THE THREE DIMENSIONAL STRUCTURE OF

AZURIN, A BLUE COPPER PROTEIN, AT 3Å

RESOLUTION

A thesis presented in partial fulfilment of the

requirements for the degree of Doctor of Philosophy

in Chemistry at Massey University.

GILLIAN E. NORRIS

1982
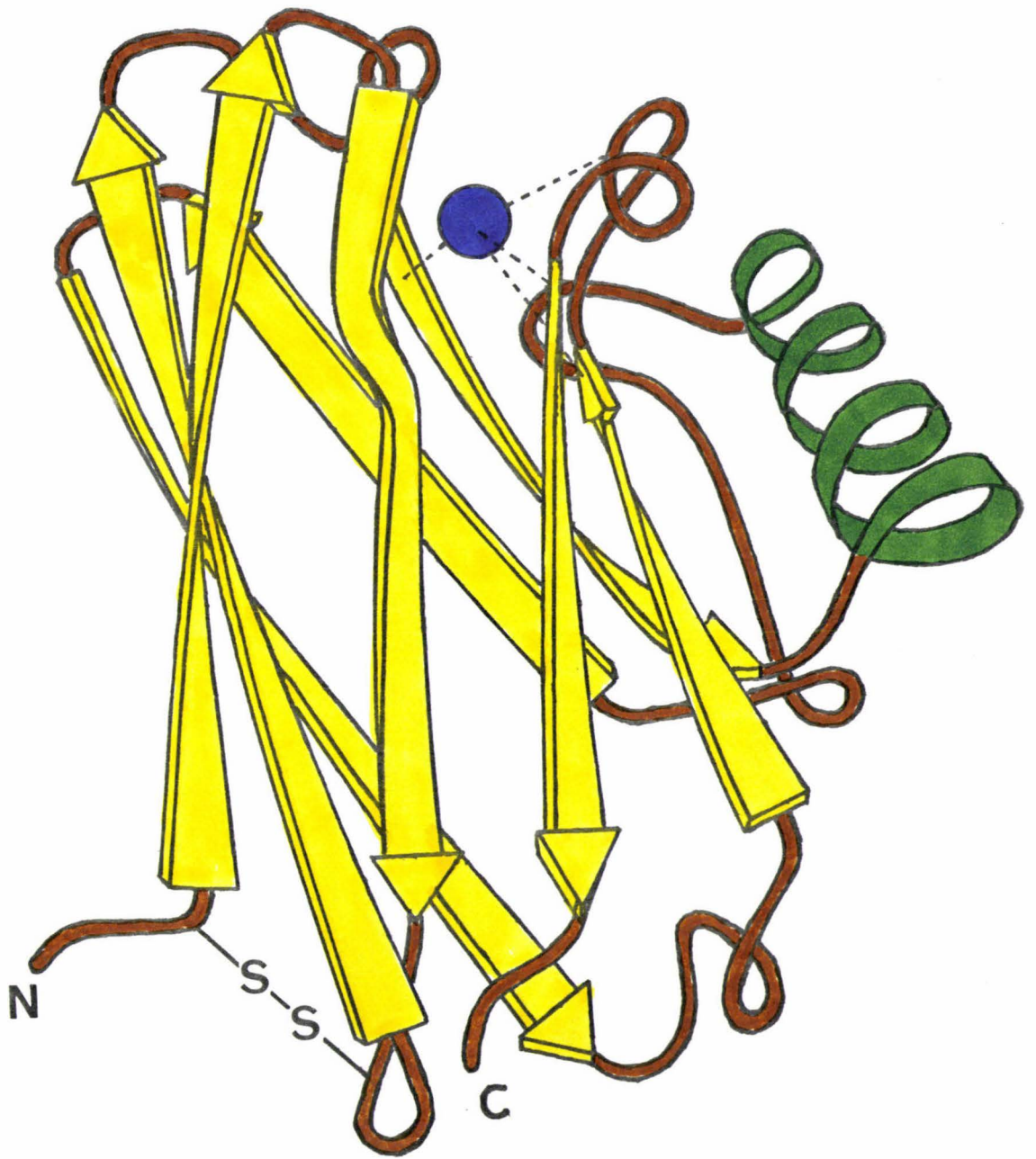
AZURIN

ACKNOWLEGEMENT

I sincerely thank my supervisors, Dr E.N. Baker and Dr.
S.V. Rumball for their help and encouragement throught this work.

Thanks are also due to the following:

Dr R.P. Ambler, of the University of Edinburgh, for bacterial
    cultures.

Dr G.G. Pritchard, for help with the microbiological aspects of
    this work.

Dr G.G. Midwinter, and Mr J.W. McLean for help with the purification
    procedures.

Dr K.L. Brown, Chemistry Division, D.S.I.R., Petone, for much help
    with data collection on the four-circle X-ray diffractometer.

The Staff of the Chemistry Department, Auckland University, for
    collection of data on the CAD-4 diffractometer.

Dr B.F. Anderson, for help with computing, contouring and proof-
    reading.

Drs. E.W. Ainscough and A.M. Brodie for helpful discussions.

Mrs H. Baker for help with contouring.

Mrs J. Trow and my husband, Dr R. Norris, for preparation of
    diagrams.

Miss K. Stalker for typing.

Finally, I would like to express my appreciation to my husband
and family, particular my mother, for their constant help,
encouragement and understanding during what must have seemed like an
interminable period.

The award of a University Grants Committee Postgraduate Scholarship
is gratefully acknowledged.

ABSTRACT

Purifications and crystallizations of two electron transfer
proteins, azurin and cytochrome c', from Alcaligenes denitrificans
and Alcaligenes sp. NClB 11015 have been carried out. The azurin
crystals from Alcaligenes denitrificans were found suitable for
high resolution X-ray structure analysis. They are orthorhombic,
space group $C222_1$ (with marked tetragonal pseudosymmetry), cell
dimensions a = 75.0Å, b = 74.1Å, c = 99.5Å, with two molecules
per asymmetric unit. A 3Å resolution electron density map of azurin
was calculated. Four isomorphous heavy atom derivatives, prepared
with $KAu(CN)_2$, uranyl acetate, $Hg(NH_3)_2Cl_2$ and ($KAu(CN)_2$ + uranyl
acetate) (a double derivative) were used to calculate phases by the
method of isomorphous replacement, giving an overall figure of merit
of 0.614. The polypeptide chain could be followed unambiguously
in both protein molecules in the asymmetric unit, with the aromatic
sidechains, in particular, readily identifiable because of their
distinctive appearance.

Kendrew skeletal models were built for both molecules, the poly-
peptide chain(consisting of 129 amino acids) being found to be
folded into an eight-strand β-barrel, with an additional flap
containing a short helix. There is one disulphide bridge within the
barrel. The topology of the molecule was found to be the same as
that of plastocyanin, and a comparison of the three dimensional
structures of azurin and plastocyanin allowed the sequences to be
aligned on structural rather than purely statistical grounds. It
also established the probability that the two proteins have evolved
from a common ancestor.

The copper atom has a highly-distorted tetrahedral co-ordination
geometry, forming three shorter bonds (length approximately 2Å),
with a cysteine thiolate sulphur (Cys 112) and two histidine imidazole

nitrogens (His 46 and 117), as well as a longer bond (approximately

3$\overset{o}{A}$) with a methionine thioether sulphur (Met 121). A surprising

result was the closeness of a peptide carbonyl oxygen, that of

Gly 45, to the copper atom. At this stage of the structure analysis

it is not clear whether it should be regarded as a ligand, or not.

Reduction of the protein crystals with chromous ions was

attempted, and the results are discussed in terms of the possible

electron transfer mechanism of the protein.

The cytochrome c' crystals from both species of bacteria

are hexagonal, space group P $6_1$22 (or P $6_5$22), cell dimensions

a = b = 54.7$\overset{o}{A}$, c $\sim$ 185$\overset{o}{A}$ γ = 120°, with one subunit (molecular weight

14,000) in the asymmetric unit. No structural work has been carried

out on these.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This thesis describes an investigation into the structure of a copper-containing electron transfer protein, and is aimed particularly at elucidating the structural and functional role of the copper atom.

1.1 The General Co-ordination Chemistry of Copper

Copper has a wide and varied co-ordination chemistry. It has two stable oxidation states, Cu(I) and Cu(II), each with particular chemical and co-ordination properties.

Copper(I)

The cuprous ion has the outer electronic configuration $3d^{10}$, so that its compounds are diamagnetic, and, except where colour results from the anion or charge transfer bands, are colourless. Cu(I) salts such as Cu_2SO_4 are unstable in water, decomposing to give copper and the cupric salt, although the Cu(I) state can be stabilised by the formation of complexes with suitable ligands (as shown, for example, in the $Cu(CN)_4^{3-}$ ion). Under the Pearson classification of "hard" and "soft" acids and bases (1,2) the Cu(I) ion is an extremely soft acid, and as such preferentially binds to "soft" bases such as R_2S, RS^-, and CN^-. In metalloproteins, groups which might be expected to bind to Cu(I) are the side-chains of methionine and cysteine, through their thioether and thiolate sulphur atoms respectively (both soft bases). The imidazole side-chain of histidine is regarded as an intermediate base and hence may also bind to Cu(I) (3,4).

Tetrahedral, four-co-ordinate geometry dominates the co-ordination chemistry of copper(I), with two and three co-ordination much less common. The co-ordination number rarely exceeds 4, although five-co-ordination is known. For example, the work of Gagné (5) has

shown that ligands that can enforce an approximately square planar geometry on copper(I), such as
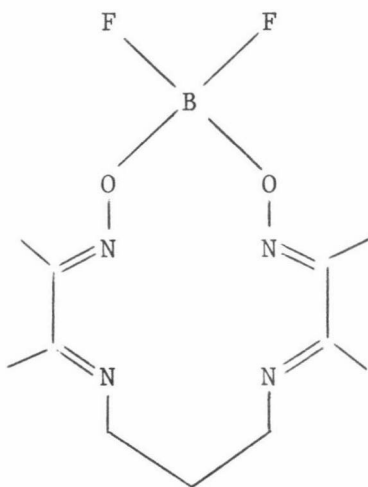
can encourage the Cu(I) to co-ordinate a fifth ligand such as CO rather readily. Model complexes with similar co-ordination geometry may then be pertinent to the further understanding of copper proteins such as haemocyanin, which is able to reversibly bind dioxygen.

Copper(II)

The dipositive (cupric) state is the most important one for copper. Most cuprous compounds are fairly readily oxidised to cupric compounds, many of which are soluble in water. In addition, a large number of complexes of Cu(II) are known. As it is near the middle of the Pearson scale, Cu(II) preferentially binds to bases such as $RCOO^-$, RNH_2 and H_2O, although it generally forms more stable complexes with nitrogen than with oxygen donors. It is also able to bind "softer" ligands such as RS^- because of its "intermediate" nature (1,2). Thus potential ligands for Cu(II) in proteins could include the hydroxyl groups of serine and threonine, the phenolic oxygen of tyrosine, the amino groups of lysine and arginine, the carboxyl groups of glutamic and aspartic acids, the imidazole group of histidine, the thiolate and thioether sulphur atoms of cysteine and methionine, and even main chain carbonyl oxygen and deprotonated amide groups.

The d^9 electronic configuration of Cu(II) leads to variability or plasticity in its co-ordination sphere (6). Thus the co-ordination geometry of Cu(II) complexes is to a considerable extent determined by the steric and electronic requirements of the ligands. Co-ordination numbers of 4,5 and 6 are the most common. For the first of these, the preferred geometry is square planar, but it is seldom completely regular. Distortions may arise from interligand repulsion, spin-orbit splitting, and/or Jahn-Teller interactions. For example, bis-chelate complexes frequently adopt square planar stereochemistry, but can be forced towards tetrahedral geometry by steric interaction between the ligands (distortions which can also modify the redox properties of the complex) (7). Although distortion toward a tetrahedron is quite frequent, no regular tetrahedral complexes of Cu(II) are known (7). In 5 co-ordinate complexes, a trigonal bipyramidal arrangement is rare, square pyramidal geometry being much more common, the axial fifth ligand usually a greater distance from the metal than the four in-plane ligands. There is, however, a strong tendency for the copper to increase its co-ordination number to 6, the most common, where the preferred geometry is distorted octahedral with the two axial ligands being more weakly bound due to the Jahn-Teller effect (8).

1.2 The Biochemistry of Copper

Although copper is found only in very low levels in most biological systems, it is essential to almost every form of life, from bacteria to man, in the role of a prosthetic group for a number of vital enzymes and proteins. These appear to have two main functions, viz. electron transfer, and oxidations involving molecular oxygen. (For examples see Table 1.1). Many reviews have been published on copper proteins (see for example references 9,10,11,12,13).

Table 1.1.   Copper Proteins and their Functions.

Function	Examples of Protein
Electron Transfer (between proteins)	Azurin, Plastocyanin, Stellacyanin Rusticyanin
Oxidations utilizing O_2 to form:	
(i) H_2O	Cytochrome c oxidase Laccase, Ascorbate Oxidase, Ceruloplasmin
(ii) H_2O_2	Amine Oxidases
(iii) oxygenated substrates	Tyrosinase
Superoxide dismutation	Superoxide dismutase
Transport of oxygen	Haemocyanin,
Copper Transport	Ceruloplasmin

The ability of copper to function in proteins such as oxidases and electron transfer proteins depends on its favourable chemical properties, since even when it is strongly chelated it can vary its oxidation state and retain the ability to complex small molecules such as oxygen. However, special mechanisms are required for its utilisation in biological systems, for reasons of stability and solubility. These can be provided by the protein matrix, which is able to furnish a specific environment (for example a hydrophobic pocket or a particular set of ligands) for the metal, to stabilize it in a particular state. In addition, the properties of such a metal atom bound to a protein can be influenced by changes remote from it, brought about by changes in protein structure which then change the environment of the metal (14).

1.3 Types of Copper Centres found in Copper Proteins.

Copper atoms in proteins have customarily been divided into three classes. These are:

Type I Copper

These are characterised by a very intense absorption band at about

600 nm (ε_{max} = 3000-5000 $M^{-1}cm^{-1}$) so that proteins containing such a prosthetic group are termed "blue" copper proteins. They include both electron transfer proteins and oxidases, and may contain only a single copper atom (as in the azurins and plastocyanins), or more than one copper atom per molecule where other types of copper are in association with the blue copper. Two explanations of the intense 600 nm absorption band have been proposed. One treats the band as arising from one or more allowed d-d transitions in a non-centro-symmetric centre (15,16), and the other attributes the strong absorption to a charge transfer process, probably of a ligand to metal type. Spectroscopic studies of cobalt(II) derivaties of stellacyanin, plastocyanin and azurin have established that the charge transfer interpretation is preferable (17,18.19). As other chemical and spectroscopic studies indicated that the thiolate sulphur of cysteine was a ligand to the "blue" copper (20,21), the band was assigned to a $\sigma S^- \rightarrow Cu$ charge transfer. Suggestions that it may have components of a methionine $S \rightarrow Cu$ charge transfer transition have been made on the basis of model studies (22). Further optical spectral studies on single crystals of plastocyanin have suggested, however, that the reason for the intensity of the band is more probably extreme geometrical distortion of the site (23).

The electron spin resonance (E.S.R.) parameters for Type I Cu(II) ions are also very distinctive. Although the parameters $g_{\parallel}$ and $g_{\perp}$ have typical values of approximately 2.23 and 2.05, similar to those found in many small molecule copper complexes with planar or tetragonally distorted stereochemistries (24), the hyperfine splitting constant, $A_{\parallel}$, has typical values of approximately $63 \times 10^{-4} cm^{-1}$, about 1/3 the value of $A_{\parallel}$ values usually found in inorganic Cu(II) complexes (9,13). The only small-molecule complexes in which these unusual spectroscopic properties are approached are those in which the copper site is in a highly distorted tetrahedral or square pyramidal co-ordination geometry (25,26).

A third characteristic feature of Type I copper proteins is their relatively high redox potentials (0.3 to 0.8 V). Redox potentials for copper reduction couples range from -0.01 V for the $Cu(NH_3)_4^{2+}/Cu(NH_3)_2^{+}$ couple, to +1.12 V for the $Cu(CN)_4^{2-}/Cu(CN)_4^{3-}$ couple (7), although they are normally in the region of 0.16 V. The highest values are observed for the softest ligands, and for sites with low symmetry, for example copper co-ordinated to the highly hindered ligand bis-2,9-dimethyl-1,10-phenanthroline, for which complex the redox potential is 0.59 V (14). Thus the asymmetry of the copper binding site, and the presence of "soft" ligands such as sulphur which tend to stabilize Cu(I), are proposed as the major reasons for these high potentials. Table 1.2 shows some of these properties for various blue copper proteins (13,27,28).

Type II Copper

This form of copper is present in all blue multi-copper oxidases, where it is essential to the protein function. Type II copper atoms are often referred to as "non blue" having spectroscopic and E.S.R. parameters similar to those found in low molecular weight Cu(II) complexes (13). Apart from their lack of the unusual spectroscopic properties characteristic of type I copper, the Type II copper atoms do have chemical properties which distinguish them from other forms of copper. For example they have an unusual anion affinity, which may enable them to stabilize intermediates such as peroxides that may be formed during the reduction of oxygen. Type II copper atoms are thought to be bound solely to nitrogen ligands, as in superoxide dismutase, where an X-ray structure determination has shown the copper to be bound to 4 histidine residues (29). Type II sites show a considerable variability in redox behaviour, and their function in proteins is not yet fully understood (7).

Type III Copper

This form of copper is found in all multi-copper oxidases where it

Table 1.2   The Properties of Some Blue Copper Proteins

Protein	Source	Mol.Wt.	$E_O(pH)$ (V)	Blue Band (nm)	ϵ $M^{-1}cm^{-1}$	ESR parameters $g_{\parallel}$ $g_{\perp}$ $A_{\parallel}(cm^{-1})$
Azurin	Bacteria	13,900	0.330(6.4)	625	5700	2.260,2.052,0.0064
Plastocyanin	Plants Green Algae	10,700	0.350	598	4500	2.226,2.053,0.0063
Stellacyanin	Lacquer Tree	20.000	0.184(7.1)	617	3550	2.287,2.077,0.0035
Umecyanin	Horse radish roots	14,600	0.283(7.0)	606		2.317,2.050,0.0035
Rusticyanin	Thio. ferro-oxidans	16,500	0.680(2.0)	597,450	1950,1060	2.229,2.064,0.0045
Laccase	Fungi	62,000	0.767	610	4600	
Ascorbate Oxidase	Squash Cucumber	140,000		610	9400	
Ceruloplasmin	Blood Serum (Human)	134,000	0.490 0.580	610	10,000	

is essential to the reduction of oxygen. Type III copper atoms are characterized by an absorption band at 330 nm, and in particular by their lack of any E.S.R. spectrum, remaining non-paramagnetic over a wide range of temperatures (13). They have therefore often been referred to as E.S.R. non-detectable. Type III copper atoms appear to always be found in pairs, probably forming a centre in which the two copper atoms are in close proximity and are strongly anti-ferro-magnetically coupled. They can thus act as two-electron donor or acceptor units (10).

The copper in tyrosinase, which is considered to be type III, is E.S.R. non-detectable in the resting, reduced, or oxygenated enzyme. However, upon addition of mercaptoethanol, a paramagnetic species, probably a mixed valence Cu(I)-Cu(II) pair, is formed, suggesting the involvement of sulphur ligands (30). Table 1.3 summarises the number and nature of the copper atoms found in each molecule of some copper proteins.

Table 1.3   The Nature and Number of Copper Atoms in Copper Proteins

Protein	Copper Content (Atoms/molecule)			
	Total	Type I	Type II	Type III
Azurin, Plastocyanin, Stellacyanin	1	1		
Laccase	4	1	1	2
Ascorbate Oxidase	8-10	3	1	4
Cytochrome c oxidase	2	type uncertain		
Tyrosinase	2			2
Haemocyanin[a]	2			2
Ceruloplasmin	6-8	2	2	2-4
Superoxide Dismutase	1		1	

(a) per subunit

1.4 The "Blue" Copper Site

A great deal of the early speculation on the nature of the "blue" copper binding sites was derived from spectroscopic studies. Thus the oxidised form (Cu(II)) of Type I sites (especially those of azurin, plastocyanin and stellacyanin) has been extensively studied using chemical (21,31), amino acid sequence (31,32,33), resonance Raman (34,35), E.S.R. (36,11), nuclear magnetic resonance (NMR) (37, 38, 39), X-ray photoelectron spectroscopy (XPS) (20,40), electronic spectral, circular dichroism, magnetic circular dichroism (19), infrared (41) and metal (cobalt, nickel and manganese) substitution (17,18,42,43) methods. From such studies three basic structural models were proposed, these being shown in Figure 1.1.

(a) CuN_3S (b) CuN_4S (c) CuN_2S_2

Figure 1.1   Proposed Models for Copper Co-ordination in Blue Copper Proteins

All of these models emphasise the importance of N from the imidazole group of the histidine side-chain, and S from the side-chain of cysteine as possible ligands in the binding of copper. Some studies have implicated a peptide oxygen or nitrogen as being involved in the co-ordination of copper (34,41,44). More recently, results of resonance Raman studies suggested methionine might provide a thioether ligand for the copper (45).

The Raman spectra of some of the blue proteins have been interpreted in terms of 5-co-ordinate geometry ((b) in Figure 1.1) around the copper atom, because of the multiplicity of the bands (35). Thus a trigonal bipyramidal geometry is proposed with the two axial ligands being more weakly bound than the three equatorial ligands, which include the thiolate sulphur. Such molecular symmetry (C_{2v}) may be lowered by non-equivalent axial ligands or distortions to the bond angles, so that it becomes more compatible with the observed Raman intensity patterns of these proteins. Variability among the different blue copper proteins with respect to these "weakly" bound axial ligands, but not to the equatorial ligands, may then account for differences observed in their Raman intensities as well as their redox potentials (see Table 1.2) (35). Five co-ordinate geometry has also been suggested for the Cu(II) state of blue copper proteins from a comparative study of Cu(II) and Co(II)-substituted anion complexes of bovine carbonic anhydrase (46). E.S.R. data, however, appear to be in conflict with such a model (22). Compressed trigonal bipyramids which possess C_{2v} symmetry, such as $[Cu(bpy)_2I]^+$, $[Cu(bpy)_2(tu)]^{2+}$ (where bpy is 2,2'-bipyridyl, and tu is thiourea) are well-characterized and are particularly relevant because they are analogous to the model, CuN_4S, proposed for the blue proteins, with the S in the trigonal plane. In contrast to the E.S.R. spectra of the blue copper proteins, which show $g_\parallel > g_\perp$, such complexes show $g_\perp > g_\parallel$. An elongated trigonal-bipyramidal structure, however, would show $g_\parallel > g_\perp$ (as would elongated octahedral, tetrahedral and planar structures when the singly occupied orbital is $d_{x^2-y^2}$, or d_{xy}, d_{yz}, d_{xz}). A high degree of anisotropy would, however, be expected in the copper hyperfine coupling constants $A_\parallel$ and $A_\perp$, whereas in blue copper proteins they are more isotropic (34). Such small $A_\parallel$ values relate to distorted structures (47), and as intermediate 5- co-ordinate geometries have received little attention as model complexes for blue copper proteins (48), it is felt they cannot be completely discounted.

Generally, spectra have been consistent with a distorted or flattened tetrahedral structure that involves at least one cysteine or thiolate ligand (19,49) ((a) or (c) of Figure 1.1). Such a model is compatible with both E.S.R. and other spectra of the blue copper proteins (22,34). X-ray absorption fine structure (EXAFS) methods have predicted unusually short Cu-S bond lengths of 2.10 ± 0.02Å, as well as shorter than normal Cu-N distances in the first co-ordination sphere of the copper (50). Such a Cu-S bond length approaches that predicted for a hypothetical tetrahedral Cu(II) bis(thiosemicarbazonato) complex, 2.16Å, providing further evidence for an approximately tetrahedral co-ordination geometry for Cu in blue copper proteins (50).

1.5 Azurin

Azurins are "blue" proteins that contain a single Type I copper atom, and have a molecular weight of approximately 14,000. They are found mainly in _Pseudomonas_ (Ps.) and _Alcaligenes_ (Alc.) bacteria, in addition to two species of _Bordetella_, and a _Paracoccus_ species, both of which are closely related to _Pseudomonas_ and _Alcaligenes_ (51,52,53). Although they are known to have an electron transfer function, it is not entirely clear what that function is. It is generally believed that azurin receives an electron from a small c-type cytochrome, cytochrome c-551 (a small acidic cytochrome of 82 residues which is generally found in bacterial species possessing azurins (54)), and passes it on to a cytochrome oxidase (55). Thus the sequence for electron transfer may be

$$\text{cytochrome c-551} \longrightarrow \text{azurin} \longrightarrow \text{cytochrome oxidase} \longrightarrow O_2 \quad .$$

Such a role for azurin would be analogous to that for plastocyanin , which receives an electron from the membrane-bound cytochrome f (a small c-type cytochrome), and passes it on to P_{700}, the double chlorophyll pigment of photosystem I (56). Although azurin and cytochrome c-551 have been shown to interact in vitro, with rates that

are among the fastest known for protein-protein electron exchange
(55,57), the biological significance of this has been queried on the
grounds that azurin interacts with other acidic cytochromes, for
example cytochrome c-553 from the red algae Plocamium coccineum, and
cytochrome f from parsley, at least as fast as with cytochrome c-551
(58).Azurin and cytochrome c-551 have also been shown to be able to
donate electrons to a nitrate reductase, found in Pseudomonads, which
can accomplish both a four-electron reduction of $O_2 \rightarrow H_2O$, and the
single-electron reduction of nitrite $\rightarrow$ NO, the latter process being
considered as the physiological function (59).

Comparison of the amino acid sequences of 9 different azurins
show that they all have either 128 or 129 amino acids, and that out
of these, 47 are invariant and 32 conservatively substituted (32).
As they all have similar spectroscopic and redox properties, and the
same metabolic function, they clearly belong to one homologous family
of protein molecules.

Chemical and Spectroscopic Studies on Azurin

Azurins have the characteristic optical and E.S.R. spectra shown
by other blue copper proteins. As is shown in Table 1.4, however,
the azurins from different bacterial species show minor variations in
both spectral parameters and redox potentials, which may reflect minor
variations in the ligand co-ordination geometry of the"blue"copper site
(13,60).

A variety of spectroscopic techniques have been invoked to probe
the nature of the copper site in"blue"proteins as a whole, as is
discussed in section 1.4. Many of them have focussed on azurin, and
together with other probes of the general structure of azurin are
summarised in Table 1.5.

Table 1.4   Properties of some Azurins from Different Bacterial Species

Source	Molecular Weight	ε at 625 nm($M^{-1}cm^{-1}$)	E_o (pH) V
Ps. aeruginosa	13,900	5700	0.330 (6.4)
			0.330 (7.0)
Ps. fluorescens	13,944	3500	
Bordetella pertussis	14,600	3500	0.395
Paracoccus denitrificans	13,790		0.230
Alc. sp.	14,000	5400	0.230 (6.8)
			0.260
Alc. faecalis	13,900	4600	0.266

At the time this work was begun, no X-ray crystallographic studies had been carried out on azurin, nor on any of the blue copper proteins. The objectives of this work were therefore to determine the three dimensional structure of azurin by X-ray crystallography, in order to

(i)   elucidate the nature of the copper site in a blue copper protein and its functional significance, and

(ii) try to identify possible macromolecular binding sites on the azurin molecule and obtain some insight into the electron transfer process.

1.6 The Special Requirements of Electron Transfer Proteins

As the primary function of azurin is to transfer electrons from one centre to another, it must have specifically evolved to carry out this task efficiently. The copper atom is essential to this function, and the protein may modify its properties by:

(i)   using specific ligand groups and/or co-ordination geometry to tune the E^o value of the metal to the required value, or

Table 1.5   Chemical and Spectroscopic Studies on Azurin

Residue	Interpretation with respect to the Structure	Method	Reference
Cys 112	Cu ligand. Hydrophobic site – delocalised Cu–S bond	XPS. Hg binding, metal replacement, ^{1}H NMR	41, 42 21 18, 61 62
Cys 3–Cys 26	Buried disulphide bridge	Not reduced by CO_2^-	64
His 35, His 83	CG of one < 7.5Å from Cu and inaccessible to solvent. Other is remote from Cu and accessible to solvent.	^{1}H NMR ^{13}C NMR	62, 64 44
His 46, His 117	Cu ligand; bound through ND1 of imidazole	^{13}C NMR ^{1}H NMR	44 37, 62
Trp 48	CG < 8.4Å, CD2 < 9.8Å from Cu, located in hydrocarbon–like environment	^{13}C NMR ^{1}H NMR phosphorescence ODMR	44 62 65 65

Table 1.5   (Continued)

Residue	Interpretation with respect to the Structure	Method	Reference
Tyr 72	Remote from Cu; $> 20\overset{o}{A}$ from Tyr 108; inaccessible to solvent, not adjacent to disulphide bridge; not H bonded.	^{13}C NMR fluorescence phosphorescence	44 66 65
Tyr 108	CG $< 7.5\overset{o}{A}$ from Cu; CD further away; hydrophobic environment; not adjacent to S-S bridge, not H bonded.	^{13}C NMR fluorescence phosphorescenc	44 66 65
Phe 110,111,114	Close to Cu	^{13}C NMR	44
Arg 79	Remote from Cu	^{13}C NMR	44
NH or CO (amide or carbonyl)	Cu ligand	IR ^{13}C NMR	41 44
amide NH	2 very close, 1 reasonably close to Cu	^{1}H NMR	62
Met 121	Cu ligand	^{1}H NMR Raman	62 45
Cu	Water is not present in the copper first co-ordination sphere Protein structure around Cu insensitive to pH over range 4.7-9.3.	^{1}H NMR fluorescence	67, 68 66 63

(ii) creating a specific environment for the metal, which may affect its properties (for example a hydrophobic environment).

The protein must be able to provide a co-ordination geometry which can accommodate the two different oxidation states of the metal ion (and their possibly different steric requirements) with a minimum of structural change (14). To do this, it may enforce a co-ordination geometry on the metal near to that of the transition state between the two oxidation states, not only decreasing the need for structural rearrangement during redox activity, but also reducing the activation energy required for electron transfer, thereby increasing the rate. Such a state has been termed the "entatic" state (69). As well as this, the protein must provide a specific site or sites for the ready donation and acceptance of an electron to and from its physiological partners, plus a pathway for that electron which is directional and can be controlled. Thus it may have unusual surface structural features, such as an acidic or basic "patch", to ensure specificity of inter-action.

At present, the current theories of electron transfer mechanisms are:- inner sphere transfer, via a ligand common to both electron centres (i.e. the formation of an intermediate complex in which the two metal centres are connected by a common bridging group); outer sphere, via ligands on each of the reactants (i.e. ligands are not directly shared between the two redox centres but rather must overlap or be connected in some way to allow the delocalization of an electron from one centre to another); and quantum mechanical tunneling (71) (the reactants have sufficiently close low-lying energy levels and are of a distance such that a finite probability of electron transfer occurs - this distance being recently computed as approximately 8 $\overset{o}{A}$ (70)). Outer sphere mechanisms seem more likely if the protein metal centre is buried and does not include a ligand which reaches the protein

surface. Ligands to the metal centre that are able to delocalize the electron that is to be transferred are advantageous, and the larger the ligand, the larger the possible electron transfer distance. For example, in cytochromes $[porphyrin]^{2-}$ is an ideal ligand for the spreading of charge away from the central metal, and orbital overlap with another protein metal ligand may allow electron transfer over distances up to 15Å. Other possible ligands able to spread charge over long distances are thiolate and disulphide groups, imidazole, and the aromatic side-chains of tyrosine and tryptophan. The advantages of such ligands for metal centres involved in electron transfer are:-

(a) Smaller local charge changes on the transfer of an electron.

(b) Smaller bond length changes on redox change.

(c) Less requirement for the close approach of the reacting metal atoms.

It is also possible to consider electron motion over relatively long distances via a rapid series of hops. Thus the transfer of an electron from metal atom to metal atom would involve the jumping of an electron from one organic group to another along a chain of groups such as quinones, or the aromatic side-chains of tyrosine, phenylalanine and tryptophan (14). The hop mechanism is like that suggested by Dickerson et al. for cytochrome c (72,73). Electrons may also be able to pass through hydrophobic regions of proteins, as is likely in cytochrome b_5, cytochrome c, and the $(Fe_4S_4)_2$ proteins. All these proteins possess hydrophobic channels where the electron may move effectively in a vacuum (14). Thus the observation of extended relays of parallel or near-parallel aromatic side-chains, hydrophobic channels, surface structural features, some of which may or may not be conserved in proteins with the same function, may provide a clue as to the mechanism of electron transfer.

1.7   The Relationship of Azurin to other Blue Copper Proteins

Amino acid sequence data is available for blue copper proteins other than azurin. The most extensively studied is plastocyanin, the small (molecular weight ∿ 10,700) electron transfer protein found in the photosynthetic electron transfer chain of plants, green algae, and some blue-green algae. Altogether the amino acid sequences for 12 species of plastocyanin (including 2 algal sequences) are known, in addition to the 9 species of azurin (32,74). Other blue copper proteins for which full or partial sequence data is available include stellacyanin (75), human ceruloplasmin (76,77), and bovine mitochondrial cytochrome c oxidase (78).

Several analytical and statistical comparisons of the amino acid sequences of azurins and plastocyanins have been carried out, aimed at estimating the amount of structural similarity between the two proteins (32,56,74). These suggested that the two families of proteins were probably monophyletic, and also suggested certain amino acids as ligands, these being subsequently confirmed by crystallographic studies (see section 1.8). Although the azurin and plastocyanin families are individually highly homologous (36% of residues in azurins are invariant, and 28% of residues in plastocyanin are invariant), even the best alignment of 9 azurin species with 8 plastocyanin species reveals very little homology; only 9 residues are common to both families, and a further 17 conservatively substituted (32,56,74). Furthermore, substantial deletions from the azurin sequence are required for it to match that of plastocyanin, as azurins have about 30 more residues than plastocyanins. Of course some of the differences may arise from functional differences,since although both are electron transfer proteins, they have different physiological partners with which to interact.

When,in the past, functionally-similar but distantly-related proteins have been compared, similarities have not always been apparent

at the level of primary structure, yet have been demonstrated by the comparison of three dimensional structures. For example, without the three dimensional structures, the problem of where to place the massive deletions in the cytochrome c sequence to align it with the much smaller cytochrome c-551 sequence was insurmountable, but with the structure in hand, it became trivial. It was the folding pattern revealed by the electron density map, following X-ray studies, that made it clear which part of the cytochrome c molecule had to be "removed" to make it fit the c-551 map (79). The structure of plastocyanin was reported after this work on azurin was begun, and it was then clear that a comparison of the crystal structures of azurin and plastocyanin should enable a proper alignment of the amino acid sequences to be made, taking into account their three dimensional structures. It should also clarify whether the two proteins have a common ancestor, or whether their similarities have arisen from convergence, where the same functional requirements have led to the evolution of similar active sites.

Comparison of the amino acid sequences of azurins and plastocyanins with other copper proteins has also yielded some interesting observations. Figure 1.2 shows the homology found between the sequences of segments from azurin, subunit II of cytochrome oxidase, the 19K dalton and 50K dalton fragments of ceruloplasmin (ceruloplasmin 1 and 2 respectively), stellacyanin and plastocyanin. The sequences in the upper half of the figure are found near the carboxy terminus, both in the multi-copper oxidases, and the single-copper blue proteins, and include 3 of the residues shown subsequently to be liganded to Cu in plastocyanin and azurin. A second region of similarity, including the fourth ligand for copper (His 37 in plastocyanin), is shown in the lower half of the figure. Analyses have suggested that, overall, subunit II of cytochrome oxidase is more like azurin, and the segments of ceruloplasmin more like plastocyanin (80,81). It is thus proposed that the binding

Azurin	CYS	SER-PHE-PRO-GLY	HIS	TRP- . -ALA-MET	MET	LYS-GLY-GLU
Cyt. oxidase	CYS	GLY- . -SER-ASN	HIS	. - . -SER-PHE	MET	PRO-ILE-VAL
Plastocyanin	CYS	ALA- . -PRO- .	HIS	GLN-GLY-ALA-GLY	MET	VAL-GLY-LYS
Ceruloplasmin (1)	CYS	HIS-VAL-THR-ASP	HIS	ILE-HIS-ALA-GLY	MET	GLU-THR-THR
Ceruloplasmin (2)	CYS	LEU-THR-THR-ASP	HIS	TYR-THR-GLY-GLY	MET	LYS-GLN-LYS
Stellacyanin	CYS	GLY-VAL-PRO-LYS	HIS	CYS-ASP-LEU-GLY-GLN-LYS-VAL-HIS		

Azurin	MET	ALA-LYS-ALA-VAL	MET	GLY	HIS	ASN	TRP	VAL	LEU
Cyt. oxidase	LEU	THR-VAL-LYS-THR	MET	GLY	HIS	GLN	TRP	TYR	TRP
Plastocyanin	PHE	. - . - . - . - . - .	PRO	.	HIS	ASN	VAL	VAL	PHE
Stellacyanin	TYR	. -ASP-ARG-ARG	PHE	.	HIS	ASN	VAL	ASP-LYS	

Figure 1-2   Copper-Binding Sequences

site for type I copper is similar in all the small blue electron
transfer proteins and large multi-copper oxidases, and that all
these enzymes may have evolved from the same ancestral gene (80,81).

1.8   Recent X-ray Studies on Blue Copper Proteins

While this work was in progress, successful X-ray crystallographic
analyses were reported on both plastocyanin from poplar leaves (56)
and azurin from Ps. aeruginosa (82). While these results reduced
some of the novelty of the present work, they also offered the
opportunity for valuable comparative studies.

