	1.0	78.7	4647	0.24	0.32	1.5
11	53.0° - 61.0°	1.0	77.8	4598	0.23	0.29	1.3
12	60.5° - 68.5°	1.0	79.3	4710	0.24	0.26	1.4
13	68.0° - 76.0°	1.0	80.1	4731	0.20	0.23	2.0

- Image plate 5 was not processed as it was an inverted image
- Total range of 97 ° (0.5 ° overlap per image).

The integrated intensities of each image from WEIS were combined and converted into a form which could be read by INTAV3 (see appendix I) and INTAV4 (see appendix I), and these programs were then used to scale and merge the 20 image plate data files. Table 2.3.1.5 gives the output statistics from INTAV4 for the synchrotron data.

Table 2.3.1.5 Final statistics from INTAV4 for the Photon factory data collected from *Alcaligenes* sp cytochrome c'.

Source of data set	synchrotron (c and a axes)
No. of observations	75,683
R_{merge}	0.075

The reflections were then indexed into the unique part of reciprocal space using the program HEXRED (see appendix I) giving a total of 14,674 reflections to a resolution of 1.8 Å. Table 2.3.1.6 gives the Photon factory data statistics as a function of resolution and Figure 2.3.1.1 shows the % completeness versus resolution for this data set. This figure shows the quality of the data set to high resolution with very little reduction in completeness at the 1.8 Å resolution limit.

Table 2.3.1.6 Statistics for Photon Factory data as a function of resolution.

d_{min} (Å)	Reflections	Completeness in shell (%)	Cumulative completeness (%)	$I > 3\sigma_I$ (%)
7.95	240	99.0	99.0	95.0
5.66	374	100.0	99.7	92.5
4.63	451	99.6	99.6	95.6
4.01	521	100.0	99.7	97.5
3.59	578	99.2	99.5	97.8
3.28	639	100.0	99.7	95.8
3.04	661	98.3	99.7	93.9
2.84	716	99.6	99.7	90.6
2.68	750	98.8	99.6	88.1
2.54	777	98.2	99.4	85.3
2.43	814	98.1	99.2	83.8
2.32	835	96.6	98.9	78.9
2.25	862	96.1	98.6	74.5
2.15	874	94.5	98.3	68.4
2.08	918	95.7	98.0	62.4
2.01	896	91.0	97.4	55.1
1.95	940	93.0	97.0	50.1
1.90	942	80.4	95.5	44.1
1.85	932	87.6	95.0	38.8
1.80	954	87.4	94.4	31.1
Total	14674		94.4	71.4

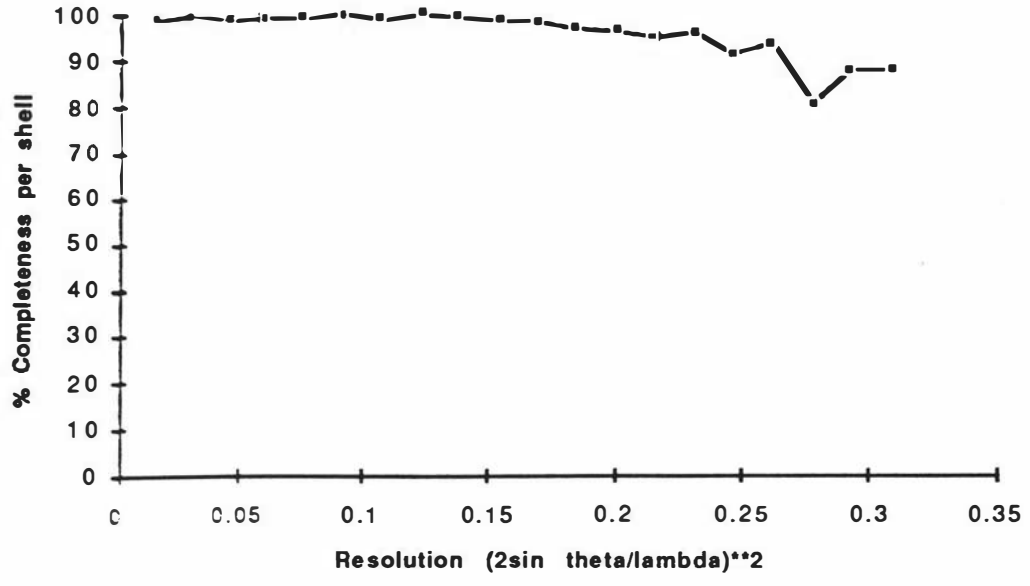


Figure 2.3.1.2 Plot showing the % intensities greater than 3σ as a function of resolution for the Photon Factory set.

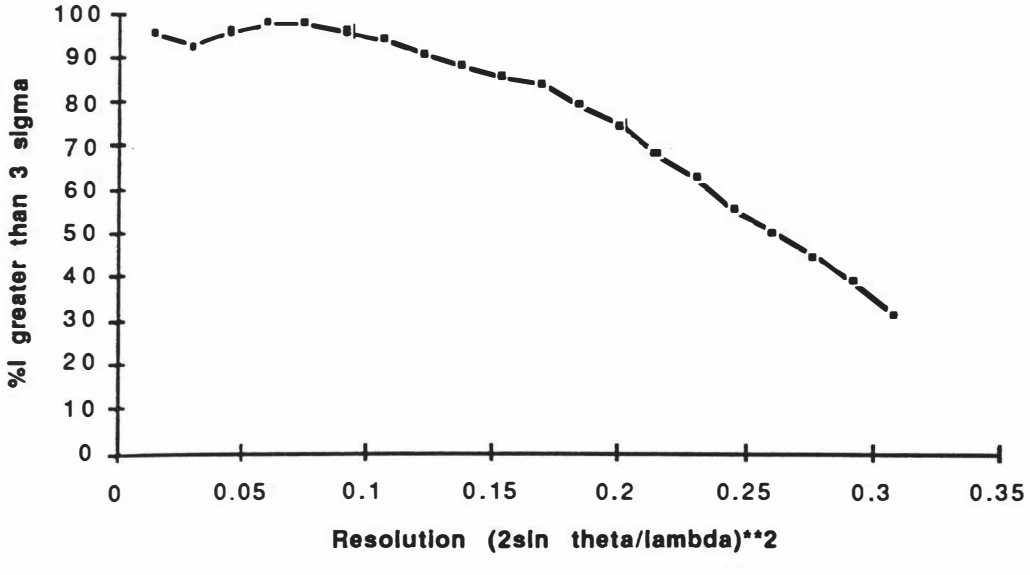


Figure 2.3.1.1 Plot giving % completeness against resolution for the *Alcaligenes* sp cytochrome c' data set from the Photon factory.

2.3.2 Anomalous data from *Alcaligenes* sp.

Early in this study the opportunity arose to collect several data sets at the University of Oregon. Two crystals soaked in heavy atom compounds (K_2PtCl_4 and ethyl mercury chloride) were chosen, in the hope that they might provide isomorphous derivatives. In the event, they showed no heavy atom substitution at all, but the data were of such quality, that they showed reliable anomalous scattering from the single iron atom and were consequently used as native data sets later in the structure analysis.

Data Collection

The data were collected at the University of Oregon (USA), using a Xuong-Hamlin multiwire area detector (Xuong *et al.*, 1985) with graphite-monochromated CuK_α radiation from a Rigaku RU200 rotating anode generator. The maximum resolution of the data collected was 2.8 Å. This was done in order to maximise the completeness of the data and in particular to optimise the measurement of anomalous differences. The two detectors were placed at distances from the crystal of 117.04 cm, ϕ = -24.0 °, and 110.80 cm, ϕ = 15.0 ° to give essentially complete coverage of the Friedel pairs. Each frame had a step of 0.1 ° and counts were accumulated for 45 seconds per step. Although the spacegroup of the crystals was either $P6_122$ or $P6_522$ the data was collected in $P6_1$.

Data processing

The data collected from the two crystals from *Alcaligenes* sp were processed with the supplied detector software (Howard *et al.*, 1985). Both sets of data were then processed into the correct space group (ie. the appropriate equivalent reflections were merged). Table 2.3.2.1 shows the data processing statistics from the better of the two crystals, since this data set was used for subsequent work. Table 2.3.2.2 and Figures 2.3.2.1 and 2.3.2.2 give statistics of the data set collected as a function of resolution.

Table 2.3.2.1 Data processing statistics for the Xuong-Hamlin data from a native crystal of *Alcaligenes* sp cytochrome c'.

Resolution (Å)	2.82
Number of observations	15,232
Completeness (%)	94.0
Unique reflections	4,037
R_{merge}[a]	0.022
R_F[b]	0.024

(a) R_{merge} = $\Sigma|I-\bar{I}|/\Sigma\bar{I}$, where the summation is over all redundant measurements (but not Friedel pairs)

(b) R_F = $\Sigma||I^+-I^-|/\Sigma(I^++I^-)$ where I^+ and I^- are Friedel pairs, and the summation is performed after merging redundant measurements.

Table 2.3.2.2 Statistics for the Xuong-Hamlin data as a function of resolution.

d_{min} (Å)	Reflections	Completeness in shell (%)	Cumulative completeness (%)	$I > 3\sigma_I$ (%)
8.73	104	59.6	59.6	100.0
6.22	281	98.9	84.0	96.1
5.09	348	100.0	90.9	94.5
4.42	395	99.5	93.8	96.5
3.95	438	100.0	95.4	97.3
3.61	476	98.2	96.0	96.2
3.34	518	100.0	96.8	94.4
3.13	544	99.4	97.3	90.4
2.95	566	98.5	97.5	89.2
2.80	367	61.0	93.3	86.4
Total	4037		93.3	93.4

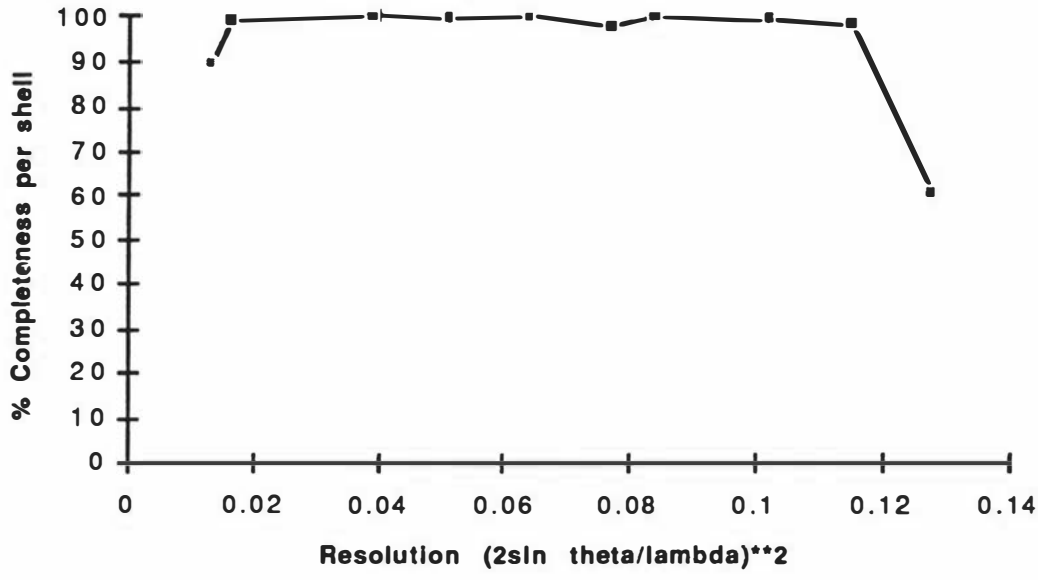


Figure 2.3.2.1 Plot of % completeness versus resolution for the Xuong-Hamlin data collected from *Alcaligenes* sp cytochrome c'.

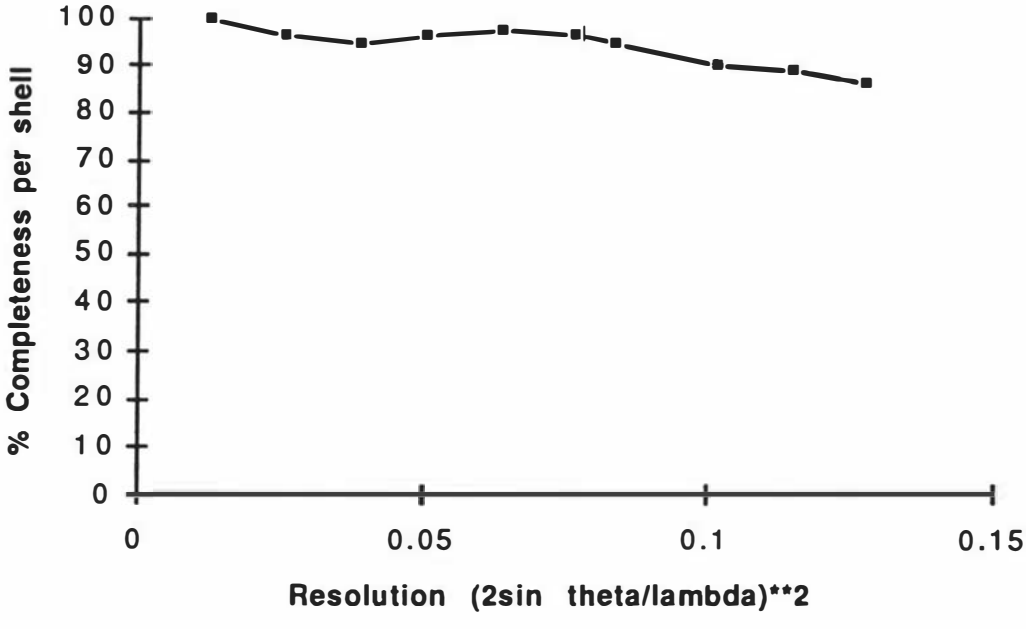


Figure 2.3.2.2 Plot of % intensities greater than 3σ as a function of resolution for the Xuong-Hamlin data collected from *Alcaligenes* sp cytochrome c'.

2.3.3 Heavy atom derivative data from *Alcaligenes* sp.

Various heavy atom compounds were used in soaking experiments in attempts to obtain isomorphous heavy atom derivatives. The complexes tried are given in section 4.5.1, but of these only one yielded a suitable derivative (K_2PtCl_4).

Data Collection

Details of the heavy atom soaking are given in section 4.5.1. All the crystals of the potential heavy atom derivatives were mounted and aligned in the same fashion with the c-axis along the capillary tube. All data were collected on an R-Axis IIC imaging plate system, with graphite-monochromated CuK_α radiation from a Rigaku RU200 rotating anode generator operating at 50 KV and 100 mA. The unit cell and orientation matrix of each crystal were determined by collecting three still images at ϕ angles of 0 °, 30 ° and 60 °, and auto-indexing the observed reflections, using the supplied software (see appendix I). In each case, two initial images (each with a 2.0 ° oscillation range) were collected and processed with the software provided (see appendix I), and these data were then scaled against the native data to determine whether the crystal was a possible derivative. If the intensity difference was greater than 10 % it was considered a possible derivative, and data collection was continued. The only apparent substitution occurred with crystals soaked in K_2PtCl_4. As this was the only derivative obtained, two different sets of soaking conditions were tried (0.5 mg ml^{-1} (1 mM) and 2 mg ml^{-1} (4 mM) each for a period of 4 days) in order to see if varying levels of substitution could be obtained to provide more phase information. Although it was apparent that both conditions gave similar substitution, both sets of derivative data were used in phase determination (see chapter 4). The data collection for PT1 (2 mg ml^{-1} K_2PtCl_4) comprised 30 frames ($\Delta\phi$ = 2.0 °), each with an exposure time of 60 minutes, while for PT2 (0.5 mg ml^{-1} K_2PtCl_4) the data set comprised 15 frames with $\Delta\phi$ and the exposure time the same. The crystal to detector distance was set at 120.0 cm to give a minimum spot - spot distance of approximately 1 mm. This distance depends on the longest cell length that the crystal has, following the equation:

$$\delta = \frac{\lambda D}{l}$$

where λ= Cuk$_\alpha$ 1.5418 Å,

l = Max. cell edge length (approximately 185 Å),

D = Detector distance,

δ = Minimum spot-spot distance.

Data Processing

All data processing was carried out using the software provided with the image plate system (see appendix I) on a VAXstation 4000 computer system. The crystal orientation and unit cell dimensions were first determined from the three still images taken with each data collection. These were further refined using the first oscillation image, in order to give the best possible prediction of reflection positions. The software also allowed for the possibility of crystal slippage during data collection. Each reflection was integrated within a box whose dimensions were chosen to get the best separation of reflection and background. After integration, the reflections were then merged into a single file, and the frames were scaled using an inverse scale factor and a temperature factor. Reflections with intensities greater than 1σ were used in the scaling, and refinement was deemed complete when the ratio of shift/standard deviation was less than 0.001. The symmetry equivalent reflections were then averaged if their intensities were greater than 1σ. Reflections were rejected if the difference between a certain reflection intensity and that of its equivalent reflection was greater than 0.3 times the mean reflection intensity of the data set plus 0.1 times the mean intensity of the particular reflections. All the above processes are described in more detail in the users manual (Instruction manual for R-AXIS IIC and software: Manual number ME201LR1). Table 2.3.3.1 gives the data processing statistics for the two K_2PtCl_4 derivatives. Partial reflections are included in the data sets and in the calculations based on these data sets. Table 2.3.3.2 and Figures 2.3.3.1 and 2.3.3.2 give data processing statistics as a function of resolution.

Table 2.3.3.1 Data processing statistics for platinum derivatives from *Alcaligenes* sp.

	PT1	PT2
Resolution (Å)	2.32	2.31
Number of observations	46,282	21,007
Unique reflections	7,062	7,046
Reflections removed (%)	0.2	0.0
Completeness (%)	97.4	93.3
R_{merge}[a]	0.035	0.037
R_{iso}[b]	0.151	0.138

(a) $R_{merge} = \Sigma|I-\bar{I}|/\Sigma\bar{I}$, where the summation is over all redundant measurements (but not Friedel pairs)

(b) $R_{iso} = (\Sigma||F_{PH} - F_P|| / \Sigma|F_P|$, $|F_{PH}|$ and $|F_P|$ are the measured structure amplitudes of the derivative and native protein in electrons.

Table 2.3.3.2 Statistics for the PT1 data as a function of resolution.

d_{min} (Å)	Reflections	Completeness in shell (%)	Cumulative completeness (%)	$I > 3\sigma_I$ (%)
7.36	256	88.1	88.1	91.8
5.23	457	99.1	94.8	94.3
4.78	566	99.5	96.8	97.7
3.71	655	100.0	97.9	97.6
3.32	734	100.0	98.5	95.9
3.03	776	98.8	98.5	94.2
2.81	844	99.6	98.8	91.8
2.63	884	98.2	98.9	87.1
2.48	923	97.3	98.6	85.8
2.35	967	91.5	97.6	77.6
Total	7062		97.6	90.3

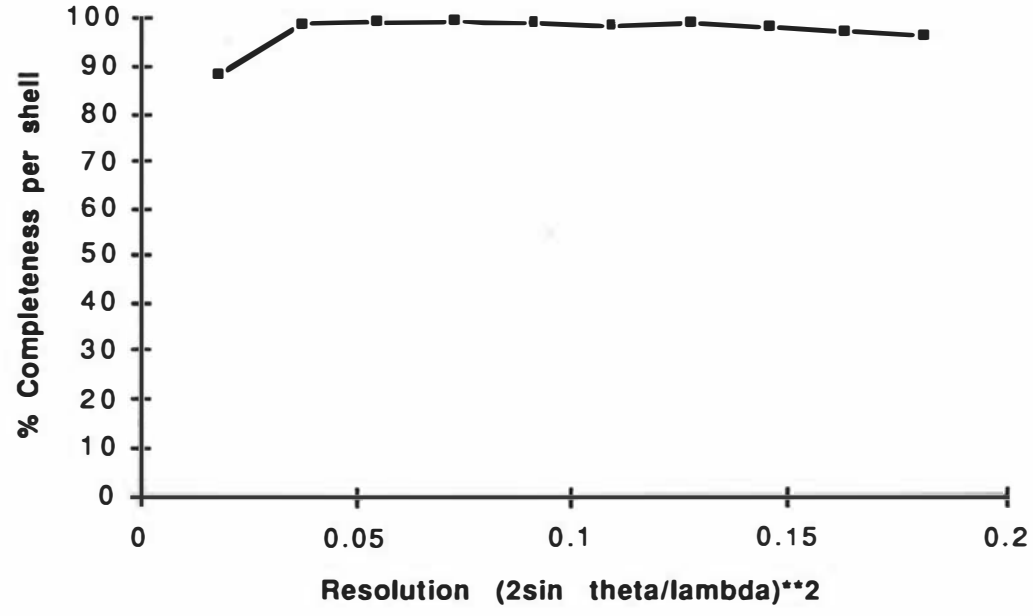


Figure 2.3.3.1 Plot of % completeness as a function of resolution for the heavy atom derivative data set from *Alcaligenes* sp cytochrome c' (PT1).

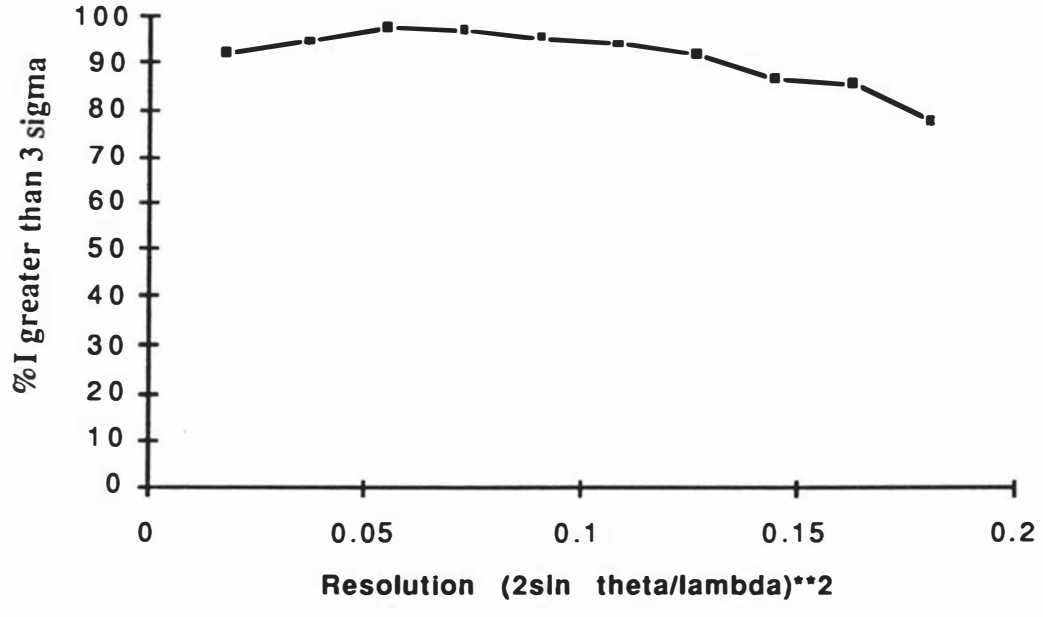


Figure 2.3.3.2 Plot of % intensities greater then 3σ as a function of resolution for the heavy atom derivative data set collected (PT1).

Scaling with native Alcaligenes sp data

The derivative data sets were scaled against the *Alcaligenes* sp native data collected at the Photon factory in Japan, using the program SCALEIT (CCP4 suite, see appendix I). Scale and anisotropic temperature factors were determined over five cycles of refinement or until all of the shifts were less than ABS (value used was 0.01) times the standard deviation of the parameter (two scale and six beta parameters in this case for anisotropic refinement). Reflections were rejected from the scale determination if F_{PH} was greater than a certain value or less than three times the standard deviation. Tables 2.3.3.3 and 2.3.3.4 show the scales applied to the derivative data sets as a function of resolution, together with the R_{iso} values.

Table 2.3.3.3 Scale and R_{iso} values for PT1 derivative.

Resolution (Å)	No. of reflections	Scale	R_{iso}
11.3	93	1.048	0.251
6.5	301	1.084	0.178
5.1	410	1.017	0.152
4.3	474	1.027	0.136
3.8	538	1.016	0.132
3.4	609	1.033	0.132
3.1	640	1.017	0.145
2.9	684	1.039	0.149
2.7	704	1.050	0.150
2.6	728	1.041	0.155
2.5	755	1.051	0.164
2.4	750	1.055	0.177
2.3	379	1.055	0.162
Total	7062	1.035	0.151

Table 2.3.3.4 Scale and R_{iso} values for PT2 derivative.

Resolution (Å)	No. of reflections	Scale	R_{iso}
11.3	83	0.990	0.246
6.5	296	1.010	0.159
5.1	402	0.970	0.152
4.3	463	0.992	0.126
3.8	533	0.985	0.126
3.4	600	0.997	0.119
3.1	635	0.980	0.132
2.9	690	1.004	0.136
2.7	703	1.019	0.130
2.6	740	1.023	0.141
2.5	758	1.031	0.144
2.4	779	1.047	0.162
2.3	437	1.028	0.151
total	7046	0.998	0.138

As can be seen the derivative data scales against the native data quite consistently, with the isomorphous differences relatively constant over the full resolution range. Figure 2.3.3.3 shows R_{iso} as a function of resolution.

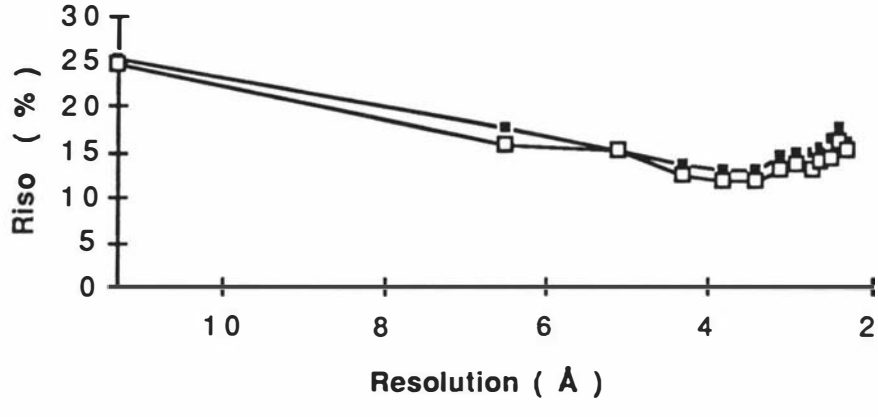


Figure 2.3.3.3 Graph of R_{iso} as a function of resolution for the K_2PtCl_4 derivatives (PT1 filled squares and PT2 empty squares).

2.3.4 R-Axis data collection - *Alcaligenes denitrificans.*

Data collection

Data were collected from a single crystal of *Alcaligenes denitrificans* cytochrome c', using the R-Axis goniometer and following the same procedures as used for the heavy atom derivative data collection (see section 2.3.3). A total of 30 frames covering a total of 60 ° were collected, with each frame being exposed for 50 minutes. Table 2.3.4.1 compares the cell dimensions between the two species of cytochrome c'. These revealed a high degree of isomorphism.

Table 2.3.4.1 Cell dimensions of *Alcaligenes* sp and *Alcaligenes denitrificans* cytochromes c'.

Axis	*Alcaligenes* sp	*Alcaligenes denitrificans*
a	54.4 Å	54.6 Å
b	54.4 Å	54.6 Å
c	181.1 Å	180.4 Å

Data Processing

The data were processed as for the heavy atom derivative data in section 2.3.3. Data processing statistics are given in Table 2.3.4.2. It can be seen that although the value for R_F is less than the value for R_{merge} the data still gave rise to a measurable anomalous signal from which the iron positions could be experimentally determined (see section 4.4.5.1). Table 2.3.4.3 gives the data processing statistics as a function of resolution. Figures 2.3.4.1 and 2.3.4.2 show the % completeness against resolution and % intensities greater than 3σ respectively. It can be seen that the completeness falls off with increasing resolution from about 2.25 Å. Although in later refinement all of the data were used, it is more realistic to quote the resolution of the final structure as about 2.15 Å

Table 2.3.4.2 Data processing statistics for *Alcaligenes denitrificans* cytochrome c' data.

	Alcaligenes denitrificans
Resolution (Å)	2.15
Number of observations	60,526
Unique relections[a]	8,220
% reflection removed	0.09
% completeness (to 2.15 Å)	89.5
R_{merge}[b]	0.032
R_F[c]	0.028

(a) 6,105 reflections with anomalous differences.

(b) R_{merge} $=\Sigma|I-\bar{I}|/\Sigma\bar{I}$, where the summation is over all redundant measurements (butnot Friedel pairs)

(c) R_F $= \Sigma|I^+-I^-|/\Sigma(I^++I^-)$ where I^+ and I^- are Friedel pairs, and the summation is performed after merging redundant measurements.

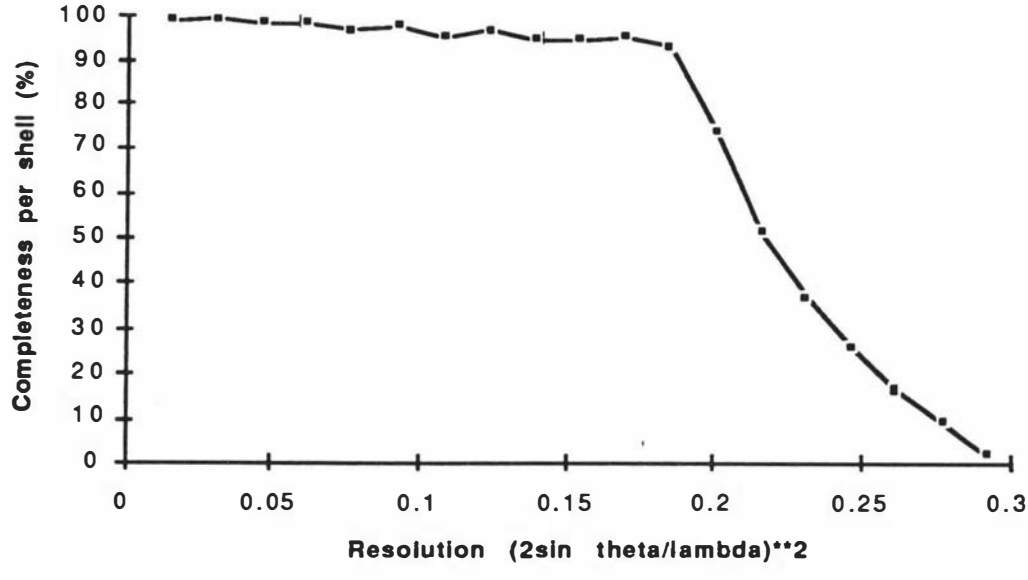


Figure 2.3.4.1 Plot of % completeness vs resolution for R-axis data set collected from *Alcaligenes denitrificans* cytochrome c'.

Table 2.3.4.3 Statistics for R-Axis *Alcaligenes denitrificans* cytochrome c' data as a function of resolution.

d_{min} (Å)	Reflections	Completeness in shell (%)	Cumulative completeness (%)	$I > 3\sigma$ (%)
7.95	237	98.8	98.8	95.8
5.66	367	99.7	99.3	95.9
4.63	442	98.5	99.0	94.1
4.01	505	98.5	98.8	95.8
3.59	559	97.1	98.3	96.6
3.28	612	97.8	98.2	93.0
3.04	632	95.4	97.7	91.0
2.84	691	96.9	97.5	90.6
2.68	716	95.0	97.1	87.3
2.54	746	94.9	96.8	83.0
2.43	789	95.3	96.6	82.3
2.32	802	93.6	96.2	78.7
2.23	656	73.8	93.8	77.6
2.15	470	51.1	89.5	70.6
2.08	352	37.1	84.5	63.9
2.01	250	25.6	79.3	54.8
1.95	168	16.6	74.0	47.0
1.90	100	9.5	68.8	39.0
1.85	29	2.2	63.0	45.8
Total	9118		63.0	83.8

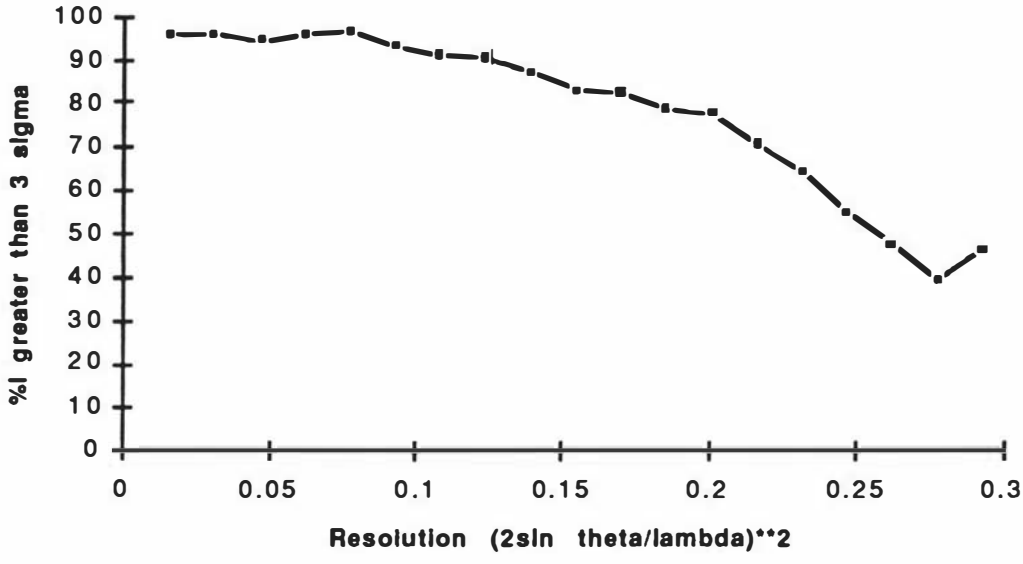


Figure 2.3.4.2 Plot of % intensities greater than 3σ as a function of resolution for the R-axis data set collected from *Alcaligenes denitrificans* cytochrome c'.

Chapter Three

Molecular Replacement

Together with isomorphous replacement (chapter 4) the method of molecular replacement is of major importance in the structure solution of biological macromolecules. This technique was first introduced in the early 1960's (Rossman and Blow, 1962) to use structural redundancy, a consequence of noncrystallographic symmetry, for phase determination. The application of this method to the determination of protein structures using the known structures of related proteins forms an alternative approach to isomorphous replacement. The approach however, does depend on a fairly closely related protein structure being available. The concept involves three stages;

(1) Determination of the orientation of the molecule in its unit cell by rotation of the search model,

(2) Finding the position of the orientated model in the unit cell ("translation"),

(3) Computation of an initial set of phases which can be used to calculate electron density maps for improvement or correction of the model.

3.1 The rotation function.

The rotation function is used to find the correct orientation of the model in the crystal lattice. The search is carried out by looking for maximum overlap between the Patterson functions of the search and target structures. At the relative orientation that superposes the model on to the target, the product of the two Patterson functions (which are maps of interatomic vectors) should have a large value. The function was originally formulated (Rossman and Blow, 1962) as;

$$R(C) = \int_{u} P_1(x)P_2(x')\, dV \qquad \text{(Eqn 3.1)}$$

where P_1 and P_2 are the two Patterson functions, C is a matrix defining a rotation of x with respect to x', and U is a volume of integration, usually spherical, centred at the origin. A maximum in the rotation function R(C) indicates a potential orientation for the

search molecule in the target crystal. For mathematical convenience this can be reformulated as an integral over all space

$$R(C) = \int_{\text{all space}} P_1(x)U(x)P_2(Cx)\, dV \qquad \text{(Eqn 3.2)}$$

where U(x) is equal to one inside the sphere of integration and zero elsewhere. In reciprocal space this becomes

$$R(C) = \sum_h I_{obs}(h)F_m^2(Ch) \qquad \text{(Eqn 3.3)}$$

where $I_{obs}(h)$ are the observed intensities and $F_m(Ch)$, the Fourier transform of $U(x)P_2(Cx)$, is a continuous function defined over all reciprocal space. This formulation, first used by Lattman (Lattman and Love, 1972), is clear and flexible in application but unfortunately slow to compute.

More usually the rotation function is computed by the fast Fourier method (Crowther, 1972). Instead of rotating spherical volumes, it was more appropriate to expand the Patterson density in a sphere in terms of Bessel spherical harmonics.

3.2 The translation function.

The second part of molecular replacement consists of a translation function. Once the search model has been orientated correctly in the cell, a translation function is used to translate the model into the correct position. The translation function is theoretically related to the rotation function. When the reference model, once correctly orientated, is translated in the unknown cell, the symmetry related molecules move concurrently and the intermolecular vectors change. The correlation between cross-vectors of the model and its symmetry equivalents as it is moved through the unknown unit cell and the cross-vectors inherent in the unknown Patterson function is calculated. The T function was introduced by Crowther and Blow (1967). It is a derivative of the general form of the translation function

$$T(x) = \sum_h Iobs(h)F_c^2(h,t) \qquad \text{(Eqn 3.4)}$$

$F_c^2(h,t)$ is the calculated intensity for the rotated model shifted from its origin by the vector t. If the unit cell contains two molecules related by a two-fold axis Eqn 3.4 can be written

$$F_c^2(h,t) = F_c(h,t)F_c^*(h,t) = (F_{m1} + F_{m2})(F_{m1}^* + F_{m2}^*) \qquad \text{(Eqn 3.5)}$$

F_{m1} and F_{m2} are the molecular transforms of the two molecules sampled at point h in reciprocal space, and * is the complex conjugate. If the reference molecule is shifted by t from its initial position only the phase of its molecular transform changes and F_m becomes $F_m e^{2\pi iht}$. The calculated intensities can now be expressed as follows.

$$F_c^2(h,t) = (F_{m1}(h)e^{2\pi iht_1} + F_{m2}(h)e^{2\pi iht_2})(F_{m1}^*(h)e^{2\pi iht_1} + F_{m2}^*(h)e^{-2\pi iht_2}) \qquad \text{(Eqn 3.6)}$$

These molecular transforms and their positions t_1 and t_2 are related by two-fold symmetry ($t_1 = -t_2$). The above expression can be substituted into the above equation (Eqn 3.4).

$$T(t) = \sum_h I_{obs}(h)[\ F_{m1}^2 + F_{m2}^2 + F_{m1} F_{m2}^* e^{2\pi iht} + F_{m1}^* F_{m2} e^{-2\pi iht}\ ] \qquad \text{(Eqn 3.7)}$$

where t is equal to t_1-t_2.

Since only the cross vector terms are required, ie those dependent on the relative positions of the molecules, the first two F terms can be omitted as these represent the self vectors for the two molecules. For simplicity one of the cross terms is also omitted. This results in what is known as the T function.

$$T(t) = \sum_h I_{obs}(h)F_{m1} F_{m2}^* e^{-2\pi iht} \qquad \text{(Eqn 3.8)}$$

Most translation functions use a slight variation of the T function called the T2 function.

The T2 function combines all the symmetry

$$T2(t) = \sum_h I_{cross}(h) \times 2\sum_i \sum_{j<i} F_{m1}F_{mj}^* e^{-2\pi iht} \qquad \text{(Eqn 3.9)}$$

where $I_{cross}(h)$ are the observed intensities with the self vectors subtracted, F_{m1} is the Fourier transform of one of the molecules in the crystal, h is the reciprocal vector *hkl* and t is the translation vector between molecule m1 and mj.

The T2 function applies phase shifts so that all the peaks converge at the same place, defined by a single translation vector t for the reference molecule. This means the signal/noise ratio is much better.

3.3 Computational methods.

Three different program suites were available for the application of molecular replacement methods viz ALMN/TFFC, X-PLOR and AMORE.

3.3.1 ALMN.

Cross-rotation searches were performed using the 60-Bessel-function version of ALMN (CCP4 suite, see appendix I), based on Crowther's (1972) fast rotation algorithm. A Patterson function is a collection of inter-atomic vectors; some will be vectors between atoms within the same molecule, i.e. self-vectors, and some will be vectors between neighbouring molecules, i.e. cross-vectors. In the rotation function the aim is to include as many self-vectors and to eliminate as many cross-vectors as possible to discriminate between the noise from the intermolecular vectors and signal from the self-vectors. The coordinates of the models were placed in a hypothetical orthogonal cell with no symmetry (spacegroup P1) with cell edges 100 Å in length so that most of the cross-vectors were larger than the chosen radius of integration (Lifchitz, 1983).

The origin peak was removed by keeping the radius of integration smaller than the smallest native unit cell dimension minus twice the resolution limit minus the diameter of the origin peak (Lifchitz, 1983).

The atomic temperature factors of the search models were used unchanged in the calculation as these tend to be related between structures and their inclusion lends to the generation of more accurate structure factors (Dodson, 1985).

Structure factors were calculated from the search models using SFALL (CCP4 suite, see appendix I). These calculated structure factors were used to compute cross-rotation functions with the observed structure-factor amplitudes from the *Alcaligenes* sp. crystal. No self-rotation functions were calculated since the asymmetric unit cell only contained a monomer. The calculated structure factors and the observed data were both converted into normalised structure amplitudes using the program ECALC (CCP4 suite, see appendix I), to remove the possibility that the rotation function might be dominated by a few large reflections (Dodson, 1985; Tickle, 1985). These were then combined in a cross-rotation function.

The radius of integration used was between 3 Å and 30 Å (inner sphere 3 Å, outer sphere 30 Å). Data within the resolution ranges 15.0 - 3.5 Å and 10.0 - 4.0 Å were used. The Eulerian angle system was used to define the rotation angles (ALMN also

outputs spherical polar angles and direction cosines). A representation of this system is given in Figure 3.3.1.1. The Eulerian angles (α,β and γ) are defined as;

(i) A rotation α of the coordinate system about Z
(ii) A rotation β about the new direction of Y
(iii) A rotation γ about the new direction of Z.

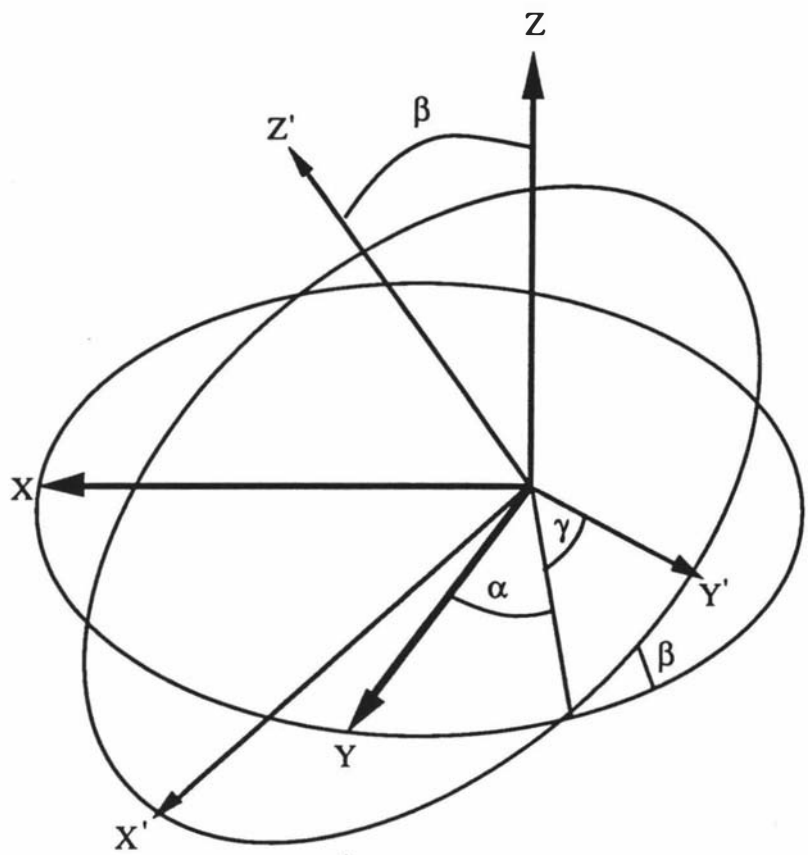


Figure 3.3.1.1 Graphical representation of the Eulerian set of angles used in cross rotation programs.

3.3.2 TFFC.

This program computes a full-symmetry Translation Function. TFFC uses structure factor information (amplitudes and phases) from all molecules in the unit cell whose positions have already been determined or are about to be determined. Three important features which further improve the signal/noise ratio of the T2 function are the shell-scaling of the F(obs) and F(calc) values by the difference Wilson plot method, the use of normalised amplitudes (E-values) and the subtraction of the transform of the intramolecular vector set from the transform of the Patterson function of the target structure.

3.3.3 X-PLOR.

X-PLOR is a versatile program suite for computational structural biology. It was the first program to incorporate X-ray diffraction data and molecular dynamics for refinement (Brünger *et al.*, 1987). Since its initial release it now contains programs designed for structure determination as well as refinement, including rotation and translation programs for molecular replacment.

Rotation program

X-PLOR performs conventional rotation searches via a real-space Patterson search method. A reference Patterson map P2 is calculated on a specified grid from the observed intensities. The Patterson map P1 to be rotated is calculated from a search model and the largest vectors compiled into a list. The vectors are rotated using the Eulerian angle system α, β and γ (Rossman and Blow, 1962; see also Figure 3.3.1.1), although other angle systems can be used. The rotation search is normally restricted to the asymmetric unit of the rotation function (Rao *et al.*, 1980).

The product function (RF) is calculated at each sampled orientation, where

$$RF(\Omega) = < P_{obs} P_{model}(\Omega) >$$

P_{model} and P_{obs} are the rotated vectors of P1 and the interpolated values of the Patterson map P2 respectively. Once the rotation search has finished, all the sampled grid points are sorted, their product correlation values compared and a single peak search is carried out

using a matrix metric method (Brünger, 1990). For two rotation matrices Ω^1, Ω^2, the metric is defined as

$$m(\Omega^1,\Omega^2) = \min_{s=1\ n_s} \sqrt{Tr\{(\Omega^1 - O_s\Omega^2)(\Omega^1 - O_s\Omega^2)\}}$$

where n_s = number of symmetry operators and o_s = rotational part of the symmetry operators. Two RF grid points are considered to be in the same cluster if the corresponding rotation matrices yield $m(\Omega^1,\Omega^2)<\varepsilon$. For example if ε is set to 0.25, two grid points are said to be in the same cluster if they are not mutually different by more than 10 °. This peak search removes grid points that are close to grid points with larger RF values.

Patterson Correlation (PC) Refinement

Once the highest peaks of the rotation function have been determined, PC-refinement of the highest peaks is carried out. PC-refinement (Brünger, 1990) is a filtering process whereby a defined number of the top rotation angles from the cross-rotation function are correlated and minimised against an energy function. The target function for Patterson refinement is proportional to the negative correlation coefficient between the squared amplitudes of the observed and calculated normalised structure factors. The result should be that out of the sets of angles input, one set should have a greater correlation than the rest. Also, the rotation angles in each set are refined simultaneously.

Translation program

The search routine computes the structure factors F_{calc} of the translated primary molecule and the symmetry-related molecules by applying appropriate phase-shift operators in reciprocal space to the calculated structure factors of the original (not translated) molecule and its symmetry mates. The search is carried out over a 3D grid that can be specified in fractional or orthogonal coordinates. The asymmetric unit of the translation function is equivalent to the unit cell of its Cheshire group (Hirshfeld, 1968).

3.3.4 AMORE.

AMORE (CCP4 suite, see appendix I) (automated molecular replacement) is a molecular replacement package which is based on improved algorithms in the rotation function (in terms of generation of radial quadrature instead of Bessel expansion), and organised for automation and rapid solution (Navaza, 1994). The package is set up as follows;

Data processing

Data processing is performed in two parts:
(1) SORTING; This program sorts, packs and assesses the quality of the measured data.
(2)TABLING; This calculates the continuous Fourier coefficients derived from various search models. The coefficients are calculated by Fourier transformation of the electron density derived from the atomic coordinates of the molecule, which has been placed in a cell with dimensions approximately four times the size of the model. To avoid using an excessively large cell in cases where there is an unfavourable position of the molecule, the centre of mass of the atoms is shifted to the origin and the coordinates are rotated in order to align their principal axes of inertia with the cell axes; this is the initial reference position.

Rotation program

Calculation of the cross rotation function is carried out using ROTTING. This program calculates the fast rotation function (Crowther, 1972). The integration radius is usually taken as the distance from the centre of mass to the farthest atom (the inner radius is set to zero) although this radius may give cross vector overlap. The structure factors are calculated by simple interpolation (Lattman and Love, 1972) of the model, taking a P1 cell of size equal to the smallest box containing the model, plus the integration radius, plus the desired resolution over which the rotation function is to be calculated.

Translation program

The translation function is computed using TRAING, and is based on the translation function defined as the overlap of the observed and calculated Patterson

functions, for a given orientation of the search model (Crowther and Blow, 1967; Harada *et al.*, 1981). For each Eulerian set of angles output from the rotation function the model is rotated by the specified angles and translated within its Cheshire cell (Hirshfeld, 1968).

Rigid body refinement

Least-squares refinement of the positional parameters of the various solutions using FITING. First proposed by Huber and Schneider (1985), this uses the algorithms described by Castellano *et al.* (1992).

3.4 Search models.

Three 3D structures of cytochrome c' have already been determined, two of them at high resolution. A 1.67 Å model of cytochrome c' from *Rhodospirillum molischianum* (Weber *et al.*, 1981; Finzel *et al.*, 1985) has been deposited in the Brookhaven Data Bank (Bernstein *et al.*, 1977; entry 2CCY) and a 1.8 Å resolution structure from *Chromatium vinosum* (McRee *et al.*, 1993) has also been deposited (entry P1BBH). The third structure reported is a 2.8 Å resolution structure of cytochrome c' from *Rhodospirillum rubrum* (Yasui *et al.*, 1992). At the time of the molecular replacement studies, only two of the structures were available. These were the 1.67 Å model for *Rhodospirillum molischianum* cytochrome c' (Finzel *et al.*, 1985) and the 2.8 Å model for *Rhodospirillum rubrum* cytochrome c' (Yasui *et al.*, 1992; coordinates kindly supplied by Dr M. Yasui). Both form dimers, and in both cases there is one dimer in the crystal asymmetric unit. Sequence identity between the two species is low, however, ~ 25 %, and superposition of their monomer units shows correspondingly imprecise agreement; *rms* agreement of 2.5 Å when 90 % of C_α atoms are superimposed (omitting sites of insertions/deletions). The agreement is worst in the loops joining the four helices, but even if just the C_α atoms of the four helices are compared the *rms* agreement is still only 1.35 Å (for 71 C_α atoms), since the angular dispositions of the helices differ by small amounts (1.0 - 8.3 °) in the two structures (Yasui *et al.*, 1992). The amino acid sequence of the *Alcaligenes denitrificans* cytochrome c' is not known, but the high isomorphism between the *Alcaligenes denitrificans* and *Alcaligenes* sp crystals suggested that the two structures would be very similar. The *Alcaligenes* sp cytochrome c' sequence differs substantially from those of the two known structures, having only 29 % identity with *R. molischianum* and 27 % with *R. rubrum* (Ambler, 1981) There was no reason to choose either of these known structures in preference, therefore, and each was used as the basis for the construction of separate search models.

The two models were placed in the same orientation, by using the fact that the cytochrome c' molecules form dimers. In each case, the two-fold axis relating the monomers was aligned parallel to the z-axis (Moss, 1985) to improve signal/noise ratio in the rotation function. From these structures various search models were constructed (Tables 3.4.1.1 and 3.4.1.2). The idea of using various search models is to find the "best" model which fits the target structure. For example, sidechains that are known from sequence alignment to be different can be trimmed to alanine (C_β). It is best not to remove too much model as structure is normally conserved better than sequence and corresponding chains in related proteins may be chemically different but still have a similar fold. Also removing too many correct atoms reduces the signal that is being sought.

Table 3.4.1.1 Models used in molecular replacement studies using ALMN and X-PLOR.

Model name[a]	Residues	Atoms[b]	Comments
J_all	1 - 122	921	all sidechains
J_con	1 - 122	766	similar/identical sidechains[c]
J_poly	1 - 122	647	truncated to C_β
U_all	3 - 127	976	all sidechains
U_con	3 - 127	888	similar/identical sidechains[c]
U_poly	3 - 127	648	truncated to C_β

a - Models U_all, U_con, and U_poly derived from *Rhodospirillum molischianum* cytochrome c', J_all, J_con, and J_poly from *Rhodospirillum rubrum* cytochrome c'.

b - Includes haem group.

c - Sidechains (s/c's) were either all truncated to C_β atoms or were retained only where they were identical, or similar in geometry. In the latter case, "similarity" equated Phe with Tyr (minus O_η), Ser with Thr (minus $C_{\gamma 2}$), Thr with Val, Asn with Asp, Gln with Glu, Val with Ile (minus $C_{\delta 1}$), Lys with Arg (as far as C_δ) and Leu with Phe or Tyr (as far as $C_{\delta 1}$ and $C_{\delta 2}$).

Table 3.4.1.2 Search models for molecular replacement used in AMORE.

Model name [a]	Residues	Atoms [b]	Comments
Mol1	3-127	648	All sidechains truncated at C_β
Mol2	3-28; 40-63; 80-125	632	Loops AB and BC deleted. Similar or identical sidechains retained (51 s/c's)[c]
Rub1	1-122	647	All sidechains truncated at C_β
Rub2	1-25; 38-60;72-122	629	Loops AB and BC deleted. Similar or identical sidechains retained (51 s/c's)[c]

a - Models Mol1 and Mol2 derived from *Rhodospirillum molischianum* cytochrome c', Rub1 and Rub2 from *Rhodospirillum rubrum* cytochrome c'.

b - Including the haem group.

c - See Table 3.4.1.1.

3.5 Results.

3.5.1 Results using ALMN/TFFC.

Rotation function (ALMN)

In Table 3.5.1.1 the cross rotation input variables are given for the molecular replacement attempts. The values for the radius of integration and the size of the P1 box were determined as in section 3.3.1.

Table 3.5.1.1 Cross rotation input values for use in ALMN.

Resolution ranges	10.0 - 4.0 Å , 15.0 - 3.5 Å
Radius of integration	3 -3 0 Å
Size of P1 box	100 Å x 100 Å x 100 Å
β angle step	2.5 °
Data set	*Alcaligenes* sp, Photon factory data

The top five solutions from ALMN for the models derived from the *R. rubrum* cytochrome c' are given in Table 3.5.1.2, while Table 3.5.1.3 displays the solutions from the *R. molischianum* search models. *Rms* is defined as the root mean square of the Patterson density and the relative peak heights are given after the Eulerian angles. Also given in the Tables are the solutions obtained when E-values were used in the structure factor calculations.

Table 3.5.1.2 Cross rotation peaks from ALMN calculated from search models based on the *R. rubrum* structure. (numbers in parentheses are the peak height divided by the *rms* value)

10.0 - 4.0 Å	J_all (*rms* 51.1)	J_con (*rms* 46.3)	J_poly (*rms* 37.8)
	50.0 46.5 47.4 (2.79)	50.8 45.5 43.0 (2.55)	31.0 53.5 41.5 (2.37)
	9.4 52.4 55.7 (2.70)	36.5 46.0 63.5 (2.30)	36.0 55.2 38.7 (2.35)
	0.5 43.0 221.4 (2.56)	20.4 41.7 50.3 (2.20)	24.7 61.6 27.5 (2.20)
	53.2 53.9 38.0 (2.50)	359.1 45.9 222.8 (2.14)	33.3 74.5 272.4 (2.17)
	8.5 38.6 21.5 (2.44)	4.4 60.4 208.2 (2.04)	3.7 47.2 60.1 (2.24)
(E-values)	(Rms 48.1)	(Rms 46.0)	(Rms 37.1)
	49.9 46.2 47.3 (2.79)	50.9 45.2 43.0 (2.56)	34.0 55.5 38.5 (2.33)
	9.8 52.1 55.7 (2.69)	37.3 46.0 62.2 (2.37)	54.3 53.8 33.4 (2.29)
	359.9 42.9 221.9 (2.63)	20.5 41.6 50.3 (2.23)	33.2 74.5 272.4 (2.21)
	53.6 54.2 37.5 (2.52)	358.5 45.8 223.6 (2.20)	24.0 61.3 27.1 (2.20)
	8.5 38.6 41.5 (2.49)	4.9 60.5 207.8 (2.08)	1.5 44.1 221.6 (2.18)

15.0 - 3.5 Å	J_all (Rms 37.6)	J_con (Rms 31.6)	J_poly (Rms 28.7)
	46.9 49.0 53.5 (2.35)	31.0 49.0 66.5 (2.13)	31.7 48.1 63.1 (1.96)
	26.3 52.2 67.1 (2.24)	39.2 80.2 268.1 (1.94)	3.7 43.7 219.0 (1.91)
	6.0 40.9 216.5 (2.21)	56.5 45.7 223.6 (1.87)	38.4 76.7 269.8 (1.84)
	14.0 54.0 53.5 (2.21)	21.0 59.5 50.9 (1.66)	45.4 57.6 31.5 (1.83)
	29.6 86.1 93.2 (1.89)	48.5 48.8 68.2 (1.64)	30.2 56.2 36.1 (1.80)
(E-values)	(Rms 37.5)	(Rms 31.5)	(Rms 28.6)
	45.9 48.4 53.7 (2.39)	33.5 46.5 64.0 (2.18)	31.9 48.1 63.1 (1.97)
	4.7 40.8 217.4 (2.29)	38.8 79.5 268.8 (1.94)	4.0 43.4 218.6 (1.95)
	26.1 52.0 66.6 (2.27)	56.5 45.7 224.2 (1.93)	38.5 76.7 269.8 (1.84)
	14.0 53.8 53.5 (2.22)	16.0 51.0 54.0 (1.64)	46.0 57.6 31.1 (1.81)
	29.6 81.7 93.6 (1.91)	8.2 46.4 61.4 (1.62)	26.9 58.3 30.7 (1.79)

Table 3.5.1.3 Cross rotation peaks from ALMN calculated from search models based on the *R. molischianum* structure.

10.0 - 4.0 Å	U_all (Rms 31.8)	U_con (Rms 31.3)	U_poly (Rms 33.8)
	34.0 57.4 64.8 (3.05)	28.8 55.2 68.3 (2.52)	16.8 66.3 31.2 (2.62)
	24.0 54.8 64.8 (2.92)	23.0 40.6 208.7 (2.40)	23.1 55.8 70.8 (2.52)
	8.0 55.3 62.8 (2.85)	35.4 41.4 211.0 (2.27)	33.8 41.2 211.5 (2.49)
	14.2 55.1 56.2 (2.76)	32.7 64.8 274.7 (2.21)	23.7 64.0 41.4 (2.39)
	2.5 47.2 60.1 (2.67)	56.0 42.0 208.0 (2.18)	16.5 81.1 7.1 (2.25)
(E-values)	(Rms 32.0)	(Rms 31.8)	(Rms 33.8)
	33.7 57.4 64.9 (3.06)	26.5 54.8 68.0 (2.55)	16.4 65.9 30.5 (2.62)
	26.5 54.8 64.8 (2.98)	22.5 40.6 209.0 (2.42)	23.6 55.7 70.8 (2.59)
	7.4 54.8 63.0 (2.89)	34.0 41.5 211.0 (2.27)	32.8 41.1 211.9 (2.51)
	3.3 46.2 59.4 (2.78)	32.3 64.0 274.8 (2.21)	23.5 64.1 41.9 (2.36)
	14.2 65.6 56.2 (2.74)	56.5 41.8 207.4 (2.18)	16.8 81.0 6.9 (2.25)

15.0 - 3.5 Å	U_all (Rms 28.7)	U_con (Rms 25.3)	U_poly (Rms 25.6)
	31.5 57.2 66.0 (2.26)	28.9 57.0 66.8 (2.19)	24.5 57.0 68.3 (2.08)
	2.6 60.0 55.3 (1.86)	34.0 41.2 213.5 (2.02)	18.6 66.9 34.3 (2.03)
	32.5 42.9 215.4 (1.84)	20.4 40.5 209.5 (1.98)	41.3 41.5 212.5 (1.98)
	7.7 56.7 61.0 (1.73)	25.0 76.1 12.2 (1.95)	52.7 90.0 198.9 (1.93)
	51.5 45.3 208.5 (1.64)	46.0 41.3 212.1 (1.91)	7.4 90.0 18.9 (1.93)
(E-values)	(Rms 25.2)	(Rms 26.3)	(Rms 25.4)
	31.5 57.2 66.0 (2.61)	29.0 57.0 66.8 (2.15)	26.0 57.1 68.3 (2.12)
	33.5 57.1 66.5 (2.60)	34.0 41.2 213.5 (1.94)	13.9 66.5 31.5 (2.02)
	26.4 68.5 68.6 (2.13)	20.4 40.6 209.7 (1.93)	15.8 66.2 31.5 (2.02)
	32.9 42.3 215.1 (2.09)	25.2 76.0 11.9 (1.88)	41.3 41.5 212.5 (1.98)
	3.1 59.7 55.1 (2.09)	33.1 63.9 274.0 (1.85)	14.3 80.3 15.8 (1.90)

Two major points stand out in the results. The first is the small size of the rotation peaks; no solution is more than approximately 3σ above the *rms* values of the Pattersons. Previous molecular replacement studies have indicated that the "true" rotation peak should be at least 5σ above the *rms* (Dodson, 1985). As the rotation function is a combination of two Pattersons, a larger peak (superposition of self-vectors) should indicate a "better"

solution. In the present case it can also be seen that the first solution peak from each Patterson does not stand out from the rest of the solutions. This would also indicate that the highest peak may not necessarily be the correct answer.

The second point to note is that all of the search models derived from a given structure give a similar answer but that this differs depending on the original structure used. This must result from the small differences that exist between the *R. rubrum* and *R. molischianum* structures; although their general fold is the same there are differences in the loops and slight differences in how the helices are orientated with respect to one another.

The results using normalised structure factors give basically the same results indicating that the data has few reflections with large intensities.

Other features which may influence the results of the rotation function could be the spacegroup symmetry and internal symmetry in the molecule. Highly symmetrical spacegroups can cause problems in the rotation function; if there is a non-crystallographic axis approximately aligned with a crystallographic axis, any maximum of the search function will tend to smear into its symmetry equivalent (Dodson, 1985). Furthermore the overall 3D structure of cytochromes c', comprising four helices arranged in a four-α helix bundle, is reasonably symmetrical. This, combined with the high symmetry space group and the low sequence identity between the cytochromes c' themselves means that this is not the ideal system for molecular replacement.

Translation function (TFFC)

After the rotation results from ALMN were searched for recurring peaks it was decided to test two sets of peaks that were given by the models derived from both *R. rubrum* and *R. molischianum*. The translation program used was TFFC (CCP4 suite, see appendix I) which is a combination of two previous translation functions which were available in the CCP4 suite called TFSGEN and TFPART. This program uses a T2 type of function to calculate Pattersons (see section 3.2). The model that was translated in each case was the conserved structure of the appropriate model, rotated accordingly. After rotation, the iron in each case was placed at the origin; the translations should then agree with the actual iron positions as determined from the anomalous difference Patterson (see section 4.5.4) and would hence provide a check on the correctness of the rotation solution. Translation functions were calculated in both spacegroups, $P6_122$ and $P6_522$, so that a total of eight translation functions were computed. Table 3.5.1.4 shows the four sets of rotation angles tested in the translation function. These rotation angles are representative of the top two peaks found in the majority of the rotation solutions.

Table 3.5.1.4 Rotation angles used in Translation Functions.

Model	θ_1	θ_2	θ_3
J_con	50.0 °	45.0 °	50.0 °
J_con	30.0 °	50.0 °	60.0 °
U_con	30.0 °	57.0 °	65.0 °
U_con	33.0 °	40.0 °	212.0 °

As can be seen in Table 3.5.1.5 and Table 3.5.1.6 in conjunction with the iron positions given in Table 4.5.4.1, none of the rotation angles derived from the *R. rubrum* cytochrome c'gave a translation that corresponded to a shift of the iron to a calculated site. The same was true for both spacegroups tried and for both conserved models used. Also noticeable are the low peak/*rms* values which were also a good indication that the solutions were not correct.

Tables 3.5.1.5 and 3.5.1.6 Translation function results using TFFC with the rotation angles from ALMN, using models derived from *R. rubrum* cytochrome c'.

Rotation Angle	peak/*rms*	Fractional coords			Orthogonal coords		
50.0° 45.0° 50.0°							
($P6_122$)	5.15	0.3477	0.7886	0.0088	-2.53	37.16	1.52
	4.96	0.1117	0.0709	0.3931	4.15	3.34	71.19
	4.76	0.6902	0.3897	0.3929	26.95	18.36	71.16
	4.71	0.7014	0.6452	0.4677	20.60	30.40	84.71
	4.70	0.4121	0.8160	0.2225	0.22	38.44	40.30
($P6_522$)	5.05	0.1618	0.7288	0.1272	-11.02	34.34	23.04
	4.85	0.1637	0.7284	0.2306	-10.91	34.32	41.76
	4.85	0.7372	0.7537	0.2525	19.60	35.51	45.72
	4.78	0.3980	0.2108	0.3074	15.91	9.93	55.67
	4.71	0.9559	0.3105	0.1695	43.61	14.63	30.69

Rotation Angle	peak/*rms*	Fractional coords			Orthogonal coords		
30.0° 50.0° 60.0°							
($P6_122$)	5.70	0.6055	0.0275	0.0298	32.19	1.30	5.40
	4.92	0.2055	0.1814	0.3927	6.24	8.55	71.12
	4.88	0.1553	0.6920	0.0611	-10.37	32.60	11.06
	4.79	0.0999	0.3471	0.0977	-4.01	16.35	17.69
	4.74	0.9746	0.2241	0.4341	46.92	10.56	78.61
($P6_522$)	5.83	0.2186	0.1286	0.0000	8.39	6.06	0.00
	5.83	0.2186	0.1286	0.5000	8.39	6.06	90.55
	4.90	0.7721	0.5274	0.3412	27.65	24.85	61.78
	4.70	0.0888	0.0197	0.4040	4.29	0.93	73.18
	4.65	0.1475	0.1630	0.1312	3.59	7.68	23.75

Similar results were obtained using the rotation angles derived from the *R. molischianum* cytochrome c' models. Again low peak/*rms* values were found and no shifts corresponded to an expected iron position. Tables 3.5.1.7 and 3.5.1.8 give the translation function results using the rotation function solutions derived from the *R. molischianum* cytochrome c'.

Tables 3.5.1.7 and 3.5.1.8 Translation function results using TFFC with the rotation angles from ALMN using models derived from the *R. molischianum* cytochrome c'.

Rotation Angle	peak/*rms*	Fractional coords			Orthogonal coords		
30.0° 57.0° 65.0°							
($P6_122$)	4.90	0.0335	0.8736	0.1152	-21.94	41.16	20.88
	4.66	0.0081	0.6106	0.4816	-16.17	28.77	87.22
	4.59	0.0356	0.8744	0.3375	-21.85	41.19	61.12
	4.58	0.6565	0.0807	0.1505	33.52	3.80	27.31
	4.54	0.0373	0.9269	0.2607	-23.18	43.67	47.21
($P6_522$)	6.34	0.7000	0.1553	0.3786	33.87	7.32	68.56
	5.39	0.7001	0.1547	0.4430	33.92	7.29	80.23
	5.04	0.7001	0.1567	0.3160	33.82	7.38	57.22
	5.03	0.8455	0.3914	0.1219	35.34	18.44	22.07
	4.85	0.1185	0.8989	0.3170	-18.00	42.35	57.40

Rotation Angle	peak/*rms*	Fractional coords			Orthogonal coords		
33.0° 40.0° 212.0°							
($P6_122$)	5.37	0.9822	0.0563	0.0369	51.90	2.65	6.69
	5.20	0.0445	0.1556	0.3209	-1.81	7.33	58.12
	5.05	0.9529	0.0360	0.0876	50.86	1.70	15.86
	5.02	0.9909	0.0642	0.3080	52.15	3.04	55.78
	4.90	0.9815	0.0576	0.3202	52.26	2.85	58.12
($P6_522$)	5.33	0.3074	0.0175	0.4967	16.25	2.65	6.69
	5.04	0.1473	0.2806	0.1830	0.38	13.22	33.14
	4.90	0.0771	0.9842	0.2468	-25.84	46.37	44.69
	4.79	0.0162	0.9843	0.1399	-25.89	46.37	25.34
	4.77	0.0171	0.9856	0.2383	-25.88	46.43	43.15

3.5.2 Results using X-PLOR.

Rotation Function

Two models were used, corresponding to the two known cytochrome c' structures (from *R. rubrum* and *R. molischianum*). These structures were used without any modification of the sidechains. The temperature factors were changed to a default value of 25 $Å^2$ during conversion to X-PLOR format. Both models were initially placed in the same orientation as the R. *molischianum* cytochrome c' and the iron in each was set at the origin. Each model was placed in a box of 100 Å x 100 Å x 100 Å with P1 symmetry, the integration radius was between 3 Å and 30 Å and the data resolution range used was 20.0 - 3.0 Å. The spacegroup symmetry was treated as in Rao *et al.* (1980) by calculating the rotation over a restricted range depending on the spacegroup, although the program ended up calculating angles that lay outside the values of the asymmetric unit. The results from the two search models used are given in Tables 3.5.2.1. and 3.5.2.2.

Table 3.5.2.1 Highest rotation peaks using X-PLOR of the *R. rubrum* cytochrome c' model.

θ_1	θ_2	θ_3	RF-function ($\varepsilon = 0.25$)
40.0 °	50.0 °	30.0 °	2.418
50.0 °	52.5 °	17.5 °	2.403
207.5 °	50.0 °	5.0 °	2.306
210.0 °	50.0 °	52.5 °	2.299
40.0 °	47.5 °	17.5 °	2.280

Table 3.5.2.2 Highest rotation peaks using X-PLOR of the *R. molischianum* cytochrome c' model.

θ_1	θ_2	θ_3	RF-function ($\varepsilon = 0.25$)
232.5 °	65.0 °	0.0 °	2.797
240.0 °	62.5 °	12.5 °	2.737
60.0 °	52.5 °	50.0 °	2.686
237.5 °	67.5 °	42.5 °	2.685
230.0 °	60.0 °	12.5 °	2.668

There was no correspondence between the top solutions for the two models. The top solutions of the model from *R. molischianum* cytochrome c' showed a greater overlap between the model and observed Pattersons as judged from the RF values. By default the top 200 rotation solutions were input to the routine PCREFINE (X-PLOR suite, see appendix I). For each orientation the model was split into five groups (four helices and the haem group) and energy minimisation carried out. The resolution range used was the same as in the rotation function and E-values were used. The results are tabulated in Tables 3.5.2.3 and 3.5.2.4 for each corresponding model.

Table 3.5.2.3 Highest rotation peaks after PC-refinement from the model derived from *R. rubrum* cytochrome c'.

Before	After	Correlation coefficient
197.5 ° 55.0 ° 0.0 °	198.6 ° 55.9 ° 2.0 °	6.55 x10^{-2}
217.5 ° 47.5 ° 5.0 °	201.3 ° 48.1 ° 8.1 °	6.38 x10^{-2}
52.5 ° 52.5 ° 27.5 °	47.1 ° 53.6 ° 25.7 °	6.28 x10^{-2}
50.0 ° 55.0 ° 7.5 °	52.1 ° 54.9 ° 16.6 °	6.25 x10^{-2}
223.5 ° 55.0 ° 52.5 °	224.8 ° 55.7 ° 49.5 °	6.24 x10^{-2}

Table 3.5.2.4 Highest rotation peaks after PC-refinement from the model derived from the *R. molischianum* cytochrome c'.

Before	After	Correlation coefficient
200.0 ° 60.0 ° 7.5 °	196.8 ° 58.5 ° 4.4 °	6.36 x10^{-2}
242.5 ° 60.0 ° 35.0 °	241.2 ° 63.6 ° 37.3 °	6.32 x10^{-2}
252.5 ° 87.5 ° 35.0 °	253.3 ° 86.7 ° 32.2 °	6.29 x10^{-2}
237.5 ° 67.5 ° 42.5 °	234.9 ° 66.5 ° 38.7 °	6.28 x10^{-2}
77.5 ° 72.5 ° 42.5 °	79.1 ° 74.7 ° 43.8 °	6.25 x10^{-2}

For a correct rotation solution the value of the correlation coefficient for that solution should stand out from the rest. For these solutions however, the top correlation coefficient was found to be only just above the rest of the values. This indicated again that the solutions from the rotation function were probably incorrect. One other point to note however, is that after PC-refinement the same peak in both models appeared as the highest, whereas before refinement this peak did not appear as one of the top five solutions given in Tables 3.5.2.1 and 3.5.2.2.

Translation function

Translation functions were carried out using the routine TRANS (X-PLOR suite, see appendix I). The models derived from both *R. rubrum* and *R. molischianum* cytochromes c' were rotated according to the angles given by PC-refinement. At this stage it was not known if the spacegroup was $P6_122$ or its enantiomer $P6_522$ so both spacegroup operators were used in the translation function. The resolution was from 10.0 - 3.0 Å, and as in the PC-refinement of the rotation solutions the molecule was split up into five groups (four helices and the haem group). The top rotation peak from PC-refinement of both models was used. Results are given in Tables 3.5.2.5 and 3.5.2.6.

Table 3.5.2.5 Translation function shifts for model derived from the *R. rubrum* cytochrome c' (rotation angle 198.6 ° 55.9 ° 2.0 °).

Spacegroup $P6_122$
Rms of T-function = 0.022

Orthogonal (Å)	Fractional	Height
-21.20 41.52 83.43	0.051 0.551 0.461	T=0.1670
-5.07 19.96 77.32	0.119 0.424 0.427	T=0.1653
-3.69 19.16 8.14	0.136 0.407 0.045	T=0.1633
13.83 11.18 25.44	0.373 0.217 0.140	T=0.1620
23.51 16.77 25.44	0.610 0.356 0.140	T=0.1620

Spacegroup $P6_522$
Rms of T-function = 0.022

Orthogonal (Å)	Fractional	Height
3.22 27.95 47.81	0.356 0.593 0.264	T=0.1935
43.34 9.58 69.18	0.890 0.203 0.382	T=0.1805
0.92 33.54 88.52	0.373 0.712 0.489	T=0.1802
-12.45 35.93 65.12	0.153 0.763 0.360	T=0.1782
14.75 9.58 23.40	0.373 0.203 0.129	T=0.1753

Table 3.5.2.6 Translation function shifts for model derived from the *R. molischianum* cytochrome c' (rotation angle 196.8 ° 58.5 ° 4.4 °).

Spacegroup $P6_122$
Rms of T-function = 0.021

Orthogonal (Å)			Fractional			Height
17.72	20.76	22.39	0.542	0.441	0.124	T=0.1793
-5.07	8.78	84.45	0.000	0.186	0.466	T=0.1765
49.33	8.78	84.45	1.000	0.186	0.466	T=0.1765
17.52	20.76	7.16	0.542	0.441	0.051	T=0.1741
-5.07	8.78	65.12	0.000	0.186	0.360	T=0.1739

Spacegroup $P6_522$
Rms of T-function = 0.021

Orthogonal (Å)			Fractional			Height
-12.91	41.52	78.34	0.203	0.881	0.433	T=0.1689
14.29	40.72	87.50	0.695	0.864	0.483	T=0.1682
-13.83	43.12	78.34	0.202	0.915	0.433	T=0.1668
13.83	9.58	78.34	0.316	0.203	0.433	T=0.1644
-14.92	39.13	76.34	0.153	0.831	0.433	T=0.1633

As can be seen from the translation results none of the top five peaks from either model (in either spacegroup) gives a correct solution, as judged by correlation with the iron atom position. The iron atom positions are given in Table 4.5.4.1.

3.5.3 Results from AMORE.

Four models were used as search model inputs to AMORE (descriptions of the various search models used are given in Table 3.4.1.2). Two data sets were used, the R-Axis native data from *Alcaligenes denitrificans* cytochrome c' and the Xuong-Hamlin data from *Alcaligenes* sp cytochrome c'. Two resolution ranges were employed, one from 10.0 - 4.0 Å and the other from 15.0 - 3.5 Å. Consequently a total of 16 experiments were undertaken. All of the top rotation solutions found were then tested in the translation function in both enantiomeric spacegroups $P6_122$ and $P6_522$. It was expected that the correct solution would appear in most, if not all of the experiments and would be

identified by its consistency, even if it was not necessarily the top solution. This was based on the expectation that both species would give the same solution (the crystals being isomorphous) and that the four different models should be capable of giving a correct solution. AMORE was chosen for this experiment because of its ability to test a large number of rotation solutions in an economical fashion. Again, as in previous molecular replacement attempts, no clear solution was obtained. Several solutions appeared more than once, but none was found that was a strong solution for each model, and each data set, and over both resolution ranges. Correlation coefficients varied in the range 0.15 to 0.25, and crystallographic R-factors were all higher than 0.56.