1.8.1 Plastocyanin

The shape of the plastocyanin molecule resembles a slightly
flattened cylinder, the walls of which are formed by 8 strands of
polypeptide chain drawn roughly parallel to the cylinder axis, to form
a so-called β-barrel structure. The core of the molecule is hydrophobic,
with a high proportion of aromatic sidechains, 6 out of the 7
phenylalanine residues in the sequence being found there. Charged
sidechains are unevenly distributed on the surface of the molecule,
with a net negative charge (at physiological pH) being located more on
one side of the molecule than the other. The copper binding site is
embedded between the ends of strands 3,7 and 8 of the protein backbone.
In addition to the widely predicted ND1 nitrogen atoms of His 37 and
87, the sulphur atoms of Cys 84 and Met 92 turned out to be ligands
to the copper atom. Three of the four ligand-donating residues belong
to a short stretch in the amino acid sequence (-Cys-XX-His-XXX-Met-)
which forms a tight loop between strands 7 and 8. This loop, or
variations on it, can be seen repeated in the amino acid sequences
of other copper proteins in Figure 1.2. The co-ordination geometry is
irregular with bond angles deviating by as much as 23 degrees from
tetrahedral values. Accessibility of the solvent to the metal is
limited to one direction only, and blocked by the sidechain of His 87,
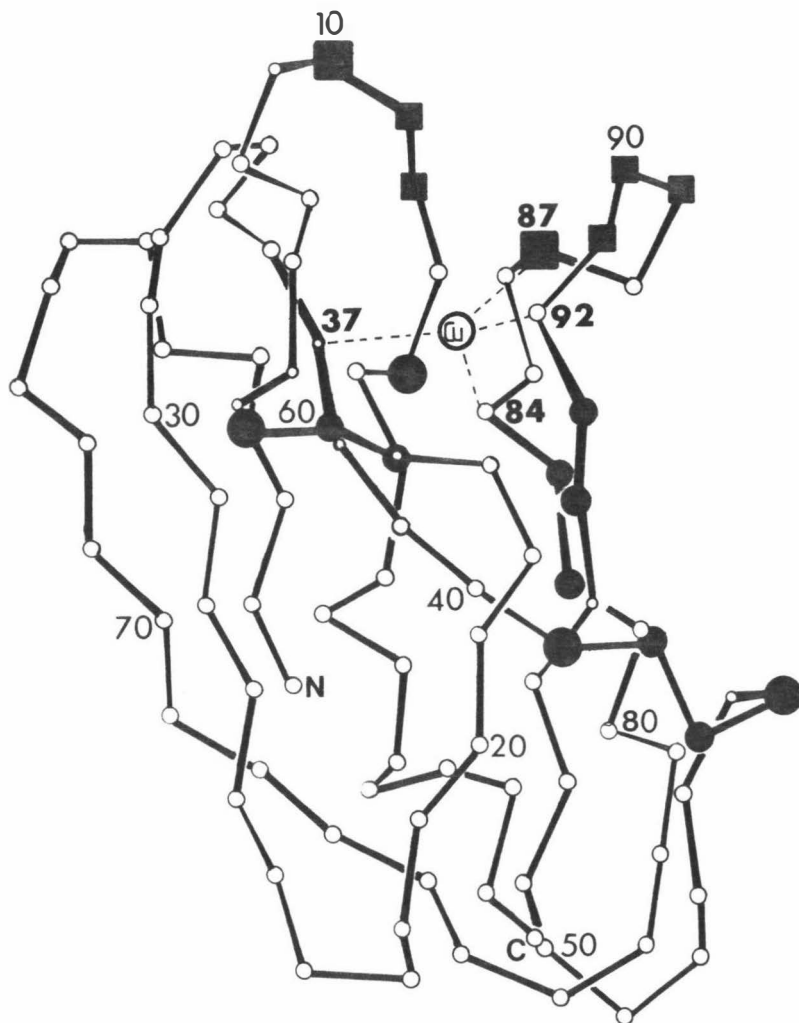
Figure 1.3   The Polypeptide Chain Folding of Poplar Plastocyanin

from X-ray Diffraction Studies at 2.7Å Resolution.

Residues marked ■ represent a hydrophobic patch found on the surface of the molecule near the copper site, and those marked ● an aromatic channel that runs from the base of the copper site down the front of the molecule.

one of the copper ligands, so that the metal-to-surface distance is
approximately 6Å. This geometry represents a compromise between the
requirements of copper in its two oxidation states, which may mean that
the reduction of Cu(II) can proceed without any changes in co-ordination.
Figure 1.3 shows the polypeptide chain folding of plastocyanin.

1.8.2 The Structure of Azurin from Ps. aeruginosa

More recently, the structure of azurin from _Ps. aeruginosa_ has
been reported at a resolution of 3Å (82). The shape of the molecule
was found to be that of a flattened pear of approximate dimensions
25 x 30 x 45 Å, with the copper atom at the small end of the pear. In
both shape and dimensions, it closely resembles poplar plastocyanin.
The single polypeptide chain folds to form an eight-stranded, somewhat-
distorted β-barrel. The most striking departure from the plastocyanin
structure is the addition of a flap between residues 53 and 78, which
contains some helical turns and hangs outside the body of the rest of
the molecule. The copper atom is co-ordinated by four residues which
are the exact analogues of the ligands in plastocyanin, His 46, Cys 112,
His 117, and Met 121. It lies in a hydrophobic pocket near an extensive
hydrophobic surface which clearly is analogous to the hydrophobic
patch in plastocyanin.

The initial 3Å map was not, however, easily interpretable, and in
order to improve the map, the electron densities of the 4 independent
molecules in the crystallographic asymmetric unit were averaged.
Although this led to some improvement, the two N-terminal residues could
not be seen, and the presence of a disulphide bridge between residues
3 and 26 made the connectivity of the β strands difficult to establish
unambiguously. A discrepancy between the topologies of azurin and
plastocyanin thus resulted (i.e. in the folding patterns of the
polypeptide chains), that would be unusual if the two proteins had a
common ancestor. Phasing was extended to 2.7Å by a method based on

direct-space averaging of the electron density of the 4 molecules, which much improved the quality of the map (83). The density <u>could</u> now be interpreted in terms of the plastocyanin topology, although the ambiguity remained since either folding pattern could be fitted reasonably well (see Figure 1.4).
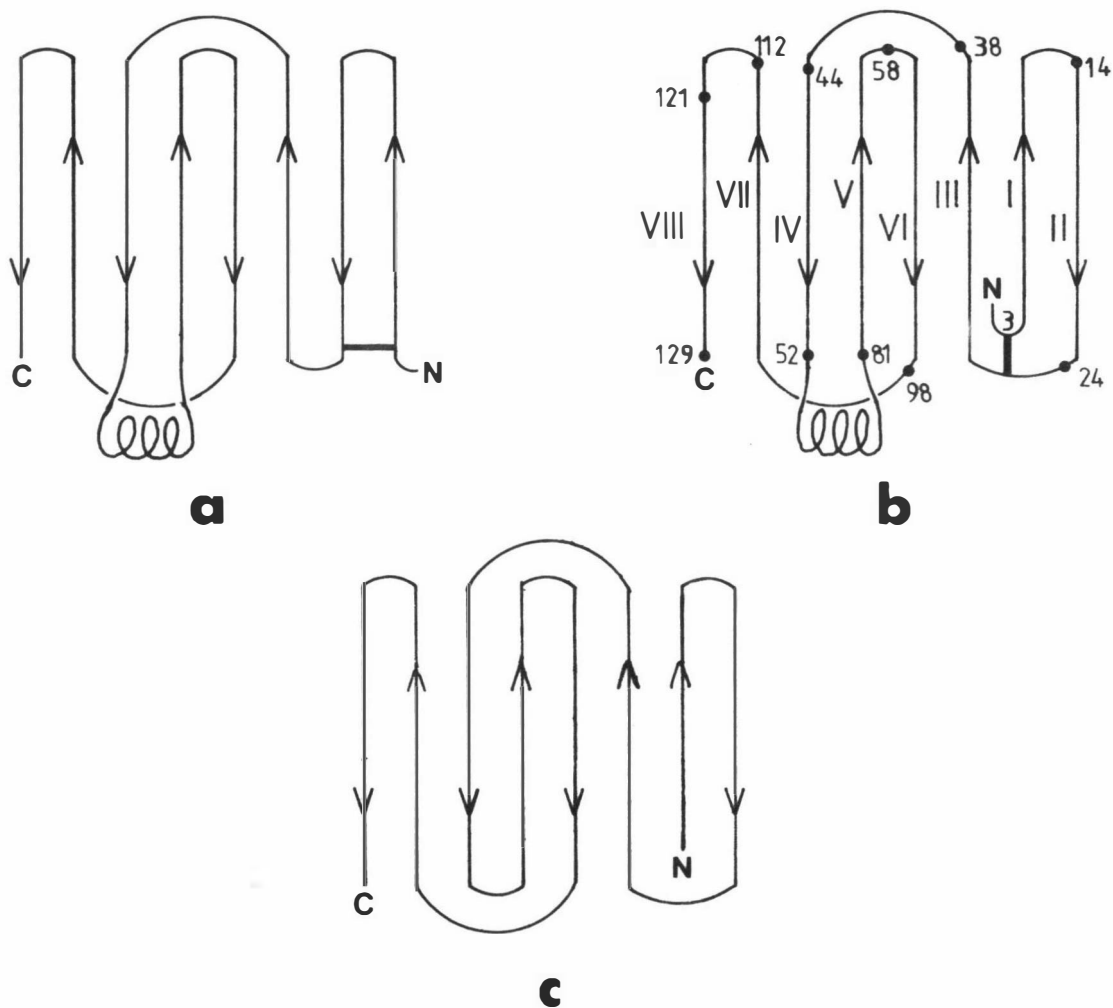
Figure 1.4   <u>A schematic representation of the topologies of azurin and plastocyanin showing the different folding patterns.</u>

(a) and (b) are the two alternative folding patterns for azurin

(c)   is the folding pattern for plastocyanin

——   represents a disulphide bridge between Cys 3 and Cys 26.

1.9   Comparative Studies on Azurin

One of the challenges posed by proteins is the difficulty in establishing the significance of different parts of such complex molecules. Some residues may be able to be changed without affecting structure or function; others may not. Some geometrical features may also be essential, either for activity or for maintenance of the three dimensional structure, or both. This problem is magnified for proteins such as hormones or electron transfer proteins which interact with macromolecular species rather than small molecules, and for which there may be no obvious active site. Binding surfaces may be extensive, and difficult to recognise.

Determination of the structure of azurin from another species should resolve the ambiguity in the structure of Ps. aeruginosa azurin, and establish whether or not azurins and plastocyanins have the same folding pattern. Comparison with the Ps. aeruginosa azurin should also help to establish which structural and geometrical features are important for the biological activity of azurin. Although 35% of amino acid residues in Alc. denitrificans azurin, for example, differ from those in Ps. aeruginosa azurin, the two proteins have very similar activities, and presumably similar binding sites; comparison may help to identify the latter. On the other hand, there are differences in E^o values between different azurins, and recent work has suggested kinetic differences (13,84,85). Reports have been published which show variations in the kinetics of low pH decolourisation of azurins from Ps. aeruginosa and Alc. faecalis (84), as well as in the electron transfer between Alc. and Ps. azurins with Ps. cytochrome c-551 (85). Such differences may be the direct result of small changes to the co-ordination geometry of the metal, and/or differences in the molecular envelope of the protein, or other localised parts of the structure. Again, a close comparison of the two may lead to a better understanding of these aspects of their functions.

Comparison of the structure of azurin with that of poplar plastocyanin should allow a proper alignment of the two sequences, based not on just statistical considerations, but on the three dimensional structures. This may throw further light on possible evolutionary relationships (including those with other copper proteins). It may also be possible to relate the differences in structure to the differences in activity, since azurin and plastocyanin have different electron transfer partners. It is hoped that the structure of azurin can ultimately be refined at high resolution, thus permitting a very detailed comparison of the copper site with that in plastocyanin, which has already been refined.

CHAPTER 2

.

EXPERIMENTAL

2.1 The Growth of the Bacteria

Alcaligenes denitrificans (Alc. denitrificans) and Alcaligenes

sp. (sometimes referred to as Pseudomonas denitrificans (86)) are

both members of a large and varied group of nitrate reducing bacteria.

While Alcaligenes is a strictly aerobic genera, Pseudomonads appear to

be able to grow under both aerobic and anaerobic conditions when the

growth medium is nitrate enriched (87).   Azurin was extracted from

both these bacteria, the culture of Alc. denitrificans, strain N.C.T.C.

8582 being obtained from the National Collection of Type Cultures,

Colindale, London, NW9 5HT, U.K., and the Alcaligenes sp. (Alc. sp.)

strain from Dr R.P. Ambler, Department of Molecular Biology, University

of Edinburgh, EH9   3JR ,   U.K.

For Alc. denitrificans, the growth medium used contained, per

litre : 10g Trypticase peptone; 7g yeast extract; 10g sodium glutamate;

1g KH_2PO_4; 0.002g $FeSO_4.7H_2O$; 0.01g $CaCl_2$; 0.008g $MgCl_2.6H_2O$; and

0.01g $CuSO_4.5H_2O$. This differed from the media previously reported for

the growth of this type of bacteria in the proportions of yeast extract,

trypticase peptone (casamino acids), sodium glutamate, and KH_2PO_4 used

(51,88). The ratio used was the optimum determined by comparing yields

of cell paste from 100 ml cultures grown in media containing varying

proportions of these four ingredients. The initial pH of the solution

was adjusted to 6.9 with 1M NaOH (approximately 10 ml/litre). A 1

litre shaken culture inoculum was grown at $37^{o}C$ for 24 hours and used

to seed the bulk medium (35 litres) in a New Brunswick 50 litre Ferma

Cell fermentor. The culture was stirred at a rate of 200 r.p.m. while

sterile air was being passed through it at a rate of 0.19 litres per

second, and was maintained at $34^{o}C$. Cell growth was monitored at

regular intervals by measuring the absorbance of the cell suspension

at 540 nm and subculturing the sample. The cells were harvested after 16 hours (when the culture was in the log phase) using Sorvall centrifuges, and could either be stored as a frozen wet paste (-20°C) or processed immediately. In the latter case yields of purified azurin were much greater.

For Alc. sp. the culture medium and growth conditions were the same as that used for denitrifying Pseudomonads (89,90). The organism was grown in 30 litre batches in large carboys under nearly anaerobic conditions at 28°C, harvested as above, and stored frozen.

2.2   The Purification of Azurin

Azurin from both species was purified using a method developed from those of Parr et al. (1976), Cusanovich et al. (1970), Ambler (1973) and Ambler and Brown (1967) (86,88,91,92).

(i)   Initial Extraction.

The wet cells, once harvested, were immediately suspended in 0.02M phosphate buffer pH 7.0 at 4°C using approximately 1 litre for every 300g of paste. They were then homogenized using an I.K.A. Ultra Turex T-45 with a 24 mm head at full speed for 6 minutes before being completely broken by a single passage through an Aminco French Pressure Cell (387 kg cm^{-2}). This was found to be the most efficient way to disrupt the cell membrane, as both Biuret and atomic absorption tests carried out on the supernatants of cell suspensions disrupted with ultrasonication, homogenization alone, or the above method showed the latter to contain the most protein as well as the highest concentration of copper. The resulting suspension, black and viscous, was collected in a cooled vessel. It was returned to room temperature before the addition of deoxyribonuclease (4 mg/litre of suspension) to break down the nucleic acids, and thus reduce the viscosity of the suspension. This occurred quite rapidly, after which the solution was cooled to 0°C and centrifuged at 34800g for 30 minutes. The very dark green

supernatant was decanted and saved, while the precipitated cell debris was extracted a second time as above. After centrifugation the supernatant was again decanted, and combined with the first extract, the precipitate being discarded.

(ii) <u>Ammonium Sulphate fractionation</u>.

The combined extracts were brought to 45% saturation with finely powdered ammonium sulphate over a period of 2 hours, centrifuged, and the pellet discarded. The supernatant was saturated with ammonium sulphate and left to stand overnight with stirring. Before centrifugation, the solution was made 10 µM with $K_3Fe(CN)_6$ which changed the reddish brown colour to blue green. After centrifugation, the resulting pellet was dissolved in a minimum of distilled water, and dialysed against distilled water made 10 µM with $K_3Fe(CN)_6$ until the conductivity of the protein solution was comparable to that of water. The supernatant from the precipitation, blue-green in colour and still containing some azurin, was passed through a pad of carboxy-methyl-cellulose (Whatman CM-32) in a Buchner funnel. Azurin was precipitated on the surface of the cellulose, remaining as a band that could then be washed off with distilled water This eluate was also dialyzed against distilled water as above, the two fractions then being combined. Stepwise fractionation of the 45% saturated supernatant was tried out in earlier preparations where finely powdered ammonium sulphate was added slowly with stirring until the solution became cloudy. It was then centrifuged, and both the resulting pellet and supernatant were retained, the process being repeated till 100% saturation was reached. However, when the pellets were dissolved in distilled water and oxidised with $K_3Fe(CN)_6$, all retained some blue colour, indicating the presence of azurin, in all fractions above 45% saturation.

(iii) Acidification of the Extract.

Fifty percent (v/v) acetic acid was added to the desalted, cooled (4°C) solution with thorough mixing until the pH reached 4.1. A white precipitate had begun to form at approximately pH 5.5, and this was removed by centrifugation at 14,600g for 30 minutes. The conductivity of the clear supernatant was then measured and compared to that of 0.05M ammonium acetate buffer,pH 4.1. The solution was diluted with distilled water until the conductivities were the same, and if any precipitate formed it was removed by centrifugation.

(iv) Chromatography on Carboxymethyl-cellulose.

The pale green solution was loaded on to a CM-32 column (5 cm diameter x 20 cm) which had been pre-equilibrated with 0.05M ammonium acetate buffer pH 4.1. Three coloured bands formed at the top of the column as the solution was passed through. These were red, blue and then a second,very narrow, lower red band   (Fig. 2-1). When loading had finished, the column was washed with the equilibrating buffer until no protein could be detected in the eluate, as judged by the absorbance at 280 nm. It was found that the final purification steps were much more effective if care was taken at this step, the quantity of buffer required for complete washing of the bands being in the vicinity of 4 litres. The choice of column is also important, as overloading of the bed will result if the pad of cellulose is too small,because of the high concentration of proteins in the cell extract, and the respiratory proteins will not adsorb. The lower red band was usually more obvious for the Alc. sp. extract than for the Alc. denitrificans where in some preparations it was hardly visible.

The column was then washed with 0.05M ammonium acetate buffer, pH 4.5, the eluate being constantly monitored for protein as above. For Alc. denitrificans, the lower red and blue bands began to diffuse very slowly through the column although they still adsorbed to the

Figure 2·1 Carboxymethyl-cellulose chromatography
during the purification of Azurin.
Three bands are evident :
1. an upper red band, cytochrome c',
2. a blue band, Azurin,
3. a lower faint red band, a c-type
cytochrome.

cellulose, and colourless protein was eluted. After the passage of approximately 1 litre of buffer, no further protein could be detected in the eluate. The pH of the buffer was then raised to 4.7 and the procedure repeated. This time, more colourless protein was washed off the column before the lower red band was eluted. Electronic spectra of this red eluate, recorded on a Shimadzu M.P.S.-5000 spectrophotometer using 10 mm quartz cells, showed it to contain c-type cytochromes which were identified by their characteristic spectra (54). The blue band was immediately behind the lower red one, but was not eluted until the pH of the buffer was raised to 5.1, when it was eluted as a broad dilute band (∿ 800 ml). The remaining red band had become more diffuse by this time, and it was removed from the column by raising the pH of the buffer to 7.0. Electronic spectra of this eluate, recorded as above, showed it to contain a mixture of c-type cytochromes, including c', which is distinguishable by its split Soret peak (91).

For <u>Alc. sp.</u> the procedures were much the same as above, except that once the column was loaded, it was washed with buffer at pH 4.3, since the first red band was eluted at pH 4.5 rather than 4.7 as for <u>Alc. denitrificans.</u> The blue band was eluted as a broad band at pH 4.7 and the remaining red band at pH 7.0 as above. Electronic spectra of the two red eluates showed the solution at pH 4.5 to contain c-type cytochromes, and the solution at pH 7.0 to contain mainly cytochrome c' as above. The cytochrome c solution from both bacterial species was discarded, while that containing cytochrome c' was kept for further purification (see Appendix I).

(v) Chromatography on Diethylaminoethyl-cellulose.

The azurin solutions were concentrated by ultrafiltration using a Dia-flow ultrafiltration cell and PM-10 membrane, before being re-equilibrated in 10 mM tris-HCl buffer, pH 8.7, by gel filtration on medium grade Sephadex G-25. Chromatography on diethylaminoethyl-

cellulose (Whatman DEAE-52) was then carried out twice to remove any nucleic acids and remaining respiratory enzyme components (66,89,92). The blue protein (azurin) did not bind to the column, being merely retarded by it, although some cytochromes remained as a red band at the top of the column. This step proved to be quite important as it removed contaminants that were difficult to remove by ammonium su⸍ hate precipitation, chromatography on CMC-32, and standard gel filtration techniques.

(vi) Precipitation with Ammonium Sulphate.

The blue solutions from step (v) were again concentrated by ultrafiltration, before finely powdered ammonium sulphate was added to them very slowly with stirring. When the solutions became cloudy they were left standing for 2 hours before the precipitates were removed by centrifugation. At 95% saturation most of the azurin precipitated, the resulting blue pellets being redissolved in a minimum of distilled water before being subjected to gel filtration.

(vii) Gel Filtration.

Gel filtration was carried out on Biogel P-10 (100-200 mesh) equilibrated in 0.1M phosphate buffer, pH 6.0. The columns were 100 cm x 2.6 cm in diameter (Pharmacia $K_{26/100}$) and a flow rate of 0.2 ml/ minute was used. Usually 4-8 ml aliquots of the protein solution were applied, the eluate being collected by a fraction collector and monitored by a u.v. absorption recorder (L.K.B. Uvicord-II measuring at 280 nm). The fractions associated with the blue peak were assessed for purity by acrylamide gel electrophoresis (93), those showing no impurities being pooled.

The homogeneity of the protein was tested during the purification by measuring the ratio of the absorbance of the oxidised protein at 625 nm to that at 280 nm. Chromatography on CM-cellulose, ammonium sulphate precipitation and gel filtration were repeated until the

spectral ratio E_{625}/E_{280} reached 0.27-0.30 for <u>Alc. denitrificans</u>

azurin, and 0.45-0.47 for <u>Alc. sp.</u> Reported values of this spectral

ratio for other azurins have ranged from 0.47-0.6 for <u>Ps. fluorescens</u>,

<u>Ps. aeruginosa</u> and <u>Alc. faecalis</u> (60,88,89,94) and 0.35 for <u>Alc. sp.</u>

(60). It is well known that the absorbance at 280 nm is mainly due

to the aromatic side chains in the protein. These are tryptophan

(which contributes the most), tyrosine, and phenylalanine (whose

contribution is negligible) (95). Comparison of the sequences of these

azurins shows that <u>Ps. fluorescens</u>, <u>Ps. aeruginosa</u> and <u>Alc. faecalis</u> all

have 1 tryptophan, and 2 tyrosines in their sequences, while <u>Alc. sp.</u>

has 1 tryptophan and 3 tyrosines, and <u>Alc. denitrificans</u> has 2

tryptophans and 4 tyrosines (32). On this basis the absorbance at

280 nm should be greater, and hence the ratio of E_{625}/E_{280} lower than

an equal concentration of azurin of the same purity from <u>Alc. sp.</u>,

<u>Alc. faecalis</u>, and the two <u>Ps.</u> species. This will be reflected in the

spectral ratios, which are found to vary between different azurins,

that for <u>Alc. denitrificans</u> being one of the lowest. Table 2.1 gives

the ratio of E_{625} to E_{280} at various stages of the purification of

azurin from <u>Alc. denitrificans</u>.

<u>Table 2.1</u> <u>Extinction Ratios of Azurin at Various Stages of</u>

<u>Purification.</u>

Step	$E_{625}^{ox}/E_{280}^{ox}$
(1) Eluate from the C.M.C. column	0.175
(2) Eluate from the D.E.A.E. column	0.213
(3) Eluate from gel filtration after ammonium sulphate precipitation	0.273

All steps were carried out at $4^{O}C$, and the concentration of the protein was calculated either by using an extinction coefficient of 3.5 $mM^{-1}.cm^{-1}$ (96) or by measuring the concentration of the Cu present in the sample by atomic absorption spectroscopy, assuming a ratio of 1 copper atom per molecule of protein.

The purified protein solution had a spectrum characteristic of azurin as shown in Fig. 2.2 (51,97), and was homogeneous by S.D.S. acrylamide gel electrophoresis. It was found that the protein could be freeze-dried, and that this process could actually be used as a purification step. A sample made up of the fractions of blue protein rejected after gel filtration procedures on the basis of acrylamide gel electrophoresis was concentrated by ultrafiltration before being freeze-dried. When the resulting blue powder was placed in distilled water, the azurin redissolved immediately to form a clear blue solution leaving an insoluble white precipitate in the bottom of the flask. The resulting protein solution had a spectral ratio of 0.27 and was successfully crystallized. The yield of azurin from Alc. denitrificans was approximately 8 mg per litre of culture.

2.3 Crystallization.

Crystallizations of azurin from Alc. sp. have previously been carried out by Suzuki and Iwasaki (97) and Strahs (98). In both cases the azurin was crystallized from ammonium sulphate solutions, the crystals of Suzuki and Iwasaki being blue needles that were too small for X-ray analysis, while those of Strahs were of sufficient size for preliminary X-ray examination. This showed them to belong to the tetragonal space group $P4_122$ with a = b = 53.2Å and c = 101Å, with one molecule per asymmetric unit. Azurin from Ps. fluorescens has also been crystallized, although again the crystals were small and unsuitable for X-ray analysis (92). Azurin from Ps. aeruginosa had previously been crystallized in this department (99), the crystals being orthorhombic, space group $P2_12_12_1$
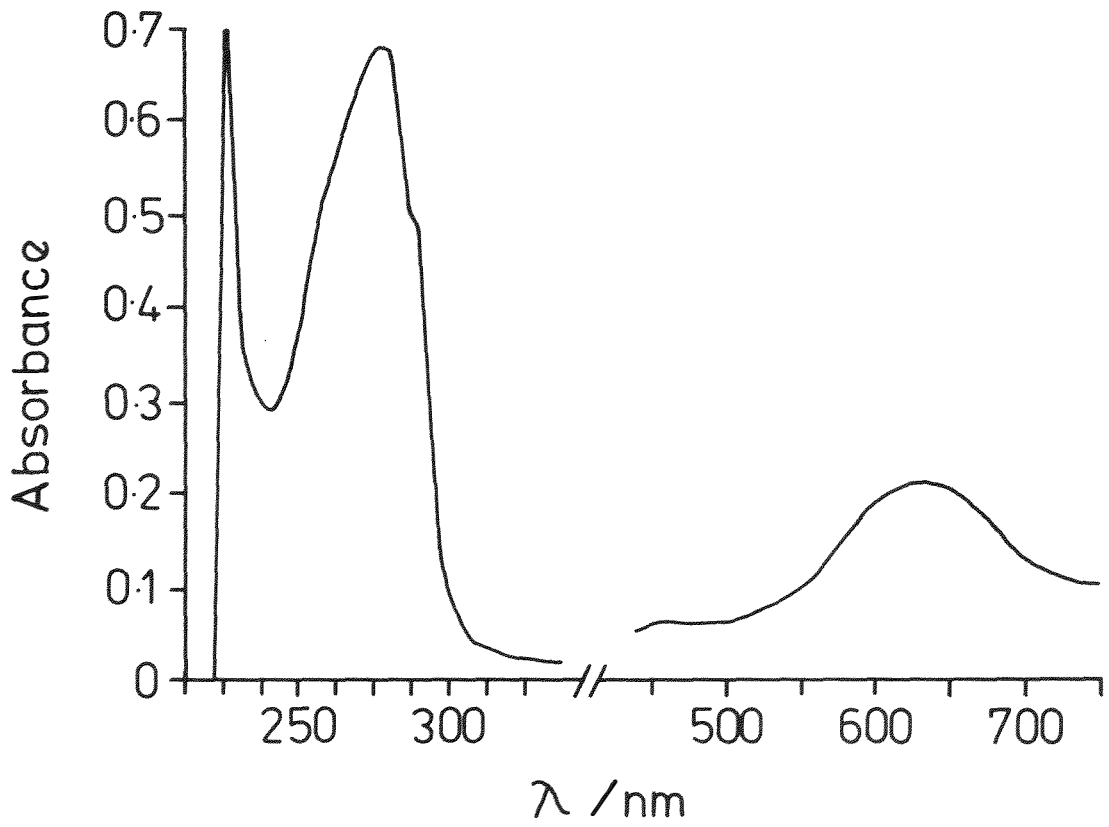
Figure 2·2   Absorption Spectra of Azurin from *Alcaligenes denitrificans*.

with a = 49.2$\overset{o}{A}$, b = 98.5$\overset{o}{A}$ and c = 109$\overset{o}{A}$, and 4 molecules per asymmetric unit. Attempts to prepare suitable heavy atom derivatives were, however, unsuccessful. While the present work was in progress, the structure of azurin from Ps. aeruginosa was reported at a resolution of 3.0$\overset{o}{A}$ (82). Here, the space group was again $P2_12_12_1$ with dimensions a = 58.85$\overset{o}{A}$, b = 78.98$\overset{o}{A}$, c = 108.47$\overset{o}{A}$ and 4 molecules per asymmetric unit.

2.3.1 Crystallization Techniques Used.

Before setting up the crystallizations, care was taken to filter all solutions and centrifuge them at high speed for at least 30 minutes to remove any dust particles or other debris. All glassware was acid-washed, well rinsed in de-ionized water and oven-dried immediately before use, while the plastic tubes sometimes used for crystallizations were "dusted" with a jet of compressed air.

(a) Free Interface Diffusion

When a protein is soluble in solution A and insoluble in solution B, and the solution containing the dissolved sample can be layered very carefully on top of the other, then transient conditions of supersaturation will be achieved in the region of the interface. As the two layers diffuse into one another toward equilibrium, nuclei are formed. If, at equilibrium, the total precipitant concentration is less than that required to produce immediate precipitation of the protein, but high enough to support the growth of crystals from the nuclei that have formed, large single crystals should result.
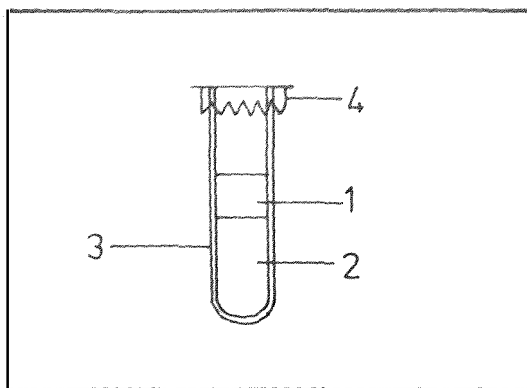
Figure 2.3   A Schematic Diagram of the Free Interface Diffusion

Technique.

1.   protein solution   3.   plastic tube (4 ml)

2.   precipitant   4.   parafilm seal

(b) Microdialysis

A protein solution may be brought slowly toward its precipitation
point by dialysis against a specific salt solution or organic solvent.
The advantage of this technique is that as the differential between
concentrations inside and outside the membrane decreases, the rate of
equilibration also decreases. As azurin could pass through 10 mm
dialysis tubing, special care had to be taken with this method. Small
samples of protein solution (10-50 $\mu\ell$) were injected very carefully,
so that no bubbles formed, into short glass capillaries or tubes
which had previously been siliconized (100). The thick walled glass
tubes, open at both ends, had a double piece of 2.5 mm dialysis
tubing, which had been prepared in the usual way (101), stretched over
one end and secured by a collar of plastic tubing cut to provide
support legs. The other end was covered by a piece of parafilm held
on with another collar. The entire assembly was then submerged in a
sealed tube containing the precipitating solution. The apparatus

could be modified to use capillaries with polyacrylamide gel plugs,
polymerized in situ, instead of dialysis membranes. The advantage
of this method is that the concentration of the acrylamide in the plug
can be used to control the rates of diffusion (102,103).
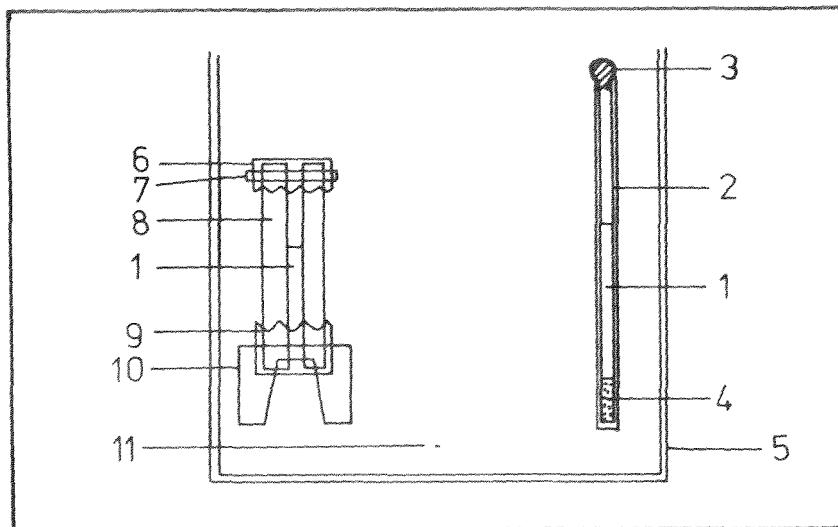
Figure 2.4   A Schematic Diagram of the Microdialysis Methods Used.

1.	Protein Solution	7.	Collar
2.	X-ray capillary tube (1.5 mm)	8.	Thick walled glass
3.	Wax seal		tubing
4.	Polyacrylamide gel plug	9.	Dialysis membrane
5.	Sealed container	10.	Support
6.	Parafilm seal	11.	Precipitant.

(c) Evaporation

This method is a most primitive approach, but is one which has
been used successfully for a few protein crystallizations (104). The
mother liquor was allowed to slowly evaporate at room temperature, so
that the precipitant and/or protein concentration attained super
saturation. The rate at which evaporation occurred was controlled to

some extent by covering the plastic tube containing the protein

solution with parafilm, and making one hole or a number of holes in

it to provide a passage for water loss.

(d) The Batch Method

In this method, small plastic or siliconized glass vials were

filled with 0.5 to 1.0 ml of the protein solution, usually at a

concentration of 10-20 mg ml^{-1}, and a salt concentration slightly less

than that at which azurin precipitated or was thought to precipitate.

The vials were then sealed with screw caps or rubber stoppers and set

aside. A range of precipitant concentrations were investigated,

accurate additions of precipitant solution being made with a Hamilton

micro-syringe.

(e) Vapour Diffusion

Evaporation of water from the protein solution is controlled in

this technique by equilibration with a more concentrated salt solution.

Thus a solution of the protein, containing a salt concentration

approximately 10% below the concentration needed for precipitation,

is equilibrated through the vapour phase with a large volume of a more

concentrated salt solution, where both solutions are in a sealed system.

Multiple depression spot plates afforded a convenient way of testing

this method on micro quantities of protein solutions under a variety

of conditions (such as pH). A diagram of such an apparatus is shown in

Figure 2.5. The technique could also be effected by placing small

quantities (100-500 $\mu\ell$) of mother liquor in plastic tubes which were

then sealed in small bottles containing the precipitant solution, as

shown in Figure 2.5 also.
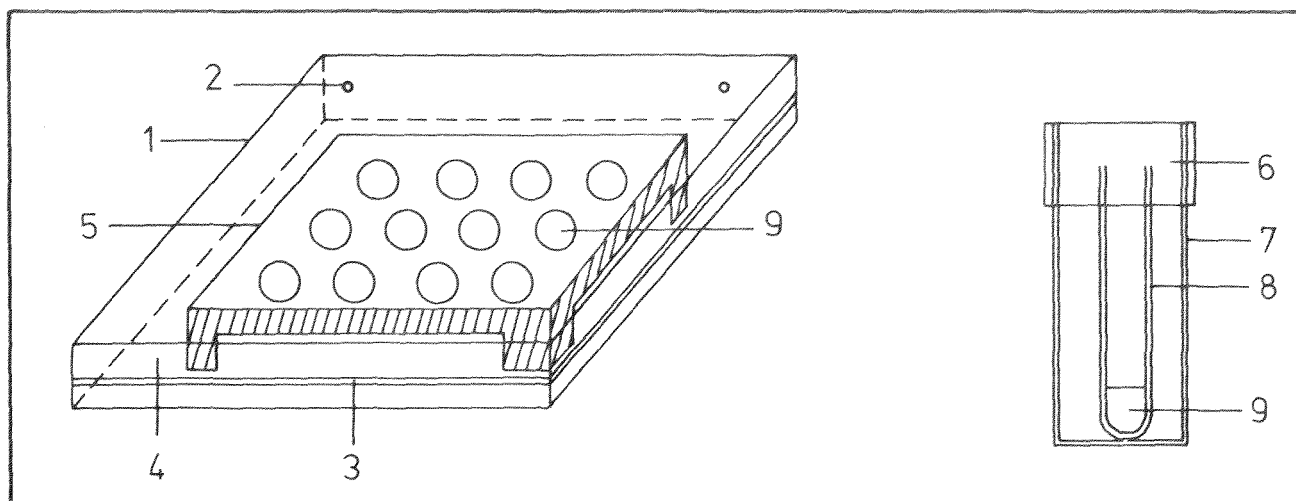
Figure 2.5   A Schematic Diagram of the Vapour Diffusion Arrangements

Used in the Crystallization of Azurin.

1.   plastic sandwich box

2.   holes sealed by sellotape to allow addition or

removal of the precipitant

3.   seal (tape)

4.   precipitant solution

5.   perspex depression plate on support legs

6.   screw cap

7.   glass vial

8.   plastic tube

9.   protein solution.

These methods have been reviewed in detail by McPherson (100)

2.3.2   The Conditions Used for the Crystallization of Azurin.

(a)   pH

With the exception of the concentration and nature of the

precipitating agent, the most important variable to be investigated

in the search for crystallization conditions is pH (100).   It has been

pointed out that the probability of obtaining an amorphous precipitate or micro crystals rather than large single crystals may depend on only a few tenths of a pH unit (102). The pH of solutions used in the crystallization of azurin from both organisms varied between 4.4 and 6.8. They were made up in either 0.1 M ammonium acetate buffers, pH 4.4, 5.0, 5.4 and 6.0, or 0.2 M and 0.1 M phosphate buffers, pH 6.0, 6.4, 7.0, 7.1, 7.2, 7.3, 7.5 and 7.8. When these buffers were saturated with ammonium sulphate the pH of the solutions made up with phosphate buffer were measured as 5.0, 5.4, 6.0, 6.1, 6.2, 6.3, 6.5 and 6.8, while that of the solutions made up with ammonium acetate buffer showed no change. (All measurements were made with standard electrodes).

(b) Temperature

Crystallizations of proteins have been reported over a wide range of temperatures, from $0^{\circ}C$ to $40^{\circ}C$. Many of the factors that govern the solubility of proteins have a marked temperature dependence, and at high ionic strength most proteins have a negative temperature coefficient of solubility, i.e. they are less soluble at higher temperatures than they are at lower ones (105). Thus crystallizations were set up at $4^{\circ}C$, $17^{\circ}C$, $37^{\circ}C$ and room temperature.

(c) Precipitants

Precipitants fall into two categories: salts such as ammonium sulphate, and organic solvents such as ethanol. Because ammonium sulphate had previously been used with some success in the crystallization of azurin (92,97,98), efforts were concentrated on optimizing conditions of crystallization using it, although some experiments were carried out using sodium citrate, sodium and potassium phosphates, and polyethylene glycol as precipitating agents. These are summarised in Table 2.2.

Table 2.2   A Summary of the Precipitants and their Concentrations used
in the Crystallization of Azurin from Alc. denitrificans
and Alc. sp.

Precipitant	Method	Conc. of precipitant in reservoir (%) (a)	Conc. of precipitant in protein solution	Conc. of precipitant overall (%)
ammonium sulphate	Vapour diffusion	62,63,64,65, 66,67,68,69, 70,71	50%	62,63,64,65,66, 67,68,69,70,71
ammonium sulphate	Batch		60,61,62,63,64, 65,66,67,68,69, 70,71,72 %	60,61,62,63,64, 65,66,67,68,69, 70,71,72
ammonium sulphate	Free interface diffusion	80 80	40% 50%	60,62,64,66 68,70,72,74
ammonium sulphate	micro-dialysis	100,90,80	60%	100,90,80
sodium citrate	micro-dialysis	100	-	100
sodium citrate	Batch (b)	-	-	-
phosphate buffer (mixed sodium & potassium salts)	Evapora-tion	-	2.4 M	-
polyethy-lene glycol [c]	Vapour diffusion	20	8%	20

(a) Concentrations of the precipitants are expressed as percentages of
saturation.

(b) finely powdered sodium citrate was added with stirring to the
protein solution until it became slightly turbid. Micro-droplets
of distilled water were then added to the solution till it became

clear again, and after centrifugation it was carefully decanted

into a screw-capped vial and left at room temperature.

(c)   the polyethylene glycol used was PEG-6000. Before use it was

dissolved in 15 volumes of acetone, then precipitated from

20 volumes of ether, and dried under vacuum (107).

(d)   Concentration of the Protein

The concentration of the protein molecules in the mother liquor

is often critical, and crystals have been reported grown at

concentrations from one to several hundred milligrams per ml, although

the most common range seems to be from 5 to 30 mg ml^{-1} (100). In

these experiments protein concentrations between 10 and 20 mg ml^{-1}

were used.

(e)   Additives

Additions of small quantities of organic solvents such as ethanol

(100), dioxane (108,109) and other salts such as LiCl (82), NaCl,

MgSO$_4$ and sodium citrate to the mother liquor were tried in order to

induce the growth of large single crystals. Although there appears

to be no obvious reason why such additives should enhance the

crystallizations of proteins, many proteins have been crystallized

with the presence of small amounts of such compounds in their mother

liquor (110,111).

2.3.3 The Crystallization of Azurin from Alc. Sp.

Azurin from this organism could be crystallized under a great

variety •f conditions, including nearly all combinations of the

conditions described in sections 2.3.1 and 2.3.2. It could not be

crystallized using polyethylene glycol as a precipitating agent, and

the use of both sodium citrate and mixed phosphate salts produced

micro crystals only. The best crystals were obtained from protein

solutions made up in phosphate buffers of pH 6.4-7.0, with a protein

concentration of 10 mg ml^{-1}, and ammonium sulphate as the precipitant.
Vapour diffusion was the most successful method of crystallization,
the crystals taking the form of either very fine needles or dark blue
rectangular plates, both forms often occurring together as shown in
Figure 2.6. The largest plates were grown in plastic tubes, and had
dimensions up to 0.5 x 0.5 x 0.1 mm. Although they appeared to be
single crystals, magnification showed them to be aggregates of very
thin plates, and all attempts to increase the size of the needle form,
or to obtain large single crystals of the plate form, failed.

2.3.4 The Crystallization of Azurin from Alc. denitrificans.

Initial screening using the crystallization techniques and conditions
discu sed in sections 2.3.1 and 2.3.2, at room temperature only, showed
small but nicely formed crystals could be obtained from protein solutions
made up in 0.2 M phosphate buffer, pH 6.0, and a protein concentration of
20 mg ml^{-1}, with ammonium sulphate as a precipitant. The best crystals
were obtained using vapour and free interface diffusion techniques,
the reservoir solutions being 85% saturated with ammonium sulphate
(Figure 2.7). The same crystal form could be obtained using mixed
sodium and potassium phosphates, or sodium citrate as the precipitating
agents, although the crystals were again small. No crystals were
obtained using polyethylene glycol as the precipitating agent, the
azurin forming an oily amorphous precipitate in its presence. In
an attempt to increase the size of the crystals, the temperature,
final concentration of precipitant, and the concentration of the protein
in the mother liquor were varied. The best crystals were grown in two
ways:

(a) by free interface diffusion, where 200 µℓ of protein solution
     (14 mg ml^{-1}), in 0.1 M phosphate buffer pH 6.0, made 40% saturated
     with ammonium sulphate, was layered on to 240 – 370 µℓ of 80%
     saturated ammonium sulphate made up with the same buffer and
     left at 37°C, or
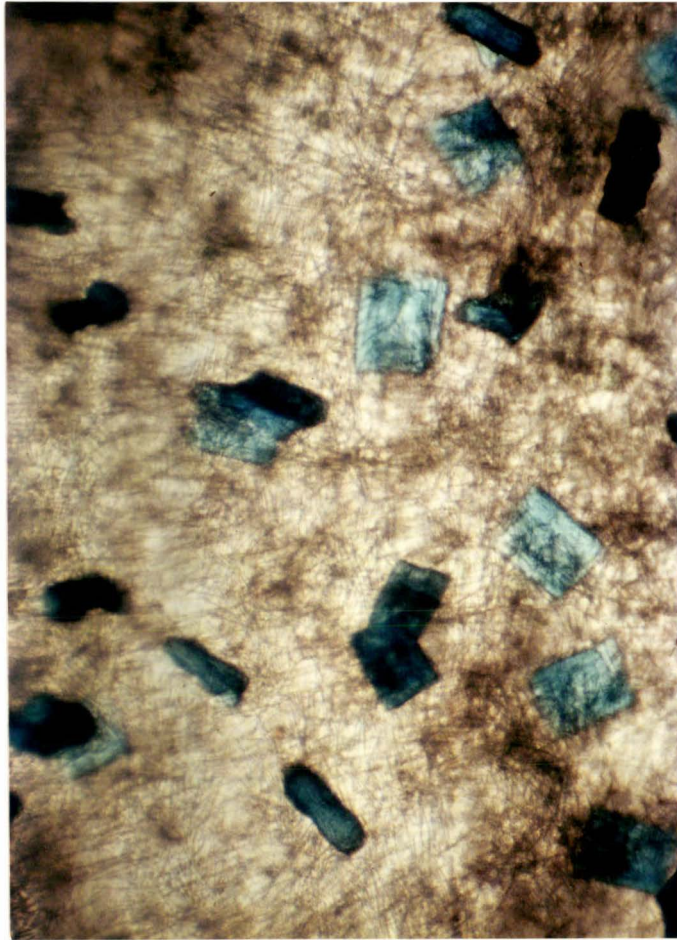
Figure 2·6 Crystals of Azurin from *Alcaligenes sp*.
Two crystal forms are visible (needles and plates). (63 ×).
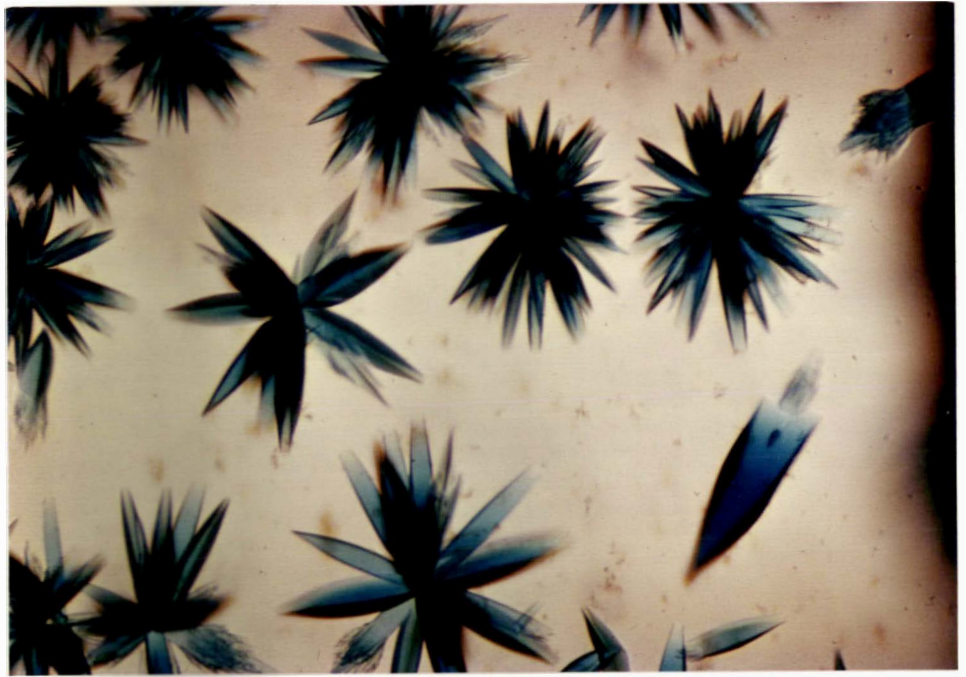
Figure 2·7 Crystals of Azurin from *Alcaligenes denitrificans* showing Spherulitic Growth Habit (63 ×).
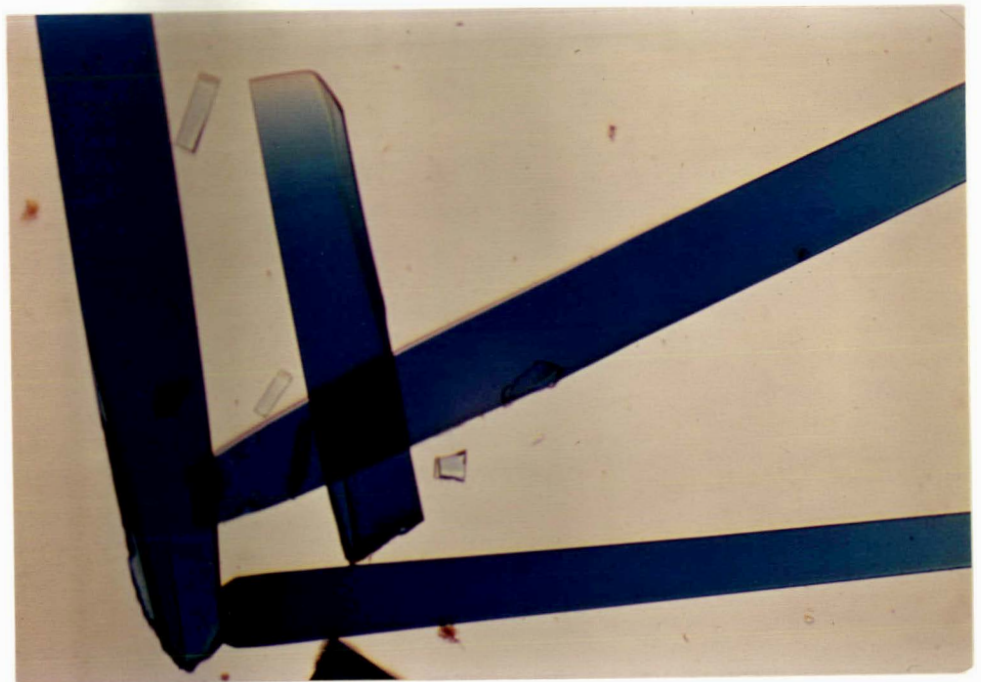
Figure 2·8 Crystals of Azurin used for data collection. (63 ×).

(b)   by vapour diffusion,where the protein solution (12 mg ml^{-1}, in

50% saturated ammonium sulphate solution made up in 0.1 M

phosphate buffer pH 6.0) was equilibrated against 62-66%

saturated ammonium sulphate at 37°C.

In both cases plastic tubes produced the best results, and crystals

appeared in 6-7 weeks. The same crystal form was always obtained, i.e.

very intensely coloured blue needles of rectangular or square cross-

section at one end, tapering to the other. The crystals tended to

grow in clusters,as shown in Figure 2.7,joined at their square ends,

but individual crystals up to 2-4 mm in length could be separated from

the clusters. Sometimes large single crystals grew with their square

ends attached to the tube. Figure 2.8 is a photograph of part of such

a crystal. As shown below, the needle axis was found to correspond to

the crystallographic c axis, while the a and b axes lie along the

diagonals of the cross section.

Figure 2.9   A schematic diagram of a crystal of azurin from Alc. denitrificans,

showing the relationship between the crystal morphology

and the crystallographic axes.

2.4 Characterization of the Crystals of Azurin from Alc. denitrificans.

The crystals grown as described above were mounted in thin-walled

glass capillaries,which contained a small reservoir of mother liquor

at one end and were sealed with wax.   Zero and first level precession

photographs (μ = 16°) were taken using Ni-filtered CuK$_{\alpha}$ radiation from

a Phillips 1130/00 X-ray generator operated at 20 mA and 40 kV.
Exposures varied from 20-40 hours depending on the size of the
crystal.

The crystals were found to belong to the orthorhombic space
group $C222_1$ with cell dimensions a = 75.0 Å, b = 74.1 Å and c = 99.5 Å.
The density, measured in a water-saturated bromobenzene/kerosene
gradient column (112) was 1.26 g cm^{-3}. Assuming the asymmetric unit
contains two molecules of weight 14,000 daltons, the crystal volume
per unit of protein molecular weight, Vm,is calculated as 2.47, a
value comparable with other protein crystals (113). If the partial
specific volume is assumed to be 0.74 cm^3 g^{-1}, the fraction of the
volume of the unit cell occupied by the solvent is thus calculated to
be 50.2%. The crystals diffract to at least 2 Å, and are radiation
stable. An interesting feature of the diffraction patterns is the
fact that they show strong tetragonal pseudosymmetry at low resolution.
The pseudotetragonal cell (a' = b' = 52.4Å, c' = 99.5Å) is approximately
half the volume of the true unit cell, and has axes which are at 45°
to the true a and b axes. It is very similar to that reported by
Strahs (98) for azurin from Ps. denitrificans (Alc. sp. of this work)
    (viz a = b = 53.2Å, c = 101Å,space group $P4_1 22$). Figure 2.10
is a schematic drawing of the relationship between these two space
groups. Figure 2.11 shows precession photographs ($\mu = 16^o$) of the
h0ℓ, 0kℓ, and hk0 zones. Systematic absences for this space group are
00ℓ absent when ℓ = 2n+1, and hkℓ absent when h+k = 2n+1.

2.5 The Preparation of Heavy Atom Derivatives.

A three dimensional crystal can be represented by a Fourier series

$$\rho(x,y,z) = \frac{1}{V} \sum_h \sum_k \sum_\ell F_{hk\ell}[\exp i \, \alpha_{hk\ell}]\exp[-2\pi i(hx + ky + \ell z)]$$

where ρ is the electron density at a point x,y,z

    V is the volume of the unit cell
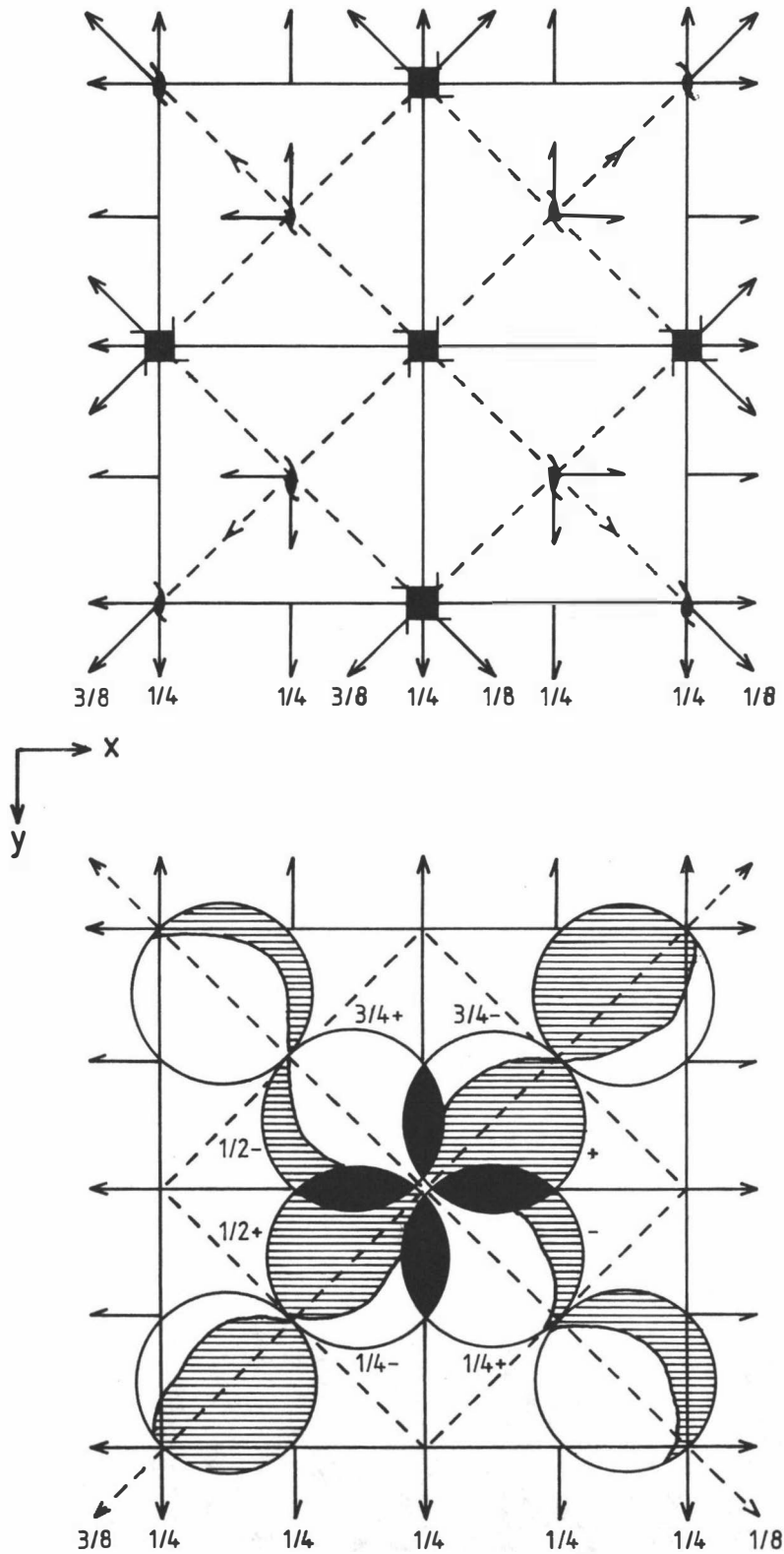
Figure 2-10. A schematic representation showing the relationship between the space group $C222_1$ (—) and the tetragonal space group $P4_122$ (---). The two cells are aligned when the origin of the tetragonal cell is shifted to $z=1/8$ and the other two axes are redefined along the face diagonals.
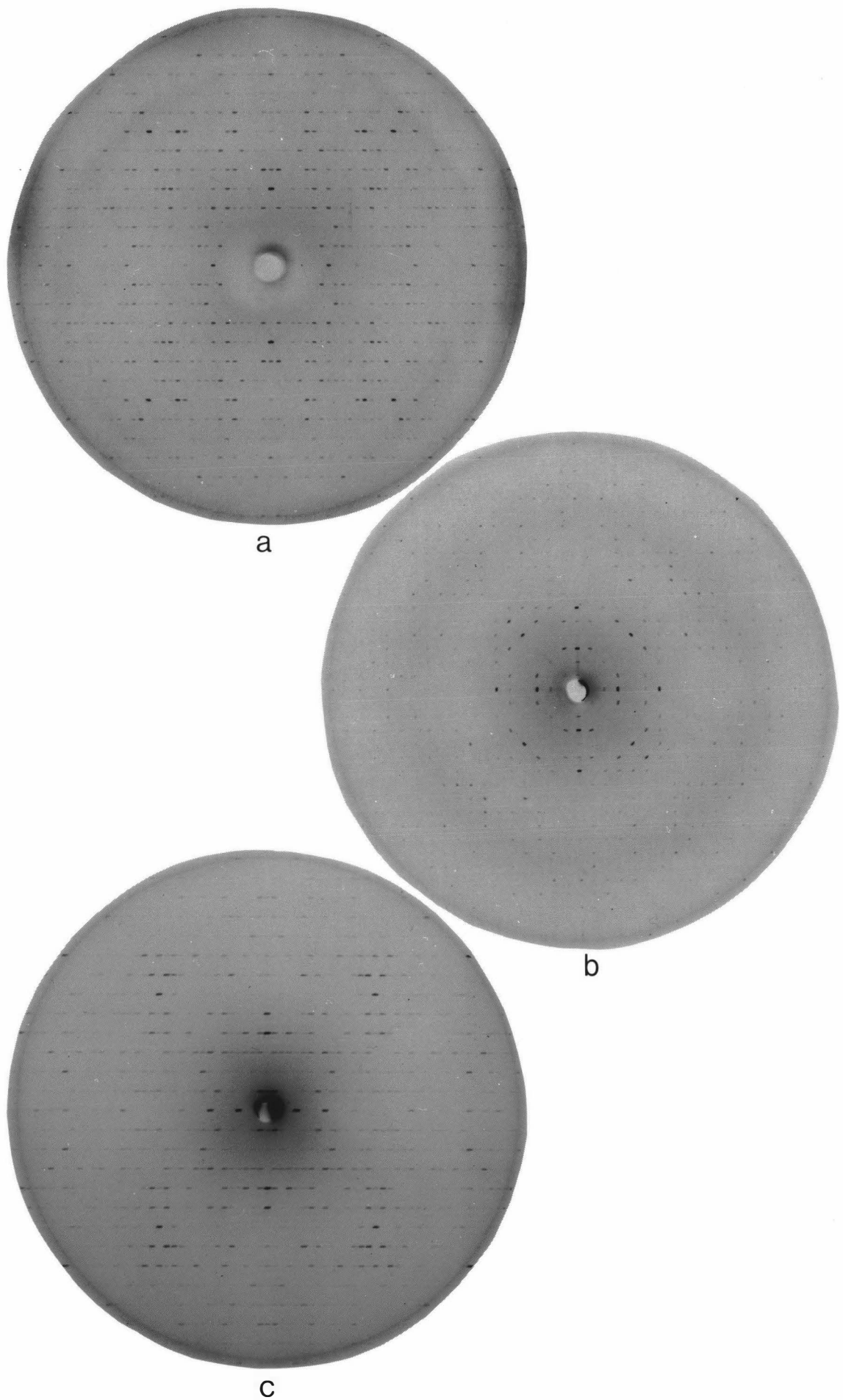
Figure   2.11   Precession photographs (μ = 16°) of (a) the h0l, (b) the hk0, and
(c) the 0kl zones of native azurin crystals.   The hk0 zone clearly
shows the peusdo-tetragonal symmetry at low resolution.

$F_{hk\ell}$ is the amplitude of the wave described by the indices $hk\ell$

$\alpha_{hk\ell}$ is the phase of the wave.

The amplitude of the wave is simply the structure factor magnitude and can be derived directly from the intensity of the reflection $hk\ell$. However the phase of the reflection cannot be observed directly, this being the basis of the so called "phase problem", the major difficulty in structure determination.

In this study, as in other protein structure analyses, the method of multiple isomorphous replacement (114,115,116,117), supplemented by anomalous scattering measurements (118,119,120), was used to obtain phases. This involves the preparation of derivatives of the protein, in which additional ions or small molecules containing at least one "heavy" atom (i.e. an atom of high atomic number) are bound to the protein without disturbing the structure of the protein or the crystal packing. The space group and cell dimensions should then be unchanged, and the intensity differences in the X-ray diffraction pattern will arise solely from the additional atoms in the structure. The heavy atom species may be covalently bound to the protein, or held by non-covalent forces, the important thing being that they residue in exactly the same place on each protein molecule (i.e. are well ordered), and that the occupancy is high (i.e. that they bind to a high proportion of the protein molecules in the crystal). The intensities of the diffraction pattern are then measured for both parent and derivative crystals to the desired resolution, and the position of the heavy atom(s) found, usually by a Patterson synthesis (see Chapter 3).

Heavy atom derivative screening was carried out on medium sized crystals, the largest crystals being saved for diffractometer data collection. Early attempts to remove the copper atom, with a view to replacing it with heavier metal atoms, did not appear to be

successful (see Appendix III), and the commonly-used trial and error method was therefore employed in searching for suitable isomorphous heavy atom derivatives. In this method protein crystals are soaked in solutions of a heavy atom reagent, so that the latter can diffuse into them to possible, unspecified, binding sites. Initially, azurin crystals were transferred to capped vials containing 75% saturated ammonium sulphate solutions, made up with 0.1 M phosphate buffer pH 6.0 and approximately 1 mg ml^{-1} of heavy atom reagent. After a soaking period of 1 week, a zero level precession photograph obtained from the crystal was compared with an appropriate photograph from a native crystal. If intensity differences were present, but were small, three steps could be taken to try to increase the amount of substitution. Firstly, the concentration of the reagent was increased to levels of up to 2.5 mg ml^{-1} as this sometimes will give rise to further binding (105). Secondly, where the heavy atom reagent was partly or completely insoluble in the mother liquor, a change of buffer from phosphate to acetate often effected an improvement in substitution. This is because many of the heavy metal ions found in such reagents are more soluble in acetate than in phosphate buffer (105). As some heavy atom reagents, for example uranyl or lanthanide complexes, form insoluble salts with phosphate ions, acetate buffers were also used with these reagents. The buffer used was 0.05 M ammonium acetate pH 5.0, and the protein crystals were always pre-soaked in at least two changes of buffer before being introduced to the soaking solutions. Where the heavy atom reagents remained insoluble, they were left as precipitates in the bottom of the tube, in most cases in contact with the crystals for longer periods of time. This was because even when a reagent appears to be insoluble, it may dissolve very slowly, allowing "heavy" atoms to diffuse into the protein. Finally, the temperature of the soaking solution was raised to 37°C in an attempt to increase both the rate of reaction, and the solubility of the heavy atom reagent.

Generally it was found that increasing the temperature did not increase substitution, but merely increased the rate at which it occurred, and thus was of no real advantage.

Where damage to the crystal, such as cracking, a change in cell dimensions, or a deterioration in the quality of the diffraction pattern was observed following soaking, the concentration of the reagent used was reduced, as was the time of soaking. Where this failed to reduce the damage, soaking was carried out at 4°C to try to control the amount of substitution by slowing down the rate of diffusion of the reagent into the crystal. However, it was found that this procedure generally offered no advantage, as crystal damage still occurred. Table 2.3 is a summary of the heavy atom soaking experiments, the middle column indicating the extent of the intensity changes, the last column giving an estimation of the amount of non-isomorphism with the native crystals. This can refer to either physical damage to the crystal itself, or to a change in the dimensions of the crystal lattice. A change in cell dimensions of 0.5% was the maximum allowable before the crystal was considered to be non-isomorphous with the native (121).

Table 2.3   Results of the Heavy Atom Soaking Experiments.

Compound	Intensity Change	Crystal Damage
Mercuric Acetate	0	0
Baker's Mercurial [a]	0	0
PCMBS [b]	0	0
Thiomersal [c]	+ + + +	xx [g]
Mercuric thiocyanate	+ + + +	xxx [g]
DMA [d]	+	xxx
K_2HgI_4	−	xxxx
PHMB [e]	0	0
phenyl mercury nitrate	0	0
ethyl mercury phosphate	0	0
mercuric chloride	+	0

Table 2.3 (Continued)

Compound	Intensity Change	Crystal Damage
methyl mercury iodide	+	xx
ethyl mercury chloride	++	0
$Hg(NH_3)_2Cl_2$	+++	0 [g]
$K_2Pt(CN)_4$	+++	0
$Pten(Cl)_2$ [f]	−	x
K_2PtCl_6	−	x
$K_2Pt(NO_2)_4$	++	xx
cis-$Pt(NH_3)_2Cl_2$	−	x
$K_2Pt(SCN)_4$	−	xxxx
$NaAuCl_4$	++++	xxx
$KAu(CN)_2$	++++	0
$K_3UO_2F_5$	0	0
uranyl acetate	++	0
$UO_2(NO_3)_2$di pyrO [*]	0	0
$UO_2(NO_3)$pyrO [*]	0	0
$UO_2(SCN)_2$CNO [*]	−	x
UO_2Br_2 + DPSO [*]	0	0
$KAu(CN)_2$ + uranyl acetate	++++	0
Samarium acetate	+	0
Lanthanum acetate	0	0
Lanthanum nitrate	0	0
Na_3IrCl_6	++	0
Lead nitrate	+	0
Lead acetate	+	0
Lead chloride	0	0
Silver sulphate	−	xxxx
Silver nitrate	−	xxxx
Thallous chloride	+	0

0	none	0	none
+	weak	x	no visible damage but did not diffract
++	moderate	xx	changes in axial lengths
+++	strong	xxx	crystal cracked
++++	very strong	xxxx	crystal completely broken up

Table 2.3   (Continued)

(a) $CH_3COO-Hg-CH_2-CH(OCH_3)$

      $CH_3COO-Hg-CH_2-CH(OCH_3)$

(b)   p-chloro-mercuri-
            benzene-sulphonic acid $Cl-Hg-\langle\bigcirc\rangle-SO_3^-$

(c) $CH_3CH_2-Hg-S-\langle\bigcirc\rangle$ CO_2Na (attached)

(d)   dimercury acetate NO_3-Hg
                                    $CHCOO^-$
                            NO_3-Hg

(e)   p-hydroxy-mercuri-benzoate $HO-Hg-\langle\bigcirc\rangle-CO_2^-$
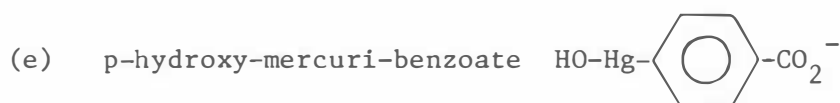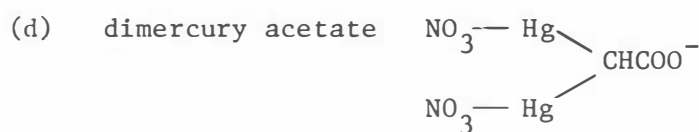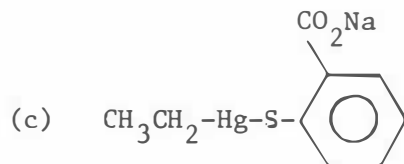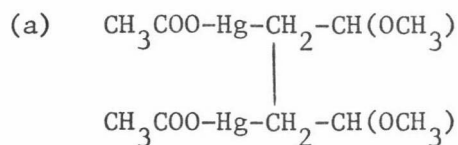
(f)   en   ethylene diamine

*   reagents kindly donated by Dr C.E.F. Rickard, Chemistry

     Department, University of Auckland.

(g)   the crystal became colourless.

     Some heavy atom compounds led to a fading out of the diffraction

pattern at low resolution, and some led to the physical destruction

of the crystal. This was the case for almost all of the platinum

compounds. In retrospect, this was probably due to the large number

of methionine residues on the surface of the molecule in one particular

region which formed a close contact with another molecule in the

crystallographic asymmetric unit. Platinum compounds are known to bind

to methionine side chains (122), and in this case such binding would

disrupt the lattice.

When the reagent Na_3IrCl_6 was first used, no substitution occurred, probably because of the insolubility of the compound. However, it was found that the reagent dissolved very slowly, and after six months, a crystal which had been left soaking was very cracked, indicating possible substitution. In order to accelerate and control substitution, 2 mg of the reagent was heated in a 75% saturated solution of ammonium sulphate made up with acetate buffer until the solution became quite brown. This was cooled and decanted into a fresh tube before intro-duction of the crystal. Soaking times of 1 to 6 weeks produced only minor intensity changes, so the experiment was repeated, this time leaving the undissolved crystals of Na_3IrCl_6 in contact with the protein crystals. Substitution was much improved after only 8 days, but the crystals were cracked, possibly because of the abrasive effect of the hard Na_3IrCl_6 crystals. In an attempt to protect the protein crystals, the heavy atom reagent was placed in a small dialysis bag, which was suspended in the mother liquor above the crystals during the soaking period, allowing any further "heavy atom" species to diffuse into the solution, and hence into the protein crystals. This had the desired effect, moderate substitution being achieved in the crystals without any cracking. Derivatives finally selected for data collection are shown in Table 2.4, together with the conditions under which they were prepared.

2.6 Data Collection.

Only a small number of azurin crystals were large enough for diffractometer data collection. In the early stages, therefore, photographic data was collected using medium sized crystals with a view to deciding which derivatives were likely to be worth collecting diffractometer data on.

Table 2.4   Preparation of the Heavy Atom Derivatives Used for Three Dimensional Data Collection.

Reagent	Concentration	Buffer	pH	Time of Soak	Temperature
$KAu(CN)_2$	0.5 mg/ml	0.1M phosphate	6.0	20 days	room
uranyl acetate	2.0 mg/ml	0.05M amm. acetate	5.0	7 weeks	room
$KAu(CN)_2$ + uranyl acetate	0.5 mg/ml + 2.2 mg/ml	0.05M amm. acetate	5.0	3 weeks	room
$Hg(NH_3)_2Cl_2$ [a]	2 mg/ml	0.05M amm. acetate	5.0	11 weeks	room
thiomersal [a]	0.5 mg/ml	0.05M amm. acetate	5.0	7 days	room
Na_3IrCl_6 [b]	2.5 mg/ml	0.05M amm. acetate	5.0	8 days	room
EtHgCl [c]	saturated	0.05M amm. acetate	5.0	7 weeks	room

(a)   the reagents caused the crystals to become colourless.

(b)   the reagent was dialysed into the solution.

(c)   ethyl mercury chloride

2.6.1 Photographic Data Collection.

Only the data for the centric zones h0ℓ and 0kℓ were collected photographically for the native and various heavy atom derivative crystals,as the hk0 zone was not easily obtained due to the difficulty of mounting needle shaped crystals with the needle axis across the capillary tube. Precession photographs were taken in 2-film packs so that the second film had a much weaker exposure than the first. The photographs were scanned using a rotating-drum OPTRONICS P1000 microdensitometer, the raster size being set at 100 x 100 μm. The resulting density data for each film was written on to magnetic tape for processing on a Burroughs B6700 computer. Intensities were then extracted using the program DENSFOR (appendix II) which predicts the position of each reflection, integrates its intensity, and makes an appropriate background correction. Film-to-film scale factors were calculated for each 2-film pack, and the films merged. Symmetry-related reflections were also merged and Lorentz and polarisation factors applied,giving a set of unique corrected intensities. Derivative data were scaled to the native data set, using a Wilson-type plot of $\log(\Sigma\ F_{PH}^{2}/\Sigma\ F_{P}^{2})$ versus $(\sin\theta/\lambda)^{2}$, giving a list of native and derivative structure amplitudes on the same scale.

2.6.2 Diffractometer Data Collection.

For diffractometer data collections, the largest available crystals were used. These were typically 0.5 mm by 0.5 mm in cross section, and were always cut (with a scalpel blade) so that the length did not exceed about 1.0 mm. The crystals were mounted in thin-walled glass capillaries in the usual way. One set of data (a second set of native data) was collected on a Nonius CAD-4 diffractometer in the Chemistry Department, University of Auckland. For all other data sets, the instrument used was a Hilger and Watts four circle X-ray diffractometer controlled by a PDP-8I computer at the Chemistry Division,

Department of Scientific and Industrial Research, Wellington. The

X-ray tube was operated at 40 kV and 25 mA, the radiation used

being CuKα.

For all crystals, a large incident beam collimator (1.4 mm)

was used, in order to ensure that the whole of the crystal was fully

bathed in the X-ray beam. The diffracted beam collimator was the

smallest that could be used without significant loss of intensity

from reflection peaks. This was either 2.5 or 3.5 mm depending on

the size of the crystal. Integrated intensities were collected by

an ω-scan (105,123), the width of the scan being typically 0.5 to

0.6 degrees, depending on the mosaic spread of the crystal. Counting

times were deliberately kept short in order to minimise radiation

damage, and to allow as much data as possible to be obtained for each

crystal; typically 40 to 50 seconds was spent counting each peak, and

5 seconds either side of it for the background.

For orthorhombic space groups, $F_{hk\ell} \equiv F_{\bar{h}\bar{k}\ell} \equiv F_{h\bar{k}\bar{\ell}} \equiv F_{\bar{h}k\bar{\ell}}$, and

$F_{\bar{h}\bar{k}\bar{\ell}} \equiv F_{\bar{h}k\ell} \equiv F_{h\bar{k}\ell} \equiv F_{hk\bar{\ell}}$. The data collection scheme used by the

Hilger and Watts diffractometer was such that ℓ was the fastest

varying index; thus the Friedel pairs $hk\ell$ and $hk\bar{\ell}$ were collected

so that the two measurements would be made close together in time.

The choice of the particular quadrant of reciprocal space within which

data was collected was made with some care. This was because the use

of long needle-shaped crystals, adhering to the wall of a glass capillary,

gave rise to a variation of peak shape in different regions of

reciprocal space; in some regions, reflection widths might be as great

as 1.0^o compared with as little as 0.3^o in others. Where possible,

data collection utilized the quadrants for which reflections were the

sharpest, thus allowing more accurate setting and smaller scans. The

data was collected in shells of reciprocal space, chosen to contain

approximately the same number of reflections.

For the native data collection, 2 crystals were used, one for data covering the range $\theta = 1^{\circ}$ to 14.5° (i.e. data to 3.08Å) and the other for data between $\theta = 13.5^{\circ}$ to 16.5° (i.e. data to 2.7Å). For each derivative, a complete high resolution data set was collected from one crystal (except for the double derivative ($KAu(CN)_2$ + uranyl acetate) for which two were used). Although the use of only one crystal to collect a whole set of high resolution data leads to greater radiation damage, errors caused by scaling data from different crystals together no longer arise. This was, in any case, forced to some extent by the need to travel to another centre to set up crystals for the data collection. The intensity measurements plus two back-grounds were written on to paper tape and subsequently transferred to magnetic tape for storage and processing. Table 2.5 shows the unit cell parameters for the native and derivative crystals, derived from diffractometer measurements.

Radiation Damage.