Two features of the cytochrome c' structure were then used to screen the solutions to find the most probable one. These were the positions of the iron atoms in the cell and the likely position of the dimer two-fold axis. The iron atom positions were determined from a Patterson map calculated using the observed anomalous differences as coefficients (see section 4.5.4) and are given in Table 4.5.4.1. These positions give an Fe-Fe distance, across the nearest crystallographic two-fold axis, of 25.6 Å which is consistent with the distance found in the cytochrome c' dimers of *R. rubrum* (Yasui *et al.*, 1992) and *R. molischianum* (Finzel *et al.*, 1985). The second criterion, the position of the monomer-monomer two-fold axis, was weaker since although both known cytochrome c' structures are dimeric and use the same helix (helix A) in dimerisation, the mode of association differs in detail, and their monomer-monomer two-fold axes differ in angular disposition by ~ 12.0 ° when the respective monomers are superimposed (Yasui *et al.*, 1992).

The molecular replacement solutions that placed the iron atom of the model to within 6 Å of a calculated position are listed in Table 3.5.3.1. The main conclusions that can be drawn are that most of the best solutions are found in spacegroup $P6_522$ and that the "conserved" model gave the largest number of acceptable solutions, with lower R-factors and higher correlation coefficients. The number of acceptable solutions was further reduced when the second criterion was applied, ie. that the monomer-monomer two-fold axis should lie close to a crystallographic two-fold axis. To test this, each model had two dummy atoms included, defining the two-fold axis found in the *R. molischianum* or *R. rubrum* structure, and solutions were rejected if the line joining these atoms was not placed within 15 ° of a crystallographic two-fold axis. Solution 7 had a well-placed iron atom and a high correlation coefficient but was eliminated by this criterion, as were others many of which placed the monomer-monomer two-fold axis near the crystallographic six-fold. Two solutions remained that fitted both criteria, numbers 3 and 11; these solutions placed the iron 2.4 Å from a calculated position.

Table 3.5.3.1 Potential molecular replacement solutions from AMORE after screening on Fe positions.

Solution[a]	Rotation angles[b]			Translations[c]					
	α	β	γ	trx	try	trz	cc[d]	R[e]	ΔFe[f]
(a) Space-group $P6_122$; polyalanine model.									
1. ADEJ41	27.9	90.0	135.0	0.464	0.312	0.396	16.1	59.0	5.14
(b) Space group $P6_122$; truncated/conserved model.									
2. ASPC351	12.7	77.0	52.3	0.586	0.122	0.095	37.4	55.0	5.09
(c) Space group $P6_522$; polyalanine model.									
3. ADEC355	29.0	55.1	245.0	0.489	0.580	0.227	20.5	62.0	2.41
4. ADEC355	32.0	50.4	249.0	0.489	0.573	0.228	18.9	62.0	2.76
5. ASPJ355	40.1	70.5	238.3	0.388	0.996	0.112	18.1	59.0	5.32
6. ASPJ45	42.6	43.2	247.6	0.484	0.683	0.236	20.1	59.0	5.85
(d) Space group $P6_522$; truncated/conserved model.									
7. ADEC355	28.6	78.8	245.8	0.635	0.453	0.469	38.8	56.0	0.76
8. ADEJ45	37.2	85.8	165.8	0.492	0.840	0.140	30.3	55.0	1.01
9. ASPJ45	38.2	85.8	167.2	0.498	0.841	0.140	29.7	55.0	1.01
10. ASPJ355	2.2	89.9	169.2	0.784	0.329	0.302	30.7	55.0	2.36
11. ASPJ355	41.3	56.1	243.4	0.454	0.571	0.214	40.5	52.0	2.43
12. ADEJ45	54.1	71.2	9.0	0.770	0.571	0.423	27.4	56.0	3.45
13. ADEC45	39.4	39.7	256.1	0.835	0.454	0.064	29.5	55.0	3.82
14. ASPC45	40.3	39.5	254.2	0.833	0.450	0.064	30.1	55.0	3.84
15. ASPJ355	25.5	90.4	348.0	0.787	0.424	0.042	33.1	55.0	4.16
16. ASPJ355	36.6	92.0	244.4	0.612	0.842	0.130	30.7	55.0	5.04
17. ASPJ45	34.4	89.8	247.7	0.620	0.837	0.130	25.0	56.0	5.45
18. ASPJ355	10.0	77.4	52.8	0.258	0.385	0.420	29.5	56.0	5.51

a - The code accompanying each solution indicates the MR experiment it derives from: ADE = *Alcaligenes denitrificans* data, ASP = *Alcaligenes* sp data; J = *R. rubrum* model, C = *R. molischianum* model; 35 = 15.0 - 3.5 Å data, 4 = 10.0 - 4.0 Å data; 1 = $P6_122$, 5 = $P6_522$.

b - Eulerian angles (degrees)

c - Fractional coordinates

d - Correlation coefficient (%)

e - Crystallographic R factor $R = \Sigma||F_o|-|F_c||/\Sigma|F_o|$ for the given resolution range (%)

f- Difference (Å) between iron position and that derived from anomalous difference Patterson.

At this stage there were two possible solutions (3 and 11) which fitted the two criteria used to determine correct molecular replacement solutions. The rotation and translation parameters from each of these two solutions were then applied to the *R. molischianum* model (being the higher resolution model of the two available). The two resulting models were subjected to rigid body refinement (using X-PLOR) before testing to see which solution was correct. In order to test the correctness of these solutions, phase information from the iron anomalous scattering was used. This is described in section 4.5.5. Initial phases were calculated then improved by solvent flattening (see section 5.1.2).

3.6 Retrospective comparison of molecular replacement results.

A retrospective look at the molecular replacement problem was undertaken after the refinement of the *Alcaligenes* sp cytochrome c' structure had been completed. Using the final *Alcaligenes* sp structure as search model (with all sidechain atoms included) gives a clear rotation solution (peak of 4.1σ, with the next highest peak 2.4σ). The correct rotation solution is also given by a truncated/conserved model (3.5σ; next highest 2.7σ) and a polyalanine model (3.2σ; next highest 2.8σ), and all three models give the correct translation, with peak heights 7.9σ, 6.8σ and 5.3σ respectively (next highest 3.6σ, 2.9σ and 2.3σ). Table 3.6.1.1 gives the molecular replacement solutions (ALMN and TFFC) using the final refined structure of the cytochrome c' from *Alcaligenes* sp.

Table 3.6.1 Molecular replacement solutions using the final refined structure of the cytochrome c' from *Alcaligenes* sp.

Model	Resolution (10-4 Å)	Resolution (15-3.5 Å)	Peak height (σ)
(i)Rotation (ALMN)			
cp_full	31.5° 51.3° 68.7°	29.8° 51.3° 69.4°	4.11
cp_con	33.5° 51.8° 68.5°	30.0° 51.3° 69.2°	3.48
cp_poly	31.7° 50.7° 67.8°	29.3° 50.8° 68.9°	3.22
(ii)Translation (TFFC)	Translation	shift	
cp_full	33.88 7.34	69.02	7.89
cp_con	33.85 7.36	69.01	6.81
cp_poly	33.87 7.37	69.08	5.33

These results can be compared with Table 3.5.1.4 which gives the best rotation solutions derived from the cytochromes c' from *R. rubrum* and *R. molischianum*. Although two of the solutions were approximately correct, it is generally accepted that the translation function will only be successful if the rotation solution is within 5 ° of its correct value (Dodson, 1985). The inaccuracy of the solutions must have been sufficient to prevent a recognisably correct translation function (ie. one which was comparable with the known iron positions) from being obtained. The translation function with the molecule orientated correctly, shifts the model to one of the calculated iron positions.

The same retrospective tests were also carried out using the model in the orientation used with AMORE. Table 3.6.2 gives the comparison when the refined structure of the cytochrome c' from *Alcaligenes* sp is placed in the *R. molischianum* cytochrome c' orientation. It is this orientation that the search models had in the original molecular replacement experiments with AMORE (section 3.5.3).

Table 3.6.2 Testing AMORE molecular replacement solutions.

Model	Resolution (10-4 Å)	Resolution (15-3.5 Å)	Peak height (σ)
Rotation			
cp_full	42.4° 52.6° 243.1°	40.3° 53.3° 243.9°	4.10
Translation	Translation	shift	
cp_full	10.64 33.41	38.78	7.19

The next highest peak in the rotation solutions was 3.45σ and in the translation solutions 3.90σ. This retrospective test confirmed solution 11 was the correct molecular replacement solution.

Chapter Four

Phase determination by isomorphous replacement and anomalous scattering.

4.1 Introduction.

The fundamental difficulty of X-ray crystallography is that calculation of the electron density requires the knowledge of the structure amplitudes, which are directly observable, and the structure factor phase angles, which are not. One of the fundamental techniques for solving the crystallographic phase problem in protein crystallography is that of isomorphous replacement. The first protein structure determined by the use of isomorphous replacement was horse haemoglobin in 1954 (Green *et al.*, 1954) and the method remains one of the few available for the determination of novel protein structures (Rossman, 1960). In protein crystallography, isomorphous replacement could be renamed isomorphous addition, since it is usually the binding of an heavy atom in a position formerly occupied by disordered solvent. If atoms of a high atomic number are added to a known position into the macromolecular structure, the corresponding changes in scattering can be used to obtain partial or complete information about the phase angles (Green *et al.*, 1954).

The method depends on the high solvent content of protein crystals which allows a heavy atom or associated complex to be diffused into the crystal through channels of disordered or partially ordered solvent. The term "isomorphous" indicates that the derivative crystal upon heavy atom uptake should not have had any alteration in either the protein structure or the crystal lattice. In practice, this is seldom completely the case, as the introduction of a compound which interacts with some of the atoms on the surface of the protein usually gives rise to small changes in the cell dimensions and the protein structure. There are two ways of preparing heavy atom derivatives of proteins. The first involves reacting the protein with heavy atom reagents and then crystallising the complex. The second, more common method, involves soaking the crystal in a mother liquor containing the heavy atom compound. The second method is more often used as this avoids having to find new crystallisation conditions for the derivative species, whose solubility may have changed.

4.2 Structure factor and phase relationships.

If $\mathbf{F_P}$ is the structure factor of the native protein, F_P is its magnitude, α_p its phase and, $\mathbf{F_{PH}}$, F_{PH}, and α_{PH} are the equivalent quantities of the heavy atom derivative, the vector relationship between $\mathbf{F_{PH}}$ and $\mathbf{F_P}$ can be written as:

$$\mathbf{F_{PH}} = \mathbf{F_P} + \mathbf{F_H}$$

Here $\mathbf{F_H}$ is a vector representing the heavy atom contribution. This relationship is also illustrated in vector form in Figure 4.2.1.

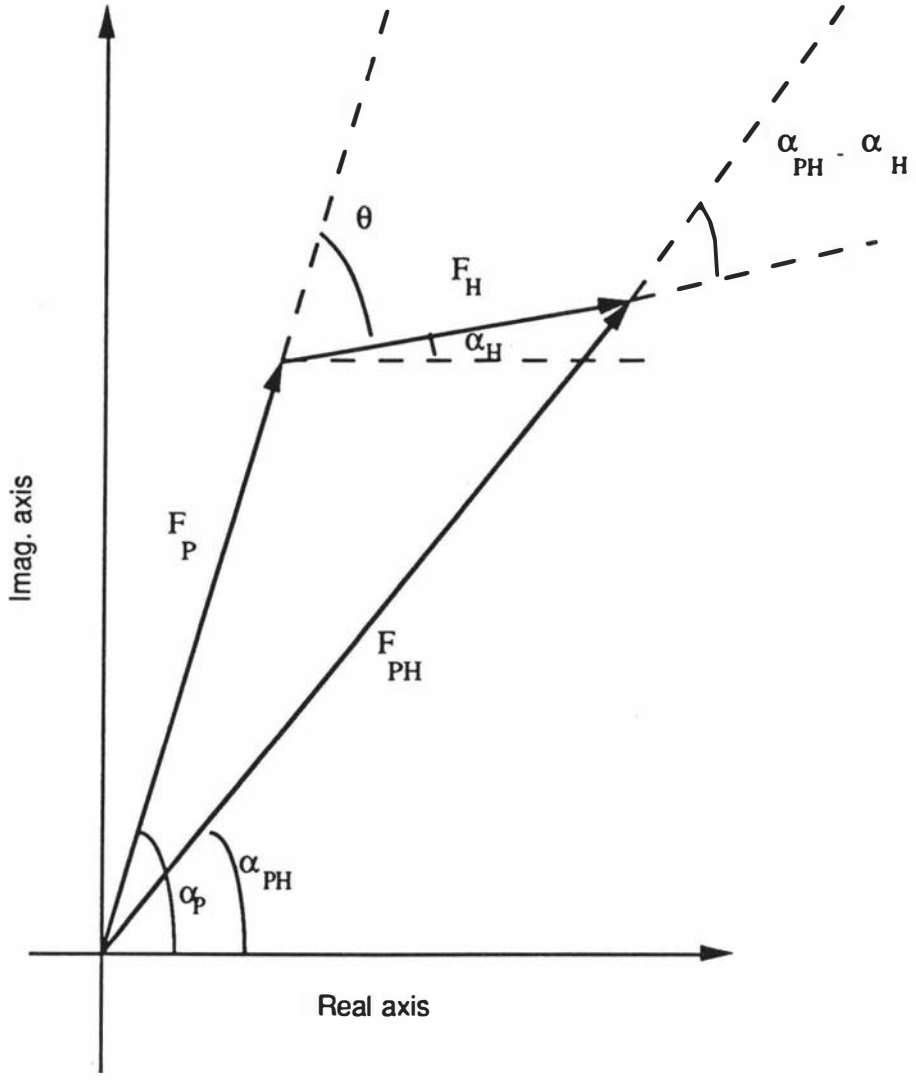


Figure 4.2.1. A vector diagram (Argand diagram) defining the structure factor amplitudes and phases referred to in section 4.2.

4.2.1 Isomorphous replacement.

Single isomorphous replacement (SIR)

If only a single derivative is available, as in this study, then an ambiguity in the phase determination exists. This ambiguity is illustrated in Figure 4.2.1.1. In principle, a three-dimensional electron density map could be calculated, using the two possible phase values for each reflection. This kind of synthesis is called an SIR map and it will contain information about the true structure plus noise (Blow and Rossman, 1961). If α_{PT} is the correct phase value of α_P, and α_{PW} the wrong one, then from Figure 4.2.2.1 we can write:

$$2F_{SIR} = F_P\exp(i\alpha_{PT}) + F_P\exp(i\alpha_{PW})$$

since $\alpha_H = \frac{1}{2}(\alpha_{PT} + \alpha_{PW})$, then:

$$2F_{SIR} = F_P\exp(i\alpha_{PT}) + F_P\exp(2i\alpha_H)\exp(-i\alpha_{PT}) \qquad \text{(Eqn 4.1)}$$

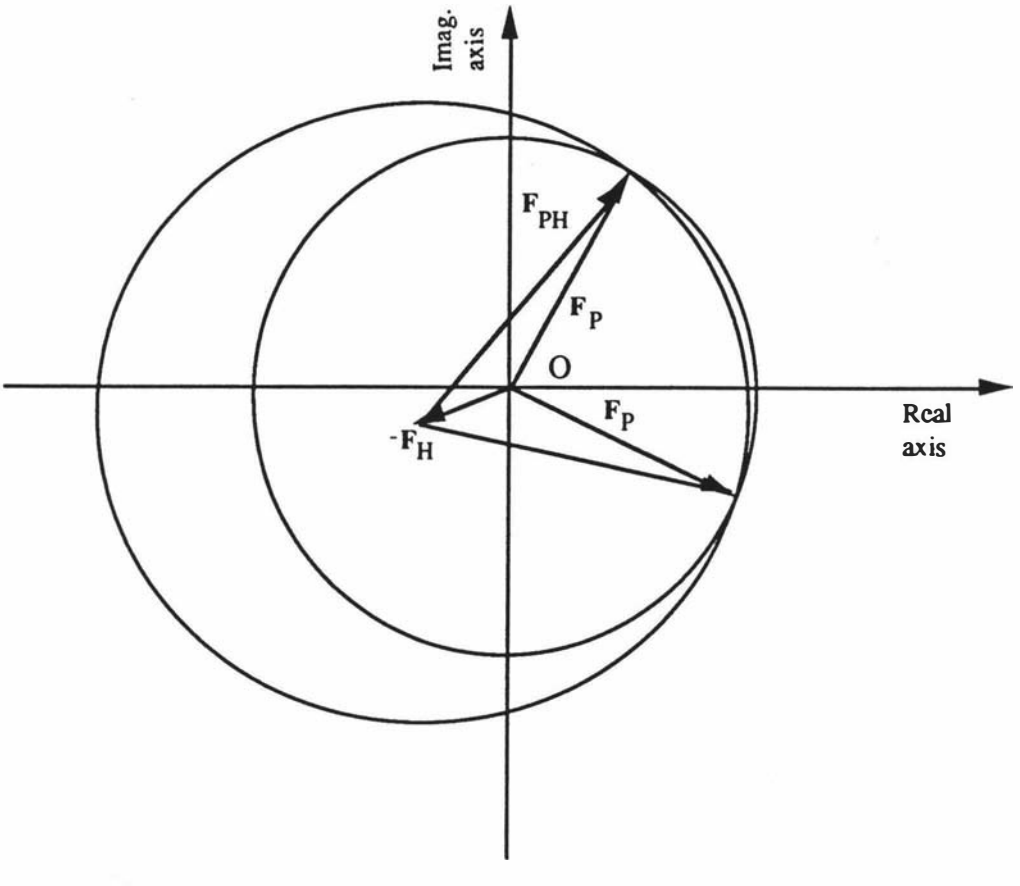


Figure 4.2.1.1 Graphic construction of SIR method, showing the two possible values for the phase angle of the native structure factor F_P. The structure factor of the protein is a vector lying on a circle of radius F_P and centre O. The intersection of the two circles represents the two possible solutions.

From Eqn 4.1 it is evident that F_{SIR} will approach the correct value of F_P if α_H approximates α_{PT}. The first term of Eqn 4.2 will give the correct electron density, while the second one will only contribute to the noise. This is illustrated in Figure 4.2.1.2.

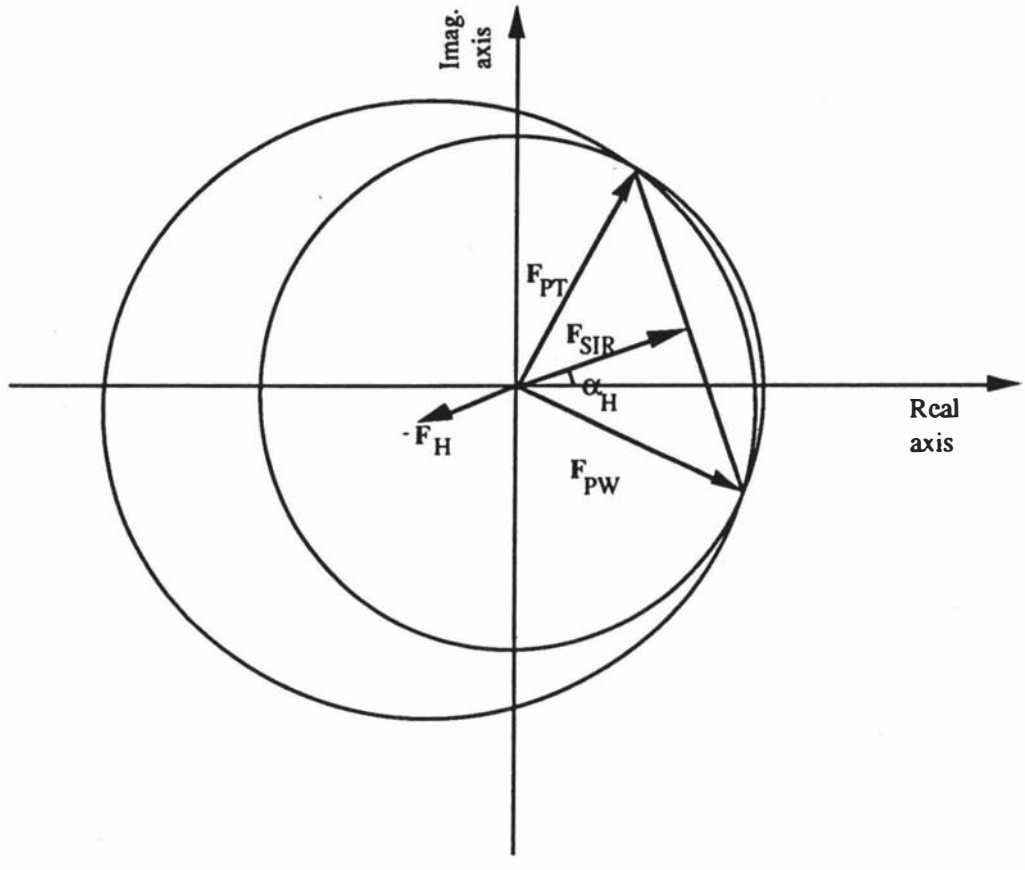


Figure 4.2.1.2 The same diagram as Fig 4.2.1.1, except now that the "true" and "wrong" value for the native structure factors are explicitly indicated as F_{PT} and F_{PW} respectively. The further the two vectors are from each other, the more the SIR phase (α_H) is different from α_P.

Multiple isomorphous replacement (MIR)

From the vector triangle $F_{PH} = F_P + F_H$ as in Figure 4.2.1., knowledge of the amplitudes F_{PH}, F_P, F_H, the phase α_H and the cosine law, α_P can be derived;

$$\alpha_P = \alpha_H + \cos^{-1} ((F^2_{PH} - F_P^2 - F^2_H) / 2F_PF_H) \qquad \text{(Eqn 4.2)}$$

where the cosine function term is derived from the equation $a^2 = b^2 + c^2 - 2bc \cos A$. Where a, b, and c are the amplitudes representing the sides of the triangle. Harker gave a geometrical solution to the problem of phase determination (Harker, 1956) as shown in Figure 4.2.1.3. Harker showed that for two heavy atom derivatives, 1 and 2, the points of intersection of the circles with radii F_{PH1} and F_{PH2} with the circle of radius F_P (native protein) give possible phase angles. In the absence of experimental errors, and assuming complete ismorphism on addition of the heavy atom, then the two derivative circles will intersect at one place, giving a unique solution for α, the phase of the protein.

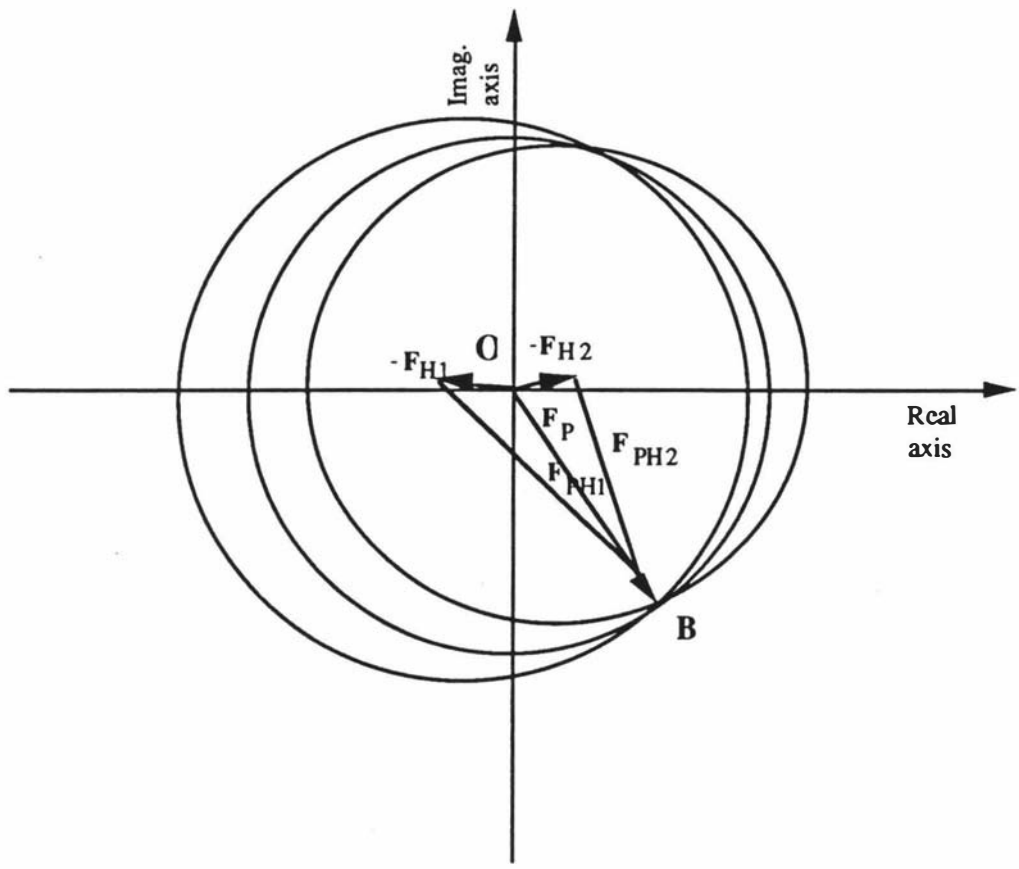


Figure 4.2.1.3 Harker construction illustrating MIR method. The three circles have only one common intersection, representing the solution.

This is based on the assumption that F_P, F_{PH1}, F_{PH2}, F_{H1} , F_{H2}, α_{H1} and α_{H2} are known.

4.2.2 Incorporation of anomalous scattering.

Electrons are bound to the nucleus by forces which depend on atomic field strength and on the quantum state of the electron. Because of this they can be considered

as oscillators with natural frequencies. If the X-ray wavelength approaches an absorption edge of a particular element the electrons in the atom can not be considered free. From radiation physics a free electron in the scattered beam differs exactly 180 ° in phase from an electron in the incident beam. This only applies to the real component $\Delta f'$. However the inner electrons in an electron cloud around an atomic nucleus are more tightly attached to the nucleus than the outermost ones (more so for heavy elements with a high nuclear charge). For these inner shell electrons the diffracted beam does not differ 180 ° in phase angle from the incident beam. This is the imaginary part of the diffracted beam and is 90 ° in front of the scattered wave. This anomalous scattering or dispersion can be represented by the expression;

$$f = f_o + \Delta f' + i\Delta f'' = f' + i\Delta f'' \qquad (\text{Eqn } 4.3)$$

where

f_o = normal scattering factor
$\Delta f'$ = real correction term
$\Delta f''$ = imaginary component

The variation of dispersion effects is evident in Table 4.2.2.1 which gives $\Delta f'$ and $\Delta f''$ for CrKα and CuKα radiation as a function of wavelength.

Table 4.2.2.1 Variation of the atomic scattering factor with wavelength.

	Atomic number	CuKα radiation (λ = 1.542 Å)		CrKα radiation (λ = 2.291 Å)	
		$\Delta f'$	$\Delta f''$	$\Delta f'$	$\Delta f''$
Fe	26	-1.1	3.4	-1.6	0.9
Zn	30	-1.7	0.8	-1.0	1.5

The structure factor may be written as two parts;

$$F(hkl) = F'(hkl) + iF''(hkl)$$

where the real part of anomalous dispersion is included in F '(hkl). When the anomalous scattering contribution is zero, Friedel's law applies and F(hkl) is equal to $F(\bar{h}\,\bar{k}\,\bar{l})$, where

F(hkl) and $F_{(\bar{h}\,\bar{k}\,\bar{l})}$ are referred to as Bijvoet pairs. When the anomalous contribution is non-zero, however, Friedel's law breaks down, ie.

$$F_{(hkl)} \neq F_{(\bar{h}\,\bar{k}\,\bar{l})}$$

If an accurate measurement of the Bijvoet pairs is undertaken, an anomalous difference Patterson synthesis can be calculated, with coefficients;

$$(\Delta F_{ano})^2 = (F^{+}_{PH} - F^{-}_{PH})^2$$

where F^{+}_{PH} represents the amplitude of the structure factor for a reflection (hkl) and F^{-}_{PH} is the amplitude for the reflection $(\bar{h}\ \bar{k}\ \bar{l})$. A map calculated with the above coefficients can be used separately or in combination with the isomorphous replacement method (see section 4.4).

Most heavy atoms have significant anomalous scattering contributions and the phase relationships for isomorphous replacement then become as shown in Figure 4.2.2.1. Anomalous scattering differences are only observed if there are anomalous scatterers amongst non-anomalous scatterers. Figure 4.2.2.1 shows the relationship between the structure factors if anomalous scattering occurs, with the vectors $\mathbf{F''}_H$ and $\mathbf{F}_H$ at right angles to each other. The latter is true if the anomalous scatterers are of the same type. If this is not the case then the analysis of the anomalous differences is more complex. In this Figure the $\mathbf{F}_{PH}$- construction has been reflected across the x-axis as normally $\alpha_{h\,k\,l} = -\alpha_{(\bar{h}\,\bar{k}\,\bar{l})}$.

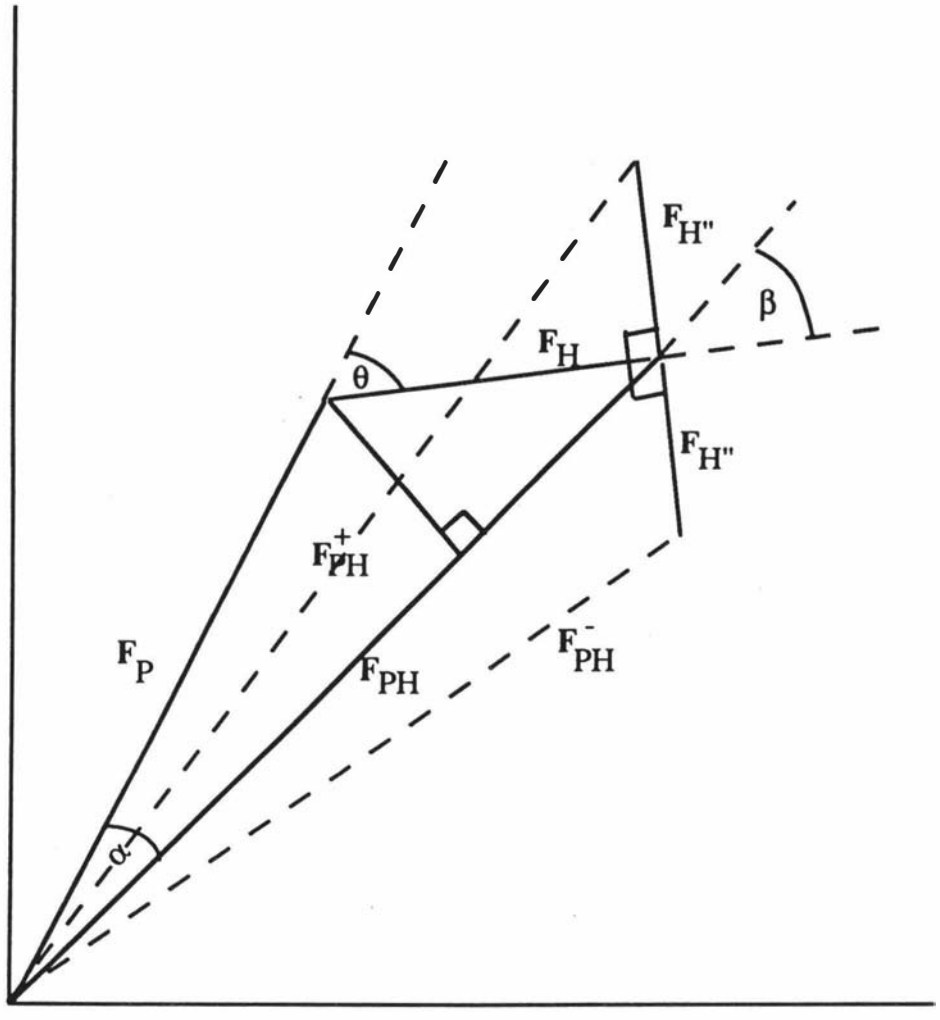


Figure 4.2.2.1 Vector diagram illustrating the relationship between F_P, F_{PH}, and F_H. F_H'' is the anomalous contribution to the heavy atom scattering, while F^+_{PH} and F^-_{PH} are the Friedel related structure factors.

Single isomorphous replacement with anomalous scattering (SIRAS)

One of the most important aspects of anomalous scattering in protein crystallography is the possibility of solving the phase ambiguity if only a single heavy atom derivative is available. This is illustrated in Figure 4.2.2.2.

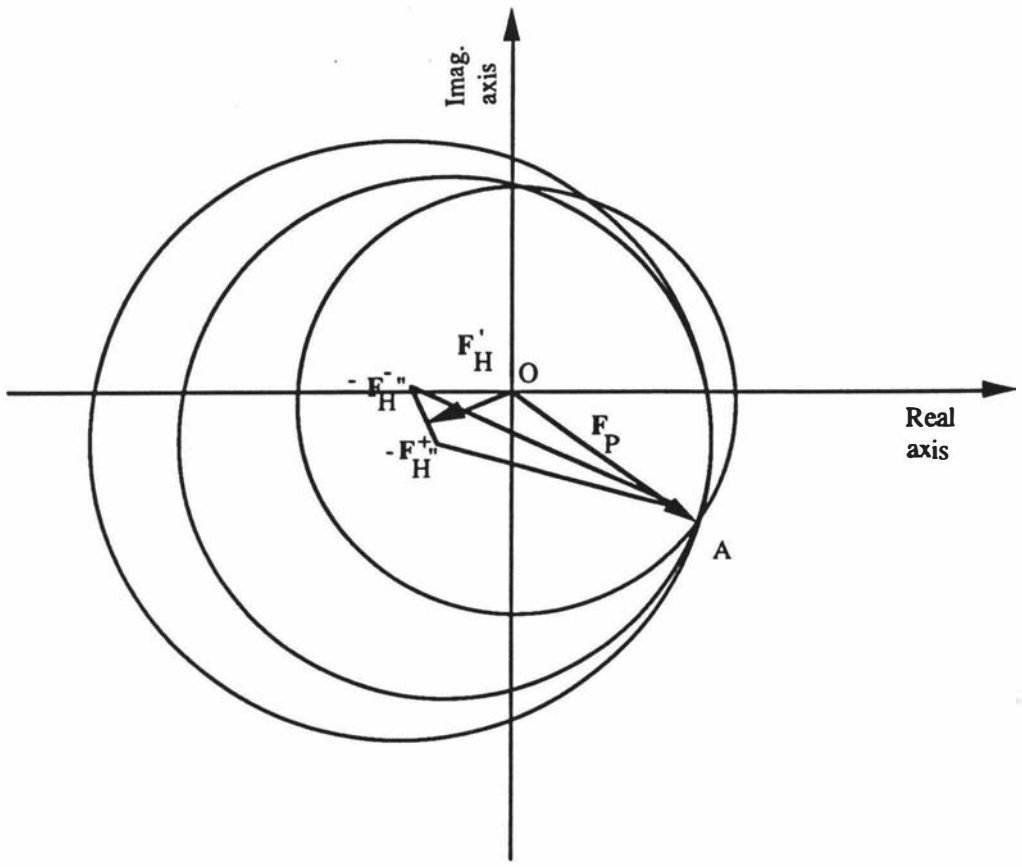


Figure 4.2.2.2 A graphical representation of the solution of the phase problem if only a single derivative is available (and its anomalous component). Two circles of radius F^+_{PH} and F^-_{PH} are drawn at the end of vectors $-\mathbf{F}_H^+$ and $-\mathbf{F}_H^-$ respectively. OA is the native protein structure factor, whose phase angle is uniquely determined by the intersection of the three circles.

Multiple isomorphous replacement with anomalous scattering (MIRAS)

The use of two or more heavy atom derivatives means the phase ambiguity present in SIR is removed. In this case, the use of anomalous scattering is used primarily in augmenting the phase information produced by the isomorphous differences to reduce errors that arise.

4.3 Practical considerations.

There are four main approaches to obtaining a derivative (Blundell and Johnson, 1976);

(i) metal ion replacement in metalloproteins,

(ii) use of heavy atom labelled inhibitors,

(iii) replacement of one amino acid with an heavy atom analogue,

(iv) covalent or coordination reaction of an heavy atom reagent with specific residues of the protein.

Concentration of the heavy atom complex and soaking time are important variables: they can range from 0.001 mol L^{-1} to 0.1 mol L^{-1} and from a few hours to several weeks. These conditions will also depend on the protein being studied, pH, mother liquor, and temperature, as follows :

1. pH. At pH >4 the acidic groups on the protein will be negatively charged, and therefore able to react with cations. At the same time, however, some metal ions are likely to form insoluble hydroxides at pH's in the alkaline range (>7) . At low pH, potentially reactive groups will be protonated and may be prevented from reacting. Basic groups such as lysine (pK_a~ 9) are protonated at neutral pH and lower.

2. Mother liquor. The composition of the mother liquor can strongly affect the heavy atom binding (Sigler and Blow, 1965). High salt concentration is not ideal as solubility is affected. In the case of ammonium sulphate, stable coordination complexes between the metal and ammonia could be formed instead of the heavy atom binding to the protein.

3. Temperature. An obvious effect of temperature is that reactions are slower at 4°C than at room temperature.

If no sites allowing direct reaction with heavy metal complexes are found, it is possible to introduce reactive groups in the protein. This can be attempted either by chemical modification or by protein engineering (Mowbray and Petsko, 1983).

4.4 Determination of heavy atom positions.

Before phases can be calculated from a heavy atom derivative, the sites in the crystal lattice where these atoms have been substituted must be determined. This is done from the observed differences in the structure amplitudes between the derivative and native reflections.

4.4.1 Use of the Patterson function.

Patterson syntheses (Patterson, 1934) are calculated using the formula;

$$P(u,v,w) = \frac{1}{V} \sum_h \sum_k \sum_l F_{H(hkl)}{}^2 \cos 2\pi(hu+kv+lw)$$

where V is the volume of the unit cell, and F_H is the structure amplitude contribution of the heavy atom. For a protein molecule containing N heavy atoms in the unit cell, the Patterson map should contain N^2 peaks arising from N possible vectors which can be drawn between each of the N atoms. Of these, N will be self vectors and concentrated in a large peak at the origin, while the remaining N^2 - N will be distributed throughout the cell. If a derivative has many sites of substitution, interpretation of the Patterson synthesis will be very difficult especially in high symmetry space groups. Difficulties also arise if there are errors in the measured intensities or if the derivative and native data have not been scaled together correctly. If the binding of the heavy atom distorts the position of the atoms in the lattice, this can destroy some of the isomorphism. This means that some of the intensity changes will result from this distortion rather than from the presence of the heavy atom. In this case, not all the peaks in a difference Patterson map are due to heavy atom-heavy atom vectors. As it is often difficult to detect low occupancy binding sites using Patterson methods, other methods such as difference Fouriers, can be used to determine low occupancy sites.
Three types of Patterson functions can be used to determine the heavy atom sites;

(1) $\Delta_{iso}{}^2$ - isomorphous difference Patterson. This has coefficients $(F_{PH} - F_P)^2$. Problems with this type of function only occur if there are several sites of low occupancy or even if there are several fully occupied sites on a large protein (Blundell and Johnston, 1976) when noise may give rise to peaks larger than the heavy atom peaks.

(2) Δ_{ano}^2 - anomalous difference Patterson with the coefficients $(F_{PH}^{+} - F_{PH}^{-})^2$. These are more noisy than isomorphous Pattersons as the differences are usually in order of magnitude of the standard deviation, but will contain the expected heavy atom peaks.

(3) F_H^2 - a combined difference Patterson with information from the isomorphous and anomalous scattering differences. The anomalous contribution must be carefully used due to the errors involved in measurement (Blundell and Johnston, 1976).

4.4.2 Use of difference Fouriers.

If estimates of the protein phases are available from other heavy atom derivatives, then difference Fourier techniques can be used to locate the heavy atom sites (Stryer *et al.*, 1964). The Fourier synthesis has the coefficients;

$$m(F_{PH} - F_P)\exp(i\alpha_P)$$

where m is the figure of merit and α_P is the calculated phase from a different derivative than F_P. Difference Fourier syntheses are most useful when there are many sites of heavy atom substitution and the resulting difference Pattersons difficult to interpret.

4.4.3 Refinement of the heavy atom positions.

A variety of methods have been used to refine heavy atom parameters and hence the phases derived from them. In "Maximum likelihood phase" refinement (Otwinowski, 1991) weights are found for all the possible values of the protein phases α_P from 0 ° to 360 °. The weighted differences between observed and calculated values of F_{PH} generated with these phases are then minimised by adjustment of the heavy atom parameters. The following parameters may be refined:

1) Relative scale and temperature factor between protein and derivative data,
2) Coordinates,
3) Real and anomalous occupancies,
4) Thermal factors (isotropic and anisotropic),
5) error estimates which are used when finding the appropriate weight for each observation.

The function being minimised is;

$$R_j = \sum_{\phi_p} \sum_{h} w(\phi_p) (F^{obs}_{PH_j} - F^{calc}_{PH_j}(\phi_P))^2$$

where j refers to the j^{th} derivative.

The centric data is used to refine the scale factor, relative occupancies and estimated errors. Since all possible phases are sampled for each reflection, the procedure eliminates incorrect sites. For reflections with well defined phases the weight of the protein phases will be very small for all values except the best phase. There are, however, many reflections where the protein phase is poorly defined, and for these reflections the "maximum likelihood" approach avoids the problem of biasing the results towards phases derived from incorrect sites (Otwinowski, 1991). Details of the refinement are given in section 4.5.2.

4.4.4 Treatment of errors.

In practice ideal conditions will not apply and errors will arise. The main two sources of error are;

(1) Error in measuring intensities of reflections will produce errors (σ) in the amplitudes F_P and F_{PH} ;
(2) Errors in locating and describing the scattering of the heavy atoms as well as lack of isomorphism will introduce errors (ε) in F_H.

As a result the two derivative circles (in the case of double isomorphous replacement) will not intersect the native circle at exactly the same point. The total error E, is a measure of the discrepancies between theory and experiment:

$$\langle E \rangle^2 = \langle \sigma \rangle^2 + \langle \varepsilon \rangle^2$$

Here $\langle \sigma \rangle$ can be estimated by comparing the values from different crystals. The estimation of $\langle E \rangle$ can be made for centric reflections from the differences between F_{PH} - F_P and F_H. For general reflections this is not possible because F_{PH} - F_P is not a correct estimate of F_H. The error estimation is refined using "maximum likelihood phase refinement".

The probability distribution of phases tends to be bimodal so a strategy for making the "best" choice is used. The "best" Fourier is calculated with phases corresponding to the centroid of the distribution weighted by the figure of merit, *m*, which is a measure of the reliability of the phase determination. If the probability of a set of phases is sharp and unimodal, *m* will be close to 1, but if the probability is almost uniformally distributed along the circle, *m* will be very small.
This Fourier has coefficients;

$$m\ F_P e^{i\alpha_{best}}$$

where *m* is related to the mean value of the cosine of the error in the phase angle for that reflection:

$$m = \langle \cos\Delta\alpha_i \rangle$$

The "maximum likelihood" approach of assigning each possible phase a likelihood and combining these, means the weights are more realistic (ie lower) and phases more accurate since spurious sites are removed (Otwinowski, 1991).

4.5 Results.

4.5.1 Heavy atom soaking trials.

Some of the principles of heavy atom derivative preparation are outlined in section 4.3. Initial trials involved soaking crystals of cytochrome c' from *Alcaligenes* sp in various heavy atom solutions (see Table 4.5.1.1) for a period of seven days. The initial mother liquor, which consisted of 95 % $(NH_4)_2SO_4$ in 0.1 mol L^{-1} phosphate buffer at pH 8.0, is unsatisfactory for heavy atom substitution due to competition of the ammonium sulphate for the heavy atoms (Petsko, 1985) to form metal-ammine complexes. This was therefore changed to a mother liquor containing 95 % $(NH_4)_2SO_4$ at pH 8.0 with Tris-HCl as the buffer. Although these were not ideal conditions to attempt heavy atom substitution, the crystals were stable in this solution. From the initial trials only the crystals soaked in K_2PtCl_4 gave any indication of substitution. With one exception the rest of the soaks either gave no substitution as judged from the size of intensity differences from the native data. In the case of the $NaAuCl_4$, the crystal was not isomorphous upon heavy atom uptake.

Table 4.5.1.1 Heavy atom complexes used in initial derivative trials.

Reagent	Concentration (mg mL^{-1})	Soaking period	Result
K_2PtCl_4	10.0	7 days	R_{iso}~30 %
	8.0	7 days	R_{iso}~ 28 %
$KAuCl_4$	0.5	7 days	crystal cracked
Ethyl mercury phosphate	10.0	7 days	no change
	1.0	7 days	no change
Samarium acetate	1.0	7 days	no change
Uranyl acetate	1.0	7 days	no change
K_2PtCl_6	saturated solution	7 days	no change
PCMB[a]	1.0	7 days	no change
Thiomersal[b]	1.0	7 days	no change
$Pt(en)_2Cl_2$	saturated solution	7 days	no change
Baker's dimercurial[c]	1.0	7 days	no change
K_2PtBr_4	1.0	7 days	no change
Trimethyl lead acetate	1.0	7 days	no change
PHMB[d]	1.0	7 days	no change

a - *p*-chloromercuribenzoate

b - EMTS (C_2H_5- $HgSC_6H_4COO^-$ Na^+)

c - 1,4-diacetoxymercuri-2,3-dimethoxybutane

d - *p*-hydroxymercuribenzoate

Trials were repeated using K_2PtCl_4 at a concentration of 0.5 mg mL^{-1} (0.001 mol L^{-1}) and 2 mg mL^{-1} (0.004 mol L^{-1}) in order to reduce the amount of substitution, for a period of four days. In both cases the crystal was used for data collection and subsequently for phase determination. Fresh solutions of K_2PtCl_4 less than 24 hours old were used in order to reduce the reaction of the heavy atom ligand with the ammonium sulphate (Petsko *et al.*, 1978). Details of the data collection and processing are given in section 2.3.3.

4.5.2 Heavy atom site determination.

For the K_2PtCl_4 derivative (two independent sets of data from different crystals) isomorphous difference Patterson maps were calculated. The Pattersons were all calculated in spacegroup P6/mmm and interpreted for $P6_122$ or $P6_522$. From these Pattersons it was possible to deduce that there were two sites, one major and one minor. The anomalous differences were also determined and used in the refinement of the heavy atom positions and in the calculation of the phases. The heavy atom parameters were refined with MLPHARE (CCP4 suite, see appendix I) and are given in Table 4.5.2.1 for each space group and each of the K_2PtCl_4 derivative sets collected. The heavy atom positions were refined using all centric data ($\alpha_P = 0$ ° or 180 °), combined together to give a data-set. Initial occupancy factors were set at 1.0 for the major site and 0.5 for the minor sites. All reflections were given an equal weight in the refinements and occupancies and the x, y, and z coordinates were refined first assuming an initial isotropic temperature factor B of 25 $Å^2$ for each site, a common value for heavy atoms that are bound to protein structures (Blundell and Johnson, 1976; Brayer *et al.*, 1978; Matthews *et al.*, 1979; Scouloudi and Baker, 1978).

Table 4.5.2.1 Refined platinum sites from MLPHARE.

Site	Occ. [a]	x/a	y/b	z/c	B $(Å)^2$ [b]
Derivative 1	$P6_122$				
A	2.654	0.759	0.086	0.114	25.0
B	1.348	0.715	0.330	0.227	25.0
	$P6_522$				
A	2.609	0.087	0.328	0.552	25.0
B	1.331	0.715	0.386	0.560	25.0
Derivative 2	$P6_122$				
A	2.530	0.759	0.086	0.114	25.0
B	1.248	0.715	0.330	0.227	25.0
	$P6_522$				
A	2.504	0.087	0.328	0.553	25.0
B	1.238	0.614	0.285	0.606	25.0

a - Occ. is the relative occupancy of the site in arbitrary units.

b - B is the isotropic temperature factor of the site, and is of the form $\exp[-B(\sin^2\theta)/\lambda^2]$ where $B = 8/3\pi^2u^2$ and u^2 is the mean square amplitude of vibration.

It can be seen from the occupancies of the platinum sites that there is one major and one minor site for each K_2PtCl_4 derivative. The major binding site for the platinum complex has been determined to be the $N_{\delta 1}$ atom of His120. This residue also binds the iron in the haem group. The positions of the platinum heavy atoms in the two soaks are slightly different when compared to each other in space group $P6_522$ (which was subsequently determined to be the correct spacegroup). These positions, however, are related by symmetry and in fact are equivalent to each other.

4.5.3 Phase calculations.

Phases for the data were calculated with the program MLPHARE (CCP4 suite, see Appendix I) using the method of Blow and Crick (Blow and Crick, 1959; Dickerson *et al.*, 1961). The probability distribution was calculated at 10 ° intervals around the phase circle. The figure of merit, *m*, was calculated for each reflection. The phasing power of a derivative was expressed in terms of the ratio F_H/E. Table 4.5.3.1 gives the phasing statistics from the K_2PtCl_4 derivatives used in calculating initial phases.

Table 4.5.3.1 Phasing statistics from K_2PtCl_4 heavy atom derivatives.

Derivative	K_2PtCl_4 (1)	K_2PtCl_4 (2)
Concentration (mg mL^{-1})	2.0	0.5
Soaking time	4 days	4 days
Heavy atom sites	A,B	A,B
Phasing power (acentric, centric)[a]	1.2, 0.9	1.3, 0.9
Cullis R (acentric, centric)[b]	0.71, 0.69	0.70, 0.66

Figure of merit **acentric** 0.3903 (5,506 reflections
centric 0.6422 (1,868 reflections)

[a] Phasing power = $(\sum F_H^2 / n)^{0.5} / \sum \| F_{PH} \pm F_P | - F_H |$

[b] Cullis R = $\sum \| F_{PH} | \pm | F_P | - F_H | / \sum \| F_{PH} - F_P \|$

Lack of closure E = $\sum \| F_{PH} \pm F_P | - F_H |$

Rms E = $(\sum \| F_{PH} \pm F_P | - F_H |^2) / n^{0.5}$

Rms F_H = $(\sum F_H^2 / n)^{0.5}$

F_H = calculated heavy atom structure amplitude.
F_P = structure factor amplitude of native crystal.
F_{PH} = structure factor amplitude of derivative crystals.

From Table 4.5.3.1 both derivatives have essentially the same phasing power. The first derivative has a slightly greater occupancy of the heavy atom sites as a result of the higher concentration used (2 mg mL^{-1} compared with 0.5 mg mL^{-1}). As the phasing powers are ~ 1.0, however, the derivatives were not good enough to allow full chain tracing (discussed in section 5.2). The Cullis R-factors for both species are ~ 0.7 which indicates a good degree of isomorphism with the native crystal. The contribution from the anomalous data was also used in phasing, but during the refinement of the heavy atom positions, the anomalous occupancies refined almost to zero, and the Cullis R-factors were around 1.0. This was later traced to an error in the R-Axis software, which scrambled some of the signs of the anomalous differences during processing. The practical effect was that most of the phasing came from the isomorphous differences.

4.5.4 Determination of iron sites.

The cytochrome c' monomer contains a single haem group, and with a monomer molecular weight of approximately 14,000 Da it was decided to see if the iron produced a detectable anomalous signal that could be used to determine the iron sites. A set of native data from *Alcaligenes* sp to a resolution of 2.8 Å and collected with CuKα radiation (data collection and processing is described in section 2.3.2) was processed and the anomalous differences calculated. These were used to calculate an anomalous Patterson which was analysed using the program VECSUM (CCP4 suite, see appendix I). Spacegroup $P6_122$ (or $P6_522$) has a multiplicity of 12, so the Patterson function will contain 11 unique vectors between a heavy atom and its 11 equivalent sites. VECSUM systematically searched the eleven Harker sections for correspondences, and produced a map showing the real space locations of possible heavy atom sites. Two equivalent sets of peaks (7.5σ above rms) related by an origin shift of 0, 0, 1/2 were obtained for both spacegroups $P6_122$ and $P6_522$. A peak, with an approximately 25 Å vector to a two-fold symmetry related equivalent, identified the most probable iron site.

Once a set of native data from *Alcaligenes denitrificans* cytochrome c' had been collected (section 2.3.4), the anomalous differences were calculated, an anomalous difference Patterson was calculated, and the iron sites determined using the same method as in the previous paragraph. Table 4.5.4.1 gives the iron coordinates for *Alcaligenes* sp cytochrome c', but both species give the same positions, as expected from the fact that

they crystallise isomorphously. From these results it was possible to verify possible molecular replacement solutions (section 3.5.3) and to calculate a set of initial phases as described in section 4.5.5.

Table 4.5.4.1 Positions of the iron sites (orthogonal coordinates) in the unit cell of the cytochrome c' from *Alcaligenes* sp in $P6_122$ and $P6_522$.

Spacegroup $P6_122$			Spacegroup $P6_522$		
34.272	6.973	22.094	20.128	6.973	22.094
31.226	26.194	82.460	38.298	13.945	142.828
16.103	13.945	142.828	23.175	26.194	82.460
-7.072	40.140	112.644	7.072	40.140	112.644
-4.025	20.918	173.010	-11.097	33.167	52.278
11.098	33.167	52.278	4.026	20.918	98.640
-11.098	33.167	38.272	-4.025	20.918	98.640
4.026	20.918	98.640	11.098	33.167	38.272
7.072	40.140	159.006	-7.072	40.140	159.006
38.298	13.945	128.822	31.226	26.194	8.090
23.174	26.194	8.090	16.103	13.945	128.822
20.128	6.973	68.456	34.272	6.973	68.456

4.5.5 Phasing from anomalous scattering.

Phase information from the iron anomalous scattering was used to test the correctness of the molecular replacement solutions obtained from AMORE (as described in section 3.5.3), and to calculate an electron density map for model building. Although the anomalous signal is weak from a single iron atom (f" = 3.4 electrons for CuK_α radiation, so that in a molecule of ~ 1000 atoms a single iron atom should give an average intensity difference of ~ 2 % (Crick and Magdoff, 1956)), data sets from both *Alcaligenes* species were available and were of high quality.

Best phases (Blow and Crick, 1959) to a resolution of 3.0 Å were calculated for each species using only the iron anomalous scattering; this gave a mean figure of merit of 0.04 and 0.05 for the *Alcaligenes* sp and *Alcaligenes denitrificans* data sets respectively. When MLPHARE (CCP4 suite, see appendix I) was used to refine the sites and occupancy it was unsuccessful. The occupancy refined to zero, probably due to the

weakness of the anomalous signal. Instead each data set was placed on an approximately absolute scale through a Wilson plot (Wilson, 1949) and an iron occupancy of 1.0 was assumed. Electron density maps calculated using these phases were of poor quality but the two maps (from *Alcaligenes* sp and *Alcaligenes denitrificans*) did show reasonable agreement (correlation coefficient 0.63). To reduce noise, a combined electron density map was calculated, simply by adding the two individual maps, point by point, with equal weight using the program OVERLAPMAP (CCP4 suite, see appendix I). From this point the phases were improved using solvent flattening and histogram matching, as is described in detail in section 5.2.

Chapter Five

Structure determination

5.1 Density modification.

The average solvent content of protein crystals is about 43 % although this varies between 27 % and 65 % (Matthews, 1968) and may even be as high as 80 %. For the majority of crystals it is expected that the protein density should be separated by regions of solvent, and since most of the solvent is disordered these solvent regions should appear flat in comparison with the protein density. Within a crystal there should be no negative density.

In practice experimentally determined electron density maps do not often meet these criteria due to errors in the phasing. The errors in the phases can be reduced by exploiting this knowledge of the character of the electron density in a protein crystal, by levelling the density in the solvent region and by truncating negative densities in the protein region. The resulting modified map is backtransformed and the resulting phases are combined by weighting with the original phases. The combined phases, which should be improved compared with the original ones, can then be used to calculate a new map and the procedure is repeated until convergence is achieved (Zhang, 1993).

5.2 Improvement of anomalous scattering phases and their use in determining the correct molecular replacement solution.

Best phases (Blow & Crick, 1959) to a resolution of 3.0 Å had been calculated for each of the Alcaligenes species using only the scattering from the iron atoms (section 4.5.5); this gave a mean figure of merit of 0.04 and 0.05 for the *Alcaligenes* sp and *Alcaligenes denitrificans* data sets, respectively. Electron density maps calculated using these phases were, not surprisingly, of poor quality but the two maps (for *Alcaligenes* sp and *Alcaligenes denitrificans*) did show a reasonable agreement (correlation coefficient 0.63). This provided the encouragement to undertake further phase improvement. Phases were subjected to the solvent flattening and histogram matching procedures in the program SQUASH (Cowtan and Main, 1993) (CCP4 suite, see appendix I). In an attempt to reduce noise, a combined electron density map was also calculated, simply by adding the two

individual maps, point by point, with equal weight, using the program OVERLAPMAP (CCP4 suite, see appendix I).

Further improved phases were obtained from the combined map by carrying out a Fourier transform on it and then treating the resulting phases with SQUASH (CCP4 suite, see appendix I). The electron density map calculated using these "squashed", combined phases, with amplitudes from the *Alcaligenes* sp data set, to a resolution of 3.0 Å, was used in the initial stages of structure building. Representative sections of this map are shown in Figures 5.2.1 and 5.2.2 (Model used in all figures is the final refined *Alcaligenes* sp cytochrome c' structure). From inspection of the electron density map it was possible to distinguish the haem and its environs and most of the helical nature of the protein. The loop regions, however, had very little traceable density.

This electron density map also verified the correct molecular replacement solution, found in AMORE (section 3.5.3). A quantitative measure of the agreement between each of the two MR solutions and the electron density maps phased by the anomalous scattering iron, was obtained by comparing the latter with maps calculated from the MR models (F_c maps), and expressing the agreement as the correlation coefficient (program OVERLAPMAP). The results of these comparisons are summarised in Table 5.2.1. Solution 3 was clearly shown to be incorrect, as its F_c map had zero correlation with the anomalously-phased maps. Solution 11, on the other hand, showed a positive correlation of 0.17 to 0.18; moreover this figure was increased for the solvent flattened combined maps, and gave a correlation coefficient of 0.24 with the final map calculated using the solvent flattened, combined anomalous phases. This was taken as evidence that solution 11 was indeed correct. The usefulness of the preceding rigid-body refinement was shown by the fact that correlation coefficients obtained from the unrefined model were significantly lower, 0.11 to 0.16. Table 5.2.3 shows the correlations of the electron density maps calculated from the molecular replacement solutions during the course of phase improvement by solvent flattening.

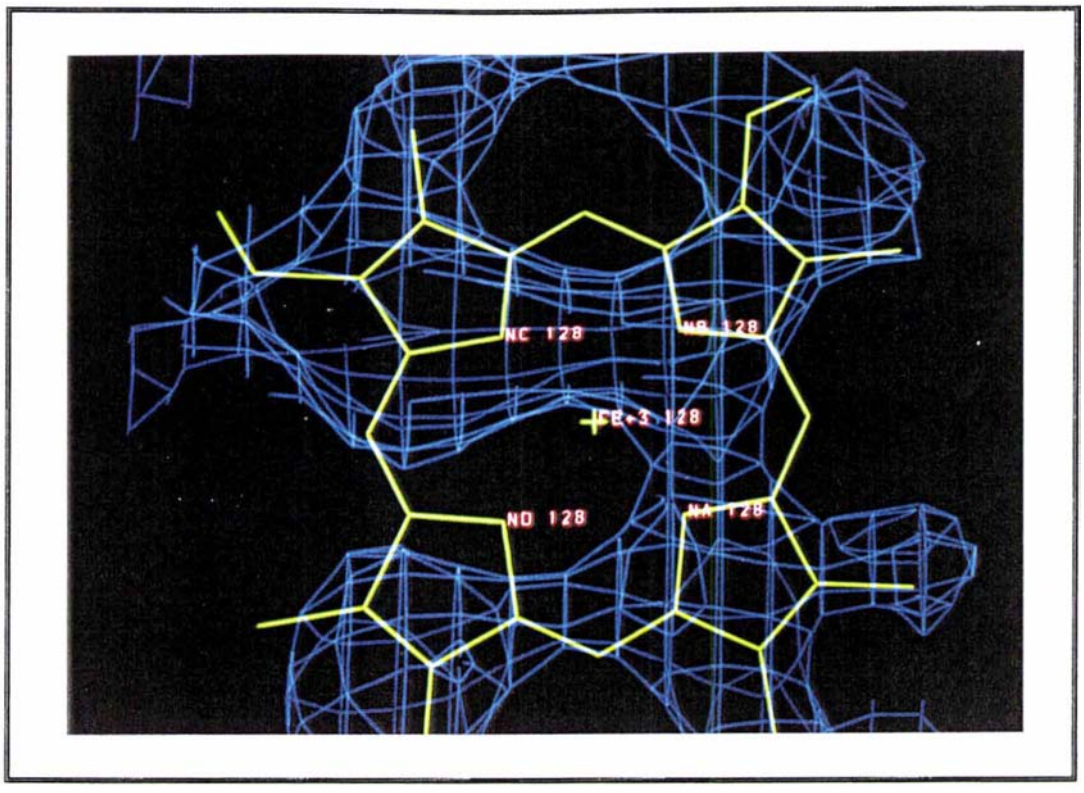


Figure 5.2.1 Section of anomalously phased map showing haem density.

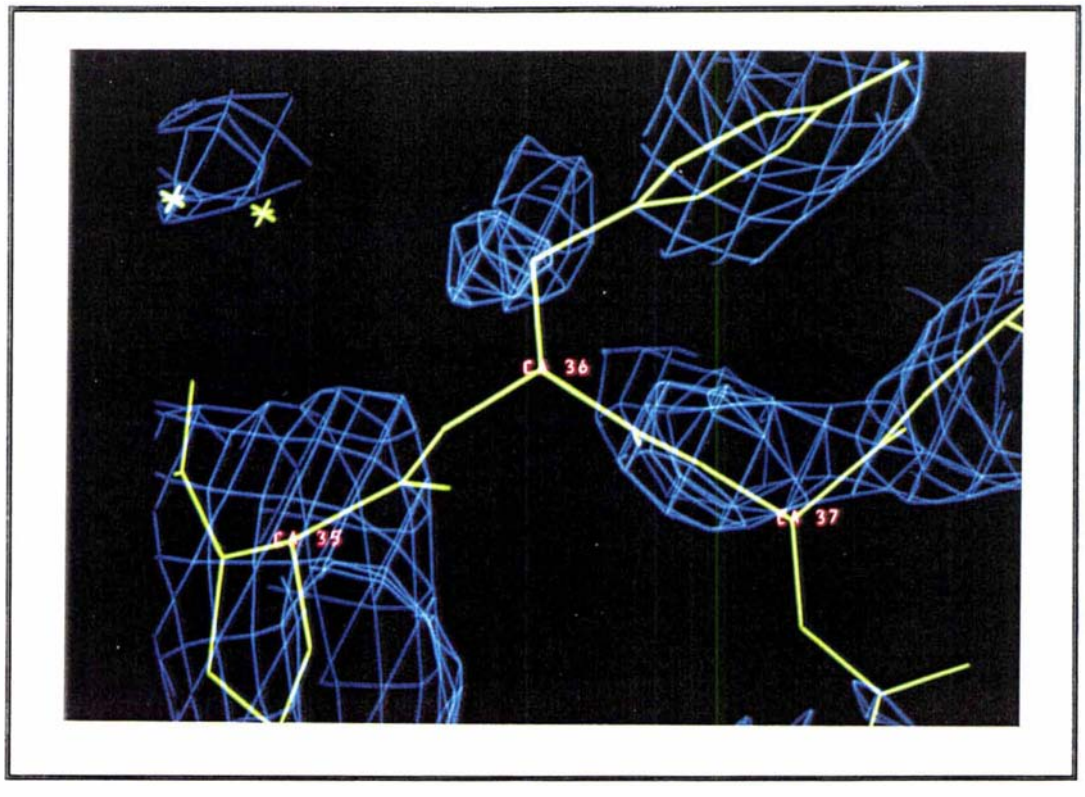


Figure 5.2.2 Section of anomalously phased map showing loop density.

Table 5.2.1 Electron density map correlations[a].

Model	Correlation coefficients					
	F_c/Asp.z	F_c/Ade.z	F_c/Asp.zsq	F_c/Ade.zsq	F_c/Comb.z	F_c/Comb.zsq
Solution 3	0.004	-0.010	-0.012	-0.004	-0.002	-0.002
Solution 11	0.189	0.172	0.199	0.184	0.200	0.242
(before XPLOR)	0.138	0.102	0.136	0.112	0.164	
Solution 11	0.220	0.231			0.243	0.303
First rebuild,						
R = 0.404						
Solution 11	0.251	0.277			0.282	0.354
Second rebuild,						
R = 0.364						

a - Code for maps: F_c = calculated map from MR model; Asp.z and Ade.z = anomalous phased maps from *Alcaligenes* sp and *Alcaligenes denitrificans* data respectively; Asp.zsq and Ade.zsq = anomalous phased maps from *Alcaligenes* sp and *Alcaligenes denitrificans* data after phase improvement with SQUASH; Comb.z = combined anomalous-phased map; Comb.zsq = combined anomalous-phased map after SQUASH.

Two rebuilds of the structure were made from this anomalous-phased, combined, "squashed" map, after each of which the model was subjected to a round of simulated annealing refinement using X-PLOR (see appendix I). This led to an R factor of 0.36 for data in the range 10.0 to 3.0 Å, and a correlation coefficient of 0.35 between the Fc map and the anomalous-phased map. Extension of the resolution to 2.0 Å and refinement using the restrained least squares program PROLSQ (see appendix I) further reduced the R factor to 0.32. At this stage, however, the density in the loop regions could still not be traced with any certainty

and attempts at rebuilding the model only made the R factor increase. To carry on with rebuilding, improved phases were required, so it was decided to see what improvement could be obtained using the single isomorphous phases with anomalous scattering.

5.2.1 Improvement of SIR phases using solvent flattening.

Initial phases were calculated as described in section 4.5.3. The electron density maps calculated using these phases provided a second independent confirmation of the molecular replacement solution. The main chain density was basically complete for the four helices and for the atoms comprising the haem group, but again the density in the loop regions was poor and these regions could not be traced with any certainty. To improve the phases, and therefore the resulting map, solvent flattening was undertaken, using SQUASH (CCP4 suite, see appendix I). The resolution range used was between 15.0 - 2.35 Å. After eight cycles of solvent flattening, the overall figure of merit had increased from 0.454 to 0.585, but an electron density map calculated with the new "squashed" phases gave no discernible improvement of the density in the loop regions. This is shown in Figure 5.2.1.1.

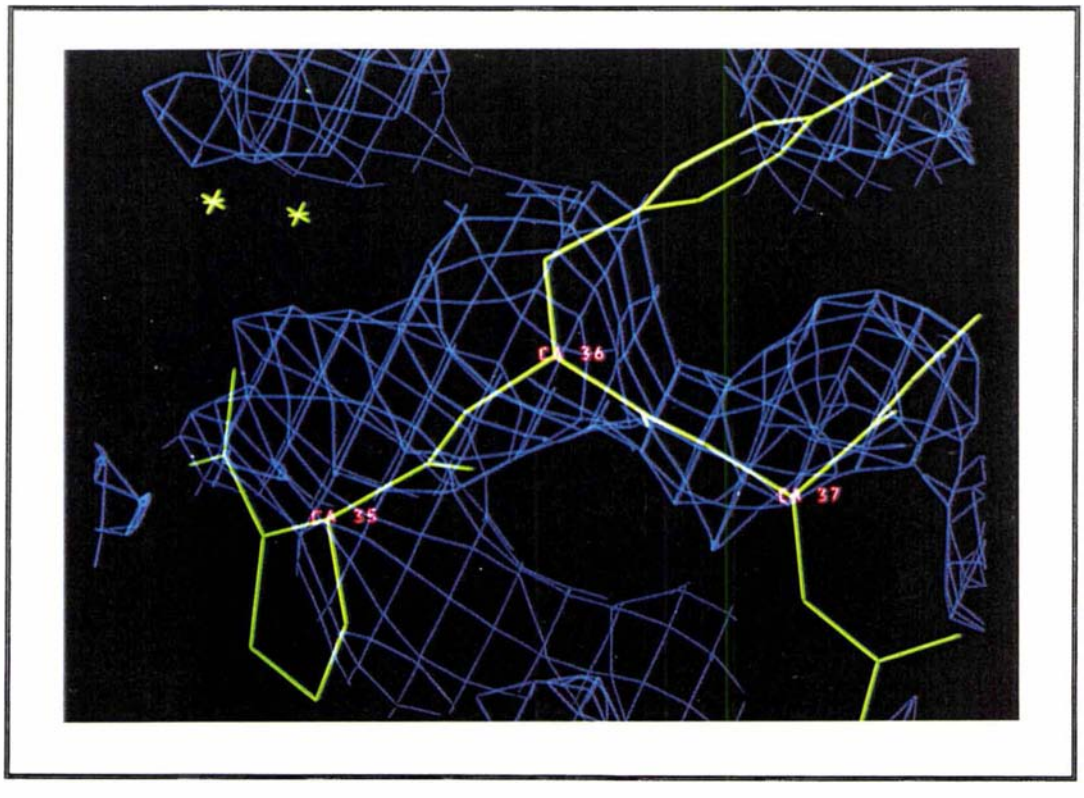


Figure 5.2.1.1 Electron density in AB loop after solvent flattening of SIR phases.

5.2.2 Phase improvement of the SIR phases using molecular averaging.

The "squashed" SIR phases were further modified using B-C Wang's method of molecular averaging and phase extension (Wang, 1985). This involves "averaging" the map by replacing the electron density at each grid point by the weighted average of the densities at surrounding grid points, within some radius. The molecular envelope is then defined by applying a threshold to the 'averaged' map such that the volume below the threshold is the estimated solvent content. Phases were calculated by molecular averaging to a resolution of 3.0 Å. Six cycles of the molecular averaging were used; this improved the figure of merit from 0.72 after the first cycle of molecular averaging to 0.90 after six cycles. However electron density maps calculated from these phases again gave no better density in the loop regions. A pictorial representation of Wang's method of noise filtering is given in Figure 5.2.2.1. While in Figure 5.2.2.2 electron density from this map is shown.

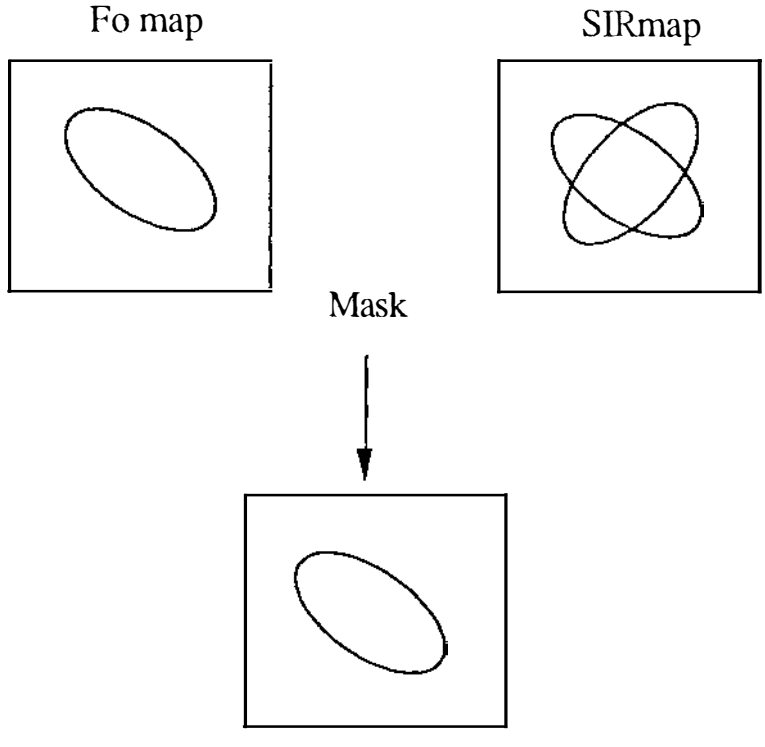


Figure 5.2.2.1 Diagram showing the two phase possibilities arising from single isomorphous replacement and the application of a solvent mask. The result is the removal of "noise" from the correct phase.

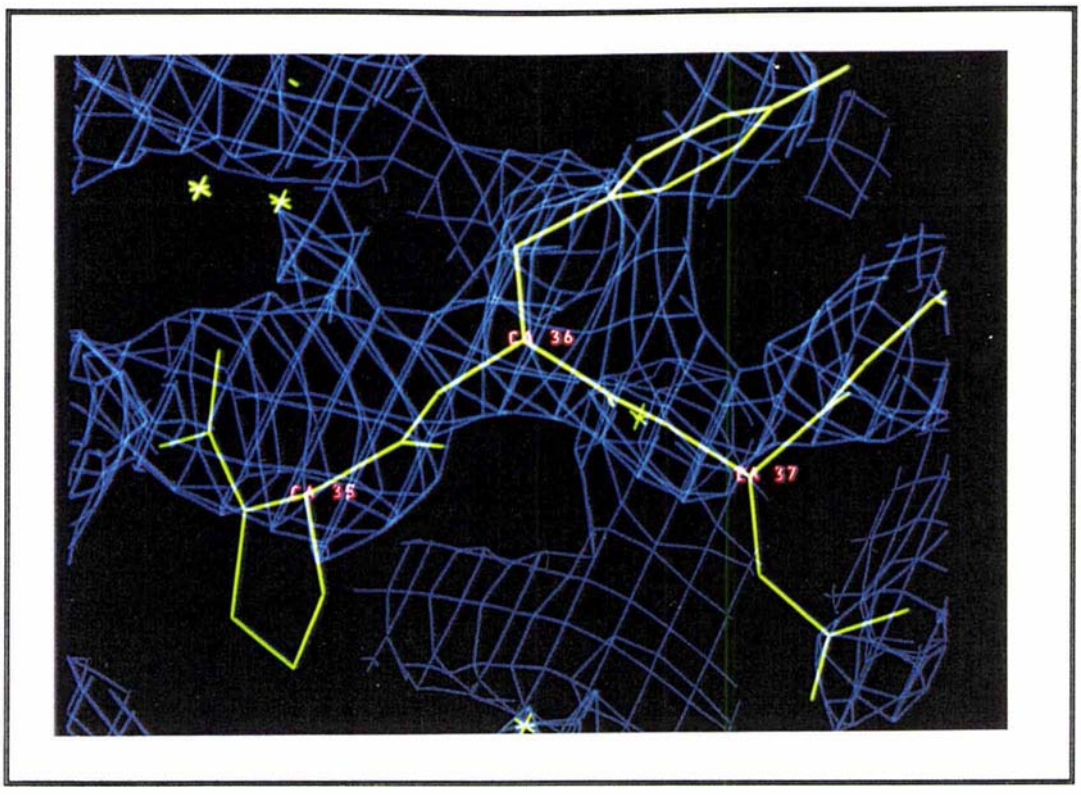


Figure 5.2.2.2 Electron density in AB loop from "Phasext" map.

5.2.3 Combining SIR phases with iron anomalous scattering phases using SIGMAA.

The next attempt at improvement involved the use of SIGMAA (CCP4 suite, see appendix I) in order to combine the phases calculated from single isomorphous replacement (after solvent flattening) and those calculated from the anomalous scattering of the iron (see section 4.5.3). Again this had the effect of improved density for the helices and the haem group, but not in the loop regions. When the anomalous scattering phases were combined with the SIR phases the overall figure of merit to 3.0 Å resolution was 0.769. Figure 5.2.3.1 gives an example of the density in one of the loop regions.