Radiation damage to the crystals was monitored by comparing the intensities of three strong standard reflections which were measured every 100 general reflections. These were chosen so that they were well distributed in reciprocal space. Generally, two of the reflections had $\chi \approx 0$ and ϕ values separated by 90°, while the third one had χ near 90°. This ensured that if crystal movement occurred it would show as uneven backgrounds in one or other of the standards. For high resolution data collections, the standard reflections were chosen from an outer shell where possible, since it was felt that high angle reflections might be the first affected by radiation damage. The corrected intensities of these reflections were plotted as a function of exposure times, and a best straight line (or series of straight lines, since discontinuities sometimes occurred) was drawn through the points. An example of such a

Table 2.5   Unit Cell Parameters for the Native and Eight Heavy
            Atom Derivatives for which Diffractometer Data were
            Collected.

Derivative	$x(Å)$	$y(Å)$	$z(Å)$
Native	75.0	74.1	99.5
$KAu(CN)_2$	75.0	74.4	99.65
$K_2Pt(CN)_4$	74.95	74.1	99.75
uranyl acetate	74.9	74.2	99.7
$KAu(CN)_2$ + uranyl acetate	74.9	74.55	99.5
Thiomersal	75.0	74.15	99.75 - 100.1 [a]
EtHgCl	75.20	74.20	99.55
Na_3IrCl_6	75.20	74.25	99.60
$Hg(NH_3)_2Cl_2$	75.27	74.56	99.50

(a)   See section 3.1.6

plot, that for the gold derivative, is shown in Fig. 2.12, which shows
how the fall-off is calculated as a function of the number of reflections
measured, or the number of hours of exposure.   Such a correction factor
was calculated for each standard reflection in all data sets, the mean
of the three being taken as the correction function.   For the native
crystals, the fall-off in the three reflections was roughly equal
(15.9%, 12.9% and 15%) while the derivative crystals showed much
greater sensitivity to radiation damage with more variation in fall-
off between standards (Table 2.6).
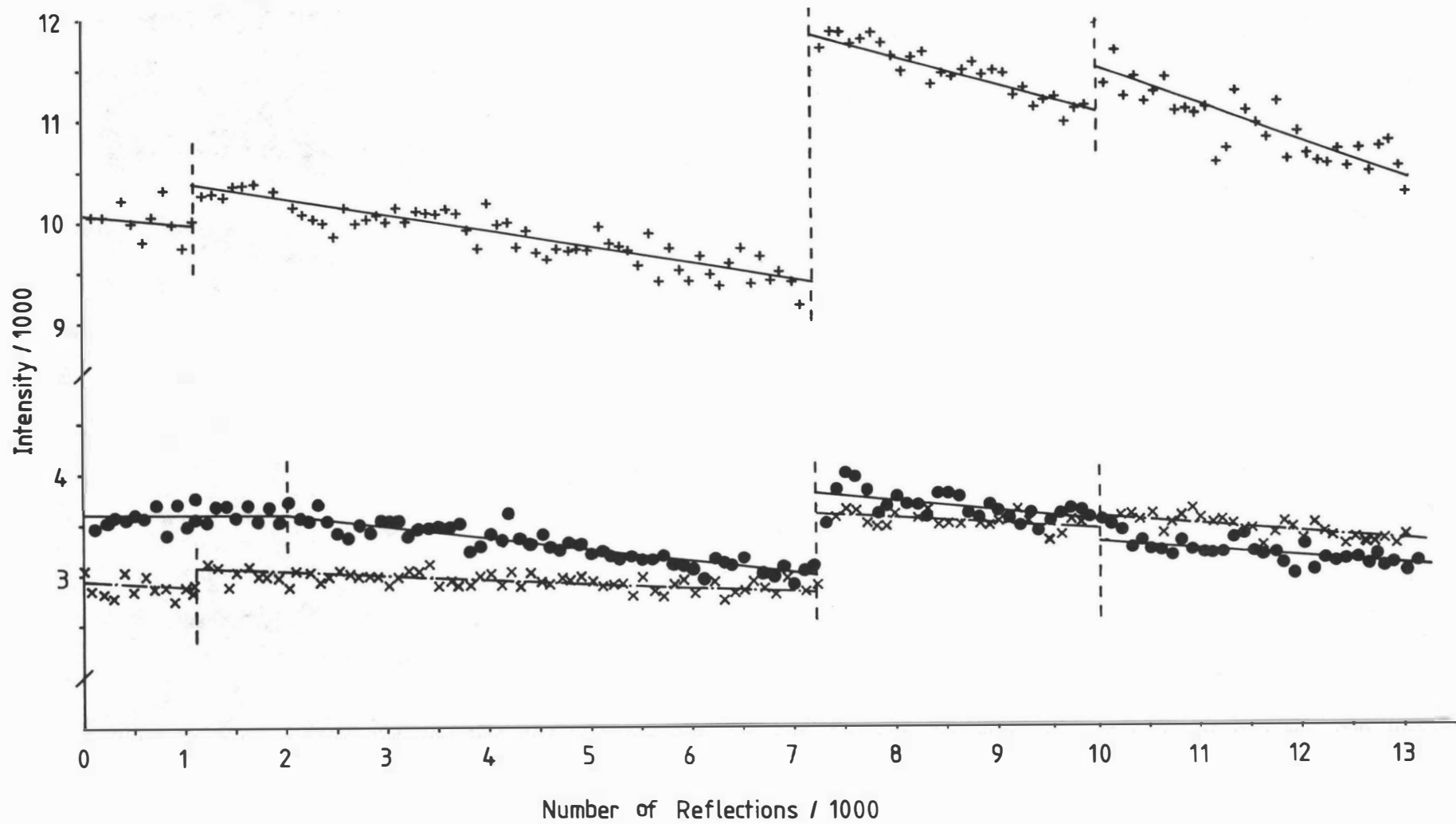
Figure 2-12. Intensity loss for three standard reflections due to radiation damage to the gold derivative.

The vertical lines represent realignment of the crystal during data collection.

+ = [14,0,17]   ● = [0,22,0]   × = [20,0,0]

Table 2.6   Statistics of Intensity Losses for the Native and

and Derivative Crystals.

Derivative	Resolution Range	No. of measurements*	Radiation Decay (%) for Three Standards		
Native	to 3.0Å	10075	15.7	12.9	15.0
KAu(CN)$_2$	to 2.9Å	13499	27.5	21.6	31.7
uranyl acetate	to 3.1Å	10625	25.7	31.9	27.3
K$_2$Pt(CN)$_4$	to 2.9Å	11701	35.7	36.0	32.9
EtHgCl	to 3.3Å	9452	28.1	33.9	39.0
Double derivative	to 4.0Å	5319	26.6	32.0	19.5
	4.0Å-3.0Å	7027	18.9	24.6	10.8
Na$_3$IrCl$_6$	to 4.0Å	3915	14.3	19.4	11.1
Thiomersal	to 3.5Å	8159	34.1	32.8	34.4
Hg(NH$_3$)$_2$Cl$_2$	to 5.0Å	2576	13.3	13.8	15.8
Chromium(II)[‡]	to 4.0Å	3929	10.5	7.9	10.1

*   In some data collections (e.g. KAu(CN)$_2$) substantial portions of the data had to be remeasured as a result of crystal movement.

‡   see chapter 6.

From the table it can be seen that for some reflections, in some of the derivatives, the overall decay in diffraction intensity exceeds 30%. It has been shown that the effects of radiation on protein crystals are not always continuous with time, as some reflections can show either increases or decreases in intensity with exposure (124,125). Thus it is not possible to correct for all the effects of irradiation by simply applying an overall correction factor.

For this reason some workers prefer to discard the crystals once the intensity of a set of reference reflections has decreased by approximately 10% (126,127). However it is advantageous if the useful life of a crystal can be extended by radiation damage corrections, since if a derivative crystal is replaced by another, which appears identical, but which actually differs significantly due to differences in substitution, errors larger than those due to the approximate corrections will be introduced into the calculations (128). As successful applications of empirical radiation decay functions ranging up to 33% have been reported (128,129), the correction functions were applied, and the data used in structure determination.

Absorption Corrections

Absorption corrections were made by the semi-empirical method of North et al. (130). At the beginning of each data set, an empirical absorption curve was obtained from an azimuthal reflection (i.e. one at $\chi = 90^{\circ}$ in the Hilger and Watts 4-circle geometry). ω-scans were made through this reflection at intervals of 10° from $\phi = 0^{\circ}$ to $\phi = 360^{\circ}$, and integrated intensities obtained, thus allowing a correction function to be calculated for the loss of X-ray intensity caused by passage through the crystal in a mean direction perpendicular to the rotation axis. Such a correction takes into account the crystal shape, the adhering liquid, and the glass capillary tube. Since the crystals were oriented with the c axis along the ϕ direction, a c axial reflection was always used. No significant difference was noted on using the low-angle 0 0 8, or the much weaker high-angle 0 0 20 reflection. For each crystal, I_{max}/I_{ϕ} was plotted against ϕ, the resulting curve being used to calculate the absorption of any reflection hkℓ (130).Maximum absorption corrections ranged from 1.20 1.59.

Data Processing.

Data processing was carried out on a Burroughs B6700 computer at the Massey University computer centre. The programs used in processing are outlined in Appendix II. Background correction of the raw diffractometer data made use of a background averaging routine, developed by Dr P.E. Nixon, and similar to that suggested by Krieger et al. (131). A subset of the backgrounds was taken that consisted of the background measurements for reflections whose ratio of total counts to total background was less than a factor k (typically between 1.2 and 1.5). Thus only the weaker reflections (between 30-50% of the reflections in a data set) were used, so that the background measurements would not contain significant contributions from mis-set strong peaks. These backgrounds were then averaged as a function of θ, ϕ and χ, the averaged values being applied to all reflections in making background corrections. The relability of the data was greatly improved by this procedure, as judged by comparison of the symmetry related reflections and by comparison with photographs. Data were then corrected for Lorentz-polarization factors, absorption and the fall-off in intensity due to radiation damage.

Equivalent reflections, and reflections which had been measured more than once, were merged by taking their weighted mean. Where more than one crystal was used for data collection (i.e. for the native and the double derivative), the data from different crystals were first put on to a common scale using the reflections from overlapping shells. Reflections with negative intensities were retained at this stage, being given intensities equal to half their standard deviation; later, for heavy atom refinement and phase calculation, all reflections for which $I_{hk\ell} < 1.5\ \sigma(I_{hk\ell})$ were omitted. Friedel pairs were kept separate for heavy atom derivatives and the extent of Friedel differences were monitored by calculating reliability factors for agreement between the Friedel pairs. These are shown in Table 2.7.

Table 2.7   Reliability Factors for Agreement Between Friedel Pairs in $\sin^2\theta$ Ranges.

$\sin^2\theta$	.005	.010	.015	.020	.025	.030	.035	.040	.045	.050	.055	.060	.065	.070	.075	.080	.085
native[*]	.037	.023	.039	.051	.034	.036	.033	.041	.041	.049	.067	.153	.186	.137	.091	.108	.107
$KAu(CN)_2$	.028	.029	.034	.039	.035	.024	.029	.035	.037	.045	.051	.058	.075	.098	.129		
$K_2Pt(CN)_4$	.025	.023	.060	.057	.066	.073	.076	.043	.037	.039		.068	.084	.119	.131		
uranyl[a]	.033	.026	.030	.034	.029	.024	.024	.034	.038	.050	.068	.074	.099				
Double Deriv.	.046	.041	.049	.053	.049	.039	.043	.049	.056	.062	.067	.080	.101	.146			
$Hg(NH_3)_2Cl_2$	.078	.074	.077	.089	.094												
EtHgCl	.020	.018	.022	.022	.022	.020	.022	.027	.032	.040	.052						
Na_3IrCl_6	.029	.024	.039	.044	.044	.038	.045	.058									
Thiomersal	.032	.026	.035	.040	.040	.(	.036	.045	.051	.070	.074	.099					

$R_F = \Sigma \mid F^+ - F^- \mid / \Sigma \tfrac{1}{2} \mid F^+ + F^- \mid$ where F^+ and F^- are the amplitudes of reflections hkℓ and hk$\bar{\ell}$

(a)   Scaling of the Friedel pairs was carried out by making the mean values of $I_{hk\ell}$ and $I_{hk\bar{l}}$ equal in small blocks of reciprocal space.

*   For native crystals, F^+ and F^- are of course, equivalent reflections.

The increase in R_F value with increasing θ is probably due mostly to the larger number of weak reflections at higher resolution. It can also be seen that for the native data equivalent reflections agree very well up to $\theta = 13.6^{\circ}$ ($\sin^2\theta = 0.055$). The poor agreement between 14.2° and 15.3° ($\sin^2\theta = 0.060 - 0.070$) arises from the merging of the data from two data sets, one measured on a Hilger and Watts 4-circle diffractometer, and the other on a CAD-4 diffractometer, and appears to be a consequence of differences in the measurement of the weaker reflections. Also calculated was the average intensity for reflections as a function of resolution. In Figure 2.13 these values are plotted for some of the heavy atom derivatives that were used in the final structure determination.
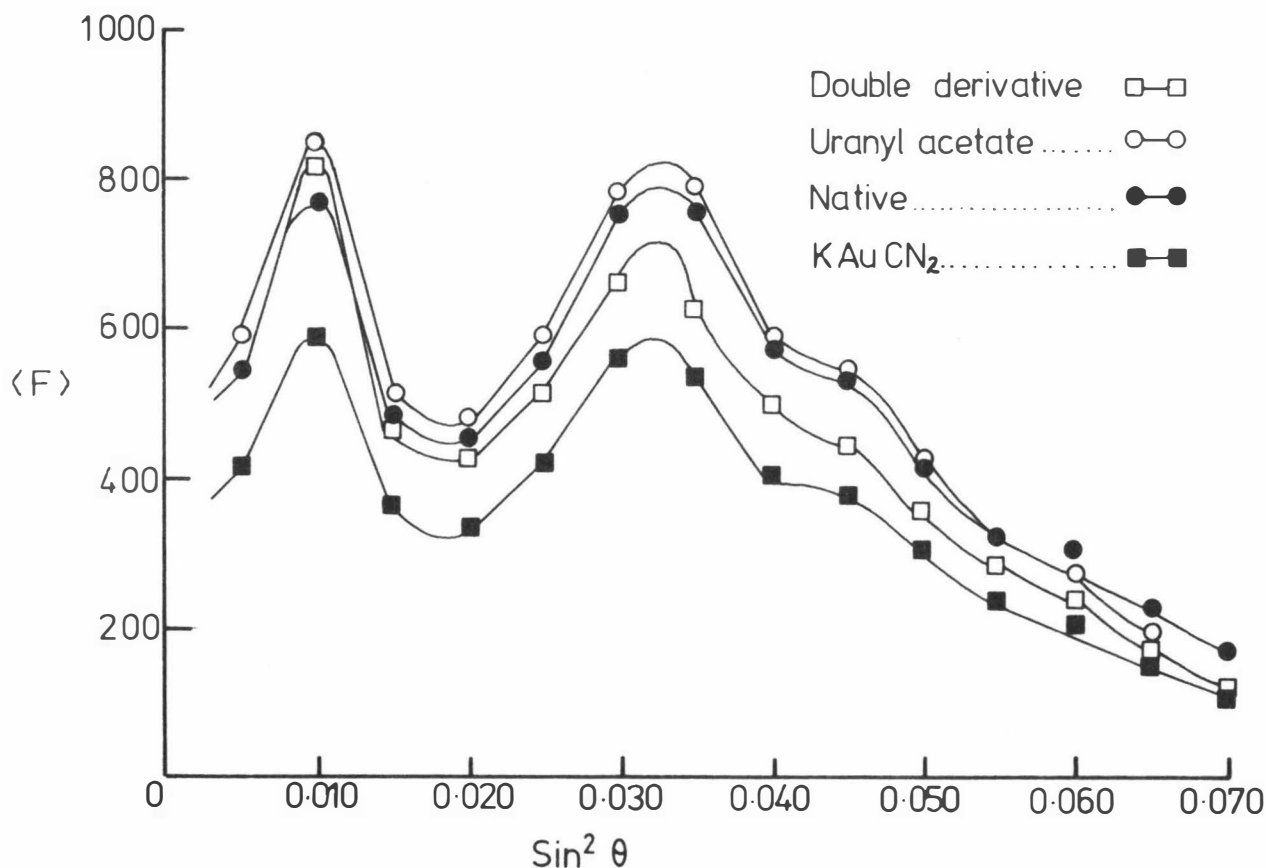
Figure 2.13   The average intensities for reflections as a function of resolution for the native and three derivative sets of data.

Native and heavy atom derivative data were scaled together following the method of Wyckoff et al. (132). In this method, derivative intensity data are scaled by a factor k_I, where

$$k_I = \frac{\Sigma F_P^2}{\Sigma F_{PH}^2} \; [1 + R_I^2]$$

and $\;\; R_I = \Sigma|F_{PH} - F_P|/\Sigma F_P \;\;\;$ (F_P = protein structure amplitude)

(F_{PH} = derivative structure amplitude)

The latter index also gives some idea of the overall change in intensity between native and derivative crystals and hence of the amount of substitution. Values for the various heavy atom derivatives are shown in Table 2.8. It can be seen that the uranyl, ethyl mercury chloride and iridium derivatives showed very little substitution, suggesting that they might be of limited value in any phase calculation.

Table 2.8   Statistics for the Derivative Data Collections.

Derivative	No. of reflections	Scale Factor	R_I
Native	6991	1.0000	–
$KAu(CN)_2$	6414	1.2093	0.164
$K_2Pt(CN)_4$	6510	1.0653	0.147
uranyl acetate	5113	0.9958	0.095
double derivative	5946	1.0994	0.231
$Hg(NH_3)_2Cl_2$	1337	1.5712	0.217
ethyl mercury chloride	4293	0.9006	0.065
Na_3IrCl_6	2003	1.2305	0.059
Thiomersal	3933	1.0649	0.195

$R_I = \Sigma|F_{PH} - F_P|/\Sigma F_P.$ Scale Factor $= \dfrac{\Sigma F_P^2}{\Sigma F_{PH}^2} (1 + R_I^2)$

CHAPTER 3

RESULTS

3.1 Determination of the Heavy Atom Parameters

Before any heavy atom isomorph can be used for phase determination, the sites of substitution of the heavy atoms themselves must be determined, from the observed differences in structure amplitudes, caused by their introduction into the protein.

3.1.1 The Patterson Function

Patterson syntheses calculated using the function

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

where V is the volume of the unit cell and

F_H is the structure amplitude due to the heavy atom.
do not yield the atomic sites directly, but give the distribution of vectors between pairs of heavy atoms (133). The essential problem is to obtain the best representation of F_H^2 from the differences between native and derivative diffraction patterns.

The relationships between the structure factors for the protein, $\underset{\sim}{F}_P$, the derivative, $\underset{\sim}{F}_{PH}$, and the heavy atom (s), $\underset{\sim}{F}_H$, are shown in Figure 3.1. From the diagram, the amplitude of $\underset{\sim}{F}_H$ is given by the cosine rule.

$$F_H^2 = F_{PH}^2 + F_P^2 - 2F_P F_{PH} \cos \alpha \qquad - (i)$$

$$= (F_{PH} - F_P)^2 + 2F_P F_{PH}(1-\cos \alpha) \qquad - (ii)$$
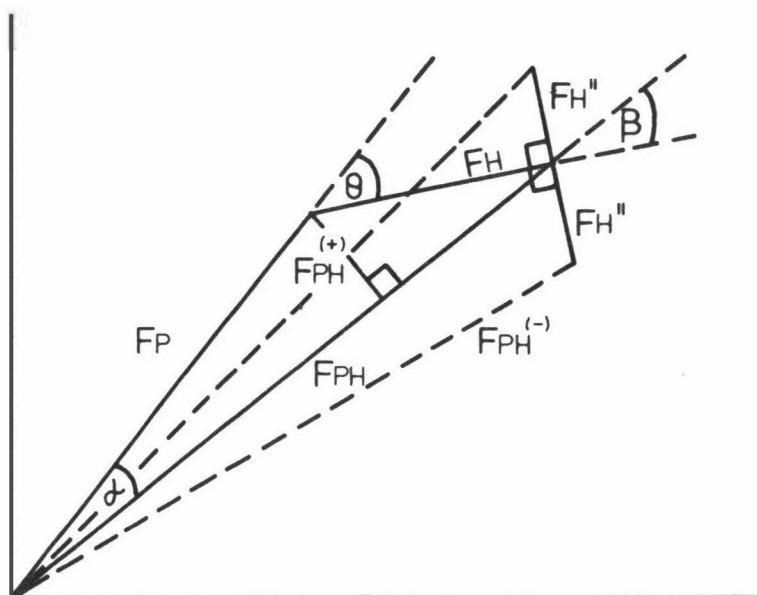
Figure 3.1   A Vector Diagram Illustrating, for an Arbitrary

Reflection, 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.

Centric data

For centrosymmetric reflections, α is likely to be 0 for all

but some weak reflections, so that $F_H^2 = (F_{PH} - F_P)^2 = \Delta_{iso}^2$,

where Δ_{iso} is the isomorphous difference. Thus, for a protein with

one or more centrosymmetric projections, a difference Patterson map

calculated with the above coefficients should contain only peaks

representing vectors between the heavy atoms. Patterson maps for

centrosymmetric projections may therefore allow an initial evaluation

of potential derivatives, although the overlap of peaks makes them

harder to interpret for derivatives with multiple sites.

Non-Centric Data

For non-centric data α is not known. However, using the coefficients $(F_{PH} - F_P)^2$ as an approximation for F_H^2 in the Patterson synthesis still produces peaks at the expected heavy atom-heavy atom vector positions, although the peaks will be smaller and there will be more background features due to protein-protein and heavy atom-protein vectors (134-137). It has been found that three dimensional Patterson syntheses in which information about anomalous scattering is included, give improved heavy atom vector maps (135,138-140). From Figure 3.1,

$$F_{PH}^{(+)2} = F_{PH}^2 + F_H^{''2} - 2F_{PH}F_H^{''} \cos(90+\beta) \qquad - (iii)$$

and
$$F_{PH}^{(-)2} = F_{PH}^2 + F_H^{''2} - 2F_{PH}F_H^{''} \cos(90-\beta) \qquad - (iv)$$

$$\therefore \quad F_{PH}^{(+)2} - F_{PH}^{(-)2} = 4F_{PH}F_H^{''} \cos(90-\beta) \qquad - (v)$$

But $F_{PH}^{(+)2} - F_{PH}^{(-)2} = (F_{PH}^{(+)} + F_{PH}^{(-)})(F_{PH}^{(+)} - F_{PH}^{(-)})$

$$\approx 2 F_{PH}(F_{PH}^{(+)} - F_{PH}^{(-)})$$

Combining these results, the anomalous difference Δ_{ano} is given by

$$\Delta_{ano} \approx (2/k) F_H \cos(90-\beta) \approx (2/k) F_H \sin\beta \qquad - (vi)$$

where $k = F_H/F_H^{''}$

As $F_P \sin\alpha = F_H \sin\beta$, $\Delta_{ano} \approx (2/k) F_P \sin\alpha$.

Thus, it can be seen, from the diagram, that the term $|F_{PH}-F_P|^2$ will be a poor estimate of F_H^2 when $\alpha \approx 90°$, while that of $|F_{PH}^{(+)}-F_{PH}^{(-)}|^2$ will be good in this situation. The two syntheses may then be combined to good advantage provided appropriate weights are used (134,135,139,140). Since $\cos\alpha = \pm\sqrt{1-\sin^2\alpha}$, combining equations (i) and (vi) gives

$$F_H^2 \approx F_P^2 + F_{PH}^2 \pm 2 F_P F_{PH} [1 - (k \Delta_{ano}/2F_P)^2]^{\frac{1}{2}} \qquad - (vii)$$

or, assuming $(1-x)^{\frac{1}{2}} \approx 1 - \frac{1}{2}x$, equation (vii) gives

$$F_{HLE}^2 \approx \Delta_{iso}^2 + (k\,\Delta_{ano}/2)^2$$

HLE refers to one of the two solutions for F_H, known as the lower estimate. For centric reflections when $\Delta_{ano} = 0$, $F_{HLE}^2 = (F_{PH} - F_P)^2$ and $F_{HUE} = (F_{PH} + F_P)^2$ where HUE refers to the other of the two solutions, the upper estimate.

In this work, conventional difference Patterson maps were calculated, using coefficients $(F_{PH} - F_P)^2$, as well as "combined" difference Patterson maps (or F_{HLE}^2 difference Pattersons). For the latter, coefficients were calculated using an empirical value of k, given by

$$k = 2 < \Delta F_{iso} > / < \Delta F_{ano} >$$

where the averaging is over non-centric reflections only (139).

For a protein molecule containing N heavy atoms in the unit cell, the Patterson map should contain N^2 peaks arising from the N possible vectors which can be drawn from each of the N atoms. Of these, N will be self vectors and concentrated in a very large peak at the origin, while the remaining N^2-N will be distributed throughout the cell. Obviously the interpretation of the Patterson synthesis of a multisite derivative, especially in a high symmetry space group, will be difficult owing to the large number of peaks. Also, if there are errors in the measured intensities, or if the parent and derivative data sets have been improperly scaled to each other, spurious peaks may appear. Where the binding of a heavy atom distorts the position of the atoms in the protein, partially destroying the isomorphism, some of the intensity changes will result from this distortion rather than from the presence of the heavy atom. Thus although the major peaks in a difference Patterson map should, in principle, be due to the heavy atom-heavy atom vectors, this is not always the case.

3.1.2   Difference Fourier Syntheses.

As it is especially difficult to detect binding sites of low occupancy, other methods such as cross-phased difference Fourier syntheses, and double difference Fourier syntheses may be used.

Cross-phased Difference Fouriers.

Co-ordinates which have been determined for one or more derivatives,by methods such as difference Patterson syntheses, are used to calculate approximate protein phase angles α_p, which are in turn used to calculate a difference synthesis with coefficients

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

where m is the mean figure of merit (see section 3.2.1) (141). The resulting electron density map should give,to a reasonable approximation, the difference between the electron density of the substituted and native protein. If the two crystal structures are isomorphous, the protein density will cancel leaving the desired heavy atom density relative to the same origin adopted for the phase calculation.

Double Difference Fouriers.

These may be calculated with the coefficients

$$(F_{H(obs)} - F_{H(calc)}) \, \exp(i\alpha_{Hc}) \qquad (142)$$

Such a synthesis should reveal any differences between the true heavy atom density and the approximation to that density calculated in terms of the heavy atom parameters determined by other methods. Additional minor sites should appear as positive peaks, and incorrect sites should be indicated by negative regions of density.

3.1.3   Heavy Atom Vectors in the Space Group $C222_1$

For the space group $C222_1$, the equivalent positions are:-

$$x \, , \, y \, , \, z \qquad\qquad \tfrac{1}{2}+x \, , \, \tfrac{1}{2}+y \, , \, z$$

$$x \, , \, \bar{y} \, , \, \bar{z} \qquad\qquad \tfrac{1}{2}+x \, , \, \tfrac{1}{2}-y \, , \, \bar{z}$$

$$\bar{x} \; , \; \bar{y} \; , \; \tfrac{1}{2}+z \qquad\qquad \tfrac{1}{2}-x \; , \; \tfrac{1}{2}-y \; , \; \tfrac{1}{2}+z$$

$$\bar{x} \; , \; y \; , \; \tfrac{1}{2}-z \qquad\qquad \tfrac{1}{2}-x \; , \; \tfrac{1}{2}+y \; , \; \tfrac{1}{2}-z$$

In Patterson maps, these give rise to a set of 6 heavy atom-heavy atom self vectors, at positions

$$0 \; , \; \pm 2y \; , \; \pm 2z \qquad\qquad \tfrac{1}{2} \; , \; \tfrac{1}{2}\pm 2y \; , \; \pm 2z$$

$$\pm 2x \; , \; \pm 2y \; , \; \tfrac{1}{2} \qquad\qquad \tfrac{1}{2}\pm 2x, \; \tfrac{1}{2}\pm 2y \; , \; \tfrac{1}{2}$$

$$\pm 2x \; , \; 0 \; , \; \tfrac{1}{2}\pm 2z \qquad\qquad \tfrac{1}{2}\pm 2x, \; \tfrac{1}{2} \; , \; \tfrac{1}{2}\pm 2z$$

i.e all the self vectors for a given heavy atom site are found on the Harker sections

$$0,v,w \; ; \; \tfrac{1}{2},v,w \; ; \; u,v,\tfrac{1}{2} \; ; \; u,0,w \; ; \; \text{and} \; u,\tfrac{1}{2},w.$$

The asymmetric unit calculated was $\frac{1}{8}$ of the unit cell, i.e. $x = 0 \rightarrow \tfrac{1}{2}$, $y = 0 \rightarrow \tfrac{1}{2}$, and $z = 0 \rightarrow \tfrac{1}{2}$. Where more than one heavy atom position was found, all sites were referred to a common origin by checking the cross vectors.

3.1.4   Refinement of the Heavy Atom Positions.

The heavy atom positions were refined using all centric data ($h0\ell$ + $hk0$ + $0k\ell$), combined together to form a pseudo-three dimensional data set. For centric data, for most reflections

$$F_H = |F_{PH} - F_P| \equiv \Delta_{iso}$$ assuming F_{PH} and F_P are on the same scale. For a very small number of reflections for which $F_H > F_P$,

$$F_H = |F_{PH} + F_P|$$

Reflections in the latter category are often referred to as "crossovers" because the sign of $F_{\sim PH}$ is opposite to that of $F_{\sim P}$ (138). A full matrix least squares program, CUCLS, was used   (see Appendix II), the function minimised being

$$E = \sum_{hk\ell} [|\Delta_{iso}| - F_H]^2$$

where Δ_{iso} is the observed isomorphous difference and F_H is the

calculated heavy atom structure amplitude. A typical centric R

factor was calculated where

$$R = \frac{\Sigma\,||\Delta_{iso}| - F_{Hcalc}|}{\Sigma\,|\Delta_{iso}|}$$

and is summed over centric data only. Since no allowance is made for

the few "crossover" terms for which Δ_{iso} is an incorrect estimation

of F_H, this R value will always make the agreement for a given heavy

atom derivative appear worse than it really is. Values of R for heavy

atom derivatives will also be inevitably higher than R values obtained

for small molecule structures because of the small size of Δ_{iso}

and the relatively large error. A good solution may give R = 40%,

but quite useable ones may give R = 60% or even higher. A totally

incorrect one will give R values of approximately 83% or higher (138).

All reflections were given an equal weight in the refinements,

and occupancies and the x,y, and z co-ordinates were refined first,

assuming an initial isotropic temperature factor B of $25\text{\AA}^2$ for each

site, a value similar to that commonly found for heavy atoms bound to

proteins (105,143,144,145). A refinement cycle in which the x,y and z

co-ordinates and the temperature factors were varied was usually

carried out, but in most cases there was little or no change in the

value of B, and difference Fourier syntheses calculated rarely revealed

any major under- or over-estimates of B (123).

3.1.5   Patterson Syntheses and Refinement from Projection Data.

From difference Patterson syntheses of the h0ℓ and 0kℓ projections

it should be possible to derive all three co-ordinates x,y and z of

a heavy atom site. The results of attempts to determine and refine

heavy atom positions from such syntheses, calculated using photographic

data, are shown in Table 3.1. Although all 5 derivatives looked

promising from these results, especially the gold and $Hg(NH_3)_2Cl_2$

derivatives, in retrospect these were the only two which were correctly

interpreted when compared with parameters obtained later from three dimensional data sets. For the gold derivative, the same main site was deduced from both data sets, but the minor site was not correctly identified in the projection Patterson map. The two sites determined for $Hg(NH_3)_2Cl_2$ were approximately the same for both data sets, but the interpretation was assisted by the fact that the relationship between the two protein molecules in the crystallographic asymmetric unit was then known and that the mercury had probably replaced the copper (see section 3.11). For all other derivatives, the sites obtained from these projection Patterson syntheses were different from those determined from full three dimensional data sets. Of these both the platinum $(K_2Pt(CN)_4)$ and ethyl mercury chloride (EtHgCl) derivatives proved difficult to interpret satisfactorily in three dimensions, the first because of probable non-isomorphism and the second because of inadequate substitution. The uranyl acetate derivative, however, was successfully interpreted from a three dimensional difference Patterson synthesis, with 4 sites, 2 major and 2 minor, and was subsequently used in the final phase calculation. These results suggest that projection difference Patterson syntheses are mainly of use in the interpretation of heavy atom sites in derivatives which have one, or at the most two, well-occupied sites of substitution. This is because the overlap of both self and cross vectors, arising from a number of sites, makes interpretation virtually impossible in projection.

3.1.6 Patterson Syntheses and Refinement in Three Dimensions

For each heavy atom derivative, two difference Patterson maps were calculated, a conventional difference Patterson using coefficients $(F_{PH} - F_P)^2$, and an HLE difference Patterson, based on the combination of isomorphous and anomalous differences. For the latter, coefficients were calculated as $\quad F_{HLE}^2 = \Delta_{iso}^2 + (k\Delta_{ano}/2)^2 \quad$ (see section 3.1.1)

Table 3.1   Heavy Atom Parameters from Projection Difference Patterson
Syntheses.

Derivative	Occupancy[a]	x/a, y/b, z/c	B(Å^2)	R		no. of reflections	
				h0ℓ	0kℓ	h0ℓ	0kℓ
$KAu(CN)_2$	0.289	0.108, 0.090, 0.054	25.0	0.543	0.616	153	161
	0.120	0.473, 0.151, 0.102	25.0				
$K_2Pt(CN)_4$	0.167	0.063, 0.106, 0.182	25.0	0.634	0.671	158	169
	0.078	0.050, 0.203, 0.102	25.0				
uranyl acetate	0.150	0.120, 0.155, 0.185	25.0	0.661	0.755	154	154
EtHgCl	0.026	0.077, 0.100, 0.128	25.0	0.594		167	188
	0.105	0.240, 0.150, 0.097	25.0				
	0.098	0.070, 0.238, 0.276	25.0				
	0.053	0.238, 0.230, 0.156	25.0				
$Hg(NH_3)_2Cl_2$		0.161, 0.020, 0.121	25.0	0.581	0.591	174	164
		0.021, 0.159, 0.126	25.0				

(a)   arbitrary units

In most cases, difference Pattersons calculated with HLE coefficients had
fewer peaks (i.e. were less noisy), making them easier to interpret, as the
peaks corresponding to the heavy atom vectors were not so confused by other
features.   This was because although isomorphous differences
$F_{PH} - F_P$ are larger than those arising from the anomalous scattering
$F_{PH}^+ - F_{PH}^-$, and can thus be measured with greater precision, they are
influenced by lack of isomorphism, by scaling errors, and by inadequacies
such as those arising from absorption effects in the different crystals used

for measurement of F_{PH} and F_P. Anomalous differences, however, although often of the same order as their standard deviations, are affected by fewer systematic errors and enhance the reliability of the F_{HLE} values.

The KAu(CN)$_2$ Derivative

This derivative was initially interpreted in terms of one major site. The six highest non-origin peaks in the three dimensional difference Patterson maps correspond to the self vectors of the same major heavy atom site deduced originally from the projection Patterson maps, at (0.146, 0.137, 0.054). Figure 3.2 shows the Harker sections 0vw, u0w and uv½ of the difference Patterson map for the gold derivative. The identified vectors are indicated (for details see legend). The site was refined by least squares using the 694 centric reflections to 2.9Å resolution, and subsequent cycles of refinement, followed by double difference Fourier syntheses, led to the identification of two minor sites. Table 3.2 shows the improvement in R, as the heavy atom model was improved.

Table 3.2   Reliability Factors During Heavy Atom Refinement

of the KAu(CN)$_2$ Derivative.

No of sites	Resolution	No. reflns.	occ$^{(a)}$	x/a	y/b	z/c	B$^{(b)}$ (Å^2)	R
1	2.9Å	694	0.410	0.145	0.132	0.053	25.0	0.570
2	"	"	0.105	0.128	0.154	0.197	25.0	0.552
3	"	"	0.057	0.382	0.411	0.501	25.0	0.545$^{(c)}$

(a) occupancy 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 $B = 8\pi^2 \overline{u^2}$

and $\overline{u^2}$ is the mean square amplitude of vibration.

(c) when allowance was made for "cross-over" terms, the final R factor

was 0.510
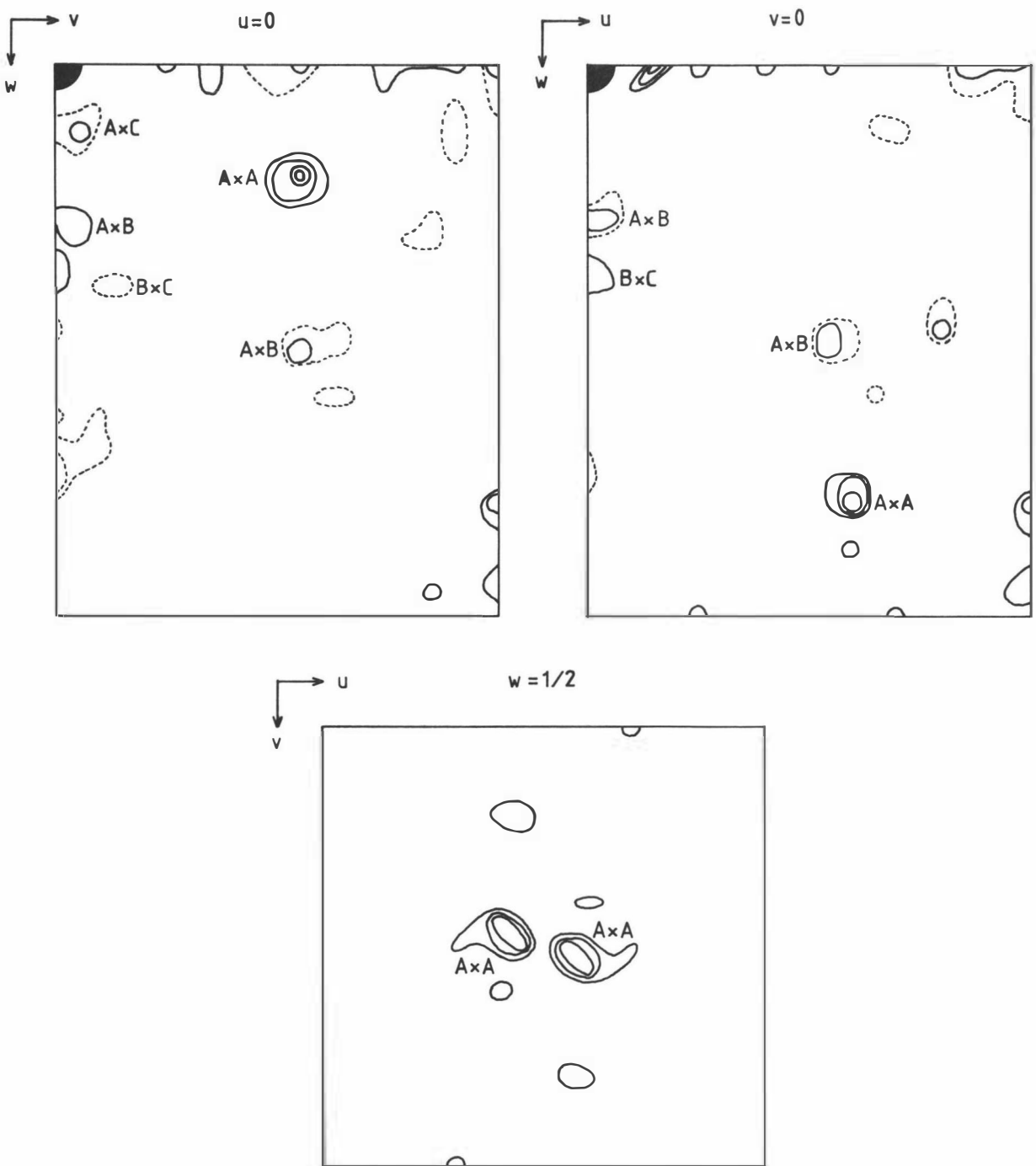
Figure 3-2. Harker sections of the difference Patterson map for the gold derivative. The three self vectors of the major site (A) are shown with some of the cross vectors between this site and the two minor sites (B and C).