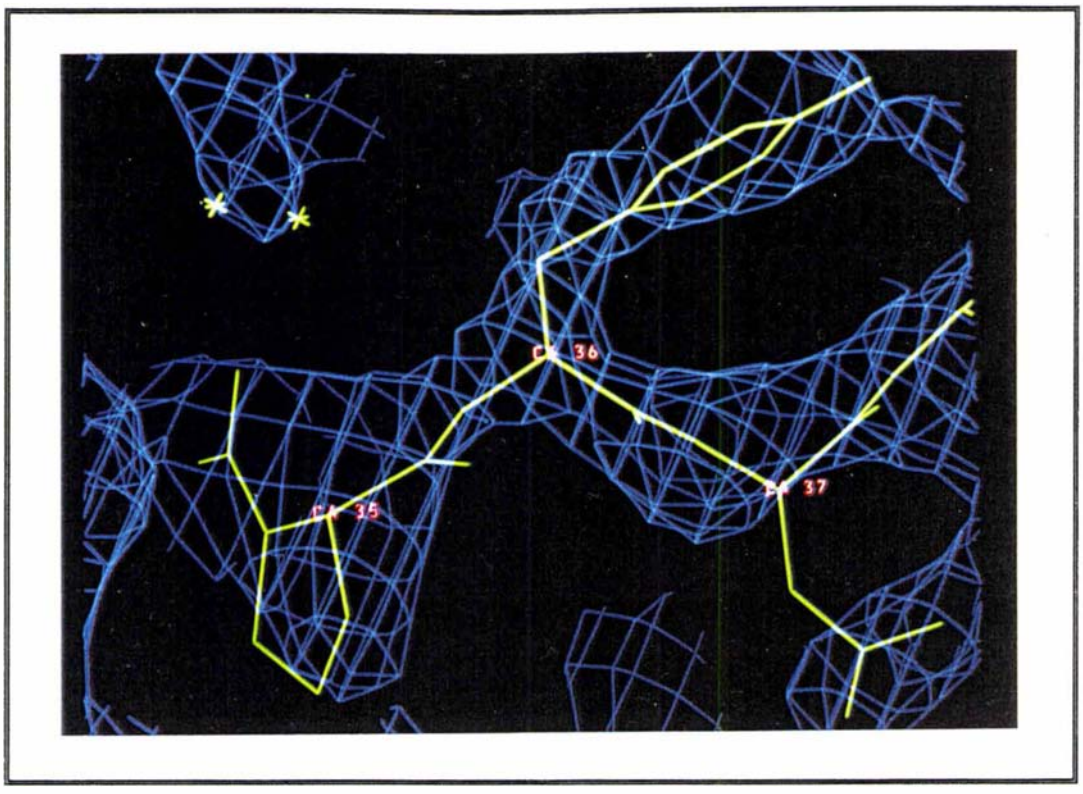


Figure 5.2.3.1 Density from the AB loop in the combined "SIR" and "Iron" phased map.

5.2.4 Combining SIR phases with anomalous scattering phases using OVERLAPMAP.

After consideration of the improvement that was obtained by simply adding the two anomalously phased maps point by point (see section 5.2) it was decided to try a similar approach to improve the SIR phases. The map calculated using phases extended by B-C Wang's method (1985) (see section 5.2.1) and the map from the combined anomalous phases (section 5.2) were added point for point using OVERLAPMAP. The correlation between these maps was 0.174. The resulting map was then itself added to the map calculated from the combination of the SIR phases with the squashed anomalous phases using SIGMAA . The correlation between these maps was 0.790. Inspection of electron density in the loop regions of this map showed density in the map from which the main chain atoms of the loops could be traced with reasonable certainty. This density was then used for rebuilding the model obtained by the initial refinement and rebuilding of the molecular replacement solution (section 5.2). Figures 5.2.4.1 and 5.2.4.2 show the density for the AB and BC loops from this combined map (these Figures can be compared with those in other sections). Table 5.2.4.1 gives the maps used in the density combination.

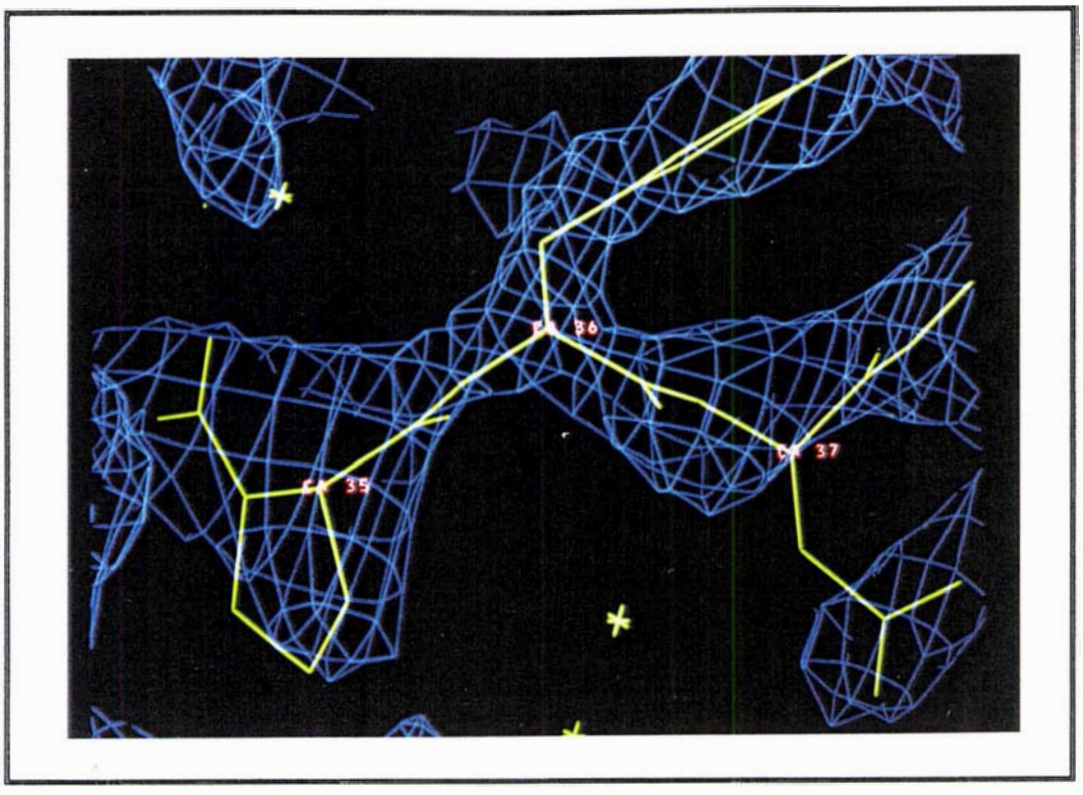


Figure 5.2.4.1 Density in AB loop from "comb2" map.

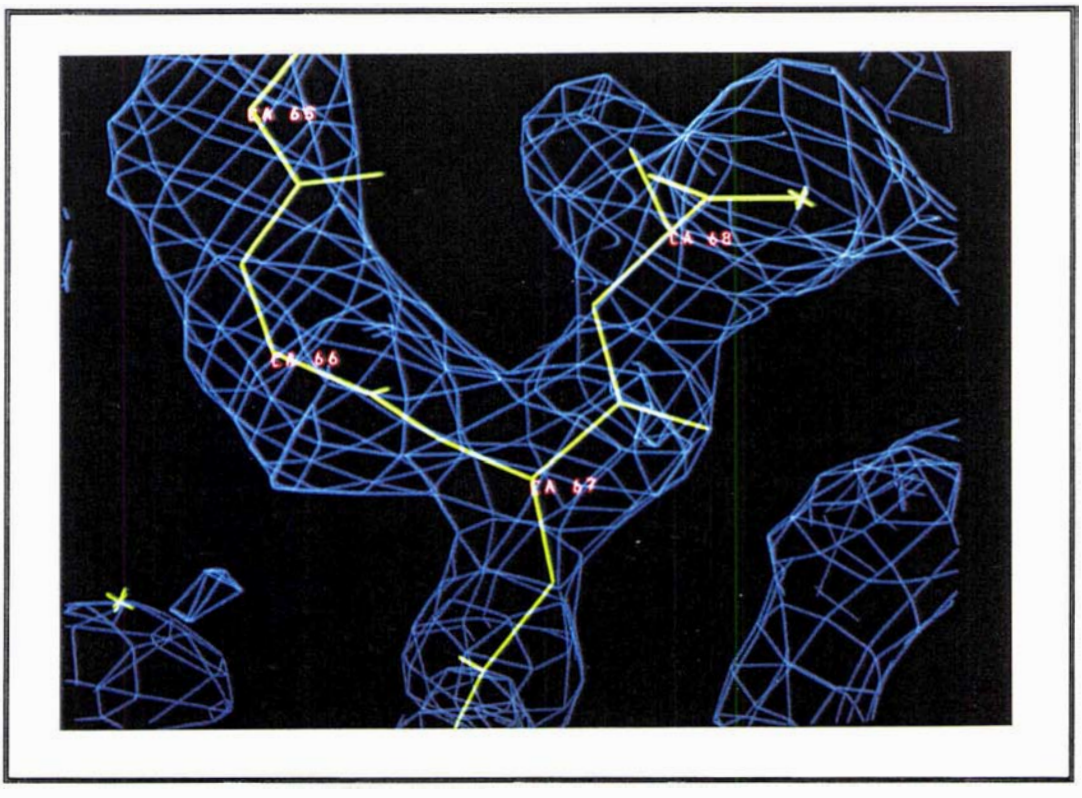


Figure 5.2.4.2 Density in BC loop from "comb2" map.

Table 5.2.4.1 Maps used in density combination using OVERLAPMAP (resolution 15.0 - 3.0 Å). ("Iron" is the squashed combined anomalous-phased map, "Phasext" is the map from Wang's method (1985) and "Sir + Iron" is the map phased from the combination of the squashed anomalous phases map with the squashed SIR phases)

Maps used	Overlapmap(1)	Overlapmap(2)
"Iron"	Comb1.map (0.174)	Comb2.map (0.790)
"Phasext"		
"Sir + Iron"		

For some sidechains the density was present but lacked the definition to orient accurately; these were then built by placing them in sterically favourable conformations. Loops AB and BC were rebuilt first and then refined. Loop CD was rebuilt last, and as the refinement progressed the sidechains were rebuilt as necessary. The progress of the refinement is discussed in section 5.4.

5.3 Refinement methods.

5.3.1 Least squares refinement with TNT.

For small molecule crystallography, full matrix least squares routines can be used in structure refinement because the ratio between the number of observations (X-ray diffraction data) and atom parameters (x,y,z,B) is relatively high, generally higher than 10:1 (e.g. 3000-6000 observations for a 30-40 atom structure). In macromolecular crystallography, this method is inappropriate because the ratio is low, perhaps as low as 2:1 (e.g. 20000 observations for a 2000 atom structure at 2.0 Å resolution), and most of the X-ray data from protein crystals at high resolution are weak. This means a method of restrained refinement is required.

Refinement of the *Alcaligenes* sp and *Alcaligenes denitrificans* cytochrome c' structures was completed using the restrained least squares program TNT 5-C (Tronrud *et al.*, 1987). Like all least squares refinement programs TNT uses structural knowledge from previous structures to restrain the stereochemistry of the model close to "ideal" values.

The function being minimised is:

$$F(x) = \sum_{x=1} (Q_o(i) - Q_c(i,x)^2$$

where $Q_o(i)$ is the i(th) observation and $Q_c(i,x)$ is the corresponding calculated value given a set of parameters "x". If the model is being restrained by several types of observations it is more convenient to use a separate term for each type. Then the function becomes;

$$f(x) = f1(x) + f2(x) + f3(x) + ...$$

where f1(x) is the sum over the first class of observations and so forth. The benefit of restrained refinement is that the ratio between contributions from X-ray data and ideal geometry can be varied.

Normally the restraints on the geometry during the refinement are kept tight. However the restraints on the geometry can be loosened by increasing the weight of the X-ray function in the minimisation. Specific changes to the weights of the

constraints can also be made e.g. by tightening and loosening the weights on the bond angles.

5.3.2 X-PLOR.

Initial refinement of the *Alcaligenes* sp structure was undertaken using simulated annealing with X-PLOR Version 3.1 (Brünger, 1992). X-PLOR is based on an energy function approach; arbitrary combinations of empirical, geometric and effective energy terms describing experimental data can be used. The combined energy function can be minimised by a variety of gradient descent, simulated annealing, and conformational searches. Crystallographic refinement, although conceptually simple, can be very time-consuming since restrained least squares refinement is easily trapped in false minima. The method of crystallographic refinement by simulated annealing (SA) (Brünger, 1991) can generate an ensemble of structures, each of which agrees with the diffraction data. SA-refinement consists of simulating the many-parameter system by molecular dynamics. Initially, the temperature is kept very high and then the system is "annealed" by reducing the temperature slowly. The "target function" that is used in SA-refinement consists of an effective potential energy (Ex) that is described by the crystallographic residual and an empirical potential energy that comprises the stereochemistry and non-bonded interactions of the macromolecule.

5.3.3 Model rebuilding.

Each time convergence was reached during refinement, either with X-PLOR or with TNT, an electron density map was calculated. The fit of the protein structure to the density was examined visually and incorrect parts were omitted from the model (or parts of structure with high B-values were omitted). The model was then subjected to refinement for three cycles to remove any bias of the electron density resulting from incorrect structure. Lastly, "omit" maps were calculated with the coefficients 2Fo-Fc and Fo-Fc, i.e.

$$(2|Fo|-|Fc|)e^{-i\alpha c}$$

and $$(|Fo|-|Fc|)e^{-i\alpha c}$$

where |Fo| = observed structure factor amplitude

|Fc| = calculated structure factor amplitude

αc = calculated phase

The first map should show the electron density of the molecule including that of the omitted residues or sidechains, while the latter map shows only the electron density of the omitted structure with minimal bias towards the current model.

The rebuilding was carried out manually, using interactive graphics systems. Two systems were used, firstly an Evans and Sutherland PS-330 graphics terminal and the program FRODO (Jones, 1978) was used to rebuild the *Alcaligenes* sp cytochrome c' structure. Secondly a Personal Iris 4D/30 and the package TOM Version 2.4 (Jones, 1982) was used for the rebuilding of the *Alcaligenes denitrificans* cytochrome c' structure.

5.3.4 Inclusion of solvent.

During refinement, solvent molecules were included in the model. No solvent molecules were added until the crystallographic R-factor had dropped to ~ 0.21 in order to avoid incorrect assignment of the solvent structure. All of the solvent molecules included in the cytochrome c' structures were treated as water. Peaks were located in the 2Fo-Fc and Fo-Fc Fourier maps and were labelled as water molecules if:

(1) the density was discrete and well defined,

(2) the peak height was above 3σ in the Fo-Fc map,

(3) the geometry of the interactions with potential hydrogen bonding partners was good.

The most well ordered solvent molecules were clearly resolved in the Fo-Fc maps and consequently included in the model early, whereas the least well ordered solvent molecules were more poorly resolved in the Fo-Fc maps and were added towards the end. The occupancies for the solvent molecules were not refined, being kept at unity; errors in the occupancy will be largely taken up by the B-values (Blundelland Johnson, 1976).

Water molecules already in the model were checked by inspecting the 2Fo-Fc and Fo-Fc maps, and by checking their B-values after refinement. Solvent molecules were removed if there was no density in the 2Fo-Fc map at 1σ and the B-value was greater than 60 $Å^2$. The geometry of the potential hydrogen bond formed was also considered. A conservative approach was undertaken as to the number of solvent molecules added to the protein structure. Inspection of the final Fo-Fc difference map showed a number of low density areas which could be interpreted as solvent molecules. However, when water molecules were placed into this density and

refined, the B-values for these sites became very high (>65 $Å^2$) so they were removed from the final model (these may be disordered or low occupancy solvent molecules).

5.4 Progress of refinement.

5.4.1 Refinement of *Alcaligenes* sp cytochrome c'.

Manual rebuilding into electron density for the cytochrome c' from *Alcaligenes* sp (ASCC) was undertaken on an Evans and Sutherland PS 300 using the package FRODO (Jones, 1978). The molecular replacement model (section 5.2) was initially fitted to the combined, anomalous phased, "squashed" map (section 5.2) with sidechains and loops being placed as best they could into the density available. Sidechains with density but ambiguity in orientation were arranged in their sterically preferred conformations, while those without reasonable density were omitted. This initial refinement of this model is described in section 5.2.

Following this initial phase, the combined density map (Comb2) obtained using OVERLAPMAP (see section 5.2.4), was used to carry on rebuilding. The model was rebuilt into this density and refinement by simulated annealing lowered the R factor to 0.272 (Free R = 0.345) for data greater than 2 σ between 5.0 and 1.8 Å resolution. Calculated phases from this model still did not provide unambiguous loop tracing. The electron density map calculated from this model was therefore combined (using OVERLAPMAP) with the map Comb2. Another rebuild followed by simulated annealing dropped the R factor to 0.252 (Free R = 0.321). By now the loop regions could be traced with certainty, although several of the lysine sidechains in the structure had little density. Two further rebuilds were made, with rounds of simulated annealing following each rebuild lowering the R factor to 0.221 (Free R = 0.275). At this stage only residues 2-122 had been included in the model, although density could now be seen for further residues at the C-terminus. The refinement was then continued using TNT Version 5-C (Tronrud *et al.*, 1987; see appendix I) in order to be able to better control the geometrical constraints during refinement, and to add the haem linkages. During the next rebuild residues 123, 124 and 125 were added at the C-terminus of the structure. Residue 125 was initially built as alanine (lysine is the correct amino acid) as no density for the sidechain was apparent. The first solvent molecules (38 in total) were also added to the model. The resolution range for the data used in refinement was then increased from 5.0 - 2.0 Å to 8.0 - 2.0 Å. B-factor refinement was then carried out concurrently with positional

refinement, and 15 more water molecules were added to the model. After a total of 65 cycles of TNT refinement, the crystallographic R factor had been lowered to 0.186 with an *rms* deviation of bond lengths of 0.014 Å and *rms* deviation of bond angles of 2.1 ° (*rms* is defined as the root mean square deviation from standard values). The model now consisted of residues 2 - 125, the haem group and a total of 61 solvent molecules. The resolution range was then increased from 8.0 - 2.0 Å to 20.0 - 2.0 Å. The N-terminal residue was added to the model. This was initially built as an alanine until a correct geometry file was constructed. The amino acid sequence (Ambler, 1973) suggested that this residue is a pyroglutamic acid. The density at the N-terminus confirmed this observation (see Figure 6.4.1). A further 25 cycles of B-factor and positional refinement were then carried out, slightly raising the R factor to 0.189. The resolution range was again increased, to 20.0 - 1.8 Å and 29 further water molecules were added. The alanine at residue 125 was changed to its correct amino acid (lysine) as density was now available to place the sidechain. After a further 65 cycles of TNT refinement the R factor was 0.198 and the model consisted of residues 1 - 125 and the haem group plus a total of 100 waters. The constraints on the geometry were then loosened, and the lysine sidechains in the model were truncated depending on the density available for the sidechains. A further 26 waters were added bringing the number of solvent molecules to 126. The R factor was lowered to 0.185 after 30 cycles of refinement had been carried out. Finally the correct N-terminal residue was modelled in the structure and those waters which had very high B-factors (>65 $Å^2$) were removed.

The final model consists of 956 protein atoms plus 89 waters and has been refined to a an R factor of 0.188 from 20.0 to 1.8 Å for 14,834 reflections with no σ cutoff. The final model has an *rms* deviation from standard bond lengths of 0.017 Å and *rms* deviation from standard bond angles of 2.4 °. A summary of the course of refinement of the cytochrome c' from *Alcaligenes* sp is given in Table 5.6.3.1.

Table 5.4.1.1 Course of refinement for the cytochrome c' from *Alcaligenes* sp.

Cycles of refinement	Free R-factor	R-factor	*Rms* deviation of bonds (Å)	*Rms* deviation of angles (°)	Comments
2 rounds XPLOR	0.382	0.304	0.020	3.4	Resolution range 5.0 - 1.8 Å, model consists of residues 2-122 and haem, all s/c's, 943 atoms.
2 rounds XPLOR	0.345	0.297	0.014	2.9	Rebuild 1, using combined density map (Comb2) to rebuild loops, residues 2-122 and haem, 943 atoms.
4 rounds XPLOR	0.321	0.255	0.013	2.8	Rebuild 2, second combined density map (Comb3) used, all loops placed in model , residues 2-122 and haem, 943 atoms.
6 rounds XPLOR	0.288	0.255	0.014	2.8	Rebuild 3, residues 2-122 and haem, 943 atoms.
7 rounds XPLOR	0.275	0.221	0.015	2.7	Rebuild 4, all side-chains of loops correctly placed, C-terminus of helix B rebuilt, residues 2-122 and haem, 943 atoms.
15		0.221	0.022	2.8	TNT refinement (xyz), resolution range 8.0 - 2.0 Å, added residues 123,124 and 125 to C-terminus. 38 waters added, all s/c's, 999 atoms.
15		0.213	0.021	2.8	Xyz refinement, added 15 waters, residues 2-125 and haem, 1014 atoms
10		0.193	0.017	2.4	XyzB refinement, residues 2-125 and haem, 1014 atoms.

Table 5.4.1.1 continued.

15		0.192	0.014	2.1	Xyz refinement, added 8 waters, residues 2-125 and haem, 1022 atoms
10		0.186	0.014	2.1	XyzB refinement.
25		0.189	0.014	2.2	Resolution range 20.0 - 2.0 Å, xyzB refinement, added alanine as N-terminal residue, added 29 waters, model consisted of residues 1-125 plus haem, 1051 atoms.
15		0.183	0.015	2.2	XyzB refinement, added 29 waters, removed 6 waters, changed Ala125 to Lys125, 1074 atoms.
50		0.198	0.015	2.8	XyzB refinement, resolution range between 20. 0 - 1.8 Å
30		0.185	0.017	2.5	xyzB refinement, loosened geometry,truncated lysine residues, added 13 water molecules, removed 5 waters, model consisted of residues 1-125, plus haem and 127 waters, 1082 atoms
10		0.188	0.017	2.4	Remodelled N-terminal residue as pyroglutamic acid, removed 37 waters. Final model consists of 1045 atoms (956 protein + 89 solvent molecules).

5.4.2 Refinement of *Alcaligenes denitrificans* cytochrome c'.

As the cytochromes c' from both species crystallise isomorphously the structures were expected to be similar. Once the refinement of the structure of cytochrome c' from *Alcaligenes* sp had reached an R factor of 0.185, and the model consisted of residues 1 - 125 plus the haem group and 125 waters, these refined coordinates (not including the water molecules) were subjected to three rounds of rigid body refinement against the *Alcaligenes denitrificans* data using TNT (Tronrud *et al.*, 1987). Each round consisted of five cycles. The structure was split into eight groups, comprising the four helices, connecting loops and haem group. After each round of rigid-body refinement, the resolution was increased (starting at 20.0 - 3.0 Å and finishing at 20.0 - 2.0 Å). This gave an R factor of 0.217.

At this stage the model was rebuilt using a Silicon Graphics Personal Iris 4D/30 with the interactive graphics package TOM Version 2.4 (Jones, 1982). and the solvent structure from the *Alcaligenes* sp cytochrome c' model was incorporated into the model. All the added water molecules were checked and were either i) refined, or ii) removed, if there was no density available for them, or iii) shifted if density was available close by. The sequence for the cytochrome c' from *Alcaligenes denitrificans* has yet to be determined so the structure was examined to see if there were any indications of residue changes in the electron density. From the density four sequence changes from that of *Alcaligenes* sp were made. Thr27 was changed to glycine, the proline at position 61 was changed to alanine, the serine at position 78 was changed to glycine and lastly Ala122 was remodelled as serine. After 23 waters were removed, 10 cycles of least squares refinement with TNT (Tronrud *et al.*, 1987) were carried out using data in the resolution range of 20.0 - 2.2 Å to give an R factor of 0.200.

The resolution was then increased to 20.0 - 1.85 Å and another rebuild was carried out. Waters were removed or shifted depending on their B-values or the presence or absence of density. This lowered the R factor to 0.169, at which stage the model consisted of residues 1-125, the haem group and a total of 103 water molecules. Finally the correct N-terminal residue was modelled in the structure, and those waters which had very high B-factors (>65 $Å^2$) were removed.

The final model consists of 953 protein atoms and 75 water molecules and has been refined to an R factor of 0.167 for 8,220 reflections (no σ cutoff) in the resolution range 20.0 to 2.15 Å. The final model has an *rms* deviation from standard bond lengths of 0.013 Å and *rms* deviation from standard bond angles of 2.1 °. A summary of the course of refinement for the cytochrome c' from *Alcaligenes denitrificans* is given in Table 5.4.2.1.

Table 5.4.2.1 Summary of the course of refinement of the cytochrome c' from *Alcaligenes denitrificans.*

Cycles of refinement	R-factor	*Rms* deviation of bonds (Å)	*Rms* deviation of angles (°)	Comments
5	0.188			Rigid body refinement - 8 groups resolution 20.0 - 4.0 Å.
5	0.187			Rigid body refinement as above, resolution 20.0 - 3.0 Å
5	0.217			Rigid body refinement, resolution 20.0 - 2.0 Å.
15	0.200	0.023	2.9	Xyz refinement, resolution 20.0 - 2.2 Å, added 103 waters from *Alcaligenes* sp model, rebuilt model, changed sequence to that suggested by density, slight adjustment in C-terminus of helix A and around residues 60-63. Model consisted of residues 1-125 plus haem, 1036 atoms.
15	0.203	0.024	2.9	Xyz refinement, increased resolution 20.0 - 1.86 Å.
15	0.169	0.014	2.2	XyzB refinement.
15	0.167	0.013	2.1	XyzB refinement, replaced N-terminal residue with correct pyroglutamic acid, removed 28 waters leaving total of 75. Final model consists of residues 1-125 plus haem (also includes 75 water molecules), 1026 (952 protein) atoms in total. Resolution range changed to 20.0 - 2.15 Å

5.5 Quality of the final models

5.5.1 Accuracy of the model for *Alcaligenes* sp cytochrome c'.

The structure of cytochrome c' from *Alcaligenes* sp has been determined and refined using data between 20.0 and 1.80 Å. The current model has an R factor of 0.188 and *rms* deviations from standard bond lengths and bond angles of 0.017 Å and 2.4 ° respectively. Residues 126 and 127 at the C-terminus have not been included in the final model. All the lysine sidechains in the model have been truncated at C_δ (except those at positions 82, 92 and 125 which are stabilised by hydrogen bonding). Although all of the aspartic acid sidechains have been included in the final model, several of these (67, 75, 98 and 121), situated on the outside of the molecule and not involved in hydrogen bonding have unclear density and high B-values. Gln33 and Gln81 have poor density and high B-values for their sidechains. Arg69 and 106 also have high sidechain B-values for the sidechain although their density is reasonable (see section 6.9 for mainchain and sidechain B-value plots). The rest of the structure has well defined continuous electron density in the final 2Fo-Fc electron density map, when contoured at 1σ. Two examples of the fit of the structure to the density are shown in Figures 5.5.1.1 and 5.5.1.2.

As a quantitative measure of the error in the structure the mean error of the coordinates was estimated from the variation in R factor with resolution as described by Luzzati (1952). The Luzzati plot is shown in Figure 5.5.1.3, with standard error lines included. Except at low resolution, where the higher R values can be attributed to the failure to fully account for the solvent structure, the error in the structure is approximately 0.18 Å. This compares with the value calculated using SIGMAA (CCP4 suite, see appendix I) of 0.16 Å. The value obtained from the graph actually represents the maximum average error and in fact the error is probably considerably less in the well defined internal regions.

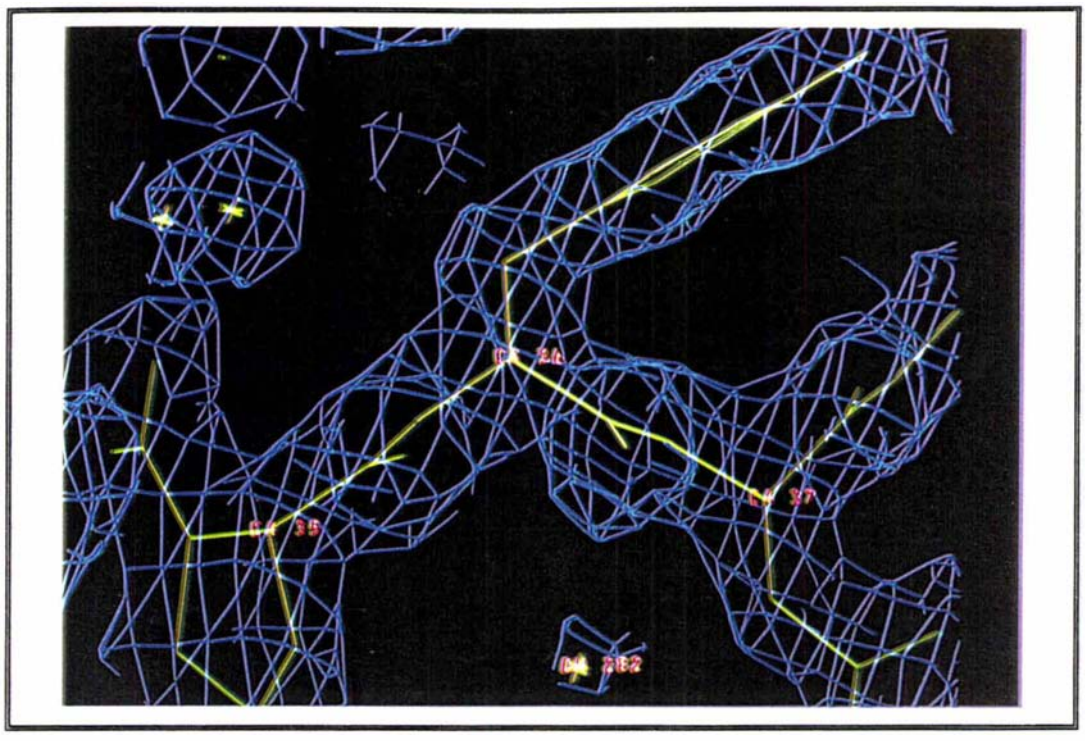


Figure 5.5.1.1 AB loop, residues 34-36 in the cytochrome c' from *Alcaligenes* sp, shown in the final 2Fo-Fc map.

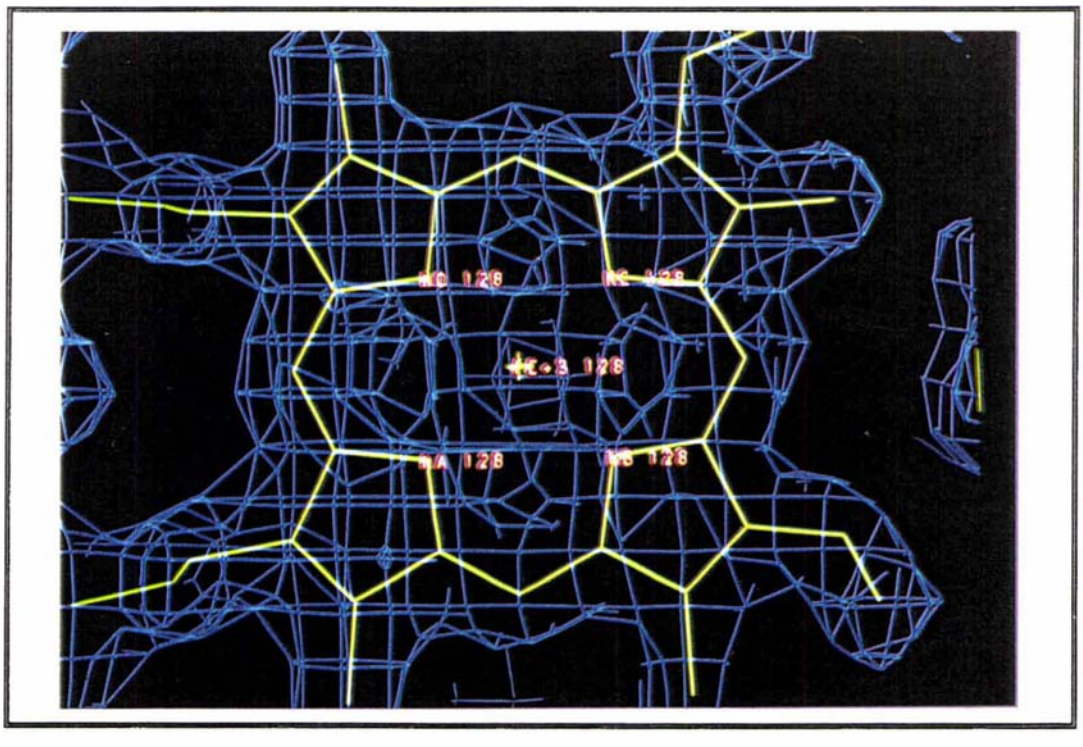


Figure 5.5.1.2 Haem pyrrole rings from the *Alcaligenes* sp cytochrome c', shown in the final 2Fo-Fc map.

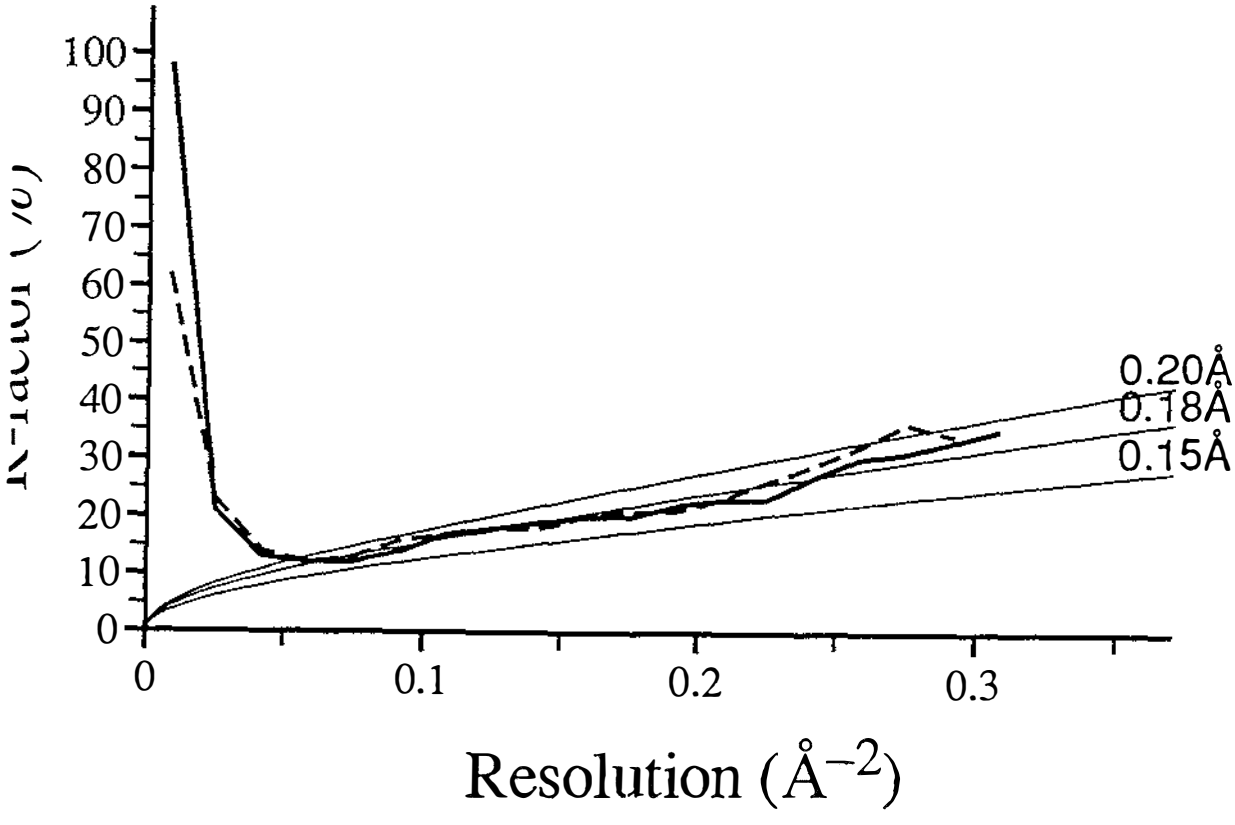


Figure 5.5.1.3 Luzzati plot for the cytochrome c' structures of *Alcaligenes* sp (thick line) and *Alcaligenes denitrificans* (dotted line).

The fit of the atomic structure to the density was quantified by calculating real space correlation coefficients using the program OVERLAPMAP (CCP4 suite, see appendix I). The correlation of the model with the data is very high (see Figure 5.5.1.4) for all the structure. The correlation coefficients for the sidechain atoms follow a similar pattern to that seen for the mainchain (see Figure 5.5.1.5). The sidechains with lowest correlation coefficients are labelled in the figure, although it should be noted that none has an average correlation coefficient lower than 0.875. The residues that have the lowest correlation coefficients are on the surface of the protein, while the residues that have the highest correlation coefficients are found in the protein interior. Figure 5.5.1.6 shows the lack of electron density around Asp75.

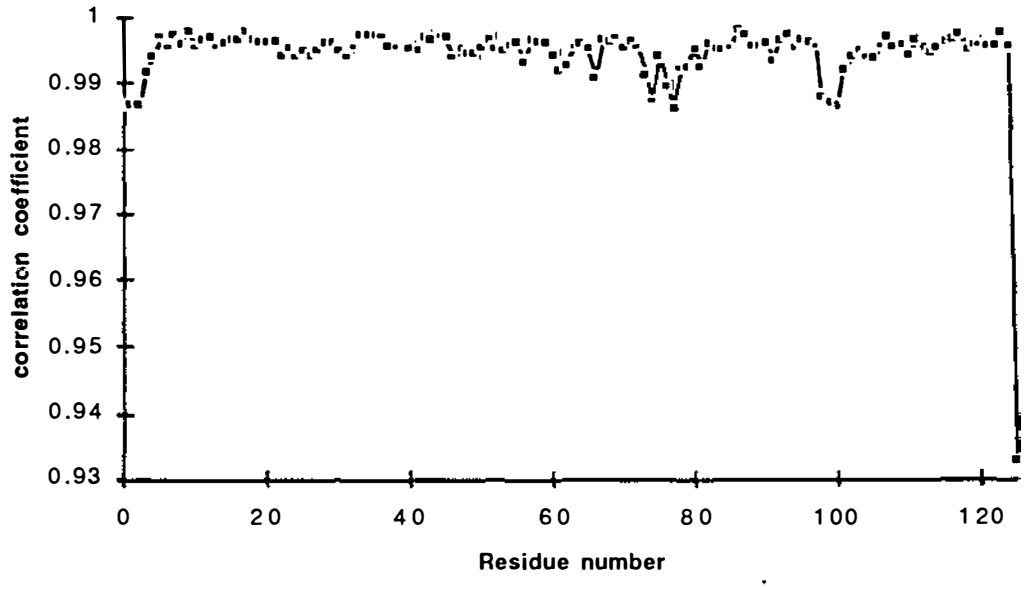


Figure 5.5.1.4 Real space correlation coefficient plot for the mainchain atoms of cytochrome c' from *Alcaligenes* sp.

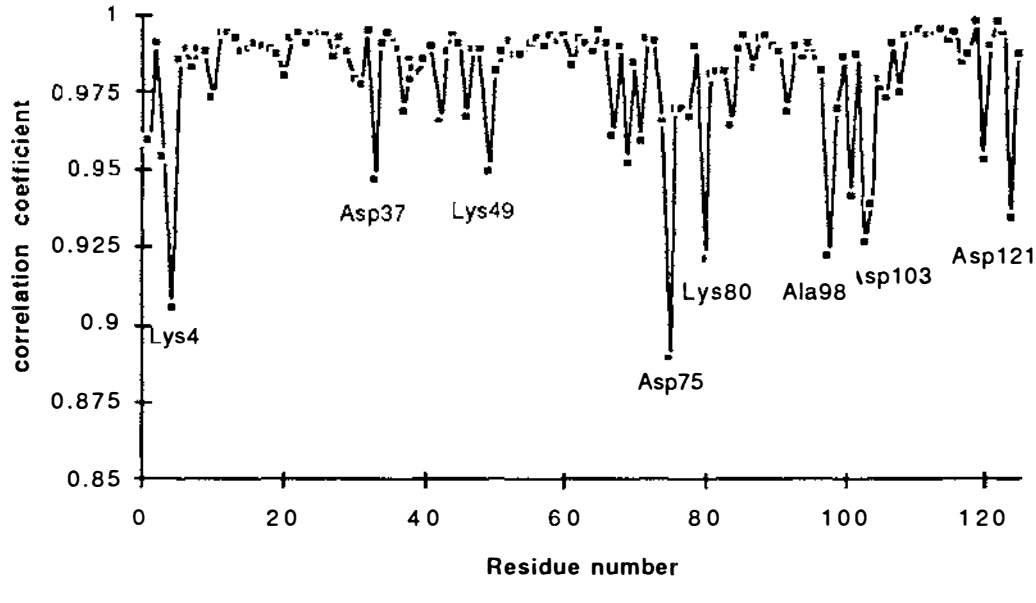


Figure 5.5.1.5 Real space correlation coefficient plot for the sidechain atoms of cytochrome c' from *Alcaligenes* sp.

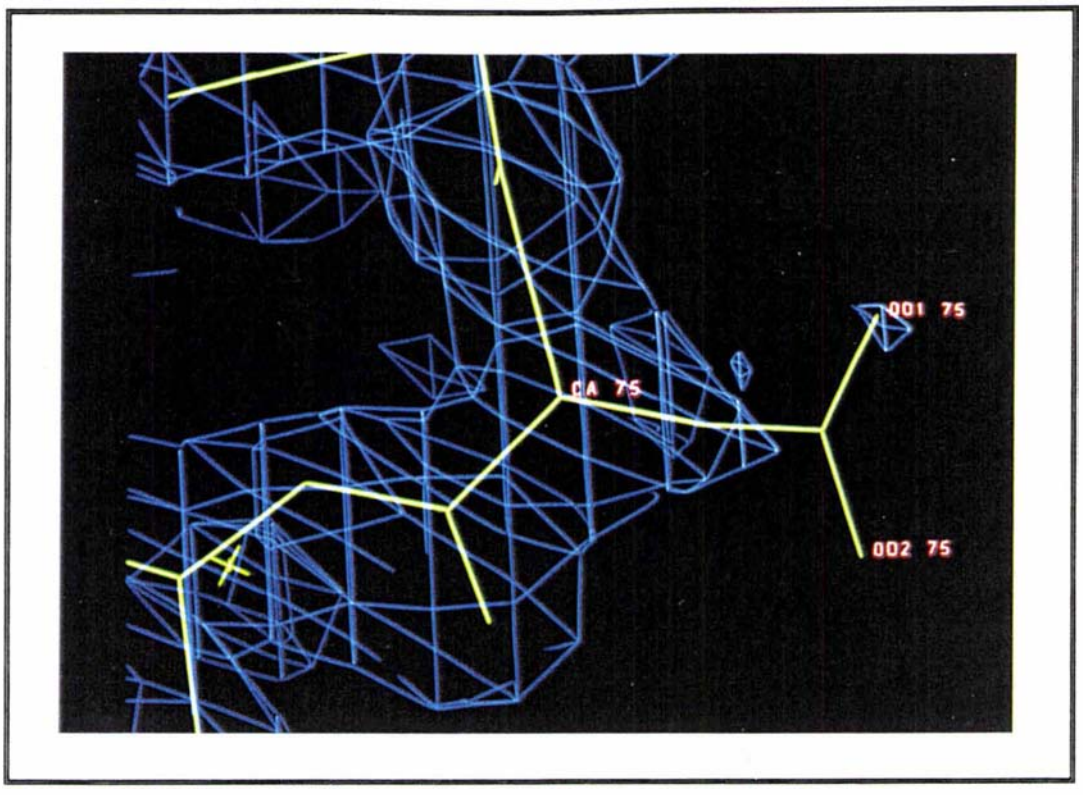


Figure 5.5.1.6 Density around Asp75 in the cytochrome from *Alcaligenes* sp, shown in the final 2Fo-Fc map.

5.5.2 Accuracy of the cytochrome c' structure from *Alcaligenes denitrificans*.

The current model of cytochrome c' from *Alcaligenes denitrificans* has an R factor of 0.167 for 8,220 reflections between 20.0 and 2.15 Å and *rms* deviations from standard bond lengths and angles of 0.013 Å and 2.1 ° respectively. The accuracy of the structure was assessed by several criteria. Firstly the mean error in the coordinates was estimated from the variation in R factor with resolution as described by Luzzati (1952). From the Luzzati plot shown in Figure 5.5.1.3 the mean error in the coordinates was estimated to be 0.18 Å. This compares with the value calculated using SIGMAA (CCP4 suite, see appendix I) of 0.16 Å. At low resolution the R factor increases, as in the cytochrome c' from *Alcaligenes* sp, because the incompleteness of the solvent structure. Figure 5.5.2.1 gives an example of how well the final model fits the density.

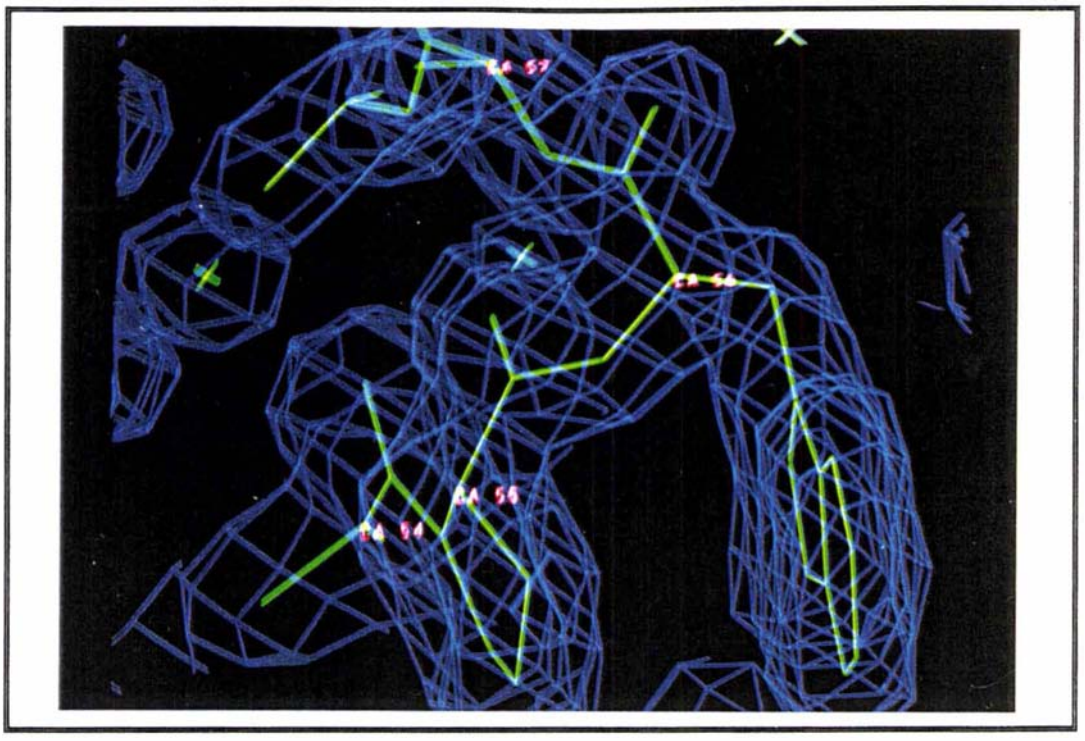


Figure 5.5.2.1 Example of the model from the *Alcaligenes denitrificans* cytochrome c' fitting the final 2Fo-Fc electron density (region around Pro55, Trp56).

The agreement of the model with the data was quantified by calculating real space correlation coefficients using OVERLAPMAP (CCP4 suite, see appendix I). All of the correlation coefficients are above 0.94 for the mainchain indicating the that model fits the density very well. Figures 5.5.2.2 and 5.5.2.3 show real space correlation coefficients for the mainchain and sidechain atoms.

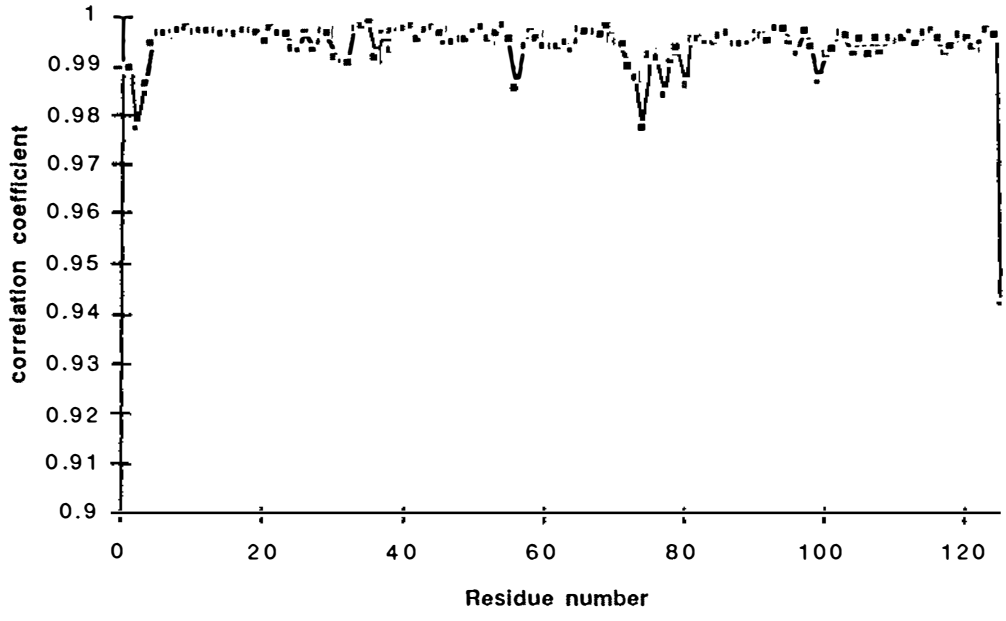


Figure 5.5.2.2 Real space correlation coefficient plot for the mainchain atoms of cytochrome c' from *Alcaligenes denitrificans*.

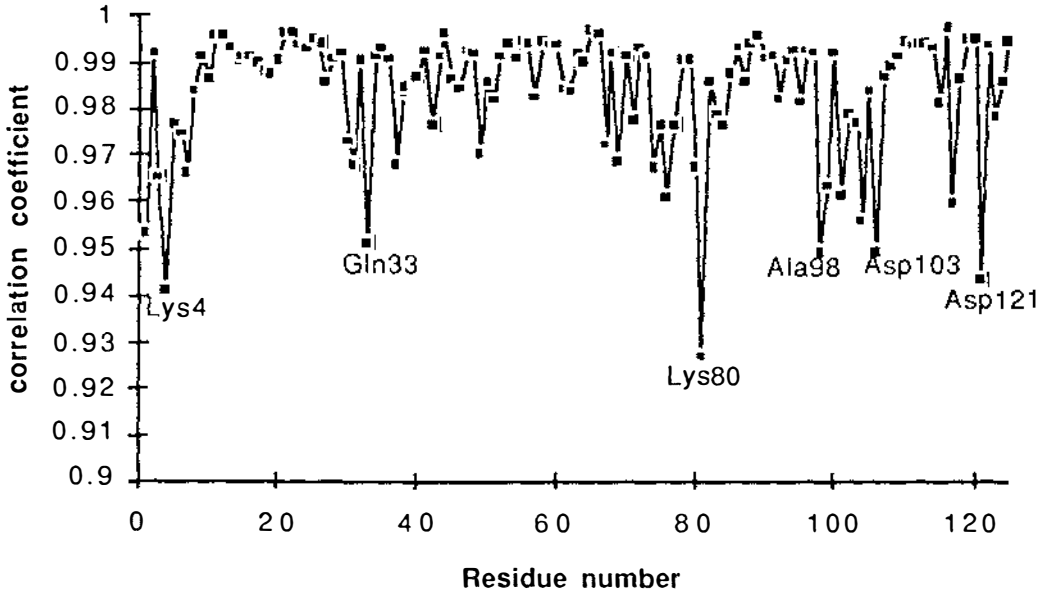


Figure 5.5.2.3 Real space correlation coefficent plot for the sidechain atoms of cytochrome c' from *Alcaligenes denitrificans*.

As in the case of *Alcaligenes* sp cytochrome c' all the lysine sidechains have been truncated at their Cδ atoms except for residues 82, 92 and 125. Some aspartic acid residues present on the surface of the protein (especially those in loop regions) have high B-values with little electron density available. Gln33 and Gln81 again have poor density for their sidechains and other residues with unclear density include Arg69 and Arg106 which also have high B-values (see section 6.9 for mainchain and sidechain B-value plots). Figure 5.5.2.4 shows Arg69 and the density for its sidechain.

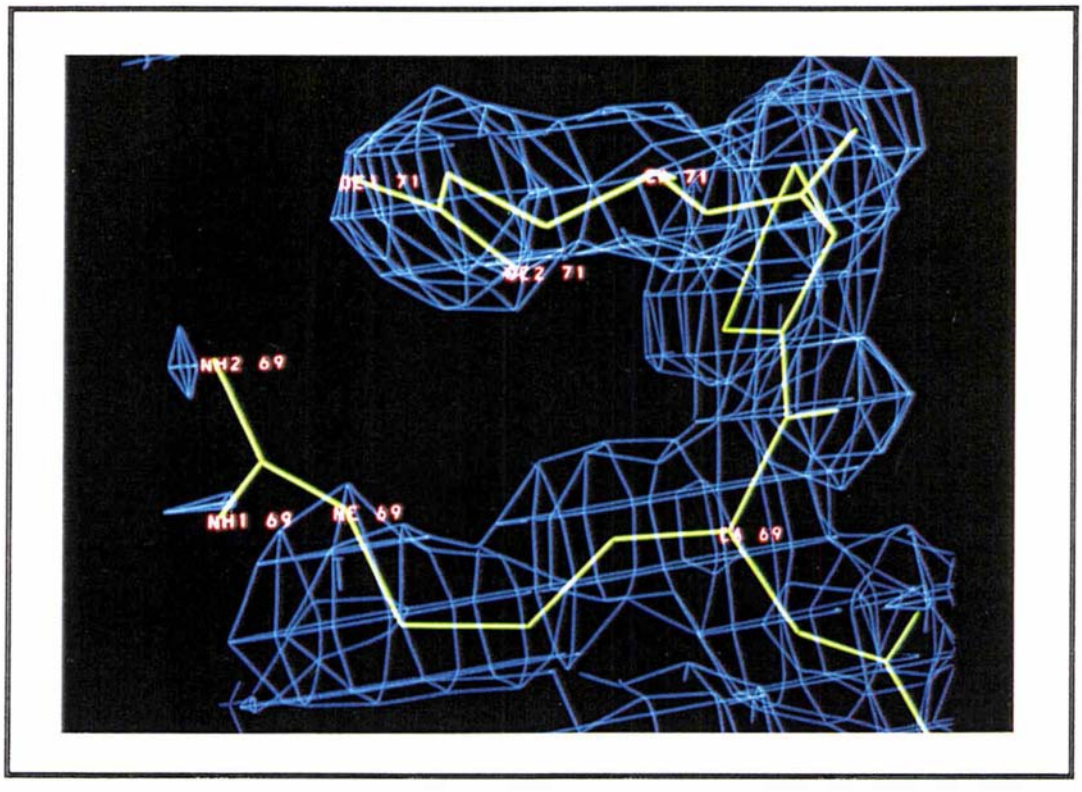


Figure 5.5.2.4 Electron density available for the sidechain of Arg69, in the final 2Fo-Fc map for *Alcaligenes denitrificans* cytochrome c'.

5.5.3 Conformational angles.

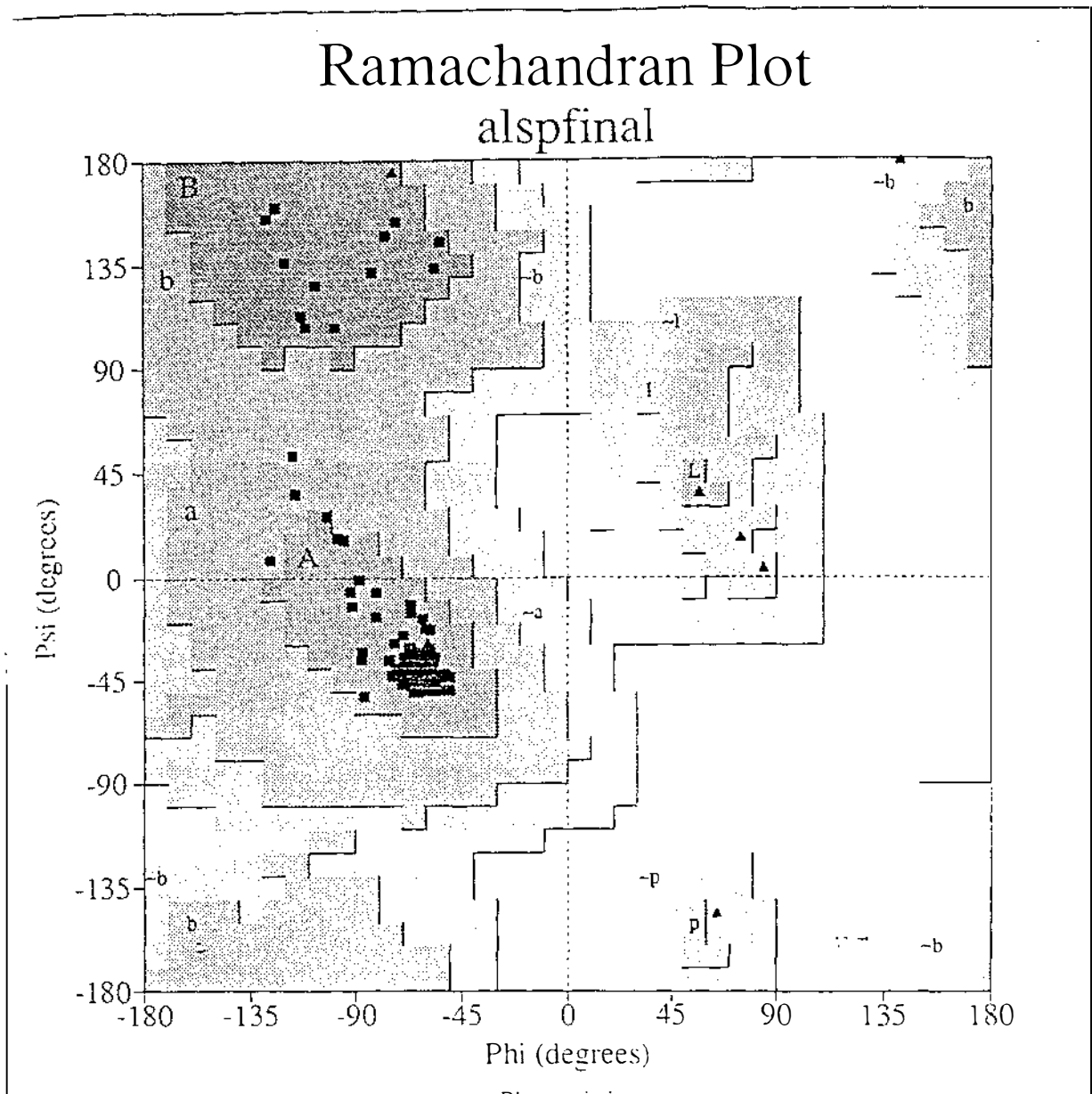
Mainchain torsion angles (ϕ, ψ)

Figures 5.5.3.1 and 5.5.3.2 show Ramachandran plots (Ramakrishnan and Ramachandran, 1965) of the mainchain torsion angles (ϕ, ψ) in the two cytochrome c' structures. The plots show that the structures have no residues with unfavourable (ϕ, ψ) angles, with 97.2 % residues in the most favoured regions in the case of

Alcaligenes sp cytochrome c', and 96.3 % in the structure from *Alcaligenes denitrificans*. The dominant α-character is shown by the clustering of residues around (-60 °, -40 °) which confirms the helical nature of these residues. A small number of residues have (ϕ, ψ) values in the β-region, these residues occurring in turns, or at the N and C-termini where the backbone exists as an extended chain. The average helical mainchain conformational angles of the cytochromes c' from *Alcaligenes* sp and *Alcaligenes denitrificans* are given in Table 5.5.3.1 (value calculated after leaving out the respective N and C-terminal residues in each helix). These values correlate very well with the helix ϕ and ψ angles (-65.3 ± 11.9 ° and -39.4 ± 11.3 ° respectively) quoted by Morris *et al.* (1992) in their assessment of protein structures.

Table 5.5.3.1 ϕ, ψ values for the helical regions in the cytochromes c' from *Alcaligenes* sp and *Alcaligenes denitrificans*.

		Alcaligenes sp	*Alcaligenes denitrificans*
Helix A	ϕ	-66.2 ° ± 8.4 °	-64.2 ° ± 7.6 °
(residues 5-30)	ψ	-37.0 ° ± 10.8 °	-39.2 ° ± 9.7 °
Helix B	ϕ	-64.5 ° ± 9.5 °	-63.1 ° ± 8.4 °
(residues 38-54)	ψ	-38.8 ° ± 10.1 °	-40.7 ° ± 9.9 °
Helix C	ϕ	-63.4 ° ± 4.4 °	-62.6 ° ± 3.9 °
(residues 76-98)	ψ	-40.0 ° ± 4.6 °	-40.5 ° ± 4.6 °
Helix D	ϕ	-62.7 ° ± 7.0 °	-62.9 ° ± 8.4 °
(residues 102-122)	ψ	-42.1 ° ± 4.5 °	-42.1 ° ± 6.4 °

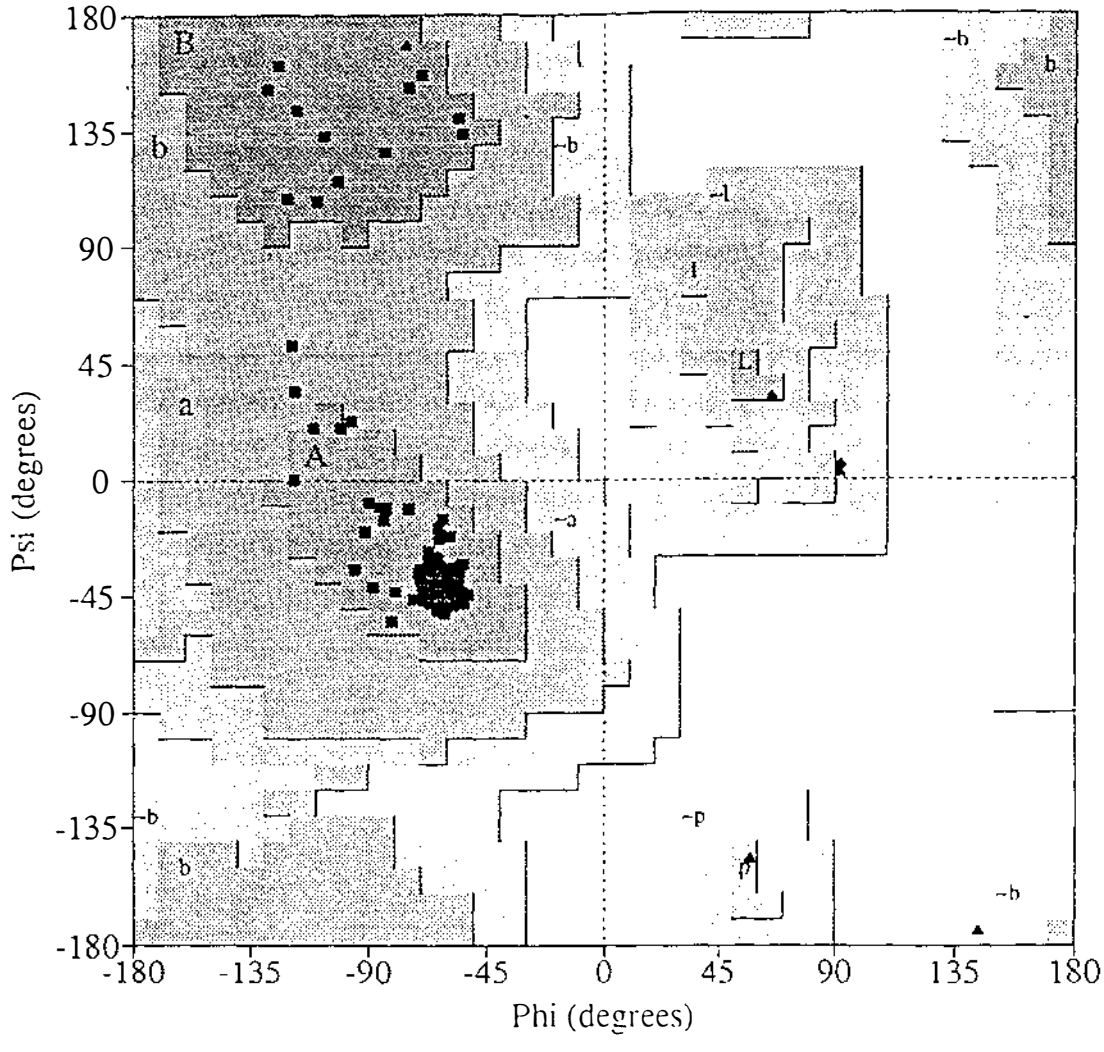


Plot statistics

Residues in most favoured regions [A,B,L]	105	97.2%
Residues in additional allowed regions [a,b,l,p]	3	2.8%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	108	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	9	
Number of proline residues	6	
Total number of residues	125	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figure 5.5.3.1 Ramachandran plot for *Alcaligenes* sp cytochrome c'.



Plot statistics

Residues in most favoured regions [A,B,L]	104	96.3%
Residues in additional allowed regions [a,b,l,p]	4	3.7%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	108	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	10	
Number of proline residues	5	
Total number of residues	125	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figure 5.5.3.2 Ramachandran plot for *Alcaligenes denitrificans* cytochrome c'.

Sidechain torsion angles ($\chi_1 - \chi_5$)

The distribution of sidechain torsion angles is similar for the cytochromes c' from *Alcaligenes* sp and *Alcaligenes denitrificans*. For example the means and standard deviations for the χ_1 (gauche +) torsion angles are -66.9 ± 13.8 ° and -66.5 ± 15.0 ° respectively. The standard deviations for the sidechain conformational angles vary more than the main-chain torsion angles, and this effect is more pronounced for those torsion angles which are further away from the backbone of the polypeptide chain. This could be due to the fewer constraining forces that the sidechains experience relative to the mainchain atoms, and the lessening of constraints on the sidechains further out from the backbone of the protein chain. Figures 5.5.3.3, 5.5.3.4 show histograms for the χ_1 (gauche -), χ_1 (*trans*) and χ_1 (gauche +) energy wells in the cytochromes c' from *Alcaligenes* sp and *Alcaligenes denitrificans* respectively. In general, the tighter the distribution of the histogram, the better the quality of the model (Morris *et al.*, 1992).

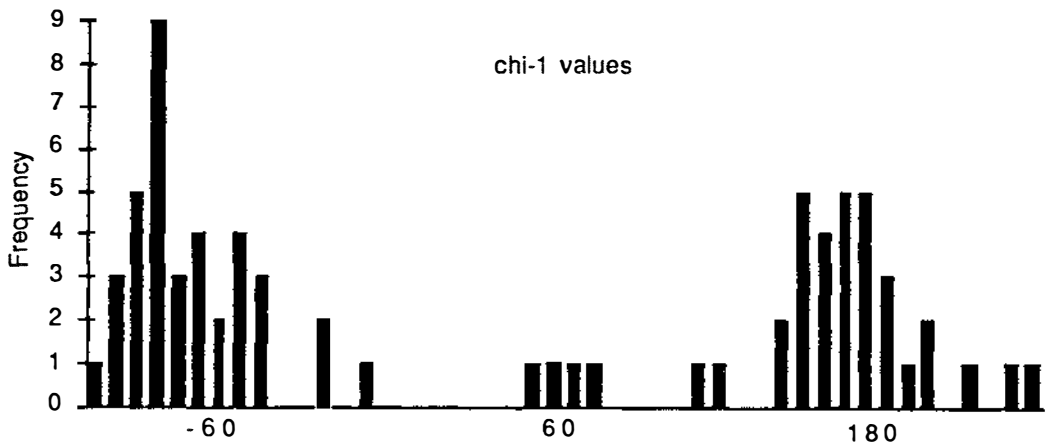


Figure 5.5.3.3 Distribution of chi-1 values in the cytochrome c' from *Alcaligenes* sp.

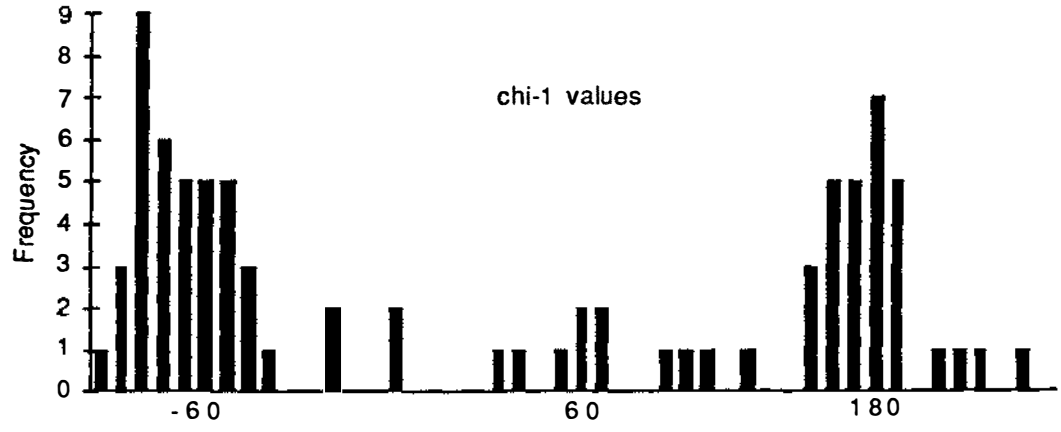


Figure 5.5.3.4 Distribution of chi-1 values in the cytochrome c' from ***Alcaligenes denitrificans***.

From Figures 5.5.3.3 and 5.5.3.4 it can be seen that there are very few residues whose sidechains have χ_1 torsion angles of approximately 60 ° (gauche -), in these cytochromes c'. The small number is due to the influence of the mainchain conformation, since sidechains of residues in helices prefer to have their sidechains in either *trans* or gauche + configurations (Janin *et al.*, 1978; Gray and Matthews, 1984). This has been suggested to result from the close contact that can occur between the sidechains and the mainchain (Gray and Matthews, 1984) as shown for example by serine residues, for which a close contact between O_γ and the (n-3) oxygen frequently occurs in an α-helix. This also occurs with threonine residues in an α-helix.

Chapter Six

Structure description

The asymmetric unit of the crystal contains a monomer of cytochrome c', with the functional dimer being created by a crystallographic two-fold axis. In this respect it differs from the cytochromes c' from *R. molischianum*, *R. rubrum* and *Chr. vinosum*, which all crystallise with a dimer in the asymmetric unit.

6.1 The overall fold.

Consistent with the previous structures of cytochrome c' that have been determined, the 125 residue monomer is based on four antiparallel helices (A to D) which form a left-twisted bundle (Weber and Salemme, 1980). These helices are connected by loops AB, BC and CD, with the shorter loops being at one end (AB and CD) and the larger loop at the other end (BC). Figure 6.1.1 shows a Molscript (Kraulis, 1991) diagram of the cytochrome c' monomer, while in Figure 6.1.2 a C_α stereo plot of the monomer is displayed. It can be seen from these figures that the bundles are well suited for the incorporation of prosthetic groups due to the packing requirements unique to bundles of α-helices (Weber and Salemme, 1980). In the bundle, the helices splay out from their point of closest approach at the end containing the AB and CD connecting loops. This has the effect of creating a hydrophobic pocket for the haem group to be incorporated into the structure. The four-helix bundle can also be thought of as comprising two antiparallel pairs of helices, A + B and C + D. The angle between A and B is 160 °, and that between C and D is 164 °; other inter-helix angles are given later in Table 7.4.5, where they are compared with other cytochromes c'.

These four helices account for approximately 75 % of residues in the cytochrome c' monomer. The assignments of residues to these helices in the two cytochromes c' are given in Table 6.1.1.

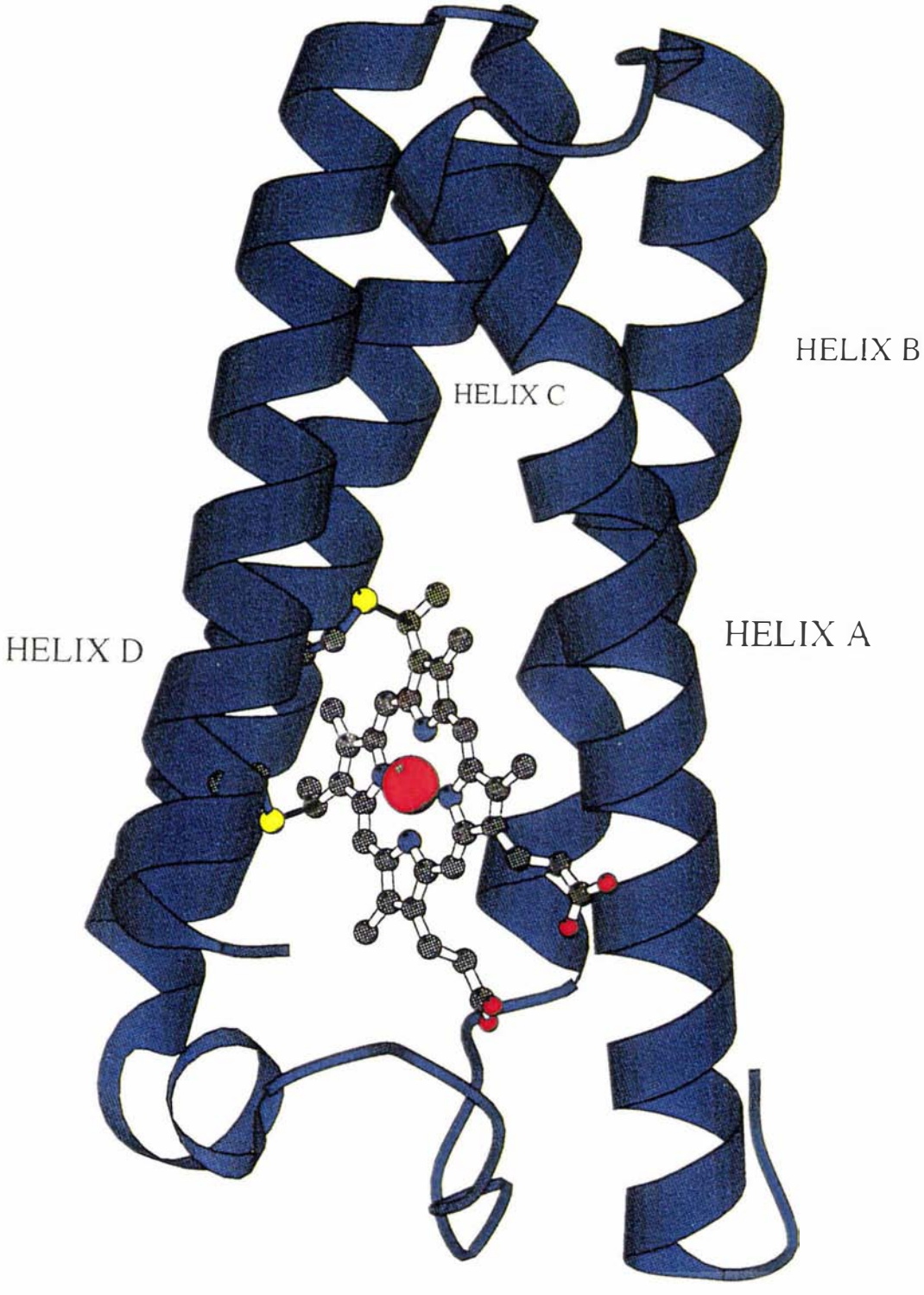


Figure 6.1.1 Molscript (Kraulis, 1991) diagram of the cytochrome c' monomer from *Alcaligenes* sp.

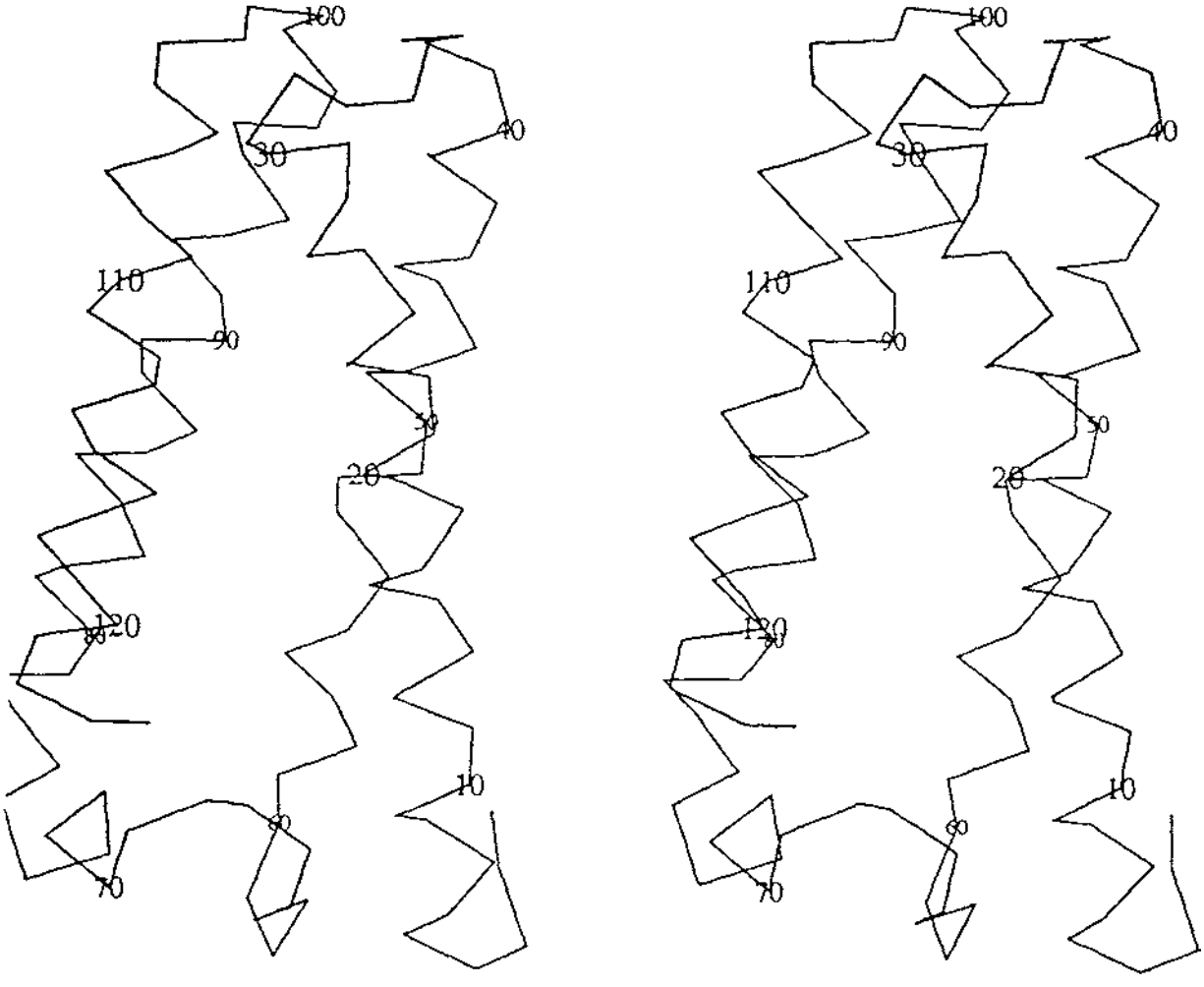


Figure 6.1.2 C_α stereo plot (Kraulis, 1991) of the cytochrome c' monomer.

Table 6.1.1 α-helix assignments.

α-helix assignment	Comments*
Helix A residues 4 - 33	Glu6 and Asp7 provide N-cap hydrogen bonds, kink induced by Pro28, ends with α_{c2} configuration (28O...33N, 29O...32N)
Helix B residues 37 - 54	Asp37 N-cap, bend around residue 45, ends with α_{c1} configuration (50O...53N, 51O...54N)
Helix C residues 75 - 100	Asp75 N-cap, ends with α_{c1} configuration (96O...99N, 97O...100N)
Helix D residues 101 - 124	Asp101 N-cap, bend at ~ residue 117 and ~ residue 120 (H bonds 115-119,118-122 long)

* Terminology α_{c1} and α_{c2} from Baker and Hubbard (1984).

6.2 Hydrogen bonding details.

In the cytochromes c' from *Alcaligenes* sp and *Alcaligenes denitrificans* most of the mainchain carbonyl groups and mainchain amide NH groups are involved in explicit hydrogen bonding. Of those groups which are not (14 % of total) all are exposed to the solvent, and their lack of hydrogen bond partners arises because no waters are modelled in their proximity. Analysis of hydrogen bonding was undertaken using the programs CONTACT and DISTANG (both CCP4 suite programs, see appendix I). Hydrogen atom positions were calculated and input into the coordinate set for all the non-amino nitrogens (amino nitrogens are excluded because of the free rotation about the N-C bond). Hydrogen bonds were assigned using the criteria of Baker and Hubbard (1984). An acceptor-donor pair are considered to be hydrogen bonded if;

1) O...N, O...O or N...N distances, without a calculated hydrogen atom, are less than 3.5 Å, or O...H or N...H, with a calculated hydrogen atom, are less than 2.5 Å

2) the hydrogen bond angle at both donor and acceptor atoms is greater than 90 ° (ie. O(N)...H-N angle > 90 ° or O(N)...H-O angle > 90 °).

6.2.1 Mainchain...mainchain hydrogen bonding.

The majority of hydrogen bonds found in both cytochromes c' are mainchain...mainchain interactions. As noted above the structure of cytochrome c' consists of four α-helices which are connected together by three loops. In an α-helix, the polypeptide chain is coiled into a helix with 3.6 residues per turn, and hydrogen bonds between the C=O groups of residues n and the NH groups of residues n+4. This gives a series of n...n+4 hydrogen bonds which extend along the helix axes.

In each of the cytochrome c' structures the four helices are arranged in antiparallel pairs to give a classic four-helix bundle. Helix A starts at Lys4 and is terminated at Gln33. The N-terminus of this helix contains what is called an N-cap (Baker and Hubbard, 1984; Presta and Rose, 1988, Richardson and Richardson, 1988; Stickle *et al.*, 1992), in which a sidechain is hydrogen bonded to one of the free NH groups in the first turn of the helix. In this case, atom $O_{\varepsilon 1}$ of Glu6 hydrogen bonds to its own mainchain amide and atom $O_{\delta 1}$ on Asp7 forms a hydrogen bond to the mainchain amide group on Lys4. There is a kink in this helix centred around Pro28, such that the regular n...n+4 hydrogen bonding found in helices is disrupted between residues 25 and 30. Table 6.2.1.1 lists the altered hydrogen bond pattern found in the region. The helix is terminated at Gln33 by an αc_2 hydrogen bonding pattern (Baker and Hubbard, 1984) with the carbonyl oxygens of Pro28 and Val29 forming hydrogen bonds to the amide mainchain groups on Gln33 and Gly32 respectively. In this configuration Gly32 is required to have an left-handed α-configuration.

Helix B starts at Asp37 with an N-cap formed by the carboxyl group of Asp37 with the mainchain amide of Gln40. There is a slight bend around Val45 which results in the n...n+4 hydrogen bond between Ala43 and Val47 being longer at 3.33 Å. Helix B ends with an αc_1 conformation (Baker and Hubbard, 1984) with two n...n+3 (O...N) hydrogen bonds from Thr50 and Leu51 to Ala53 and Leu54 respectively. Helix C starts with Asp75 providing an N-cap (involving the mainchain amide of Ser78) and terminates at Gly100 with an αc_1 configuration in which the carbonyl oxygens of Ala96 and Ala97 hydrogen bond to the mainchain amide groups on Ala99 and Gly100 respectively.

Helix D begins at Asp101 with an N-cap formed by this residue with the mainchain amide of Lys104. There are bends occurring approximately at residues Lys117 and His120, as indicated by the long n...n+4 hydrogen bonds formed from Ser115 to Cys119 and Ala118 to Ala122 (bond lengths of 3.38 Å and 3.57 Å respectively). This bending may be associated with haem binding since the two Cysteine-thioether linkages are at 116 and 119, and the His ligand at 120.

All three types of N-cap (Baker and Hubbard, 1984; Richardson and Richardson, 1988) occur in these cytochromes c'. The Asp(7)...NH(4) N-cap in helix A is an example of an n - NH(n-3) interaction, while the Glu6...NH(6) is an example of an n-NH(n) interaction. In the other three helices Asp37...NH(40), Asp75...NH(78) and Asp101...NH(104) give examples of n - NH(n+3) interactions. Baker and Hubbard (1984) reported that these hydrogen bonding motifs are quite common in globular proteins (47 helices out of a total of 71 in their sample). They also noted that the preponderance of acidic side-chains (confirmed in these structures) may suggest that the positive charge of the helix dipole (Hol *et al.*, 1978; Hol, 1985) has some influence. An example of an N-cap at the start of a helix (in this case helix B) is given in Figure 6.2.2.1.

The hydrogen bond lengths are remarkably constant along each of the four α-helices apart from at the points where kinks occur. These bond lengths are shown diagrammatically in Figures 6.2.1.1 to 6.2.1.4. The mean values and their standard deviations are further evidence; omitting the C=O groups of 22 - 25, 43, 115 and 118, the values for the C=O(n)...HN(n+4) hydrogen bonds for the four helices are 2.95 ± 0.10, 2.98 ± 0.16, 2.96 ± 0.12 and 2.98 ± 0.15 Å respectively. Figures 6.2.1.5 and 6.2.1.6 show stereoviews of helix hydrogen bonding.

Two pieces of 3_{10} helix occur, at either end of the BC loop. These involve residues 55 - 60 and 69 - 74, each with three C=O(n)...HN(n+3) hydrogen bonds. The first 3_{10} helix follows on from helix B, and is at an slight angle of 9 ° to it. The second comes just before helix C and is at an angle of 110 ° to it.

Table 6.2.1.1 Hydrogen bonding in kink found in helix A.

Residue	Alternate mainchain hydrogen bonding	Bond distance (Å)
Ser21	O...N Arg25	2.97
	O...N_{ε} Arg25	3.06
His22	O...$N_{\delta 2}$ Asn44	2.87
Phe23	O...N Met26	3.04
Gly24	O...N Thr27	3.21
	O...$O_{\gamma 1}$ Thr27	3.13
Met26	O...N Val29	3.03
	O...N Val30	3.00

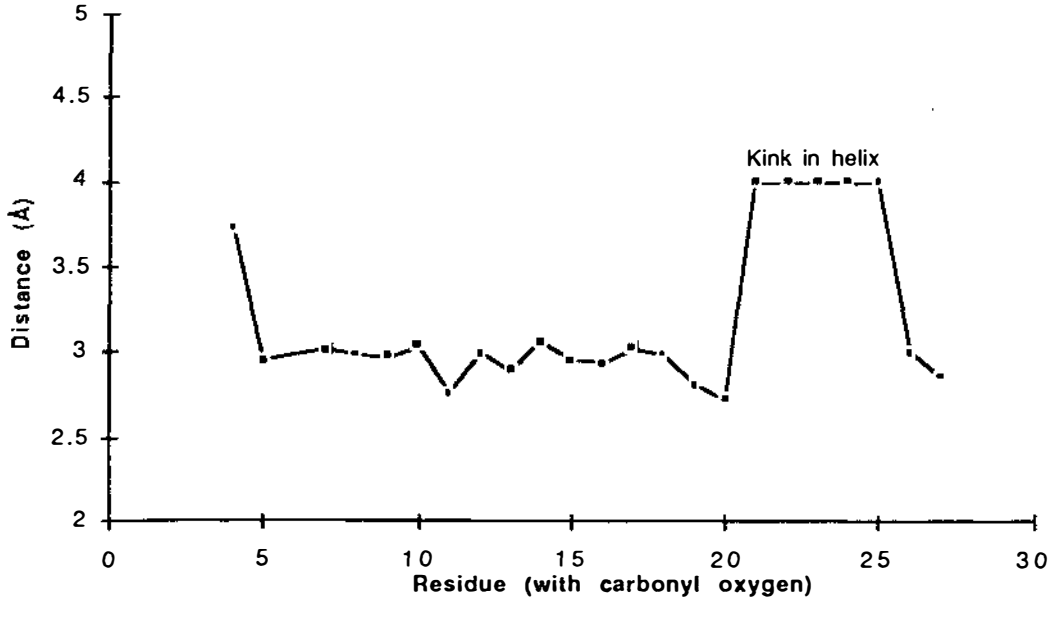


Figure 6.2.1.1 Plot of n...n+4 hydrogen bond length in helix A.

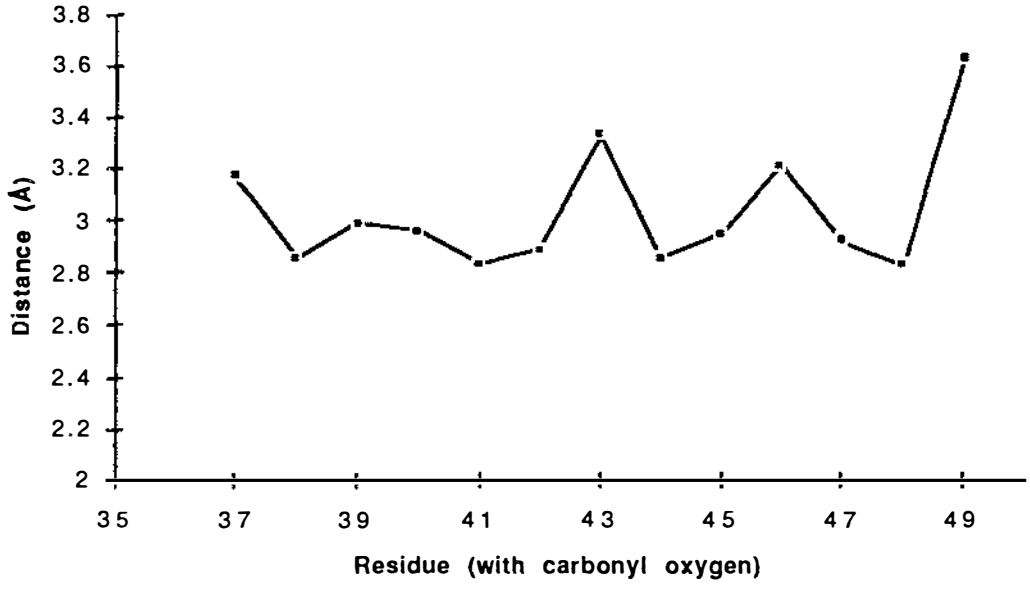


Figure 6.2.1.2 Plot of n...n+4 hydrogen bond length in helix B.

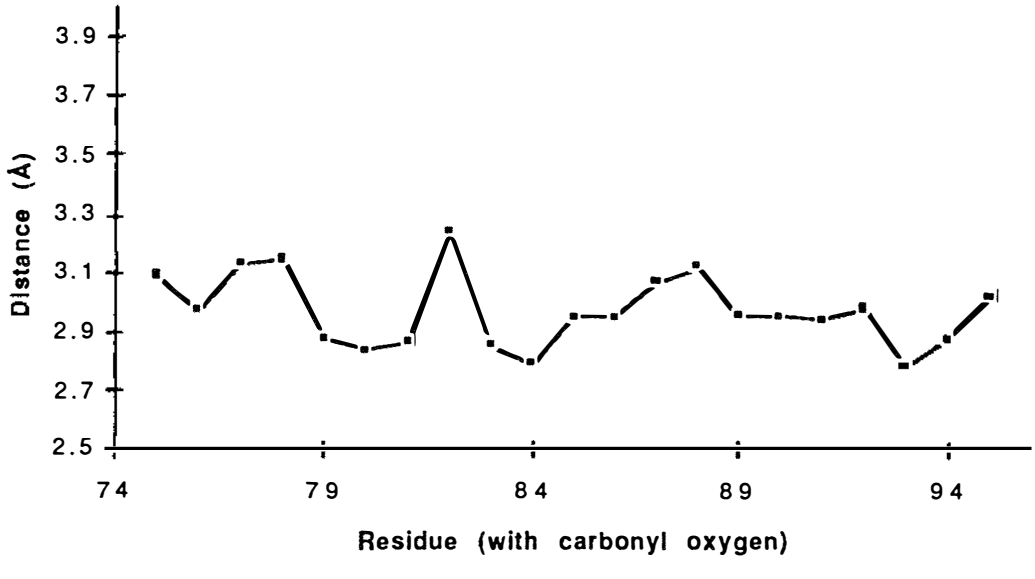


Figure 6.2.1.3 Plot of n...n+4 hydrogen bond length in helix C.

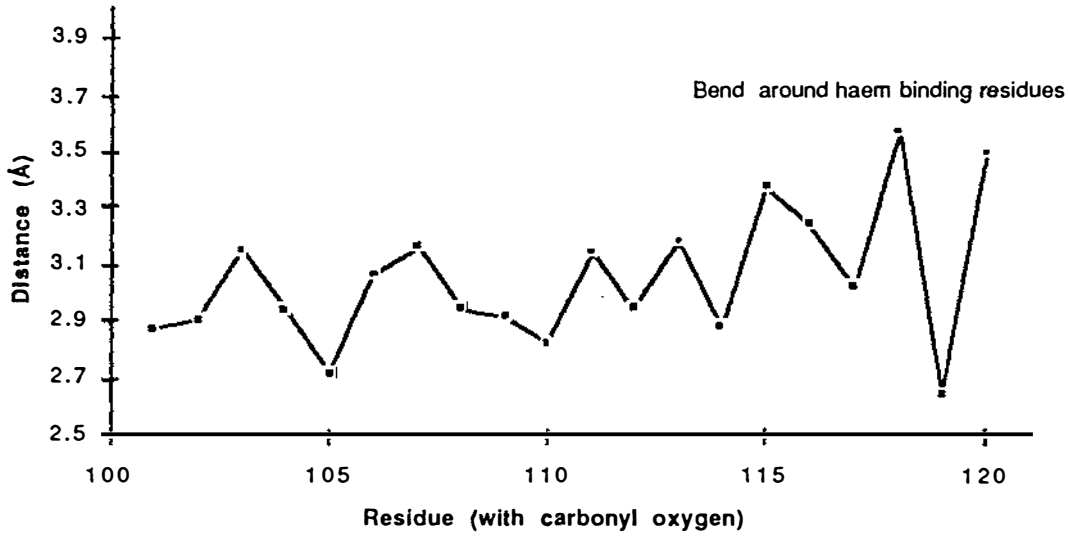


Figure 6.2.1.4 Plot of n...n+4 hydrogen bond length in helix D.

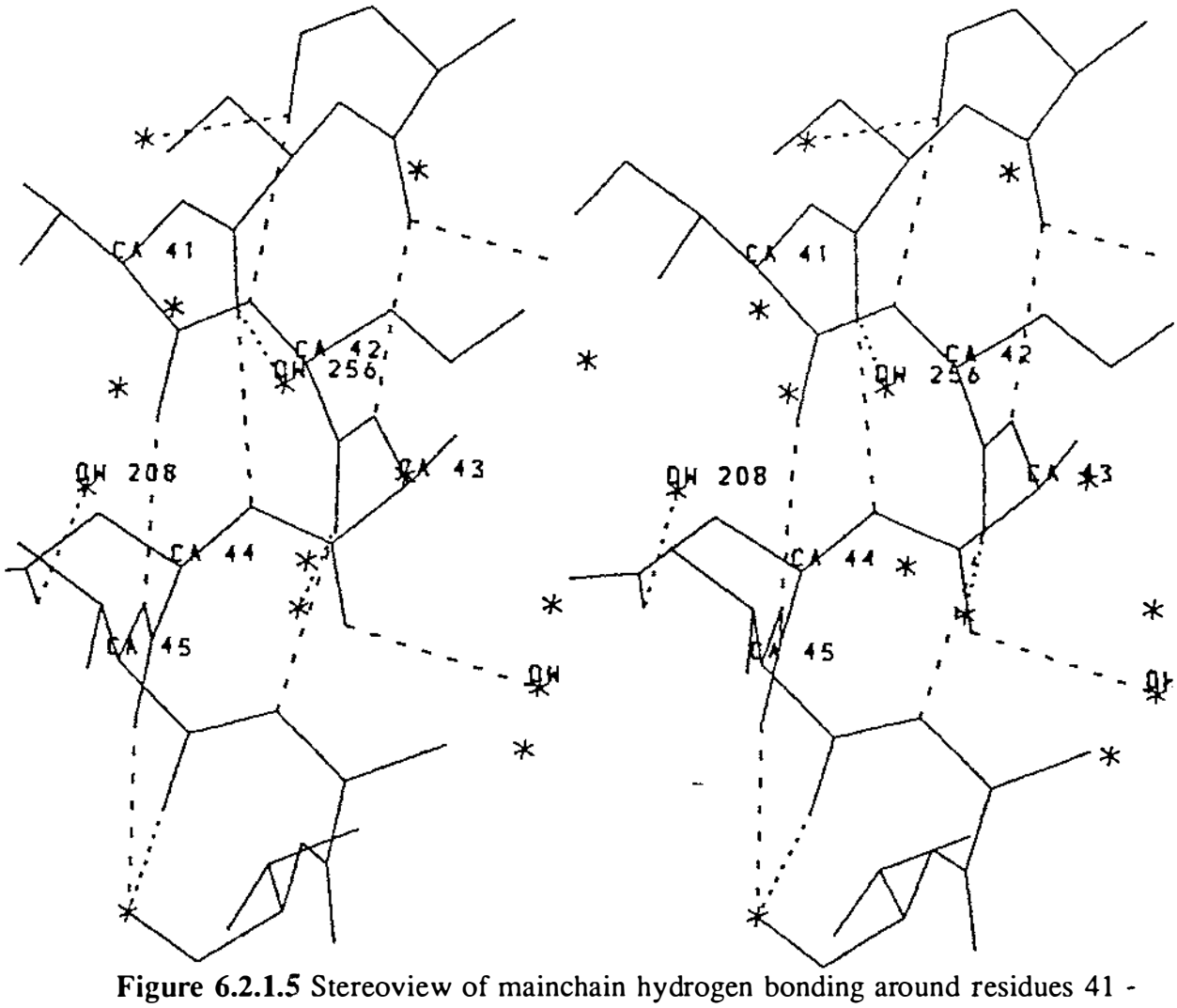


Figure 6.2.1.5 Stereoview of mainchain hydrogen bonding around residues 41 - 45 showing 1-5 helix hydrogen bonding and mainchain...solvent interactions.

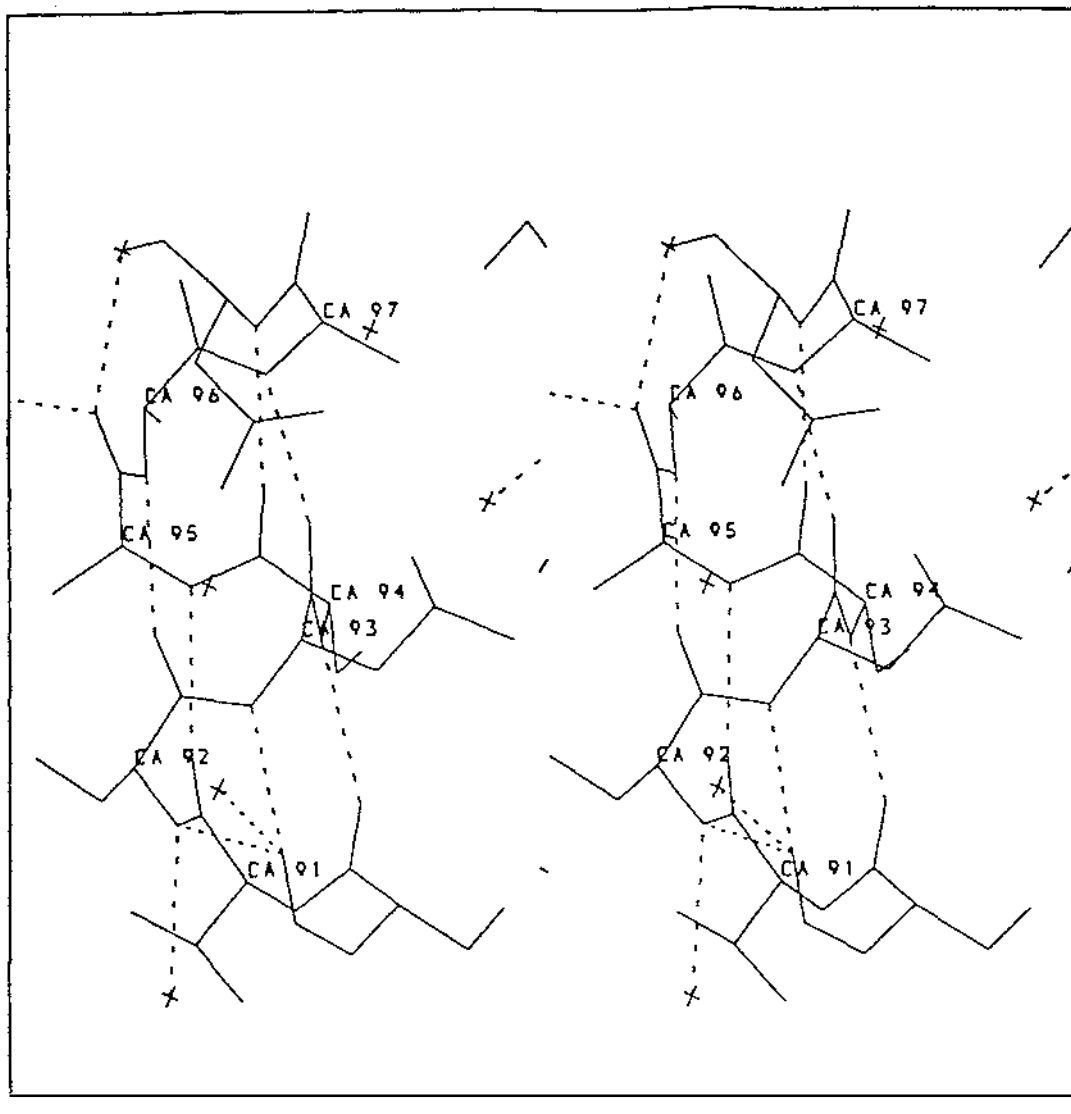


Figure 6.2.1.6 Stereoview of n...n+4 mainchain hydrogen bonds in helix C (residues 91 - 97).

6.2.2 Mainchain...sidechain hydrogen bonding.

There are 21 mainchain...sidechain hydrogen bonds in the cytochrome c' from *Alcaligenes* sp and 22 in the *Alcaligenes denitrificans* cytochrome c' structure. They can be placed into two categories; hydrogen bonds which are formed by residues sequentially no more than four residues apart, are classed as "local" interactions and those which occur between residues which are far apart in the sequence, and are classified as "cross-linking" interactions (Baker and Hubbard, 1984). Tables 6.2.2.1 and 6.2.2.2 give the mainchain...sidechain hydrogen bonds in the cytochromes c' from *Alcaligenes* sp and *Alcaligenes denitrificans.*

Approximately three-quarters (17) of the mainchain...sidechain hydrogen bonds in *Alcaligenes sp* are "local" interactions and most of these occur in the α-

helices. The number of "local" interactions is slightly different in the cytochrome c' from *Alcaligenes denitrificans* where only 13 of the 21 mainchain...sidechain hydrogen bonds are "local". This reduction in the number of "local" interactions is in part as a result of the sequence changes made in the *Alcaligenes denitrificans* cytochrome c' (see section 5.4.2), since two residues forming "local" interactions in the *Alcaligenes* sp cytochrome c' (Thr27 and Ser78) have both been replaced with glycine. The most common among the mainchain...sidechain hydrogen bonds are the n...(n-4) interactions where a sidechain interacts with the main chain carbonyl four residues earlier in the polypeptide chain (Baker and Hubbard, 1984); this almost always appears to be associated with helical regions in proteins. Many of these probably result from the favourable conformation obtained by Ser and Thr residues with χ_1 values of -60 °; such a configuration places the sidechain hydroxyl in a favourable position where it can readily hydrogen bond to a C=O group from the previous turn of the helix.

Table 6.2.2.1 Mainchain...sidechain hydrogen bonds in cytochrome c' from *Alcaligenes* sp, divided into "local" and "cross-linking" interactions. (Hydrogen bonding which is conserved in the cytochrome c' from *Alcaligenes denitrificans* is in bold)

"Local"	"Cross-linking"
$O_{\varepsilon 1}$ Glu6...N Glu6	**$N_{\eta 1}$ Arg12...O Ala58**
$O_{\partial 1}$ Asp7...N lys4	**$N_{\eta 1}$ Arg12...O Thr63**
O_{γ} Ser14...O Lys10	**$N_{\eta 2}$ Arg12...O Thr63**
$O_{\gamma 1}$ Thr17...O Gln13	**$N_{\eta 2}$ Arg12...O Gly65**
N_{ε} Arg25...O Ser21	**$N_{\delta 2}$ Asn44...O His22**
$O_{\gamma 1}$ Thr27...O Gly24	**$N_{\varepsilon 1}$ Trp73...O Gly60**
$O_{\delta 2}$ Asp37...N Gln40	
$O_{\gamma 1}$ Thr50...O Glu46	
$O_{\gamma 1}$ Thr52...O Leu48	
$O_{\gamma 1}$ Thr52...O Lys49	
$N_{\varepsilon 1}$ Trp56...O Thr52	
$O_{\gamma 1}$ Thr63...O Gly60	
$O_{\varepsilon 2}$ Glu 71...N Glu71	
O_{γ} Ser74...O Pro70	
$O_{\delta 1}$ Asp75...N Ser 78	
$N_{\delta 2}$ Asn89...O Ala85	
$O_{\delta 2}$ Asp101...N Lys104	

These interactions (see Tables 6.2.2.1 and 6.2.2.2) are predominantly formed by threonine residues, but serine, arginine and asparagine are also present in n...(n-4) interactions (see Figure 6.2.2.2). The N-cap stabilisation (Baker and Hubbard, 1984; Presta and Rose, 1988; Richardson and Richardson, 1988) of the helices provides other local hydrogen bonds, which have been described in detail in section 6.2.1 and example of this is shown in Figure 6.2.2.1.

Table 6.2.2.2 Mainchain...sidechain hydrogen bonds in cytochrome c' from *Alcaligenes denitrificans* divided into "local" and "cross-linking" interactions. (Hydrogen bonding which is conserved in the cytochrome c' from *Alcaligenes* sp is in bold)

"Local"	"Cross-linking"
$O_{\varepsilon 1}$ Glu6...N Glu6	**$N_{\eta 1}$ Arg12...O Ala58**
$O_{\partial 1}$ Asp7...N lys4	**$N_{\eta 1}$ Arg12...O Thr63**
O_{γ} Ser14...O Lys10	**$N_{\eta 2}$ Arg12...O Thr63**
$O_{\gamma 1}$ Thr17...O Gln13	**$N_{\eta 2}$ Arg12...O Gly65**
N_{ε} Arg25...O Ser21	**$N_{\delta 2}$ Asn44...O His22**
$O_{\delta 2}$ Asp37...N Gln40	**$N_{\varepsilon 1}$ Trp73...O Phe59**
$O_{\gamma 1}$ Thr50...O Glu46	**$N_{\varepsilon 1}$ Trp73...O Gly60**
$O_{\gamma 1}$ Thr52...O Leu48	N_{ζ} Lys125...O Ala68
$O_{\gamma 1}$ Thr52...O Lys49	
$O_{\gamma 1}$ Thr63...O Gly60	
$O_{\varepsilon 2}$ Glu64...N Gly62	
$O_{\varepsilon 2}$ Glu71...N Glu71	
O_{γ} Ser74...O Pro70	
$O_{\delta 1}$ Asp75...N Ser 78	
$N_{\delta 2}$ Asn89...O Ala85	
$O_{\delta 2}$ Asp101...N Lys104	

"Cross-linking" hydrogen bonds, which form between residues which are sequentially far apart, play an important role in drawing together different parts of the structure and stabilising the final folding pattern of a protein (Alber *et al.*, 1987). There are very few "cross-linking" hydrogen bonds in either species. Only three residues seem to be involved. Arg12 (which is in helix A and is conserved in all members of the cytochrome c' family) makes a number of strong hydrogen bonds with mainchain C=O groups of the BC loop. Asn44 provides one "cross-linking" interaction from helix B to helix A, but this residue is not conserved in other species. Trp73, just prior to helix C, hydrogen bonds to a region of the BC loop that is

strongly conserved in structure in the other cytochrome c' structures; this residue is also conserved and the hydrogen bond is probably important.

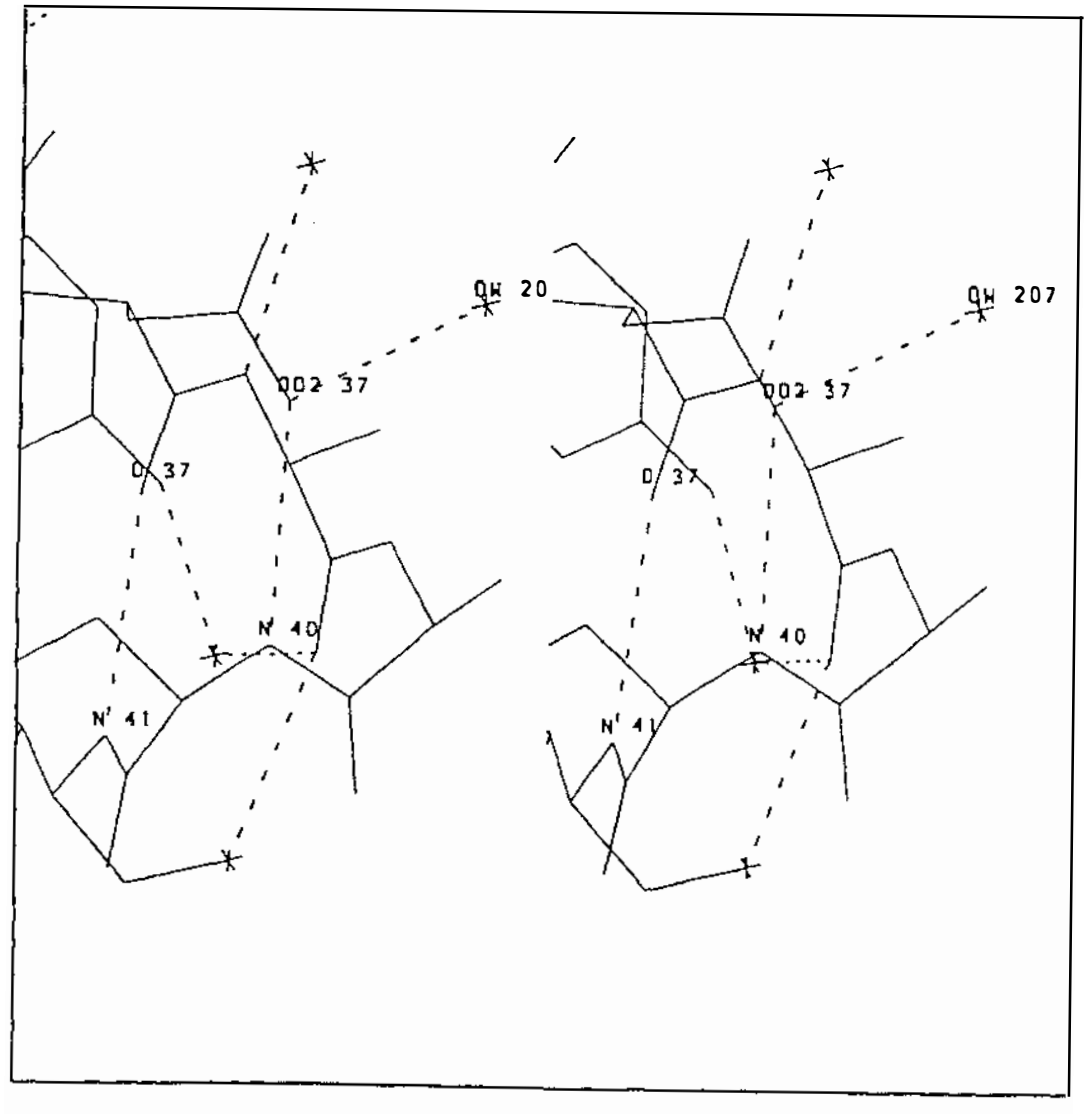


Figure 6.2.2.1 Stereodiagram of N-cap at beginning of helix B (Asp37 providing sidechain...mainchain hydrogen bond).

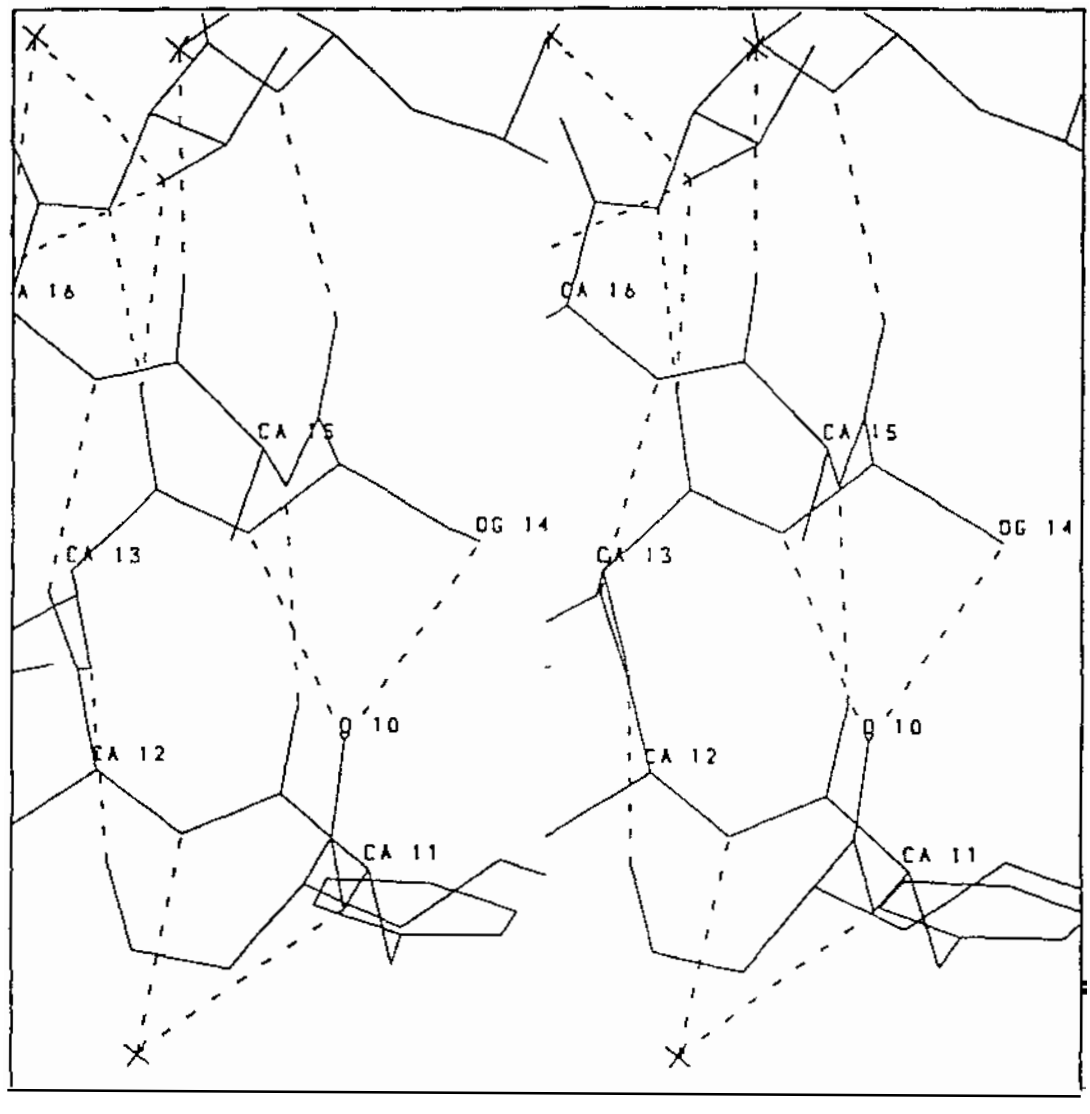


Figure 6.2.2.2 Stereodiagram showing backbonding interaction between sidechain of Ser14 and the carbonyl oxygen of Lys10. A similar interaction between Thr17 and C=O(13) can be seen at top of picture.

6.2.3 Sidechain...sidechain hydrogen bonding.

There are only eight sidechain...sidechain hydrogen bonds in the cytochrome c' from both *Alcaligenes* sp and *Alcaligenes denitrificans*. These are listed in Table 6.2.3.1. Not included is the hydrogen bonding of the haem propionates (which will be discussed later).

Although there are only eight sidechain...sidechain hydrogen bonds, several of these may be important in stabilising the protein. They are the "cross-linking" hydrogen bonds from residues which are not close in the sequence. For instance $N_{\delta 2}$ Asn44...$N_{\delta 1}$ His22 forms a hydrogen bond between helix A and B, and N_{ζ} Lys92...$O_{\delta 1}$ Asp111 links helices C and D. Figure 6.2.3.1 gives an example of a sidechain...sidechain "local" interaction occurring between Gln83 and Glu87.

Table 6.2.3.1 Sidechain...sidechain hydrogen bonds.

Alcaligenes sp/ *Alcaligenes denitrificans*
O_{γ} Ser21 - $N_{\eta 2}$ Arg25
$N_{\delta 1}$ Asn44 - $N_{\delta 1}$ His22
$O_{\gamma 1}$ Thr63 - $N_{\eta 1}$ Arg12
$O_{\delta 2}$ Asp75 - O_{γ} Ser78
$O_{\varepsilon 1}$ Gln83 - $N_{\varepsilon 2}$ Glu87
$O_{\varepsilon 1}$ Gln83 - O_{γ} Ser115
$O_{\delta 1}$ Asn89 - N_{ζ} Lys92
$N_{\delta 2}$ Gln89 - O_{γ} Ser115
$O_{\delta 1}$ Asp111 - N_{ζ} Lys92

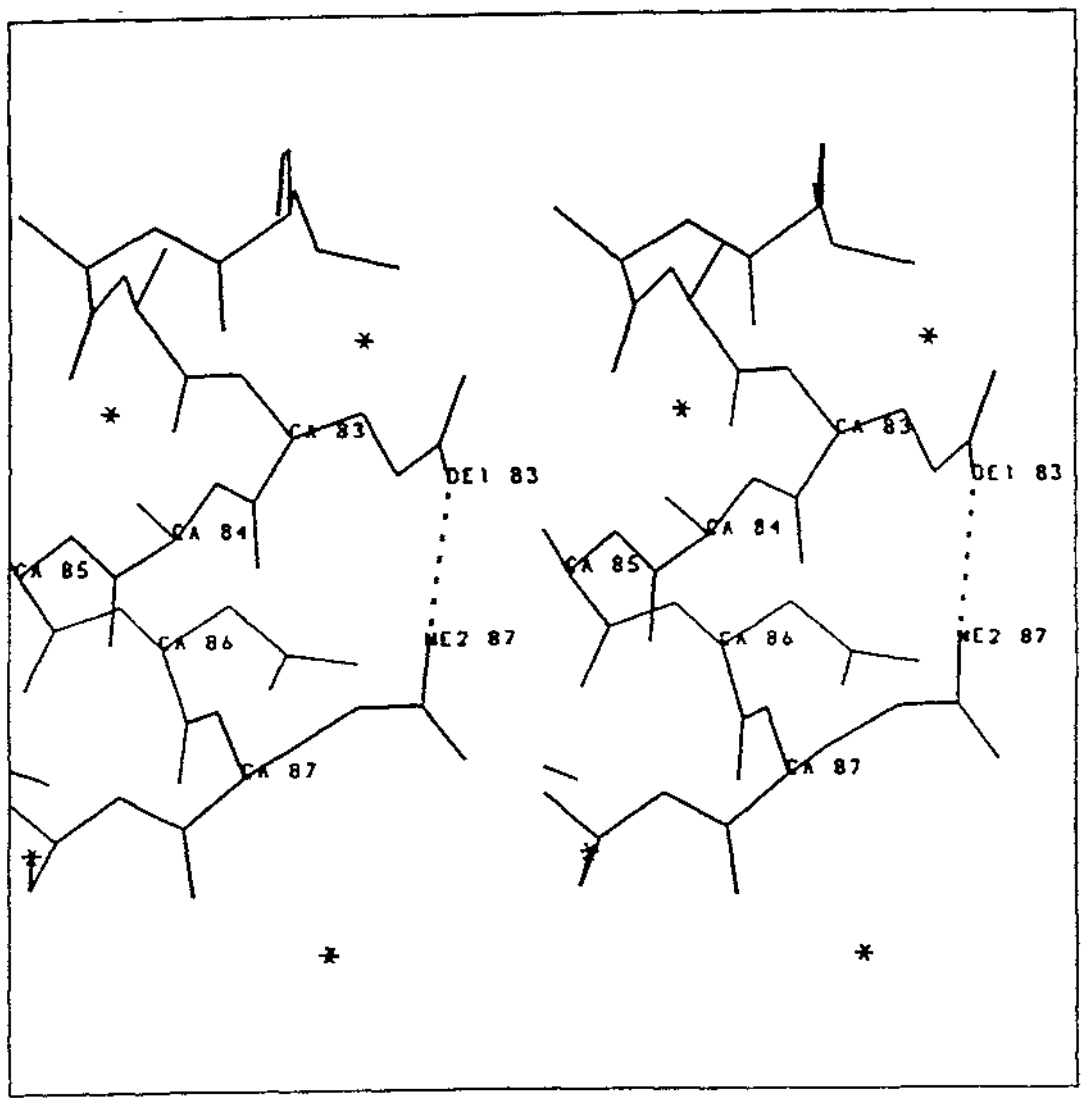
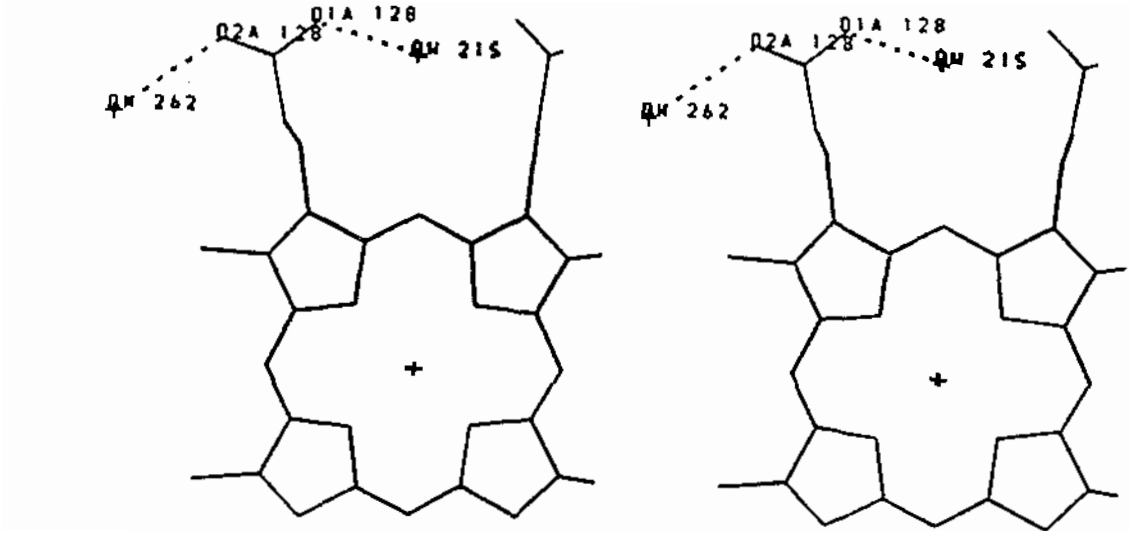


Figure 6.2.3.1 Stereoview of "local" sidechain...sidechain interaction between Gln83 and Gln87.

6.2.4 Haem hydrogen bonding.

A series of conserved hydrogen bonds to the haem propionates, involving Arg12, Gln13 and Thr63, is a feature of the cytochromes c' from *Alcaligenes* sp and *Alcaligenes denitrificans*. Of these residues, Arg12 is conserved in all species, while Gln13 and Thr63 are conserved in the majority of cases (Ambler, 1981). N_{ε} and $N_{\eta 2}$ from Arg12 form hydrogen bonds with the propionate on ring D (hydrogen bonding of Arg12 is shown in Figure 6.2.4.3). The sidechain amide nitrogen of Gln13 also hydrogen bonds to the propionate on ring D, and Thr63 forms a sidechain...sidechain hydrogen bond to Arg12 to provide extra stability. The propionates on rings A and D also hydrogen bond to solvent molecules. One water molecule (262 in *Alcaligenes* sp numbering, 241 in *Alcaligenes denitrificans* numbering) hydrogen bonds to O2A, while a second (215 in *Alcaligenes* sp, 214 in *Alcaligenes denitrificans*) bridges between the two propionic acid sidechains. Other hydrogen bonds are formed between the NH group of Asp67 and O2D; the latter hydrogen bond probably plays a role in stabilising the BC loop. These hydrogen bonding interactions are the same for both cytochrome c' structures. Table 6.2.4.1 lists the hydrogen bonds to the haem group for both species. While in Figures 6.2.4.1 and 6.2.4.2 the hydrogen bonding of propionate carboxylate sidechains are shown.



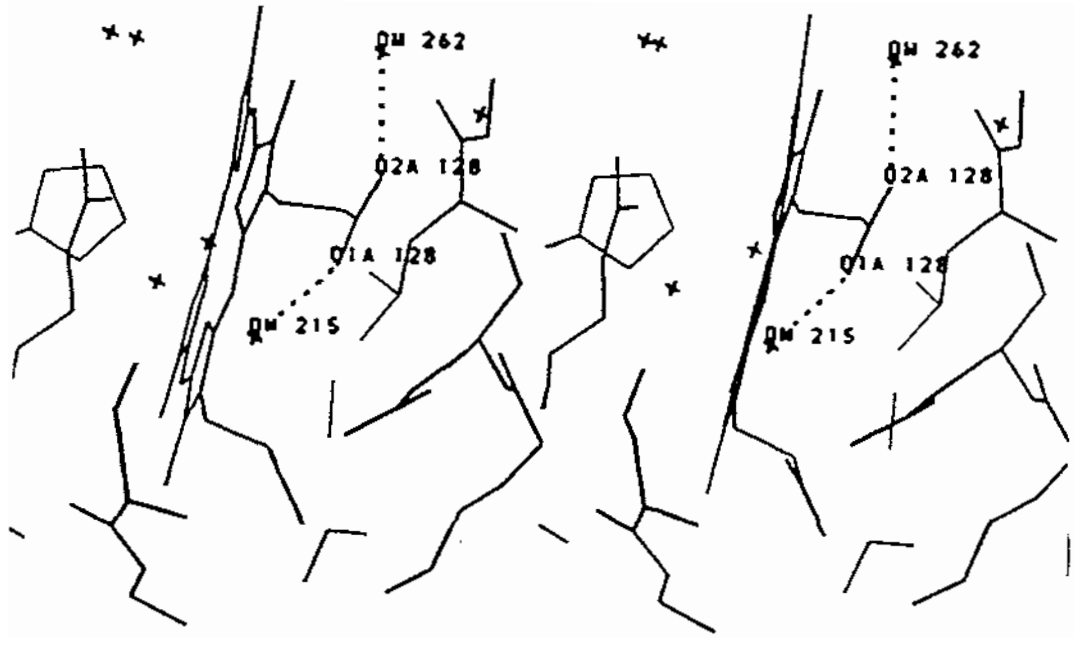


Figure 6.2.4.1 Stereodiagrams showing two views of the hydrogen bonding involving the exposed propionate sidechain from ring A.

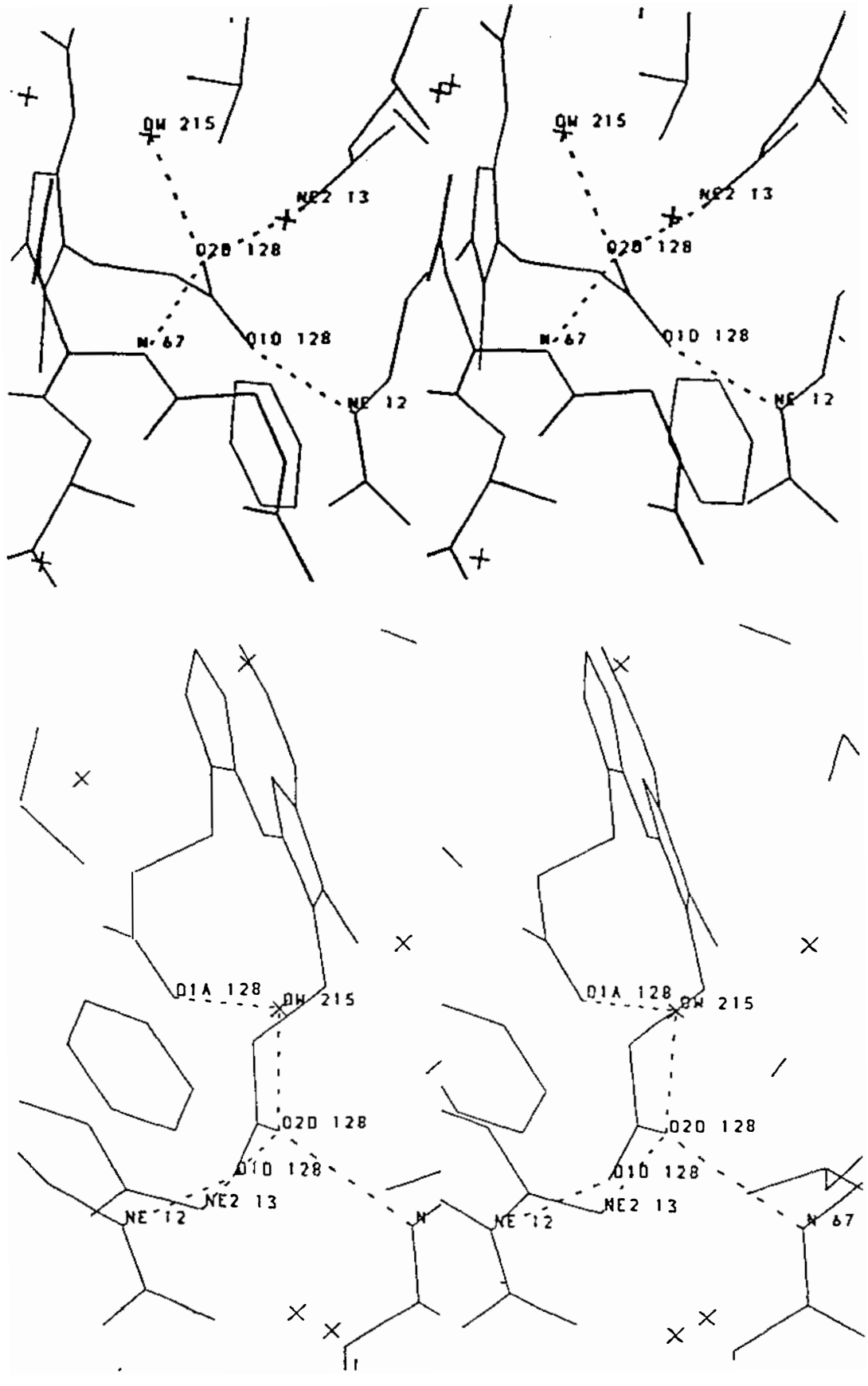


Figure 6.2.4.2 Stereodiagrams showing two views of the hydrogen bonding involving the buried propionate sidechain from ring D.

Table 6.2.4.1 Protein...haem and solvent...haem interactions in *Alcaligenes* sp and *Alcaligenes denitrificans* cytochromes c'.

Interaction	Distance (Å)	Angle (°)[a]
Alcaligenes sp		
O1D...$N_{\eta 2}$ Arg12	3.32	135.8
O1D...N_{ε} Arg12	2.75	169.0
O2D...$N_{\varepsilon 2}$ Gln13	2.72	164.5
O2D..N Asp67	2.78	160.9
O1A...HOH215	2.72	
O2D...HOH215	2.61	
O2A...HOH262	3.00	
Alcaligenes denitrificans		
O1D...$N_{\eta 2}$ Arg12	3.33	135.0
O1D...N_{ε} Arg12	2.82	168.0
O2D...$N_{\varepsilon 2}$ Gln13	2.72	171.3
O2D...N Asp67	2.86	159.5
O1A...HOH214	2.69	
O2D...HOH214	2.63	
O2A...HOH241	3.00	

a - If the donor atom is a nitrogen whose hydrogen atom position is unambiguous, the angle given is that subtended at the calculated hydrogen atom position. If the donor is an oxygen, the angle is: acceptor...oxygen -- carbon atom.

As can be seen from Table 6.2.4.1, the hydrogen bonding is exactly conserved between the two *Alcaligenes* species providing further indication of their structural similarity. Figure 6.2.4.3 shows the hydrogen bonding of the conserved Arg12 residue.

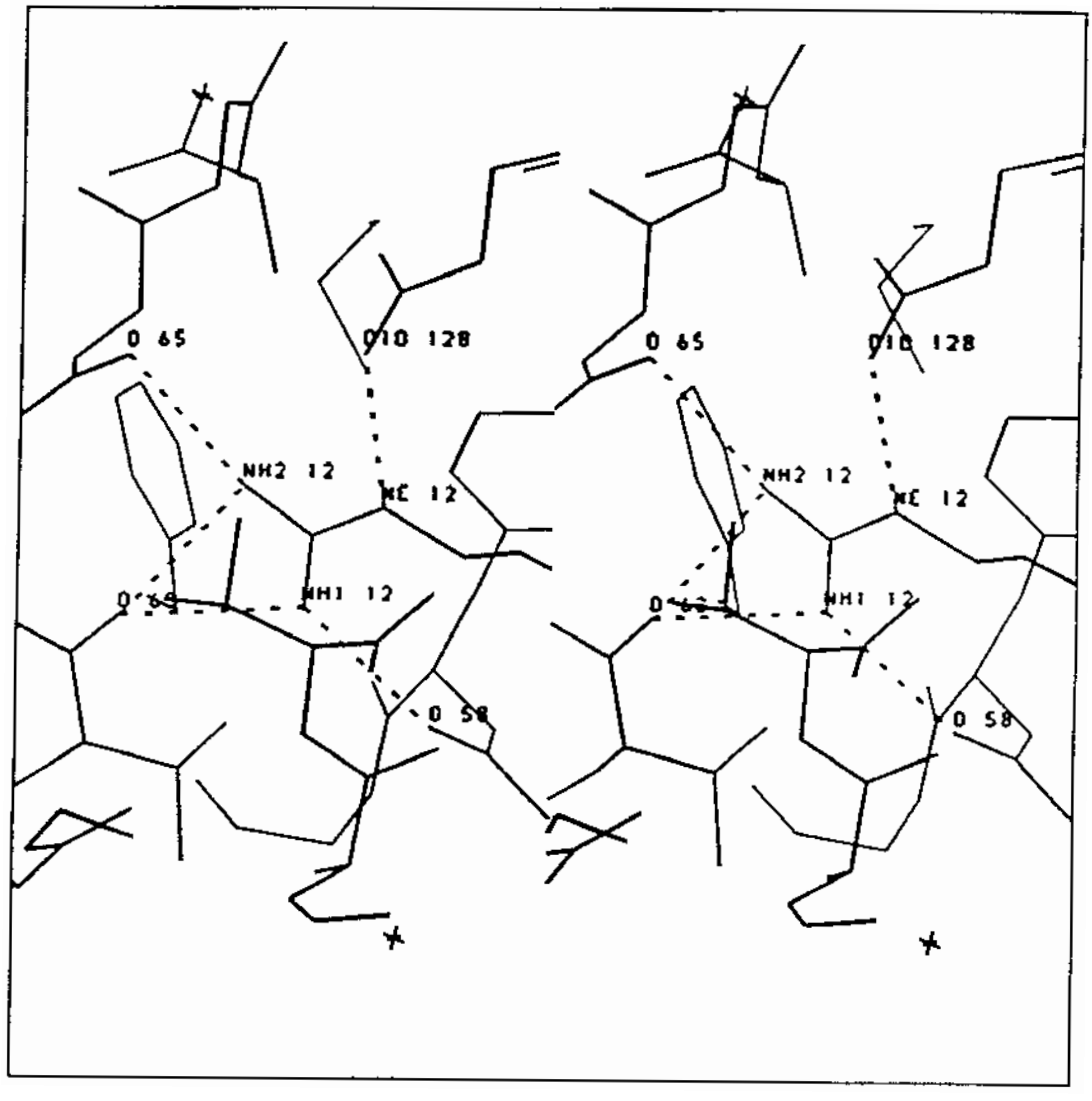


Figure 6.2.4.3 Stereodiagram showing hydrogen bonding of the conserved Arg12 residue (ASCC numbering).

6.3 Loops and turns.

Apart from the four helices, and the two 3_{10}-helices in the BC loop, there is very little other secondary structure. Only two turns are present in the structure, a type II turn (Lewis *et al.*, 1973) at residues 60 - 63 and a type II' turn at residues 65 - 68. Both are in the long BC loop, and both appear to be important in contributing to the haem stabilisation. The first turn occurs just after the first piece of 3_{10}-helix and contains the highly conserved (Ambler, 1981) threonine at position 63 (*Alcaligenes* sp numbering). The hydroxyl of this residue hydrogen bonds back to the mainchain carbonyl of Gly60 thus stabilising the turn, and also receives another hydrogen bond from the invariant Arg12 (which in turn hydrogen bonds to one of the haem propionates). The second turn occurs before the second piece of 3_{10}-helix; in this turn the NH of the peptide linking residues 2 and 3 of the turn (residue numbers 66 and 67) hydrogen bonds to the haem propionate; this is probably the reason why this turn has the rather uncommon II' configuration. Table 6.3.1 lists the turns present in the cytochromes c' from *Alcaligenes* sp and *Alcaligenes denitrificans*. Figures 6.3.1 and 6.3.2 show stereoviews of the hydrogen bonding stabilising the turns. Note in the second turn (residues 65 - 68) the extensive interactions of this part of the BC loop with haem (67N), Arg12 (65O) and the C-terminus of helix D (67O and 69N) plus the bridging water (OW213).

Table 6.3.1 Turns present in the BC loop.

Residues	$\phi_{i+1}(^\circ)$ $\psi_{i+1}(^\circ)$[a]	$\phi_{i+2}(^\circ)$ $\psi_{i+2}(^\circ)$[a]	assignment
60 - 63	-50 132	76 16	Type II
65 - 68	63 -145	-97 16	Type II'

a - torsion angles refer to the cytochrome c' from *Alcaligenes* sp (similar conformational angles are present in the *Alcaligenes denitrificans* structure).

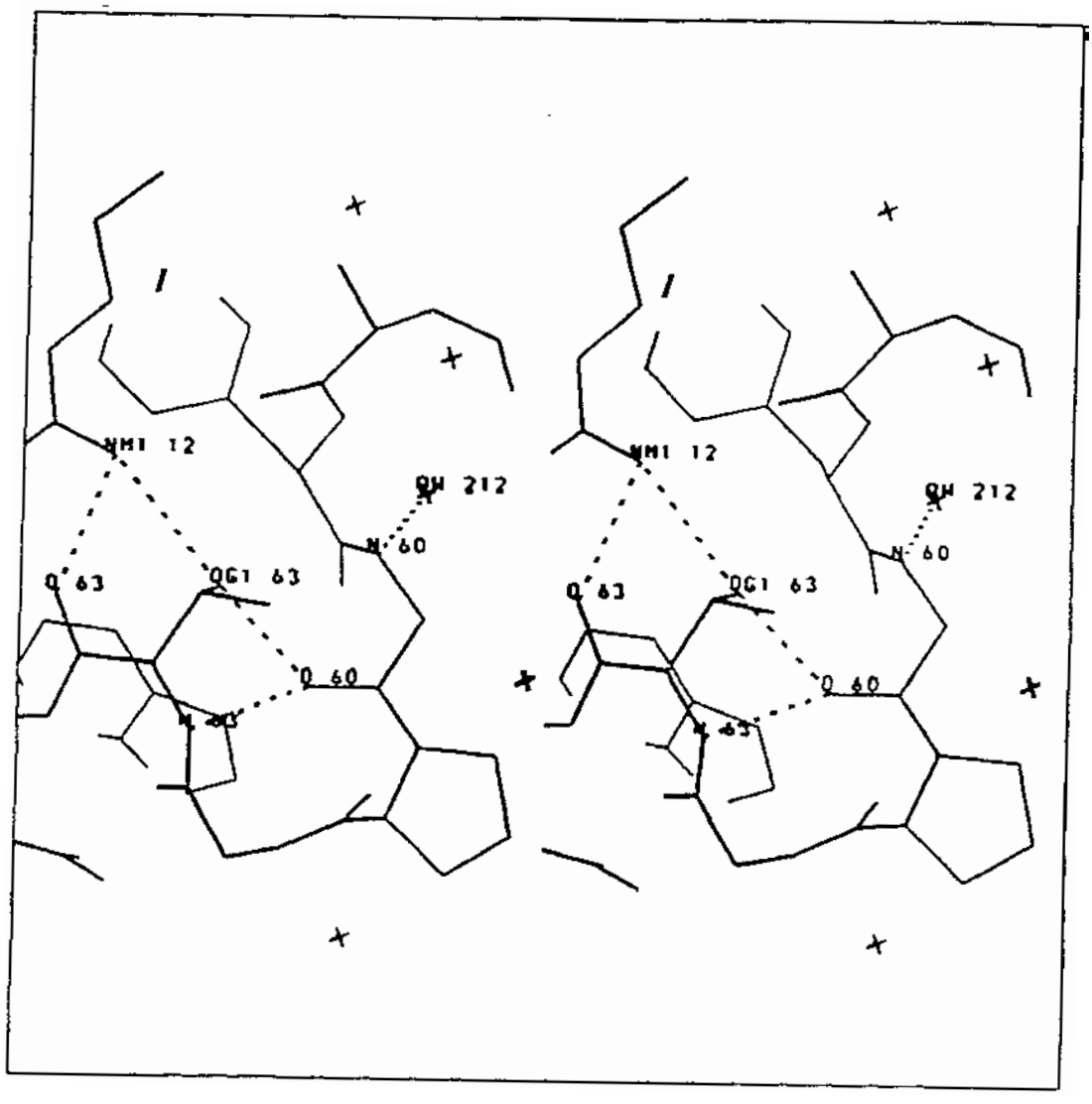


Figure 6.3.1 Stereoview of the type II turn involving residues 60-63 showing hydrogen bonds.

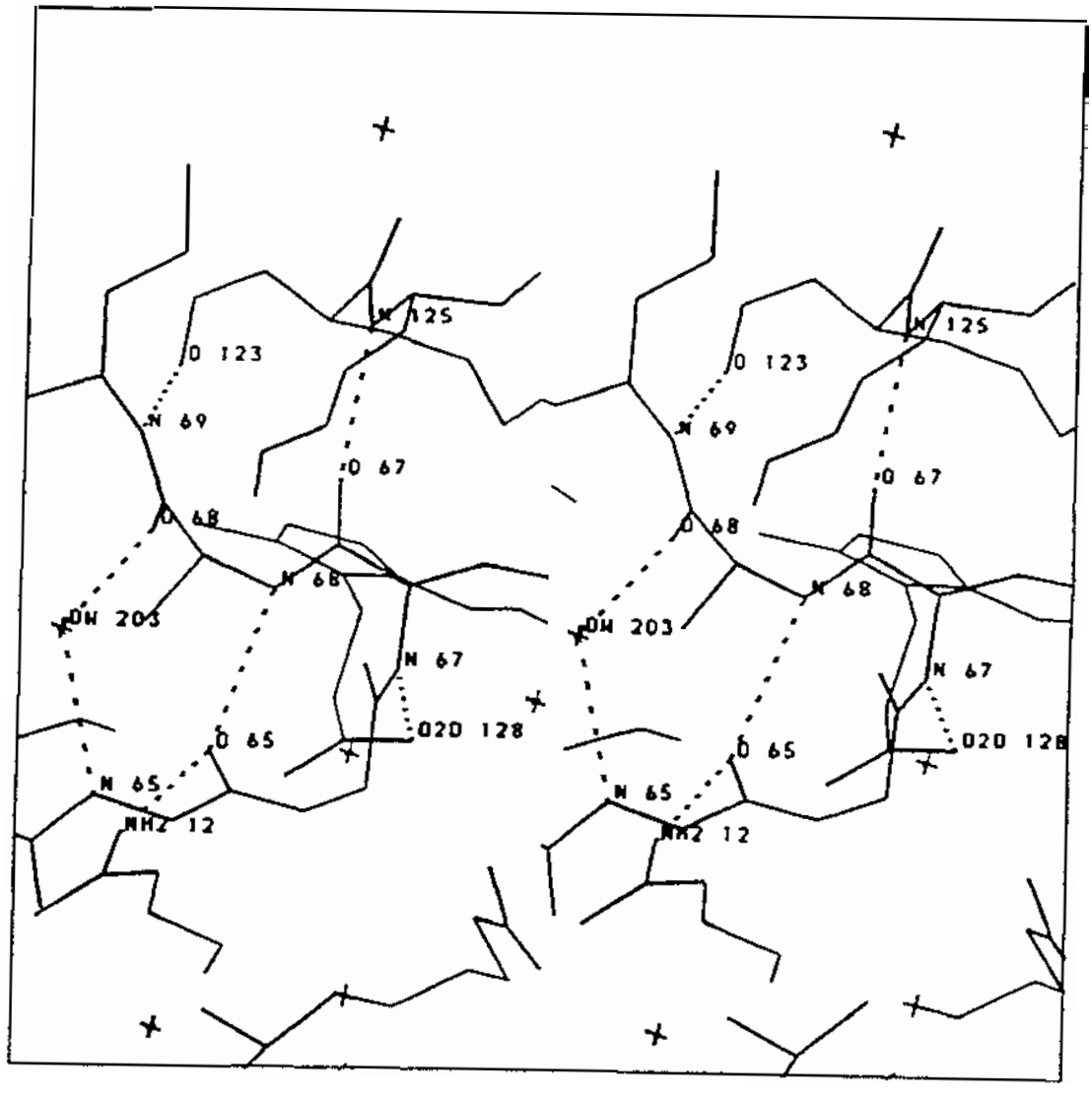


Figure 6.3.2 Stereoview of type II' turn involving residues 65-68 showing the ntework of various stabilising hydrogen bonds.

6.4 Modified N-terminus.

The N-terminal residue of the cytochrome c' from both species is post-translationally modified. The residue has been shown chemically to be pyroglutamic acid (pyrrolidone carboxylic acid), in which the sidechain carboxyl group is covalently linked to the N-terminal α-amino group to give a cyclic structure. The electron density for this group is very well defined in both cytochromes c' so there is no doubt about the assignment. Figure 6.4.1 shows electron density around the pyroglutamic acid N-terminus from the *Alcaligenes* sp cytochrome c'. The pyroglutamic acid residue is situated in a pocket, between residues Arg25 and Gly32 of the symmetry related second molecule of the dimer molecule. The ring of the pyroglutamic acid also packs against the face of Phe2, approximately parallel to it (Hunter *et al.*, 1991), and it is stabilised by the formation of two hydrogen bonds to solvent molecules. Further stabilisation is added by the packing against the C-terminal end of helix A of the symmetry related dimer molecule, with intermolecular contacts between the NH group of Phe2 and O Gly32 on another symmetry related molecule (see Table 6.10.3).

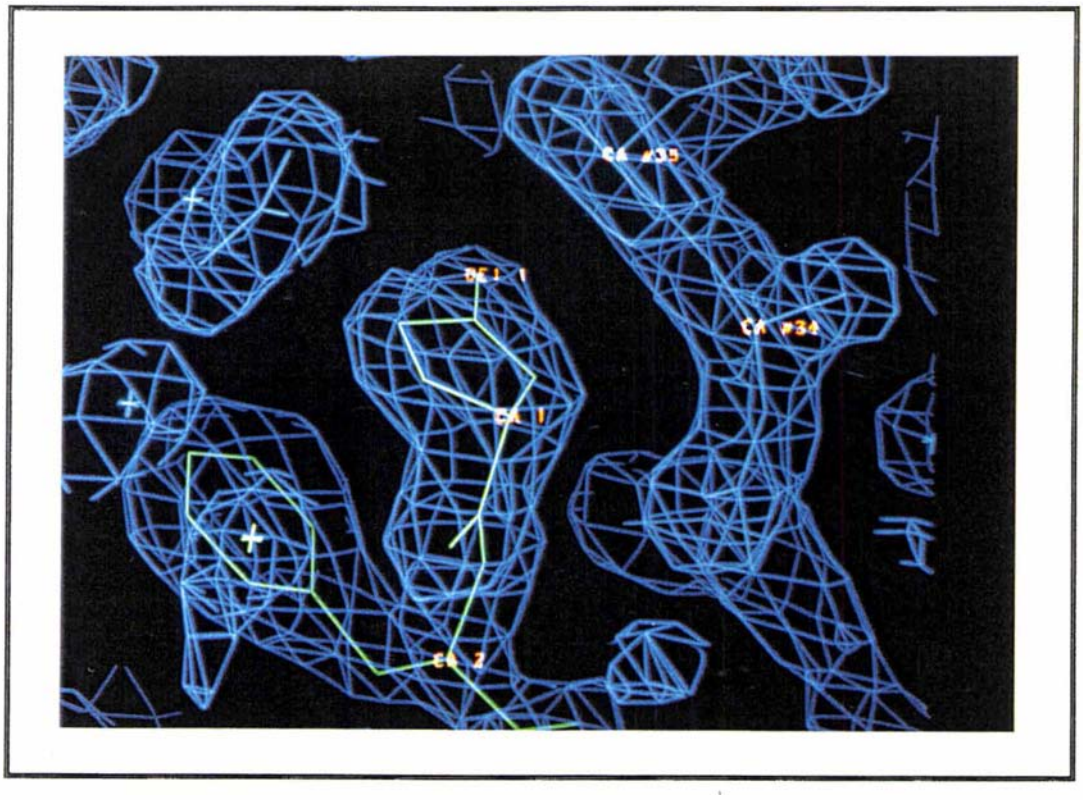


Figure 6.4.1 Electron density from the final 2Fo-Fc map of the modified N-terminus in the cytochrome c' from *Alcaligenes* sp.

6.5 Analysis of the haem geometry.

The haem is located near the carboxy terminus of Helix D, where it sits in a pocket at the splayed open end of the four α-helix bundle (same end as the BC-loop). As in c-type cytochromes the protoporphyrin IX prosthetic group is attached to Helix D via two thioether linkages formed through the condensation of the haem vinyl groups and the cysteine side-chains of residues 116 and 119 .

The iron is penta-coordinate, being covalently bound by the four in-plane pyrrole nitrogens contributed by the four pyrrole rings, and by the single axial ligand provided by $N_{\varepsilon 2}$ of the imidazole side-chain of His120. The sixth ligand position opposite the histidine, is empty. Leu16 from helix A approaches the face of the haem in place of the sixth ligand and preventing a water molecule from binding (distance of the closest contact is between the Leu sidechain and the haem ligand is 3.74 Å). Figure 6.5.1 shows a protoporphyrin IX prosthetic group with labelling of the atoms and Figure 6.5.2 shows the coordination of the iron atom. In Table 6.5.1 the details of the haem geometry in the two structures are provided.