Positions determined from double difference Fourier syntheses were always cross checked with the difference Patterson maps by looking for cross vectors between the proposed site and the one major site. Only those for which such vectors could be found were accepted as minor sites. Fifteen of the twenty-two highest peaks in the Patterson map could be explained by this 3-site model, with cross vectors appearing as peaks smaller than the self vectors of site 1, and the self vectors of the minor sites hardly above the general background level of the map.

Because there are 2 molecules in the asymmetric unit, and because the diffraction pattern at low resolution showed evidence of tetragonal pseudo-symmetry, it was assumed there must be some non-crystallographic symmetry relationship between them. If heavy atoms are substituted on each protein molecule at identical sites, then they too should be related by such symmetry. Examination of the heavy atom positions showed that sites 1 and 2 were in fact related by an approximate 2-fold axis along the diagonal between x and y, at $z = 1/8$, although the occupancies of these sites were uneven (one approximately 4 times the other (Table 3.2)). Similarly uneven occupancies have been noted in other cases where heavy atom sites are related by non-crystallographic symmetry (146). Assuming such a local symmetry, site 3 should be related to a fourth site at (.086, .109, .261), but no peak was found at this position. When B was refined either isotropically or anisotropically for each or all of the gold sites there was no improvement in R factor.

The absolute configuration of the gold sites was found by calculating protein phases based on the gold derivative, using the combined isomorphous and anomalous differences (see Section 3.2.4), and sites with co-ordinates first x, y, z, then x, $\bar{y}$, z. (147). Phases from each of these calculations were then used to calculate difference Fourier syntheses for the uranyl derivative. Figure 3.3 shows difference Fourier sections for this derivative using the two configurations, that giving rise to the
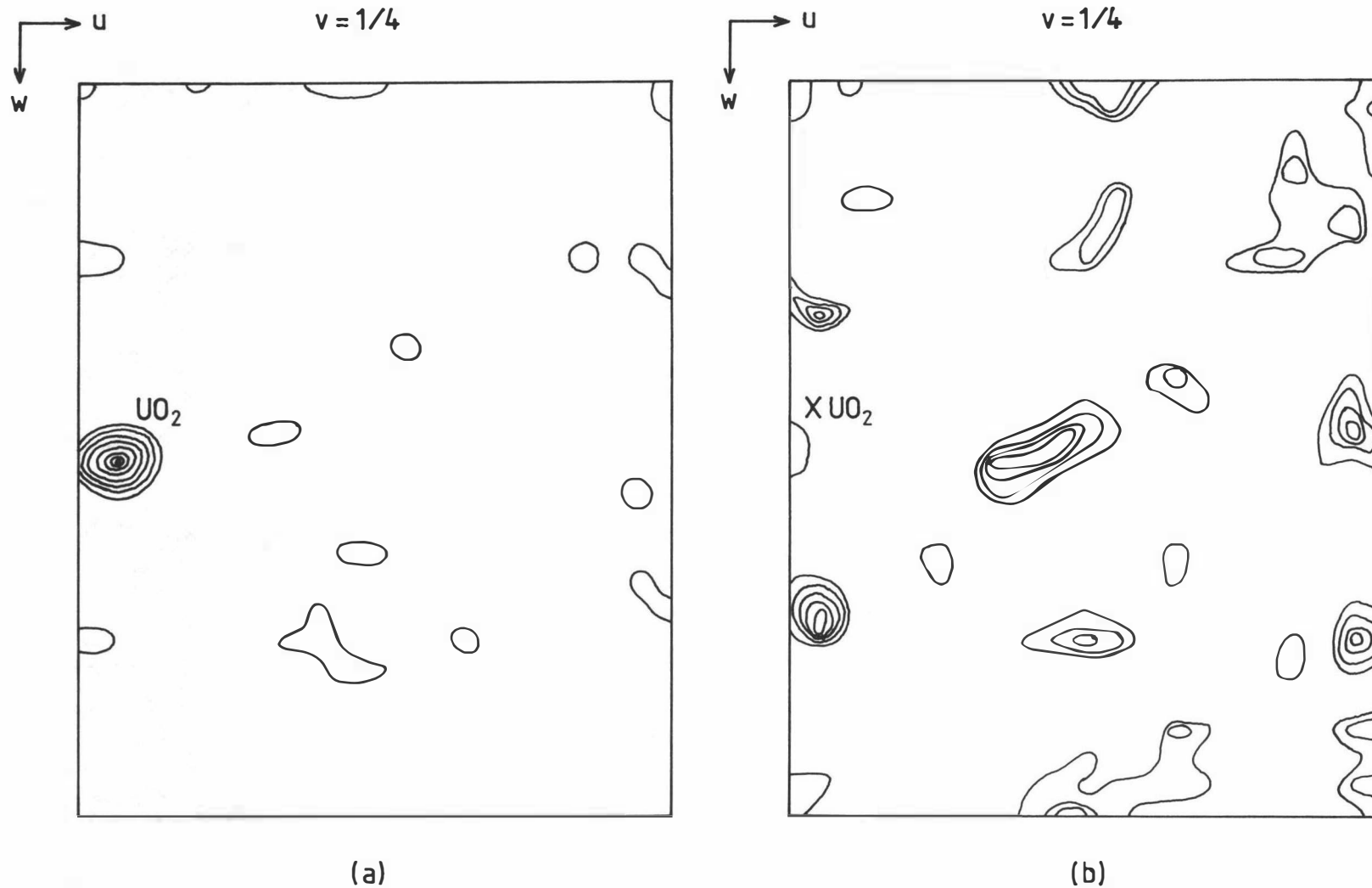
Figure 3-3. Cross-sections of two difference Fouriers ($F_{PH} - F_P$) showing one of the uranyl sites. The phases were calculated using the $AuCN_2^-$ solution for the anomalous component with : (a) the correct hand ; (b) the wrong hand.

more clearly defined peaks being taken to represent the correct hand
(in this case x, y, z).

The Uranyl Acetate Derivative.

The quality of the difference Patterson map calculated with
coefficients $(F_{PH} - F_P)^2$ for this derivative was obviously not as good
as for the gold derivative, probably because of the smaller isomorphous
differences (R_I = .095 , see Table 2.8). Although the difference
Patterson map calculated with F^2_{HLE} coefficients was an improvement,
it was not immediately interpretable, and a difference Fourier was
therefore calculated using phases derived from the combined anomalous
and isomorphous differences of the gold derivative. This map revealed
two major sites with approximately equal occupancies, that seemed to
be related by the same non-crystallographic symmetry that related
the two main sites in the $KAu(CN)_2$ derivative, viz. an approximate
2-fold axis along the diagonal between the x and y axes, at z = 1/8.
They were refined, and a subsequent double difference Fourier synthesis
calculated to check for minor sites. Table 3.3 is a summary of the
refinement as the heavy atom model improved.

Table 3.3   Reliability Factors During Heavy Atom Refinement of the
             Uranyl Acetate Derivative.

No of sites	R	resolution	No. of reflns	occ[a]	x/a	y/b	z/c	B($\mathring{A}^2$)
				0.113	0.023	0.264	0.262	25.0
2	0.585	3.1$\mathring{A}$	563	0.138	0.261	-0.001	0.007	25.0
4	0.560[b]	3.1$\mathring{A}$	563	0.029	0.478	0.272	0.132	25.0
				0.039	0.278	0.480	0.122	25.0

a   occupancy in arbitrary units.

b   when cross-overs were taken into account, the R factor was 0.558

When these positions were compared with both the F^2_{HLE} and $(F_{PH} - F_P)^2$

Patterson syntheses, it was seen that,for the two major sites, self vectors occurred on the Harker sections 0,v,w, u,0,w and u,v,½, and that the cross vectors between them fell on areas of very high density in the maps. Although the self vectors of the minor sites could not be identified above the background level of the maps, they were related by the local two-fold axis, and cross vectors between them and the two main sites were compatible with the difference Patterson maps. Figure 3.4 shows some of these sections from the F^2_{HLE} Patterson map, the peaks representing self and cross vectors being indicated. It was noted that site 2 (.261, -.001, .007) was very close to the crystallographic 2-fold axis parallel to x at y = 0, z = 0, so that a crystallographically-related site would be at (.261, .001, -.007),only 1.4Å away. Double difference maps showed residual density around site 2 and various attempts were made to improve the description of the site. These included: splitting it into 2 sites, one on the crystallographic 2-fold axis and one a short way from it; refining its temperature factor anisotropically; refinement of the x, y or z co-ordinates individually. All attempts,however, only increased the value of R.

The Double Derivative (uranyl acetate + $KAu(CN)_2$)

Where there are two successful derivatives which have heavy atoms attached to the protein molecule at a single site or a few sites, soaking the crystals in solutions containing both reagents often produces a diffraction pattern which shows strong intensity changes compared to the native, but which has a different intensity distribution to that of either derivative. Such a derivative was produced by soaking azurin crystals in solutions of $KAu(CN)_2$ + uranyl acetate. The intensity changes were very strong, but the total changes in axial lengths were slightly greater than 0.5% (Table 2.5), indicating a possible lack of isomorphism. Data was first collected to 4.7Å resolution and a difference Patterson synthesis calculated using all data,and
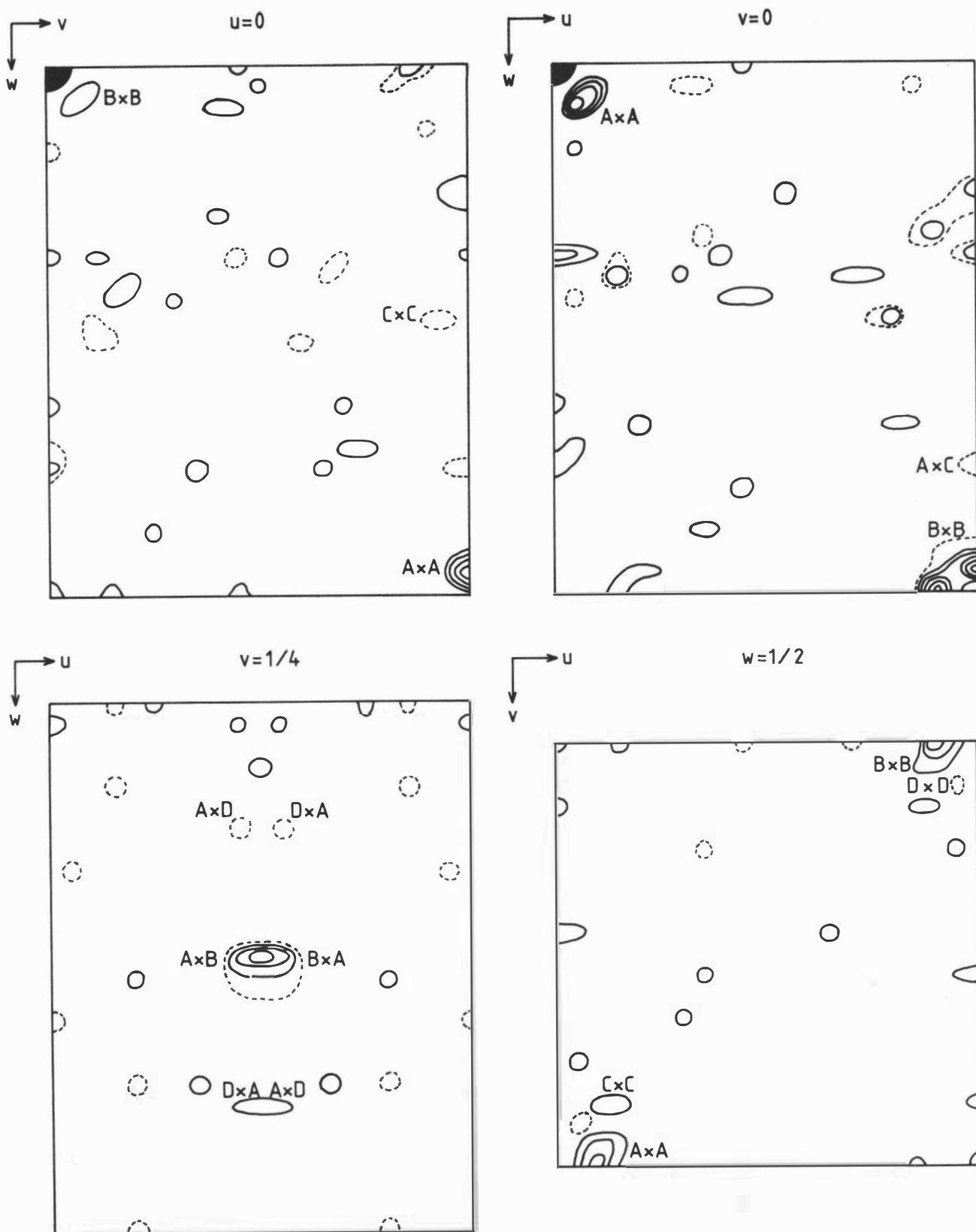
Figure 3-4. The three Harker sections (u=0, v=0, w=1/2) of the F^2_{HLE} difference Patterson map for the uranyl derivative. The three self vectors for each of the two major sites (A and B) are clearly defined. However the self vectors representing the minor sites (C and D) are not well resolved from the background and only three of the six can be seen. Cross vectors which fall on these sections and the section v=1/4 are also shown.

coefficients $(|F_{PH} - F_P|)^2$. The self vectors of the main gold site and two principal uranyl sites were immediately obvious, as were the cross vectors between the two uranyl sites, these being in the same position as in the original maps. Although there were a large number of background peaks, the self peaks for the second gold site were just detectable, and this 4-site model gave a centric R value, R = 0.585. Cross vectors between the gold and uranyl sites all fell on areas of high density. A double difference Fourier synthesis calculated with phases from these 4 positions gave a map which had many peaks of similar height. Cross checking with the F_{HLE}^2 difference Patterson map, for cross vectors between major and minor sites, gave 4 possible minor sites which, when included in refinement, gave R = 0.540. As it was not known whether they were gold or uranyl sites, they were designated X and given gold scattering factors. Their co-ordinates were:

$$X_1 \quad (.228, \quad .002, \quad .165)$$
$$X_2 \quad (.161, \quad .351, \quad -.001)$$
$$X_3 \quad (.163, \quad .370, \quad .489)$$
$$X_4 \quad (.422, \quad .392, \quad .231)$$

None of these sites are related by the non-crystallographic two-fold axis, and two of the sites, X_1 and X_4, had higher occupancies than the remaining two (0.151 and 0.120, cf. 0.076 and 0.066 respectively).

Three dimensional data to 2.9Å was then collected on a different, larger crystal, after soaking in a fresh heavy atom solution, and a difference Patterson map was calculated. A combined difference Patterson synthesis was also calculated. These both confirmed the presence of the two main gold and uranyl sites found in the low resolution data set. Cross phased difference Fourier syntheses were not calculated because the heavy atom positions used for phasing were the same as some of those in the double derivative (105,149). Thus

cycles of refinement and subsequent double difference Fourier syntheses
were used,as already described,to improve the heavy atom model, minor
sites being accepted only if they proved to be compatible with the
difference Patterson maps. Table 3.4 shows how the R factor improved
as new minor sites were introduced into the refinement.

Table 3.4   The Reliability Factor During Heavy Atom Refinement of
            the Double Derivative.

No of sites	R	Resolution	No. of reflns	occ(a)	x/a	y/b	z/c	B(Å^2)
				.441	.144	.130	.054	25.0
4	0.640	2.9Å	635	.129	.132	.141	.196	25.0
	0.530	4.7Å		.159	.027	.250	.261	25.0
				.215	.254	-.019	.011	25.0
				.074	.146	.351	.020	25.0
8	0.612	2.9Å	"	.063	.120	.077	.420	25.0
	0.483	4.7Å		.119	.423	.383	.280	25.0
				.056	.143	.358	.505	25.0
10	.594	2.9Å	"	.088	.393	.119	.245	25.0
	.458	4.7Å		.066	.345	.123	.401	25.0

(a) occupancy in arbitrary units.

These minor,previously unidentified,positions were initially
given B values of 25Å^2 as shown, and gold scattering factors.
Of the six, four were related by the local 2-fold axis, these being
sites 5 and 9, and sites 7 and 8. A cycle of least squares refinement
was carried out in which the temperature factors as well as the x, y
and z co-ordinates were varied while the occupancies were held

constant. This resulted in a slight drop in the R value to 0.591, but
the B values of some of the minor positions became negative suggesting
that the occupancies given to these atoms were too small. In the
final cycle of refinement, the temperature factors were held at their
refined values, apart from those that had become negative which were
set at $5.0\text{\AA}^2$, and the occupancies alone were refined. The final centric
R factor was 0.586, which was reduced to 0.560 when the "cross-overs"
were taken into account. All sites, together with their occupancies
and temperature factors, are shown in Table 3.5.

The table shows that the major gold and uranyl sites have very
similar co-ordinates to those found for the major sites in each of the
individual derivatives. They differ only by 0.2 and $0.9\text{\AA}$ respectively
for the two major gold positions, and by 1.0 and $1.5\text{\AA}$ respectively
for the two major uranyl sites, although the degree of substitution for
the uranyl sites in the double derivative is somewhat greater. The
minor sites found in this derivative bear no similarity to those found
in the individual derivatives, although they are consistent (with one
exception) with the sites found from the initial, low resolution double
derivative data set.

Such differences in binding can be due either to small differences
in soaking conditions, such as small changes in the concentration
of reagents or pH, or to local conformational changes that may occur
when a heavy atom binds to the protein at a specific site and indirectly
affects other regions in the protein, so that other potential binding
sites become either exposed or removed.

The $Hg(NH_3)_2Cl_2$ Derivative

Crystals soaked in solutions containing $Hg(NH_3)_2Cl_2$, at a
concentration of approximately 2 mg ml^{-1}, became gradually colourless
over a period of up to 2 months. Photographic comparison showed
changes of 0.36% in the a axial length, and 0.62% in the b axial

length, indicating possible non-isomorphism. Difference Pattersons
of the h0ℓ and 0kℓ projections obtained from photographic data
showed, however, two major sites of substitution, and the change of
colour suggested these might be the copper sites, the copper being
replaced by mercury. Accordingly it was felt that the collection of
a three dimensional data set was warranted, even if it only proved
useful at low resolution. Data was collected to 5$\overset{\circ}{A}$ resolution, and a
difference Fourier synthesis was calculated using phases calculated
from the combined isomorphous and anomalous differences of the gold
and uranyl derivatives. The main peaks of the resulting map had the
following co-ordinates.

$$(.018, \quad .158, \quad .121)$$
$$(.165, \quad .018, \quad .125)$$
$$(.190, \quad .290, \quad .357)$$

There was a suggestion that the third peak (.190, .290, .357) might
consist of two overlapping peaks, but it was initially treated as a
single peak. A difference Patterson map proved consistent with these
sites; all self and cross vectors were present and 20 out of 26 of
the highest peaks in the map were accounted for. Refinement of those
three sites gave R = 0.62, but a subsequent double difference Fourier
synthesis showed a large peak of residual density adjacent to site 3.
It was noted that this site lay almost exactly on a non-crystallo-
graphic two-fold axis, and hence could be made up of two separate
sites either side of it. When refinement was carried out, splitting
site 3 into two, the centric R factor dropped to 0.524. Subsequent
calculation of a double difference Fourier synthesis failed to reveal
any further minor sites. When cross-overs were taken into account,
the centric R factor reduced to 0.503.

The $K_2Pt(CN)_4$ derivative

The difference Patterson synthesis showed no outstanding peaks, but rather a sea of peaks of about the same height. Heavy atom positions deduced from the HLE difference Patterson, which had many of the same features, were refined using centric data, but all gave similar R factors, approximately 0.78-0.80. To help locate heavy atom sites, a difference Fourier map was calculated using phases calculated from the combined isomorphous and anomalous differences of the gold derivative. Although the highest peaks could be related to corresponding vectors in the HLE difference Patterson map, and cross vectors fell on areas of positive density, no single site refined satisfactorily. A second cross-phased difference Fourier synthesis, this time using combined phases from the uranyl and gold derivatives (mean figure of merit <m> = 0.70), showed the same peaks as the previous map, and eventually a model with 12 sites, all of which were compatible with the difference Pattersons,was refined to give R = 0.66. All sites were then examined for possible partners related by the non-crystallo-graphic symmetry axis, and this suggested a further 7 sites. Introduction of these into the refinement reduced the R value slightly, to 0.629. Because of the large number of sites and poor refinement, this derivative was viewed with mistrust (150) and was not eventually used in the phase determination. This was the only platinum compound of the many tried that did not cause the actual physical destruction of the protein crystals during soaking. The $Pt(CN)_4^{2-}$ ion is quite stable, its ligands being less likely to be replaced by those from protein side chains than those of ions like $PtCl_4^{2-}$. Anionic binding thus becomes the most likely mode of interaction with the protein (105). However, it must be assumed that the binding of the ion to the azurin crystals either occurred at multiple sites with low occupancy, or caused some non-isomorphism with the native crystals.

The Thiomersal Derivative.

Crystals placed in solutions containing thiomersal (1 mg ml^{-1})
turned colourless after soaking for approximately two weeks at room
temperature. This suggested that the copper atom had either been
reduced or replaced by the mercury. The crystals diffracted strongly,
but the photographs showed that the c axial length had increased from
99.5 to 100.1 Å. It was found that by soaking the crystals in more
dilute solutions, for shorter times, the intensity distribution of the
diffraction pattern showed changes that, while not so pronounced, were
still considerable, and were not accompanied by such a big change in
the c axial length. Using a crystal soaked under the conditions shown
in Table 2.3, a three dimensional data set was collected. However,
during data collection the crystals, which were initially light blue,
became completely colourless, and when the axial lengths were remeasured
after the data collection had been completed, it was found the c axis
had increased from 99.75 to 100.1 Å. (Not surprisingly the crystal had
to be realigned during the data collection, and some reflections had
uneven backgrounds). Despite these difficulties and the probable non-
isomorphism, it was decided to calculate a difference Patterson to low
resolution ($\sim$ 6Å), as well as a difference Fourier synthesis phased by
other derivatives, to try to determine the heavy atom binding sites.
The two main peaks in the cross phased Fourier map had co-ordinates
(-.01, .16, .095 and .17, -.01, .155). They were thus very near the
copper sites (see sections 3.3 and 3.4) and were related by the non-
crystallographic two-fold axis. No other major sites of substitution
could be identified. Vectors corresponding to the 2 sites were found
in the difference Patterson map, but refinement was unsatisfactory,
R being 0.72 for data to 3.1Å and 0.68 for data to 6Å.

The Na$_3$IrCl$_6$ Derivative.

Data to 4Å only was collected for this derivative, to provide an

initial evaluation. As can be seen from Table 2.7, the isomorphous
differences were very small (smaller than had appeared likely in
photographs). This was reflected in the difference Fourier synthesis
phased by the uranyl and gold derivatives, which showed no outstanding
peaks. Some of the peaks from this map could be related to vectors in
the difference Patterson map for this derivative, but no refinement
gave R less than 0.70. This derivative was thus not used in structure
determination.

Ethyl Mercury Chloride (EtHgCl)

When precession photographs from crystals soaked in EtHgCl were
compared with native photographs, there appeared to be significant
intensity differences between the two isomorphs. However, when a
three dimensional data set was collected from a crystal soaked under
apparently identical conditions, the intensity differences proved to
be very small, similar to those for the iridium derivative. It was
assumed that slight variations in the soaking procedure, such as the
number of crystals soaking, or the concentration of reagent, were
responsible for the different results. A cross phased difference
Fourier map (phased by the gold derivative) contained only one large
peak. This had co-ordinates very similar to those of the main gold
site, and so was treated with some scepticism. As it could be related
to peaks in both the HLE and difference Patterson maps, it was refined,
but refinement converged at R = 0.75. No minor sites could be
detected and the derivative was not used any further.

The refined heavy atom parameters for each of these derivatives
are summarized in Table 3.5.

Table 3.5   The Refined Heavy Atom Parameters

site	occ.	x/a	y/b	z/c	$B(\mathring{A}^2)$	site	occ.	x/a	y/b	z/c	$B(\mathring{A}^2)$
$\underline{KAu(CN)_2}$											
A	.410	.1448	.1318	.0532	25.0	A'	.105	.1276	.1523	.1976	25.0
B						B'	.057	.3795	.4137	.4975	25.0
Uranyl Acetate											
A	.113	.023	.264	.262	25.0	A'	.138	.261	-.001	.007	25.0
B	.029	.478	.272	.132	25.0	B'	.039	.278	.480	.122	25.0
Double Derivative $(KAu(CN)_2$ + uranyl acetate)											
A	.441	.144	.130	.0538	27.0	A'	.129	.132	.141	.196	11.0
B	.159	.027	.250	.261	19.0	B'	.215	.254	-.019	.0108	34.0
C	.074	.146	.351	.020	17.0	C'	.088	.393	.119	.245	5.0
D	.119	.423	.383	.280	6.0	D'	.063	.120	.077	.420	5.0
E	.056	.143	.358	.505	5.0						
						F'	.066	.345	.123	.401	5.0
$\underline{Hg(NH_3)_2Cl_2}$											
A	.220	.019	.159	.115	25.0	A'	.355	.172	.023	.133	25.0
B	.074	.199	.287	.354	25.0	B'	.127	.218	.282	.349	25.0

Table 3.5   (Continued)

site	occ.	x/a	y/b	z/c	$B(\overset{O}{A}{}^2)$	site	occ.	x/a	y/b	z/c	$B(\overset{O}{A}{}^2)$
Thiomersal											
A	.307	-.018	.160	.095	25.0	A'	.239	.171	-.011	.159	25.0
Na$_3$IrCl$_6$											
A	.008	.032	.225	.231	25.0	A'	.057	.211	.026	.017	25.0
B	.088	.153	.297	.118	25.0						

' denotes a site related by the non-crystallographic two-fold axis between the two molecules in the asymmetric unit.

B is the isotropic temperature factor of the site, and is of the form

$\exp[-B(\sin^2\theta)/\lambda^2]$ where $B = 8\pi^2 \overline{u}{}^2$ and $\overline{u}{}^2$ is the mean square amplitude of vibration.

occ. is the relative occupancy of the site in arbitrary units.

3.2 Methods of Phase Determination

X-ray diffraction experiments yield only the amplitudes of scattered X-rays, but not their phase. Information from the change in scattering amplitude of the total structure, caused by the incorporation of one or more atoms of high atomic number and known position into the protein molecule, can be used to obtain partial or complete information about the phase angles (151).

3.2.1 The Isomorphous Replacement Method

The native crystal will give a total scattering amplitude F_P with a resultant phase angle α_P. Additional scattering material, in known positions in the isomorphous crystal, results in a total scattering amplitude F_{PH} and phase angle α_{PH}. These amplitudes are experimentally measured while the phases are not. But, as the positions and occupancy of the sites of heavy atom substitution are known for a particular heavy atom derivative, the heavy atom scattering can be calculated from the equation

$$\underset{\sim hk\ell}{F} = \sum_{j=1}^{N} f_{j\ hk\ell} \exp 2\pi i\ (hx_j + ky_j + \ell z_j)$$

where the summation is over N atoms in the unit cell, the jth atom having fractional co-ordinates $(x_j\ y_j\ z_j)$ and f_j being proportional to the number of electrons associated with the jth atom.

The three vectors, $\underset{\sim P}{F}$ (whose magnitude only is known), $\underset{\sim PH}{F}$ (whose magnitude only is known) and $\underset{\sim H}{F}$ (whose magnitude and phase are known) are related as shown in Figure 3.5.

From Fig. 3.5

$$\underset{\sim PH}{F} = \underset{\sim P}{F} + \underset{\sim H}{F}$$
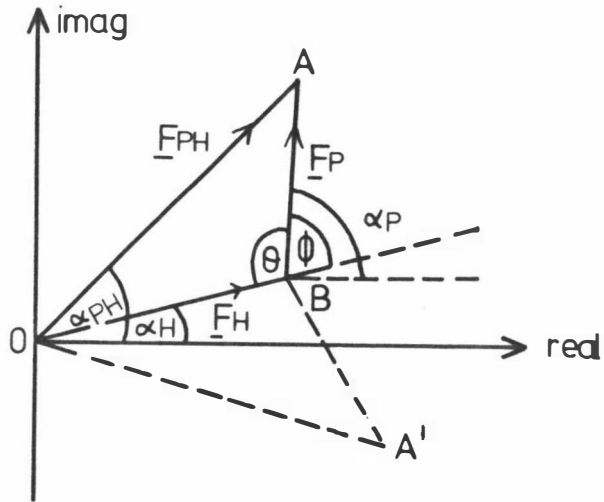
Figure 3.5   A Vector Representation of $\underset{\sim}{F}_H$, $\underset{\sim}{F}_P$ and $\underset{\sim}{F}_{PH}$. If α_H is known, the directions of F_P and F_{PH} can be deduced, giving two ambiguous solutions arranged symmetrically about $\underset{\sim}{F}_H$.

As the lengths of the three sides of the triangle are known, the cosine rule gives

$$F_{PH}^2 = F_P^2 + F_H^2 - 2F_P F_H \cos\theta$$

Therefore $\cos\theta = -\dfrac{F_{PH}^2 - F_P^2 - F_H^2}{2F_P F_H}$

But $\cos\theta = \cos(180-\phi) = -\cos\phi$

From the diagram it can be seen that $\alpha_P = \alpha_H + \phi$

As it is possible to construct another vector solution by reflecting triangle OAB about $\underset{\sim}{F}_H$ to give triangle OA'B, there are two possible

solutions for α_P,

i.e. $\alpha_P = \alpha_H \pm \phi$

which are symmetrical about α_H, the phase of the replaceable group.

3.2.2 The Method of Multiple Isomorphous Replacement

For non-centrosymmetric crystals this ambiguity can be resolved by having another isomorphous crystal, with replaceable atoms in different positions. This is the basis of the multiple isomorphous replacement method, and Harker (115) introduced a convenient graphical solution which is shown in Figure 3.6. For two heavy atom derivatives, 1 and 2, the points of intersection of the circles with radii F_{PH_1} and F_{PH_2} with the circle of radius F_P (the native protein) give possible phase angles. In the absence of all experimental errors, and assuming that the introduction of the heavy atom causes no perturbation of the protein structure, then the two derivative circles should intersect the parent circle at the same point, giving a unique solution for α, the phase of the protein.

The main disadvantage of this method is that it does not take into account the various errors involved in the analysis, for the circles rarely intersect at a single point. Errors can arise from many sources, for example from measurements in intensities, a lack of isomorphism, or incomplete or imperfect refinement. Blow and Crick (153) showed how the probable error in such phase determinations could be assessed. They noted that each angle around the parent phase circle could be assigned a probability of being the true phase angle. The probability is high at a given angle if many derivative circles intersect nearby, and low if not. If an arbitrary protein phase angle α is chosen, and the vector triangle constructed by adding the known heavy atom component, F_H, to the native protein component, F_P, with phase α (Figure 3.7), the third side of the triangle is given by
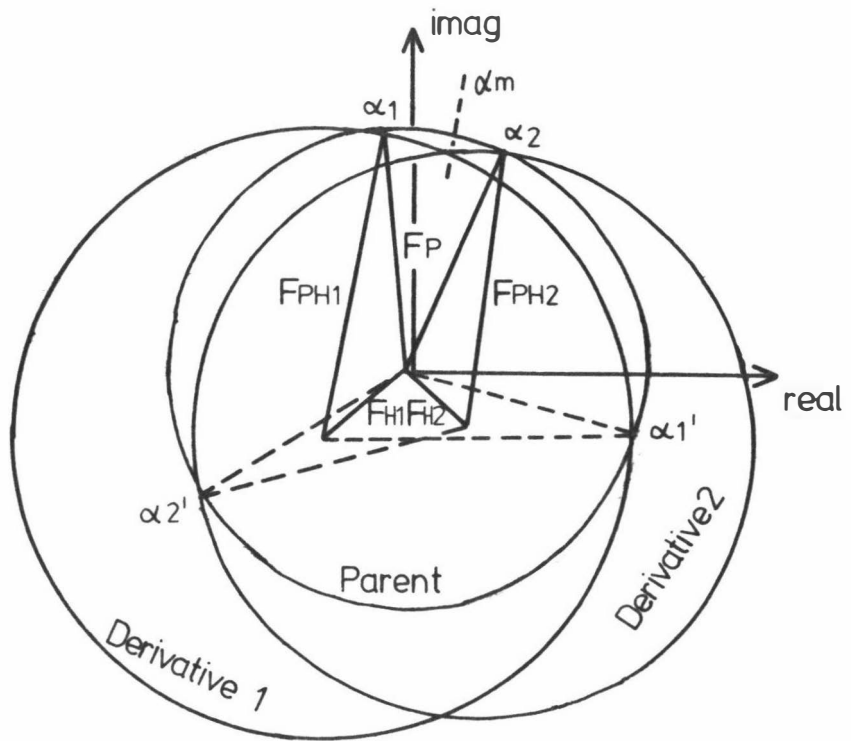
Figure 3.6   Harker Construction for a Double Isomorphous Replacement. α_m is the "most probable" phase for F_P

$$D_H(\alpha) = F_P^2 + F_H^2 + 2F_P F_H \cos(\alpha_H - \alpha)$$

In general, this will not be equal to the observed value F_{PH}, but will differ from it by a "lack of closure"

$$\varepsilon_H(\alpha) = F_{PH} - D_H(\alpha)$$

The probability of the phase angle being correct for a single isomorphous replacement is

$$P(\alpha) = N \exp[-\varepsilon_H(\alpha)^2/2E^2]$$

where N is a normalizing factor, so that the sum of all the probabilities is unity, i.e. $\int_o^{2\pi} P(\alpha) \, d\alpha = 1$.
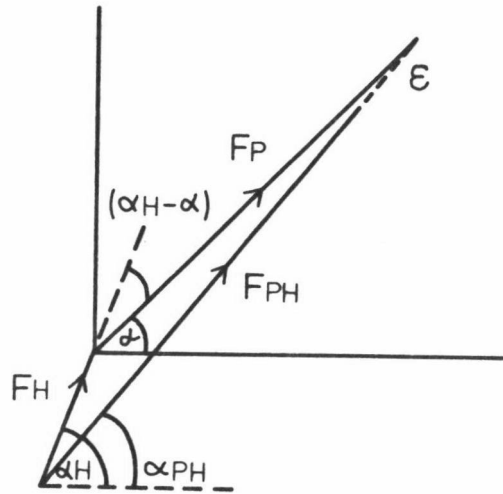
Figure 3.7   Vector Diagram Illustrating the Lack of Closure ε of an

Isomorphous Replacement Phase Triangle.

and E^2 is the mean square error in F_{PH}. E^2 can be calculated from centric data by means of the equation

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

where the summation is over n centric reflections. When a number, j, of heavy atoms are used, the total probability of a given phase angle α being correct is given by

$$P(\alpha) = N \exp\left[-\sum_i \varepsilon_j(\alpha)^2/2E_j^2\right]$$

The assumption that all errors reside in the F_{PH} values is a limitation that imposes excessive reliance on the accurate measurement of F_P. Thus care must be taken to obtain the most accurate data set possible for the native crystals.

From the above arguments it would seem the most obvious phase
to choose would be the most probable, i.e. that α_p for each
reflection should be the one with the highest value of $P(\alpha)$. Blow
and Crick showed, however, that the centroid of the probability
distribution is a much better choice, since it results in a Fourier
synthesis having the least mean square error in electron density over
the cell (153). Fig. 3.8 shows a native protein phase circle for
a given reflection. Derivative circles intersect the native circle
at points A, B, C, D, E and F, and from the values of F_H, F_{PH} and F_P for
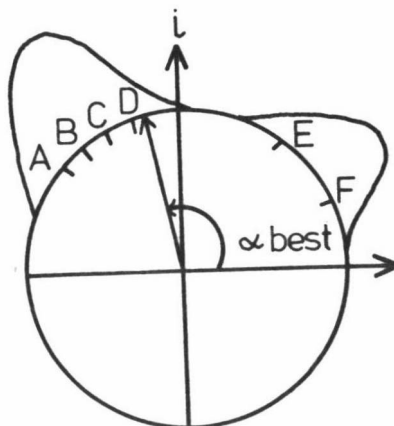
Figure 3.8   The Assessment of Errors in Isomorphous Replacement

these isomorphs, the probability that any phase angle is correct can
be calculated and plotted around the phase circle. The probability
is obviously highest near clusters of intersections, but the "best"
phase angle is determined by drawing a vector from the origin to the
centroid of the probability distribution, the vector having a
magnitude mF, where m is the figure of merit for the reflection.