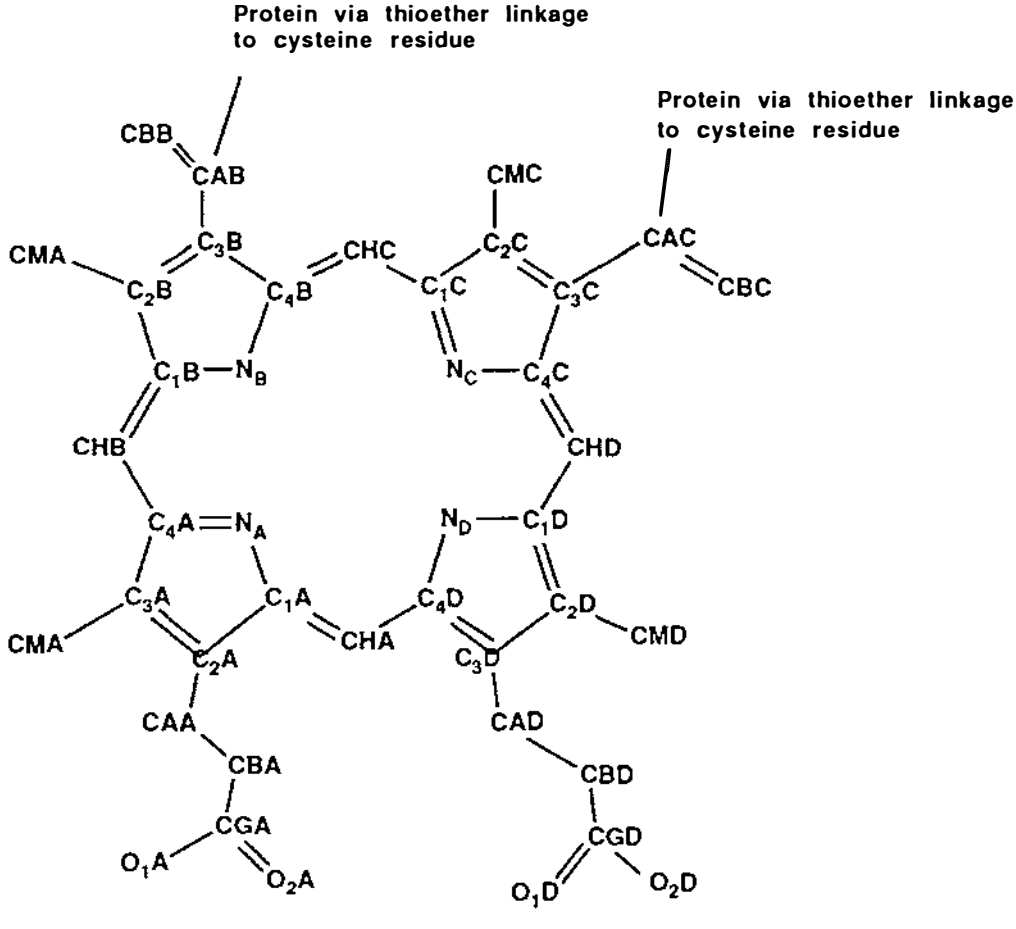


Figure 6.5.1 Protoporphyrin IX haem group.

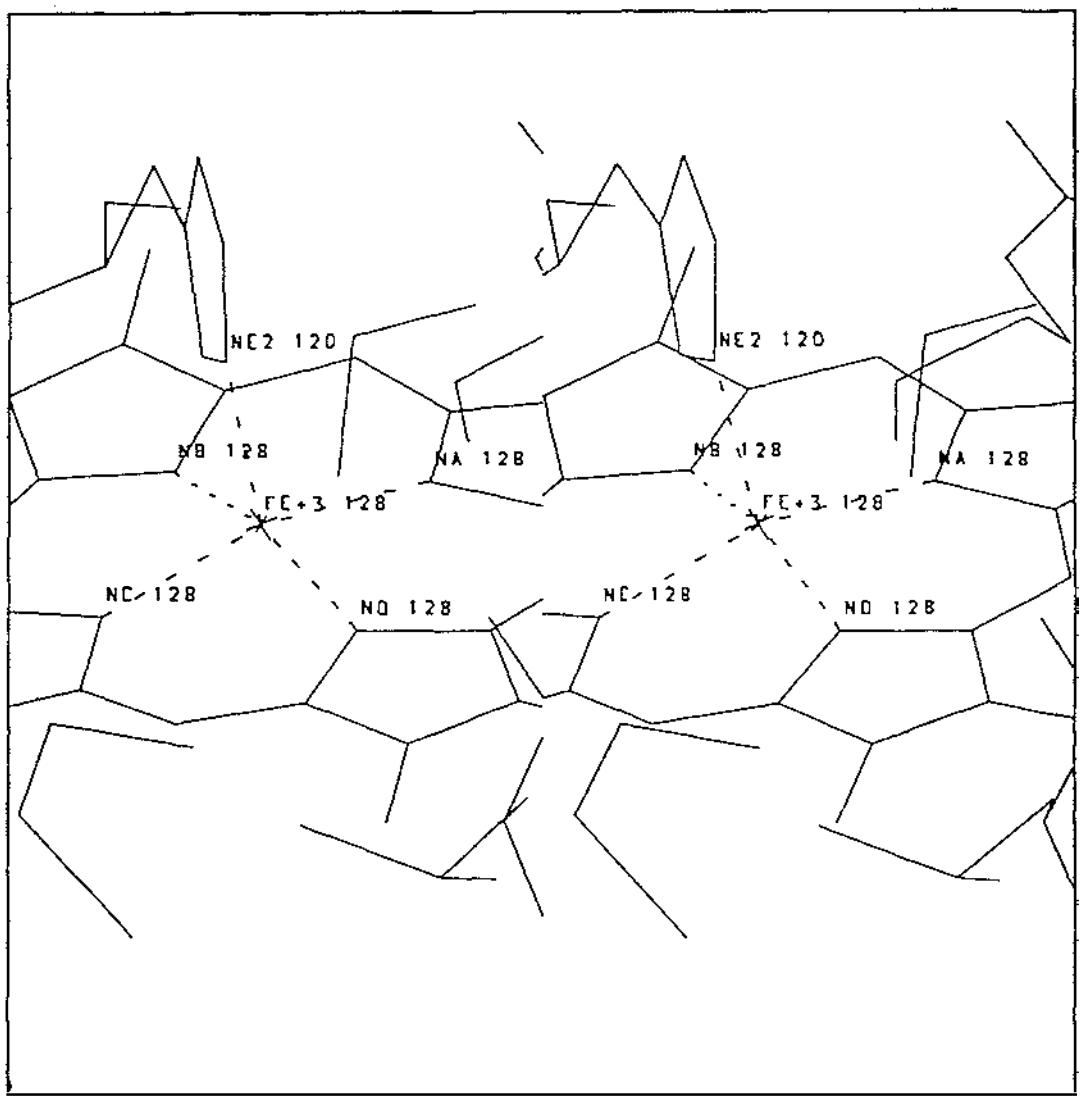


Figure 6.5.2 Stereodiagram showing coordination of the iron atom.

Table 6.5.1 Porphyrin stereochemistry and haem ligand binding geometry of the cytochromes c' from *Alcaligenes* sp and *Alcaligenes denitrificans*.

	Alcaligenes sp	*Alcaligenes denitrificans*
(a) Distances (Å)		
Fe to N-plane[a]	0.33	0.33
Fe to haem plane[b]	0.32	0.32
Fe to His$N_{\varepsilon 2}$	1.94	2.02
Fe to N (average)	1.98	1.99
Fe $-N_A$	1.97	1.94
$-N_B$	1.95	2.00
$-N_C$	1.99	2.01
$-N_D$	2.02	2.01
(b) Angles (°)		
imidazole plane to N_A-N_C-Fe plane	56.8	55.2

a - defined by the least-squares fit of the four in-plane pyrrole nitrogens.
b - defined by the least-squares fit of the 24 atoms in the porphyrin core.

The dihedral angle between the plane of the axial ligand histidine and the plane defined by N_A, N_C and Fe is 56.8 ° (55.2 ° in the cytochrome c' from *Alcaligenes denitrificans*), an orientation that minimises steric overlap of the pyrrole nitrogens and the imidazole hydrogens of the]histidine (Collins *et al.*, 1972). In contrast to cytochromes c, $N_{\delta 1}$ from His120 does not hydrogen bond back to a mainchain carbonyl oxygen; instead in both *Alcaligenes*sp and *Alcaligenes denitrificans* it is hydrogen bonded to a poorly ordered water molecule (the temperature factors for this water molecule are reasonable, at 33.15 $Å^2$ and 32.85 $Å^2$, in the two species, but the shape of the density suggests positional disorder). The $N_{\delta 1}$ ...water hydrogen bond lengths are 2.93 and 2.92 Å respectively. One difference between cytochromes c and c' is the two different haem orientations, as noted by Finzel *et al.* (1985). The guanidinium group of Arg124 packs plane-to-plane against pyrrole ring A, approximately 3.5 Å from it, and its conformation allows the protons on N_{ε} and $N_{\eta 2}$ to be directed towards the plane of the His120 imidazole ring suggesting amino-aromatic hydrogen bonding (Burley and Petsko, 1986a; Burley and Petsko, 1986b; Perutz *et al.*, 1986). This arrangement is shown in Figure 6.5.3.

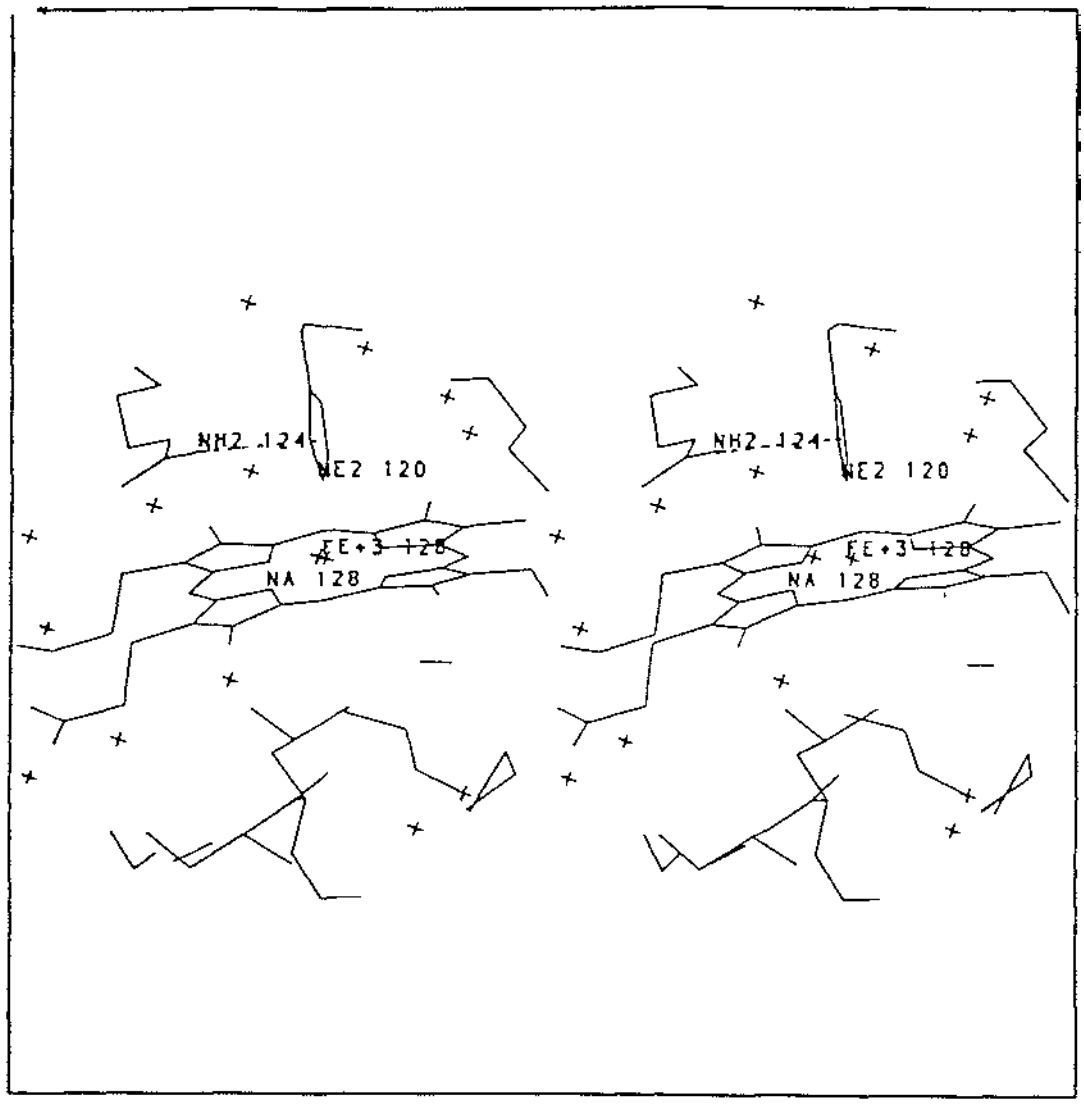


Figure 6.5.3 Stereodiagram showing that the Arg124 sidechain is planar with respect to the haem group and edge on to the imidazole ring of His120.

The haem Fe is shifted out of the pyrrole nitrogen plane by 0.32 Å in both cytochromes c', towards the proximal histidine ligand. This value is consistent with the other cytochrome c' structures that have been determined. The value is also similar to those in other five-coordinate haem systems found, such as

deoxymyoglobin (Takano, 1977a; Takano, 1977b) which has an Fe to N-plane distance of 0.42 Å. Likewise, the His64Leu and His64Val mutants of sperm whale myoglobin have distances of 0.21 Å and 0.27 Å respectively (Quillan *et al.*, 1993). Other high spin cytochromes have slightly smaller shifts of around 0.15 Å, as seen for example in erythrocruorin (Weber, 1978; Steigemann and Weber, 1979).

A plane of best fit for the haem group was calculated (program MEANPLANE) using the 24 atoms in the porphyrin core. This gave an *rms* deviation of 0.10 Å for both species. Table 6.5.2 shows the deviations from planarity of the haem group in the cytochrome c' from *Alcaligenes* sp (similar deviations are found in the *Alcaligenes denitrificans* cytochrome c'). The pyrrole rings in both species exhibit a saddle-shaped distortion from planarity which is also consistent with other cytochromes c'. The deviations from the plane show that rings A and C are bent in one direction away from the iron, and rings B and D are bent in the other. This is shown in Table 6.5.2. The Fe to His $N_{\varepsilon 1}$ distances are very similar in all the cytochromes (Pettigrew and Moore, 1990). The only major difference is the shift of the iron out of the haem plane by approximately 0.30 Å in the cytochromes c', as compared to the small shifts less than 0.05 Å found in low spin cytochromes c (Pettigrew and Moore, 1990). This is due to the difference in coordination number, five-coordinate as opposed to six-coordinate.

Table 6.5.2 Saddle shaped disortion from planarity of the haem group. (values are in Å, a negative sign indicates that the atom is below the plane of the porphyrin ring away from the iron atom)

Equation of plane is 0.8878x + 0.4463y - 0.1122z - 19.6320 = 0

Atom	Ring A	Ring B	Ring C	Ring D
N	0.0694	-0.0429	-0.0461	0.0331
C1	-0.0650	0.1243	-0.1243	0.0668
C2	-0.1705	0.1781	-0.0778	0.0852
C3	-0.1624	0.0650	0.0308	0.0259
C4	-0.0304	-0.0784	0.0180	0.0271
CH	-0.0710	0.1913	-0.2159	0.1077

6.6 Analysis of the haem environment.

The haem group is situated in the cytochrome c' monomer with one face buried in a pocket lined with aromatic sidechains, while the other face is partially exposed to the solvent. The residues involved in the haem binding from helix D cover most of the porphyrin face on the same side as the proximal histidine, but the atoms of the pyrrole rings A and B and the proximal ligand His120 are exposed to the solvent. The total surface area of the haem group exposed to solvent is 104.2 $Å^2$ (SURFACE (CCP4 suite) probe radius of 1.4 Å). Sidechains from helices A, C, D and the BC loop make a network of hydrophobic and aromatic interactions with the haem. There are a total of 273 van der Waal contacts (less than 4.5 Å) between the atoms of the haem group and the protein (CONTACT, see appendix I). A listing of these contacts is given in Table 6.6.1. Hydrophobic residues from helix A form the majority of packing interactions, most of these being with pyrrole ring D. Haem pyrrole rings C and D are found in a pocket lined with the aromatic residues Trp56, Phe59, Phe79 and Phe 86. The sidechain of Arg124 is co-planar with pyrrole ring A, at a distance of ~ 3.5 Å and within hydrogen bonding distance (amino aromatic hydrogen bond) of His120, as discussed in section 6.5. This also probably stabilises the proximal histidine.

The haem propionates are not fully buried in the protein with the propionate acid sidechain from ring A having a solvent accessible surface of 60.7 $Å^2$. The propionate sidechain from ring D is more buried than that of A with an negligible accessible surface area of 0.7 $Å^2$. This means the carboxyl oxygens on the propionate sidechain from ring A are able to form hydrogen bonds with the solvent molecules. This is in contrast to the mitochondrial c-type cytochromes (Salemme, 1977; Mathews, 1985) in which the haem propionates are totally buried from the solvent and are stabilised by hydrogen bonding from internal sidechains. Figure 6.6.1 shows the residues within van der Waals contact of the haem group, while in Table 6.6.1 the van der Waal contacts are listed.

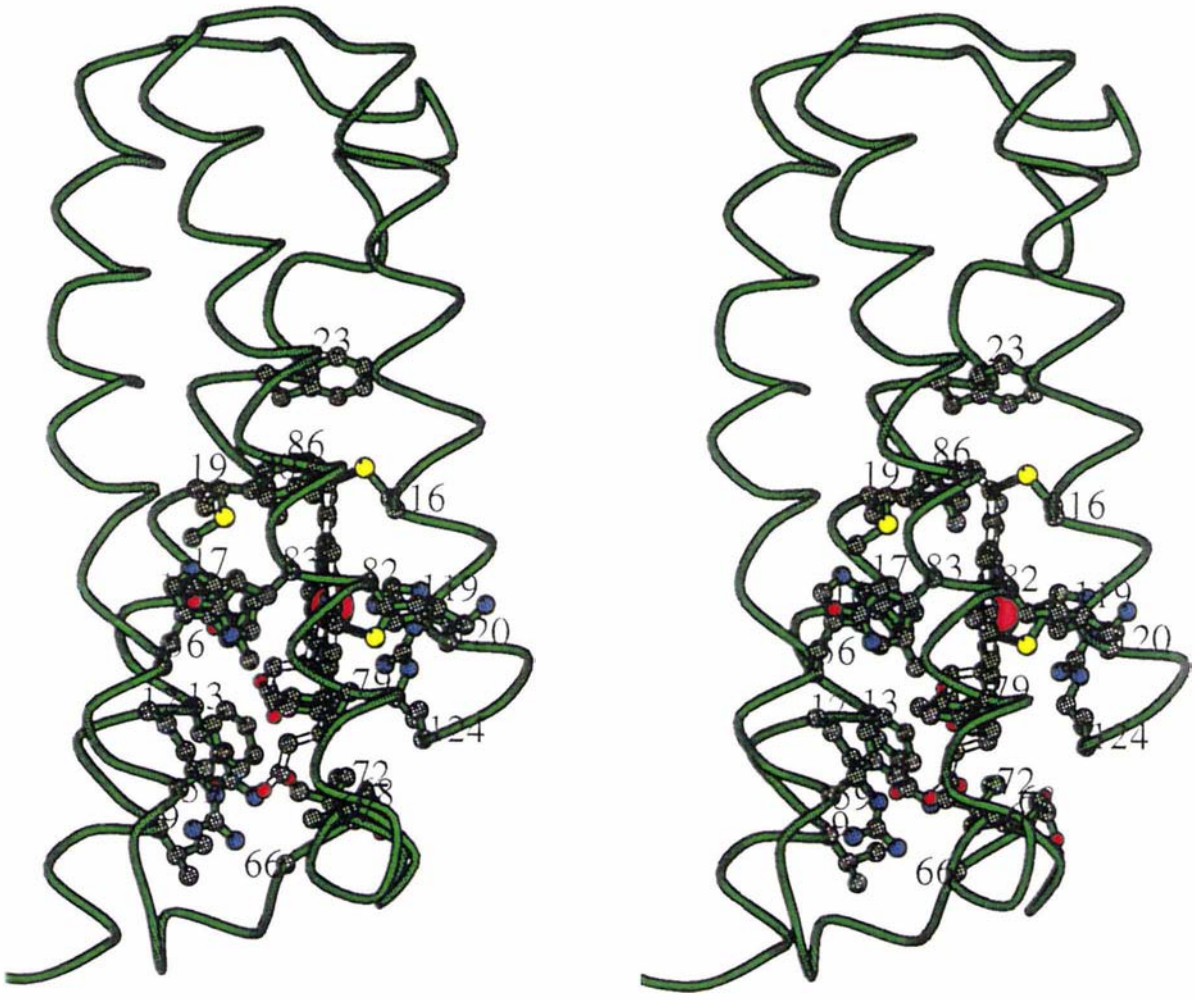


Figure 6.6.1 Stereoview of Molscript (Kraulis, 1991) diagram showing residues within van der Waals distance of the haem group.

Table 6.6.1 Haem packing contacts in *Alcaligenes* sp (the same contacts are made in the *Alcaligenes denitrificans* cytochrome c').

Residue	No. of contacts[a]	Closest distance (Å)	Residue	No. of contacts[a]	Closest distance (Å)
Val 9	1	$C_{\gamma 1}$...O1D (4.31)	Ala 68	12	C_{β}...O1D (3.76)
Arg 12	7	N_{ε}...O1D (2.76)[b]	Ile 72	1	$C_{\delta 1}$...CMD (3.59)
Gln 13	26	$N_{\varepsilon 2}$...O2D (2.66)[b]	Phe 79	4	$C_{\varepsilon 1}$...CAD (3.76)
Leu 16	35	$C_{\delta 1}$...C4D (3.42)	Lys 82	3	C...CBC (3.94)
Thr 17	4	$O_{\gamma 1}$...O2A (3.21)	Gln 83	3	N...CBC (3.71)
Met 19	1	C_{γ}...CBB (4.20)	Phe 86	10	C_{γ}...CMC (3.48)
Ala 20	3	C_{β}...CMA (3.75)	Val 112	1	$C_{\gamma 1}$...CBB (4.10)
Phe 23	5	C_{γ}...CBB (4.14)	Ser 115	1	O...CMC (3.96)
Trp 56	22	$C_{\eta 2}$...C3C (3.59)	Cys 116	26	S_{γ}-CAB (1.81)[c]
Phe 59	6	$C_{\varepsilon 2}$...CMD (3.90)	Cys 119	17	S_{γ}-CAC (1.82)[c]
Gly 65	1	O...O1D (3.77)	His 120	40	$N_{\varepsilon 2}$-Fe (2.01)[c]
Gly 66	6	C...O2D (3.77)	Tyr 123	2	O...CMD (3.51)
Asp 67	16	N...O2D (2.79)[b]	Arg 124	24	$N_{\eta 1}$...C2A (3.48)

a - Sidechain atom...haem atom contacts < 4.5 Å.

b - Hydrogen bonds.

c - Covalent bonds.

6.7 Dimer interface.

The dimer interface in the cytochromes c' from *Alcaligenes* sp and *Alcaligenes denitrificans* is formed by residues in helices A and B of one molecule and the same helices in the two-fold symmetry-related molecule; they are related by the symmetry operator y - x, y, 1/2 - z. Between them these helices bury a total surface area of 1184 $Å^2$ (as calculated by the method of Connolly, 1983), and create an AA'BB' four-helix bundle.

The N-terminal residues (Pyr1 and Phe2) from each of the monomers fold around and seal off either end of the interface (shown in Figure 6.7.1). This conformation is stabilised by the packing of Pyr1 and the aromatic ring of Phe2 against Ala8 and Tyr11 in the same molecule, and Ser21 and Arg25 on the other. There are a number of well-defined solvent molecules which form stabilising contacts in this part of the dimer interface. Next to the N-terminal residues in the interface, Tyr11 on one molecule, and His22 on the other, form a strong hydrogen bond. This pair of residues, at each end of the interface, seal off a region from which solvent is totally excluded. This central part of the interface, between Tyr11 and His22 on helix A and Val47 and Leu54 on helix B, is marked by shape and size complementarity. Leu18 on one molecule fits between Tyr11 and Leu18 on the opposing molecule into a hole created by the small Ala sidechain at position 15, as is shown diagrammatically in Figure 6.7.2. Table 6.7.1 lists the residues involved in the dimer interface with their closest contact distance and number of contacts made by each. From Figure 6.7.2 and Table 6.7.1, it is noticeable that the major stabilising interactions in the interface are hydrophobic, with additional stabilisation coming from the hydrogen bonds on the edge of the interface.

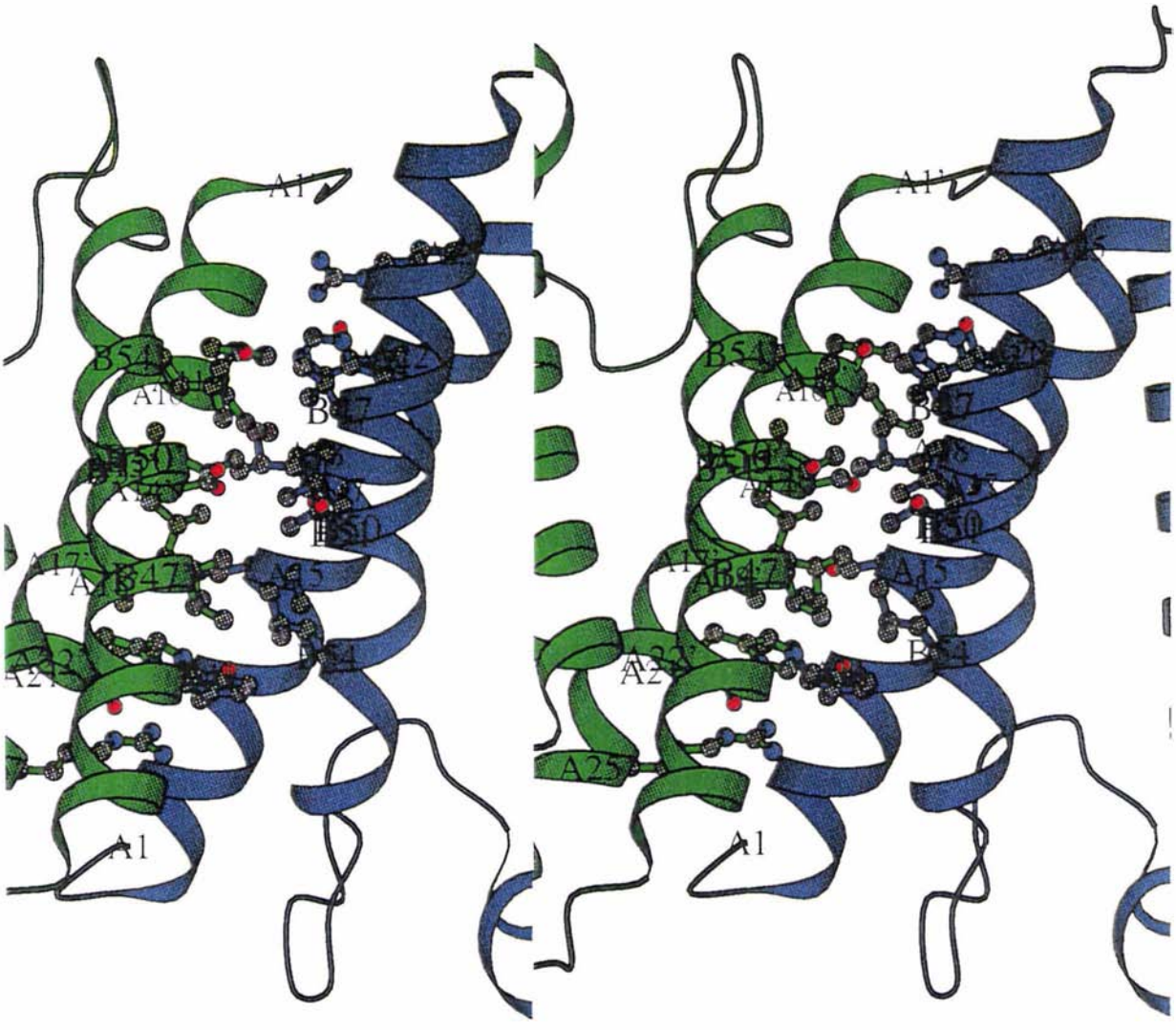


Figure 6.7.1 Stereoview of Molscript (Kraulis, 1991) diagram of the dimer interface.

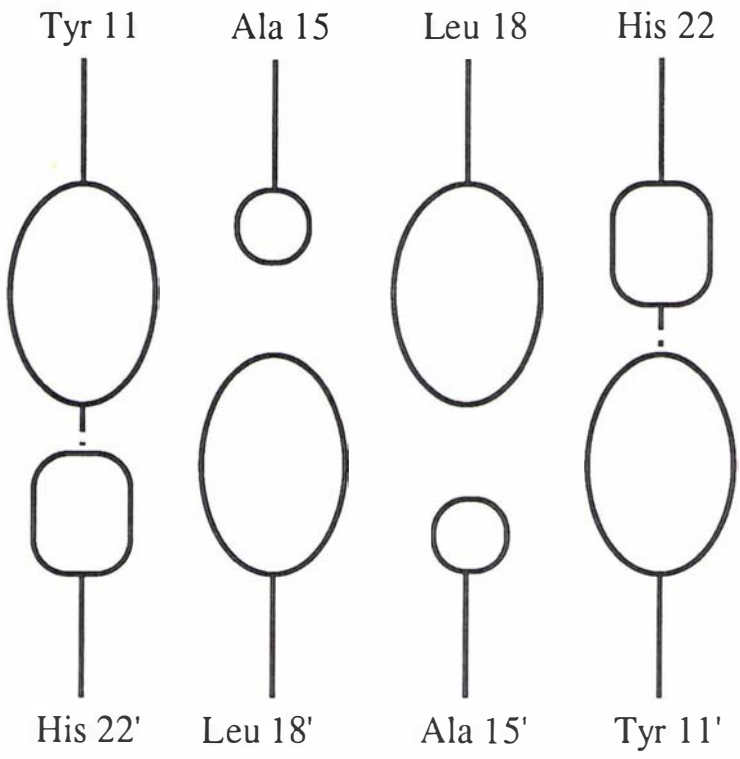


Figure 6.7.2 Schematic diagram showing shape recognition at interface.

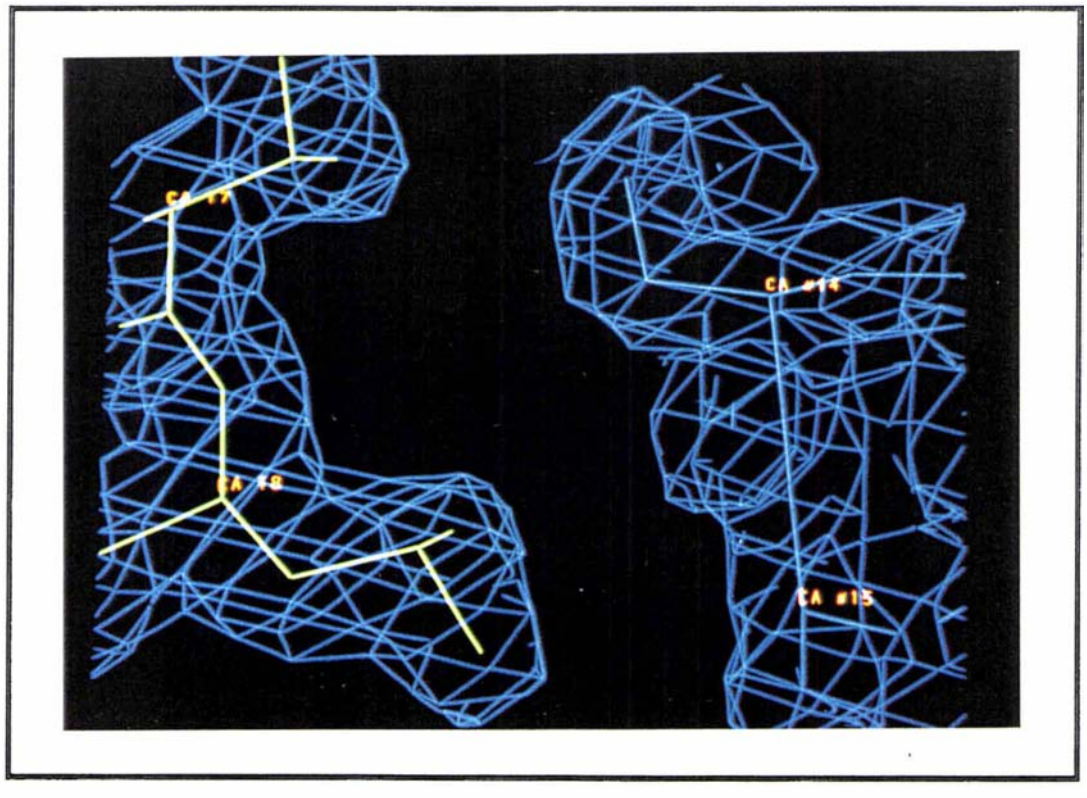


Figure 6.7.3 Electron density from final 2Fo-Fc map around dimer interface.

Table 6.7.1 Dimer interface interactions in the cytochromes c' from *Alcaligenes* sp and *Alcaligenes denitrificans*.

Residue	No. of contacts[a]	Contacts with	Closest contact distance (Å)
Pyr 1	6	Arg 25	C_{γ}...Arg 25 $N_{\eta 1}$ (3.70)
Phe 2	3	Ser 21, Arg 25	$C_{\varepsilon 2}$...Arg 25 $N_{\eta 1}$ (4.03)
Lys 10	2	Thr 17	C_{γ}...Thr 17 $C_{\gamma 2}$ (4.11)
Tyr 11	25	Leu 18, Ser 21, His 22, Arg 25	O_{η}...His 22 $N_{\varepsilon 2}$ (2.67)
Ser 14	22	Ser 14, Thr 17, Leu 18	O_{γ}...Ser 14 O_{γ} (3.72)
Ala 15	5	Leu 18	C_{β}...Leu 18 $C_{\delta 1}$ (4.08)
Thr 17	6	Lys 10, Ser 14	$C_{\gamma 2}$...Ser 14 O_{γ} (3.75)
Leu 18	18	Tyr 11, Ser 14, Ala 15	$C_{\delta 2}$...Tyr 11 $C_{\delta 1}$ (3.42)
Ser 21	4	Tyr 11	O_{γ}...Tyr 11 $C_{\delta 1}$ (3.58)
His 22	8	Tyr 11, Leu 54	$N_{\varepsilon 2}$...Tyr 11 O_{η} (2.67)
Arg 25	8	Tyr 11	$N_{\eta 2}$...Tyr 11 $C_{\varepsilon 1}$ (3.56)
Val 47	2	Leu 54	$C_{\gamma 1}$...Leu 54 $C_{\delta 1}$ (4.21)
Thr 50	2	Thr 50	C_{β}...Thr 50 $C_{\gamma 2}$ (4.33)
Leu 51	2	Leu 51, Leu 54	$C_{\delta 2}$...Leu51 $C_{\delta 2}$ (3.92)
Leu 54	4	His 22, Val 47, Leu 51	$C_{\delta 1}$...Val47 $C_{\gamma 1}$ (4.21)

a - Sidechain...sidechain contacts < 4.5 Å

6.8 Thermal parameters.

The thermal parameters of the two cytochromes c' follow very similar patterns. Figures 6.8.2 and 6.8.3 show plots of the mainchain and sidechain B-values of both structures. The average overall mainchain B-value for *Alcaligenes denitrificans* is 6.2 Å^2 higher than that for the *Alcaligenes* sp cytochrome c' and the overall sidechain B-value is 5.8 Å^2 higher. The actual values are given in Table 6.8.1. The higher value is probably due to the effective resolution of the *Alcaligenes denitrificans* data being lower. The distribution of B-values correlate with general structural features of proteins, with the sidechains on the outside of the structure having higher B-values, and those on the inside generally having lower B-values. Lower B-values are found in the residues comprising the four helices due to the packing constraints. Within each helix the highest B-values are found at its N and C-termini. The lowest B-values in the structure are found for the residues of helices A and B which form the dimer interface (average mainchain B-value of 13.3 Å^2 for residues Tyr11 to His22 and Val47 to Leu54 in the *Alcaligenes* sp cytochrome c' and 19.8 Å^2 in the *Alcaligenes denitrificans* cytochrome c').

In general the B-values found in the connecting loops are slightly higher than those found in the "core" elements such as the α-helices (16.7 Å^2 as opposed to 16.1 Å^2 in the helices for *Alcaligenes* sp cytochrome c', and 23.8 Å^2 as opposed to 21.9 Å^2 for *Alcaligenes denitrificans* cytochrome c'). Only two small regions, however, have significantly higher B-values. These are residues 60 - 63 in the BC loop (24.3 and 30.3 Å^2 for mainchain atoms respectively in the two structures) and residues 75 - 80 at the N-terminus of helix C (26.7 and 30.5 Å^2 for mainchain atoms). Even these are not particularly high, and the general conclusion is that the whole structure, including the loops, is relatively rigid. Figures 6.8.1 and 6.8.2 show similar patterns to those found in the cytochrome c' structures solved previously (Finzel *et al.*, 1985; McRee *et al.*, 1993). Table 6.8.1 gives the average B-values for the two Alcaligenes species while Table 6.8.2 lists the average mainchain B-values for the helices in the two cytochromes c'. It can be seen from Figures 6.8.1 and 6.8.2 that the largest differences in the thermal parameters between *Alcaligenes* sp and *Alcaligenes denitrificans* occur in the loop regions.

Table 6.8.1 Average mainchain and sidechain B-values for the cytochromes c' from *Alcaligenes* sp and *Alcaligenes denitrificans*.

	Average mainchain B-value	Average sidechain B-value
Alcaligenes sp	16.7 Å^2	26.4 Å^2
Alcaligenes denitrificans	22.9 Å^2	32.2 Å^2

Table 6.8.2 Average mainchain B-values in the α-helices.

	ASCC (Å^2)[a]	ADCC (Å^2)[b]
Helix A	14.1	21.3
Helix B	14.7	19.9
Helix C	19.4	24.8
Helix D	16.4	21.5

a - *Alcaligenes* sp cytochrome c'.

b - *Alcaligenes denitrificans* cytochrome c'.

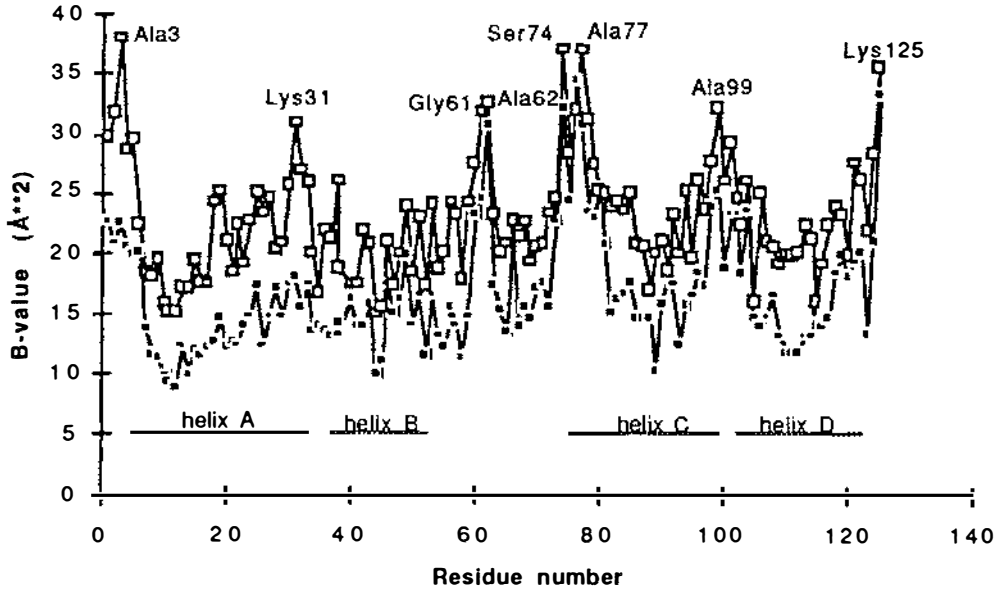


Figure 6.8.1 Mainchain B-values for the cytochromes c' from *Alcaligenes* sp and *Alcaligenes denitrificans* (filled black squares-*Alcaligenes* sp, hollow squares-*Alcaligenes denitrificans*).

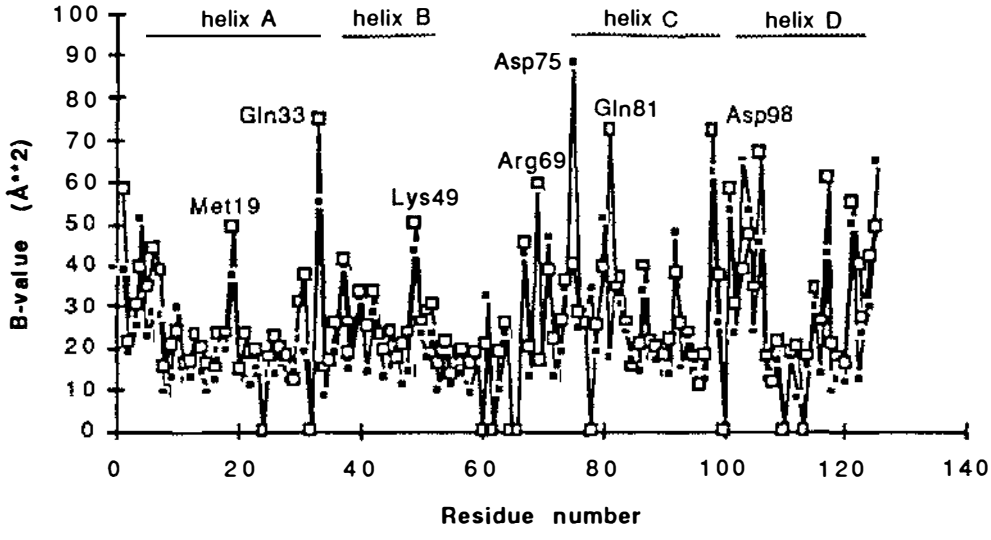


Figure 6.8.2 Sidechain B-values for the cytochromes c' from *Alcaligenes* sp and *Alcaligenes denitrificans* (filled black squares-*Alcaligenes* sp, hollow squares-*Alcaligenes denitrificans*). (glycine residues have a sidechain B-value of zero)

6.9 Solvent structure.

All the solvent molecules in the structures of cytochrome c' from *Alcaligenes sp* and *Alcaligenes denitrificans* have been modelled as waters, on the basis of the size of their electron density peaks and the enviroment of each site. The final *Alcaligenes* sp model includes 89 water molecules, and the final *Alcaligenes denitrificans* model 75. Figures 6.9.1 and 6.9.2 show the solvent structures for the cytochromes c' for the two proteins.

Of the 89 solvent molecules in the cytochrome c' from *Alcaligenes* sp, 84 form explicit hydrogen bonds, and 25 are in the second coordination sphere (compared to 67 and 12 respectively in the cytochrome c' from *Alcaligenes denitrificans*). There are no internal water molecules. The solvent molecules of the first hydration layer are bound to sidechains on the protein surface or situated in cavities and grooves in the protein. Due to the helical nature of cytochrome c' a number of deep grooves and channels exist between the helices and at the dimer interface. The majority of the water molecules are concentrated in these grooves or in areas between protein molecules in the crystal lattice (especially around residues 30 - 40). These molecules are then able to form hydrogen bonds to supplement and strengthen the direct intermolecular interactions which occur (see section 6.10). In other such proteins, such water-mediated interactions generally outweigh the number of direct intermolecular contacts (Sheriff *et al.*, 1987). The other

solvent molecules present in the structure probably act as a type of molecular adhesive, filling in surface irregularities between the protein molecules at lattice contacts (Salemme *et al.*, 1988). Table 6.9.1 shows the number of water molecules making a given number of hydrogen bonds and their mean B-values.

Table 6.9.1 Number of hydrogen bonds and mean B-values.

	ASCC		ADCC	
(a) Number of waters making 1, 2, 3 or 4 hydrogen bonds.				
	Number	Mean B (Å^2)	Number	Mean B (Å^2)
1	24	42	28	54
2	25	40	18	49
3	25	33	12	45
4	9	30	8	30
(b) Number of waters making a given number of hydrogen bonds to protein.				
	Number	Mean B (Å^2)	Number	Mean B (Å^2)
0	25	42	12	52
1	33	41	34	49
2	18	31	13	43
3	7	26	7	38
4	1	33	1	20

It can be seen from Table 6.9.1 that the number of hydrogen bonds a solvent molecule forms is correlated with its B-value, as generally the greater the number of hydrogen bonds formed by a solvent molecule the lower its B-value. In Table 6.9.2 an analysis of the water hydrogen bonds in the cytochrome c' from *Alcaligenes* sp is given. Table 6.9.3 gives an analysis of the water hydrogen bonding in the *Alcaligenes denitrificans* cytochrome c'.

Table 6.9.2 Analysis of water hydrogen bonds (ASCC).

Partner	Number	<Distance> (Å)	*Rms*	<Angle> (°)[a]
mainchain O	45	3.02	0.29	124.3
mainchain N	9	3.00	0.28	160.6
sidechain O	23	3.05	0.35	119.6
sidechain N	17	3.20	0.34	143.1

a - If the donor atom is a nitrogen whose hydrogen atom position is unambiguous, the angle given is that subtended at the calculated hydrogen atom position. If the donor is an oxygen, the angle is: acceptor...oxygen -- carbon atom.

Table 6.9.3 Analysis of water hydrogen bonds (ADCC).

Partner	Number	<Distance> (Å)	*Rms*	<Angle> (°)[a]
mainchain O	38	2.98	0.32	124.1
mainchain N	9	2.95	0.26	160.6
sidechain O	25	3.18	0.35	125.9
sidechain N	13	3.10	0.33	145.6

a - If the donor atom is a nitrogen whose hydrogen atom position is unambiguous, the angle given is that subtended at the calculated hydrogen atom position. If the donor is an oxygen, the angle is: acceptor...oxygen -- carbon atom.

The data in Tables 6.9.2 and 6.9.3 are similar to those that have been previously determined for highly resolved protein structures such as in actinidin (Baker, 1980) and erythrocruorin (Steigemann and Weber, 1979) and in general analyses of water hydrogen bonding in proteins (Baker and Hubbard, 1984; Thanki *et al.*, 1988; Thanki *et al.*, 1991).

There are four main sites in which the solvent molecules occur in the *Alcaligenes* species. These areas are (1) between the N-terminus and helix A, (2) near the BC loop (stabilising the loop), (3) in a groove between helices C and D and (4) occurring in areas of intermolecular contact. These areas are illustrated in Figures 6.9.2 and 6.9.3. The difference between the solvent structures in *Alcaligenes* sp and

Alcaligenes denitrificans is very small, as would be expected from the isomorphism of the two crystal forms, and the method by which the water molecules were added. As mentioned earlier, slightly fewer water molecules have been included in the *Alcaligenes denitrificans* model than in that for *Alcaligenes* sp (75 as opposed to 89). Of these, there are 58 cases where the position in one structure is within 1.0 Å of the position in the other. These water molecules which are "common" to the two structures are not confined to any particular areas of the cytochrome c' monomer. The water molecules which differ in position by more than 1.0 Å (i.e. those which are "unique" to either structure) are few in number and their average B-values are higher (43.9 Å^2) than the "common" solvent molecules (average 38.2 Å^2). The differences in the positions of the water molecules which are "common" between *Alcaligenes* sp and *Alcaligenes denitrificans* are weakly correlated with the average B-value of the solvent pair (see Figure 6.9.1). It can be seen that in general the lower the average B-value, the smaller the difference in position between the water molecules. This suggests that the differences between the solvent positions are smallest when the water molecules are better ordered.

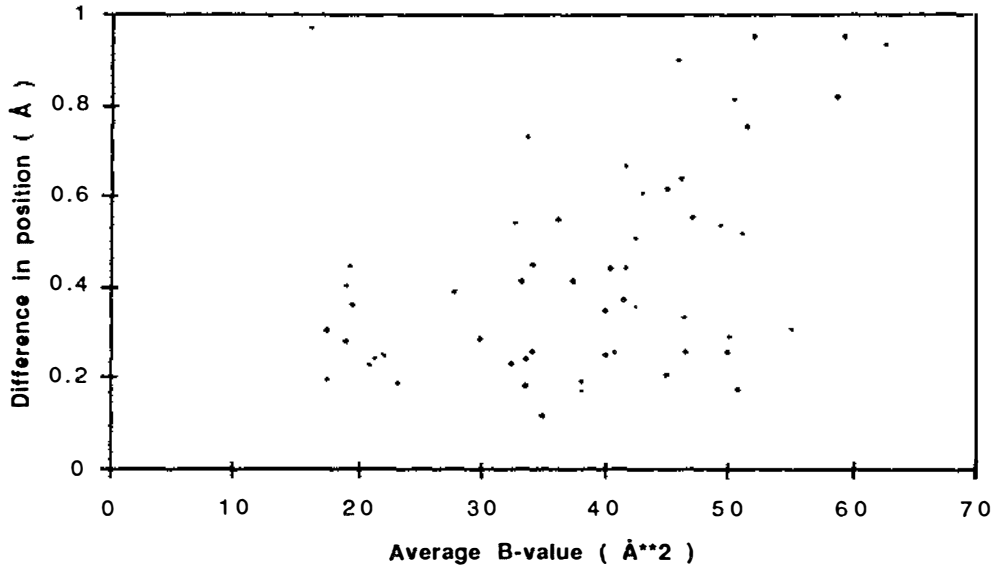


Figure 6.9.1 Plot of correlation of "common" solvent molecules with average B-values.

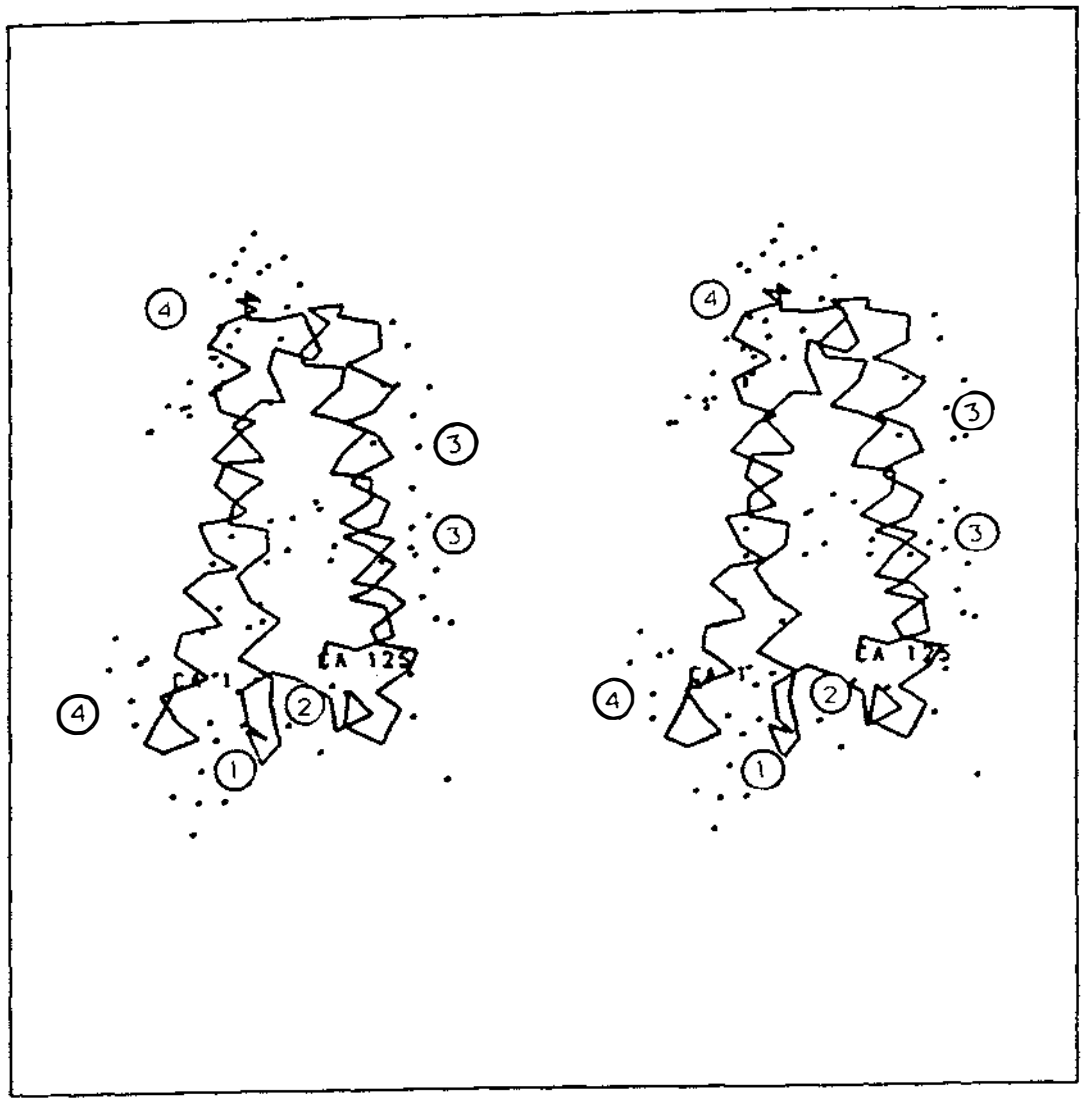


Figure 6.9.2 Solvent structure in the cytochrome c' from *Alcaligenes* sp.

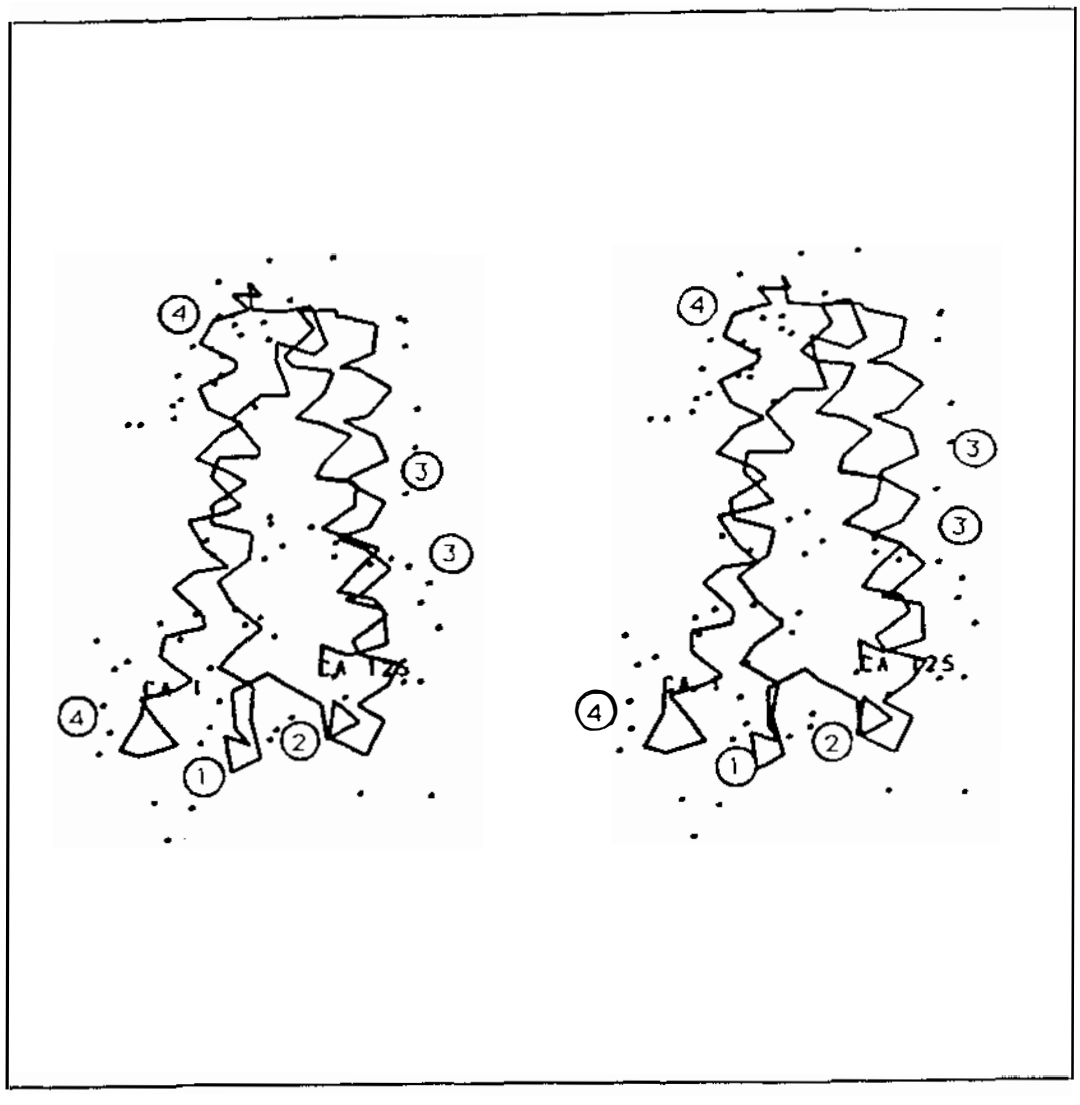


Figure 6.9.3 Solvent structure in the cytochrome c' from *Alcaligenes denitrificans*.

6.10 Crystal packing.

As the cytochromes c' from *Alcaligenes* sp and *Alcaligenes denitrificans* crystallise isomorphously (Norris *et al.*, 1979), the crystal packing is basically identical. The asymmetric unit contains a cytochrome c' monomer, with the functional dimer being created by a crystallographic two-fold axis. The crystal packing of the cytochrome c' molecules from *Alcaligenes* sp and *Alcaligenes denitrificans* is shown in Figures 6.10.1 and 6.10.2.

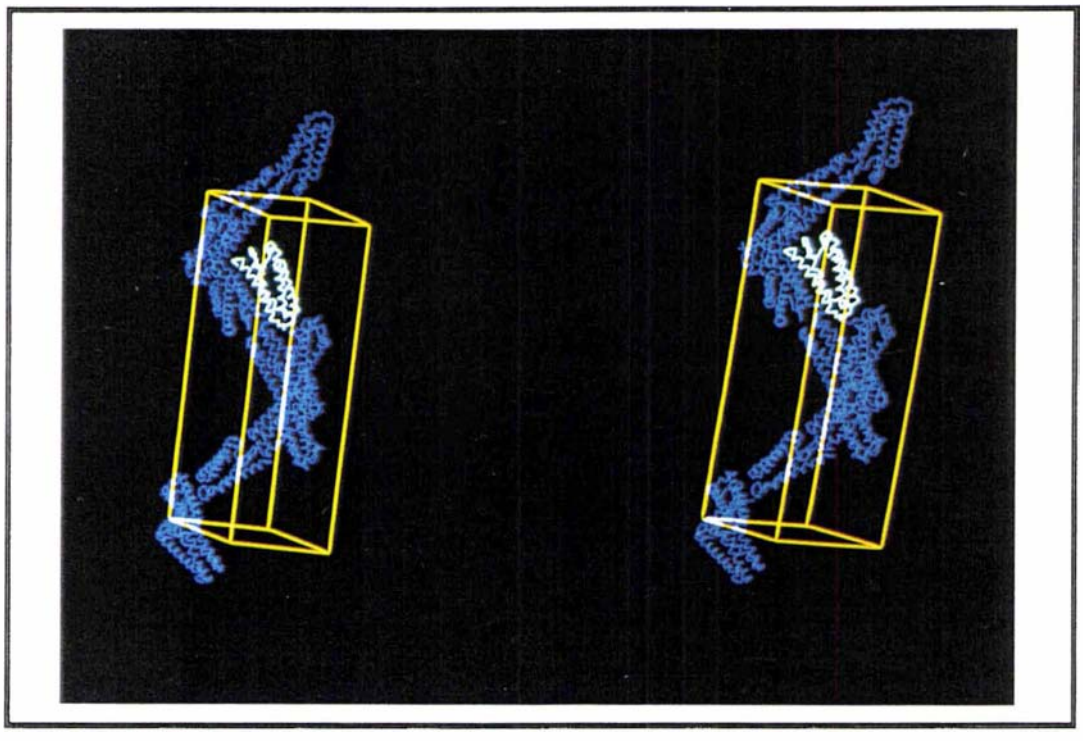


Figure 6.10.1 Crystal packing of the cytochrome c' from *Alcaligenes* sp.

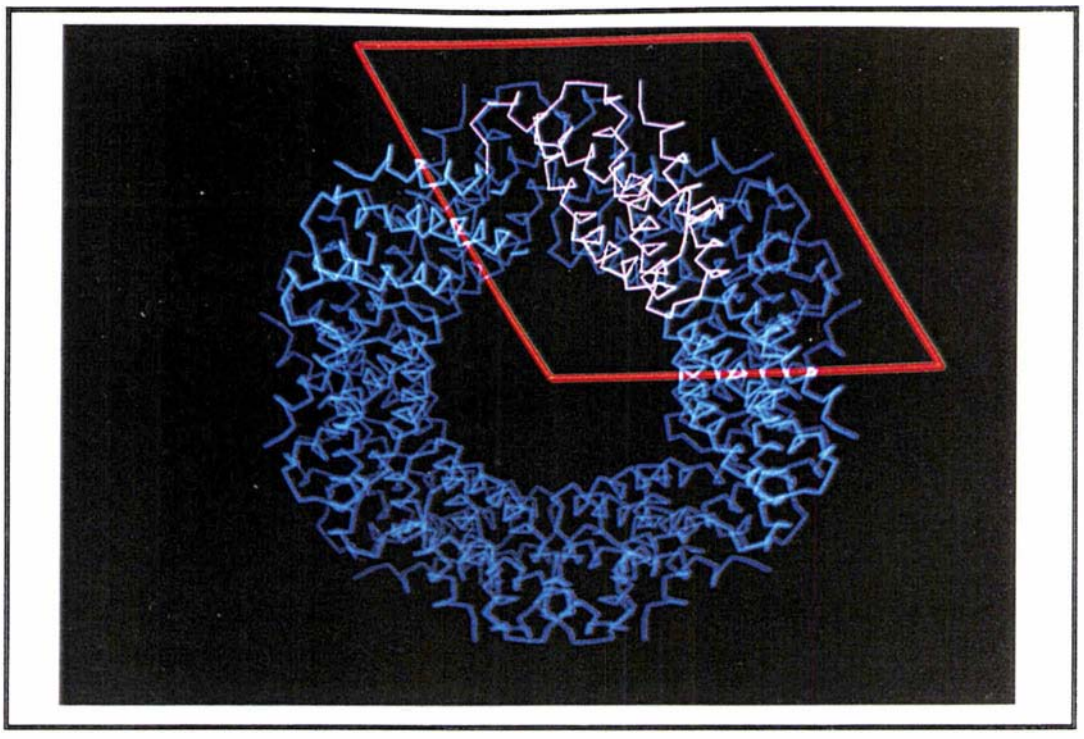


Figure 6.10.2 Looking down six fold axis.

It can be seen from Figure 6.10.1 that the dimers twist about the 6_1 axis to form long strings. This results in the majority of intermolecular contacts being formed between residues on helices C and D since A and B are involved in dimer contacts. Table 6.10.1 lists the intermolecular contacts for the *Alcaligenes* sp cytochrome c' (the contacts for *Alcaligenes denitrificans* cytochrome c' are essentially the same).

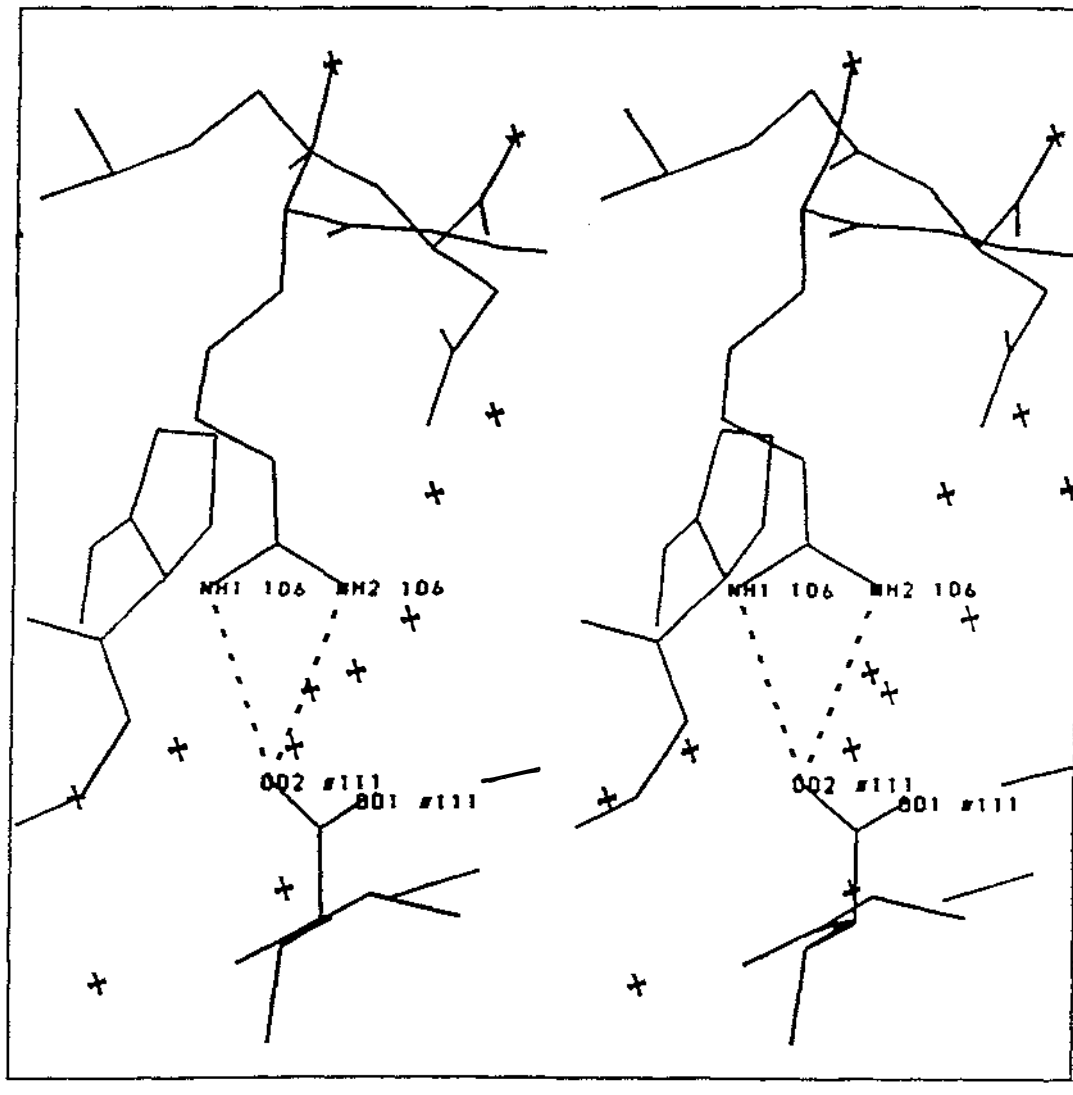


Figure 6.10.3 Salt bridge between Arg106 and Asp111(symmetry related molecule).

Table 6.10.1 Intermolecular hydrogen bond contacts (less than 3.5 Å).

Contact	Distance (Å)	Symmetry operator	Translation Tx Ty Tz
Phe2 N...O Gly32	2.94	y, y-x, 1/6+z	0 0 0
Glu6 Oe2...Nz Lys82	2.92	-x, y-x, 1/3-z	1 0 0
Tyr11 OH... Nε2 His22	2.66	y-x, y, 1/2-z	0 0 0
His22 Nε2...OH Tyr11	2.66	y-x, y, 1/2-z	0 0 0
Gly32 O...N Phe2	2.94	y, y-x, 1/6+z	0 0 0
Gln40 Oε1...Nε2 Gln40	2.98	y, x, 2/3-z	0 0 0
Gln40 Nε2...Oε1 Gln40	2.98	y, x, 2/3-z	0 0 0
Glu64 Oe2...Nz Lys92	2.51	-x, y-x, 1/3-z	1 1 0
Lys82 Nζ...Oε2 Glu6	2.92	-x, y-x, 1/3-z	1 0 0
Asp88 Oδ1...Nζ Lys125	2.77	-x, y-x, 1/3-z	1 0 0
Asp88 Oδ2...Nζ Lys125	2.81	-x, y-x, 1/3-z	1 0 0
Lys92 Nζ....Oε2 Glu64	2.51	-x, y-x, 1/3-z	1 1 0
Arg106 Nη1...Oδ2 Asp111	3.25	-x, y-x, 1/3-z	1 0 0
Arg106 Nη2...Oδ2 Asp111	3.25	-x, y-x, 1/3-z	1 0 0
Oδ2 Asp111...Arg106 Nη1	3.25	-x, y-x, 1/3-z	1 0 0
Oδ2 Asp111...Arg106 Nη2	3.25	-x, y-x, 1/3-z	1 0 0
Nζ Lys125...Asp88 Oδ1	2.77	-x, y-x, 1/3-z	1 1 0
Nζ Lys125...Asp88 Oδ2	2.81	-x, y-x, 1/3-z	1 1 0

The density for the residues involved in salt bridges with symmetry related molecules is well defined, and the B-values of these sidechains are a reflection of this. The lysine residues, at positions 82, 92 and 125 are the only ones present in these cytochrome c' structures which have good density for all the sidechain atoms. The majority of the intermolecular interactions are salt bridges and these are predominantly found between residues in helices C and D (as A and B form hydrophobic dimer interactions).

Chapter Seven

Structural comparisons with other cytochromes c'

The aim of this chapter is to compare the two *Alcaligenes* cytochrome c' structures determined in this study with other cytochrome c' structures. As well as the two models that were available for the molecular replacement studies, a third cytochrome c' structure has since been published, ie. a 1.8 Å resolution structure from *Chr. vinosum* (McRee *et al.*, 1993), PDB entry 1BBH. Although the overall structures of all the cytochromes c' are grossly similar, there are differences in detail which will be discussed in this chapter.

7.1 Sequence alignment.

The sequence alignment given in Figure 7.1.1 was obtained by first aligning the residues of helices A and D. This was done as the initial step because they have sufficient sequence identity to be sure of the structural correspondence and both have residues that interact with the haem. A structural superposition using the equivalences in helices A and D then allowed residues in the other helices to be matched, and finally the alignment to be extended into the loop regions. A full superposition of the structures using least square methods was then done, the *rms* deviations were analysed and residues which deviated by more than 3σ were removed from the superposition.

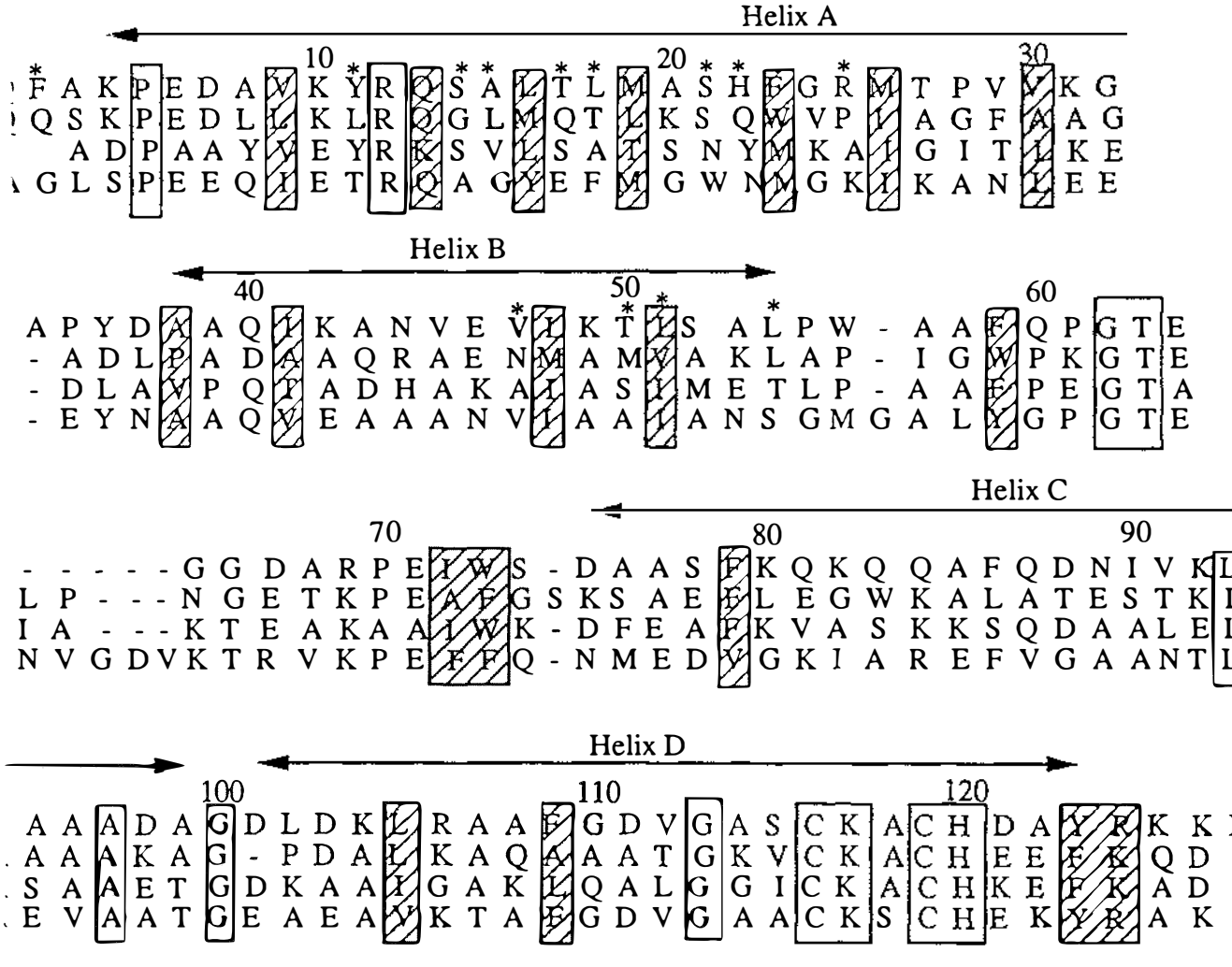


Figure 7.1.1 Sequence alignment of cytochromes c' based on structural superpostion. (* indicates dimer contact formed by the residue, clear boxes show conserved residues, while the shaded boxes show alignment based on sidechains that are identical or similar in geometry).

7.2 Structural superposition.

In the previous section a sequence alignment of the cytochromes c' whose structures have been solved was presented; in this section the alignment will be quantified by structural superposition using least-squares methods. Table 7.2.1 shows the regions that are most closely conserved structurally between *Alcaligenes* sp cytochrome c' and each of the other cytochromes c'. Those residues which deviate more than 3 σ are not included.

Table 7.2.1 *Rms* deviations and residues used in the sequence alignment of cytochromes c'.

				Rms difference (Å)
Alcaligenes sp	4-31	40-64	66-125	1.37
Chr. vinosum	4-31	38-62	71-130	
Alcaligenes sp	6-29	40-64	66-125	1.94
R. molischianum	6-29	39-63	68-127	
Alcaligenes sp	6-29	40-65	66-122	2.07
R. rubrum	6-29	37-63	66-122	

One important point stands out from Table 7.2.1 ie. that, the two structures used in the molecular replacement studies have the largest *rms* deviations against the final *Alcaligenes* sp cytochrome c' structure (1.94 and 2.07 Å respectively). This may account for some of the difficulties in achieving a correct molecular replacement solution. In Figures 7.2.1, 7.2.2, and 7.2.3 the *rms* deviations of CVCC, RMCC, and RRCC with respect to ASCC are plotted against residue number. These superpositions were carried out using least-squares methods (LSQKAB; see appendix I) and from the sequence alignment given in Table 7.2.1. The AB loop was also included in these figures and the positions of the insertions in this loop (and elsewhere in the molecule) are shown.

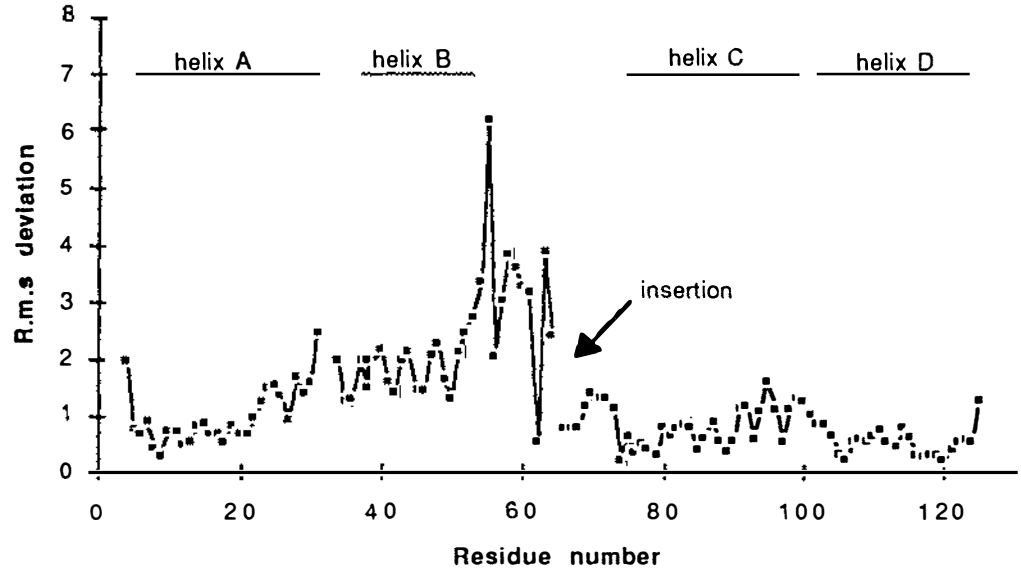


Figure 7.2.1 *Rms* deviation as a function of residue number for CVCC.

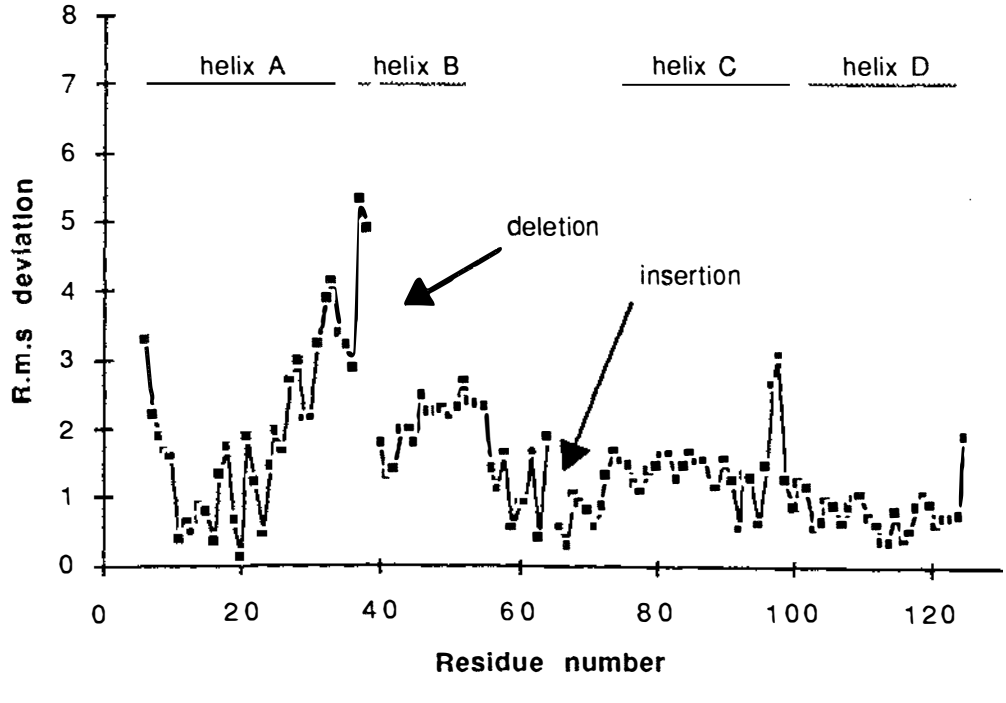


Figure 7.2.2 *Rms* deviation as a function of residue number for RMCC.

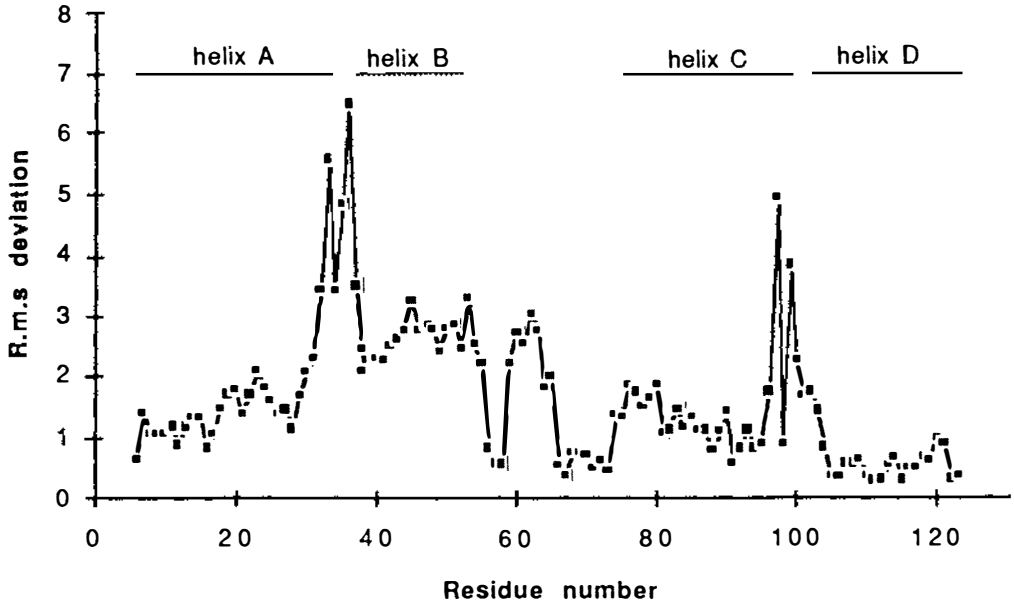


Figure 7.2.3 *Rms* deviation as a function of residue for RRCC.

From Figures 7.2.2, 7.2.3, and 7.2.4 it can be seen that the residues forming the loop between helices A and B shows the greatest *rms* deviation compared with that of the *Alcaligenes* sp structure. This could be due to helix B having a quite variable length in cytochromes c' (see Table 7.6.1.1), with a corresponding lack of common structural features in this loop. The BC loop is interesting in that it is structurally similar in cytochromes c' yet contains few conserved residues. It is also the largest loop, but its conformation appears to be conserved because of its role in helping form the bottom of the haem pocket.

7.3 Secondary Structure.

The helices in the four cytochromes c' vary slightly in length from one species to another, as shown in Table 7.3.1. Here the helices are defined as starting at the residue contributing the first C=O...NH bond, continuing for as long as successive (ψ, ϕ) angles have approximately helical angles and finishing with the residue which contributes the NH group to the last C=O...NH bond. This means where there is a break in hydrogen bonding (as in helix A at Pro28) the helix is deemed to continue if the conformation remains helical.

Table 7.3.1 Residues involved in α-helices in cytochromes c'.

	ASCC	RMCC	RRCC	CVCC
Helix A	4 - 33	5 - 33	2 - 30	5 - 32
Helix B	37 - 54	39 - 54	37 - 52	35 - 52
Helix C	75 - 100	78 - 100	75 - 99	80 - 103
Helix D	101 - 124	103 - 126	101 - 121	106 - 129

Helix A

ASCC and RMCC match up to residue 25 after which they diverge as RMCC has a Pro at 25, but ASCC has Pro at 28 so that the helix bends at different residues. The helix ends with the same hydrogen bonding pattern 29O...32N, 28O...33N.

ASCC and CVCC match quite well up to residue 30.

ASCC and RRCC match reasonably well until residue 28. RRCC has no Pro in first helix.

Helix B

Starts earlier in ASCC, RRCC and CVCC than in RMCC (in which a β-turn 37 - 40 takes the place of the first part of the helix. Finishes at approximately residue 54 and then forms a turn of 3_{10} helix before the BC loop.

Helix C

Start of helix is dependent on the number of insertions/deletions in BC loop. This helix is similar in all the cytochromes c'. Ends with a what is described as a one residue turn before helix D.

Helix D

ASCC and RRCC match very well (including an N-cap at the beginning). CVCC also has a cap at the start of the helix (O_γ104...N106). All the cytochromes c' have a bend near the end of the helix to accommodate haem binding.

In Table 7.3.2 the residues involved in 3_{10} helices in cytochromes c' are given. These helices occur at the beginning and end of the BC loop. The first is an extension of helix B but the second, preceeding helix C is basically at right angles to this helix. In the reports on the structures of RRCC and CVCC these helices have not been mentioned, but analysis of the hydrogen bonding pattern shows that they do occur. The occurrence of these helices in each structure suggests that they are part of a common structural motif in this loop, in all cytochromes c'.

Table 7.3.2 Residues forming 3_{10} helices in cytochromes c'.

ASCC	RMCC	RRCC	CVCC
55 - 58	53 - 58	53 - 56	53 - 58
69 - 75	71 - 76	72 - 75	74 - 80

7.4 Helix orientations.

All the cytochrome c' structures are based on the packing of four helices around the haem group. In order to compare their "core" structures the *Alcaligenes* sp structure was superimposed on each of the other cytochrome c' structures using the C_α or mainchain atoms (C_α, C, O, N) for the four helices, together with all the atoms of the haem group. The residues used from each of the four helices are given in Table 7.4.1 and root-mean-square (*rms*) differences are given in Table 7.4.2. The results indicate that this "core structure" (helices and haem) is most similar in *Chr. vinosum* cytochrome c' when compared with that of *Alcaligenes* sp cytochrome c' with an *rms* of 0.97 Å if the C_α atoms of the four helices plus all the atoms of the haem are superimposed (1.05 Å if the main-chain atoms of the helices are used).

Table 7.4.1 Residues used in the structural superpositions of cytochromes c'.

	Helix A	Helix B	Helix C	Helix D
R. rubrum	4-27	38-50	79-95	102-122
R. molischianum	6-29	40-52	82-98	104-124
Chr. vinosum	6-29	39-51	84-100	107-127
Alcaligenes sp	6-29	41-53	79-95	102-122

Table 7.4.2 *Rms* deviations between the "core" elements of the cytochromes c' from *R. rubrum*, *R. molischianum* and *Chr. vinosum* compared with that from *Alcaligenes* sp (numbers of atoms used in the superpositions are in parentheses). (Å)

	R. rubrum	*R. molischianum*	*Chr. Vvnosum*
(i) C_α			
Four helices (75)	1.56	1.44	1.07
Four helices + haem (117)	1.27	1.22	0.97
(ii) Main			
Four helices (375)	1.54	1.43	1.08
Four helices + haem (417)	1.46	1.36	1.05

The structure with the greatest overall difference is that from *R. rubrum* which has an *rms* difference of 1.27 Å for four helices (C_α atoms only) plus the haem group. These differences in the "core" secondary structural elements indicate that although the overall folding is conserved, the packing of the four helices must differ from species to species. This is consistent with the predominance of hydrophobic interactions (and lack of hydrogen bonding interactions) since hydrophobic interactions are less directional and therefore more malleable. To assess which helices are the most structurally conserved between the cytochromes c', individual *rms* deviations were calculated for each each of the four helices, after using all of the "core" elements together (C_α atoms of the helices, plus the haem group; 117 atoms) to define the superposition. Table 7.4.3 gives the results of these comparisons. It can be seen that helices A and D have the smallest *rms*

difference while the non-haem-binding helices B and C superimpose less well; helix B is in fact substantially worse.

Table 7.4.3 *Rms* differences for individual helices after a full structural superposition.

	Helix A	Helix B	Helix C	Helix D
R. rubrum / Alcaligenes sp	1.06 Å	2.56 Å	1.35 Å	0.56 Å
R. molischianum / Alcaligenes sp	1.34 Å	2.26 Å	1.45 Å	0.77 Å
Chr. vinosum / Alcaligenes sp	0.90 Å	1.82 Å	0.68 Å	0.54 Å

Table 7.4.4 compares the *rms* deviations in the "core" structures of all the four cytochrome c' structures. From this table it can be seen that the cytochrome c' structures of *R. Rubrum* and *Chr. vinosum* (excluding that of *Alcaligenes* sp) are the most similar to each other while that of *R. molischianum* cytochrome c' differs the most from the other two.

Table 7.4.4 *Rms* differences between the helices and haem arrangement of the cytochromes c' from *Chr. vinosum, R. rubrum, R. molischianum* and *Alcaligenes* sp.

Helices (C_α atoms) + haem (all atoms)

	ASCC	RRCC	RMCC	CVCC
ASCC	-	1.27 Å	1.22 Å	0.97 Å
RRCC	1.27 Å	-	1.18 Å	1.02 Å
RMCC	1.22 Å	1.18 Å	-	1.26 Å
CVCC	0.97 Å	1.02 Å	1.26 Å	-

To further analyse the packing of the helices in the four structures, helix axes were calculated by least-squares from the C_α positions. Table 7.4.5 gives the inter-helix angles for ASCC, RRCC, RMCC, and CVCC for both the monomer and the dimer. The conclusions are in agreement with the results of the superpositions, ie. that the main differences in structure can be traced to the non-haem binding helices. The largest variations in the inter-helical angles occur between helices B and C where the angle in RRCC is 170.5 ° while that in CVCC is 149.0 °. Another large change occurs between helices B and D where the angle in CVCC is 37.0 ° while that in RRCC is 20.6 °. In Table 7.4.4, a superposition of the helices and haem group of RRCC onto that of CVCC, resulted in an *rms* difference of 1.02 Å which is a little surprising since inspection of the inter-helical angles show large differences in the A-C pair of 10.6 °, the

B-C pair of 21.5 ° and the B-D pair of 16.4 °. This could be explained by the fact that calculations of angles between helices do not require any positional knowledge, and although the helices may superimpose structurally thay may not have similar inter-helix angles. The inter-helical angles for the ASCC and RMCC monomers are fairly similar indicating that the RMCC monomer was the better model to use for molecular replacement.

Table 7.4.5 Inter-helix angles and distances in RRCC, RMCC, CVCC, and ASCC.

Helix pair	RRCC		RMCC		CVCC		ASCC/ADCC	
	angle[a]	distance[b]	angle[a]	distance[b]	angle[a]	distance[b]	angle[a]	distance[b]
[Intra-subunit]								
A-B	155.6	10.0	152.8	9.7	152.0	4.9	160.0	5.8
A-C	28.4	12.2	36.1	12.7	41.0	10.4	35.0	10.3
A-D	154.8	10.3	147.1	8.3	146.0	7.9	149.0	8.5
B-C	170.5	8.2	163.2	7.6	149.0	4.8	163.0	5.1
B-D	20.6	13.9	25.6	11.4	37.0	8.8	22.0	9.6
C-D	166.5	8.2	167.6	7.9	162.0	5.2	164.0	5.0
[Inter-subunit]								
A-A'	157.6	7.9	127.3	6.4	136.0	4.2	133.0	4.7
B-B'	160.4	11.0	151.3	9.6	155.0	5.0	129.0	7.4
C-C'	143.3	27.8	127.5	25.4	120.0	16.5	111.0	24.9
D-D'	122.0	21.1	105.5	18.3	80.0	30.5	86.0	39.1
m^1-m^2 (ref c)	134.2	21.1	105.5	18.3	97.0	19.6	112.0	22.1

(a)- Angles between two helix axes (°).

(b)- Closest C_α-C_α distance between helix axes (Å).

(c)- Angles (°) and closest distances (Å) between the two monomer axes defined as least-squares lines by the C_α atoms of four α-helices in each monomer.

Comparison of the inter-helical angles of the ASCC dimer with those from RMCC and CVCC show that the angles A-B and A-A' are similar. Similarly for the angle B-B' all the values are ~ 155 ° except for ASCC in which the angle is 129.0 °. This probably reflects the difference in helix B orientation in ASCC.

The inter-subunit angle of A-A', i.e. the angle between Helix A in one monomer and its equivalent in the other monomer (Table 7.4.5) seems to fall consistently at a value of approximately 130 °, except for RRCC which has an angle of 157.6 °. RRCC also has a very different helix A - helix A' interface from the others (see later). The other two inter-helical angles C-C' and D-D' show no consistency between any of the structures but this is not surprising as helices A, A', B and B' form the dimer interface.

In Table 7.4.6 the range of inter-helical angles and distances is presented. This further indicates that the main differences in the monomer arise from the differences in helices B and C. The largest range of angle differences occur in the helix pairs B-C and B-D (21.5 ° and 16.4 ° respectively) while the largest range in distances occur between helices A and B (5.0 Å), and B-D (5.1 Å).

Table 7.4.6 Range of inter-helix angles and distances in RRCC, RMCC, CVCC, and ASCC.

Helix pair	Δ angle (°)	Δ distance (Å)
[intra-subunit]		
A-B	8.0	5.1
A-C	12.6	2.4
A-D	8.8	2.4
B-C	21.5	3.4
B-D	16.4	5.1
C-D	5.6	3.2
[inter-subunit]		
A-A'	30.3	3.7
B-B'	31.4	6.0
C-C'	32.3	11.3
D-D'	42.0	20.8
m^1-m^2	37.2	3.8

7.5 Connecting loops.

A four-α-helix bundle requires that the helices are connected by loops. It is in these loops that the main differences in the structures of cytochromes c' occur. There are three loops, of varying size. The largest loop occurs between helices B and C, while the smallest "loop" is the one residue turn occurring between helices C and D. Figures 7.5.1, 7.5.2 and 7.5.3 show superpositions of the AB, BC and CD loops from CVCC, RMCC, and RRCC on that from ASCC.

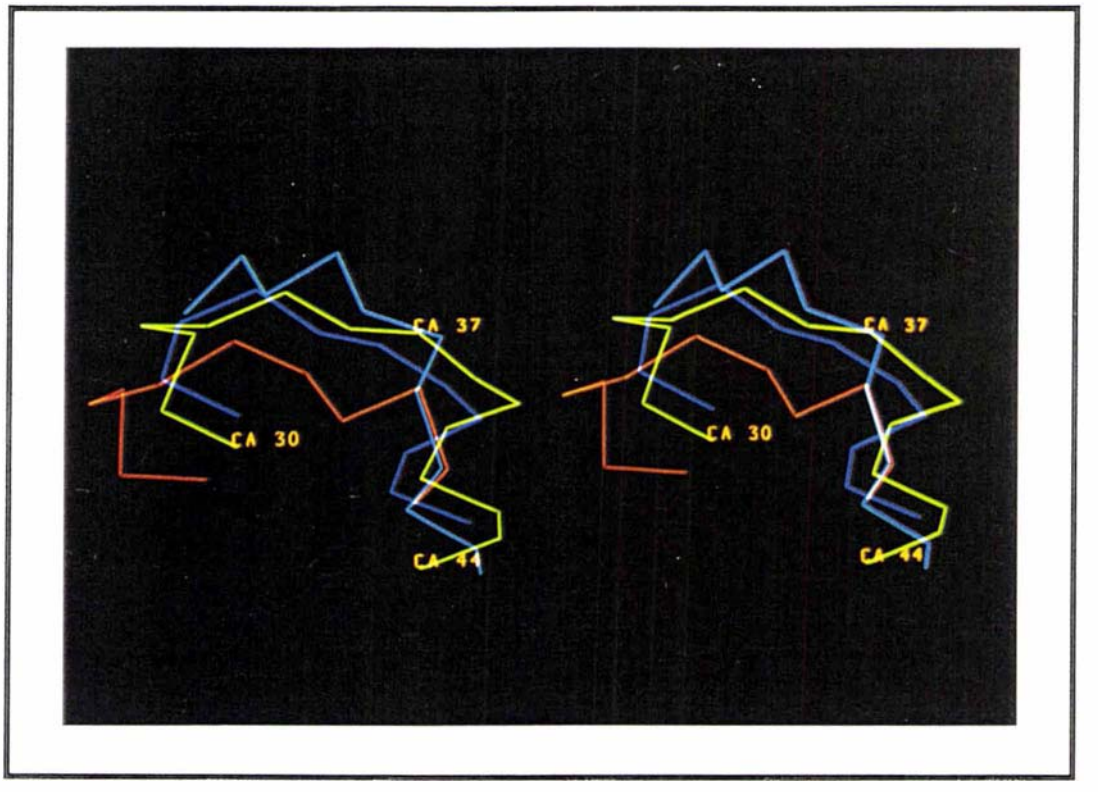


Figure 7.5.1 AB loops from ASCC (green), RMCC (red), CVCC (blue), and RRCC (light blue).

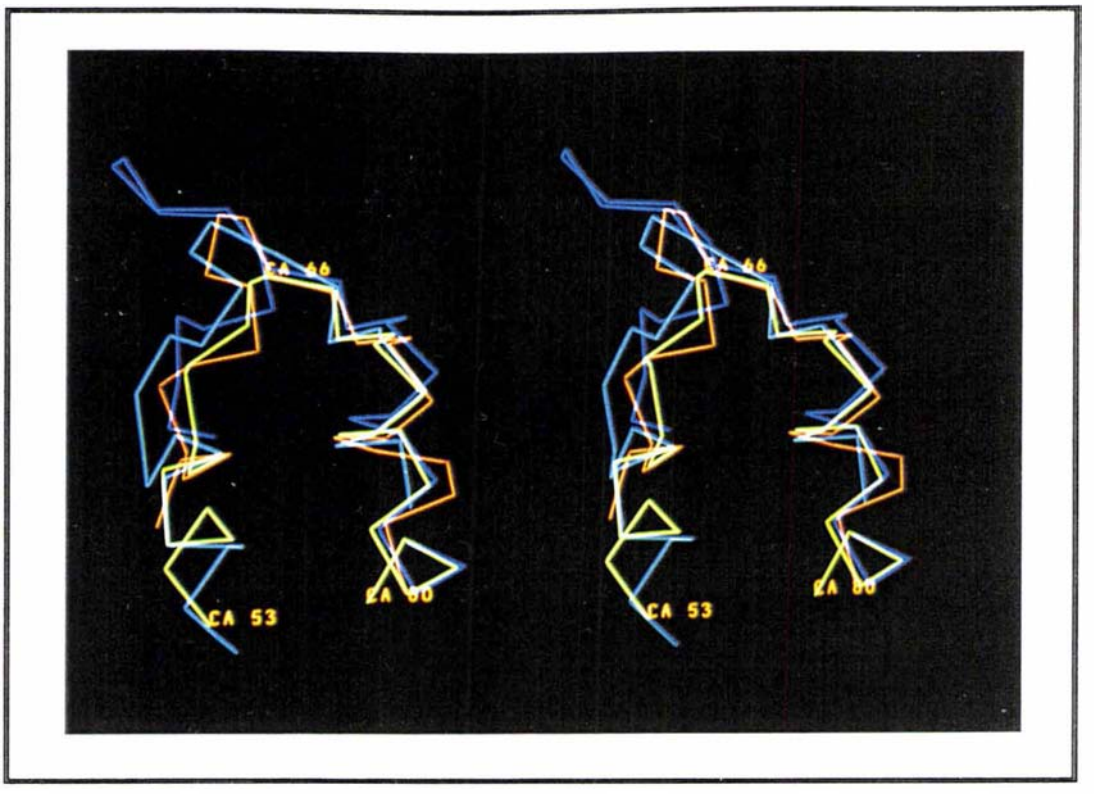


Figure 7.5.2 BC loops from ASCC (green), RMCC (red), CVCC (blue), and RRCC (light blue).

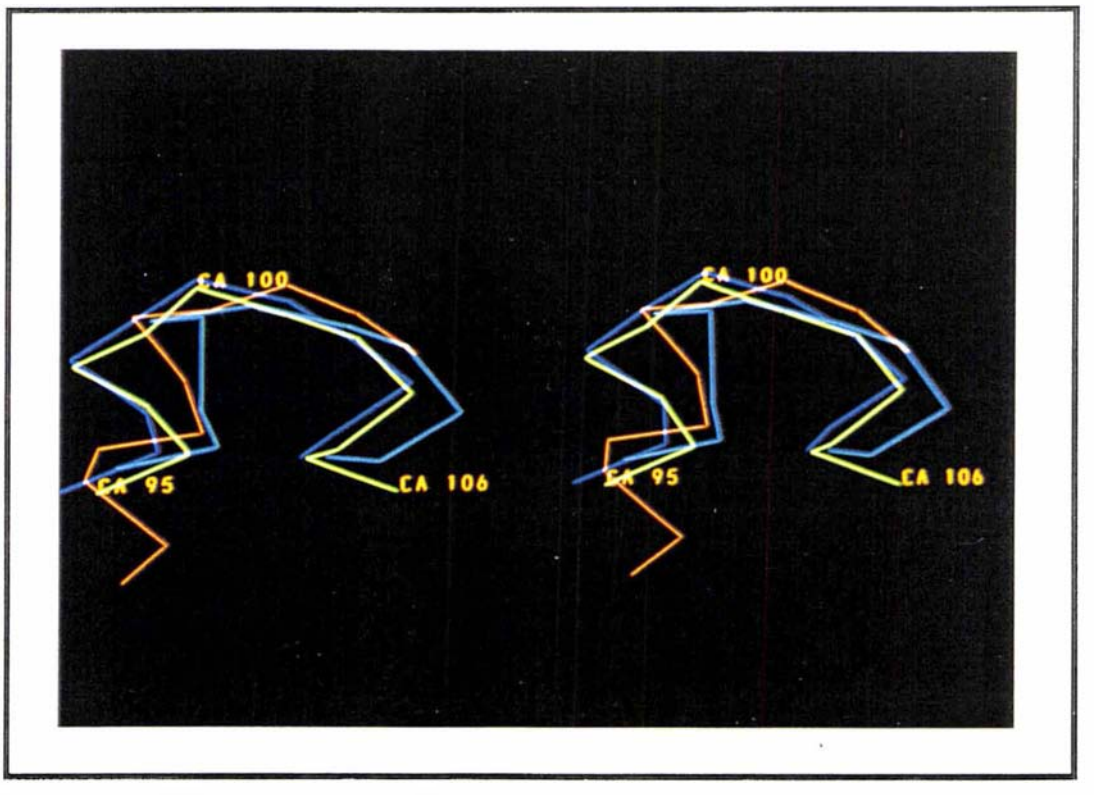


Figure 7.5.3 CD loops from ASCC (green), RMCC (red), CVCC (blue), and RRCC (light blue).

It can be seen from the above superpositions and from the plots of *rms* deviation as a function of residue number (see section 7.2), that the largest differences occur in the AB loop. In particular this may be linked to variations in the length of helix B, with the differences occurring at its N-terminus. On the other hand the BC loop is conformationally quite similar in all the cytochromes c', the *rms* deviations for this loop being basically the same as for the "core" elements such as the helices. In spite of this, the sequence alignment in this chapter (see Figure 7.1.2) shows that it is in the BC loop that most insertions or deletions are accommodated.

The explanation for this apparent contradiction is that the long BC loop has a common framework in all the cytochromes c'. This framework consists of a 3_{10} helix immediately following helix B, two β-turns in the centre of the loop where it contacts the haem, and a second 3_{10} helix immediately preceeding helix C. These elements are present in all four structures (see Table 7.5.1).

Table 7.5.1 Secondary structure within the BC loop (numbers refer to residue numbers). Types of turns follow the notation of Lewis *et al.* (1973).

ASCC	RMCC	RRCC	CVCC
55-58 3_{10} helix	53-58 3_{10} helix	53-56 3_{10} helix	53-58 3_{10} helix
60-63 Type II	59-62 Type II	58-61 Type IV	59-62 Type II'
65-68 Type II'	65-68 Type II	69-72 Type IV	66-69 Type II'
69-74 3_{10} helix	71-76 3_{10} helix	72-75 3_{10} helix	74-80 3_{10} helix

Insertions are accommodated in between the framework elements. For example, in CVCC five extra residues are inserted between what would be residues 64 and 65 if using ASCC numbering (sequence alignment is given in Figure 7.1.2). From the alignment it seems that insertions/deletions also occur in the other cytochrome c' sequences in this region.

7.6 Sidechain hydrogen bonding.

7.6.1 Sidechain...mainchain.

The number of sidechain...mainchain hydrogen bonds varies in number between the species from 14 in RMCC to 35 in CVCC. These interactions have been split into two groups, "local" interactions and "cross-linking" interactions as defined in Baker and Hubbard (1984). Table 7.6.1.1 shows the number of such interactions in the cytochromes c'.

Table 7.6.1.1 Number of "local" and "cross-linking" mainchain...sidechain interactions in cytochromes c'.

Species	No. of "local" interactions	No. of "cross-linking" interactions	Total number of interactions
ASCC	17	5	22
ADCC	13	8	21
RMCC	9	5	14
RRCC	23	7	30
CVCC	29	6	35

From Table 7.6.1 there are clearly more "local" interactions than "cross-linking" interactions. The "cross-linking" interactions are therefore probably not a major stabilising factor in the structure. Conserved mainchain...sidechain hydrogen bonds are associated with conserved residues in cytochromes c', and the latter are invariably involved with haem binding. The most important appear to be the hydrogen bonds involving Arg12; these are illustrated schematically in Figure 7.6.1.1, and are summarised in Table 7.6.1.2.

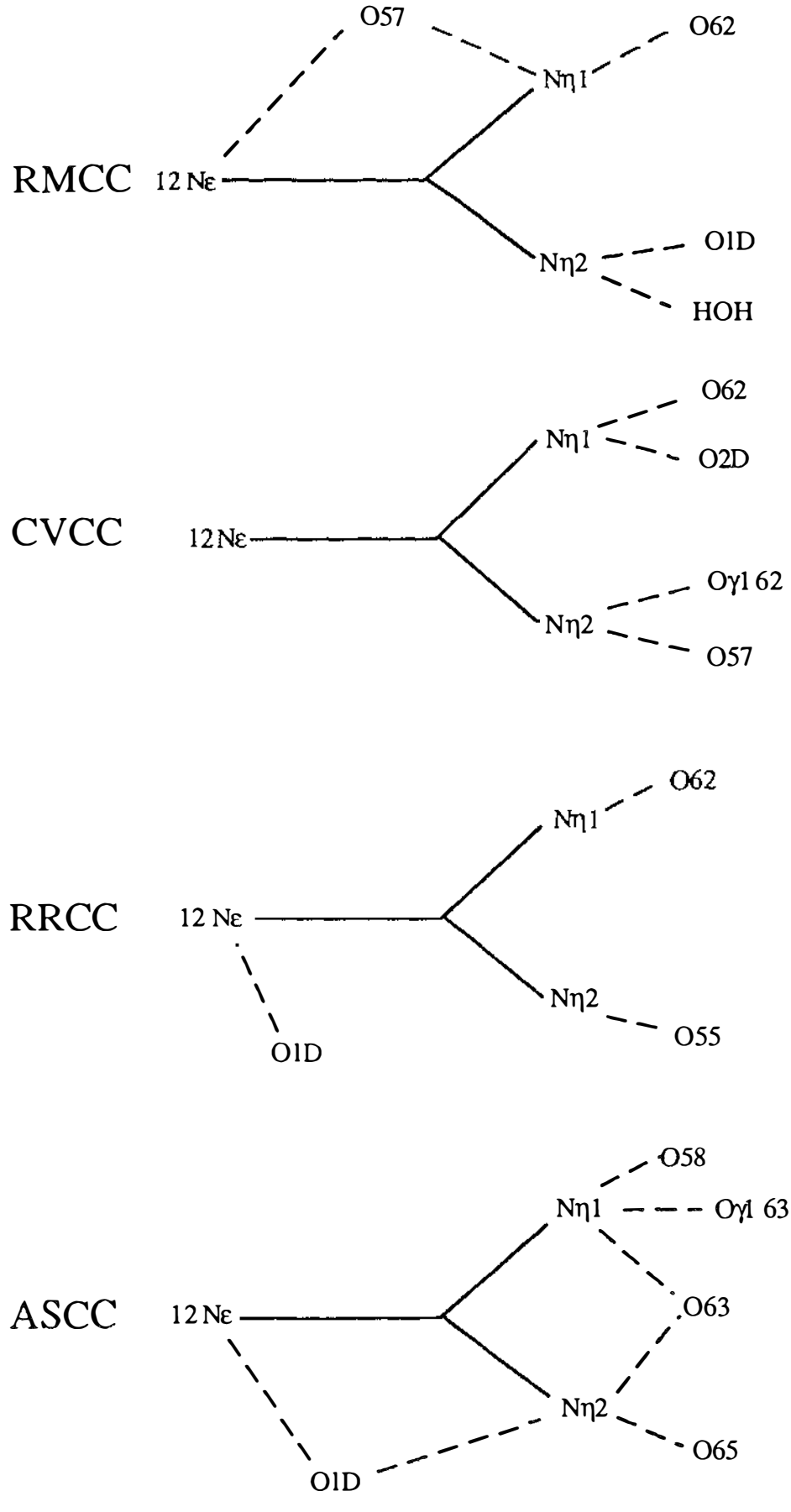


Figure 7.6.1.1 Hydrogen bonding involving Arg 12 in cytochromes c'.

Table 7.6.1.2 Conserved sidechain...mainchain and sidechain...sidchain hydrogen bonds involving Arg12. (* denotes that a hydrogen bond is conserved, the number in parentheses being the corresponding residue number)

ASCC/ADCC	CVCC	RMCC	RRCC
$N_{\eta 1}$ Arg 12...O Ala 58	* (O Leu 57)	* (O Gly 57)	* (O Ala 55)
$N_{\eta 1}$ Arg 12...O Thr 63	* (O Thr 62)	* (O Thr 62)	
$N_{\eta 1}$ Arg 12...Oγ1 Thr 63	(*)[a]		
$N_{\eta 2}$ Arg 12...O Thr 63	* (O Thr 62)		* (O Thr 62)

a - In the CVCC structure the equivalent threonine can actually form the same corresponding hydrogen bond if the sidechain is rotated 180 °.

From Table 7.6.1.2 it can be seen that Arg12 (ASCC/ADCC numbering) consistently forms hydrogen bonds to one of the haem propionate carboxylate oxygens and to a carbonyl oxygen in a structurally conserved part of the BC-loop. An exception is in CVCC, in which the arginine has been modelled in an alternate conformation from that in ASCC/ADCC and the N_{ε} atom does not form a hydrogen bond. The hydrogen bond involving the carbonyl oxygen of residue 58 in the sequence (ASCC/ADCC numbering) is strictly conserved in the other cytochromes c'. From the sequence alignment given in Figure 7.1.1 of this chapter, the identities of these residues also show some conservation in this region, and the formation of the hydrogen bond to residue 58 probably marks the beginning of a structural motif involved in haem binding in cytochromes c'.

7.6.2 Sidechain...sidechain hydrogen bonding.

Sidechain...sidechain interactions present in cytochromes c' are given in Table 7.6.2.1, and the breakdown of "local" and "cross-linking" interactions is shown in Table 7.6.2.2. Again, "local" interactions are those hydrogen bonds which involve groups within four residues of each other and "cross-linking" interactions involve groups more than four residues apart (Baker and Hubbard, 1984).

Table 7.6.2.1 Sidechain...sidechain interactions in cytochromes c'. (shaded boxes indicate monomer-monomer hydrogen bonds)

ASCC/ADCC	RMCC	CVCC	RRCC
O_{η} Tyr11-$N_{\varepsilon 2}$ His22	$O_{\delta 1}$ Asp7-N_{ζ} Lys10	O_{γ} Ser4-$O_{\varepsilon 2}$ Gln6	$O_{\varepsilon 2}$ Glu8-N_{ζ} Lys22
O_{γ} Ser21-$N_{\eta 2}$ Arg25	$O_{\delta 1}$ Asp39-N_{ε} Arg43	$O_{\varepsilon 1}$ Glu7-O_{γ} Ser4	O_{η} Tyr9-$O_{\varepsilon 1}$ Gln37
$N_{\varepsilon 2}$ His22-O_{η} Tyr11	$O_{\delta 2}$ Asp39-$N_{\eta 2}$ Arg43	O_{η} Tyr16-O_{η} Tyr58	N_{ζ} Lys11 -$O_{\gamma 1}$ Thr66
$N_{\delta 1}$ Asn44-$N_{\delta 1}$ His22	$N_{\delta 2}$ Asn46-$O_{\varepsilon 1}$ Gln22	N_{ζ} Lys25-$O_{\delta 1}$ Asn22	O_{γ} Ser15-$N_{\delta 2}$ Asn19
$O_{\gamma 1}$ Thr63-$N_{\eta 1}$ Arg12	$O_{\delta 1}$ Asn46-$N_{\varepsilon 2}$ Gln42	N_{ζ} Lys27-$O_{\varepsilon 1}$ Glu31	$N_{\delta 2}$ Asn19-O_{γ} Ser15
$O_{\delta 2}$ Asp75-O_{γ} Ser78	$N_{\varepsilon 1}$ Trp58-$O_{\gamma 1}$ Thr70	N_{ζ} Lys27-$O_{\varepsilon 2}$ Glu31	O_{η} Tyr20-O_{η} Tyr20
$O_{\varepsilon 1}$ Gln83-O_{γ} Ser115	$O_{\gamma 1}$ Thr62-$N_{\eta 1}$ Arg12	$O_{\varepsilon 1}$ Gln38-$N_{\delta 2}$ Asn29	N_{ζ} Lys22-$O_{\varepsilon 2}$ Glu8
$O_{\delta 1}$ Asn89-N_{ζ} Lys92	N_{ζ} Lys76-$O_{\varepsilon 2}$ Glu124	$O_{\delta 1}$ Asn51-$N_{\delta 2}$ Asn96	$N_{\delta 2}$ Asn36-$O_{\gamma 1}$ Thr38
$N_{\delta 2}$ Gln89-O_{γ} Ser115	$O_{\varepsilon 1}$ Glu92-$N_{\varepsilon 2}$ Gln110	$O_{\gamma 1}$ Thr62-$N_{\eta 2}$ Arg12	$N_{\delta 1}$ Asn36-$O_{\gamma 1}$ Thr38
$O_{\delta 1}$ Asp111-N_{ζ} Lys92		$O_{\delta 2}$ Asp83- O_{η} Tyr128	$N_{\varepsilon 2}$ Gln37-$O_{\gamma 1}$ Thr51
		N_{ζ} Lys86-$O_{\varepsilon 2}$ Gln90	$N_{\varepsilon 2}$ Gln37-O_{γ} Ser 47
		N_{ζ} Lys111-$O_{\varepsilon 2}$ Glu108	$O_{\delta 1}$ Asp40-O_{γ} Ser47
		N_{ζ} Lys122-$O_{\varepsilon 2}$ Glu126	O_{γ} Ser47-$O_{\delta 1}$ Asp40
		N_{ζ} Lys127-$O_{\delta 2}$ Asp83	O_{γ} Ser47-$N_{\varepsilon 2}$ Gln37
			$O_{\gamma 1}$ Thr51-$N_{\varepsilon 2}$ Gln37
			$O_{\varepsilon 1}$ Glu58-$O_{\gamma 1}$ Thr60
			N_{ζ} Lys84-N_{ζ} Lys80
			$N_{\varepsilon 2}$ Gln87-$O_{\delta 2}$ Glu88
			O_{γ} Ser86- $O_{\gamma 1}$ Thr115

From Table 7.6.2.1 it can be seen that there are no conserved side-chain...side-chain interactions in cytochromes c'. The only "cross-linking" sidechain...sidechain bonds formed are between the helix pairs A and B, and C and D.The other feature from this table is that RRCC forms the most sidechain ...sidechain hydrogen bonds with the majority of these involved in stablising the dimer interface.

Table 7.6.2.2 Number of "cross-linking" and "local" interactions formed.

Species	No. of "local" interactions	No. of "cross-linking" interactions	Total number of interactions
ASCC	3	7	10
ADCC	3	7	10
RMCC	4	5	9
RRCC	5	14	19
CVCC	8	6	14

From Table 7.6.2.2 there are few sidechain...sidechain hydrogen bonds overall, and this probably reflects the predominance of hydrophobic interactions in cytochromes c'.

7.6.3 Haem hydrogen bonding.

The only potentially hydrogen bonding groups on the haem are the two carboxylate groups of the propionate sidechains. These are involved in hydrogen bonds to several sidechains which seem to be among the few conserved features of all cytochromes c'. The more buried of the propionate groups (on the D ring) receives hydrogen bonds from the sidechains of Arg12 and Gln13 (ASCC/ADCC numbering) on helix A and one from peptide NH of Asp67 (ASCC/ADCC). The more exposed propionate on ring A generally forms solvent hydrogen bonds as a result of the fact that it is more exposed. In CVCC there are no close contacts to the outermost carboxylic oxygen, but in the other cytochromes c' either a solvent hydrogen bond or a van der Waals contact is formed. A common feature is the bridging water molecule between the two propionate sidechains, which is present in ASCC/ADCC, CVCC and RMCC. In RRCC there is no solvent model but it is likely that a better refined structure would reveal a water molecule in this position. Figure 7.6.3.1 shows schematically the haem hydrogen bonding in cytochromes c', and in Table 7.6.3.1 a list of the haem hydrogen bonds in ASCC/ADCC, CVCC, RMCC and RRCC is given.

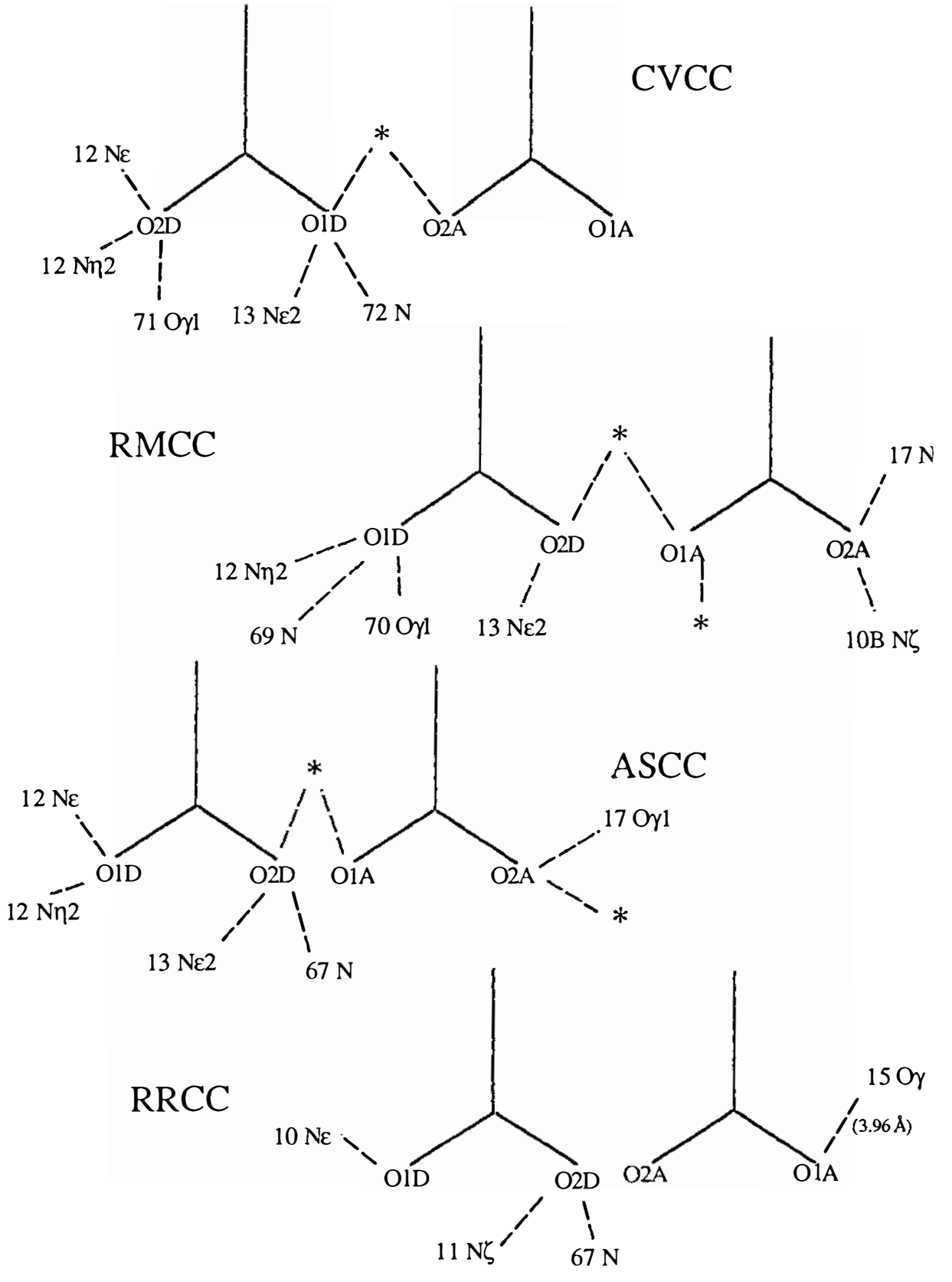


Figure 7.6.3.1 Haem hydrogen bonding.

Table 7.6.3.1 Haem hydrogen bonding. (a * indicates conserved hydrogen bond, a blank means no conserved hydrogen bond)

ASCC/ADCC	CVCC	RMCC	RRCC
O1A Haem .. HOH			
O2A Haem...HOH	*	*	
O1D Haem...$N_{\eta 2}$ Arg 12	*	*	
O1D Haem...N_{ε} Arg 12	(*)[a]	(*)[a]	*
O2D Haem...$N_{\varepsilon 2}$ Gln 13	*	*	* (N_{ζ} lys 11)
O2D Haem...N Asp 67	* (N Arg 72)	* (N Glu 69)	* (N Glu 67)

a - If the the arginine sidechain is remodelled in these structures to what occurs in the ASCC structure, these hydrogen bonds are possible.

From Table 7.6.3.1 the conservation in hydrogen bonding is generally confined to the more buried propionate sidechain. The involvement in RMCC of a sidechain from the other monomer in hydrogen bonding the ring A propionate does not occur in any of the other cytochromes c'. The other hydrogen bonds to the carboxylate oxygens of the propionate sidechain of this ring, are all with solvent molecules. The ring D propionate sidechain shows more conservation of hydrogen bonding. It is interesting to note that N_{ε} Arg12 in ASCC/ADCC and RRCC forms a hydrogen bond to one of the carboxylic oxygens of this propionate sidechain, whereas in RMCC it forms a hydrogen bond to Gly57 (RMCC numbering) and in CVCC it does not form a hydrogen bond.

It is possible remodel the sidechains of Arg12 in these two structures (CVCC and RMCC), with little rearrangement such that the atoms N_{ε}, $N_{\eta 1}$ and $N_{\eta 2}$ are arranged as in the ASCC structure. These hydrogen bonds have been included in Figure 7.6.3.1 and Table 7.6.3.1. In RMCC, turning round the guanidium group of Arg12 gives poorer hydrogen bonds and a steric clash with the carbonyl oxygen in residue 57, possibly due to changes in the C-terminal end of helix B and the start of the BC loop. The formation of a hydrogen bond to the O2D carboxylic oxygen by the peptide nitrogen on residue 67 (ASCC/ADCC numbering) could help in stabilising the BC loop just prior to the 3_{10} helix. The formation of solvent hydrogen bonds with the carboxylic oxygens of the propionate sidechains contrasts with the mitochrondrial

cytochromes c (Salemme, 1977; Mathews, 1985), in which the haem propionate sidechains are completely buried and stabilised by internal hydrogen bonds.

7.7 Haem geometry and environment.

7.7.1 Haem stereochemistry.

The haem is obviously central to the function of the protein, and it also determines many of the protein's physicochemical properties. The structures of cytochromes c' from *R. rubrum*, *R. molischianum*, *Chr. vinosum*, *Alcaligenes* sp and *Alcaligenes denitrificans* now offer five independent views of the haem geometry, four of them at resolutions better than 2.0 Å. The geometry appears to be highly conserved, with favourable side-chain torsional angles and an energetically favourable His-haem interaction. Data for the cytochrome c' structures solved previously are given in Table 7.7.1.1, while data for the structures solved in this study are presented in Table 7.7.1.2.

It can been seen from Tables 7.7.1.1 and 7.7.1.2 that the stereochemistry of the porphyrin system is very much the same thoughout the cytochromes c'. All of the cytochromes c' have their iron displaced from the pyrrole nitrogen plane by ~ 0.30 Å. In contrast with the shifts observed in the low spin six-coordinate cytochromes c, where the iron atoms are essentially in-plane, with displacements of only ~ 0.05 Å from the haem plane (Matsuura *et al.*, 1982; Takano and Dickerson, 1981a; Takano and Dickerson, 1981b). This is due both to the occupation of electrons in the $d_x2\text{-}_y2$ orbital in the high spin cytochromes c' and to the five coordinate geometry of the iron. This displacement from the nitrogen plane has also been found in model porphyrin complexes with admixed spin states (as found in cytochromes c') such as Fe(TPP)($OClO_3$) (Reed *et al.*, 1979) and Fe(OEP)($OClO_3$) (Masuda *et al.*, 1979).

Table 7.7.1.1 Porphyrin stereochemistry and haem ligand binding geometry in RRCC, RMCC, and CVCC.

	R. rubrum	*R. molischianum*	*Chr. vinosum*
(a) Distances (Å)			
Fe to N-plane	0.32	0.22	0.27
Fe to haem plane	0.34	0.23	0.29
Fe to His $N_{\varepsilon 2}$	1.90	2.02	2.02
Fe to N (average)	2.07	2.01	2.04
Fe - N_A	2.06	1.95	2.04
- N_B	2.06	2.04	2.07
- N_C	2.05	1.99	2.07
- N_D	2.10	2.04	2.06
(b) Angles (°)			
His imidazole plane to N_A - N_C - Fe plane.	57.4	42.9	51.8

Table 7.7.1.2 Porphyrin stereochemistry and haem ligand binding geometry of ASCC and ADCC.

	Alcaligenes sp	*Alcaligenes denitrificans*
(a) Distances (Å)		
Fe to N-plane	0.33	0.33
Fe to haem plane	0.32	0.32
Fe to His Nε2	1.94	2.02
Fe to N (average)	1.98	1.99
Fe -N_A	1.97	1.94
-N_B	1.95	2.00
-N_C	1.99	2.01
-N_D	2.02	2.01
(b) Angles (°)		
imidazole plane to N_A-N_C -Fe plane	56.7°	55.2°

Calculation of a least-squares plane through the haem group (defined by the 24 atoms in the porphyrinato core) shows that in ASCC and ADCC the haem exhibits an asymmetrical folding pattern tending towards saddle-shaped but also having a sightly ruffled conformation (pyrrole rings A and C below the plane and B and D above the plane, with the pyrrole rings in a ruffled conformation to one another). In CVCC, the haem is saddle-shaped, as in ADCC and ASCC with pyrrole rings A and C below the plane and B and D above the plane. In RMCC and RRCC although the haem is saddle shaped, pyrrole rings A and C lie above the plane while B and D are below.

In contrast to cytochromes c, myoglobins and haemoglobins, the $N_{\delta 1}$ atom of the axial histidine does not hydrogen bond back to a mainchain carbonyl oxygen atom. In haem proteins when this hydrogen bond occurs it is believed to induce a partial imidazolate character to the histidine, thereby decreasing the ligand field slightly below that of the "free" imidazole (Landrum *et al.*, 1980). The lack of such specific hydrogen bonding in cytochromes c' may result in an imidazole of slightly greater field strength which may bring the spin state into the admixed category (Reed *et al.*, 1981). In ASCC and ADCC a solvent molecule forms a hydrogen bond to this atom, whereas in CVCC, RMCC, and RRCC no interactions are formed (although the $N_{\delta 1}$ aotm is exposed to the solvent).

The dihedral angles between the plane of the axial ligand histidine and the N_A-N_C-Fe vector plane are around 45 °, ranging between 42.9 ° (in RMCC) and 57.4 ° (in RRCC), with ASCC and ADCC close to the latter value. This orientation minimises steric overlap of the pyrrole nitrogens and the histidine imidazole hydrogens (Collins *et al.*, 1972). In reference to the Cys-X-Y-Cys-His haem binding motif, Finzel *et al.* (1985) proposed that different classes of c-type cytochromes can differ in relative haem orientation. In the majority of c-type cytochromes both χ_1 of the cysteine closest to the C-terminus and χ_2 of the histidine prefer approximately staggered conformations (-60 °, -60 °). The alternative arrangement found in one of the haems in cytochrome c_3 (Higuchi *et al.*, 1984) results from the rotation of both sidechains to another sterically acceptable conformation with Cys $\chi_1 = 180$ ° and His $\chi_2 = 180$ °. This arrangement is found in all the cytochrome c' structures (seeTable 7.7.1.3).

Table 7.7.1.3 Alternative χ_1 values for Cys119 and χ_2 values for His120 in cytochromes c' (ASCC/ADCC numbering).

	χ_1 Cys (°)	χ_2 His (°)
ASCC	-173.46	162.33
ADCC	-173.57	161.39
RMCC	-174.05	166.89
RRCC	179.03	171.75
CVCC	-177.06	165.86

Table 7.7.1.4 shows the Fe - Fe distances and haem plane - haem plane angles in the five cytochrome c' dimers. The Fe - Fe distances are very consistent between species with the Fe - Fe distance in CVCC being the shortest, possibly because the orientation of helix B allows closer haem contact. The haem plane - haem plane angles are also reasonably consistent between the species.

Table 7.7.1.4 Fe - Fe distances and haem plane - haem plane angles for ASCC, ADCC, RRCC, RMCC, and CVCC.

	Fe - Fe distance (Å)	Haem plane - haem plane angle (°)
RRCC	25.0	68.0
RMCC	24.0	42.0
CVCC	23.4	32.1
ASCC	25.3	53.0
ADCC	25.2	54.5

In all cytochromes c' the iron is five-coordinate, with four pyrrole nitrogens and one axial histidine as ligands. The sixth axial position is vacant and is only used when cytochromes c' bind ligands such as CO, alkylisocyanides and CN^- (Kassner, 1991; Wood , 1984). Helix A runs along the distal face of the haem, with the sidechain of residue 16 (*Alcaligenes* sp numbering) coming into closest contact with the sixth axial ligand site. In RMCC this position is occupied by Met16, while in RRCC the equivalent

residue is Leu14. In the cytochromes c' from the *Alcaligenes* species the residue in this position is Leu16, which makes a closest approach of 3.42 Å to the haem iron. Table 7.7.1.5 gives the closest distance between the iron and the "sixth" ligand residue (calculated using the program CONTACT, see appendix I).

Table 7.7.1.5 Closest contact distance for the "sixth" ligand in cytochromes c'.

	"Sixth" ligand residue	Contact (Å)
ASCC/ADCC	Leu 16	3.42 ($C_{\delta 1}$...C4D)
CVCC	Tyr 16	3.36 (C_{β}...NA)
RMCC	Met 16	3.72 (S_{δ}...Fe)
RRCC	Leu 14	3.52 (O...C3A)

CVCC is different in that the amino acid which "replaces" the sixth ligand is aromatic, Tyr. It has been proposed that upon ligand binding the plane of the phenol ring has to shift, causing conformational changes in the first helix, that results in dimer dissociation (McRee *et al.*, 1993). McRee *et al.* (1993) also hypothesise that all cytochromes c' having an aromatic residue at this position in the dimer will dissociate upon binding of an extra ligand to the iron atom.

7.7.2 Haem environment.

The environment of the haem in cytochromes c' is quite highly conserved despite the low overall sequence identity. The haem group is situated in a cleft between the four helices of the monomer with one face buried in a pocket lined with aromatic sidechains, and the other face partially exposed to the solvent. In ASCC and ADCC the surface area of the haem exposed to solvent (using a 1.4 Å probe radius) is 104.2 $Å^2$, while in RMCC, CVCC and RRCC the exposed surface areas of the haem groups are 64.2 $Å^2$, 120.6 $Å^2$ and 138.2 $Å^2$ respectively. The solvent accessible surface areas for the proximal histidine residue compare alittle more closely with values of 29.0 $Å^2$, 25.8 $Å^2$, 44.1 $Å^2$ and 18.7 $Å^2$ for ASCC, CVCC, RMCC and RRCC respectively. Tables 7.7.2.1, 7.7.2.2, 7.7.2.3, and 7.7.2.4 present the haem packing contacts in ASCC, RMCC, CVCC, and RRCC respectively (the haem contacts in ADCC are essentially the same as for ASCC).

Table 7.7.2.1 Haem packing contacts in ASCC.

Residue	No. of contacts[a]	Closest distance (Å)	Residue	No. of contacts[a]	Closest distance (Å)
Val 9	1	Cγ1...O1D (4.31)	Ala 68	12	Cβ...O1D (3.76)
Arg 12	7	Nε...O1D (2.76)[b]	Ile 72	1	Cδ1...CMD (3.59)
Gln 13	26	Nε2...O2D (2.66)[b]	Phe 79	4	Cε1...CAD (3.76)
Leu 16	35	Cδ1...C4D (3.42)	Lys 82	3	C...CBC (3.94)
Thr 17	4	Oγ1...O2A (3.21)	Gln 83	3	N...CBC (3.71)
Met 19	1	Cγ...CBB (4.20)	Phe 86	10	Cγ...CMC (3.48)
Ala 20	3	Cβ...CMA (3.75)	Val 112	1	Cγ1...CBB (4.10)
Phe 23	5	Cγ...CBB (4.14)	Ser 115	1	O...CMC (3.96)
Trp 56	22	Cη2...C3C (3.59)	Cys 116	26	Sγ-CAB (1.81)[c]
Phe 59	6	Cε2...CMD (3.90)	Cys 119	17	Sγ-CAC (1.82)[c]
Gly 65	1	O...O1D (3.77)	His 120	40	Nε2-Fe (2.01)[c]
Gly 66	6	C...O2D (3.77)	Tyr 123	2	O...CMD (3.51)
Asp 67	16	N...O2D (2.79)[b]	Arg 124	24	Nη1...C2A (3.48)

a - Sidechain atom...haem atom contacts < 4.5 Å.

b - Hydrogen bonds.

c - Covalent bonds.

Table 7.7.2.2 Haem packing contacts in RMCC.

Residue	No. of contacts[a]	Closest distance (Å)	Residue	No. of contacts[a]	Closest distance (Å)
Leu 9	1	$C_{\delta 1}$...O2A (3.22)	Thr 70	14	$O_{\gamma 1}$...O1D (2.68)[b]
Arg 12	10	$N_{\eta 2}$...O1D (3.05)[b]	Phe 82	8	O...CBC (3.58)
Gln 13	24	$N_{\varepsilon 2}$...O2D (3.11)[b]	Gly 85	6	C_{α}...CBC (3.82)
Met 16	27	S_{δ}...Fe (3.72)	Trp 86	43	$C_{\eta 2}$...C4C (3.33)
Gln 17	8	$N_{\varepsilon 2}$...O2D (2.79)	Leu 89	7	$C_{\delta 1}$...CBB (3.77)
Leu 19	14	$C_{\delta 2}$...C4B (3.57)	Val 117	4	O...CMC (3.80)
Lys 20	3	C_{δ}...CMA (3.77)	Cys 118	24	S_{γ}-CAB (1.78)[c]
Trp 23	6	$C_{\delta 1}$...CMB (3.54)	Cys 121	17	S_{γ}-CAC (1.89)[c]
Met 47	1	C_{ε}...CBB (3.95)	His 122	41	$N_{\varepsilon 2}$-Fe (2.02)[c]
Trp 58	9	$C_{\zeta 1}$...C2D (3.71)	Phe 125	2	O...CMD (3.98)
Ala 68	4	C...O1D (3.82)	Lys 126	17	C_{γ}...C3D (3.61)
Glu 69	20	N...O1D (2.83)[b]	Lys 10B	4	C_{ε}...O2A (3.22)

a - Sidechain atom...haem atom contacts < 4.5 Å.

b - Hydrogen bonds.

c - Covalent bonds.

Table 7.7.2.3 Haem packing contacts in CVCC.

Residue	No. of contacts[a]	Closest distance (Å)	Residue	No. of contacts[a]	Closest distance (Å)
Ile 9	1	$C_{\gamma 2}$...O2D (4.19)	Val 73	8	$C_{\gamma 2}$...CMD (3.74)
Arg 12	8	$N_{\eta 1}$...O2D (2.97)[b]	Ile 87	1	$C_{\gamma 2}$...CBC (4.26)
Gln 13	27	$N_{\varepsilon 2}$...O1D (3.03)[b]	Ala 88	1	N...CBC (4.48)
Tyr 16	91	C_{β}...NA (3.36)	Phe 91	9	$C_{\delta 2}$...CMC (3.58)
Glu 17	4	C_{α}...CMA (4.04)	Val 117	1	$C_{\gamma 1}$...CBB (3.95)
Met 19	7	S_{δ}...CBB (4.17)	Ala 120	1	O...CMC (3.97)
Gly 20	2	C_{α}...CMB (3.95)	Cys 121	24	S_{γ}-CAB (1.86)[c]
Met 23	4	C_{γ}...CBB (4.39)	Cys 124	17	S_{γ}-CAC (1.87)[c]
Ile 46	1	$C_{\gamma 2}$...CBB (4.39)	His 125	40	$N_{\varepsilon 2}$-Fe (2.01)[c]
Trp 58	8	O_{η}...CMD (3.40)	Arg 129	18	$N_{\eta 2}$...NA (3.74)
Thr 71	10	$C_{\gamma 2}$...CMD (3.74)			
Arg 72	12	N...CGD (3.64)			

a - Sidechain atom...haem atom contacts < 4.5 Å.

b - Hydrogen bonds.

c - Covalent bonds.

Table 7.7.2.4 Haem packing contacts in RRCC.

Residue	No. of contacts[a]	Closest distance (Å)	Residue	No. of contacts[a]	Closest distance (Å)
Val 7	2	$C_{\gamma 1}$...O2D (3.55)	Thr 66	9	O...O2D (3.55)
Arg 10	8	N_{ε}...O1D (3.00)[b]	Glu 67	7	C_{α}...O2D (3.68)
Lys 11	22	C_{δ}...O2A (3.33)	Ala 68	7	N...O2D (3.79)
Leu 14	40	O...C3A (3.52)	Phe 79	11	$C_{\varepsilon 1}$...CBC (3.63)
Ser 16	3	O_{γ}...CBA (3.21)	Ala 82	5	O...CMC (3.79)
Thr 17	7	$O_{\gamma 1}$...C2B (3.99)	Ser 83	6	O_{γ}...CBC (3.20)
Ser 18	6	O_{γ}...CMA (3.31)	Ser 86	5	C_{β}...CMC (3.64)
Met 21	3	C_{γ}...CMB (4.03)	Leu 112	2	$C_{\delta 2}$...CBB (3.65)
Ile 45	1	$C_{\varepsilon 2}$...CBB (3.74)	Thr 115	1	O...CMC (3.99)
Met 49	10	S_{δ}...CHC (3.79)	Cys 116	26	S_{γ}-CAB (1.79)[c]
Leu 52	6	$C_{\delta 2}$...C3C (4.00)	Cys 119	18	S_{γ}-CAC (1.78)[c]
Phe 56	6	C2...CMD (3.79)	His 120	46	$N_{\varepsilon 2}$-Fe (1.90)[c]

a - Sidechain atom...haem atom contacts < 4.5 Å.

b - Hydrogen bonds.

c - Covalent bonds.

Table 7.7.2.5 "Conserved" residues involved in haem packing.

ASCC/ADCC	CVCC	RMCC	RRCC
Val 9	Ile 9	Leu 9	Val 7
Arg 12	Arg 12	Arg 12	Arg 10
Gln 13	Gln 13	Gln 13	Lys 11
Leu 16	Tyr 16	Met 16	Leu 14
Thr 17	Glu 17	Gln 17	Ser 15
Met 19	Met 19	Leu 19	Thr 17
Ala 20	Gly 20	Lys 20	Ser 18
Phe 23	Met 23	Trp 23	Met 21
	Ile 46	Met 47	Ile 45
Trp 56			
Phe 59	Tyr 58	Trp 58	Phe 56
Gly 66	Thr 71	Ala 68	Thr 66
Asp 67	Arg 72	Glu 69	Glu 67
Ala 68	Val 73	Thr 70	Ala 68
Phe 79		Phe 82	Phe 79
Lys 82	Ile 87	Gly 85	Ala 82
Gln 83	Ala 88	Trp 86	Ser 83
Phe 86	Phe 91	Leu 89	Ser 86
Val 112	Val 117		Leu 112
Ser 115	Ala 120	Val 117	Thr 115
Cys 116	Cys 121	Cys 118	Cys 116
Cys 119	Cys 124	Cys 121	Cys 119
His 120	His 125	His 122	His 120
		Phe 125	
Arg 124	Arg 129	Lys 126	

From Tables 7.7.2.1, 7.7.2.2, 7.7.2.3, and 7.7.2.4 it can be seen that the haem environment is very hydrophobic. The cytochrome c' making the most close contacts (less than 4.5 Å) is RMCC (314 contacts), while ASCC and ADCC make only 273 contacts. It is obvious, there is a pronounced conservation of hydrophobic contacts, since even if a residue is not conserved in its identity between the cytochrome c' sequences, it is almost invariably replaced by another which can contribute similar

contacts. A second point, related to this, is that the sequence alignment given in Figure 7.1.1 has correctly matched these residues on the basis of the protein folding, without consideration of the haem packing contacts. Table 7.7.2.5 lists the haem packing residues for ASCC/ADCC and their corresponding residues in the other cytochromes c'.

7.8 The dimer interface.

The dimer interface in cytochrome c' structures has the same basic pattern, in which helices A, A', B and B' form an antiparallel bundle similar to that of the monomer. Table 7.8.1 shows the residues which form contacts at the dimer interface.

The most striking feature from the table is the number of hydrogen bonds in the dimer interface of RRCC, where 10 residues from one monomer form hydrogen bonds to the dimer-related monomer. Residues which have sidechains involved in hydrogen bonding are Asp2, Glu8, Tyr9, Asn19, Tyr20, Lys22, Glu30, Gln37, Asp40, and Ser47. This may make the interface in RRCC more rigid than in the other cytochromes c'. In CVCC, McRee *et al.* (1993) report that the interface is formed by residues Thr11, Ala14, Glu17, Phe18, Trp21, and Lys25 in helix A, while helix B contributes Ala41, Val45, Ala48, Ile49, Ser52, Gly53, Met54, and Leu57. The present analysis reveals a slightly different interface composition (see Table 7.8.1). However, it does confirm their hypothesis that the specificity is formed by shape complementarity rather than polar interactions (only three hydrogen bonds).

In RMCC, the interaction of the subunits reflects the hydrophobic interactions between the residues of helices A and B. A salt link between the side-chain of Lys10 and one of the haem propionates is the only non-hydrophobic interaction.

The interface in ASCC and ADCC has less pronounced hydrophobic packing with fewer van der Waal contacts than in CVCC and RMCC. The hydrogen bonds formed at the interface in ASCC and ADCC follow the same pattern as those in the interfaces of CVCC and RMCC, with hydrogen bonding at the edge of the interface.

Overall the RRCC interface is very different from that of the other cytochromes c' with hydrogen bonding providing most of the stabilisation between the two monomers. This could be the reason why the interhelix angle between A and A' for RRCC is approximately 20 ° different from that in the other cytochromes c'. Generally, however, the interface pattern can said to be stabilised by hydrophobic packing, coupled with hydrogen bonding at the periphery of the interface.

Table 7.8.1 Dimer interface contacts in ASCC, RRCC, RMCC, and CVCC. (numbers in parentheses are the number of contacts[a] and the shaded boxes indicate that a hydrogen bond is formed)

ASCC	RMCC	CVCC	RRCC
Pyr1 (7)	Asp7 (16)	Leu3 (2)	Ala1 (11)
Phe2 (3)	Lys10 (18)	Ala14 (9)	Asp2 (16)
Lys10 (2)	Leu11 (28)	Gly15 (8)	Ala5 (7)
Tyr11 (25)	Gln13 (9)	Glu17 (32)	Glu8 (10)
Ser14 (22)	Gly14 (13)	Phe18 (28)	Tyr9 (24)
Ala15 (5)	Leu15 (6)	Trp21 (33)	Ser12 (21)
Thr17 (6)	Gln17 (26)	Lys25 (7)	Val13 (3)
Leu18 (18)	Thr18 (17)	Asn44 (2)	Ser15 (6)
Ser21 (4)	Lys20 (1)	Val45 (10)	Ala16 (11)
His22 (8)	Ser21 (3)	Ala48 (8)	Asn19 (34)
Arg25 (8)	Gln22 (8)	Ile49 (7)	Tyr20 (12)
Val47 (2)	Met49 (16)	Asn51 (1)	Lys22 (11)
Thr50 (3)	Val50 (3)	Ser52 (13)	Ala23 (12)
Leu51 (2)	Lys52 (3)	Gly53 (5)	Ile26 (12)
Leu54 (5)	Leu53 (16)	Met54 (5)	Thr27 (3)
	Ile56 (14)	Leu57 (14)	Gln30 (19)
			Asp31 (2)
			Ala33 (4)
			Val34 (13)
			Pro35 (7)
			Gln37 (18)
			Asp40 (3)
			His41 (5)
			Ala44 (1)
			Ser47 (10)
			Thr51 (16)
			Ala54 (7)
			Ala55 (3)

a - contacts less than 4.5 Å were recorded.

7.9 Solvent structure.

The ASCC model includes 89 water molecules and the ADCC model 75. The CVCC dimer has 221 water molecules included while the RMCC dimer has 212 water molecules. No waters have been placed in the RRCC structure probably because of the limited resolution (2.8 Å) and a reasonably high R-factor (21.4 %) compared to the other structures. The solvent structures of ASCC, ADCC, CVCC, and RMCC were compared by superimposing the structures (and their water molecules) by least-squares methods using the program LSQKAB (CCP4 suite, see appendix I). The residues used in the superpositions (C_α atoms + haem) from each cytochrome c' are given in Table 7.1.1. Of the 89 solvent molecules in ASCC, there were 15 within 2.0 Å of a corresponding solvent molecule in RMCC, and 18 in CVCC. Of these water molecules which "overlap" in atomic position, there are only four which match to within 2.0 Å in all three structures. Figures 7.9.1 and 7.9.2 show the solvent molecules from ASCC/RMCC and ASCC/CVCC which are within 2.0 Å of each other. 14 and 9 solvent molecules from RMCC and CVCC respectively, agree to within 2.0 Å with solvent positions in ADCC. There is only one water molecule which agrees in all three structures, this being the water molecule which is situated bridging between the carboxyl groups of the propionate acid sidechains. Figures 7.9.3 and 7.9.4 show the position of the solvent molecules of ADCC/RMCC and ADCC/CVCC that are within 2.0 Å of each other. It can be seen from Figures 7.9.1, 7.9.2, 7.9.3, and 7.9.4 that these "conserved" water molecules from ASCC/ADCC and/or CVCC and RMCC are localised into two regions. The first region where the majority of the "conserved" water molecules are situated is the groove between helices C and D, from residue 104 to residue 118 in helix D (this corresponds to residue 81 to residue 96 in helix C). The other conserved region is the area in an approximate line between residue 28 in helix A and residue 44 in Helix B. The rest of the "conserved" water molecules are scattered around the remainder of the structure.

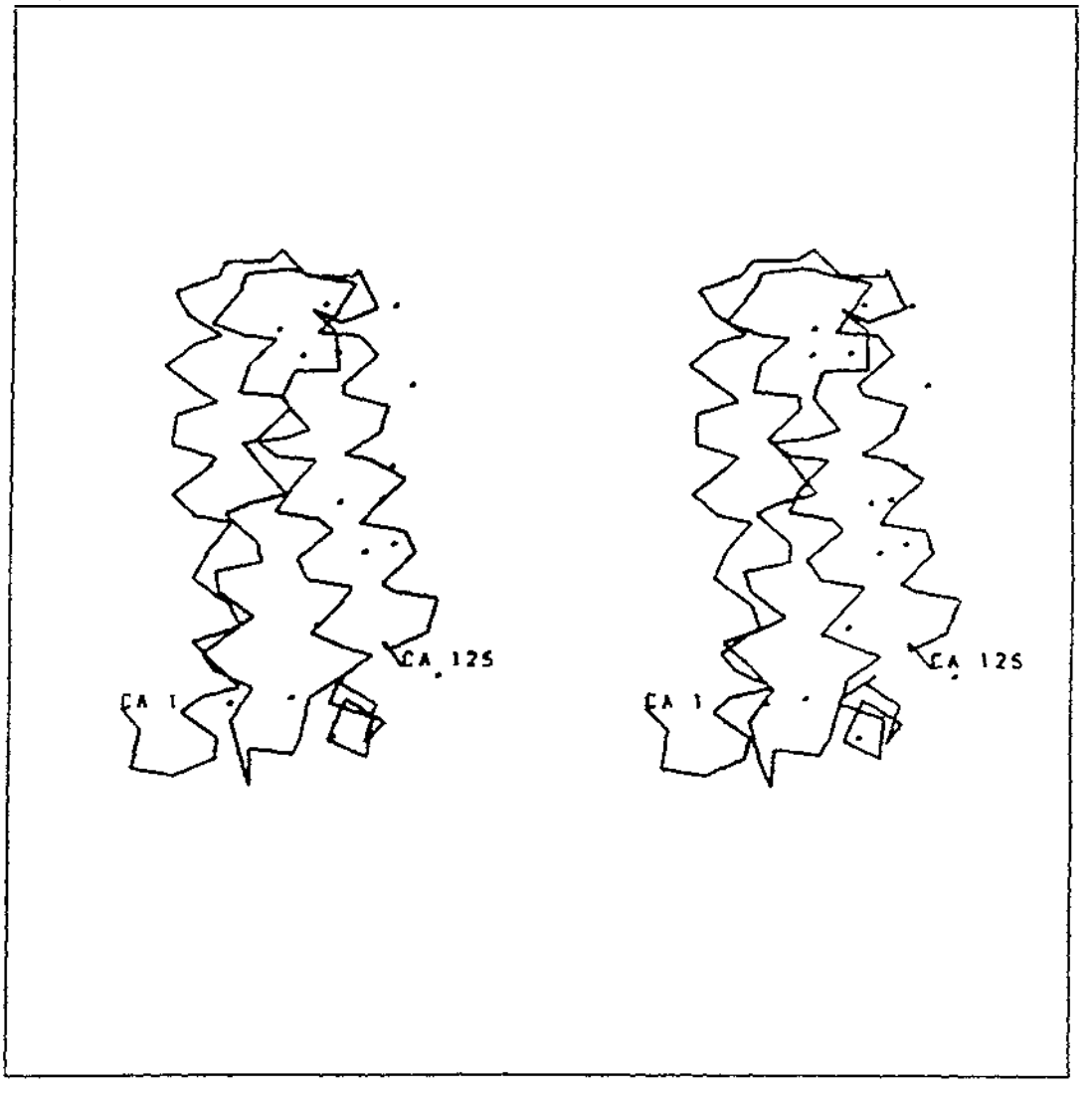


Figure 7.9.1 "Conserved" solvent positions in ASCC/RMCC.

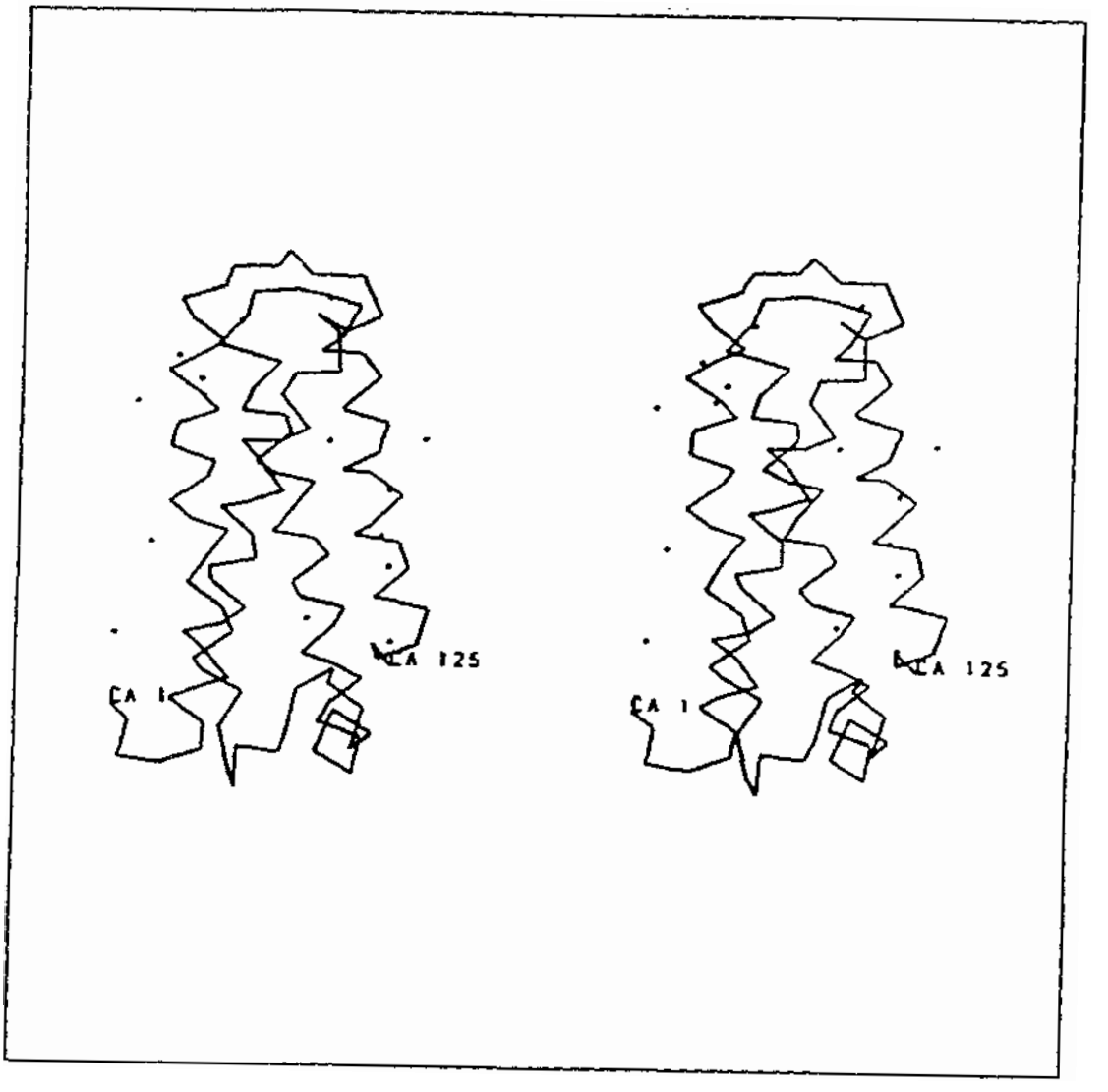


Figure 7.9.2 "Conserved" solvent positions in ASCC/CVCC.