Coefficients required for the "best" Fourier synthesis are thus $mF_P \exp i\alpha_B$, where the figure of merit acts to down-weight contributions from reflections with poorly determined phases. It is the mean cosine of the uncertainty in the phase angle, and it is near unity if the intersections of the circles cluster near a single point.

$$m = \frac{\int_0^{2\pi} P(\alpha) \exp (i\alpha) \, d\alpha}{\int_0^{2\pi} P(\alpha) \, d\alpha.}$$

This means that it is a measure of the precision of a phase determination rather than its accuracy, and depends on the values of E_j used in the calculations (154).

$$m \cos \alpha_B = \sum_i P(\alpha_i) \cos \alpha i / \sum P(\alpha_i)$$

$$m \sin \alpha_B = \sum_i P(\alpha_i) \sin \alpha i / \sum P(\alpha_i)$$

3.2.3 The Use of Anomalous Scattering Information.

The atomic scattering factor of an atom can be expressed as

$$f = f_o + \Delta f' + i \, \Delta f''$$

where f_o is the normal scattering factor, far from the absorption edge, and $\Delta f'$ and $\Delta f''$ are correction terms which arise from dispersion effects. $\Delta f'$ usually corresponds to a small reduction in the normal in-phase scattering, and $\Delta f''$, normally negative, corresponds to scattering with phases $\pi/2$ ahead of the normal scattering. The way in which anomalous scattering data are used to obtain information about the phase of F_P is illustrated in Figure 3.9. Given the heavy atom parameters and the real and anomalous contributions to the heavy atom scattering, F_H and F_H'', circles of radii equal to the observed amplitudes F_{PH}^+ and F_{PH}^- are drawn with respective centres marked + and -.
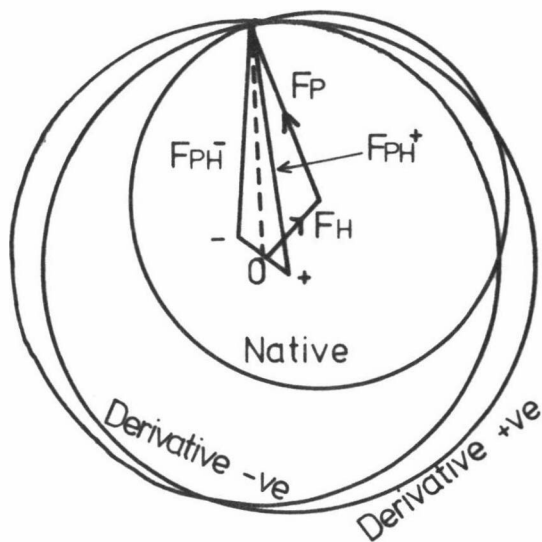
Figure 3.9   Harker Construction for a Single Isomorphous Replacement

with Anomalous Scattering in the Absence of Errors.

In the absence of all errors, the intersection of these circles will

give 2 alternative solutions for the phase angle for F_{PH}. When the

phase circle for the native protein is added to the diagram, F_P is

simply the vector from O to the intersection of F_{PH}^{+} and F_{PH}^{-}, and

the point where the three circles intersect therefore gives the phase

angle, α, for the protein. In practice, the circles usually do not

coincide because of experimental errors and non-isomorphism. Figure

3.10 shows a vector diagram illustrating the treatment of errors in

this method. To satisfy the anomalous scattering measurements, F_{PH}^{-}

and F_{PH}^{+} must meet at a point, i.e. $(\varepsilon^{+} - \varepsilon^{-}) = 0$, where ε^{+} and ε^{-}

represent the lack of closure for F^{+} and F^{-}. Thus the overall phase
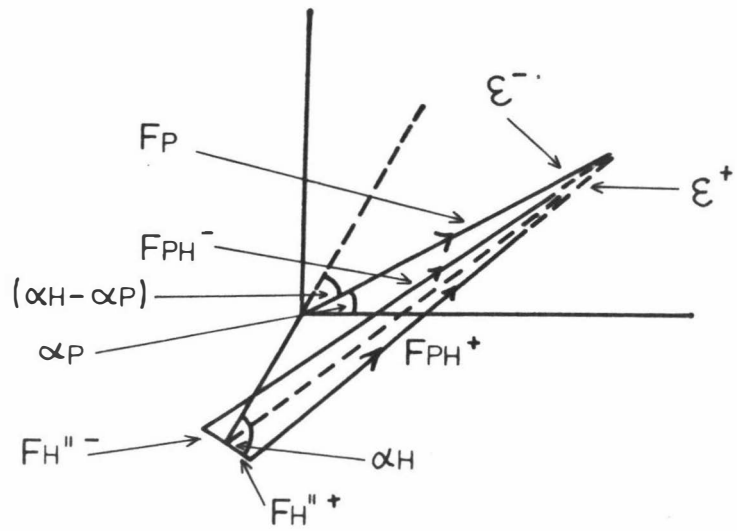
Figure 3.10   A Vector Diagram Illustrating the Lack of

Closure in the Anomalous Scattering Method.

probability distribution for one isomorphous derivative, including the anomalous scattering data, is given by

$$P(\alpha) = N \exp[-\{(\varepsilon^+)^2 + (\varepsilon^-)^2\}/2E^2]$$

where E is the total r.m.s. error in determining ε^+ and ε^-. Because this equation does not take into account the greater accuracy of anomalous differences, (147,152,155) the lack of closure terms can be rewritten in the form

$$\varepsilon^{+2} + \varepsilon^{-2} = \tfrac{1}{2}[(\varepsilon^+ + \varepsilon^-)^2 + (\varepsilon^+ - \varepsilon^-)^2]$$

In this expression, the total isomorphous lack of closure

$$\varepsilon = (\varepsilon^+ + \varepsilon^-) = 2(F_{PHcalc} - F_{PH}) \quad \text{since} \quad F_{PH} = [(F_{PH}^+ + F_{PH}^-)/2]$$

The anomalous scattering lack of closure

$$\varepsilon' = (\varepsilon^+ - \varepsilon^-) = -F_{PH}^+ + F_{PH}^- - \frac{2F_P F_H''}{F_{PH}} \sin(\alpha_{PH} - \alpha_H)$$

Thus the probability distribution for the phase of F_P, derived from anomalous scattering measurements from crystals containing heavy atoms of only one kind, is given by

$$P_{anom}(\alpha) = \exp[-(\varepsilon^+ - \varepsilon^-)^2/2E'^2]$$

where E' is the root mean square error in the anomalous scattering measurements. Matthews (147) has shown that this analysis can be extended without difficulty to treat derivative crystals containing anomalously scattering heavy atoms of more than one kind. E' was generally estimated from agreement between equivalent reflections, and was usually found to be about 1/3 E, confirming the greater accuracy of the differences.

The combined expression is thus

$$P(\alpha) = N \exp\left[-\frac{\sum\limits_j \varepsilon_j(\alpha)}{2 E_j^2} - \frac{\sum\limits_i (\varepsilon^+ - \varepsilon^-)^2}{2 E_i^2}\right]$$

where j is the total number of heavy atoms used in the phase calculation, and i the number for which anomalous differences are to be considered.

3.2.4   The Determination of the Absolute Configuration of the Molecule.

Before using anomalous scattering data for phase determination, it is essential to have the co-ordinates of the heavy atom substituents in their correct absolute configuration (139,156), since the use of the wrong hand will lead to nonsensical phase angles. That is, as well as the ambiguity in phase (corresponding to vectors $F_H \exp i\alpha_H$ or $F_H \exp{-i\alpha_H}$ in Figure 3.5),   there is also an ambiguity arising from the two possible absolute configurations. This can be expressed analytically with the following equations. From Figure 3.5, using the cosine rule

$$F_P^2 = F_H^2 + F_{PH}^2 - 2F_H F_{PH} \cos(\alpha_H - \alpha_{PH})$$

thus $$\cos(\alpha_{PH} - \alpha_H) = \frac{F_P^2 - F_{PH}^2 - F_H^2}{2\ F_{PH}F_P}$$

Taking $\cos(\alpha_{PH} - \alpha_H) = \cos\gamma$ (say) then,

$$\alpha_{PH} = \alpha_H + \gamma \quad \text{or} \quad \alpha_H - \gamma \quad \text{corresponding to the phase ambiguity.}$$

and $$\alpha_{PH} = -\alpha_H + \gamma \quad \text{or} \quad -\alpha_H - \gamma \quad \text{corresponding to the configurational}$$

ambiguity. The phase ambiguity is overcome by using more than one derivative,as discussed in Section 3.2.2 . The problem of configurational ambiguity may be overcome by consideration of the anomalous scattering information. Hence, when anomalous scattering information is incorporated into the equation, from equation (3-vi) and Figure 3.1.

$$\sin\beta = \frac{(F_{PH}^+ - F_{PH}^-)}{2\ F_H''}$$

or $$\sin(\alpha_{PH} - \alpha_H) = \frac{(F_{PH}^+ + F_{PH}^-)}{2F_H''}$$

where the absolute sign of $F_{PH}{}^{+}$ and $F_{PH}{}^{-}$ is fixed by the choice of the right-handed reciprocal lattice.

Taking $\quad \sin(\alpha_{PH} - \alpha_{H}) = \sin \gamma'$, then

$$\alpha_{PH} = \alpha_{H} + \gamma' \quad \text{or} \quad \alpha_{H} + 180 - \gamma' \quad \text{or}$$

$$- \alpha_{H} + \gamma' \quad \text{or} - \alpha_{H} + 180 - \gamma'$$

If $\alpha_{PH} = \alpha_{H} + \gamma = \alpha_{H} + \gamma'$ is the correct solution (i.e. $\gamma = \gamma'$) then these equations will also give $-\alpha_{H} + \gamma$ as a possible solution. The second solution is totally wrong, but the phase equations allow no way of distinguishing the true solution from the false one. In this work, the ambiguity in phase was resolved by using cross-phased difference Fourier syntheses to relate all derivative sites to the same origin in the unit cell once the "hand" of one derivative had been established (105,128) (See section 3.1.6).

3.2.5 Phase Calculation.

"Best" phases for the native protein data were calculated by means of the computer program PHASE (see appendix II) using the method of Blow and Crick (153,158), with anomalous scattering measurements for each derivative (apart from the $Hg(NH_3)_2Cl_2$ derivative) also being taken into account (147,155). The probability distribution was calculated at ten degree intervals around the phase circle, and a reflection was omitted from the output file when the root mean square (r.m.s) value of ε/E at the most probable phase exceeded 3.0, (where ε = the lack of closure, and E is the r.m.s. lack of closure). This ensured that reflections with improbable phase angles were excluded from any Fourier summation that might follow. The figure of merit, m, was calculated for each reflection, and as an overall check on the accuracy of the estimation of the E values, and therefore on the values of m, the overall ratio, Q, of the lack of closure to E was calculated, where

$$Q_j = < \varepsilon_j / E_j >$$

The mean was taken over all reflections, and over each contributor
(isomorphous or anomalous) to the overall phase probability
distribution. If the E values are correctly estimated, Q should
be approximately equal to unity.

The phasing power of a derivative is often expressed in terms
of the ratios F_H/E and F_H''/E' (where F_H is the r.m.s. calculated
heavy atom contribution, and F_H'' is the r.m.s. calculated anomalous
contribution). These ratios for all derivatives used in the phase
determination are shown plotted against resolution in Figure 3.11.
The decrease in phasing power of a derivative with increasing
resolution is shown by the decrease of these ratios. Over the whole
θ range, the ratio of F_H to E is larger than unity for the gold
derivative up to approximately 3.4$\overset{o}{A}$, and for both the uranyl and the
double derivative up to about 4$\overset{o}{A}$. Although the phasing power of these
derivatives might seem inadequate beyond 3.4$\overset{o}{A}$ and 4$\overset{o}{A}$ respectively,
it should be remembered that the ratios represent averages, so there
will be a proportion of the reflections at higher resolution for
which the ratio will be greater than unity, and for which significant
contributions will then be made to the phasing (159). Similarly,
values of F_H''/E' are all low. However, evidence of the value of the
anomalous contributions was shown in the case of the gold and uranyl
derivatives, where "cross phasing" of other derivatives using the
wrong "hand" produced difference Fourier maps containing spurious
peaks, and in which the peak heights of known positions were greatly
reduced (see section 3.1.6, Figure 3.3). Figure 3.12 shows a plot of
the mean figure of merit <m> against resolution for phase calculations
using each individual derivative, as well as for a double phase
determination using the gold and the uranyl derivatives, and a "combined"
phase determination using the uranyl, gold, mercury and double
derivatives. A histogram showing the number of reflections in ranges
of the figure of merit is given in Figure 3.13.
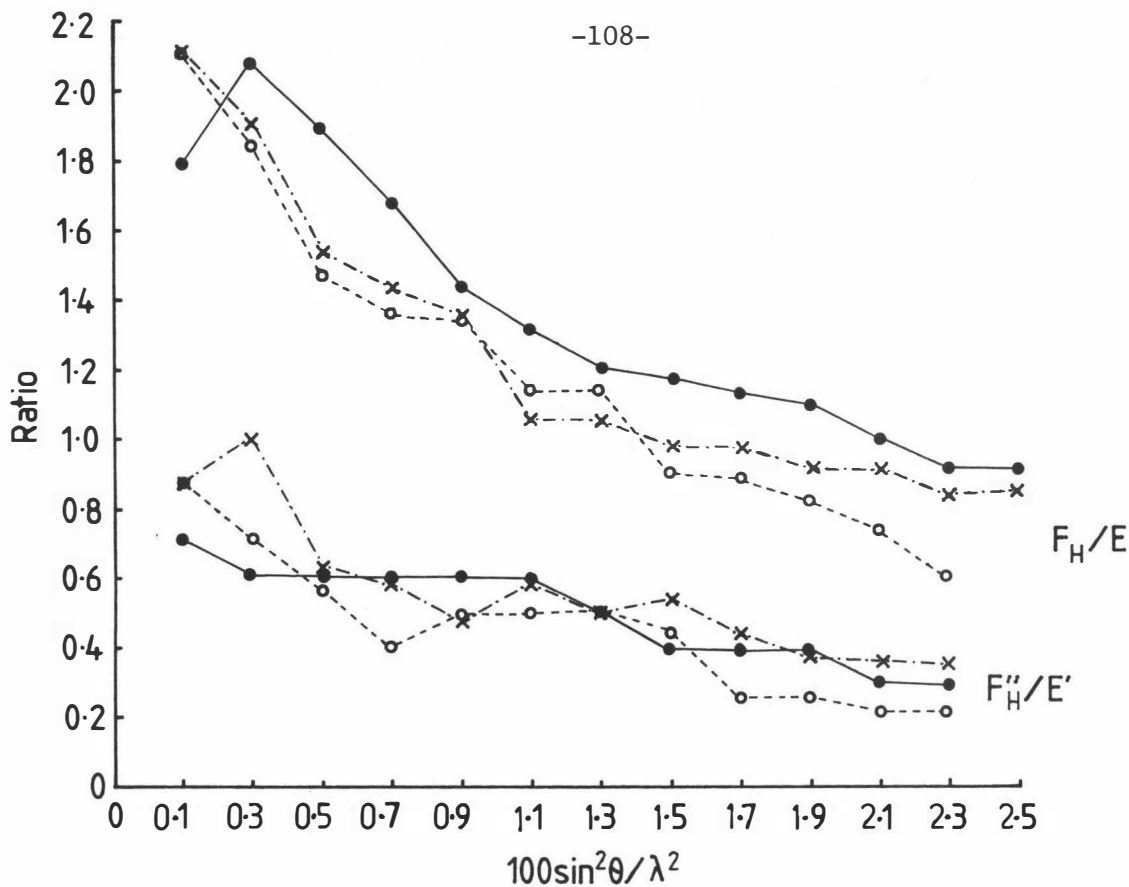
Figure 3-11. Some phasing parameters for azurin. The ratios F_H/E and F_H''/E' (see text for definition of these terms) are plotted against resolution for each derivative (• gold; ○ uranyl; × double).
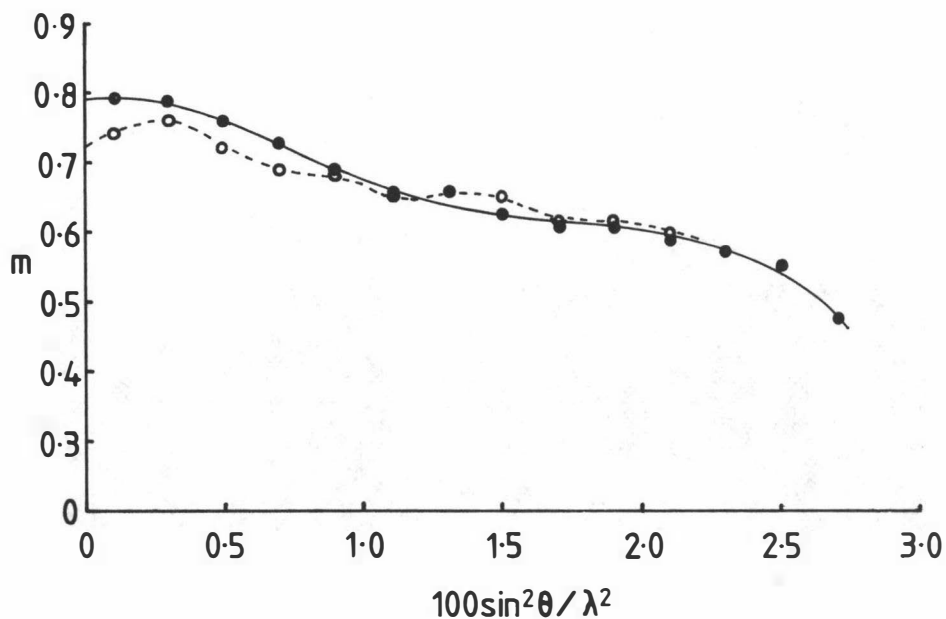
Figure 3-12. Variation of figure of merit with resolution for the phase calculations used in the structure determination.

○ Au + UO$_2$ <m> = 0·655 ; Q = 0·84

• Au + UO$_2$ + Hg + double derivative <m> = 0·614 ; Q = 0·82
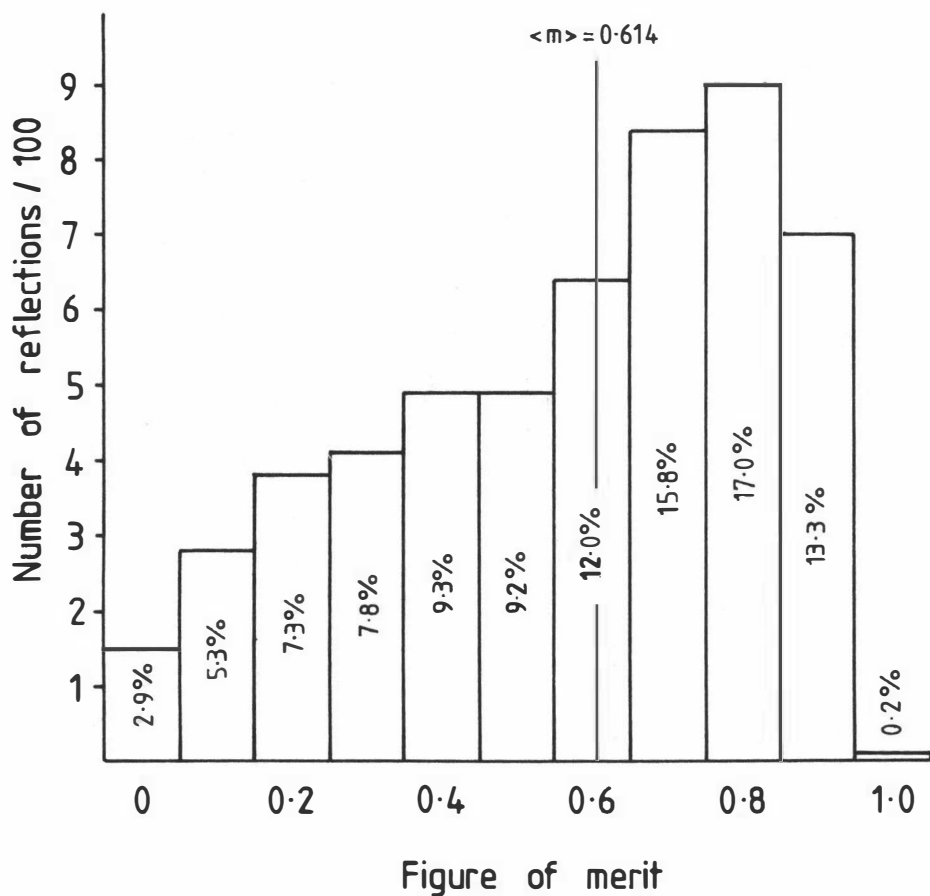
Figure 3-13. The figure of merit distribution resulting from phase calculation using four heavy atom derivatives.

3.3 The Calculation of an Electron Density Map at 6$\overset{\circ}{A}$ Resolution.

Phases were calculated using combined isomorphous and anomalous differences from the gold and uranyl derivatives. For the 759 reflections in the 6$\overset{\circ}{A}$ data set, the mean figure of merit, <m>, for the "best" phases was 0.71, with Q = 0.66. This latter value suggests that the E values were somewhat overestimated, and that <m> may therefore have been under-estimated. An electron density map at a nominal resolution of 6$\overset{\circ}{A}$ was then calculated, omitting the lowest resolution data (spacing > 20$\overset{\circ}{A}$) since this data is strongly influenced by solvent effects. The scale of the map was 4$\overset{\circ}{A}$/cm with sections calculated along the z axis at intervals of 2$\overset{\circ}{A}$. The density was sampled approximately every 2$\overset{\circ}{A}$ in the x and y directions and the map contoured by means of the computer program CONTOUR (appendix II) at arbitrary electron density levels of 200, 400, 600, 800 and 900. These contours were transferred by hand to mylar sheets which were separated by balsa spacers. The maximum density was calculated to be 0.52 e.$\overset{\circ}{A}^{-3}$, which occurred at the position (.158,.023,.119). A peak with density 0.41 e.$\overset{\circ}{A}^{-3}$ was found at (.022,.158,.123), the non-crystallographic diad-related position, and in retrospect these two positions represented the copper site in each of the molecules in the asymmetric unit. A small portion of the electron density map is shown in Figure 3.14, where the density representing these two molecules has been coloured red and yellow to allow easier identifica-tion. They are seen as essentially separate entities, the boundaries being clear along the diagonal between x and y, although their densities merge at two other points. Clear areas representing the solvent of crystallization are readily visible, these being taken as an indication that the phases were indeed reasonable (105).

A solid balsa wood model, on a scale of 2$\overset{\circ}{A}$ cm^{-1}, was also built by transferring the contours of the "yellow molecule" of Figure 3.14 to appropriate thicknesses of balsa, which were then glued together.

Figure 3.14 A portion of the electron density
map at 6 Å resolution.
The two molecules in the asymmetric
unit can clearly be seen (red and
yellow).

A photograph of the model is shown in Figure 3.15. As the density of the two different molecules in each asymmetric unit seemed to merge, as mentioned above, arbitrary cuts in the density had to be made to separate them. The non-crystallographic two-fold axis along the diagonal between x and y, at z = 1/8, and the crystallographic symmetry elements of the cell, were all taken into account when choosing these points. When the two molecules were severed adjacent to what was later shown to be the "flap" region of the "yellow" one, all these symmetry requirements were fulfilled, and the molecules were closely similar.

The resulting molecular envelope has an exaggerated pear shape of approximate dimensions 38 x 40 x 28 Å, with the area of highest density in the neck region (Figure 3.15). This was assumed to be the position of the copper atom, and comparison of this position with that of the copper in azurin from Ps. aeruginosa (82), showed them both to be in similar regions of the molecule. Another feature of the model is the so-called "flap" region, which stood out from the main barrel of the molecule, and contained a short length of α-helix (approximately 10Å) which was visible as a rod of density about 5Å in diameter.

The map was extended in the -z and +z directions so that an overall impression of the packing of the protein molecules in the unit cell could be gained. A schematic diagram of this packing is shown in Figure 3.16 for the xy projection. In these diagrams, the two molecules in each asymmetric unit are represented by shaded and plain shapes respectively. The shapes are approximate and have been chosen to display the packing to the best advantage. Each molecule of one kind, i.e. shaded or plain, appears to form close contacts with 6 molecules of the other kind, as well as with two of the same kind. Thus the unit cell seems to be arranged as clusters of protein molecules separated by relatively large areas of solvent. Solvent

Figure 3-15. Balsa model of structure at 6 Å
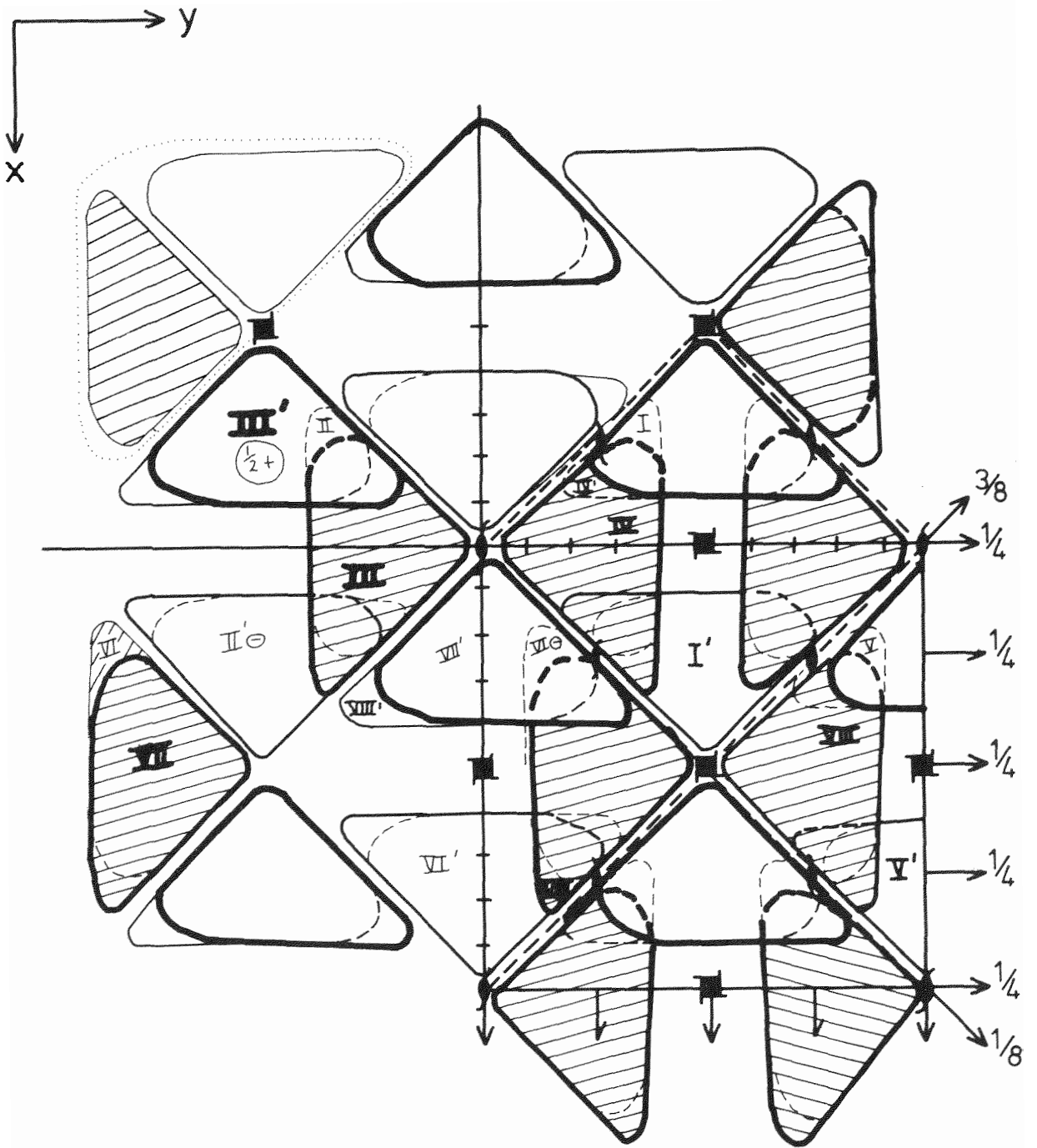
Figure 3·16 The packing of the molecules of Azurin within the crystal. The two molecules in each crytallographic asymmetric unit are cross-hatched and outlined. (........ encloses asymmetric unit)

spaces between molecules in each cluster do exist, but are not readily visible, as the diagrams represent projections of the whole unit cell.

This map was also used to check possible heavy atom positions, which, when related to the map, could usually be found near areas of density on the exterior of the molecules (although one of the minor double derivative sites seemed to lie in a solvent region between two crystallographically-related molecules).

3.4 Calculation of the 3$\overset{\circ}{A}$ map.

The continuing search for derivatives to improve the phasing of the electron density map was discouraging, as derivatives which produced changes in the intensity distribution usually seemed to produce changes in the axial lengths as well. However with one fairly good derivative ($KAu(CN)_2$), two less satisfactory derivatives (the uranyl and double derivatives), which nevertheless gave reasonable phasing to at least 4$\overset{\circ}{A}$ resolution, and a fourth derivative, $Hg(NH_3)_2Cl_2$, for which only low resolution data were available, it was decided to calculate a trial map at a nominal resolution of 3$\overset{\circ}{A}$ (the limit of the native, $KAu(CN)_2$ and double derivative data). For the 5335 reflections of the final 3$\overset{\circ}{A}$ data set, $<m>$ was 0.614, and $Q = 0.82$. This was equivalent to a weighted mean error in the phase angle of 52° (158), with 58% of the reflections having an error in phasing equal to or less than this value.

A "mini-map" on a scale of 2$\overset{\circ}{A}$ cm^{-1} was calculated for the molecule with the copper atom at (.336,.525,.371), referred to as molecule II. It was contoured as before and the contours were transferred to mylar sheets by hand. The sheets were separated by balsa wood spacers, the sections being calculated along the z direction at intervals of 1$\overset{\circ}{A}$. The maximum density was 0.68 e.$\overset{\circ}{A}^{-3}$ and the error in the electron density map, calculated by the method of Dickerson et al. (158), was

0.06 e.$\overset{\circ}{A}^{-3}$. The map was surprisingly clear, and with the help of

the amino acid sequence (32) appeared to permit a complete tracing

of the polypeptide backbone, with the identification of most protein

side chains. At one point, however, the chain was required to pass

through very weak density between what were thought to be residues 45

and 46, there also appeared to be breaks in the chain around residues

75-77, and detailed interpretation of other parts (e.g. residues

22-28) was unclear.

For comparison, a "mini-map" of the other molecule in the

asymmetric unit (with the copper at (.019,.160,.122), and referred to

as molecule I) was calculated on the same scale. The quality of the

second map seemed to be better around the copper site, but it was

found that although the tracing of the main chain was again unambiguous

for most of its course, there appeared to be breaks between residues

33 and 34, and residues 96 and 99, and a few other regions were

unclear. In general the mini-maps were sufficiently encouraging to

calculate a large scale map for detailed interpretation.

It was decided that, rather than average the density for the two

molecules in the asymmetric unit as was done, for example, for Ps.

aeruginosa azurin (82), the density for each molecule should be

interpreted in turn. While averaging the density might give a map

with a better appearance, such an "averaged" map might obscure real

differences in structure, and adversely affect regions for which the

density was clear in one molecule, but distorted in the other. Thus

an electron density map was calculated for molecule I on a scale of

2 cm $\overset{\circ}{A}^{-1}$ with sections calculated along the y direction at intervals

of 0.75$\overset{\circ}{A}$. It was sampled every 0.507$\overset{\circ}{A}$ in x and 0.480$\overset{\circ}{A}$ in z to give

a printout on the desired scale, and was contoured by hand at

intervals of 0.102 e $\overset{\circ}{A}^{-3}$ starting from a minimum value of 0.102 e $\overset{\circ}{A}^{-3}$.

The contours were transferred to acetate sheets which were hung in

a Richards optical comparator (160) using a vertical mirror (161).

3.5 Interpretation of the Electron Density map.

A skeletal model of the molecule was built by fitting Kendrew model components (Cambridge Repetition Engineers, Cambridge, England) to the electron density, starting at the residues liganded to the copper atom, which was instantly recognisable as the highest density in the map (0.68 e $\AA^{-3}$). It was found that the map could really be divided into two parts; firstly the β-barrel structure, the loop around the copper atom (residues 112-121) and the α-helix, where both main chain and sidechains were clear and easy to fit to the density; and secondly the loops between the β strands and between the flap and the β-barrel, for which the electron density was often poorly resolved and difficult to fit unequivocally. Many of the aromatic sidechains were easily identified as clearly-resolved flattened discs of electron density of the appropriate size (for example, Phe 111 and Tyr 15 in Figure 3.17). The four histidine residues were also easily identified, as were most of the leucine and isoleucine sidechains. These acted as well defined "markers" in the model building. Sometimes contours corresponding to the lower density of 0.07 e.$\AA^{-3}$ had to be drawn (as dotted lines) on to the sheets to aid the fitting of the residues to the density. Ambiguities could usually be solved by referring to the density of Molecule II in the mini-map, since often molecule II had good density where that of molecule I was poor. No reference was made to the conformation obtained for the azurin from Ps. aeruginosa (82) or that determined for plastocyanin (56). This was because of the apparent difference in the topologies of the two proteins, one of the objectives of this work being to provide an independent verification of any similarities or differences between them. Comments on the fit to the density are in Table 3.6.

Later, a second electron density map was calculated to cover the other molecule in the asymmetric unit (molecule II), and a complete Kendrew model was built for it also. As is discussed later,

Figure 3.17   Some Sections of the Electron Density Map

Showing the Aromatic Sidechain of Phe 111

The density belonging to the sidechain of

Phe 111 is clearly seen in the centre of

the photograph as a flattened disc.
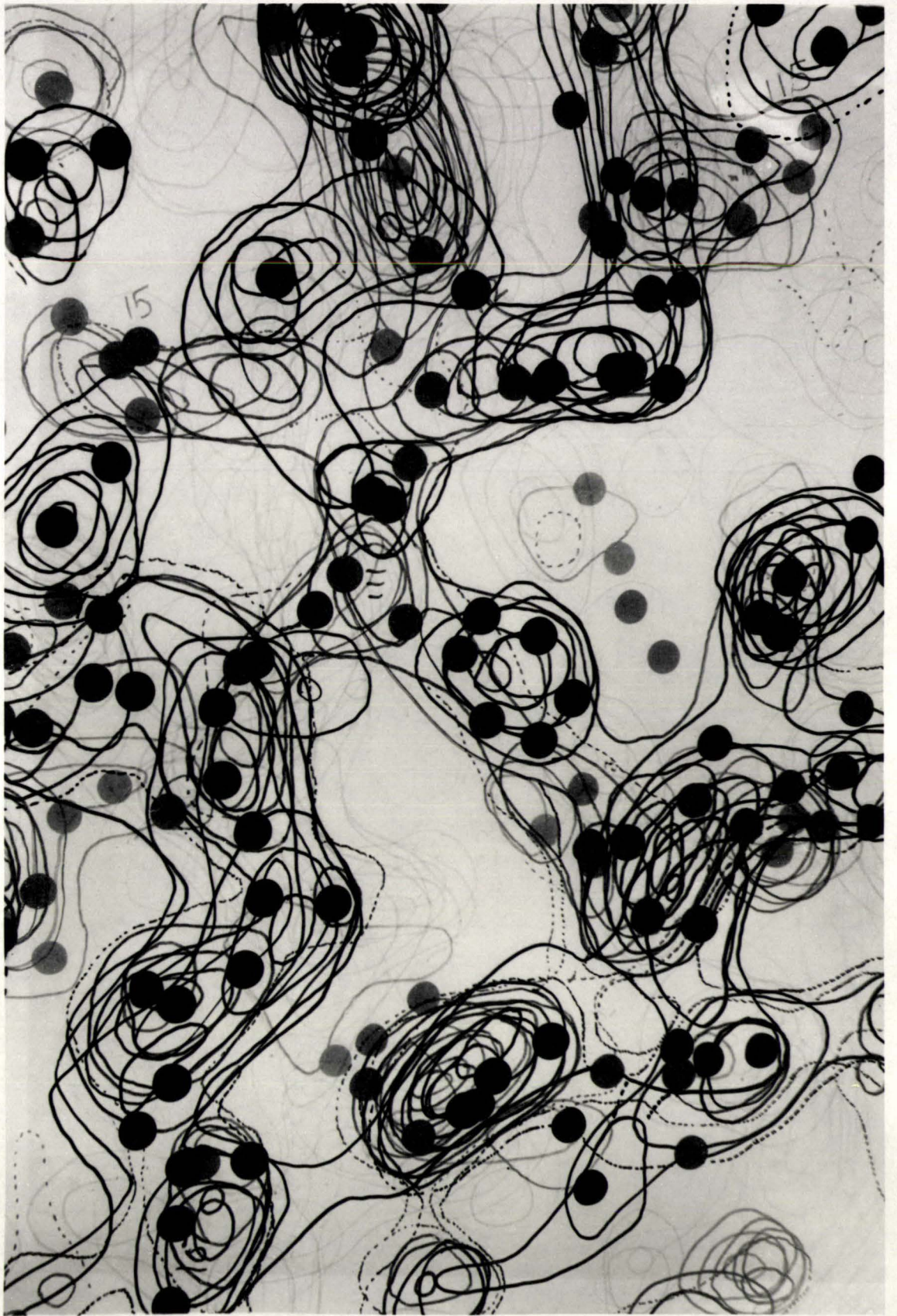
Table 3.6   A Description of the Fitting of Residues to the Electron Density in Molecules

I and II.

Sidechains with no Density	Sidechains with Weak or Discontinuous Density	Reasonable Density but Hard to Fit	Very Good Density and a Reliable Fit

MOLECULE I

Sidechains with no Density	Sidechains with Weak or Discontinuous Density	Reasonable Density but Hard to Fit	Very Good Density and a Reliable Fit
Ala 12	Lys 18	Mainchain	Mainchain
Lys 34 (beyond CB)	Glu 19 (beyond CB)	residues 23-27	residues 31-37
Lys 38 (beyond CG)	Met 20 (no density CE)	40-44	46-50
Ala 40	Lys 28	77-81	58-64
Val 73	Thr 51	99-104	82-96
Val 80	Glu 57 (beyond CB)		108-112
Asp 98	Val 59	Sidechains of	114-117
Leu 102	Gln 70 (too close to	Ala 5	122-126
Ala 109	neighbour)	Leu 17	
	Asp 77	Asp 23	Sidechains of
	Thr 78 (very poor fit)	Ala 42	
	Phe 97	Val 43	Glu 4, Glu 8
	Lys 101	Met 44	Ser 9, Asp 11
	Thr 103	Ala 54	Met 13, Tyr 15
	Trp 118	Ala 66	Phe 29, Val 31
	Lys 126	Ala 69 (very poor fit)	His 32, Leu 33
	Asn 129	Tyr 72 (ring too close to main chain)	His 35, His 46
			Trp 48, Val 49
		Ala 75	Leu 50, Asp 55
		Ile 81	Asp 62, Met 64
		Val 99	His 83, Val 86
		Ser 100	Lys 85
			Residues 88-95

Table 3.6   Continued

Sidechains with no Density	Sidechains with Weak or Discontinuous Density	Reasonable Density but Hard to Fit	Very Good Density and a Reliable Fit

MOLECULE I   Continued

| | | Glu 106 (2 possible conformations). | Tyr 110, Phe 111 Cys 112 Residues 114-117 Lys 122 |

MOLECULE   II

| Gln 2 (no density beyond CB) Gln 14 Lys 34 (no density beyond CG) Met 64 (CE) Ala 75 Ala 82 Lys 122 (beyond CB) Lys 126 (beyond CG) | Val 21 Cys 26 (very little density for CB, break between CA and SG) Phe 29 Val 36 Lys 38 Lys 56 Asn 65 (weak beyond CB) Lys 74 (break between CG and CE) Ile 81 Ala 82 Pro 104 Glu 106 Tyr 110 (no density OH) Asn 129 | Mainchain Residues   23-26 36-38 40-43 52-54 80-82 123-124 Sidechains of Asp 16 (sidechain too close to mainchain). Leu 17 Residues 40-44 Residues 51-54 Val 86 | Mainchain Residues   4-10 29-35 46-50 83-99 108-117 Sidechains of Asp 11,   Tyr 15 Lys 18,   Glu 19 Val 22,   Asp 23 Lys 24,   Val 31 His 32,   His 35 Lys 41,   His 46 Asn 47,   Trp 48 Leu 50,   Val 59 Ala 60,   Asp 62 Leu 68,   Asp 77 |

Table 3.6   Continued

Sidechains with no Density	Sidechains with Weak or Discontinuous Density	Reasonable Density but Hard to Fit	Very Good Density and a Reliable Fit

MOLECULE II   Continued

			Thr 78, Arg 79
			His 83, Lys 85
			Ile 87, Glu 91
			Residues 92–102
			Tyr 108, Ala 109
			Residues 111–121
			Leu 125, Leu 127

the two interpretations are essentially in agreement. Figure 3.18 shows the electron density for a piece of extended chain (residues 96, 97 and 98) from this molecule.

Co-ordinates of the atoms in both molecules were obtained by placing markers on to the density on the appropriate sections of the electron density map. This was done by optimizing the fit of each residue to the density, and then placing markers where atoms in the residue were reflected on to the map. Two co-ordinates were read off the sections by the use of a standard grid, while the third co-ordinate (that in the direction along which sections were taken) was measured by

(i) estimating which section the reflection fell on, using the parallax effect, and

(ii) direct measurement from the model.

Throughout, a prime consideration was that the atoms should be as well placed in the density as possible. Figures 3.19 and 3.20 show photographs of the Kendrew skeletal model of Molecule II, and a smaller Labquip model of molecule I built to a scale of 1 cm $\overset{o}{A}^{-1}$ from co-ordinates measured from the larger model.

3.6 Regularization of the Co-ordinates.

Once the co-ordinates had been recorded, they were regularized by the computer program MODFIT (appendix II) which uses an iterative least squares procedure to optimize the agreement of bond lengths and angles in a protein molecule with standard values. Before regularization, the average deviation of bond lengths and angles from standard values was $0.095\overset{o}{A}$ and 8.1^{o} respectively for Molecule II, and $0.111\overset{o}{A}$ and 9.5^{o} respectively for Molecule I. After regularization, these values were $0.01\overset{o}{A}$ and 1.5^{o} for both molecules. Table 3.7 gives the results of the regularization for both molecules.

Figure 3.18   Some Sections of the Electron Density Map

Showing Part of One of the β-strands in

Azurin, Residues 96-99.

The superposition of this part of the structure

shows how carbonyl groups fit into "bumps" in

the mainchain density. The dotted line represents

a hydrogen bond between the C=O of residue 28,

and NH of residue 99. (The density below this

belongs to the sidechain of Val 99).
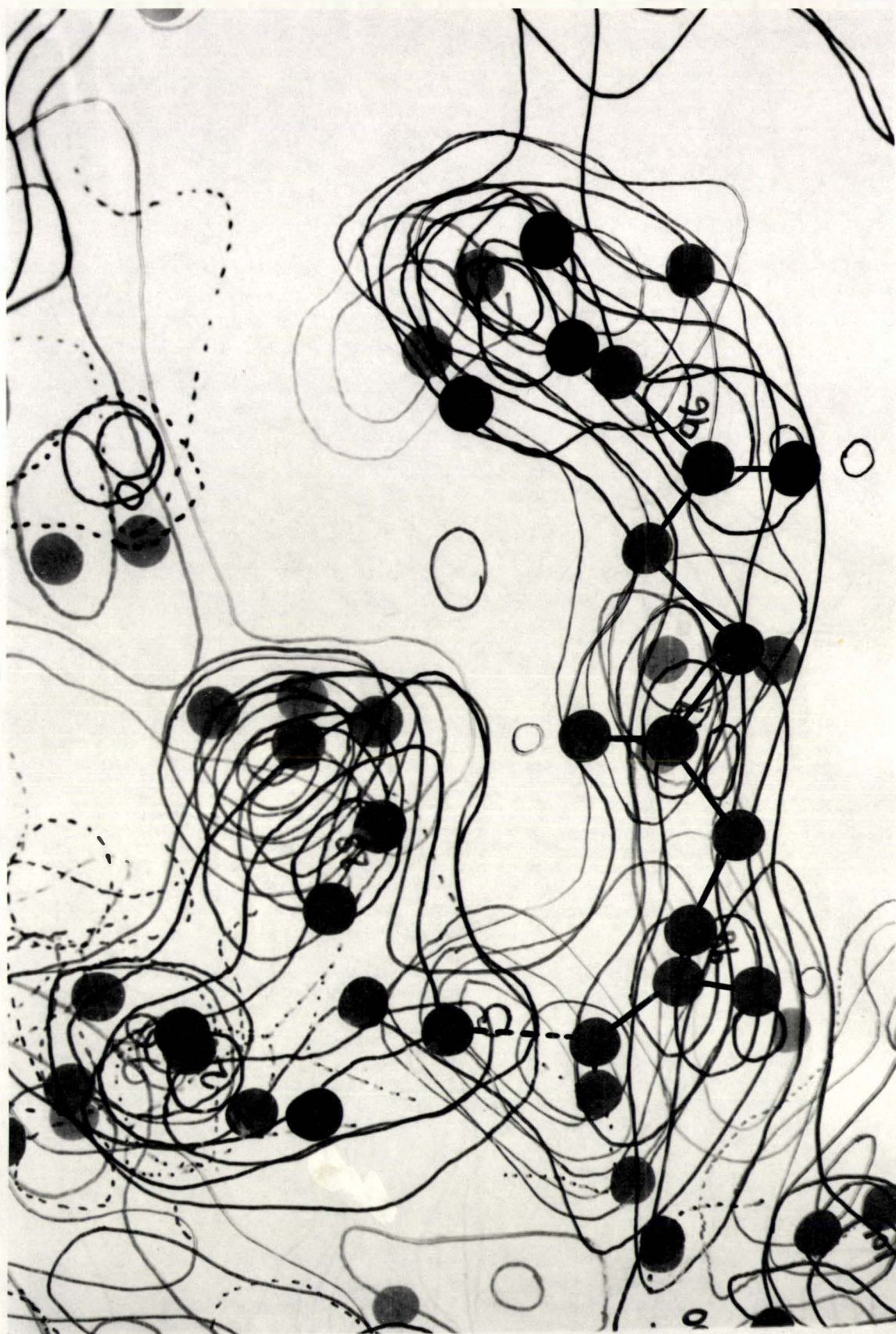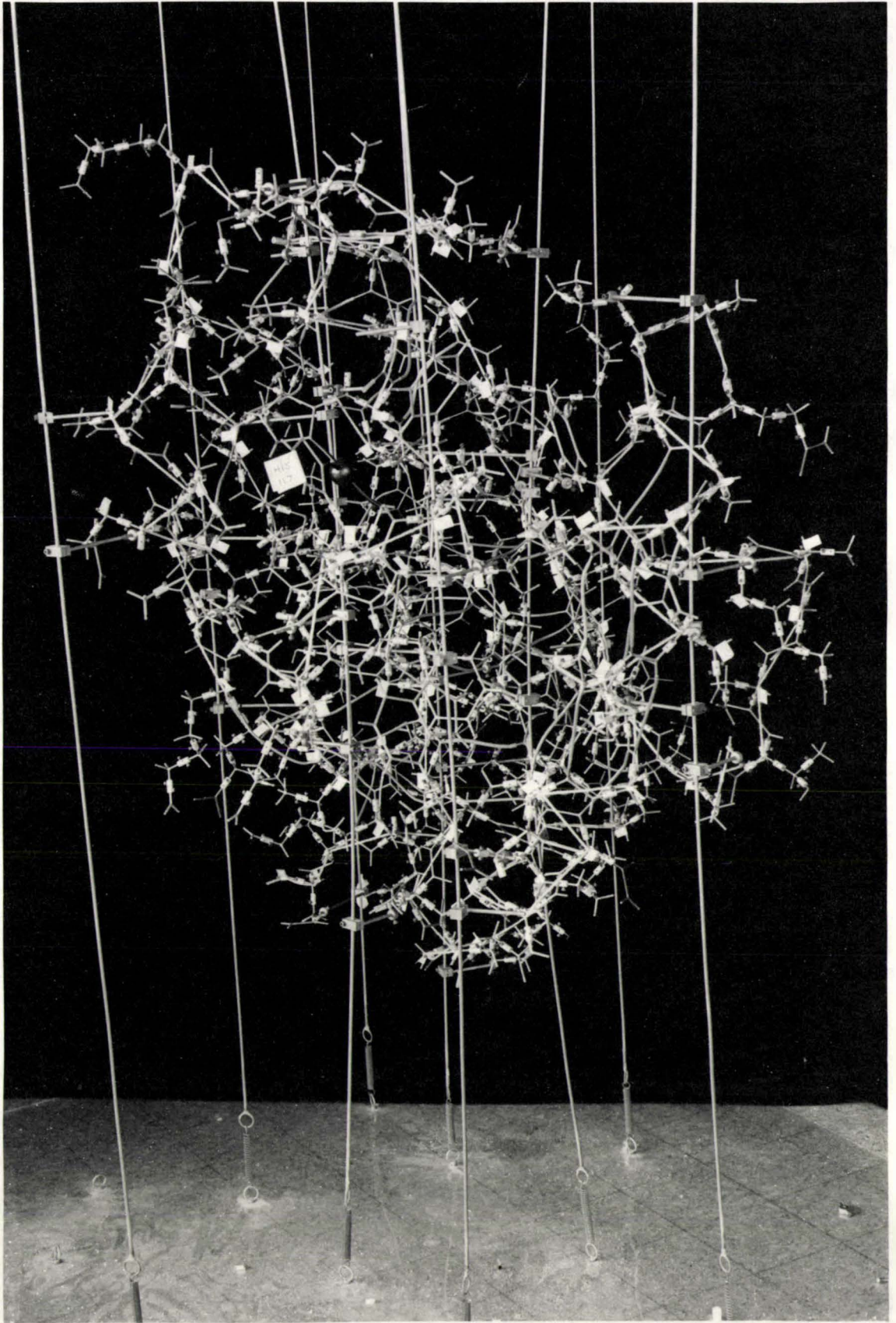
Figure 3-19. A photograph of the Kendrew model.
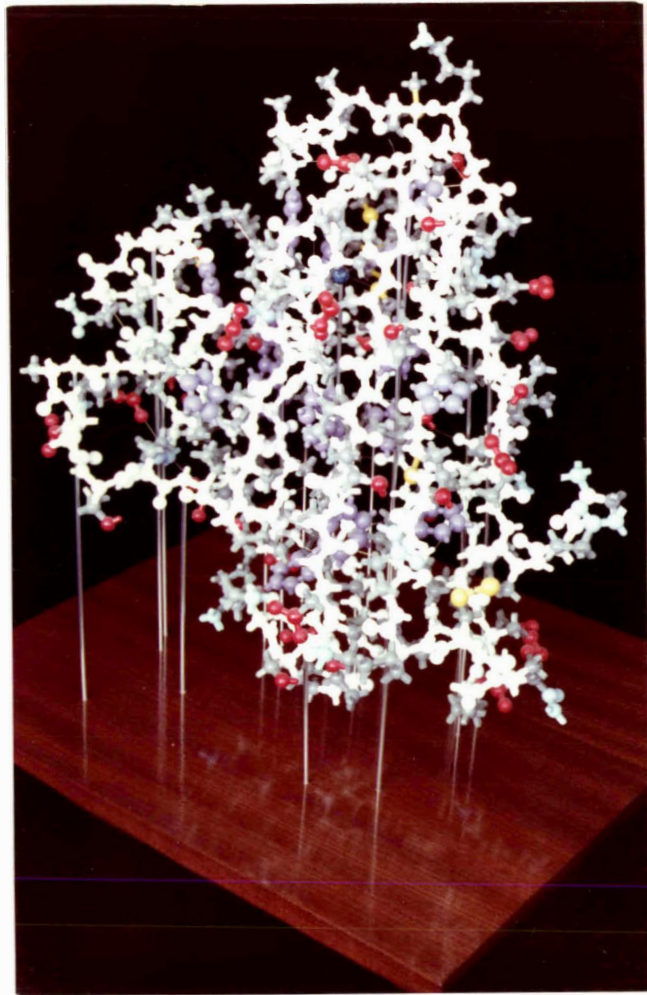
Figure 3·20 Labquip model of Azurin at 3Å
resolution.
Note the barrel shape of the
molecule and external flap on
the left-hand-side.

Table 3.7   Regularization of the Co-ordinates for Molecule I and
             Molecule II

	Molecule I				Molecule II			
	$\Delta x(\text{Å})$	$\Delta y(\text{Å})$	$\Delta z(\text{Å})$	overall	$\Delta x(\text{Å})$	$\Delta y(\text{Å})$	$\Delta z(\text{Å})$	overall
r.m.s. shift	0.125	0.114	0.123	0.209Å	0.099	0.104	0.099	0.175Å
maximum shift	0.515	0.609	0.477	0.699Å	0.585	0.480	0.454	0.633Å

3.7 Breaks in the Main Chain.

In molecule I, breaks in the main chain density occurred between
residues 29 and 30, 97 and 98, and 109 and 110, although the main
chain was unbroken in molecule II at these points. Similarly, the
density between residues 45 and 46, and 87 and 88, was very weak in
molecule II, although good in molecule I.

3.8 Differences in Interpretation.

(a) Residues 68-72

These residues constitute the major disagreement between the two
interpretations, and are shown schematically in Figure 3.21. In molecule I,
the side chain of Leu 68 occupies the density assigned to the
aromatic ring of Tyr 72 in molecule II . The side chain of Tyr 72
in Molecule I occupies (with a poor fit) density assigned to Asp 71
in molecule II. There is a general readjustment of the chain between
residues 67 and 73. Overall, the molecule II fit is better and
appears structurally more likely.
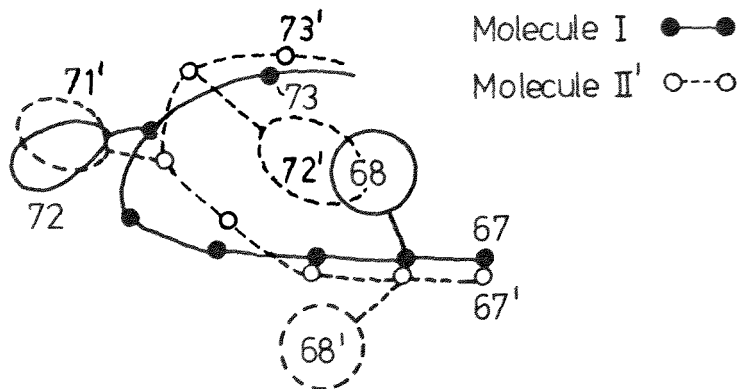
Figure 3.21   A Schematic Diagram of the Differences in Interpretation

For Residues 67-73.

(b)   Residues 77-79

The sidechain of Asp 77 in molecule II occupies part of the site

assigned to Arg 79 in molecule I (although in each case the guanidinium

group of Arg 79 is in the same place). Residue 78 is differently

placed in the two molecules and there is readjustment of the mainchain.

The molecule II interpretation appears better. The differences in

interpretation are shown schematically in Figure 3.22.

(c)   Residues 100-105

A major difference in the orientation of the sidechain of

Pro 104 causes differences between the two mainchain conformations.

Molecule I ●—●

Molecule II ○—○
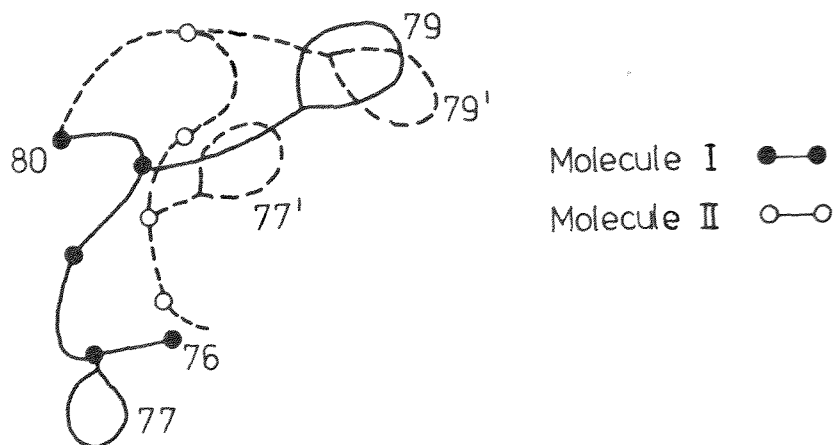
Figure 3.22   A Schematic Diagram of the Differences in Interpretation

For Residues 76-80.

3.9   Other difficult regions.

In addition to those mentioned in Table 3.6 , parts of the structure that were difficult to build included:

(a)   Residues 23-26

In molecule I, the loop between residues 23 and 26 was very difficult to build. This was not helped by the fact that the mainchain density was very weak for residues 26-27. In molecule II, residues 23 and 24 fitted well, as did 27 and 28, but the proper conformation for residues 25 and 26 was still uncertain, with no density for CB of Cys 26.   Indeed, in both molecules the disulphide bridge, Cys 3-Cys 26,was poorly resolved, although the density for the sulphur atoms was strong.

(b)   Residues 40-44

This part of the structure had strong density in both molecules, but it proved very hard to fit the model to the density. The residues form a tight loop at the "top" of the molecule. In both molecules (I and II), this loop makes close contacts with neighbouring molecules; Ala 40 with Ala 40 from a molecule related by crystallo-graphic symmetry, and Val 43 and Met 44 with residues 118-120 of a non-crystallographically related molecule. In addition, the two main gold sites are   near Val 43 (see Table 3.8) , and may cause distortions in the density in this region.

(c)   Residues 51-55

These residues are part of another loop, linking strand 4 of the β-barrel to the beginning of the α-helix in the flap region. The density for the mainchain atoms was strong in both molecules, but difficult to interpret. In molecule II, some very strong density around Glu 53 remained unaccounted for.

(d)   Residue 81

Great difficulty was experienced at Ile 81 in both molecules. This residue comes at a sharp turn linking the "flap" to strand 5 of the β-barrel. No really convincing fit could be achieved in either molecule.

3.10   The Nature of the Heavy atom Binding Sites.

Once the models for both molecules had been built, the heavy atom positions could be plotted on the electron density map and hence related to the actual protein sidechains or mainchain groups involved in binding. Table 3.8 shows the details of these sites. The occupancies listed here are absolute occupancies, calculated using the correct absolute scale factor for the native data, obtained from a structure factor calculation. If the heavy atom sites on

each molecule were exactly related by the non-crystallographic two-fold symmetry   the occupancies of each pair of sites on the two molecules in the asymmetric unit might be expected to be the same. For many sites there are, in fact, marked differences, both in occupancies and in the exactness with which the positional symmetry is obeyed. These differences may arise from local differences in sidechain orientations, due perhaps to packing effects, or even to small differences in the conformation of external groups far away from the copper site, and hence   not likely to affect the "working" parts of the molecule. It must be stressed, however, that some differences in occupancy may also be more apparent than real, arising simply from imperfect refinement of the heavy atom sites, non-isomorphism, or errors in the data.

The $KAu(CN)_2$ Derivative.

(a) For the main gold site, the degree of substitution on molecule I is much greater than on molecule II. In the former, the sidechain of Trp 118 adjoins the density of Val 43' from molecule II, to form a pocket which is further enclosed by density which appears to belong to the sidechain of Met 44'. In the equivalent site on molecule II (across the pseudo-axis), Trp 118' and Val 43 appear to be oriented differently, and as Met 44 seems to have a slightly different conformation, a "pocket" is not formed.

(b) For the minor gold site, there is substitution on molecule II but not on molecule I. Again the gold site appears to be an intermolecular pocket formed by the close approach of two type II molecules, related by the two-fold axis parallel to x at $y = \frac{1}{2}$, $z = \frac{1}{2}$, near the sidechain of Lys 38. This sidechain appears to have a different orientation in the two molecules, presumably arising from packing effects, the inexactness of the non-crystallographic two-fold axis resulting in slight differences in the crystal packing of the two molecules.

Table 3.8   The Nature of the Heavy Atom Binding Sites.

Derivative	No. of sites	Occupancy (Atoms)		Binding Site
		Mol I	Mol II	
$KAu(CN)_2$	2 major	0.5	0.13	Intermolecular pocket between Trp 118 of one molecule, and Val 43 and Met 44 of the other non-crystallographically related molecule.
	1 minor		0.07	Intermolecular pocket between two crystallographically related molecules.
Uranyl Acetate	2 major	0.15	0.17	Between the carboxyl groups of Glu 91 from two crystallographically related molecules.
	2 minor	0.03	0.05	Adjacent to the carboxyl group of Glu 106.
$Hg(NH_3)_2Cl_2$	2 major	0.44	0.71	Replace Cu.
	2 minor	0.09	0.15	Between His 32 of one molecule, and His 32' of a non-crystallographically related one.
Double Derivative	5 major	0.52	0.15	As for main gold sites.
		0.19	0.26	As for main uranyl sites.
		(0.08)	0.14	Between the carboxyl groups of Asp 77 from one molecule, and Asp 77 of a crystallographically related one.
	5 minor	0.09	0.11	Between the carboxyl groups of Asp 16 from two non-crystallographically related molecules.
		0.05		Intra-molecular pocket near Val 99 and Ser 100.
		0.08		Inter-molecular pocket, no near neighbour.
Thiomersal	2 major	0.73	0.57	Approximately 2.4Å from Cys 112 SG, and 3.4Å from the copper atom.
Na_3IrCl_6	2	0.07		Between the carbonyl oxygen and amide nitrogen of Ala 66 and the amide nitrogen of Glu 67.
			0.11	Between the amide nitrogen and carbonyl oxygen of Ala 40' and the amide nitrogen of Ala 42'.

The Uranyl Acetate Derivative.

(a) For the uranyl derivative, all sites involve sidechain carboxyl

groups, a preference which has been noted before for the UO_2^{2+}

cation (105). The two major sites,which have very similar

occupancies,are both formed by pairs of Glu 91 sidechains related

by crystallographic two-fold symmetry. A schematic diagram of the

two sites is shown below.

Molecule I
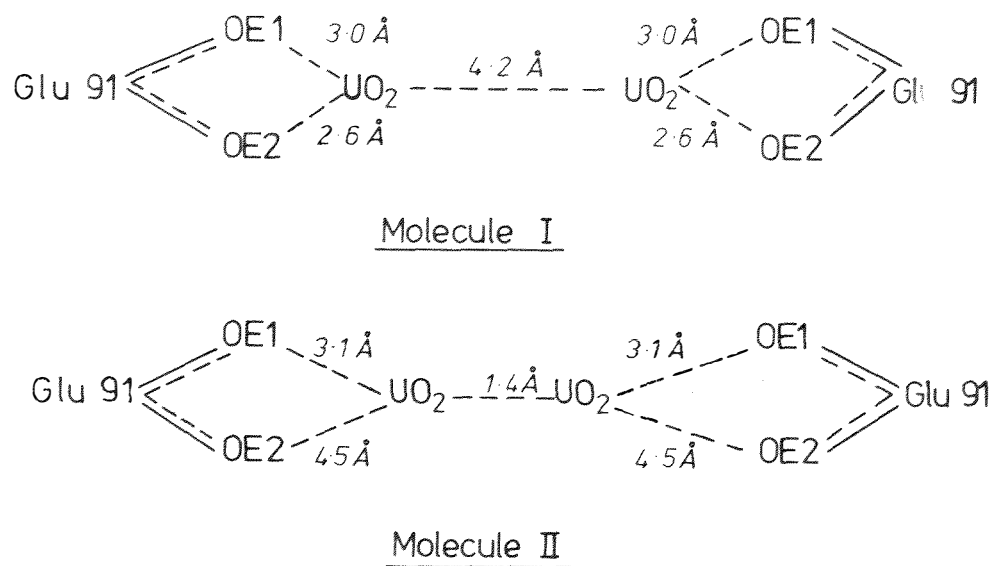
Molecule II

Figure 3.23   A Schematic Diagram Showing the Two Main Uranyl

Binding Sites on Molecule I and Molecule II.

The fact that the carboxyl groups in the two molecules appear, at

this stage, to be different distances from the diad probably arises

from the inexactness of the non-crystallographic symmetry. That is,

the carboxyl group of Glu 91 is (slightly) differently oriented

with respect to the crystallographic 2-fold axis (and hence to its

diad-related partner) in Molecule I compared with Molecule II.

(b) The two minor sites have markedly different substitution.
There appears to be a difference, however, in the orientation of the
sidechain of Glu 106 in the two molecules (in one case it is hydrogen
bonded to the hydroxyl group of Tyr 108), and since the carboxyl
group of Glu 106 is the only protein group within bonding distance
of the uranyl site, this difference may explain the uneven substitution.

The $Hg(NH_3)_2Cl_2$ Derivative.

(a) The difference in substitution of the major mercury site was
unexpected, as the fact that the crystals became colourless suggested
complete replacement of the copper atoms in both molecules. Allowing
for the fact that a mercury atom (Z = 80) has replaced a copper atom
(Z = 29), the occupancies correspond to about 71% replacement in
molecule I and 44% in molecule II. There is no obvious reason for this
discrepancy, or for the different occupancies, but possible causes
may be the non-isomorphism of the derivative, producing misleading
results in the refinement, or a difference in the amount or rate of
diffusion of the complex into the copper site for each of the molecules.

(b) The pair of minor sites is almost co-incident with a non-
crystallographic two-fold axis, the distance between them being only
1.5Å. They lie close to 2 histidine residues, His 32, and His 32',
from two non-crystallographically related molecules,as shown
schematically in Figure 3.24.

From the diagram it is clear that both mercury sites are nearer
molecule II than molecule I. This is probably a direct reflection of
the slight differences in orientation of the two molecules in the
asymmetric unit with respect to the non-crystallographic diad.
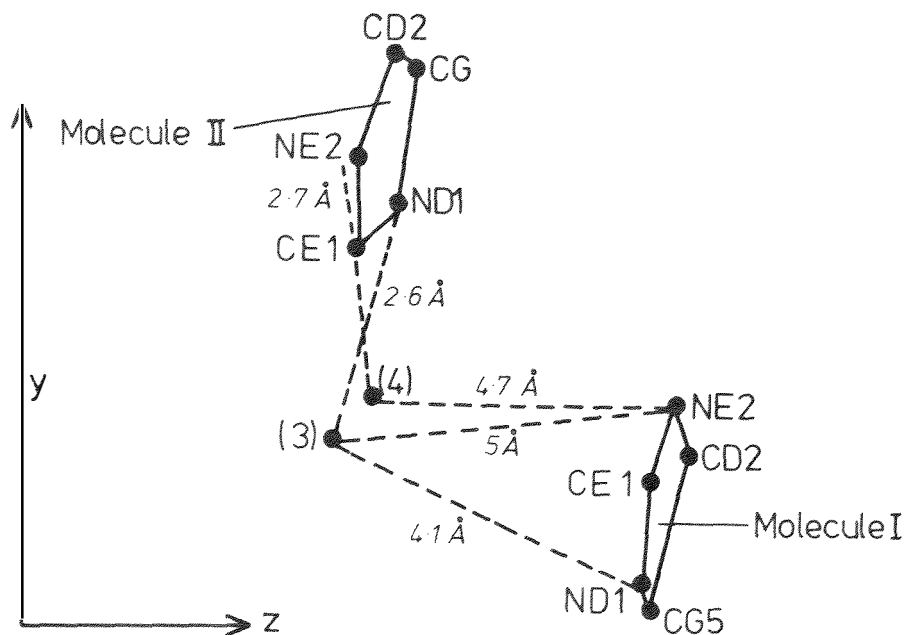
Figure 3.24   A Schematic Diagram of the Minor Mercury Site for

the $Hg(NH_3)_2Cl_2$ Derivative.

The Double Derivative.

(a)   The major gold and uranyl sites are very similar to those in

the individual derivatives, although the substitution of the uranyl

site is now somewhat greater. Reasons for the difference in occupancy

of site 3 are not clear. In molecule II, where the occupancy is

higher, the site lies between two Asp 77 side chains, of molecules

related by crystallographic 2-fold symmetry. In molecule I, no such

relationship exists as there is a difference in interpretation

around Asp 77 suggesting a structural difference which may or may not

be real. The nature of the site, however, suggests that the heavy

atom binding is the uranyl ion, which is commonly found bound to

the carboxylate groups of glutamate or aspartate (105).

(b)   Of the minor sites, site 4, which has approximately equal

substitution on both molecules, is most likely to be that of a uranyl

ion, while site 5 is more likely to be that of an $Au(CN)_2^-$ ion.
The latter is found on Molecule I only, in the vicinity of Val 99
and Ser 100 which are part of a tight loop formed by residues 99-103.
The site is at least 6Å away from any sidechain. It can only be
assumed that differences in the enviroment of this loop in molecule II
cause a disappearance of the potential binding site. Site 6
appears to be non-specific in that it has no near neighbours, but
seems to be in a solvent space between different molecules.

The Nature of the Thiomersal Site.

Soaking the crystals in solutions of thiomersal renders them
colourless as already described (section 3.1.6), indicating that
the copper atom has either been reduced or removed. The former
explanation was accepted after examination of a cross-phased difference
Fourier map showed there was no negative density at or around the
copper site. The binding site of the mercury compound was identified
from this map as being approximately 3.4Å from the copper atom, and
2.4Å from the SG atom of Cys 112 in both molecules in the asymmetric
unit. As this reagent has previously been shown to bind to cysteine
residues (105,162), it is assumed that the binding site on azurin
is the thiolate sulphur of Cys 112, despite the size and shape of
the thiomersal molecules, and the internal hydrophobic nature of the
S^- site. Some disruption does, however, occur as reflected in the
increase of the c axial length on the binding of the complex.

3.11   A Comparison of the Structures of the Two Molecules.

The two molecules were compared using a non-linear least squares
procedure (program NOLIN6, see Appendix II) based on the method of
Rao and Rossmann(163). A set of 3 atoms in each molecule was chosen
to provide an initial orientation matrix. The rotational and
translational parameters relating the two molecules were then refined
by least squares, using 474 equivalent atoms (N, C_α, C, O and C_β)

from each molecule. Only the more reliable parts of the structure
were used in this refinement, and, in addition, atom pairs deviating
by more than three times the standard deviation were discarded
during refinement. Overall, the r.m.s. deviation of the 474 atoms
used in this refinement was 1.3Å. Figure 3.25 is a histogram
showing the differences between average mainchain and sidechain
positions of each residue in the two molecules. The ordinate in
the negative direction represents the average displacement of the
sidechain atoms for each derivative divided by the mean of this figure
for all 129 residues of the protein. Similarly in the positive
direction it represents the average displacement of the 4 mainchain
atoms divided by the mean of this figure summed over all 129
residues. The abscissa represents the residue number.

Many of the apparent differences between the two molecules
undoubtedly represent minor errors in the structure at this stage
and are likely to disappear or be greatly reduced on refinement.
The major differences between the two molecules result from
differences in interpretation during model building, and may or may
not represent structural differences. The histogram may therefore
give some indication of the reliability of different parts of the
structure.

Major differences appear for residues 1-3, 11-13, 21-26, 65-72,
75-81 and 100-105. Differences at the chain termini are not unexpected,
and the other differences are all associated with external loops
reasonably distant from the copper atom, and generally with poorly
resolved density which was difficult to interpret. None of these
sections of mainchain are involved in extensive hydrogen bonding or
other interactions with other parts of the structure. On the other
hand, the sections of mainchain involved in strands of the β-barrel
structure (shown cross-hatched in Figure 3.25) differ very little
(mean deviation is ∿ 0.7Å).

MAIN CHAIN ATOMS
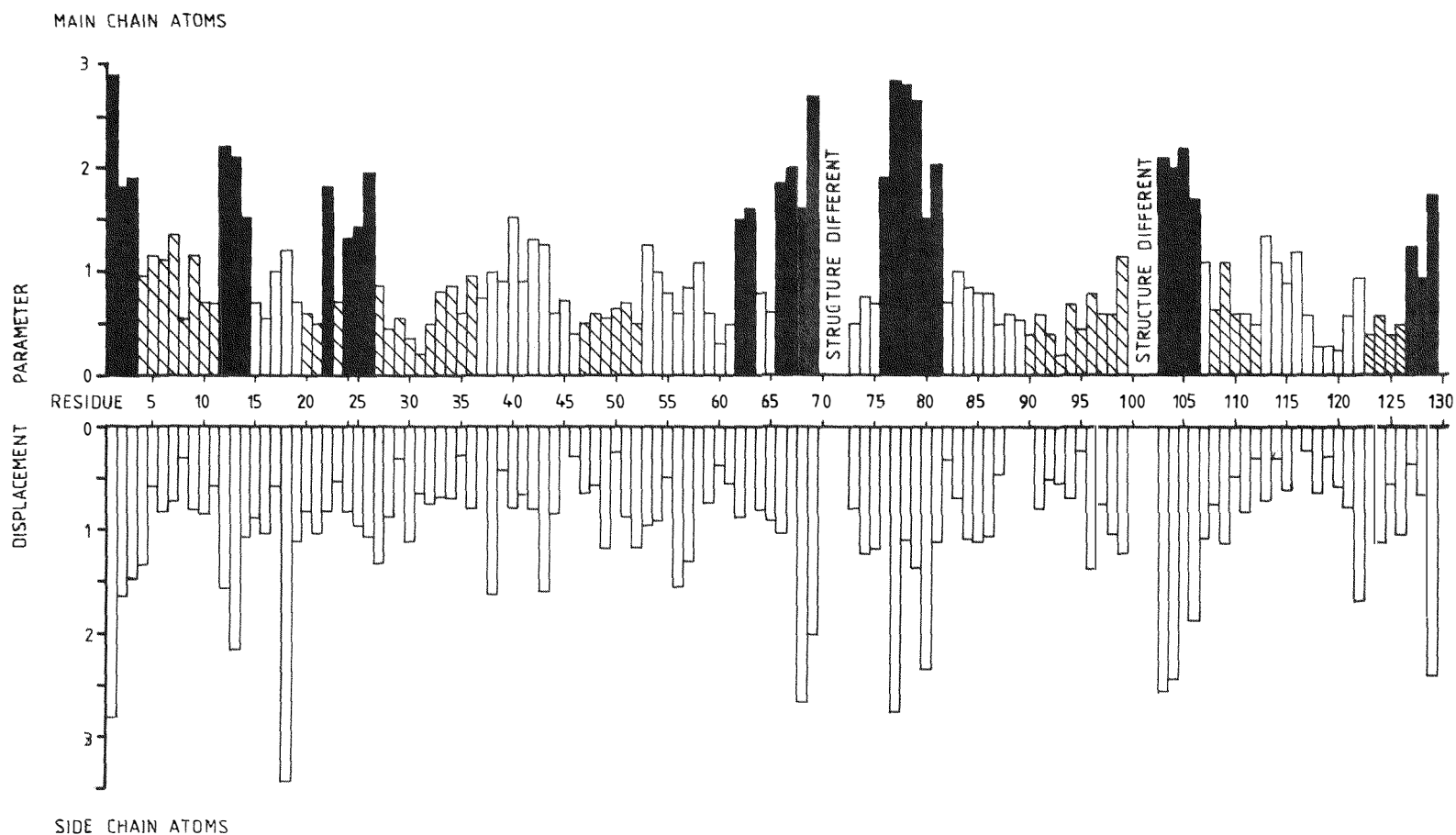
SIDE CHAIN ATOMS

Figure 3-25. The relative displacement between corresponding residues in the two molecules of the asymmetric unit after superposition by the rotation function. Displacement parameter = (mean displacement of main or side chain atoms in residue) / (average of mean displacement over all residues)
Hatched areas represent the main chain residues involved in β structure
Black areas represent the residues showing the greatest displacement

Sidechains which show the greatest differences in orientation, apart from those not included in the histogram (residues 70-72 and 100-102, for which the interpretations are different in the two molecules), are Ala 1, Gln 2, Ala 12, Met 13, Lys 18, Lys 38, Val 43, Lys 56, Leu 68, Ala 69, Asp 77, Val 80, Leu 102, Thr 103, Glu 106, Lys 122, Asn 129. Most of these are residues on the surface of the molecule. Some have well defined densities and orientations that are clearly different. These include Met 13 and Lys 18, and are presumably affected by lattice contacts. Others have poorly resolved density and differences in conformation may not be real. Leu 68, Ala 69 and Val 80 are all hydrophobic, but are in parts of the structure where the interpretation is different for the two molecules. Leu 102 is internal, but there is no density for the sidechain in molecule I, so the difference is probably not real.

The Position of the Non-Crystallographic Two-fold Axis.

The above superposition of the molecules yields three rotation and three translation parameters which define the non-crystallographic symmetry relating the two molecules. In this case, the two molecules chosen were not part of the same asymmetric unit, so that these parameters reflect the four-fold symmetry of the pseudo-tetragonal cell.

If X_{Ii} and X_{IIi} denote the position vectors of the ith equivalent atom in molecules I and II, relative to two arbitrarily chosen axial systems, then

$$X_{IIi} = CX_{Ii} + \underline{t}$$

where C is a rotation matrix expressed in terms of a rotation κ around an axis, the position of which is defined by two spherical co-ordinates ψ and ϕ. The relationship of these angles to the cartesian X_1, X_2 and X_3 axes is described by Rossmann and Blow (164). X_1 is parallel to a, X_2 parallel to b, and X_3 parallel to c. The

vector $\underline{t}$ = (t_1, t_2, t_3) where t_1 represents the translation along X_1, t_2 along X_2 and t_3 along X_3.