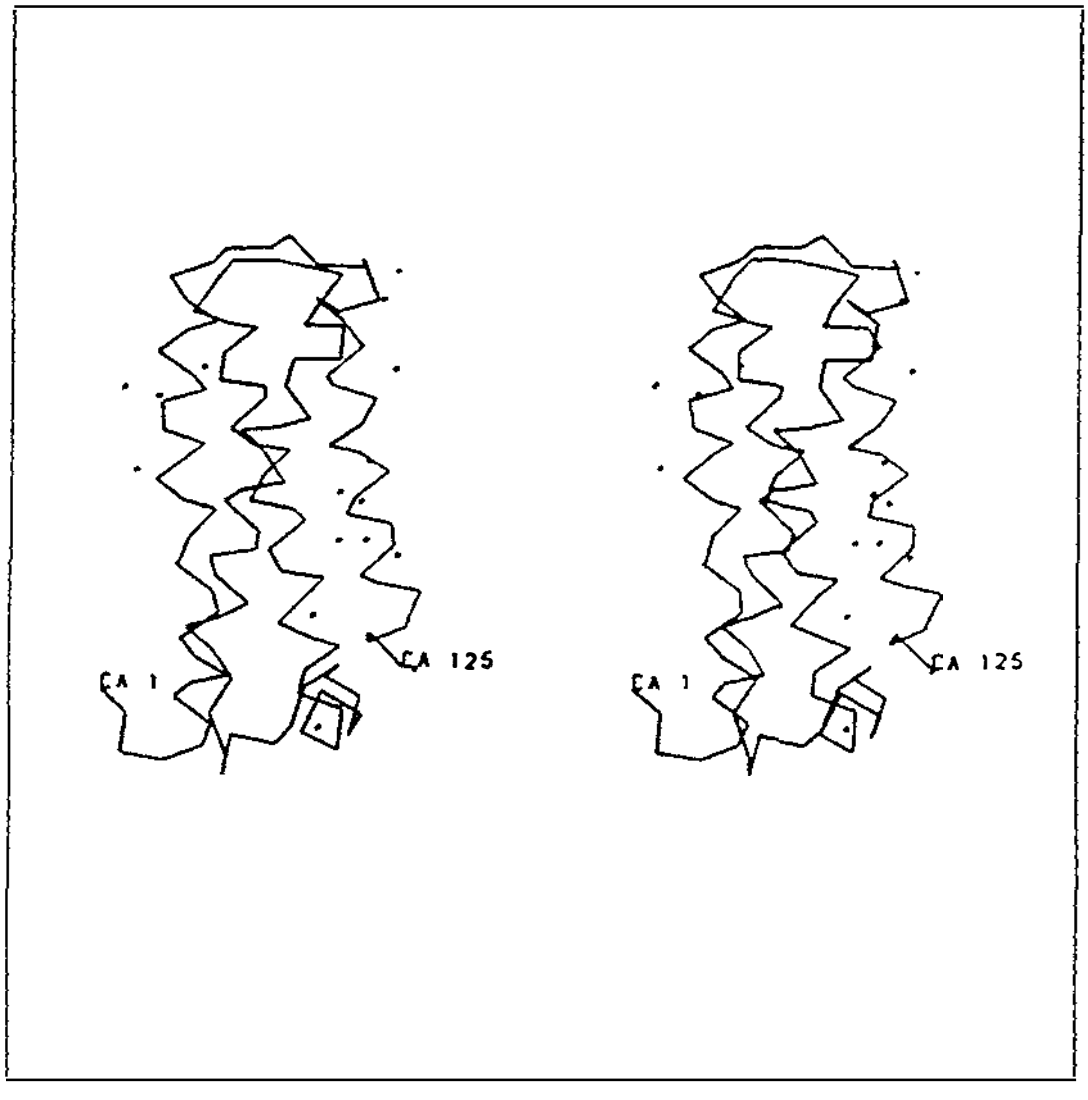


Figure 7.9.3 "Conserved" solvent positions in ADCC/RMCC.

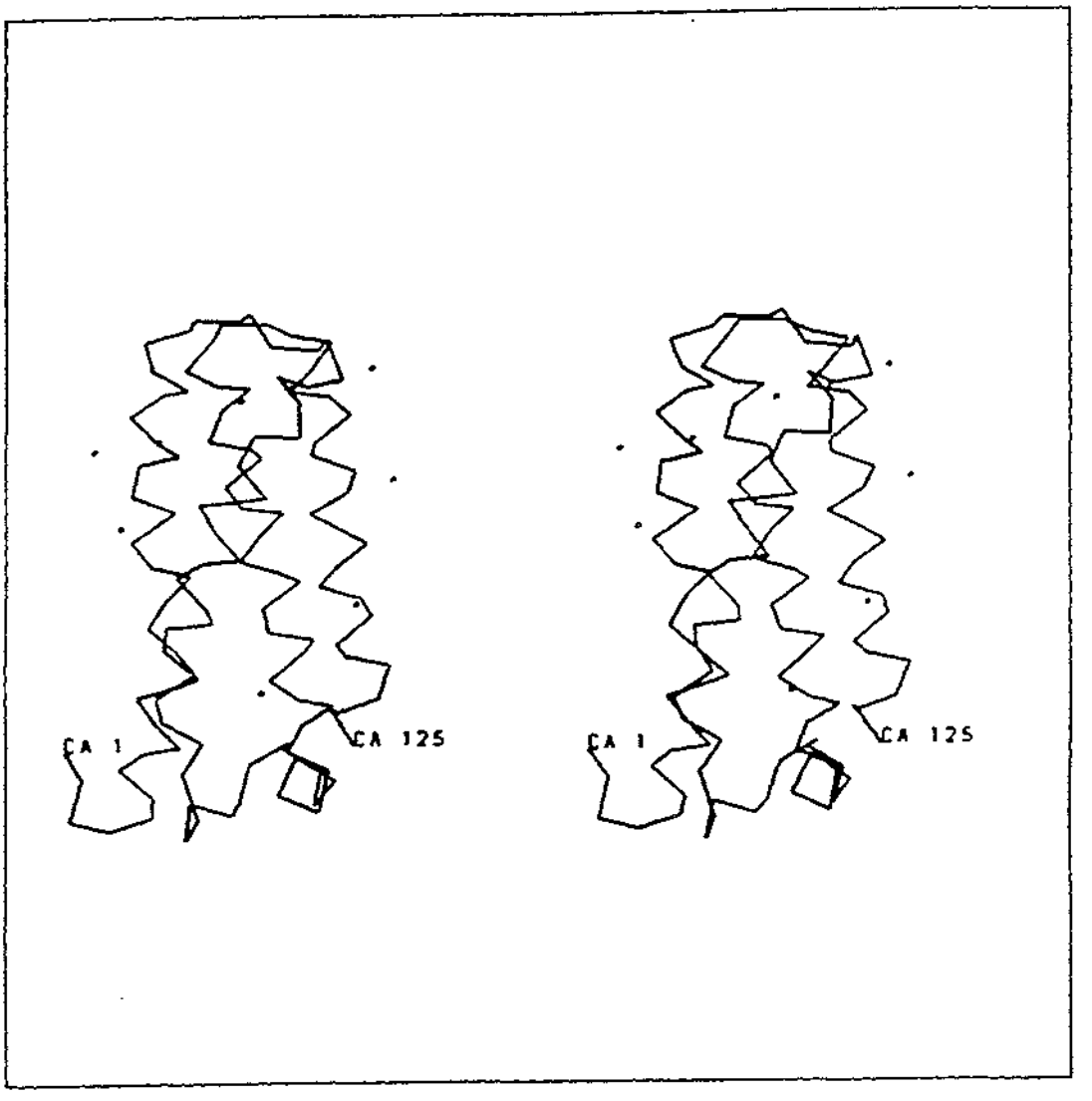


Figure 7.9.4 "Conserved" solvent molecules in ADCC/CVCC.

Chapter Eight

General discussion

This crystallographic study has presented the cytochrome c' structures from two *Alcaligenes* species, *Alcaligenes* sp and *Alcaligenes denitrificans* determined at high resolution. These structures have then been compared with those of three other cytochrome c' structures previously determined. The structure analysis itself proved to be much more difficult than expected, and this chapter is thus split into three sections. One part discusses the structure determination, and the second part discusses structural aspects. Lastly future work that could be undertaken, to resolve or extend some of the questions raised, is discussed.

8.1 Structure determination.

Three aspects of the structure determination deserve special comment;

(i) The failure of the molecular replacement calculations to produce an unequivocal solution was unexpected and points to some unfavourable aspects of the problem,

(ii) The use of additional information from a variety of sources was critical to the final identification of the correct solution,

(iii) The value of even very weak phase information (in this case the iron anomalous scattering) was highlighted.

The failure of the molecular replacement method to give a clear solution can be attributed to a number of factors. The two available search models (the cytochromes c' from *R. rubrum* and *R. molischianum*) had a relatively low sequence identity with the *Alcaligenes* sp cytochrome c' (no sequence has been determined for the *Alcaligenes denitrificans* cytochrome c'), ~ 29 % if pairwise comparisons are made, but only 10 % if all three species are compared. The four-helix bundle structure also has a considerable amount of internal symmetry at low resolution and without sidechains, and this may be the reason why multiple solutions, with similar rotation angles, were often obtained. This was also compounded by the hexagonal

spacegroup. The correct rotation solution using AMORE ($\alpha = 41^\circ, \beta = 56^\circ, \gamma = 243^\circ$) was present among the top solutions in most of the rotation searches, (see Table 3.5.3.1) but only in one case did it lead to the correct translation. Moreover in most rotation searches there were a number of other peaks of similar magnitude, with angles within 10 - 15 ° of the correct rotation solution. It is noteworthy that the truncated/conserved model, in which structurally homologous sidechains were included as far as possible, gave significantly better results than the polyalanine model. The inclusion of some sidechains, in addition to the haem group, may have been essential to break some of the internal symmetry in the molecule. When using ALMN a similar pattern emerged. An approximately correct solution did appear among the highest solutions but was not close enough to the actual correct orientation to give a correct translation solution. Attempts were made to use a homology model, constructed by modelling the *Alcaligenes* sp sidechains into the *R. molischianum* structure, but this was no more successful as a search model.

The extent to which deficiencies in the search model handicapped the molecular replacement searches can be examined in retrospect. Comparison of the final refined *Alcaligenes* sp structure from this study with the two other cytochrome c' structures that were available at that time, show that when the four helices are superimposed on those of the *R. molischianum* and *R. rubrum* structures (71 C_α atoms) the *rms* deviations are 1.56 Å and 1.44 Å respectively. Inclusion of the haem group reduces these figures slightly to 1.27 Å and 1.22 Å, but the agreement is still poor. In addition it appears that helices B and C, in particular, are somewhat differently oriented in the *Alcaligenes* sp structure (see Table 7.2.1). Inclusion of all main chain atoms of the helices, plus the other atoms included in the truncated/conserved search model used for molecular replacement (conserved sidechains and some loop atoms) gives *rms* deviations of 1.77 Å and 1.82 Å for the *R. molischianum* and *R. rubrum* cytochrome c' structures respectively.

The conclusions from these comparisons are that the *Alcaligenes* sp structure differs significantly from <u>both</u> models and that neither model is appreciably better than the other. This is consistent with the failure of either model to give better results in the molecular replacement calculations. It is also clear why the truncated/conserved model gave better results than the polyalanine model as the A-B and B-C inter-helix loops have different conformations in the final structure from those in either search model.

Using the final *Alcaligenes* sp structure as search model (with all sidechain atoms included) gives a clear rotation solution (peak of 4.1σ, with the next highest peak 2.4σ). The correct rotation solution is also given by a truncated/conserved model (3.5σ; next highest 2.7σ) and a polyalanine model (3.0σ; next highest 2.0 σ),

and all three models give the correct translation, with peak heights 7.9σ, 6.8σ and 5.3σ respectively (next highest 3.6σ, 2.9σ and 2.3σ). The peak heights for the correct rotation solutions are smaller than would be expected (Dodson, 1985). During the molecular replacement attempts using ALMN, the small peak heights even for those solutions which were near the correct orientation, cast doubt on the reliability of these solutions. Although the symmetry of the four-helix bundle may have added an extra complication (especially in the rotation function), the main conclusion is that the primary source of difficulty in the molecular replacement analysis was the substantial deviation of both search models from the target structure.

In circumstances where no clear molecular replacement solution emerges, as in the present study, it was important to use as much additional information as possible, to reduce the number of possible solutions. This may include the use of a variety of search models, if more than one is available, and any chemical or other information. In the cytochrome c' analysis, the knowledge of the iron position and the ability to place the monomer-monomer two-fold axis were essential in selecting out potential solutions that could be tested. They were still not sufficient, however, to define an unequivocal solution. In particular, an experiment in which the iron atom of one of the search models was placed on the true iron position (derived from the anomalous difference Patterson) and the model was rotated round this position, still did not give a convincing solution. In retrospect it may perhaps have been possible to place the search model so that the iron atom was correctly positioned and orient the monomer so that the crystallographic two-fold created a reasonable dimer. The problem, though, of verifying the solution and refining it satisfactorily may still have remained. Given the significant deviations of the actual structure from either of the search models, especially in the inter-helix loops, it may not have been possible to refine from such an approximate solution.

The key factor in identifying the correct solution in the present case (once the number of solutions that were feasible had been reduced to manageable proportions) was the phase information that could be derived solely from the iron anomalous scattering. The anomalously-phased map as shown in Figures 5.2.1 and 5.2.2 is of remarkably good quality, given the weakness of the original phase information, and this provided a powerful external test of the MR solutions. Crucial to making the most of this information were the noise reduction that resulted from the combination of the electron density maps for the two different (but isomorphous) species of cytochrome c' and the use of the density modification procedures in the program SQUASH. Averaging of poorly-phased electron density maps for the same protein in different unit cells has previously been used and density modification methods

have been widely used to help solve other protein structures (Varghese *et al.*, 1983; Moore *et al.*, 1994). A special feature in this study was that the phasing came from a single position and in such a case a considerable proportion of reflections had no phase information at all (because the iron anomalous contribution is too small to be significant). It is this class of reflection that can be rescued by density modification. Furthermore, an advantage when phasing from anomalous differences, is that reflections of this type are randomly distributed through reciprocal space, not concentrated at higher resolution as is the case when phasing by isomorphous differences, as in the single isomorphous replacement method.

There have been other cases in which intrinsic metal ions have been located from their anomalous scattering at a single wavelength and then used to help solve the structures. In cytochrome b_{562} (a 110 aa protein of approximately 12,000 Da molecular weight) the iron positions were obtained from anomalous difference Pattersons using the scattering of the iron atoms (Czerwinski and Mathews, 1974). Then, as in this present study, native protein phases were computed from the observed anomalous scattering. These initial phases were combined with those from a heavy atom derivative and an electron density map calculated (Mathews *et al.*, 1979). Other iron-containing proteins in which the anomalous scattering has been used for phase derivation include the four-haem cytochrome c_3 from *Desulfovibrio desulfuricans* (Pierrot *et al.*, 1982). In this instance the iron positions were determined and preliminary electron density maps calculated, prior to combining the phases from the iron scattering with phases derived from heavy atom derivatives. The [4Fe-4S] ferredoxin from *Bacillus thermoproteolyticus* had phases calculated from the [4Fe-4S] cluster (Fukuyama *et al.*, 1988), and in the high-potential iron-sulphur protein isolated from *Rhodocyclus tenuis* initialphases were calculated from the iron-sulphur cluster and combined with phases from heavy atom derivatives (Rayment *et al.*, 1992).

A more powerful method for making use of anomalous scattering in phase determination and structure solution is the multiwavelength anomalous dispersion (MAD) method (Hendrickson, 1984; Hendrickson, 1985; Smith *et al.*, 1987). Proteins whose structures have been determined using MAD include the basic "blue" copper protein from cucumbers (Guss *et al.*, 1988) and crambin (Hendrickson and Teeter, 1981). The "blue" copper protein obtained from cucumber seedlings is a small (96 residue) copper containing protein of approximately 10,100 Da molecular weight which is involved in electron transfer, while crambin is a sulphur containing protein of 46 residues with an unknown functional role. In both of these cases, anomalous scattering has been used to derive phase information, using the technique of multiple wavelength anomalous dispersion. In contrast to the present study, which

used phase information from a single wavelength, MAD uses the wavelength dependence of the anomalous scatterer over a series of wavelengths. The anomalous contribution to the atomic scattering factor is a function of the atomic absorption coefficient for the anomalously scattering element. The absorption coefficient must be measured at the absorption edge of the element and at some distance from the edge. The wavelengths are chosen to optimise the difference in intensity between the Bijvoet pairs and between the diffraction at the selected wavelengths. Fundamentally, the protein should contain an element that gives a sufficiently strong anomalous signal. One Se atom (atomic number 34) in a protein of not more than approximately 150 amino acid residues is sufficient for a successful application of MAD phasing (Hendrickson *et al.*, 1990; Leahy *et al.*, 1992), but it is also dependent on the quality of the data set. Iron-containing proteins whose structures have been determined using MAD include erythrocruorin (Hoppe and Jakubowski, 1975) and lamprey haemoglobin (Hendrickson *et al.*, 1988).

The use of the iron atom of the *Alcaligenes* cytochromes c' in MAD phasing certainly could have provided an initial set of phases from which the structure could have been solved. In the *R. rubrum* cytochrome c' structure analysis, phasing was derived from MIR methods. Earlier in their study, however, the positions of the two irons were determined using MAD data (Harada *et al.*, 1986). The iron atom in cytochrome c' is certainly a strong enough anomalous scatterer, and this, combined with the small size of the protein (127 residues), suggests that MAD would have been beneficial in this study. Unfortunately no beam-time at a suitable synchrotron radiation facility could be obtained when needed.

8.2 Structural aspects.

Functional role of cytochromes c'

Two obvious characteristics distinguish the subclass family of cytochromes c' from cytochromes c, i) their iron coordination (five-coordinate in cytochromes c' as opposed to six-coordinate in cytochromes c) and ii) the positioning of the covalent haem-binding sequence (near the N-terminus for cytochromes c and near the C-terminus for cytochromes c'). Despite their differences in iron coordination and method of haem attachment they are both proposed to have a role in electron transfer. If the function is electron transfer there should be one or more binding sites for their electron transfer partners.

In the case of cytochromes c, the surface area around the haem edge has been implicated as the interaction site for other redox proteins. In mitochondrial cytochromes c, the lysines present in the structure are clustered predominantly around the front face of the exposed haem edge. It has been proposed that it is these charged sidechains which play a role in the formation of the cytochrome c:oxidoreductase complexes (Salemme, 1977; Mathews, 1985; Pettigrew and Moore, 1987). The lysines are presumed to orient the molecule such that the haem prosthetic group makes direct and alternate interactions with the prosthetic groups of the oxidoreductases (Salemme, 1977; Mathews, 1985; Pettigrew and Moore, 1987).

The 'blue' copper electron transfer proteins azurin and plastocyanin offer different kinds of recognition sites. Azurin has a hydrophobic patch (Norris *et al.*, 1983), surrounding the exposed edge of His117, consisting of 5 Met, 1 Phe, 1 Pro, 1 Trp and 1 Ala residues. Five of these residues are invariant in azurin sequences while the other four remain hydrophobic. The conserved and hydrophobic nature suggests a binding surface either for cytochrome c-551 or for cytochrome oxidase. Electron transfer could then occur *via* the exposed haem edge of the cytochrome to the imidazole ring of His117 and hence to the Cu atom. The involvement of this hydrophobic patch was confirmed by Van Pouderoyen *et al.* (1994) who showed that mutation of one of the methionine residues to a lysine inhibited electron transfer.

Plastocyanin (Guss and Freeman, 1983) has both a hydrophobic patch and an acidic patch for binding electron transfer partners. The hydrophobic patch consists of 3 Leu, 2 Ala, 4 Pro, 1 Phe and 1 Val residues. Residues which form the hydrophobic patch have their sidechains (or part of them) in contact with the solvent. The acidic patch is concentrated in two kinks of the polypeptide, with the negatively charged

sidechains of six residues (consisting of 3 Asp and 3 Glu) directed into the solvent, forming an elongated acidic patch.

A direct demonstration of the use of recognition sites on a 'blue' copper protein is given by the recent crystal structure analysis of a ternary electron transfer complex (Chen *et al.*, 1994) comprising a quinoprotein (methylamine dehydrogenase), a 'blue' copper protein (amicyanin) and cytochrome c-551i. In this complex the iron in the haem group of the cytochrome is 24.8 Å from the copper atom. This is a longer distance than in the analogous co-factor distance. For example, the distance from the iron to the centre of the flavin ring in the flavocytochrome b_2 is about 16 Å (Xia and Mathews, 1992). The three redox centres that are present in the ternary complex are in a straight line. The interface between the cytochrome and amicyanin has been analysed and found that it is more charged than the other interface present and less hydrophobic in nature.

No direct crystallographic study of an electron transfer complex involving cytochrome c has been determined, but it would not be unreasonable to suggest that cytochrome c' might bind with electron transfer partners in similar fashion to the proposed electrostatic complexes that have been modelled with ferrocytochrome c (Salemme 1976; Poulos and Kraut, 1980). In these complexes the planes of the haems of the two partners are approximately parallel in each case. The interhaem distance and Fe - Fe separation are 8.5 Å and 21.6 Å repectively in the proposed cytochrome c:cytochrome b_5 complex, and 17 - 18 Å and 25 - 26 Å respectively in the proposed cytochrome c:cytochrome c peroxidase complex. In both of these complexes, it is the lysine residues of the cytochrome c which are believed to form the majority of the charge interactions with oppositely charged residues on the electron transfer partner.

The two *Alcaligenes* cytochrome c' structures studied here do not have an obvious ring of lysines around the periphery of the haem face, so it is difficult to determine surface residues that possibly could be involved in electron transport. In the *Alcaligenes* cytochromes c' there do not seem to be any hydrophobic or acidic patches. However, an analysis of the exposed haem edge identifies a concentrated area that includes several charged residues with sidechains that are exposed to the solvent. These are Lys4, 10, 117, 125, His120 and Arg124. Two of these residues have disordered sidechains (Lys 4 and 117) but it maybe this mobility which could provide non-specific recognition of a redox partner. Interaction of these charged residues with residues on an electron transfer partner might then allow the haem groups of the respective cytochromes (if a cytochrome is the partner) to be in close enough proximity to allow the transfer of electrons. However, a more detailed

electrostatic analysis of the protein would allow a more conclusive theory to be put forward.

The internal packing of residues around the haem group in cytochromes c' is similar to that of the cytochromes c with the groups in van der Waals contact being bulky hydrophobic sidechains. Throughout all the mitochondrial cytochromes c, sequence comparisons show that the residues which are in contact with the haem are either invariant or conservatively replaced (Mathews, 1985; Pettigrew and Moore, 1990). There are 11 aromatic residues in total, and of these, four are totally invariant, and another three residues are always aromatic. The remaining four aromatic residues appear to play a structural role, providing bulk and hydrophobicity for the haem groups. In the cytochromes c' a slightly similar pattern is found to that present in cytochromes c. Both cytochromes have hydrophobic residues around the haem, including a number of aromatic residues. However in contrast to cytochrome c, the hydrophobic residues in cytochrome c' are not as highly conserved in either position or identity. The aromatic residues that pack round the haem in cytochrome c' are frequently contributed by different parts of the structure. This suggests that in cytochrome c', although the environment needs to be hydrophobic, the exact packing around the haem group is much less important than in cytochrome c. This may be because cytochromes c have a more highly conserved function than cytochrome c'.

Cytochromes c' bind CO, alkylisocyanides and CN^- with rate constants and binding affinities with values which are 10^2 to 10^6 times smaller than those of other high-spin haem proteins such as the cytochrome peroxidases (Kassner, 1991). This decreased affinity for ligands is probably due to the steric interactions at the haem coordination site where helix A is positioned such that residue 16 (*Alcaligenes* sp numbering) has its sidechain projected towards the sixth coordination site. The distances between the sidechain of this residue and the haem varies between 3.54 Å and 3.74 Å in the different cytochromes c' (see Table 7.7.1.5) implying that the sidechain would have to move to allow a ligand to bind. The *Chr. vinosum* cytochrome c', which has the shortest contact with the haem iron actually, undergoes dimer dissociation upon addition of a sixth ligand (McRee *et al.*, 1993). The addition of this ligand is assumed to require a conformational change in helix A to remove steric hindrance (McRee *et al.*, 1993) thereby causing the dimer to dissociate.

In contrast, in sperm whale myoglobin (Takano, 1977a; Takano, 1977b) the distal histidine is 6.06 Å from the iron. In the myoglobin mutants (Quillan *et al.*, 1993) in which the distal histidine has been replaced with various other amino acids, the distance between the iron atom and the sidechain in the sixth ligand position are 5.87 Å for the His->Leu mutant, and 5.95 Å for the carbon monoxide derivative of

the His->Leu mutant (remembering that the *Alcaligenes* cytochromes c' also have a leucine in an analogous position to the myoglobin mutant).

When CO or NO is bound in haem proteins such as myoglobin (Kuriyan *et al.*, 1986) and erythrocurorin (Steigemann and Weber, 1979) the bound ligand takes up a slightly bent conformation in which the Fe-C-O angle is between approximately 120 ° and 140 °. For the sperm whale myoglobin mutants ,this angle is more linear ranging between 150 ° and 170 °. An important point to note is that in the leucine mutant of myoglobin there is no displacement of the sidechain upon binding of the sixth ligand due to the distance between the sidechain of this residue and the iron atom. To achieve a comparable distance (~ 6.0 Å) in the *Alcaligenes* sp cytochrome c', the leucine sidechain has to be placed into a gauche (-) conformation which gives a distance of 5.03 Å from the iron. However, this conformation causes a steric clash with atoms in residue 13 of helix A. If the sidechain is then placed in a trans conformation (180 °) a clash occurs with atoms of pyrrole ring A in the haem group. Although no crystallographic structure of cytochrome c' with a sixth ligand bound has been determined, it is probable from the stereochemistry of the sixth site, and from other six-coordinate structures that, with either CO or NO bound, the conformation of the ligand may tend to that which is found in the wild-type myoglobin as a result of steric reasons. The suggested Fe-C-O angle would therefore tend to be in the range 120 ° to 140 °.

Their ability to bind physiological compounds such as CO and NO could suggest a role in their removal from the body. Williams (1991) suggested that cytochromes c' play a role in maintaining low concentrations of NO by signalling cells to the presence of NO, thus implying that the cytochromes c' are connected to the chemotaxis system. However, their slow ligand binding properties, mentioned above, suggest that ligand binding is probably not their primary function, and that they do probably function primarily as electron transfer proteins rather than in the removal of toxic compounds.

Predominance of hydrophobic interactions.

The number of "cross-linking" hydrogen bonds which are found in the cytochrome c' structures is very small (five in each case). The majority of these involve residues which are conserved throughout the sequences of cytochromes c', and are involved with haem binding. There are thus very few "cross-linking" hydrogen bonds between different helices and those that do exist join helix

"hairpins", ie. link helix A to helix B or helix C to helix D. This raises the question of how the overall topology is maintained when there are so very few hydrogen bonds to define the relative orientation of sequentially different parts of the structure. The answer lies in the fact that the stability of the four-α-helix bundle comes from the hydrophobic interactions rather than from the more directional hydrogen bonding. This presumably is the reason why so few residues are absolutely conserved (hydrophobic interactions are less specific and more malleable, so replacements are possible). Likewise the variations in the helix packing and helix orientation can be traced to the same factor.

The hydrophobic surface that is buried inside a folded protein contributes directly to the stabilisation energy of the molecule. The sequence of *Alcaligenes* sp cytochrome c' contains a high proportion of non-polar residues such as valine and leucine. The latter are a feature of most all-α proteins, and especially those existing as a four-α-helix bundle (Cohen and Parry, 1990; Paliakasis and Kokkinidis, 1992). The helices in this four-α-helix bundle are also arranged anti-parallel so that the helix dipoles stabilise each other (Hol *et al.*, 1978; Hol, 1985).

Relationships with designed four-helix bundles.

A number of attempts to design and engineer four α-helix bundle proteins (DeGrado *et al.*, 1989; Cohen and Parry, 1990; Hecht *et al.*, 1990; Morii *et al.*, 1991) have been reported. With the exception of the protein Felix (Hecht *et al.*, 1990) these do, however, deviate significantly from the natural ones by having small numbers of different amino acids. Paliakasis and Kokkinidis (1992) pointed out that a four-α-helix bundle is characterised by a pattern of hydrophilic and hydrophobic residues which is repeated every seven residues, this point having also been made earlier by Crick (1953) and by Cohen and Parry (1990). The core positions at positions a and d (see Figure 8.2.1) of these heptads consist overwhemingly of the amino acids leucine and alanine. The rationale behind the design of helix bundles is that a polypeptide is devised that is capable of forming amphiphilic α-helices, that can form a tetrameric aggregrate with the hydrophobic leucine sidechains projecting toward the interior of the tetramer and the hydrophilic residues projecting towards the aqueous exterior. The arrangement of the leucine residues is normally based on the packing that is observed in naturally occurring four-helix bundles (Chothia *et al.*, 1977; Chothia, 1984). The hydrophilic sidechains (such as glutamic acid and lysine) are arranged so they can form favourable ion pairs along one face of the helix.

Designed helices often have negatively charged residues at the N-terminus and positively charged residues at the C-terminus to stabilise the helix formation by favourably interacting with the helical dipoles (Hol, 1985). In the design strategy of Regan and DeGrado (1989) the sequences linking the helices contained a single proline residue to aid in helix termination, as well as two arginine residues to promote a reversal in the overall peptide chain direction .

The sequence of the *Alcaligenes* sp cytochrome c' (Ambler, 1973) has a large proportion of alanine and leucine residues present, which should be enough to identify it as an all-α protein. This is confirmed in the present study. Figure 8.2.1 shows a helical wheel representation of the residues invloved in the α-helix structure (adapted from Cohen and Parry, 1990). It can be seen from Figure 8.2.1 that the structure of cytochrome c' meets the design principles for synthetic four-α-helix bundles very well. Hydrophobic interactions between residues at the positions a and d dominate the bundle, with very few polar residues occurring in these positions. Also there are very few hydrophobic residues present on the outside of the structure in contact with the solvent. Those hydrophobic residues which do occur on the outside of the helices are involved in interactions of the dimer interface with residues of the symmetry related dimer molecule. Further stabilisation can occur with interactions between residues at positions e - e and g - g (Cohen and Parry, 1990) but this does not occur in this case.

The structural analysis of proteins such as this, which forms a four-α-helix bundle, has provided a framework for the design of different four-α-helical proteins (DeGrado *et al.*, 1989; Cohen and Parry, 1990). This has given an improved understanding of protein folding and should lead to a more rational *de novo* design of proteins which satisfy the requirements of a particular topology.

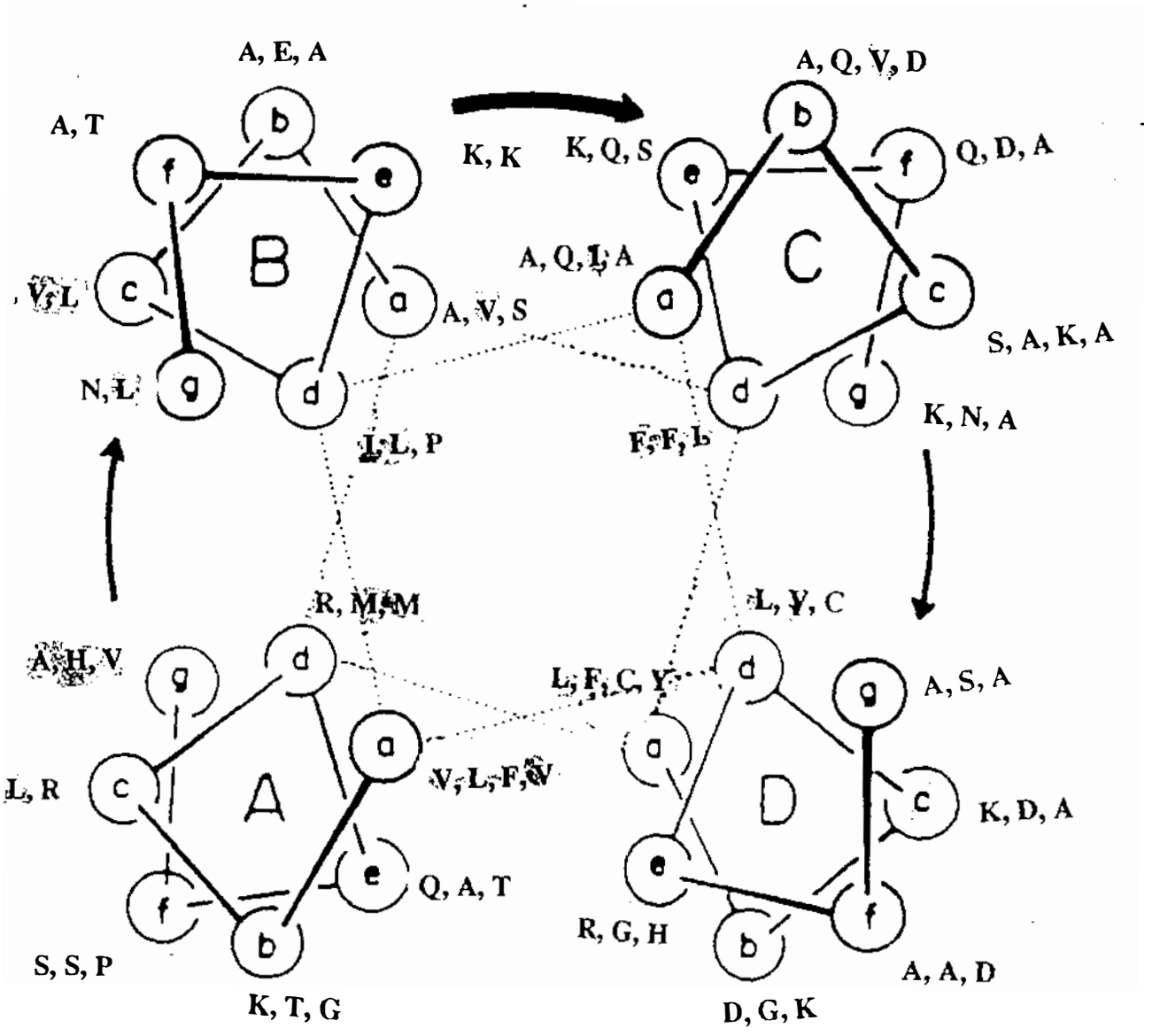


Figure 8.2.1 Helical wheel representation of the residues involved in helices in the *Alcaligenes* sp cytochrome c' (shaded are the hydrophobic residues).

Modified N-terminus.

The presence of a blocked N-terminus, consisting of a pyroglutamic acid (cyclised glutamic acid), in both of the *Alcaligenes* species, is unusual and has not been found in any other cytochrome c' structures. It does, however, occur in cytochromes c (Ambler, 1980; Pettigrew and Moore, 1987). The blocked N-terminus could play a part in the stabilisation of helix A by nullifying the positive charge that would normally be found at the N-terminus and which could destabilise the first helix by repulsion between the α-amino group and the helix N-terminus. This is one of the suggestions that was made by Regan and DeGrado (1989) for stabilising α-helix bundles. It is perhaps significant that cytochrome c also has an important helix (carrying the haem attachment site) beginning close to the N-terminus. This could be a reason why cytochromes c often have blocked N-termini. Alternatively, Awadé *et al.* (1994) have proposed that the occurrence of an pyroglutamic acid at the N-terminus of polypeptides may minimise their degradation, or provide them with a particular function. The N-terminal residue also plays a role in the dimer interface of cytochromes c' by sealing off the ends. This is not true of the other cytochromes c' however, suggesting that is not a major reason for the presence of a blocked N-terminus.

Haem stereochemistry.

The haem binding in cytochromes c' differs from that in cytochromes c due to the alternative conformation of the C-terminal cysteine and histidine residues (Finzel *et al.*, 1985). Both sidechains are rotated into an alternative conformation with Cys119 $\chi_1 = \sim 180$ ° and His120 $\chi_2 = \sim 180$ ° (as compared with values of - 60 ° and - 60 ° respectively in cytochromes c). The effect of this change is that the sidechain maintains an energetically favourable histidine - haem interaction. In this orientation the $N_{\delta 1}$ atom of His 120 cannot hydrogen bond to the mainchain carbonyl oxygen of the previous turn of the helix (as in cytochrome c and myoglobin). To do so would lead to breakage of the Fe-N(His) bond.

The $N_{\delta 1}$ atom of the His ligand in the cytochromes c' from *Alcaligenes* sp and *Alcaligenes denitrificans* instead hydrogen bonds to what has been modelled as a water molecule. This is consistent with a proposed explanation for the mixed high-

spin/intermediate-spin state of the iron atom in cytochromes c' (Maltempo, 1974; Maltempo and Moss, 1976; Maltempo *et al.*, 1974). It has been shown that an intermediate spin state depends on the presence of an unusually weak axial field ligand (Scheidt and Reed, 1981). This is shown in Figure 8.2.2. In general, increased ligand field strength is correlated with the extent of deprotonation of the imidazole $N_{\delta 1}$ nitrogen (Landrum *et al.*, 1980) and this is in turn influenced by hydrogen bonding interactions. Hydrogen bonding of His $N_{\delta 1}$ to a strong proton acceptor, such as a carbonyl oxygen, as in myoglobin (Takano, 1977a; Takano, 1977b), enhances the ligand field strength of the imidazole group. Where the hydrogen bond partner is a weaker proton acceptor, such as a water molecule as in cytochrome c', the ligand field strength is expected to be correspondingly weaker. Following these arguments, and comparing only five-coordinate species, deoxymyoglobin has a weak enough ligand field to be high spin. Cytochrome c' has an even weaker ligand field (due to the weaker hydrogen bonding of the His ligand), and therefore it adopts the mixed spin state. In support of this argument, the mixed spin state of cytochrome c' only occurs at a neutral pH. At high pH a transition to a pure high spin state occurs, and this is assumed to be due to the deprotonation of the His ligand.

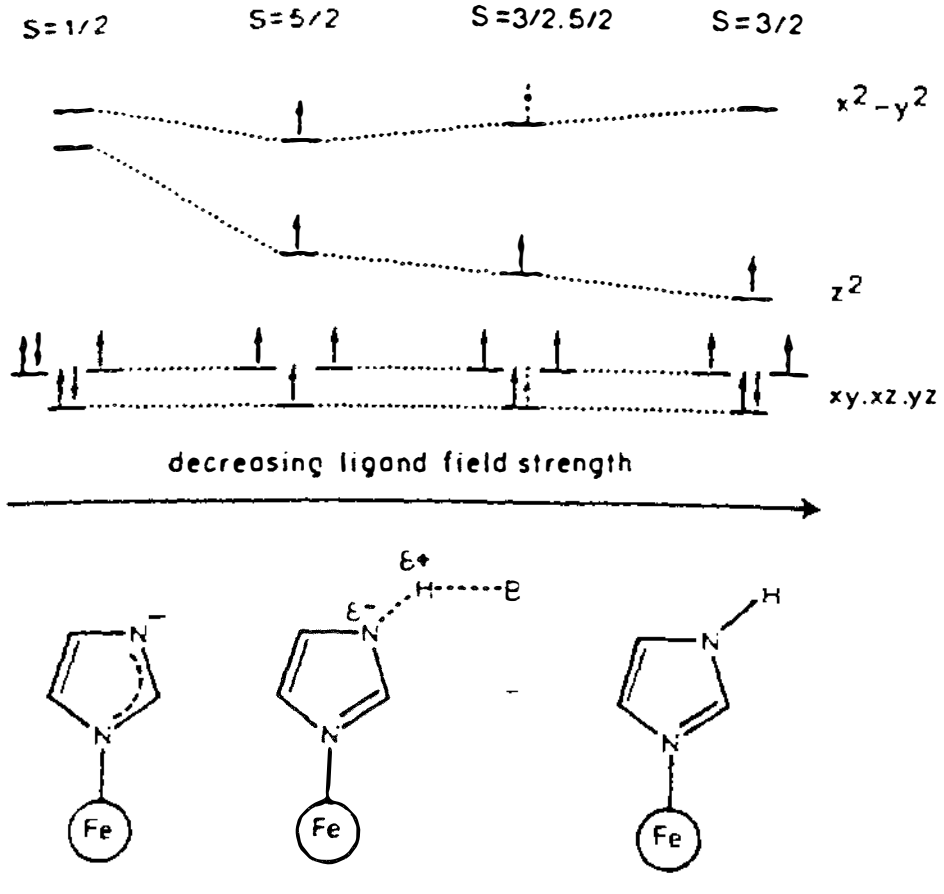


Figure 8.2.2 Correlation between spin state and ligand field strength in ferric porphyrins (from Weber, 1982).

Although the above explanation is plausible, the present structures of *Alcaligenes* sp and *Alcaligenes denitrificans* cytochromec' suggests an alternative reason for the weak ligand field strength of the His (and hence for the mixed spin state). In section 6.5 it was noted that the sidechain of Arg124 lies approximately co-planar to that of the haem group (lying above ring A), and perpendicular to the imidazole ring of the axial His residue (approximately 3.6 Å from the centre of the imidazole ring). The orientation is such that amino-aromatic hydrogen bonding between Arg124 and the imidazole ring seems possible. The $N_{\varepsilon 1}$ atom of His120, when covalently bound to iron is a strong donor of its lone pair of electrons, but this should leave the imidazole ring with a slightly positive charge ($\delta+$). The presence of a positive sidechain (eg. Arg124), so close to His120 and within amino-aromatic hydrogen bonding distance, should lessen the donation of the electrons from the His to the iron, thereby providing a possible reason for the mixed spin-state properties in cytochromes c'. Two further observations are consistent with this hypothesis. Firstly, in the cytochrome c' sequences there is always a Lys or Arg in this position next to the His (Ambler, 1981; Pettigrew and Moore, 1990). Secondly, as noted above, the exsistence of the mixed spin state is dependent upon the pH (Weber, 1982). Depending on the cytochrome c' the mixed spin state is shifted to the high spin state at pH values varying between 7.1 and 9.0 (Bartsch, 1978). At this pH the guanidine sidechain of the Arg124 present in this study would be protonated, further giving credence to this theory.

8.3 Future work.

This study brings the total number of X-ray structures of cytochromes c' solved to five. All have been of the oxidised native protein, however, and many other aspects should be explored if their properties are to be fully analysed.

Site directed mutagenesis

All structures apart from the *Chr. vinosum* cytochrome c', have an aliphatic residue at the "sixth" ligand coordination site (*Chr. vinosum* has a tyrosine at this position). McRee *et al.* (1993) have proposed that it is the presence of an aromatic sidechain in this position that causes the dimer to dissociate upon ligand binding to the vacant sixth coordination site. A most interesting study would be site directed mutagenesis of the residue that is present opposite the vacant sixth coordination site, from the wild-type sequence (leucine), which is present in both of the *Alcaligenes* species, to various residues, and the comparison with a resulting X-ray structure. Mutations which could be tried include Leu16His (as in six coordinate haem proteins) to see if the residue in the distal position in cytochromes c' is in fact able to bind, or Leu16Tyr as in the *Chr. vinosum* cytochrome c' to see whether the *Alcaligenes* dimers could also be induced to dissociate. The proposed role of the Arg that is present next to the axial His (discussed above), could be tested by mutation of this residue to a non-basic residue and analysis of the magnetic properties. Lastly the effects of mutating the aromatic residues, which make up the bulk of the hydrophobic interactions in the protein interior near the haem, and substituting them with hydrophilic residues or other hydrophobic residues with shorter sidechains, could be undertaken to see what (if any) structural changes occur in the interior.

Bound ligand in sixth coordination position

A crystal structure analysis of one of the *Alcaligenes* cytochromes c' with a ligand such as CO bound at the sixth coordination site would allow analysis of the conformational changes that the four-α-helix bundle undergoes when the iron

becomes hexacoordinate. Binding of another ligand (eg. CO) could be achieved either by seeing if crystals of native oxidised cytochrome c' placed in a saturated CO environment would absorb any CO, or by crystallising the protein in an inert environment consisting of CO. This would test the suggestion of McRee *et al.* (1993) that helix A must undergo a conformational change if a sixth ligand is allowed to bind. It could also provide answers to the question of why some of the cytochrome c' dimers dissocciate when an extra ligand binds, and whether the sidechain of the residue opposing the vacant sixth ligand site is displaced, thereby resulting in a conformational change in helix A.

Removal of modified N-terminus

The removal of the pyroglutamic acid at the N-terminus would be interesting to check the effect this blocking group has on the stability of the first helix, and any changes in the packing or shape of the dimer interface that could result from this removal. Mutation of this residue into an alanine could be tried to see if any changes in the dimer interface do occur when the blocking group is changed.

Reduction of iron

No structural analysis has yet been carried out on a cytochrome c' in its reduced form. If the true function of cytochromes c' is indeed electron transfer, it would be expected that reduction should be achieved with minimum change to the iron coordination geometry (Guss *et al.*, 1986; Shepard, 1991). It is likely that the iron in the protein could be reduced by treatment of the existing cytochrome c' crystals with a reducing agent such as sodium dithionite.

Appendix I

The following are brief descriptions of the computer programs and packages used during the course of these studies, and referred to in the text.

I.1 CCP4 (Collaborative Computational Project 4) Suite Programs.

The CCP4 suite is a collection of programs, associated data, and subroutine libraries which can be used for macromolecular structure determination (SERC (UK) Collaborative Computing project 4, Daresbury Laboratory, UK, 1979).

ALMN

A fast rotation function program which works in Eulerian angles. An extension of the Crowther rotation program by Eleanor Dodson (York University). The program calculates rotation function overlap values using FFT techniques. The position of all peaks is given for all symmetry related positions in both Eulerian and polar angles.

AMORE (Navaza, J. (1994) AMoRe: an automated package for Molecular Replacement. *Acta. Cryst.* **A50**, 157-163.)

Program which contains all the routines to run a complete molecular replacement analysis. It reformats the native data, generates structure factors from the model, calculates the rotation function, and allows for rigid body refinement. Also incorporates a translation function. The main characteristics of this package are;
(i) adequate functions are computed by powerful and fast algorithms.
(ii) many potential solutions are explored
(iii) the information coming from models already positioned is automatically incorporated into the procedure
(iv) the correlation coefficient is used as the main criterion of selection
(v) there is a high degree of automation

BAVERAGE

Program used to read a PDB file, tabulate the average B-values residue by residue (mainchain and sidechain separately) and the *rms* deviations of the B-values from this mean.

CONTACT

Program for computing various types of contacts in protein structures. Can also analyse water hydrogen bonding. Program calculates the hydrogen position for those nitrogen atoms for which the hydrogen position is unambiguous. The angles O...H-N and donor...O-C are calculated. Intermolecular contacts can be calculated by the addition of a symmetry card.

DISTANG

A general purpose program for calculating bond distances and angles within a protein molecule, and between symmetry equivalent molecules (if using a symmetry card) within a crystal lattice.

ECALC

Used to calculate normalised structure amplitudes for input to molecular replacement and direct methods programs, and to calculate origin-removed Patterson coefficients. Removes reflections with very large structure amplitudes, which might dominate calculations.

ENVELOPE

Program produces a molecular envelope from an averaged map generated using a reciprocal space equivalent of B.C. Wang's algorithm. The output file is then used in FLATMAP.

FLATMAP

This program is used to flatten solvent regions of the map and optionally attenuate the negative electron densities. After adding the equivalent of F(000) to the electron density at all grid points, the density at grid points in the solvent region is set to the average solvent density, and negative densities are truncated.

FFT

The FFT (Fast Fourier Transform) program may be used to calculate Fourier, difference Fourier, Patterson and difference Patterson maps from reflection data. The program has been developed by Birkbeck College to allow for the generation of equivalent reflections using spacegroup symmetry elements. Also allows the input of selected data items from a single binary reflection data file so that a number of different types of Fourier transforms may be calculated without having to prepare special files.

HELIXANG

Calculates contacts, angles and separation between helices using least-squares methods.

LSQKAB

Optimises the fit of a subset of atomic coordinates from one file to the same subset of another file. The program assumes both sets of coordinates are in protein data bank format. The centroid of the working subset is first moved to that of the reference set and then a rotation matrix is derived. Output options include a list of all differences between atom pairs, differences between atom pairs greater than the *rms* difference or output of a PDB file of coordinates.

MLPHARE (Otwinowski, Z.O. (1991). Maximum Likelihood refinement of heavy atom parameters. In: Proceedings of the CCP4 study weekend, edited by Wolf, W., Evans, P.R. and Leslie, A.G.W., pp80-86, SERC Daresbury Laboratory, Warrington, UK.)

This program refines heavy atom positions, occupancies, B-values and error estimates, then uses these refined parameters to generate phase information using the "maximum likelihood" approach.

OVERLAPMAP

Overlapmap reads in two maps (or three if appropriate) and either:
(a) averages them,
(b) excludes points present in one map from the second map and outputs the rest of the second map,
(c) correlates the density.

PROCHECK (Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.)

Assesses the overall stereochemical quality of a given protein structure, as compared with well-refined structures at the same resolution, and gives an indication of its local, residue-by-residue, reliability.

PROLSQ (Konnert, J.H. and Hendrickson, W.A. (1980) Incorporation of stereochemical information into crystallographic refinement. In "Computing in Crystallography" (Diamond, R., Ramasehan, S. andVenkatesan, K. eds.) pp13.01-13.23, India Academy of Sciences, Bangalore, India.)

The Hendrickson-Konnert program is used for restrained least-squares refinement of a protein structure. The program PROTIN analyses the protein geometry and produces an output file containing geometrical restraint information. SFALL calculates X-ray contribution to the matrix using FFT methods. PROLSQ combines these two contributions and outputs optimum shifts for xyzB for each atom.

SCALEIT

The program SCALEIT calculates and applies a derivative-to-native scaling function using either (a) an overall scale factor, (b) a scale and isotropic temperature factor, or (c) a scale and anisotropic temperature factor. The program has the facility to refine scale factors, to apply input scale factors or just to analyse the agreement between derivative and native amplitudes.

SFALL

This program calculates structure factors and X-ray gradients for refinement using inverse and forward fast Fourier techniques. If coordinates are input it builds an "atom map" on a suitable grid over the asymmetric unit of the space group. Atomic density is distributed as a Gaussian centred on the atom position. The value of the Gaussian depends on the distance from the atom centre, the atom type, and temperature factor. Its value is calculated at each gridpoint with a given radius for each atom and the values summed to give an "atom" map. This "map" is then used for an inverse Fourier transform which gives the structure factor. X-ray gradients are estimated by "differential synthesis". A difference map is calculated, and this is convoluted with the atomic density.

SIGMAA (Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta. Cryst.*, **A42** , 140 - 149.)

The program SIGMAA can be used to combine a set of calculated phases with a set of previously determined phases for which the phase-probability profiles are held in the form of Hendrickson-Lattman coefficients. It calculates weighted Fourier coefficients either using the calculated phases from a model structure, or combining phase probabilities from isomorphous phases with those from one or more structures.

SQUASH (Zhang, K.Y.J. (1993). SQUASH: Combining constraints for macromolecular phase refinement and extension. *Acta. Cryst.* **D49**, 213-222.)

This program combines the constraints of correct electron-density distribution, solvent flattening, correct local shape of the electron density and equal molecules to produce an integrated procedure for macromolecular phase refinement and extension. The constraints on electron density are satisfied simultaneously by solving a set of non-linear equations.

SURFACE

This program calculates surface areas and can be used to calculate volumes. Water accessible surface area calculated with a probe radius of 1.4 Å.

TFFC (Tickle, I.J. (1992). Molecular replacement, Procedings of the CCP4 study weekend, edited by E.J. Dodson, S. Gover and W. Wolf, pp. 20-32. Warrington: SERC Daresbury Laboratory.

A full-symmetry Translation Function. TFFC uses structure factor information (amplitudes and phases) from all molecules in the unit cell whose positions have already been determined or are about to be determined. Three important features which further improve the signal/noise ratio of the T2 function are the shell-scaling of the F(obs) and F(calc) values by the difference Wilson method, the use of normalised amplitudes (E-values) and the subtraction of the transform of the intramolecular vector set from the transform of the Patterson function of the target structure.

VECSUM

The purpose of this program is to deconvolute a Patterson function. Program is spacegroup general with a Patterson function produced by FFT taken as input. It requires an estimate of the number of major sites. Program calculates a "symmetry function" or "superposition function" or a combination of both. Then a map of the possible sites is output, much like an electron density map.

I.2 Other Programs used

FRODO (Jones, T.A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Cryst.* **11**, 268-272.)

Macromolecular graphics modelling program designed for use on an Evans and Sutherland PS300 graphics system.

HEXRED (B. F. Anderson, Massey University)

Program which reduces a hexagonal data set into its correct part of reciprocal space while also keeping the Friedel pairs if required.

MEANPLANE

Program used to calculated least-squares equations of best fit through "planes" of atoms. Also can calculate angles between the "planes" of atoms.

MULTIWIRE DETECTOR SOFTWARE (Howard, A.J., Nielsen, C. and Xuong, N.H. (1985). Software for a diffractometer with multiwire area detector. *Meth. Enzymol.* **114**, 452-472.)

Software package for data collection based on the methods of Howard *et al.* (1985).

R-AXIS SOFTWARE PROGRAMS (Higashi, T. (1990). The processing of diffraction data taken on a screenless Weissenberg camera for macromolecular crystallography. *J. Appl. Cryst.* **22**, 9-18.)

The R-axis II is controlled through a computer interface by a Rigaku software package that is made up of three major divisions, MEASUREMENT MODE, GRAPHIC MODE and DATA PROCESS MODE. In the last division, refinement of parameters and data integration is carried out. All the above processes are described in the users manual (Instruction manual for R-AXIS IIC and software: Manual number ME201LP1). The processing procedure consists of extracting the reflection intensity data from the oscillation frames by optimising the boundary between the peak and background areas. This is achieved by manipulating the size of a box surrounding the reflectionto get the best separation. The reflections are then merged into a single file, and the frames are scaled using an inverse scale factor and a temperature factor. Reflections greater than 1σ are used in the scaling, and refinement is deemed complete when the ratio shift/standard deviation was less than 0.001. The symmetry-equivalent reflections were then averaged together if the reflections were bigger than 1σ. To reject nonconforming full reflections, if the difference between a certain reflection intensity and its equivalent reflection is greater than 0.3 times the mean reflection intensity of the data set, plus 0.1 times the mean intensity of the equivalent reflections, the reflections are rejected.

SYNCHROTRON SOFTWARE PROGRAMS (Sakabe, N. (1991). X-ray diffraction data collection system for modern protein crystallography with a Weissenberg camera and an imaging plate using synchrotron radiation. *Nuclear Instrumen. and Methods in Physics Research* **A303**, 448-463.)

An X-ray diffraction data collection system has been constructed in the Photon Factory for protein crystallography. It allows the collection of data sets by Weissenberg photography by rotation of a sample around a direct axis. The programs used in processing the data are discussed below (programs Intav3 and Intav4 are included in this section).

WEIS

Data processing program for extracting intensities from Weissenberg images. Initial cell dimensions, space group and setting parameters used for data collection are input. After establishing an initial orientation using fiducial spots, a set of well defined reflections is chosen, and a least-squares refinement of the setting parameters is carried out. Low resolution data is used initially, but once DR and DS (the *rms* deviations between the calculated and observed spot positions in the x and y directions) have been minimised, refinement is carried out using higher resolution data (typically 2.5 Å). Once refinement is finished, two files are output, one containing the positions of all the expected reflections, and the other containing the refined setting parameters. After the observed and predicted positions have been aligned, the reflection intensities are determined using a profile fitting method. The image is divided into nine regions and a mean spot profile is determined for each region. Once the average peak profile for these regions has been determined and corrected for background absorbance, the intensities of all the reflections are calculated. Lorentz and polarisation corrections are also applied at this time.

WEISDATAJ

Program used to take the integrated intensities from WEIS and combine the images for scaling together by INTAV3 and INTAV4.

INTAV3

This program uses structure factors (F's) to calculate scale factors between different sets of data. The scale factors are calculated in a two step process. In the first pass through the data all common reflections are identified and a scale factor is calculated. Tighter criteria for the selection of the common reflections are applied in the second pass. Finally the scale factors are output after the reflections that agree poorly have been rejected. Reflections may be given unit weights or weights according to their standard deviations.

INTAV4

Scale factors calculated in INTAV3 are applied in INTAV4 and the data are merged and a unique data set output. A variety of acceptance criteria can be applied allowing reflections that agree poorly to be rejected. These are:

- accept as observed if $F_{obs}/sig > 2.0$
- accept all values if $|F_{max} - F_{min}|/sig < 4.0$

-reject lowest if $|F_{ave} - F_{min}|/sig > 2.0$

-reject highest if $|F_{max} - F_{ave}|/sig > 2.0$

-for two measurements reject the smaller if $|F_{max} - F_{min}|/sig > 4.0$

When the data is merged in INTAV4 the reflections can either be given equal weight (unit weights) or they can be weighted according to their standard deviations.

TOM Version 2.4 (Jones, T.A. (1982). TOM: A graphics fitting program for macromolecules, In "Computational Crystallography" D. Sayre, ed., Clarendon Press, Oxford pp. 303-317)

Macromolecular graphics modelling program designed for use on a Silicon Graphics IRIS graphics system. An extension of FRODO.

TNT Version 5-C (Tronrud, D.E., Ten Eyck, L.F. and Matthews, B.W. (1987). An efficient least-squares refinement program for macromolecular structures. *Acta Cryst.* **A43**, 489 - 501).

This is a suite of programs whose main function is restrained least-squares refinement. It minimises the disagreement between a model and a set of X-ray crystallographic data while restraining the model to also agree with the principles of ideal stereochemistry.

X-PLOR (Version 3.1 A system for X-ray Crystallography and NMR (1992) , Axel T. Brünger, Yale University Press)

X-PLOR is a program system for computational structural biology. X-PLOR stands for the exploration of conformational space of macromolecules restrained to regions allowed by combination of empirical energy function and experimental data. The program is based on an energy approach. The constant energy function can be minimised by a variety of gradient descent, simulated annealing, and conformational search procedures. The suite incorporates structure refinement packages as well as a molecular replacement system.

Appendix II

Crystallisation studies of aldehyde dehydrogenase

II.1 Pathway of Alcohol Metabolism.

When ethanol (alcohol) is taken orally, it is absorbed into the bloodstream from the stomach and small intestine. Although, a small amount (~ 5 %) is excreted unchanged in the breath or urine, most alcohol is metabolised by the liver. Here ethanol is initially converted to acetaldehyde, which is a more reactive and toxic compound, and then the acetaldehyde is converted to acetate, a normal metabolite, at which point the detoxification of alcohol can be considered complete. Much of the acetate leaves the liver, to be metabolised further by muscle or brain.

II.1.1 Individual Reactions.

1. Ethanol ---> acetaldehyde

This reaction is catalysed mainly by the enzyme alcohol dehydrogenase (ADH), which requires NAD^+ as a cofactor.

$$CH_3CH_2OH + NAD^+ \text{-----> } CH_3CHO + NADH + H^+$$

Many of the mammalian alcohol dehydrogenases have been well characterised in terms of amino acid sequence, 3D structure and kinetic mechanism.

2. Acetaldehyde ---> acetate

This reaction is catalysed by aldehyde dehydrogenases (ALDH), although, it can also be catalysed by aldehyde oxidase and xanthine oxidase. However, because of a high K_m for acetaldehyde, these enzymes are probably not physiologically significant.

$$CH_3CHO + NAD^+ + H_2O \text{------> } CH_3COOH + NADH + H^+$$

This reaction involving the conversion of acetaldehyde to acetate is less well studied than the conversion of ethanol. The three dimensional structure of aldehyde dehydrogenase

is as yet unknown. Aldehyde dehydrogenases are found in almost every tissue of the body (Dietrich, 1966) and in all subcellular compartments (cytosol, microsomes and mitochondria). This distribution has probably arisen because acetaldehyde is very toxic and needs to be removed rapidly from the body. One type of treatment for alcoholics is to use drugs which inhibit aldehyde dehydrogenase (Kitson, 1977). This treatment allows acetaldehyde to accumulate when ethanol is taken and unpleasant symptoms of nausea, dizziness and palpitations arise. The unpleasantness of the symptoms is sufficient to deter patients from ingestion of alcohol.

II.2 Isolation and Purification of Sheep Liver Cytosolic Aldehyde Dehydrogenase.

II.2.1 General.

(a) Sources of the Enzyme.

The existence of aldehyde dehydrogenase (aldehyde: NAD^+ oxidoreductase, EC 1.2.1.3.) was first described by Racker (1949). Dietrich (1966) showed aldehyde dehydrogenase to be present in all rat tissues. It is now believed that the enzyme has a ubiquitous distribution in mammals (Weiner, 1982).

The first homogeneous mammalian preparation was reported by Feldman and Weiner (1972), and the enzyme has now been purified from a variety of species including rat (Siew *et al.*, 1976), rabbit (Duncan, 1977), cow (Leicht *et al.*, 1978), human (Greenfield and Pietruszko, 1977) and sheep (Crow *et al.*, 1974).

Liver, the primary site of ethanol metabolism in mammals (Hawkins and Kalant, 1972), has represented the major source of aldehyde dehydrogenase in the studies that have been reported (Pietruszko and Yonetani, 1981). Aldehyde dehydrogenase has also been isolated from erythrocytes (Agarwal *et al.*, 1985), and eye lens (Hoe *et al.*, 1985), and the isoenzyme patterns of aldehyde dehydrogenase have been reported in hair root (Harada *et al.*, 1981).

(b) Multiplicity

As chromatographic techniques and materials improved, it was found that aldehyde dehydrogenase has a multiple subcellular distribution, and the properties of the enzyme from different cellular locations can differ. In rat liver, three major sites of activity exist, mitochondrial, microsomal and cytosolic (Koivula and Koivusalo, 1975). In sheep (Crow *et al.*, 1974) , horse (Eckfelt and Yonetani, 1976a) , human (Greenfield and Pietruszko, 1977) and baboon liver (Alderman *et al.*, 1982), only the cytosolic and mitochondrial forms of the enzyme are significant.

The mitochondrial isoenzyme is considered to be responsible for the major part of acetaldehyde oxidation (Marjanen, 1972; Svanas and Weiner, 1985; Harrington *et al.*, 1987). However, the cytosolic isoenzyme can make a significant contribution towards acetaldehyde oxidation when the concentration of aldehyde reaches millimolar levels (Ryzlak and Pietruszko, 1987) .

This aldehyde dehydrogenase composition is further complicated by multiple forms of the enzyme in each subcellular fraction. In sheep liver mitochondria seven isoenzymes were found, whilst two isoenzymes were found in the cytosol of the species (Agnew *et al.*, 1981). Two mitochondrial isoenzymes and one cytosolic isoenzyme have been found in baboon liver (Alderman *et al.*, 1982) and human brain (Ryzlak and Pietruszko, 1987). Multiple subcellular forms of rat liver aldehyde dehydrogenase have also been reported, including one isoenzyme associated with the outer mitochondrial membrane (Tottmar *et al.*, 1974). The function of the multiple forms of the enzyme is unknown.

II.3.2 Structural Properties.

Most mammalian aldehyde dehydrogenases are tetrameric, of approximate molecular weight 200,000 Da and subunit molecular weight approximately 50,000 Da (Weiner, 1980). Exceptions to this are the rat liver mitochondrial matrix and outer membrane enzymes which have molecular weights of 320,000 Da and 67,000 Da respectively (Siew *et al.*, 1976).

Protein sequencing studies have shown that in both mitochondrial and cytosolic isoenzymes the tetrameric enzyme is formed from identical subunits (Hempel and Jörnvall, 1987). Each polypeptide is approximately 500 amino acid residues long (Fairwell *et al.*, 1984; Von Bahr-Lindström *et al.*, 1984; Hempel *et al.*, 1984). The human cytosolic

isoenzyme has an acylated N-terminus (Hempel *et al.*, 1984), which is not found in the mitochondrial isoenzyme (Hempel *et al.*, 1985). The latter does not show N-terminal processing, probably because of its requirement for transport into this organelle (Hempel *et al.*, 1985). Structural similarities were found between the amino terminal sequences of human aldehyde dehydrogenase and other targeting sequences responsible for protein uptake into mitochondria (Brawn *et al.*, 1987). Cytosolic isoenzymes from different species have a greater sequence homology than cytosolic and mitochondrial isoenzymes from the same species (Von Bahr-Lindström *et al.*, 1984; Hempel *et al.*, 1985; Hempel *et al.*, 1983; Johansson *et al.*, 1988). The two isoenzymes are believed to fold into a similar overall conformation (Hempel and Jörnvall, 1987).

The mammalian isozymes of aldehyde dehydrogenase are divided into three classes based on their amino acid sequences (Lindahl and Hempel, 1991). Class 1, which is the cytosolic form, and Class 2, the mitochondrial form, share ~ 70 % sequence identity. The Class 3 isozymes, which are constitutive/inducible aldehyde dehydrogenases, and differ in the fact that they exist as homodimers of approximately 2 x 50,000 Da (Jones *et al.*, 1988), have only about 30 % sequence identity with the Class 1 and 2 enzymes. Each of the isoenzymes have different roles although these have yet to be fully understood. The cytosolic enzyme may have a role in the oxidation of retinal to retinoic acid (Yosida *et al.*, 1992) and perhaps a role in regulation of UV absorption (Algar *et al.*, 1990).

The exact region in the enzyme associated with coenzyme binding has yet to be established, since the existence of a "Rossman fold" is not apparent from the sequence data (Hempel and Jörnvall, 1987). All the NAD^+-linked dehydrogenases for which the tertiary structures are known have revealed a hydrophobic cleft formed by two mononucleotide-binding units, each composed of three alternating helical and β-sheet structures (Rossman *et al.*, 1976). The possible lack of this feature in aldehyde dehydrogenases is interesting with regard to monomer association in the tetrameric enzyme and the predicted asymmetry of monomer associations (Pietruszko and Mackerell, 1987) .

One residue that is conserved in all the various sequences that have been determined is Cys302, which is sensitive to disulphiram addition (Kitson, 1975). Other conserved amino acids that could be involved in the active site are Glu268 and Glu487 (Johansson *et al.*, 1988).

Kinetic experiments, although extensive, have left some interesting questions about the number of catalytic sites per tetramer. There are suggestions of only a single site (Dickinson and Haywood, 1986) and of dual sites (Blackwell *et al.*, 1983). A review evaluating the arguments in favour of single or dual catalytic sites has been presented by Duncan (1985).

Although many kinetic and chemical studies have been carried out , there is still no X-ray structure to correlate the results, and to clarify the questions that have been raised. These questions include whether there are two active sites per tetramer or a single site as mentioned above, and whether there are any conformational changes upon binding of NAD^+/NADH. A second question raised is whether the esterase and dehydrogenase catalytic sites are different or the same. As mentioned earlier the 3D structures are expected to be similar despite the lower sequence identity of the Class 3 isozymes (Rose, 1990). Preliminary crystallographic analyses of a Class 3 rat liver aldehyde dehydrogenase (Rose, 1990) and of a Class 2 dehydrogenase, bovine liver mitochondrial aldehyde dehydrogenase (Hurley and Weiner, 1992) have been reported, but as yet no 3D structure has been determined. The Class 3 aldehyde dehydrogenases differ substantially from the Class 1 and 2 aldehyde dehydrogenases, both in catalytic properties and in primary and quaternary structures. Secondary structural predictions, however, suggest that the subunit tertiary structures of all these aldehyde dehydrogenase classes are largely similar (Lindahl and Evces 1984; Hempel *et al.*, 1990).

Subnote:	Class 1	Cytosolic
	Class 2	Mitochondrial
	Class 3	Constitutive/Inducible e.g. rat microsomal, tumors.

II.3 Purification of Sheep Liver Cytosolic Aldehyde Dehydrogenase.

Initial purifications of cytosolic aldehyde dehydrogenase followed the procedure of Dickinson *et al.* (1981), in which ammonium sulphate is used as the precipitating agent. However, SDS polyacrylamide electrophoretic gels on the final enzyme sample, indicated that the protein was not of sufficient purity to be used in crystallographic trials. It was decided that a review of the purification of cytosolic aldehyde dehydrogenase should be attempted. It was hoped that an improvement in the purity and a shortening in the length of time taken for the purification scheme could be achieved.

II.3.1 Revised purification scheme.

Initially it was decided to change the precipitating agent to polyethylene glycol for ease of preparation of the liver (Motion, 1986). When using ammonium sulphate, the salt has to be removed by dialysis before the sample is applied to the first ion exchange column; this is not the case for polyethylene glycol.

Next an OSTSORB DEAE ion exchange resin was used instead of the traditional DEAE-cellulose resins. The new resin provided a greater capacity and improved flow rate and separation. From FPLC work it was decided that a step elution from the column using sodium chloride provided best purification.

The second column uses pH gradient ion-exchange chromatography on a DEAE-sephacel (Pharmacia) ion-exchange resin. This step is also used in the purification of Dickinson *et al.* (1981). The main role of this step is to remove any mitochondrial aldehyde dehydrogenase which may be present.

Finally, affinity chromatography was used in the last step of the purification of the enzyme. This was achieved using Sigma 5' AMP-sepharose 4B. This column has an affinity for enzymes requiring NAD^+ or ATP cofactors. The enzyme can be eluted by the addition of NAD^+ which has a greater affinity for the column thereby removing the enzyme.

At this stage, the purification procedure for cytosolic aldehyde dehydrogenase consisted of the following steps;

(i) Homogenisation

1.0 - 1.3 kg of chopped liver (in 250 g portions) is homogenised by sonication for 30 seconds in 350 cm^3 of 0.005 M phosphate buffer (pH 7.4) containing 0.25 mol L^{-1} sucrose, 0.1 % v/v 2-mercaptoethanol and 0.01 mol L^{-1} ethylenediamine tetraacetic acid (EDTA) using a Janke & Kunkel Ultra Turrax T45 sonicator.

(ii) Centrifugation

The homogenate is centrifuged for 15 minutes at 13,800 x g to remove cell wall material, whole cells and connective tissue. The precipitate is discarded. The supernatant is then strained through glass wool to remove fat, and recentrifuged at 23,330 x g for 30 minutes to sediment the mitochondria. The precipitate is discarded.

(iii) Precipitation of Enzyme using Polyethylene Glycol

After the centrifugation procedures, the supernatant volume is made up to 1500 cm^3 with 0.005 mol L^{-1} phosphate buffer (pH 7.4) containing 0.01 mol L^{-1} EDTA, powdered A.R. grade polyethylene glycol 8000 is added over 30 minutes with stirring to give a mixture containing 12 % w/v (120 g L^{-1}) polyethylene glycol. Stirring for a further 30 minutes to allow equilibration, is followed by centrifugation for 15 minutes at 23,330 x g. The precipitate is discarded. Further polyethylene glycol is added to the supernatant as before, to increase the concentration to 20 % w/v (a further 80 g L^{-1}) and the mixture equilibrated with stirring for 30 minutes. Following centrifugation for 15 minutes at 23,330 x g, the precipitate is redissolved in about 250 cm^3 of 0.01 mol L^{-1} pH 6.2 Bis-Tris buffer containing 0.005 mol L^{-1} NaCl and 0.01 mol L^{-1} EDTA.

(iv) Ion-exchange reisin Chromatography

After the precipitate has been redissolved, the pH of the solution is lowered to 5.7 with 1 mol L^{-1} HCl with stirring. This is then centifuged for 10 minutes at 23,330 x g, the precipitate discarded and the solution is adjusted to pH 6.2 with 4 mol L^{-1} NH_3 solution. The adjusted solution is loaded on a 4 x 20 cm column of OSTSORB DEAE ion-exchange resin. The column is washed with 0.01 mol L^{-1} pH 6.2 Bis-Tris buffer until A_{280} of the effluent in 1 cm-path-length cells is negligible. The enzyme is removed from the column by step elution using the same buffer containing 0.150 mol L^{-1} NaCl (about 500 cm^3). The fractions (about 450 cm^3 in total) are assayed for aldehyde dehydrogenase activity using the appearance of NADH at 340 nm and the active fractions are pooled together. The solution is then concentrated down to a volume of about 50 cm^3 using a diaflo ultrafiltration apparatus with an Amicon XM-50 filter and the sample is then dialysed against 0.01 mol L^{-1} pH 6.5 Bis-Tris buffer for 24 hours to remove the sodium chloride and to adjust the pH.

(v) pH-gradient Ion Exchange Chromatography

The enzyme solution, once dialysed, is added to a column (2 x 16 cm) of Pharmacia DEAE sephacel which had been equilibrated with the same buffer.

The enzyme is eluted using a pH gradient produced by running 500 cm^3 of 0.01 mol L^{-1} pH 4.6 sodium acetate buffer into 500 cm^3 of pH 6.5 Bis-Tris buffer. Fractions of 8 - 10 cm^3 are collected and assayed for aldehyde dehydrogenase activity. The active fractions

are pooled together and concentrated to 50 cm^3 as described before, and dialysed against pH 6.5 Bis-Tris buffer. The enzyme can then be frozen and stored for later use.

(vi) Affinity Chromatography

As required, the enzyme can then be dialysed (24 hours) against 0.03 mol L^{-1} pH 6.0 phosphate buffer. 20 mg of total protein is loaded on to a 1 x 4 cm Sigma 5' AMP-sepharose 4B column pre-equilibrated with the same buffer. The column is washed with the same buffer (about 20 cm^3). The enzyme is eluted with 0.03 mol L^{-1} pH 8.0 phosphate buffer containing 0.003 mol L^{-1} NAD^+ (about 10 - 15 cm^3).

The active fractions are pooled together and dialysed against 0.01 mol L^{-1} pH 6.5 Bis-Tris buffer so that the samples can be frozen. Typically between 100 - 200 mg of aldehyde dehydrogenase can be purified from a purification.

II.3.2 Further modification.

In the above purification scheme two separate columns were used; i) for the initial clean-up and ii) for the pH gradient step before the use of the affinity resin. It was decided to combine both of these into a single step. The OSTSORB DEAE resin was equilibrated with 0.01 mol L^{-1} Bis-Tris at pH 6.2. The crude extract from the PEG precipitation step was loaded on the column and the equilibration buffer plus 0.005 mol L^{-1} NaCl was washed through the column until A_{280} was neglible. A second wash with 0.01 mol L^{-1} Bis-Tris, this time at pH 5.7, was then applied, until A_{280} was again negligible; the cytosolic aldehyde dehydrogenase remained bound while any mitochondrial enzyme was eluted from the column. The cytosolic aldehyde dehydrogenase was then removed from the column with a final wash of 0.03 mol L^{-1} sodium acetate pH 5.0. The fractions were assayed for activity and the active fractions pooled. The pooled fractions were reduced in volume by ultrafiltration and the enzyme dialysed with 0.05 mol L^{-1} pH 7.4 phosphate buffer. The enzyme was then ready to be loaded onto the affinity resin. The affinity resin was changed from that used previously, by substituting a p-hydroxyacetophenone ligand. Acetophenone is a reversible inhibitor of mitochondrial aldehyde deydrogenase and so could be incorporated into a solid matrix for use as a specific binder of aldehyde dehydrogenase. Figure II.3.2.1 shows the structure of the acetophenone inhibitor.

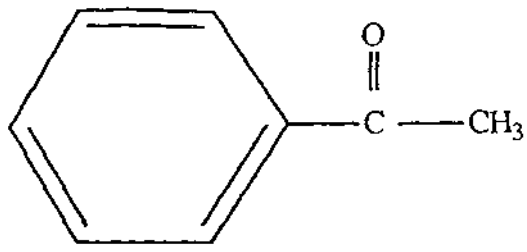


Figure II.3.2.1 The aldehyde dehydrogenase inhibitor acetophenone.

After loading the enzyme on the affinity resin the column was washed with the equilibration buffer to remove any non-binding contaminants. The cytosolic aldehyde dehydrogenase was removed by adding 0.01 mol L^{-1} hydroxyacetophenone to the equilibration buffer. The hydroxyacetophenone binds preferentially to the column thereby eluting the enzyme. Once the fractions had been collected the active fractions were pooled together and reduced in volume by ultrafiltration and dialysed into 0.02 mol L^{-1} Bis-Tris pH 6.5 to remove any hydroxyacetophenone and for storage. Figure II.3.2.2 shows a diagram of the preparation of the solid matrix of the affinity resin.

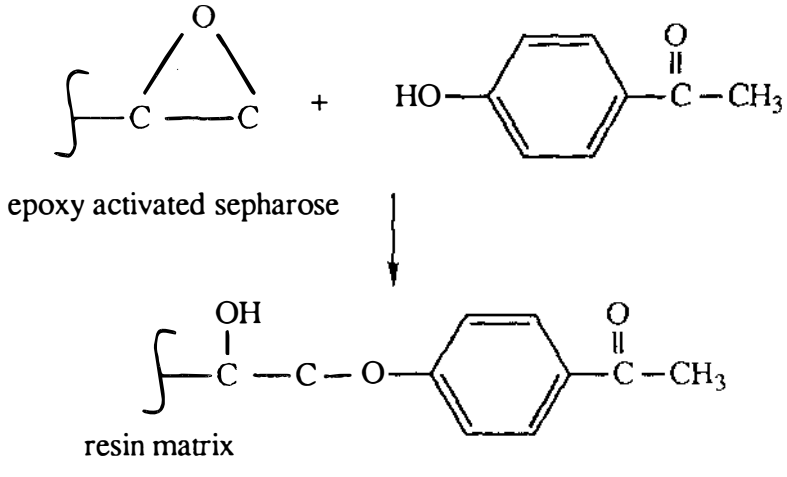


Figure II.3.2.2 The solid matrix of the p-hydroxyacetophenone affinity resin.

II.3.3 Results.

The use of this improved purification scheme produces cleaner cytosolic aldehyde dehydrogenase as well as reducing the time that the purification takes. An indication of the improved quality of the final product is shown in SDS and native gels which show only one very minor possible contaminant (see Lane 1 in Figure II.3.3.1). In contrast Figure II.3.3.2 shows an SDS gel of aldehyde dehydrogenase purified (Lanes 4 and 6) using the revised method. The original procedure left the 'purified' ALDH with several bands of contaminants.

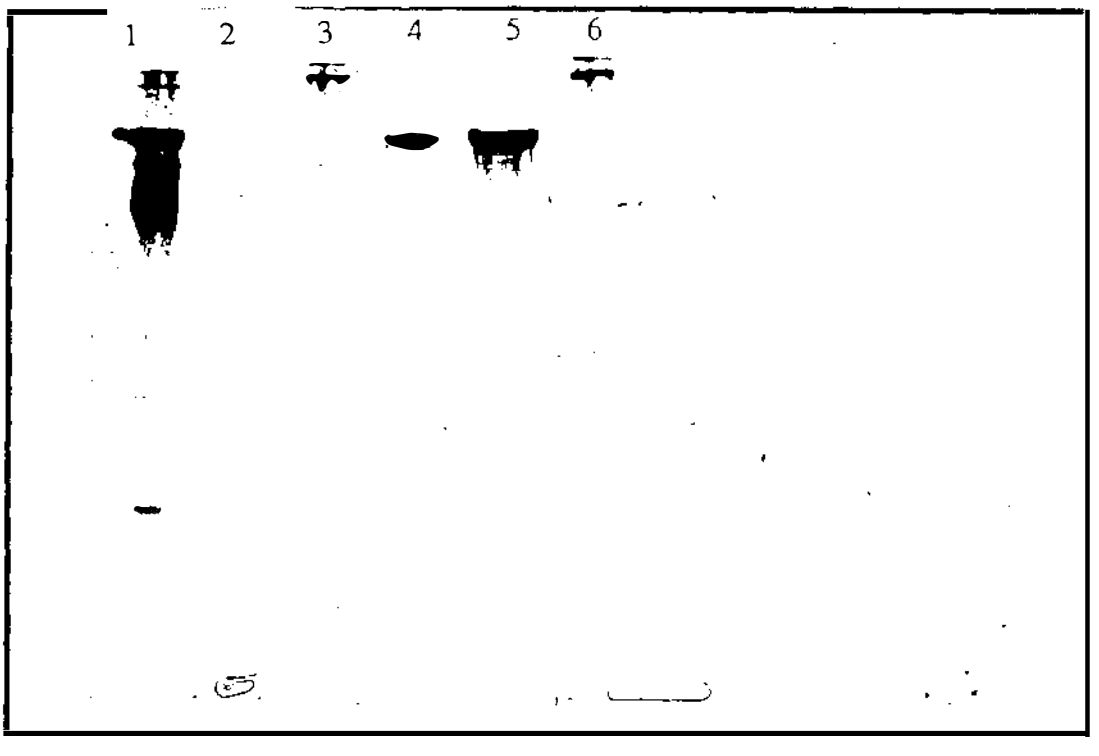


Figure II.3.3.1 Native gel showing purity of cytosolic sheep liver aldehyde dehydrogenase using revised procedure (ALDH in lanes 4 and 5).

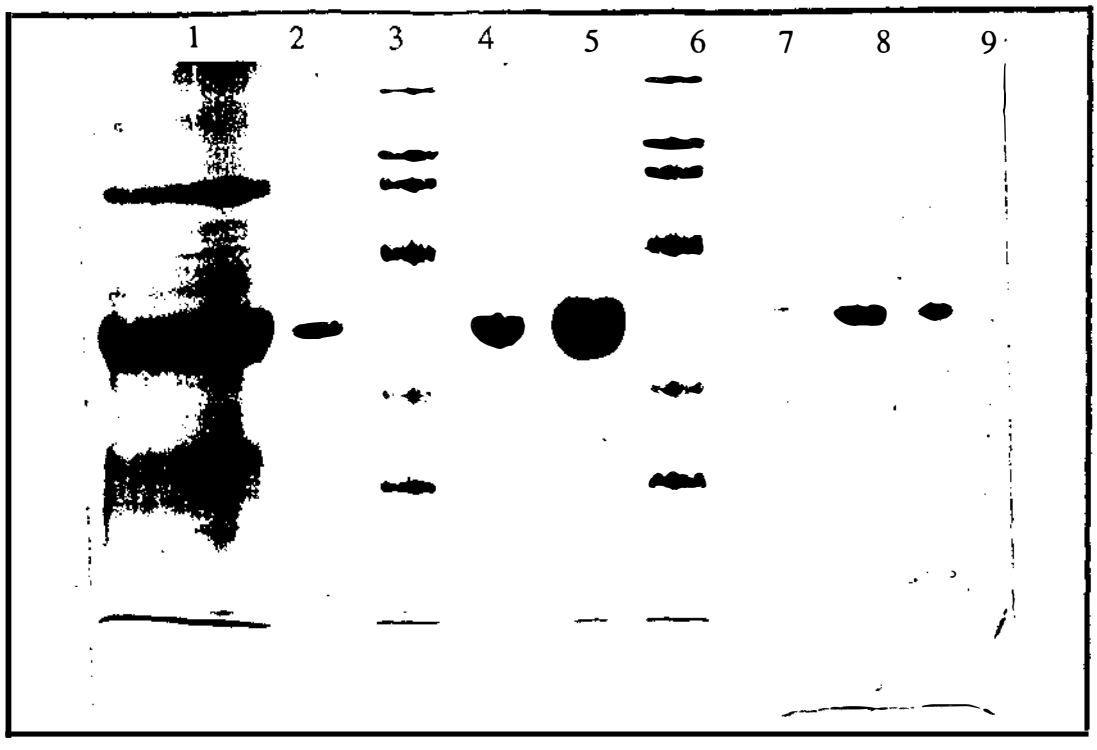


Figure II.3.3.2 SDS gel showing purity of cytosolic sheep liver aldehyde dehydrogenase (Lanes 4 and 5) using revised method.

II.4 Crystallisation of sheep liver cytosolic aldehyde dehydrogenase.

II.4.1 Initial crystallisation experiments.

Initial crystallisation trials on the enzyme were carried out by Dr J.P. Hill. A factorial method, whereby a large range of conditions (i.e. pH, precipitant etc) can be surveyed for the formation of crystals, was used initially (Jancarik and Kim, 1991). The hanging drop method (McPherson, 1990) was used to set up the crystallisations; this involves placing a drop of dissolved protein, to which an equal volume of well mix (containing buffer, precipitant, metal salt etc) has been added, on a cover slip. The cover slip is then inverted and sealed over a much larger reservoir of the well solution. This gives a vapour diffusion gradient through which the concentration of the precipitant in the protein drop changes. The advantage of this set up is that a wide range of conditions can be tried with minimal use of protein. The typical drop size is 5 μl, with protein concentrations of 7

mg ml^{-1}, so that each experiment requires approximately 1.4 mg of protein (assuming 40 wells as required by the factorial method).

It was found from the initial factorial experiments that small twinned crystals were formed from a well mix which contained 0.2 mol L^{-1} $MgCl_2$, 0.1 mol L^{-1} acetate buffer pH 4.6 and 11.5 % wt/v polyethylene glycol 4000 (these are final concentrations). From these conditions various matrices were set up around the desired region to further refine crystallisation conditions. Seeding experiments were set up to explore the possibility of growing larger crystals from those already formed (Stura and Wilson, 1990). Other trials centred around using various concentrations of different divalent metal cations, especially those that have been identified as being kinetically important in the function of aldehyde dehydrogenase (Dickinson and Hart, 1982), and replacing the acetate buffer with citrate buffer of varying concentrations.

From these trials it was found that the best crystals were grown from a solution that contained 0.2 mol L^{-1} $MgCl_2$, 0.1 mol L^{-1} acetate buffer pH 5.0, 0.001 mol L^{-1} NAD^+, and 9 % wt/v PEG 4000 (at a temperature of 4 °C).
Other observations included;

1. Only PEG could be used as a precipitant
2. Only Mg^{2+} produced protein crystals (and no crystals were produced without divalent metal ions)
3. enzyme solubility is lower at 4 °C than at room temperature and 37 °C

From this it was decided to repeat the seeding experiments that had been previously tried. Both macroseeding and microseeding (Stura and Wilson, 1990; McPherson, 1990) were tried. A range of enzyme concentrations were tried ranging from 2 mg ml^{-1} --> 12 mg ml^{-1}, and the PEG concentration was varied from 7 - 11 % wt/v. It was found that it was important not to use high enzyme concentration as otherwise crystallisation began as soon as the seeds were transferred.

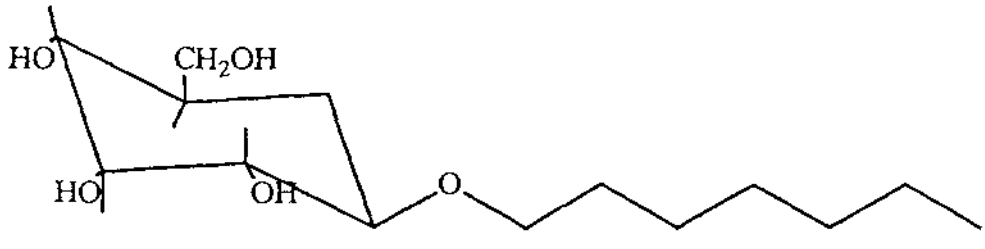
Once the new purification regime had been set in place, more crystallisation trials were set up to observe the effect the improved purity would have on crystallisation. The conditions from which crystals of aldehyde dehydrogenase formed were already known from the earlier work, so attempts were made to improve the quality of the ALDH crystals. These modifications included;

1. Addition of various salts e.g $MgSO_4$ (all trials also had 0.005 mol L^{-1} $MgCl_2$).
2. Use of various molecular weight PEG's (ranging from 400 - 20,000 MW).

3. Matrix continuation (from known conditions).
4. Addition of Zn^{2+}.
5. Expanded matrix using PEG 6000.
6. Batch crystallisations in single vials using known conditions.
7. Use of either buffered enzyme or enzyme dialysed against water.
8. Addition of detergent (β-octyl-glucoside).
9. Free interface diffusion technique (McPherson, 1990).
10. Sitting drops.
11. Use of inhibitors - Choral Hydrate
    - Disulphiram
12. Fast screen factorial experiments (Jancarik and Kim, 1991) and screen containing 0.001 mol L^{-1} NAD^+

Despite a large number of crystallisation experiments using a variety of techniques, and the addition of various compounds, no crystals of suitable quality for X-ray diffraction were obtained. The best crystals grew upon the addition of β-octyl-glucoside to the previously determined conditions for crystallisation.

The use of detergents in protein crystallisation (McPherson *et al.*, 1986) was initially developed for crystallisation of membrane proteins. The rationale was that the detergent would reduce hydrophobic interactions between molecules and this would mean ionic and electrostatic interactions would be enhanced. Ionic bonds are more directional than hydrophobic bonds as they require complementary counterions with which to form an interface. This is desirable for formation of crystalline order. Figure II.4.1.1 gives the structure of β-octyl-glucoside used in the crystallisation trials.



Octyl-β-D-glucopyranoside (β-OG)

Figure II.4.1.1 Structure of β-octyl-glucoside.

The detergent matrix tried was;

PEG 6000	10 - 16 % (wt/v)
β–OG	0 - 5 % (wt/v)
$MgCl_2$	0.17 mol L^{-1}, 0.19 mol L^{-1}, 0.22 mol L^{-1}

constant conditions: 0.005 mol L^{-1} DTT, 0.001 mol L^{-1} NAD^+, 0.1 mol L^{-1} acetate buffer, pH 5.0

The best crystals were grown from ;

12 % wt/v PEG 6000
0.17 mol L^{-1} and 0.19 mol L^{-1} $MgCl_2$
2.0 - 2.5 % wt/v β–OG

The enzyme concentration was 2.5 mg ml^{-1} in all cases. Fig II.4.1.2 shows a photograph of the best crystals of aldehyde dehydrogenase using detergent.

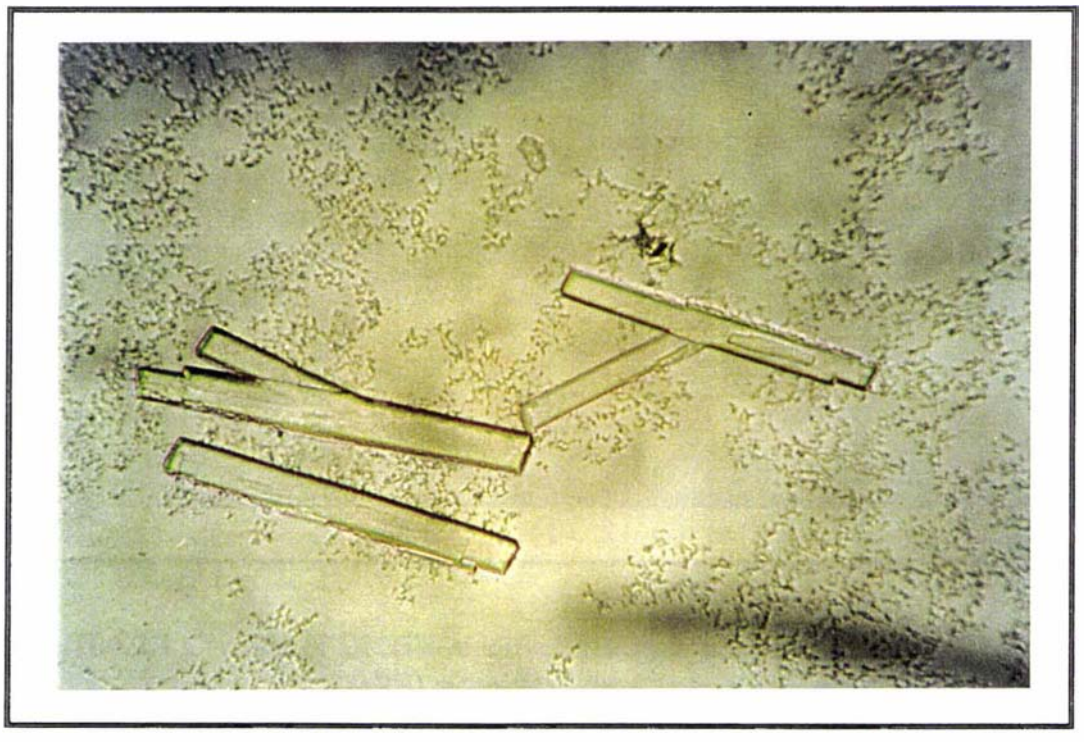


Figure II.4.1.2 Crystals of sheep liver cytosolic aldehyde dehydrogenase under the "best" conditions. (x250 magnification)

II.4.2 Summary.

Although crystals suitable for X-ray diffraction studies were unable to be grown in this study, the conditions determined provided the basis for the subsequent successful crystallisation of ALDH. This was based on the binding of a compound DMNB (3,4-dihydro-3-methyl-6-nitro2*H*-1,3-benoxazin-2-one) which had been developed to provide a reporter group for probing the active site (Kitson and Freeman, 1993; Kitson and Kitson, 1994). This reporter group probably restrains a fairly flexible structure when bound, thereby allowing the formation of larger, better-ordered crystals. The details of the crystallisation and preliminary X-ray studies of aldehyde dehydrogenase are given in Baker *et al.* (1994). Figure II.4.2.1 shows crystals of aldehyde dehydrogenase with DMNB bound.

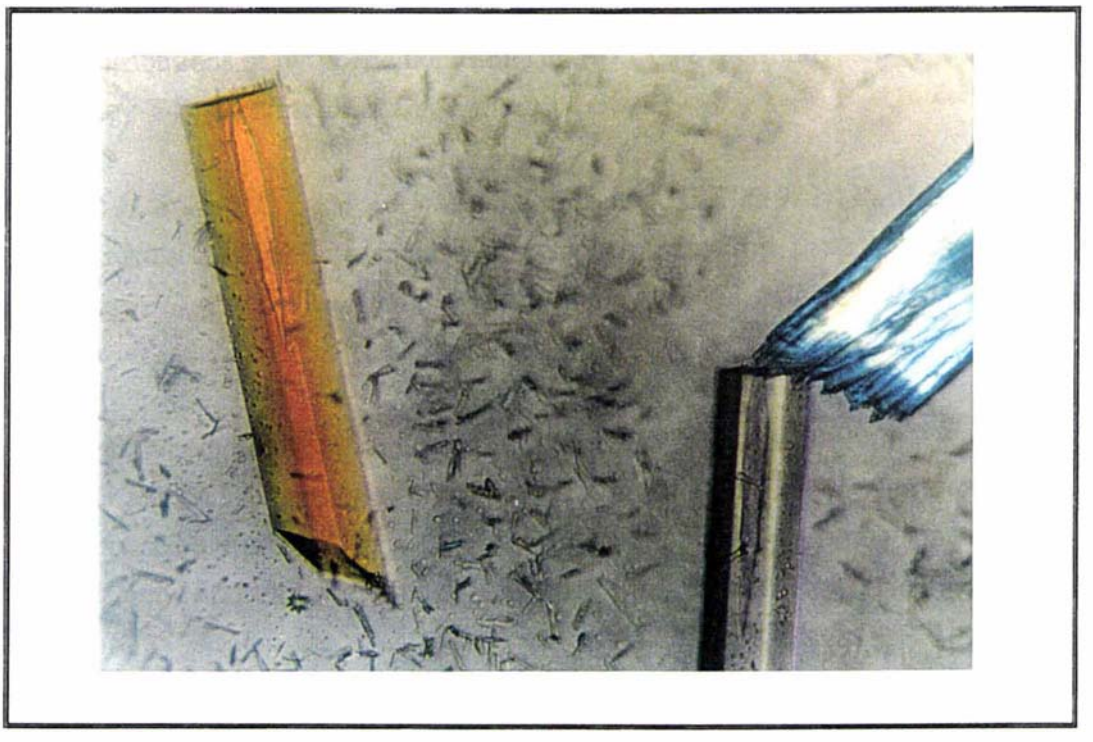


Figure II.4.2.1 Crystals of DMNB-modified aldehyde dehydrogenase.
(x250 magnification)

Bibliography

Agarwal, D.P., Müller, C., Korencke, C., Mika, U., Harada, S. and Göedde, H.W. (1985). Changes in erythrocyte and liver aldehyde dehydrogense isozymes in alcoholics. *Enzymology of Carbonyl Metabolism 2: Aldehyde Dehydrogenase, Aldo-Keto Reductase, and Alcohol Dehydrogenase* (ed.) A.R. Liss, pp 113-127.

Agnew, K.E.M., Bennett, A.F., Crow, K.E., Greenway, R.M., Blackwell, L.F. and Buckley P.D. (1981). A reinvestigation of the purity, isoelectric point and some kinetic properties of the aldehyde dehydrogenases from sheep liver. *Eur. J. Biochem.* **119**, 79-84.

Alber, T., Dao-pin, S., Wilson, K., J.A., Cook, S.P. and Matthews, B.W. (1987). Contributions of hydrogen bonds of Thr 157 to the thermodynamic stability of phage T4 lysozyme. *Nature (London)* **330**, 41-46.

Alderman, J., Sonny, C.G., Gordon, E.R. and Lieber, C.S. (1982). Partial characterisation of hepatic aldehyde dehydrogenase from the baboon. *Enzymology of Carbonyl Metabolism: Aldehyde Dehydrogenase and Aldo/Keto Reductase* (ed.) A.R. Liss, pp 77-89.

Algar, E.M., Abedinia, M., VandeBerg, J.L., and Holmes, R.S. (1990). Purification and properties of baboon corneal aldehyde dehydrogenase: Proposed uvr protective role. *Enzymology and Molecular Biology of Carbonyl Metabolism 3* (eds.) H.Weiner *et al.*, Plenun Press, New York, pp 53-61.

Almassy, R.J. and Dickerson, R.E. (1978). *Pseudomonas* cytochrome c-551 at 2.0 Å resolution: Enlargement of the cytochrome c family. *Proc. Natl. Acad. Sci. USA* **75**, 2674-2678.

Ambler, R.P. (1973). The amino acid sequence of cytochrome c' from *Alcaligenes* sp NCIB 11015 *Biochem. J.* **135**, 751-758.

Ambler, R.P. (1980). The structure and classification of cytochromes c In: *From cyclotrons to cytochromes.* (eds.) Robinson A.B. and Kaplan, N.O., Academic Press, London, New York, pp 263-279.

Ambler, R.P. and Bartsch, R.G., Daniel, M., Kamen, M.D., McLellan, L., Meyer, T.E. and Van Beeumen, J. (1981). Amino-acid sequences of bacterial cytochromes c' and c-556. *Proc. Natl. Acad. Sci. USA* **78**, 6854-6857.

Ambler, R.P. and Brown, L.H. (1967). The amino acid sequence of *Pseudomonas f lourescens* azurin *Biochem. J.* **104**, 784-825.

Awadé, A.C., Cleuziat, Ph., Gonzales, Th. and Robert-Baudouy, J. (1994). Pyrrolidone carboxyl peptidease (Pcp): An enzyme that removes pyroglutamic acid (pGlu) from pGlu-peptides and pGlu-proteins. *Proteins: Structure, Function, and Genetics* **20**, 34-51.

Baker, E.N. (1980). Structure of Actinidin, after refinement at 1.7 Å resolution. *J. Mol. Biol.* **141**, 441-484.

Baker, E.N. and Hubbard, R.E. (1984). Hydrogen bonding in globular proteins. *Prog. Biophys. molec. Biol.* **Vol. 44**, 97-179.

Baker, H.M., Blackwell, L.F., Brown, R.L., Buckley, P.D., Dobbs, A.J., Hardman, M.J., Hill, J.P., Kitson, K.E., Kitson, T.M. and Baker, E.N. (1994). Crystallisation and preliminary X-ray diffraction studies on cytosolic (class 1) aldehyde dehydrogenase from sheep liver. *J. Mol. Biol.* **241**, 263-264.

Barakat, R. and Strekas, T.C. (1982). pH variation of midpoint potentials for three photosynthetic bacterial cytochromes c'. *Biochim. Biophys. Acta* **679**, 393-401.

Bartsch, R.G. (1978). Cytochromes. In: *The Photosynthetic bacteria.* (eds.) Clayton, R.K., Sistrom, and W.R. Plenum, New York pp. 249-279.

Bartsch, R.G. and Kamen, M.D. (1960). Isolation and properties of two soluble haem proteins in extracts of the photoanaerobe *Chromatium. J. Biol. Chem.* **235**, 825-831.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr, Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977). The protein data bank; a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **122**, 535-542.

Blackwell, L.F., Bennett, A.F., and Buckley, P.D. (1983). Relationship between the mechanisms of the esterase and dehydrogenase activities of the cytoplasmic aldehyde dehydrogenase from sheep liver, an alternate view. *Biochemistry.* **22**, 3784-3791.

Blow, D.G. and Rossman, M.G. (1961). The single isomorphous replacement method. *Acta Cryst.* **14**, 1195-1202.

Blow, D.G. and Crick, F.H.C. (1959). The treatment of errors in the isomorphous replacement method. *Acta Cryst.* **12**, 794-802.

Blundell, T.L. and Johnson, L.N. (1976). In: *Protein Crystallography* (eds.) Blundell, T.L. and Johnson, L.N., Academic Press, New York.

Bodo, G. (1955). Crystalline cytochrome c from the king penguin. *Nature* **176**, 829-830.

Brayer, G.D., Delbaere, L.T.J. and James, M.N.G. (1978). Molecular structure of crystalline *Streptomyces griseus* protease at 2.8 Å resolution. *J. Mol. Biol.* 124, 243-259.

Brawn, T., Bober, E., Singh, S., Agarwal, P., and Goedde, H.W. (1987). Evidence for a signal peptide at the amino terminal end of human mitochondrial aldehyde dehydrogenase. *FEBS Letters* **215**, 233-236.

Brünger, A.T. (1990). Extension of Molecular Replacement: A New Search Strategy based on Patterson Correlation Refinement. *Acta Cryst* **A46**, 46-57.

Brünger, A.T. (1991). Simulated annealing in crystallography. *Rev. Phys. Chem.* **42**, 197-223.

Brünger, A.T. (1992). X-Plor version 3.1 A system for X-ray Crystallography and NMR. (ed.) Axel T. Brünger, Yale University Press.

Brünger, A.T., Kuriyan, J. and Karplus, M. (1987). Crystallographic R factor Refinement by Molecular Dynamics. *Science* **235**, 458-460.

Burley, S.K. and Petsko, G.A. (1986a). Dimerisation energetics of benzene and aromatic amino acid sidechains. *J. Am. Chem. Soc.* **108**, 7995-8001.

Burley, S.K. and Petsko, G.A. (1986b). Amino-aromatic interactions in proteins. *FEBS Letts.* **203**, 139-143.

Castellano, E.E., Oliva, G. and Navaza, J. (1992). Fast rigid-body refinement for molecular-replacement techniques. *J. Appl. Cryst.* **25**, 281-284.

Chen, L., Dureley, C.E., Mathews, F.S., and Davidson, V.L. (1994). Structure of an electron transfer complex: Methylamine dehydrogenase, Amicyanin, and cytochrome c-551i. *Science* **264**, 86-90.

Chothia, C. (1984). Principles that determine the structure of proteins. *Annu. Rev. Biochem.* **53**, 537-572.

Chothia, C., Levitt, M. and Richardson, D. (1977). Structure of proteins: Packing of α-helices and pleated sheets. *Proc. Natl. Acad. Sci. USA* **74**, 4130-4134.

Cohen, C. and Parry, D.A.D. (1990). α-helical coiled coils and bundles: How to design an α-helical protein. *Proteins: Structure, Function, and Genetics* **7**, 1-15.

Colby, J. and Dalton, H. (1978). Resolution of the methane monooxygenase of *Methylococcus capsulatus* (Bath) into three components. *Biochem. J.* **171**, 461-468.

Collins, D.M., Countryman, R. and Hoard, J.L. (1972). Stereochemistry of low spin iron porphyrins. 1. Bis(imidazole)-α,β,γ,5-tetraphenylporphinatoiron(III) chloride. *J. Amer. Soc.* **94**, 2066-2072.

Connolly, M.L. (1983). Analytical molecular surface calculation. *J. Appl. Cryst.* **16**, 548-558.

Cowtan, K.D. and Main, P. (1993). Improvement of macromolecular density by the simultaneous application of real and reciprocal space constraints. *Acta Cryst.* **D49**, 148-157.

Corker, G.A. and Sharpe, S.A. (1975). Influence of light on the E.P.R. detectable electron transport components of *Rhodospirillum rubrum*. *Photochem. Photobiol.* **21**, 49-61.

Crick, F.H.C. (1953). The packing of α-helices: Simple coiled-coils. *Acta Cryst.* **6**, 689-697.

Crick, F.H.C. and Magdoff, B.A. (1956). The theory of the method of isomorphous replacement for protein crystals. I. *Acta Cryst.* **9**, 901-908.

Crow, K.E., Kitson, T.M., MacGibbon, A.K.H. and Batt R.D. (1974). Intracellular localization and properties of aldehyde dehydrogenases from sheep liver. *Biochim. Biophys. Acta* **350**, 121-128.

Crowther, R.A. 1972. The Fast Rotation Function. In: *The Molecular Replacement Method*. (ed.) Rossman, M.G., Int. Sci. Rev. Ser. New York: Gordon and Breach. pp 173-178.

Crowther, R.A., and Blow, D.M. (1967). A method of positioning a known molecule in an unknown crystal structure. *Acta Cryst.* **23**, 544-548.

Cusanovich, M.A., Bartsch, R.G. and Kamen, M.D. (1968). Light induced absorbance changes in *Chromatium* chromatophores. *Biochim. Biophys. Acta* **153**, 397-417.

Cusanovich, M.A., Tedro, S.M. and Kamen, M.D. (1970). *Pseudomonas denitrificans* cytochrome cc'. *Arch. Biochim. and Biophys.* **141**, 557-570.

Czerwinski, E.W. and Mathews, F.S. (1974). Location of the iron atom and the non-crystallographic symmetry elements on cytochrome b_{562}. *J. Mol. Biol.* **86**, 49-57.

Czjek, M., Payan, F., Guerlesquin, F., Bruschi, M. and Haser, R. (1994). Crystal structure of cytochrome c_3 from *Desulfovibrio desulfuricans* Norway at 1.7 Å resolution. *J. Mol. Biol.* **243**, 653-667.

Davenport, H.E., Hill, R. and Whatley, F.R. (1952). A natural factor catalysing reduction of methaemoglobin by isolated chloroplasts. *Proc. Roy. Soc. (London)* **B139**, 327-345.

DeGrado, W.F., Wasserman, Z.R. and Lear, J.D. (1989). Protein design, a minimalist approach. *Science* **243**, 622-628.

Deklerk, H. and Kamen, M.D. (1966). A high potential non-haem iron protein from the faculative photoheterotroph *Phodopseudomonas gelatinosa. Biochim. Biophys. Acta* **112**, 175-178.

Delwiche, C.C. and Bryan, B.A. (1976). Denitrification. *Ann. Rev. Microbiol.* **30**, 241-262.

Dickerson, R.E. (1980). Cytochrome c and the evolution of energy metabolism. *Sci. Am.* **242**, 3:98-109.

Dickerson, R.E., Kendrew, J.C. and Strandberg, B.E. (1961). The crystal structure of myoglobin: Phase determination to a resolution of 2.0 Å by the method of isomorphous replacement. *Acta Cryst.* **14**, 1188-1195.

Dickerson, R.E., Kopka, M.L., Weinzierl, J., Varnum, J., Eisenberg, D. and Margolish, E. (1967). Location of the haem in horse heart ferricytochrome c by X-ray diffraction. *J. Biol. Chem.* **242**, 3015-3018.

Dickerson, R.E., Takano, T., Eisenberg, D., Kallai, O.B., Samson, L., Copper, A. and Margoliash, E. (1971). General features of the horse and bonito proteins at 2.8 Å resolution. *J. Biol. Chem.* **246**, 1511-1535.

Dickerson, R.E. (1980). The cytochromes c. An exercise in scientific serendipity. In: *The evolution of protein structure and function.* (eds.) Sigman, D.S. and Brazier, M.A., Academic press, London New York, pp 397-547.

Dickinson, F.M. and Hart, G.J. (1982). Efects of Mg^{2+}, Ca^{2+} and Mn^{2+} on sheep liver aldehyde dehydrogenase. *Biochem. J.* **205**, 443-448.

Dickinson, F.M. and Haywood, G.W. (1986). The effects of Mg^{2+} on certain steps in the mechanism of the dehydrogenase and esterase reactions catalsed by sheep liver aldehyde dehydrogenase. *Biochem. J.* **233**, 877-883.

Dickinson, F.M., Hart, G.J. and Kitson, T.M. (1981). The use of pH-gradient ion-exchange chromatography to separate sheep liver cytoplasmic aldehyde dehydrogenase from mitochondrial enzyme contamination, and observations on the interactions between cytoplasmic enzyme and disulfiram. *Biochem. J.* **199**, 573-579.

Dietrich, R.A. (1966). Tissue and subcellular distribution of mammalian aldehyde-oxidising capacity. *Biochem. Pharmacol* **15**, 1911-1922.

Dodson, E.J. (1985). Molecular replacement: The methods and its problems. In *Molecular Replacement*. Proc. of the Daresbury Study Wekend, (ed.) P. Machin, pp.33-45. SERC Daresbury Laboratory, Warrington, England.

Duine, J.A. and Frank, J. (1980). Studies on methanol dehydrogenase from *Hyphomicrobium x. Biochem. J.* **187**, 213-219.

Duncan, R.J.S. (1977). The actions, of progesterone and diethylstilbestrol on the dehydrogenase and esterase activities of a purified aldehyde dehydrogenase from rabbit liver. *Biochem. J.* **161**, 123-130.

Duncan, R.J.S. (1985). Aldehyde dehydrogenase : An enzyme with two distinct catalytic activities at a single type of active site. *Biochem. J.* **230**, 261-267.

Eckfelt, J.H. and Yonetani, T. (1976a). Subcellular localization of the E_1 and E_2 isozymes of horse liver aldehyde dehydrogenase. *Arch. Biochem. Biophys.* **173**, 273-281.

Egami, F., Ishimoto, M. and Tamiguichi, S. (1961). The electron transfer from cytochromes to terminal electron acceptors in nitrate respiration and sulphate respiration : in "*Haematin Enzymes*" I, (ed.) J. Falk *et al.*, Pergamon Press pp. 392-406.

Emptage, M.H., Zimmerman, R., Que, L., Münck, E., Hamilton, W.D. and Orme-Johnson, W.H. (1977). Mössbauer studies of cytochrome c' from *R. rubrum*. *Biochim. Biophys. Acta* **495**, 12-23.

Emptage, M.H., Xavier, A.V., Wood, J.M., Alsaadi, B.M., Moore, G.R., Pitt, R.C., Williams, R.J.P., Ambler, R.P., and Bartsch, R.G. (1981). NMR studies of *R. rubrum* cytochrome c'. *Biochemistry* **20**, 58-64.

Falk, J.E. (1964). Porphyrins and metalloproteins; their general, physical and coordination chemistry, and laboratory methods. (ed.) J.E. Falk, Elsevier, Amsterdam.

Fairwell, T., Krutzsch, H., Hempel, H., Jeffery, J. and Jörnvall, H. (1984). Acetyl-blocked N-terminal structures of sorbitol and aldehyde dehydrogenases. *FEBS Letters* **170**, 281-289.

Feldman, R.I. and Weiner, H. (1972). Horse liver aldehyde dehydrogenase. I. purification and characterization. *J. Biol. Chem* . **247**, 260-266.

Finzel, B.C. Weber, P.C., Hardman, K.D. and Salemme, F.R. (1985). Structure of ferricytochrome c' from *Rhodospirillum molischianum* at 1.67Å resolution. *J. Mol. Biol.* **186**, 627-643.

Fukuyama, K., Nagahara, Y., Tsukihara, T., Katsube, Y., Hase, T. and Matsubara, H. (1988). Tertiary structure of Bacillus thermoproteolyticus [4Fe-4S] ferredoxin: Evolutionary implications for bacterial ferredoxins. *J. Mol. Biol.* **199**, 183-193.

Gray, T.M. and Matthews, B.W. (1984). Intrahelical hydrogen bonding of serine, threonine and cysteine residues within α-helices and its relavance to membrane bound proteins. *J. Mol. Biol.* **175**, 75-81.

Green, D.W., Ingram, V.M. and Perutz, M.F. (1954). The structure of haemoglobin IV. Sign determination by the isomorphous replacement method. *Proc. Roy. Soc. (London)* **A225**, 287-307.

Greenfield, N.J. and Pietruszko, R. (1977). Two aldehyde dehydrogenases from human liver. *Biochem. Biophys. Acta.* **483**, 35-45.

Guss, J.M. and Freeman, H.C., (1983). Structure of oxidised poplar plastocyanin at 1.6 Å resolution. *J. Mol. Biol.* **169**, 521-563.

Guss. J.M., Harrowell, P.R., Murata, M., Norris, V.A. and Freeman, H.C. (1986). Crystal structure analyses of reduced (Cu^{I}) poplar plastocyanin at six pH values. *J. Mol. Biol.* **192**, 361-387.

Guss, J.M., Merritt, E.A., Phizackerley, R.P., Hedman, B, Murata, M., Hodgson, K.O. and Freeman, H.C. (1988). Phase determination by multiple-wavelength x-ray diffraction: Crystal structure of a basic "blue" copper protein from cucumbers. *Science* **241**, 806-811.

Hahn, K.W., Klis, W.A. and Stewart, J.M. (1990). Design and synthesis of a peptide having a chymotrypsin-like esterase activity. *Science* **248**, 1544-1547.

Harada, S., Agarwal, D.P. and Goedde, H.W. (1981). Aldehyde dehydrogenase deficiency as cause of facial flushing reaction to alcohol in Japanese. *Lancet.* **2**, 982-

Harada, S., Yasui, M., Murakawa, K., Kasai, N. and Satow, Y. (1986). Crystal structure analysis of cytochrome c' by the multiwavelength anomalous diffraction method using synchrotron radiation. *J. Appl. Cryst.* **19**, 448-452.

Harada, Y., Lifchitz, A., Berthou, J. and Jolles, P. (1981). A translation function combining packing and diffraction information: An application to lysozyme (high-temperature form). *Acta Cryst.* **A37**, 398-406.

Harker, D. (1956). The determination of the phases of the structure factors of non-centrosymmetric crystals by the method of double isomorphous replacement. *Acta Cryst.* **9**, 1-9.

Harrington, M.C., Henehan, G.T.M. and Tipton K.F. (1987). The roles of human aldehyde dehydrogenase isoenzymes in ethanol metabolism. In: *Enzymology andMolecular Biology of Carbonyl Metabolism* (ed.) A.R. Liss, 111-125.

Haser, R., Pierrot, M., Frey, M., Payan, F., Astier, J.P., Bruschi, M. and Le-Gall, J. (1979). Structure and sequence of the multihaem cytochromes c_3. *Nature (London)* **282**, 806-810.

Hawkins, R.D. and Kalant, H. (1972). The metabolism of ethanol and its metabolic effects. *Pharmacol. Rev.* , **24** , 67-157.

Hecht, M.H., Richardson, J.S., Richardson, D.C. and Ogden, R.C. (1990). *De novo* design, expression and characterisation of Felix: a four-helix bundle of native like sequence. *Science* **249**, 884-891.

Hempel, J.R. and Jörnvall, H. (1987). Functional topology of aldehyde dehydrogenase structures. In: *Enzymology and Molecular Biology of carbonyl metabolism.* (ed.) A.R. Liss, 1-14.

Hempel, J.H., Von Bahr-Lindström, H. and Jörnvall, H. (1984). Aldehyde dehydrogenase from human liver; Primary structure of the cytoplasmic isoenzyme. *Eur. J. Biochem* . **141**, 21-35.

Hempel, J., Kaiser, R. and Jörnvall, H. (1985). Mitochondrial aldehyde dehydrogenase from human liver: Primary structure, differences in relation to the cytosolic enzyme and functional correlations. *Eur. J. Biochem.*. **153**, 13-28.

Hempel, J., Nicholas, H. and Jörnvall, H. (1990). Thiol Proteases and Aldehyde Dehydrogenases: Evolution from a common Thiolesterase Precursor? *Proteins:Structure, Function, and Genetics* **11**, 176-183.

Hempel, J.H., Von Bahr-Lindström, H. and Jörnvall, H. (1983). Structural relationships among alcohol dehydrogenases. *Pharm. Biochem.& Behav.* **18. suppl.1**, 117-121.

Hendrickson, W.A. (1984). Measurement and use of anomalous scattering. *Acta Cryst.* **A40**, C-3.

Hendrickson, W.A. (1985). Analysis of protein structure from diffraction measurement at multiple wavelengths. *Trans. Am. Cryst. Assoc.* **21**, 11-21.

Hendrickson, W.A., Horton, J.R. and LeMaster, D. M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**, 1665-1672.

Hendrickson, W.A. and Teeter, M.M. (1981). Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature* **290**, 107-113.

Hendrickson, W.A., Smith, J.L., Phizackerley, R.P. and Merritt, E.A. (1988). Crystallographic structure analysis of lamprey haemoglobin from anamalous dispersion of synchrotron radiation. *Proteins:structure, function, and genetics* **4**, 77-88.

Henry, Y. and Bessieres, P. (1984). Denitrification and nitrite reduction - *Pseudomonas aeruginosa* nitrite reductase. *Biochimie* **66**, 259-289.

Higashi, T. (1989). The processing of diffraction data taken on a screenless Weissenberg camera for macromolecular crystallography. *J. Appl. Cryst.* **22**, 9-18.

Hill, R. (1954). The cytochrome b component of chloroplasts. *Nature (London)* **174**, 501-503.

Hill, R. and Scarisback, R. (1951). The haematin compounds of leaves. *New Phytologist* **50**, 98-111.

Hirshfeld, F.L. (1968). Symmetry in the generation of trial structures. *Acta Cryst.* **A24**, 301-311.

Ho, S.P. and DeGrado, W.F. (1987). Design of a four-helix bundle protein : Synthesis of peptides which self associate into a helical protein. *J. Am. Chem. Soc.* **109**, 6751-6758.

Hoe, S.T., Nicholson, A., Choyce, M., Ting, H. and Crabbe, M.J.C. (1985). Bovine lens aldehyde dehyrogenase: Improved purification and characterisation. *Enzymology of Carbonyl Metabolism 2: Aldehyde Dehydrogenase, Aldo-KetoReductase, and Alcohol Dehydrogenase* (ed.) A.R. Liss, 137-148.

Hol, W.G.J. (1985). The role of the α-helix dipole in protein function and structure. *Prog. Biophys. molec. Biol.* **45**, 149-195.

Hol, W.G.J., Van Duijnen, P.T. and Berendsen, H.J.C. (1978). The α-helix dipole and the properties of proteins. *Nature* **273**, 443-446.

Hon-nami, K. and Oshima, T. (1980). Cytochrome oxidase from an extreme thermophile, *Thermus thermophilus* HB8. *Biochem. Biophys. Res. Commun.* **92**, 1023-1029.

Hoppe, W. and Jakubowski, V. (1975). The determination of phases of erythrocruorin using the two-wavelength method with iron as anomalous scatterer. In: *Anomalous scattering.* (eds.) Ramaseshan, S. and Abrahams, S.C., Copenhagen:Munksgaard, pp 437-461.

Howard, A.J., Nielsen, C. and Xuong, N.H. (1985). Software for a diffractometer with multiwire area detector. *Meth. Enzymol.* **114**, 452-472.

Huber, R. and Schneider, M. (1985). A group refinement procedure in protein crystallography using Fourier transforms. *J. Appl. Cryst.* **18**, 165-169.

Hunter, C.A., Singh, J. and Thornton, J.M. (1991). π-π interactions: the geometry and energetics of phenylalanine-phenylalanine interactions in proteins. *J. Mol. Biol.* **218**, 837-846.

Hurley, T.T. and Weiner, H. (1992). Crystallisation and preliminary X-ray investigation of bovine liver mitochrondrial aldehyde dehydrogenase. *J. Mol. Biol.* **277**, 1255-1257.

Jancarik, J. and Kim, S-h. (1991). Sparse matrix sampling: a screening method for crystallisation of proteins. *J. Appl. Cryst.* **24**, 409-411.

Janin, J., Wodak, S., Levitt, M. and Maigret, M. (1978). Conformation of amino acid sidechains in proteins. *J. Mol. Biol.* **125**, 357-386.

Johansson , J., Von Bahr-Lindström, H., Jeck, R., Woenckhans, C. and Jörnvall, H. (1988). Mitochondrial aldehyde dehydrogenase from horse liver: Correlations of the same species variants for both the cytosolic and the mitochondrial forms of an enzyme. *Eur. J. Biochem.* **172**, 527-533.

Jones, D.G., Brennan, M.D., Hempel, J. and Lindahl, R. (1988). Cloning and complete nucleotide sequence of a full length cDNA encoding a catalytically functional tumour-associated aldehyde dehydrogenase. *Proc. Natl. Acad. Sci. USA* **85**, 1782-1786.

Jones, T.A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Cryst.* **11** , 268-272.

Jones, T.A. (1982). TOM (version 2.4): A graphics fitting for macromolecules, in Computational Crystallography, (ed.) D. Sayre, Clarendon Press, Oxford pp 303-317.

Kakuno, T., Bartsch, R.G., Nishikawa, K. and Hori, T. (1971). Redox components associated with chromophores from *Rhodospirillum rubrum*. *Biochem J.***70**, 75-94.

Kassner, R.J. (1991). Ligand binding properties of cytochromes c'. *Biochim. Biophys. Acta* **1058**, 8-12.

Kikuchi, G., Saito, Y. and Motokawa, Y. (1965). On cytochrome oxidase as the terminal oxidase of dark respiration of non-sulphur purple bacteria. *Biochim. Biophys. Acta* **94**, 1-14.

Kitigawa, T., Ozaki, Y.,Kyogoku, Y. and Horio, T. (1977). The pH dependence of the reasonance Raman spectra and structural alterations at haem moieties of various c-type cytochromes. *Biochim. Biophys. Acta* **495**, 1-11.

Kitson, T.M. (1975). The effect of disulphiram on the aldehyde dehydrogenases of sheep liver. *Biochem. J.* **151**, 407-412.

Kitson, T.M. (1977). The disulphiram-ethanol reaction: A review. *Journal of studies on alcohol* **vol38 no. 1**, 96-113.

Kitson, T.M. and Freeman, G.H. (1993). 3,4-dihydro-3-methyl-6-nitro2*H*-1,3-benoxazin-2-one, a reagent for labeling p-nitrophenol esterases with a chromophoric reporter group - synthesis and reaction with chymotrypsin. *Bioorg. Chem.* **21**, 354-365.

Kitson, T.M. and Kitson, K.E. (1994). Probing the active site of cytocplasmic aldehyde dehydrogenase with a chromophoric reporter group. *Biochem J.* **300**, 25-30.

Koivula, T. and Koivusalo, M. (1975). Partial purification and properties of a phenobarbital-induced aldehyde dehydrogenase of rat liver. *Biochem. Biophys. Acta* **410**, 1-11.

Konnert, J.H. and Hendrickson, W.A. (1980). A restrained-parameter thermal-factor refinement procedure. *Acta Cryst.* **A36**, 344-350.

Korzun, Z.R. and Salemme, F.R. (1977). Structure of cytochrome c-555 from *Chlorobium thiosulphatophilum*: primitive low-potential cytochrome c. *Proc. Natl. Acad. Sci. USA* **74**, 5244-5247.

Kraulis, P.J.J. (1991). Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950.

Kuhn, L.A., Siani, M.A., Pique, M.E., Fisher, C.L., Getzoff, E.D. and Tainer, J.A. (1992). The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density density measures. *J. Mol. Biol.* **228**, 13-22.

Kuriyan, J., Wilz, S., Karplus, M. and Petsko, G.A. (1986). X-ray structure and refinement of carbon-monoxy (Fe II)-myoglobin at 1.5 Å resolution. *J. Mol. Biol.* **192**, 133-154.

Kuronen, T., Saraste, M. and Ellfolk, N. (1975). The subunit structure of *Pseudomonas* cytochrome oxidase. *Biochim Biophys. Acta* **393**, 48-54.

La Mar, G.N., Jackson, J.T. and Bartsch, R.G. (1981). Analysis of field-dependent relaxation data and hyperfine shifts of cytochrome c' from *R. rubrum* in terms of the high spin iron ligation state. *J. Am. Chem. Soc.* **103**, 4405-4410.

Landrum, J.T., Hatano, K., Scheidt, W.R. and Reed, C.A. (1980). Imidazolate complexes of iron and manganese tetraphenylporphyrins. *J. Am. Chem. Soc.* **102**, 6729-6735.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993). Procheck: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291.

Lattman, E.E. and Love, W.E. (1972). A rotational search procedure for detecting a known molecule in a crystal. *Acta Cryst.* **B26**, 1854-1857.

Leahy, J.L., Hendrickson, W.A., Aukhil, I. and Erickson, H.P. (1992). Structure of a Fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science* **258**, 987-991.

Leicht, Heinz, W.F. and Freimuller, B. (1978). Purification and characterisation of aldehyde dehydrogenase from bovine liver. *Eur. J. Biochem.* **83**, 189-196.

Lemberg, R. and Barrett, J. (1973). The cytochromes. Academic Press, London, New York

Lewis, P.N., Momany, F.A. and Scheraga, H.A. (1973). Chain reversals in proteins. *Biochim. et Biophys. Acta* **303**, 211-229.

Lifchitz, A.(1983). On the choice of the model cell and the integration volume in the use of the rotation function. *Acta Cryst.* **A39**, 130-139.

Lindahl, R. and Evces, S. (1984). Rat liver aldehyde dehydrogenase. II. Isolation and characterisation of four inducible isoenzymes. *J. Biol. Chem.* **259**, 11991-11996.

Lindahl., R. and Hemple., J. (1991). Aldehyde dehydrogenases: What can be learned from a Baker's dozen sequences. In: *Enzymology and molecular biology of carbonyl metabolism 3*, (eds.) Weiner, H., Wermuth, B. and Crabb, D.W. Plenum Press, New York and London. pp 1-8.

Luzzati, V. (1952). Traitment statistique des erreurs dans la determination des structures cristallins. *Acta. Cryst.* **5**, 802-810.

MacMunn, C.A. (1886). Researches on myohaematin and the histohaematins. *Philosophical Transactions Roy Soc.* (London) **177**, 267-298.

MacMunn, C.A. (1887). Further observations on myohaematin and the histohaematins. *J. Physiol.* **8**, 51-65.

Makinen, M.W. and Churg, A.K. (1983). Structural and analytical aspects of the electronic spectra of haem proteins. In: *Iron Porphyrins part one* (eds.) Lever, A.P.B. and Gray, H.B., Addison-Wesley, London Amsterdam, pp 141-235.

Maltempo, M.M. (1974). Magnetic state of an unusual bacterial haem protein. *J. Chem. Phys.* **61**, 2540-2547.

Maltempo, M.M. (1976). Visible absortion spectra of quantum mixed-spin ferric haem proteins. *Biochim. et Biophys. Acta* **434**, 513-518.

Maltempo, M.M. and Moss, T.H. (1976). The spin 3/2, 5/2 state and quantum mixed-spin ferric haemproteins. *Q. Rev. Biophys.* **9**, 181-215.

Maltempo, M.M., Moss, T.H. and Cusanovich, M.A. (1974). Magnetic studies on the changes in the iron environment in *Chromatium* ferricytochrome c. *Biochim. et Biophys. Acta* **342**, 290-305.

Marjanen, O.L. (1972). Intracellular localization of aldehyde dehydrogenase in rat liver. *Biochem. J.* **127**, 633-639.

Masuda, H., Taga, T., Osaki, K., Sugimoto, H., Yosida, Z.I. and Ogoshi, H. (1980). Crystal and molecular structure of (Octaethylporphinato)iron (III) perchlorate: Anomalous magnetic and properties and structural aspects. *Inorg. Chem.* **19**, 950-955.

Matsuura, Y., Takano, T. and Dickerson, R.E. (1982). Structure of cytochrome c-551 from *Ps. aeruginosa* refined at 1.6 Å resolution and comparison of the two redox forms. *J. Mol. Biol.* **156**, 389-409.

Mathews, F.S. (1985). The structure, function and evolution of cytochromes. *Prog. Biophys. molec. biol.* **45**, 1-56.

Mathews, F.S., Bethge, P.H. and Czerwinski, E.W. (1979). The structure of cytochrome b_{562} from *E. coli* at 2.5 Å resolution. *J. Biol. Chem.* **254**, 1699-1706.

Matthews, B.W. (1968). Solvent content in protein crystals. *J. Mol. Biol.* **33**, 491-497.

Matthews, B.W., Fenna, R.E., Bolognesi, M.C., Schmid, M.F. and Olson, J.M. (1979). Structure of a bactriochlorophyll a-protein from the green photosynthetic bacterium *Prosthecochloris aestuarii*. *J. Mol. Biol.* **131**, 259-285.

McEwan, A.G., Greenfield, A.J., Wetzstein, H.G., Jackson, J.B. and Ferguson, S.J. (1985). Nitrous oxide reduction by *Rhodospirillaceae* and the nitrous oxide reductase of *Rhodopseudomonas capsulata*. *J. Bacteriol.* **164**, 823-830.

McPherson, A. (1990). Current approaches to molecular crystallisation (Review). *Eur. J. Biochem.* **189**, 1-23.

McPherson, A., Koszlak, S., Axelrod, H., Day, J., Williams, R., Robinson, L., McGrath , M., and Cascio, D. (1986). An experiment regarding crytsallisation of soluble proteins in the presence of β-octyl glucoside. *J. Mol. Biol.* **261**, 1969-1975.

McRee, D.E., Ren, Z. and Meyer, T.E. (1993). Atomic structure of a cytochrome c' with an unusual ligand-controlled dimer dissociation at 1.8 Å resolution. *J. Mol. Biol.* **234**, 433-445.

Meyer, T.E. and Kamen, M.D. (1982). New perspectives on c-type cytochromes. *Adv. Prot. Chem.* **35**, 105-212.

Moore, M.H., Gulbis, J.M., Dodson, E.J., Demple, B. and Moody, P.C.E. (1994). Crystal structure of a suicidal DNA repair protein: the Ada O^6-methylguanine-DNA methyltransferase from *E. coli*. *EMBO J.* **13**, 1495-1501.

Moore, G.R., McClune, G.J., Clayden, N.J., Williams, R.J.P., Alsaadi, B.M., Ångström, J., Ambler, R.P., Van Beeuman, J., Tempst, P., Bartsch, R.G., Meyer, T.E. and Kamen, M.D. (1982). Metal coordination centres of class II cytochromes c. *Eur. J. Biochem.* **123**, 73-80.

Morii, H., Ichimura, K. and Uedaira, H. (1991). Asymmetric inclusion by de novo designed proteins : Fluorescence probe studies on ampiphilic α-helix bundles. *Proteins: Structure, Function, and Genetics* **11**, 133-141.

Morita, S. (1968). Evidence for three photchemical systems in *Chtomatium* D. *Biochim. Biophys. Acta* **153**, 241-247.

Morita, S. and Conti, S.F. (1963). Localisation and nature of the cytochromes of *Rhodomicrobium vanniellii. Arch. Biochem. Biophys.* **100,** 302-307.

Morris, A.L., MacArthur, M.W., Hutchinson, E.G. and Thornton, J.M. (1992). Stereochemical quality of protein structure coordinates. *Proteins:Structure, Function, and Genetics* **12**, 345-364.

Moss, D.S. (1985). Symmetry and the rotation function. In: *Molecular Replacement.* Proc. of the Daresbury Study Wekend, (ed.) P. Machin, pp.8-13. SERC Daresbury Laboratory, Warrington, England.

Motion, R. L. (1986). Structural and Mechanistic Studies of Sheep Liver Aldehyde Dehydrogenase. Ph.D. Thesis, Massey University, Palmerston North, New Zealand.

Mowbray, S.L. and Petsko, G.A. (1983). The introduction of specific sites for heavy metal binding in a crystalline protein *J. Biol. Chem.* **258**, 5634-5632.

Nappa, M., Valentine, J.S. and Synder, P.A. (1977). Imidazolate complexes of ferric porphyrins. *J. Am. Chem. Soc.* **99**, 5799-5800.

Navaza, J. (1994). AMoRe: An automated package for molecular replacement. *Acta Cryst.* **A50**, 157-163.

Newton, N. (1969). The two haem nitrite reductase of *Micrococcus denitrificans. Biochim. Biophys. Acta* **185**, 316-331.

Norris, G.E., Anderson, B.F., Baker, E.N. and Rumball, S.V. (1979). Purification and preliminary crystallographic studies on azurin and cytochrome c' from *Alcaligenes denitrificans* and *Alcaligenes* sp NCIB 11015. *J. Mol. Biol* **135**, 309-312.

Norris, G.E., Anderson, B.F. and Baker, E.N. (1983). Structure of azurin from *Alcaligenes denitrificans* at 2.5 Å resolution. *J. Mol. Biol.* **165**, 501-521.

Olson, J.M. and Chance, B. (1960). Oxidation reduction reactions in the photosynthetic bacterium *Chromatium* - absorption spectrum changes in whole cells. *Arch. Biochem. Biophys* **88**, 26-53.

Olson, J.M. and Nadler, K.D. (1965). Engery transfer and cytochrome function in a new type of photosynthetic bacterium. *Photochem. Photobiol.* **4**, 783-791.

Otwinowski, Z.O. (1991). Maximum Likelihood refinement of heavy atom parameters In: *Proceedings of the CCP4 study weekend*, (eds.) Wolf, W., Evans, P.R. and Leslie, A.G.W., pp 80-86, SERC Daresbury Laboratory, Warrington, UK.

Paliaskasis, C.D. and Kokkinidis, M. (1992). Relationships between sequence and structure for the four-α-helix bundle tertiary motif in proteins. *Protein Engineering* **Vol. 5 no. 8**, 739-748.

Parilla, R., Ohkawa, K., Lindros, K.O., Zimmerman, U.J.P., Kobayashi, K. and Williamson, J.R. (1974). Functional compartmentation of acetaldehyde oxidation in rat liver. *J. Biol. Chem* . **249**, 4926-4933.

Parr, S.R., Barber, D., Greenwood, C., Phillips, B.W. and Melling, J. (1976). A purification procedure for the soluble cytochrome oxidase and some other respiratory proteins from *Pseudomonas aeruginosa*. *Biochem. J.* **157**, 423-430.

Patterson, A.L. (1934). A Fourier series method for the detemination of the components of interatomic distances in crystals. *Phys. Rev.* **46**, 372.

Payne, W.J. (1981). Denitrification. Wiley Interscience, New York.

Perutz, M.F., Fermi, G., Abraham, D.J., Poyart, C. and Bursaux, E. (1986). Haemoglobin as a receptor of drugs and peptides: X-ray studies of the stereochemistry of binding. *J. Am. Chem. Soc.* **108**, 1064-1078.

Petsko, G.A. (1985). Preparation of isomorphous heavy-atom derivatives *Meth. Enzymol.* **114**, 147-157.

Petsko, G.A., Phillips, D.C., Williams, R.J.P. and Wilson, I.A. (1978). On the protein crystal chemistry of chloroplatinate ions: General principles and interactions with triose phosphate isomerase. *J. Mol. Biol.* **120** , 345-359.

Pettigrew, G.W. and Moore. G.R. (1987). In: Cytochromes c: Biological aspects. Springer-Verlag, New York Heidelberg.

Pettigrew, G.W. and Moore. G.R. (1990). In: Cytochromes c: Structural aspects. Springer-Verlag, New York Heidelberg.

Pietruszko, R., and Mackerell Jr., A.D. (1987). Stoichiometry of chemical modification of human aldehyde dehydrogenase: Evidence for "Quarter of the sites" reactivity. In:*Enzymology and Molecular Biology ofCarbonyl Metabolism* (ed.) A.R. Liss, 37-52.

Pietruszko, R. and Yonetani, T. (1981). Aldehyde dehydrogenases from liver *Meth. Enz.* **71**, 772-781.

Pierrot, M., Haser, R., Frey, M., Payan, F. and Astier, J-P. (1982). Crystal structure and electron transfer properties of cytochrome c_3. *J. Biol. Chem.* **23**, 14341-14349.

Postgate, J. (1959). Sulphate reduction by bacteria. *Ann. Rev. Microbiol.* **13**, 505-520.

Postgate, J. (1961). Cytochrome c_3 : In "*Haematin Enzymes: a symposium of the international union of biochemistry, organised by the Australian Academy of Science*", (eds.) J.E. Falk *et al.*, Pergamon press, pp. 407-414.

Poulos, T.L. and Kraut, J. (1980). A hypothetical model of the cytochrome c peroxidase: cytochrome c electron transfer complex. *J. Biol. Chem.* **255**, 10322-10330.

Presta, L.G. and Rose, G.D. (1988). Helix signals in proteins, *Science* **240**, 1632-1641.

Prince, R.C., Baccarini-Melandri, A., Hauska, G.A., Melandri, B.A. and Crofts, A.R. (1975). Asymmetry of an energy transducing membrane and the location of cytochrome c_2 in *Rhodopseudomonas sphaeroides* and *Rhodopseudomonas capsulata. Biochim. Biophys. Acta* **387**, 212-227.

Prince, R.C., Leigh, J.S. and Dutton, P.L. (1974). An electron spin reasonance characterisation of *Rhodopseudomonas capsulata. Biochem. Soc. Trans.* **2**, 950-953.

Quillan, M.L., Arduini, R.M., Olson, J.S. and Phillips Jr, G.N. (1993). High-resolution crystal structures of distal histdine mutants of sperm whale myoglobin. *J. Mol. Biol.* **234**, 140-155.

Racker, E. (1949). Aldehyde dehydrogenase, a diphosphopyridine nucleotide-linked enzyme. *J. Biol.Chem.* **177**, 883-892.

Rao, S.N., Jih, J.-H., and Hartsuck, J.A. (1980). Rotation-function space groups. *Acta Cryst.* **A36**, 878-884.

Ramakrishnan, C. and Ramachandran, G.N. (1965). Stereochemical criteria for polypeptide and protein chain confromations II. Allowed conformations for a pair of peptide units. *Biophys. J.* **5**, 909-933.

Rayment, I., Wessenberg, G., Meyer, T.E., Cusanovich, M.A. and Holden, M. (1992). 3D structure of the high-potential iron-sulpur protein isolated from the purple phototropic bacterium *Rhodocyclus tenuis* determined and refined at 1.5 Å resolution. *J. Mol. Biol.* **228**, 672-686.

Rawlings, J., Stephens, P.J., Nafie, L.A. and Kamen, M.D. (1977). Near-infrared magnetic cicular dicroism of cytochrome c'. *Biochemistry* **16**, 1735-1729.

Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta. Cryst.*, **A42** , 140 - 149.

Reed, C.A., Mashiko, T., Bently, S.P., Kastner, M.E., Scheidt, W.R., Spartalian, K. and Lang, G. (1979). The missing haem spin state and a model for cytochrome c'. The mixed S=3/2, 5/2 intermediate spin ferric porphyrin: Perchlorato (meso-tetraphenylporphinato)iron (III). *J. Am. Chem. Soc.* **101**, 2948-2958.

Regan, L. and DeGrado, W.F. (1988). Characterisation of a helical protein designed from first principles. *Science* **241**, 976-978.

Richardson, J.S. and Richardson, D.C. (1988). Amino acid preferences for specific locations at the ends of α-helices. *Science* **240**, 1648-1652.

Rose, J.P., Hempel, J., Kuo, I., Lindahl, R., and Wang, B-C. (1991). Prelimary crystallographic analysis of class 3 rat liver aldehyde dehydrogenase. *Proteins: Structure, Function, and Genetics* **8**, 305-308.

Rossman, M.G. and Blow, D.M. (1962). The Detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.* **15**, 24-31.

Rossman, M., Liljas, A., Branden, C.- I. and Banaszak, L.J. (1976). Evolutionary and structural relationships among dehydrogenases, in *The Enzymes*, Vol 11, 3rd ed., (ed.) Boyer, P.D., Academic Press, New York, pp 61.

Ryzlak, M.T. and Pietruszko, R. (1987). Purification and characterisation of aldehyde dehydrogenase from human brain. *Arch. Biochem. Biophys.* **255**, 409-418.

Sakabe, N. (1991). X-ray diffraction data collection system for nodern protein crystallography with a Weissenberg camera and an imaging plate using synchrotron radiation. *Nuclear Instruments and Methods in Physics Research* **A303**, 448-463.

Salemme, F.R. (1974). Preliminary crystallographic data for cytochrome c' of *Rhodopseudomonas palustris*. *Arch. Biochem. Biophys.* **163**, 423-425.

Salemme, F.R. (1976). An hypothetical structure for an intermolecular electron transfer complex of cytochrome c and cytochrome b_5. *J. Mol. Biol.* **102**, 563-568.

Salemme, F.R. (1977). Struture and function of cytochromes c. *Annu. Rev. Biochem.* **46**, 299-329.

Salemme, F.R., Freer, S.T., Xuong,. Ng. H., Alden, R.A. and Kraut, J. (1973). The structure of oxidised cytochrome c_2 of *Rhodospirillum rubrum*. *J. Biol. Chem.* **248**, 3910-3921.

Salemme, F.R., Genieser, L, Finzel, B.C., Hilmer, R.M. and Wendoloski, J.J. (1988). Molecular factors stabilising protein crystals. *Journal of Crystal Growth* **90**, 273-282.

Sato, R. and Egami, F. (1949). Studies on nitrate reductase. *Bull. Chem. Soc. Japan* **22**, 137-141.

Scheidt, W.R. and Reed, C.A. (1981). Spin-state/stereochemical relationships in iron porphyrins: Implications for the haemoproteins. *Chem. Rev.* **81**, 543-555.

Scouloudi, H., and Baker, E.N. (1978) X-ray crystallographic studies of seal myoglobin: The molecule at 2.5 Å resolution. *J. Mol. Biol.* **126**, 637-660.

Shepard, W.E.B. (1991). Crystallographic analysis of apo and reduced azurin. Thesis, Ph.D., Massey University, New Zealand.

Sheriff, S., Hendrickson, W.A. and Smith, J.L. (1987). Structure of myohemerythrin in the azidomet state at 1.7/1.3 Å resolution *J. Mol. Biol.* **197**, 273-296.

Shidara, S. (1980). Components of the cytochrome system of *Alcaligenes* sp NCIB 11015 with special reference to particulate bound c-type cytochromes. *J. Biochem.* **87**, 1177-1184.

Siew, C., Deitrich, R.A. and Erwin,V.G. (1976). Location and characteristics of rat liver mitochondrial aldehyde dehydrogenase. *Arch. Biochem. Biophys.* **176**, 638-649.

Sigler, P.B. and Blow, D.M. (1965). A means of promoting heavy-atom binding in protein crystal. *J. Mol. Biol.* **14**, 640-644.

Smith, J.L., Pahler, A., Murthy, H.M.K. and Hendrickson, W.A. (1987). Multiple-wavelength phase dertermination in proteins crystals. *Acta Cryst.* **A43**, C-10.

Steigemann, W. and Weber, E. (1979). Structure of erythrocuorin in different liganded states refined at 1.4 Å resolution. *J. Mol. Biol.* **127**, 309-338.

Stickle, D.F., Presta, L.G., Dill, K.A. and Rose, G.D. (1992). Hydrogen bonding in globular proteins. *J. Mol. Biol.* **226**, 1143-1159.

Stryer, L., Kendrew, J.C. and Watson, H.C. (1964). The mode of attachment of the azide ion to sperm whale myoglobin. *J. Mol. Biol.* **8**, 96-104.

Stura, E.A. and Wilson, I.A. (1990). Analytical and production seeding techniques. *METHODS:A companion to methods in enzymology* **Vol.1, No.1**, August, pp. 28-49.

Suzuki, H. and Iwasaki, H. (1962). Azurin and cytochrome c in a denitrifying bacterium. *J. Biochem.* **52**, 193-199.

Svanas, G.W. and Weiner, H. (1985). Aldehyde dehydrogenase activity as the rate-limiting factor for acetaldehyde metabolism in rat liver. *Arch. Biochem. Biophys.* **236**, 36-46.

Takano, T. (1977a). Structure of myoglobin refined at 2.0 Å resolution: I. Crystallographic refinement of metmyoglobin from sperm whale. *J. Mol. Biol.* **110**, 537-568.

Takano, T. (1977b). Structure of myoglobin refined at 2.0 Å II. Structure of deoxymyoglobin from sperm whale. *J. Mol. Biol.* **110**, 569-584.

Takano, T. and Dickerson, R.E. (1981a). Conformational change of cytochrome c: Ferricytochrome c structure refined at 1.5 Å resolution. *J. Mol. Biol.* **153**, 79-94.

Takano, T. and Dickerson, R.E. (1981b). Conformational change of cytochrome c: Ferricytochrome c refinement at 1.8 Å and comparison with the ferrocytochrome structure. *J. Mol. Biol.* **153**, 95-115.

Taniguchi, S. and Kamen, M.D. (1965). The oxidase system of heterotrophically grown *Rhodospirillum rubrum. Biochim. Biophys. Acta* **96**, 395-428.

Thanki, N., Thornton, J.M. and Goodfellow, J.M. (1988). Distributions of water around amino acid residues in proteins. *J. Mol. Biol.* **202**, 637-657.

Thanki, N., Umrania, Y., Thornton, J.M. and Goodfellow, J.M. (1991). Analysis of protein mainchain solvation as a function of secondary structure. *J. Mol. Biol.* **221**, 669-691.

Tickle, I.J. (1992). Molecular replacement, Procedings of the CCP4 study weekend, (eds.) E.J. Dodson, S. Gover and W. Wolf, pp. 20-32. Warrington: SERC Daresbury Laboratory.

Tickle, I.J. (1985). Review of spacegroup general translation functions that make use of known structure information and can be expanded as Fourier series. In *Molecular Replacement.* Proc. of the Daresbury Study Wekend, edited by P. Machin, pp.22-26. SERC Daresbury Laboratory, Warrington, England.

Tonge, G.M., Harrison, D.E.F. and Higgins, I.J. (1977).Purification and properties of the methane monooxygenase enzyme system from *Methylosinus trichosporium. FEBS lett* 58, 293-299.

Tottmar, S.O.C., Pettersson, H. and Kiessling, K. H. (1974). Alcohol and aldehyde metabolising systems, **1**, (ed.) Thurman. Academic Press. pp 147-160

Tronrud, D.E., Ten Ecyk, L.F. and Matthews, B.W. (1987). An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Cryst.* **A43**, 489-501.

Tsang, D.C.Y. and Suzuki, I. (1982). Cytochrome c-554 as a possible electron donor in the hydroxylation of ammonia and CO in *Nitrosomomas europaea. Can. J. Biochem.* **60**, 1018-1024.

Ugurbil, K. and Bersohn, R. (1977). Tyrosine emission in the tryptophan azurin from *Pseudomonas fluorescens. Biochemistry*, **16**, 895-900.

Valentine, J.S., Sheridan, R.P., Allen, L.C. and Kahn, P.C. (1979). Coupling between oxidation state and hydrogen bond conformation in haem proteins. *Proc. Natl. Acad. Sci. USA.* **76**, 1009-1013.

Van Pouderoyen, G., Mazumdar, S., Hunt, N.I., Hill, A.O., and Centers, G.W. (1994). The introduction of a negative charge into the hydrophobic patch of *Pseudomonas aeruginosa* azurin affects the electron self-exchange rate and the electrochemistry. *Eur. J. Biochem.* **222**, 583-588.

Varghese, J.N., Laver, W.G. and Colman, P.M. (1983). Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å. *Nature (London)* **303**, 35-40.

Vernon, L.P. and Kamen, M.D. (1954). Hematin compounds in photosynthetic bacteria. *J. Biol. Chem.* **211**, 643-663.

Von Bahr-Lindström, H., Hempel, J., and Jörnvall, H. (1984). The cytoplasmic isoenzyme of horse liver aldehyde dehydrogenase. *Eur. J. Biochem.* **141**, 37-42.

Wang, B-C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Meth. enzymol* **115**, 90-112.

Weber, E., Steigemann, W., Jones, T.A. and Huber, R. (1978). The structure of oxy-erythrocruorin at 1.4 Å resolution. *J. Mol. Biol.* **120**, 327-336.

Weber, P.C. (1982). Correlations between structural and spectroscopic properties of the high-spin haem protein cytochrome c'. *Biochemistry* **21**, 5116-5119.

Weber, P.C., and Salemme, F.R. (1980). Structural and functional diversity in four-α-helical proteins. *Nature* **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. and Salemme, F.R. (1980). Structure of cytochrome c': a dimeric high-spin protein. *Nature (London)* **286**, 302-305.

Weber, P.C., Howard, A., Xoung, N.H. and Salemme, F.R. (1981). Crystallographic structure of *Rhodospirillum molischianum* ferricytochrome c' at 2.5 Å resolution. *J. Mol. Biol.* **153**, 399-424.

Weiner, H. (1980). Aldehyde oxidising enzymes, In: *Enzymic Basis of Detoxification*, **1**, (ed.) Jakoby. pp 261-280.

Weiner, H. (1982). Aldehyde dehydrogenase. In: *Enzymology of Carbonyl Metabolism: Aldehyde Dehydrogenase and Aldo/Keto Reductase* (ed.) A.R. Liss, pp 1-9.

Williams, R.J.P. (1991). Uncoupled and coupled electron transfer reactions. *Biochim. Biophys. Acta.* **1058**, 71-74.

Wilson, A.J.C. (1949). The probability distribution or X-ray intensities. *Acta Cryst.* **2**, 318-321.

Wood, P.M. (1983). Why do c-type cytochromes exist? *FEBS Lett.* **164**, 223-226.

Wood, P.M. (1984). Bacterial proteins with CO-binding *b*- or *c*-type haem functions and absorption spectroscopy. *Biochim. Biophys. Acta* **768**, 293-317.

Xia, Z.X. and Mathews, F.S. (1992). Molecular structure of flavocytochrome b_2 at 2.4 Å resolution. *J. Mol. Biol.* **212**, 837-863.

Xuong, N.H., Nielsen, C., Hamlin, R. & Anderson, D. (1985). Strategy for data collection from protein crystals using a multiwire counter area detector diffractometer *J. Appl. Cryst.* **18**, 342-350.

Yaoi, H. and Tamiya, H. (1928). On the respiratory pigment, cytochrome, in bacteria. *Proc. Imp. Acad. Tokyo* **4**, 436-439.

Yamanaka, T. and Okunuki, K. (1974). Cytochromes. In: *Microbial iron metabolism.* (ed.) Neilands, J.B., Academic Press, London New York. pp 349-400.

Yasui, M., Harada, S., Kai, Y., Kasai, N.,Kusunoki, M. and Matsura, Y. (1992). Three-dimensional structure of ferricytochrome c' from *Rhodospirillum rubrum* at 2.8 Å resolution. *J. Biochem.* **111**, 317-324.

Yosida, A., Hsu, L. and Dave, V. (1992). Retinal oxidation activity and biological role of human cytosolic aldehyde dehydrogenase. *Enzyme* **46**, 239-244.

Yu, C.A., Yu, L. and King, T.E. (1974). Soluble cytochrome bc_1 complex and the reconstitution of succinate cytochrome c reductase. *J. Biol. Chem.* **249**, 4905-4910.

Zhang, K.Y.J. (1993) SQUASH: Combining constraints for macromolecular phase refinement and extension. *Acta. Cryst.* **D49**, 213-222.