The rotation angles were

$\kappa = 85.16^O$, $\psi = 89.323^O$ and $\phi = 88.257^O$, while the translations were $t_1 = -38.68\overset{O}{A}$, $t_2 = 34.99\overset{O}{A}$, and $t_3 = -24.64\overset{O}{A}$. From Figure 3.16 and Figure 2.10, it can be seen that if the non-crystallographic diad was positioned exactly on the diagonal between x and y, at z = 1/8, these parameters would be

$$\kappa = 90.0^O, \quad \psi = 90.0^O, \quad \text{and} \quad \phi = 90.0^O$$

$$t_1 = -37.5\overset{O}{A}, \quad t_2 = 37.05\overset{O}{A}, \quad \text{and} \quad t_3 = -24.88\overset{O}{A}$$

(where these translations represent ½, ½ , and ¼ of the a, b and c axes respectively). The four-fold screw axis (Figure 2.10) generated by the action of crystallographic symmetry operations on the non-crystallographic diad, instead of passing through (-0.5, 0.5, 0) and being parallel to the z axis, passes through the point (-0.526, 0.468, 0) and is at an angle of 1.87^O to the z axis. The non-crystallographic diad passes through the points (.01, -.01, .127) and (.51, .49, .127). The small differences in environment of each of the molecules in the crystallographic asymmetric unit is thus not surprising , as each of the molecules will be oriented slightly differently to the symmetry elements of the unit cell.

CHAPTER 4

THE STRUCTURE OF AZURIN

The azurin molecule is based on a β-barrel structure in which eight strands of the polypeptide chain are arranged parallel or anti-parallel to form a cylindrical barrel. The β-sheet topology of the molecule (Figure 1.4b) is not identical with the β topology of any of the proteins listed by Richardson (165) although such views of the β topology are highly idealised - for instance, structural irregularities in the β strands, and differences between lengths of corresponding β strands in different structures are ignored. The core of the molecule is made up of hydrophobic sidechains, and there is a very distinctive hydrophobic patch around the edge of His 117 on the surface of the molecule (Figure 4.1).

The folding follows the same pattern as that of plastocyanin (56), rather than that suggested originally by Adman et al. for Ps. aeruginosa azurin (82). In the present structure, the two N-terminal residues were well-defined in both molecules as were the rest of the residues in strands I and II. Also, Tyr 15 and Met 13 fitted the density well, especially the former where the sidechain density was visible as a well-defined, flattened disc. With the alternative azurin folding of Adman et al. (82), Met 13 and Tyr 15 change places, as the ambiguity in interpreting the position of the disulphide bridge, Cys 3- Cys 26 causes strands I and II to change place. This ambiguity in Pseudomonas aeruginosa azurin appears to have arisen because no density was visible for the two N-terminal residues, bulky sidechain density appeared where Met 13 and Phe 15 were expected, and other hydrophobic sidechains had internal positions. When phases were extended to 2.7Å, and the electron density improved by direct space averaging (83), however, the length of strand II in
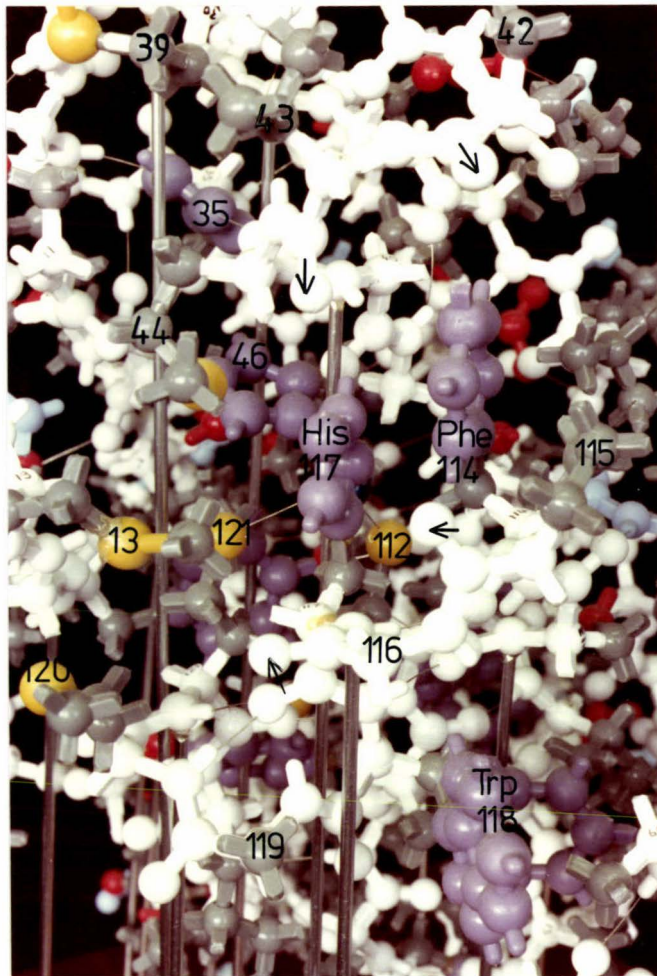
Figure 4·1 Close view of the Hydrophobic
Patch.
Arrows show carbonyls pointing
towards His 117 .

Figure 1.4a could not be spanned with the available chain. Although this led to the alternative interpretation, similar to the topology of plastocyanin (i.e. a reversal of the chain direction and numbering of the first two strands), the fit of the model to the density in the region of the middle of strand II was found to be less than satisfactory, introducing doubt about the interpretation. This present work has removed any doubt, as the good fit of most residues in strands I and II, with no ambiguity in interpretation, confirms the plastocyanin-type topology for azurin.

4.1 The Secondary Structure of Azurin

The term secondary structure refers to regular elements in the folding of a polypeptide chain, stabilized by hydrogen bonding. A hydrogen bond between two appropriate groups in a protein molecule is likely when:

(a) the distance between the H donor and H acceptor lies in the approximate ranges:

O-H.....N, 2.6-3.0Å ;   N-H.....O, 2.6-3.3Å   (167,168)

N-H.....N, 2.8-3.2Å ;   O-H.....S ∿ 3.3Å   (167,168)

N-H.....S ∿ 3.6Å   (169)
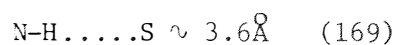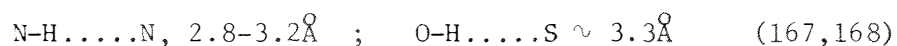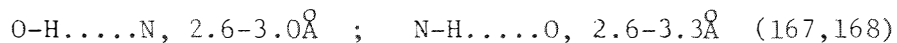
(b) the angle A-Ĥ-.....B, where AH is the proton donor, and B the proton acceptor is between 140 and 180°. Ideally, bonding is most favoured when the angle is near 180°, although in proteins N-Ĥ.....O angles are commonly about 160°. When the angle becomes less than 140° effective hydrogen bonding will be much reduced (167).

Although the hydrogen bonding pattern is similar in both molecules, it must still be regarded as tentative at this stage since the structure is unrefined, and refinement may alter the positions and orientations of groups quite significantly. Mainchain to
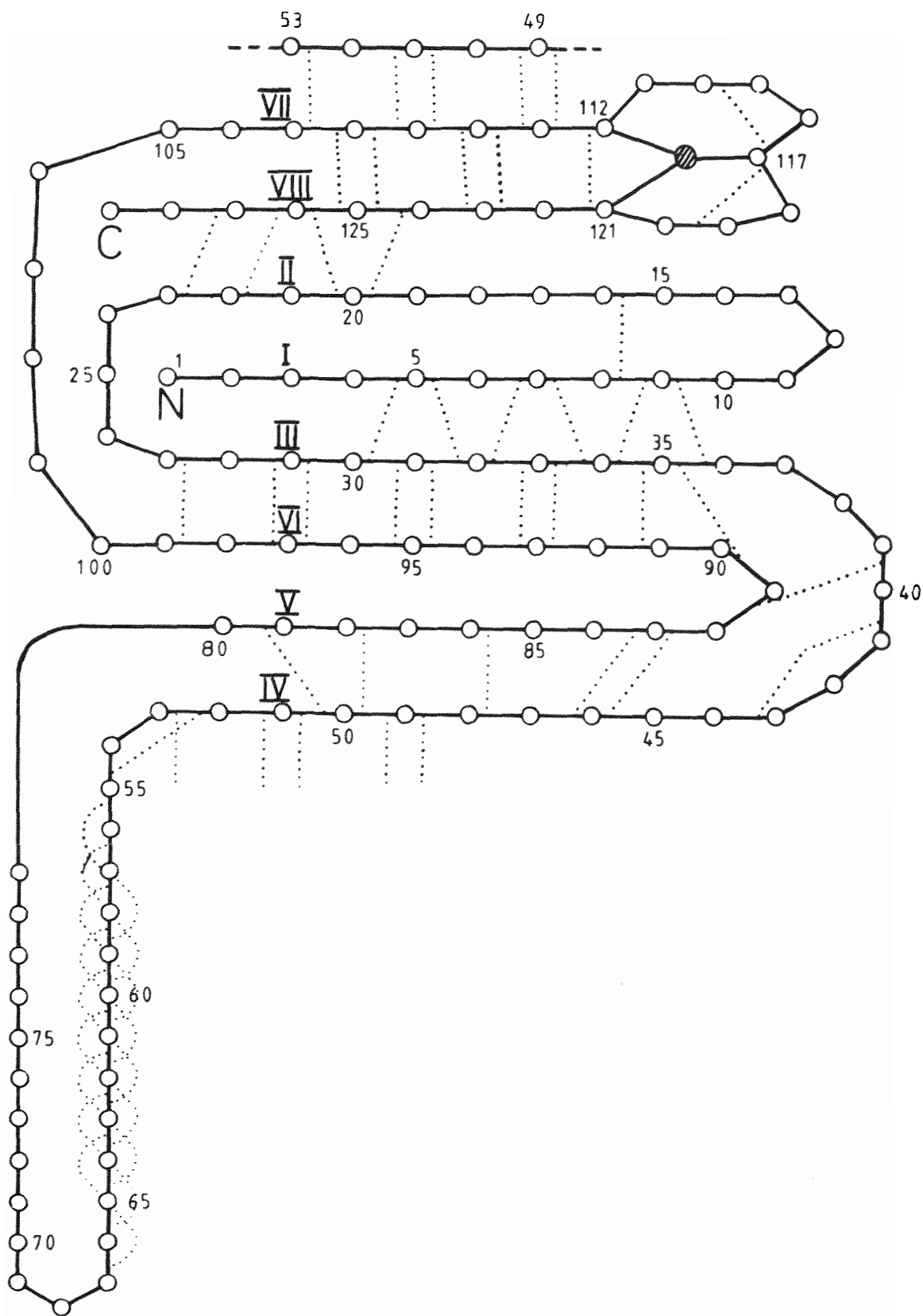
Figure 4-2. A schematic represention of the mainchain hydrogen bonding in azurin.

mainchain hydrogen bonds are shown schematically in Figure 4.2.

Table 4.1 lists mainchain to sidechain hydrogen bonds, while Table

4.2 lists sidechain to sidechain interactions.Apparent variations

in sidechain interactions between the two molecules at this stage of

the structure determination reflect either differences in interpreta-

tion (for instance between residues 68 and 72,as mentioned in

section 3.9) or local differences in the orientations of sidechains.

These should be clarified on refinement.

4.1.1 β-Structure

A classical β pleated sheet is made up of extended chains which

can either run parallel or antiparallel, and which are laced together

with hydrogen bonds between the carbonyl and amide groups. The need

to make "good" hydrogen bonds keeps the chain from being fully

extended, giving it a "pleated" look, the separation between the

chains being typically 4.1-4.7Å (170). Figure 4.3 is a schematic

diagram of a portion of parallel and antiparallel pleated sheet.

antiparallel   parallel
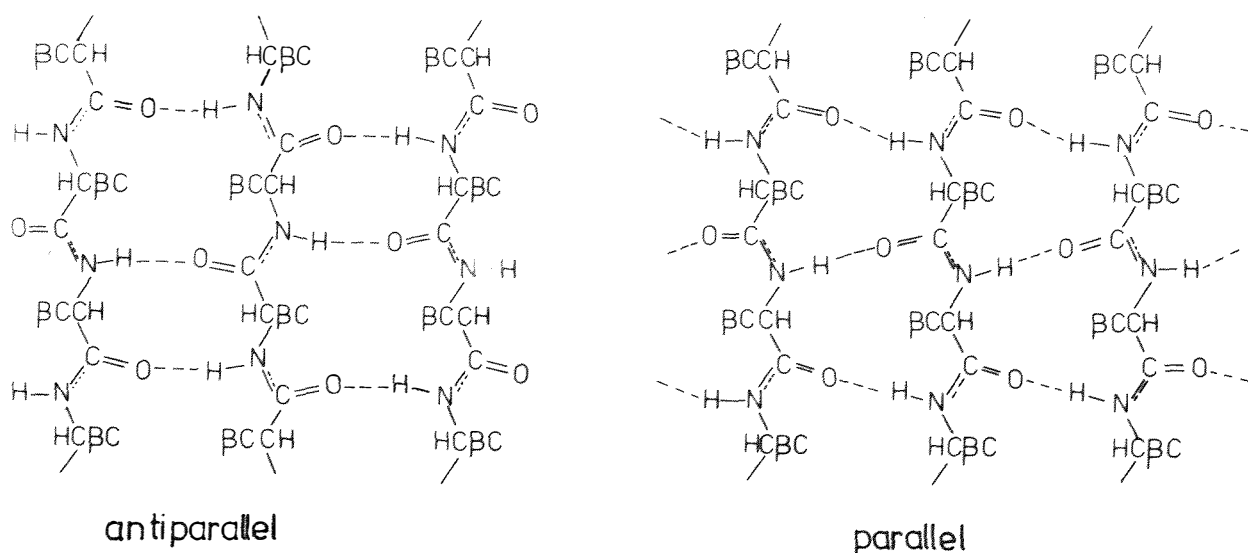
Figure 4.3   A Schematic Representation of Antiparallel Chain, and

Parallel Chain Pleated Sheet Structure.

Table 4.1   Intramolecular Hydrogen Bonds; Mainchain-Sidechain

MOLECULE I				MOLECULE II			
Mainchain		Sidechain		Mainchain		Sidechain	
Glu 8	NH	Asp 16	OD2	Glu 8	NH	Asp 16	OD2
Asn 10	C=O	His 46	NE2	Asn 10	C=O	His 46	NE2
Gln 14	C=O	Asn 10	ND2				
				Tyr 15	C=O	Gln 14	NE2
Lys 18	C=O	Tyr 110	OH	Lys 18	C=O	Tyr 110	OH
Lys 24	NH	Asn 129	OD1				
Gly 37	NH	His 35	ND1	Gly 37	NH	His 35	ND1
Lys 38	NH	Asp 11	OD1				
				Met 39	NH	Asp 11	OD1
Met 44	C=O	His 35	NE2	Met 44	C=O	His 35	NE2
Asn 47	C=O	Tyr 15	OH	Asn 47	C=O	Tyr 15	OH
Asn 47	NH	Cys 112	SG	Asn 47	NH	Cys 112	SG
Trp 48	C=O	Thr 84	OG1				
Ala 54	C=O	Lys 56	NZ				
Asp 55	C=O	Thr 51	OG1				
Glu 57	C=O	Thr 61	OG1	Glu 57	C=O	Thr 61	OG1
Ala 66	C=O	Gln 70	NE2				
Arg 79	C=O	Thr 78	OG1	Arg 79	C=O	Thr 51	OG1
				Arg 79	NH	Asp 77	OD1
Thr 84	C=O	Asn 47	ND2	Thr 84	C=O	Asn 47	ND2
Lys 85	NH	Asp 93	OD2	Lys 85	NH	Asp 93	OD2
Gly 88	NH	Glu 91	OE1	Gly 88	NH	Glu 91	OE1
Val 99	C=O	Ser 100	OG				
				Ser 113	NH	Asn 47	OD1
Asn 129	C=O	Asp 23	OD1				

Table 4.2   Intramolecular Hydrogen Bonds; Sidechain-Sidechain

	MOLECULE I				MOLECULE II		
Sidechain		Sidechain		Sidechain		Sidechain	
Glu 4	OE2	His 32	NE2	Glu 4	OE2	His 32	NE2
Glu 19	OE2	Lys 18	NZ	Glu 19	OE2	Lys 18	NZ
Asp 23	OD2	Ser 25	OG				
Gln 28	NE2	Asp 98	OD2	Gln 28	NE2	Asp 98	OD2
Thr 30	OG1	Thr 96	OG1	Thr 30	OG1	Thr 96	OG1
His 32	ND1	Ser 94	OG	His 32	ND1	Ser 94	OG
Asn 47	OD1	Ser 113	OG	Asn 47	OD1	Ser 113	OG
Asp 62	OD1	Lys 74	NZ	Asp 62	OD1	Arg 79	NH1
				Thr 84	OG1	Asp 93	OD1
Lys 85	NZ	Asp 93	OD2	Lys 85	NZ	Asp 93	OD2
				Glu 106	OE1	Tyr 108	OH

In practice, β-sheets in proteins are rarely found to be so regular. The hydrogen bonding pattern is often interrupted, for example, by so called β-bulges or kinks in the chain (171,172). The strands are almost always twisted, and Chothia (173) has shown that sheets made up from extended polypeptide chains, with a right handed twist about the axis running in the direction of the chain, have a lower free energy than sheets that are straight or have a left handed twist. The β strands occurring in the known protein structures all have right-handed twists of between 5° and 30° per residue, i.e. they would make one full 360° twist in from 12 to 72 residues. Refined protein structures have shown that the hydrogen bonds are seldom linear, the average N-Ĥ.....O angle being about 163° for antiparallel chains (174). Parallel sheets have been found in general to be less common than antiparallel or mixed sheets (165). No parallel β-structure occurs in sheets of less than five total strands, suggesting that considerable co-operativity is necessary to stabilise parallel β-structure. Examples of such structures are lactate dehydrogenase (175) and adenylate kinase (176). Antiparallel β-structure on the other hand, often occurs as a twisted ribbon of just two strands, for example in ferredoxin (177). Mixed sheets are often found, for example in carboxypeptidase (178) and thioredoxin (179). In such examples, the parallel portions of mixed sheets are usually found buried in the interior of the protein. This probably reflects a reduced stability with respect to the antiparallel sheets, which typically have one side exposed to the solvent (165).

In azurin, parts of the backbone have pronounced β character with the three strands I, III, and VI of Figure 1.4b forming a classical β-sheet, with a right handed twist. Figure 4.4 is a stereo-plot of strands I, III and VI, while Figure 4.5 is a photograph of the Labquip model viewing this region.
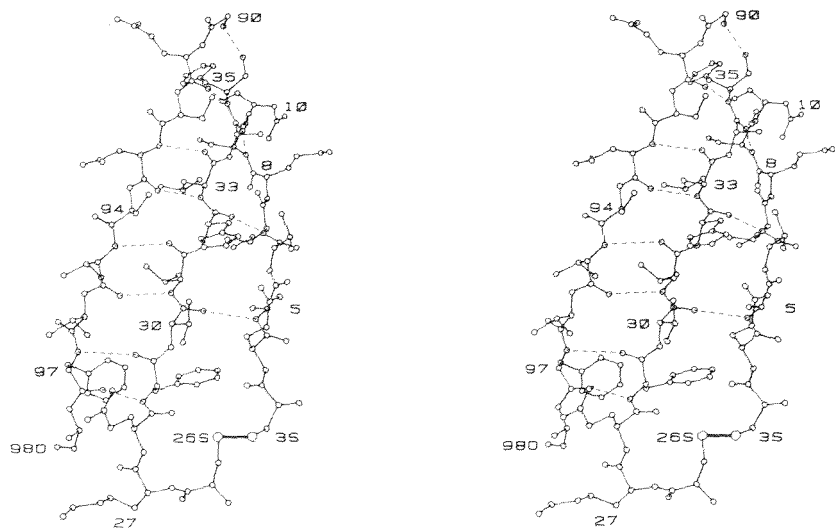
Figure 4.4   A Stereo Diagram of the β-Structure of Strands I,
              III and VI.

Strands I and III run parallel to each other, but rather than being

found in the interior of the molecule, they are part of the surface.

Figure 4.4 shows, however, that the hydrogen bonding between them

is protected from the solvent by the sidechains of the residues

which make up these strands, most of which are polar and protrude from

the surface of the molecule. The remainder of the barrel is more

distorted, with only short regions in each strand where the classical

bonding of a β-sheet is seen. This is particularly obvious in strand

II, which has a bulge between residues 16 and 19, and only forms

hydrogen bonds at its beginning (residues 13-16 hydrogen bonding

to the anti-parallel strand I) and at its end (where residues 20-23
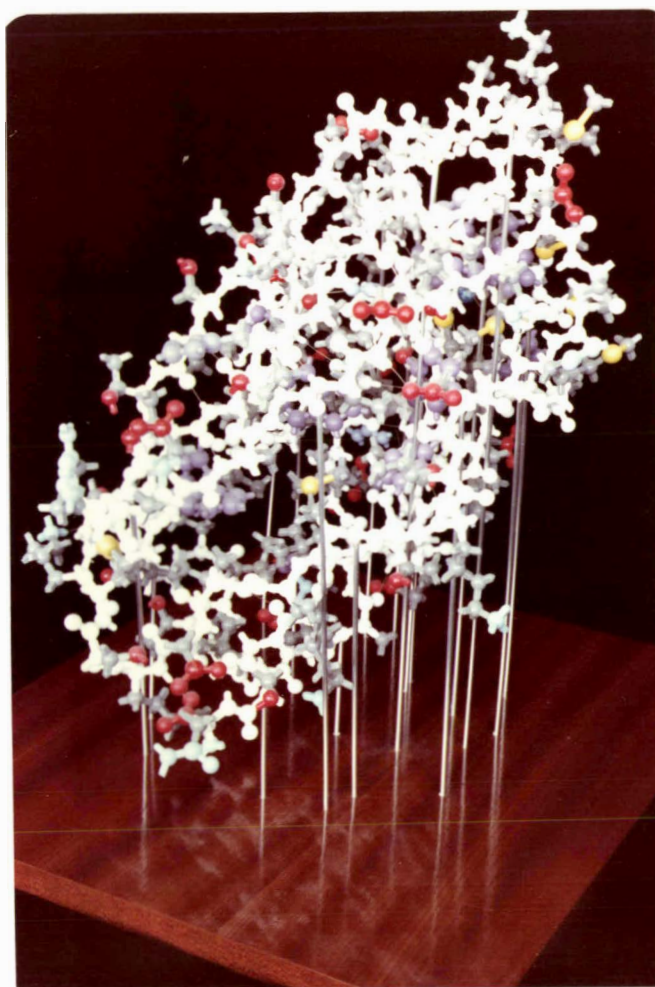
Figure 4·5 A view of the Labquip model
showing the β-sheet structure of
strands I, III and VI.
The barrel shape of the molecule
is clearly evident.

hydrogen bond to the parallel strand VIII). There is no hydrogen

bonding at all between strands V and VI, which run anti-parallel,

but where local variation in sidechain conformation has resulted

in the mainly hydrophobic sidechains pointing towards each other.

The separation between the mainchain atoms of these strands is

approximately 7Å, compared to the more usual separation of 4.5Å

typical of β-sheet structure (170) and found for example between

strands III and IV. Thus there appears at this stage to be a

reasonable gap between the strands giving access to the hydrophobic

interior of the molecule and ultimately to the copper site. It is

interesting to note that the sidechains lining this "gap" are highly

conserved in all azurins. Out of the 26 residues associated with

this opening, 15 are invariant, and 7 are conservatively substituted,

(i.e. retain their hydrophobic character). Of the remaining 4

residues, 100 is Ser in all but one species in which it is replaced

by Ala, 85 is always Lys, except in one species where it is replaced

by Ser, and 78 is always Ser or Thr except in two species where it

is replaced by Ala or Glu. Because this "gap" between strands V and

VI is only partially blocked by sidechains near the top of the molecule

(see Figure 4.6), there are some spaces, particularly around His 83,

which may allow access to the hydrophobic core. This factor, combined

with the highly conserved nature of the area, may point to some

functional significance for this region of the protein. A stereo

diagram is shown in Figure 4.6.

Connections between β strands may be either "hairpin" connections,

in which the backbone chain re-enters the same end of the β-sheet it

left, i.e.

Figure 4.6   A Stereo Diagram of the "Gap" Between Strands V and VI.

or cross-over connections, in which the chain loops around to re-

enter the sheet at the opposite end (165) i.e.

In azurin, there are no cross-over connections, but there are three

hairpin bends between nearest neighbours, (i.e. between strands I
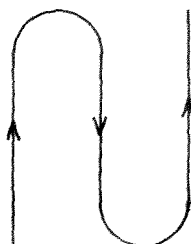
and II, V and VI, and VII and VIII), one with one intervening strand
(i.e. between strand II and III), and two with 2 intervening strands
(i.e. between strands III and IV, and VI and VII) (Figure 1.4b).

4.1.2   The α-Helix.

The α-helix involves residues 55-67, and is found in the flap
region of the molecule. It is shown as a stereo diagram in Figure
4.7. It was a prominent feature of the low resolution map, and was
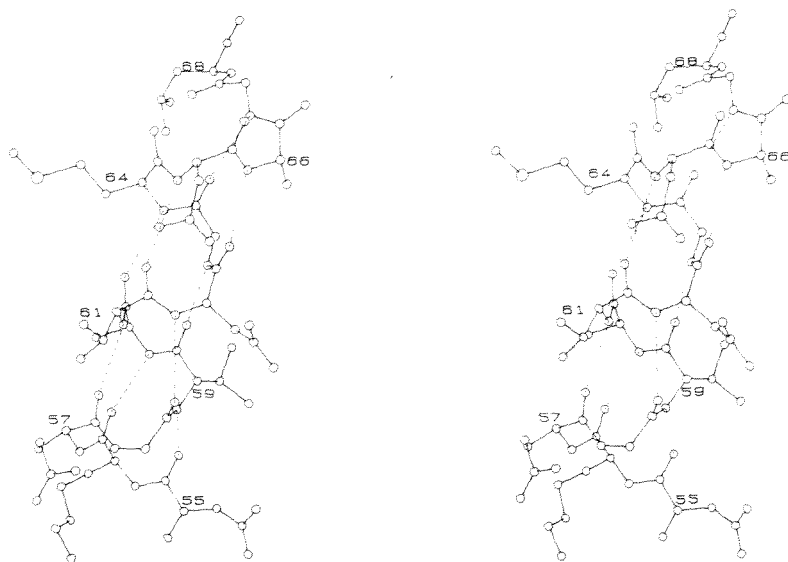
Figure 4.7   A Stereo Diagram of α-Helix

clearly resolved in the high resolution map for both molecules.
Although some of the hydrogen bonds appear a little long, the helix
is regular within the present accuracy of the analysis. Slight
distortions arise from the tilting outwards of the carbonyl oxygen
of Glu 57 to form a bifurcated hydrogen bond with both the mainchain
NH and sidechain OH atoms of Thr 61, of the carbonyl oxygen of Asn 65

to form an intermolecular hydrogen bond (see section 4.3), and of
the carbonyl oxygen of Ala 66 to form a possible hydrogen bond with
sidechain amide group of Gln 70. It is also possible that the
carbonyl oxygen of Thr 61 may form a bifurcated hydrogen bond with
both the mainchain and sidechain amide groups of Asn 65. Such
distortions are not uncommon (171), especially near the beginnings
and ends of helices as in this case. Residues whose sidechains
are exposed to the solvent include Asp 55, Lys 56, Glu 57, Thr 61,
Asp 62, Met 64, Asn 65, and Gly 67. All are polar, except for Met 64,
which is an extension of the hydrophobic patch on the surface of
the molecule (see section 4.2.3) , and Gly 67. Sidechains of the
remaining residues making up the helix point inwards, linking the
flap to the main β-barrel of the protein, and are, as expected, non-
polar, i.e. Val 59, Ala 60, and Gly 63.

4.1.3 Turns.

Hairpin bends, which involve chain reversals of approximately
180° in the polypeptide chain direction, frequently fall into a small
number of conformations. They often involve 4 residues with a
hydrogen bond between the carbonyl oxygen of residue (i) and the
amide nitrogen of residue (i + 3), and are classified as turns. Such
turns are found not only between strands of β-structure, but also
in less regular elements of polypeptide chain structure where tight
loops are formed.

Venkatachalam has categorized them into three types, distinguished,
in their general conformations, by the dihedral angles of the second
and third residues (180). One of these is a helical type of turn,
actually part of a 3_{10} helix, and is designated type III, while the
non-helical types, designated types I and II, involve almost complete
reversal in the direction of the polypeptide chain and are related
by a 180° twist in the second peptide unit. Venkatachalam also

found that any 4 residues which formed a type II turn necessarily
had a glycine residue in the third position. The allowed angles
are as follows, in the order $(\phi_2, \psi_2 ; \phi_3, \psi_3)$.

Type   I $(-60^\circ, -30^\circ : -90^\circ, 0^\circ)$ I' $(60^\circ, 30^\circ : 90^\circ, 0^\circ)$

Type   II $(-60^\circ, 120^\circ : 80^\circ, 0^\circ)$ II' $(60^\circ, -120^\circ : -80^\circ, 0^\circ)$

Type   III $(-60^\circ, -30^\circ : -60^\circ, -30^\circ)$ III' $(60^\circ, 30^\circ : 60^\circ, 30^\circ)$

Crawford et al. (181) have expanded these criteria to include other,
less regular, chain reversals in the definition of a reverse turn.
Following their approach, chain reversals in azurins have been
identified where the following conditions apply:

(1) $CA_{(i)}\ldots\ldots\ldots CA_{(i+3)} < 6\text{Å}$

(2) $\phi_2, \psi_2; \phi_3, \psi_3$ are within 20° of the values above, except that
      one may be 30° different.

(3)   There is a possibility of a $4 \rightarrow 1$ hydrogen bond (i.e.
      $O_{(i)} - N_{(i+3)} < 3.2\text{Å}$, and the geometry of the bond appears to be
      reasonable).

      Turns are then designated as follows:

RT =   reverse turn; all three criteria hold.

NR =   near reverse turn; two of the three criteria hold.

O   =   open turn; one of the three criteria hold.

Chain reversals which appear to fall into these categories in the
structure are shown in Table 4.3 for molecule I, with the values for
molecule II in parentheses.

Three of the turns listed have similar dihedral angles in
both molecules. These are between strands V and VI (88-91) and in
the loops around the copper site, (114-117) and (117-120). The
others, however, are all in regions of the molecule for which the
electron density was poorly defined, and/or there are differences in

Table 4.3   Turns Found in the Structure of Azurin from Alc. denitrificans.

Segment	Residues	d (Å)	ϕ_2	ψ_2	ϕ_3	ψ_3	H bond (Å)	Type
52–55	Lys-Glu-Ala-Asp	5.0 (5.3)	−49(−46)	−24 (129)	− 93(83)	15(−8)	2.9 (2.7)	RT I
88–91	Gly-Gly-Gly-Glu	5.5 (5.9)	−28(−58)	121 (124)	92(116)	19(−2)	3.0 (3.1)	RT II
100–103	Ser-Lys-Leu-Thr	4.8 (7.2)	−80(−114)	62 (10)	167(−108)	39(70)	3.5 (5.3)	O

Loops Around the Copper Site

Segment	Residues	d (Å)	ϕ_2	ψ_2	ϕ_3	ψ_3	H bond (Å)	Type
114–117	Phe-Pro-Gly-His	5.9 (5.4)	−50(−22)	112 (104)	76(95)	18(−22)	3.1 (2.8)	RT II
117–120	His-Trp-Ala-Met	5.5 (5.9)	−51(−56)	8 (3)	−101(−104)	−30(−24)	2.7 (3.0)	NR I

Loops in the Flap Region

Segment	Residues	d (Å)	ϕ_2	ψ_2	ϕ_3	ψ_3	H bond (Å)	Type
71–74	Asp-Tyr-Val-Lys	5.7 (7.4)	−150(−118)	79 (51)	138(−123)	−45(−65)	5.6 (5.3)	O
68–71	Leu-Ala-Gln-Asp	5.4 (4.9)	77(44)	91 (81)	− 45(110)	−22(48)	4.3 (4.6)	O

interpretation between the two molecules, – for example the loops
comprising residues 71-74 and 68-71. It is not surprising, therefore,
that the dihedral angles vary between the two molecules, and not too
much should be read into these values. The remaining hairpin bend
in the structure, that between strands I and II, does not fit into
any of the above categories.

4.2 The Tertiary Structure of Azurin.

Figure 4.8 is a stereo plot of the α-carbon positions, showing
the folding of the polypeptide chain. The molecule is shaped like a
slightly flattened pear, with dimensions of approximately 45 x 34 x 37Å.
The β-barrel structure can be clearly seen, with the strands running
approximately vertically, while the so called "flap" region sits
away from the main body of the protein. The single disulphide
bridge can be seen near the bottom of the molecule.
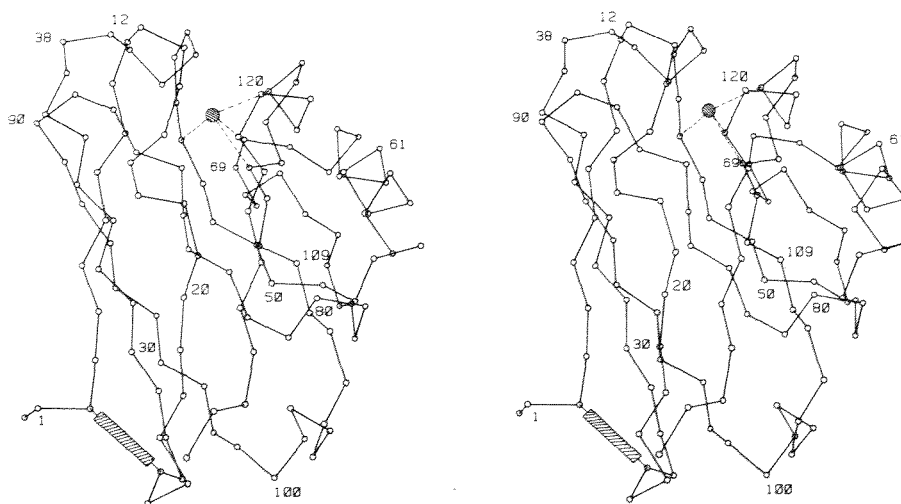
Figure 4.8   A Stereo Diagram of the α-Carbon Positions. The Copper
is Represented by the Large Circle, and the Disulphide
Bridge by ▨▨▨

4.2.1 The Copper Site.

The copper atoms in both molecules of the asymmetric unit are
situated in the narrow end of the β-barrel, as seen in Figure 4.8,
their location being completely unambiguous since the peaks
representing them were the highest in the electron density maps.
The copper atom lies approximately 6Å below the surface of the
molecular envelope, although at one point the imidazole group
providing one of the ligands to the copper, His 117, is all that
separates it from the solvent (one edge of the ring being exposed).
At 3Å resolution, the electron density of the copper atom fuses with
that from several other pieces of structure, as shown in Figure 4.9.
After model building, these were identified as the sidechains of
Cys 112, His 117, Met 121 and His 46. All these residues are
invariant in all azurins so far sequenced. Thus   three of the four
probable ligands belong to the sequence -Cys-x-x-x-His-x-x-x-x-Met-
which forms a tight loop between strands VII and VIII, incorporating
two well-defined turns (see table 4.3). In the electron density
map for molecule I, another piece of electron density fused with that
of the copper atom, and was shown, after model building to belong
to the mainchain between residues 45 and 46. The only convincing fit
to this density had the carbonyl oxygen of Gly 45 pointing towards
the copper atom at a distance of 2.7Å. Although the density was not
as strong in molecule II, it was interpreted in the same way, the
distance between the oxygen and the copper atom being increased,
but the same configuration being maintained. In view of the difficulty
in interpreting peptide orientations with certainty at 3Å resolution,
the question of whether this peptide provides a ligand for the copper
atom or not cannot be answered unequivocally at present. It is
interesting to note however, that Gly 45 is also invariant in all
azurins.

Figure 4.9   Some Sections of the Electron Density Map

Showing the Copper Site.

The density of His 46, Cys 112, Met 121 (at

the back) and His 117 (at the back) merge

with the Cu density. The density of the

peptide unit between residues 45 and 46 (at

the top) also merges with that of the Cu.

The planes of the imidazole rings of His 46

and His 35 are clearly defined (top left).

Adjacent to the Cu site, the sidechains of

Tyr 15 and Phe 114 can be seen end on.

Although atomic co-ordinates are unrefined, and therefore

subject to some uncertainty, the geometry of the copper site seems

to be very distorted in both molecules. Approximate bond lengths

and angles are shown in Table 4.4, while the geometry of the site

is shown in stereo in Figure 4.10. Figure 4.11 is a stereo plot

of the copper site for both molecules, with molecule II (green)

superimposed on molecule I.
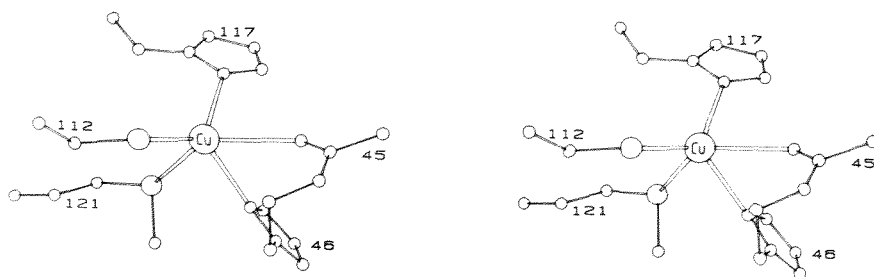
Figure 4.10 A Stereo Drawing of the Copper Binding Site in Azurin

Showing the Five Possible Ligands.

As can be seen from the diagrams, the co-ordination sphere

consists of three almost co-planar ligands, viz the ND1 atoms of His 46

and His 117, and the SG atom of Cys 112, all of which have fairly

normal bond lengths within the accuracy of the structure determination

(2.0-2.4$\overset{o}{A}$). The copper atom is displaced only 0.1$\overset{o}{A}$ out of the plane

formed by these ligands, and the bond angles between them range

between 109° and 128° for both molecules, the mean value being

120° ± 7°. A fourth ligand, Met 121, is found considerably further
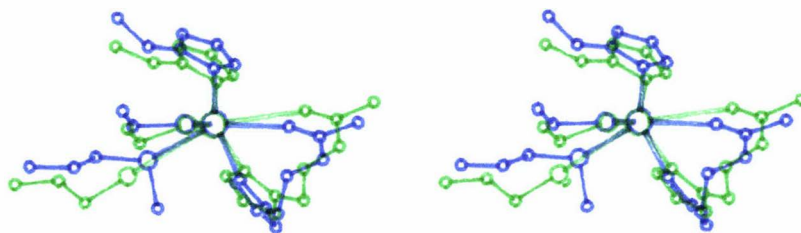
Figure 4.11   A Stereo Drawing of the Copper Binding Site of
Both Molecules in the Asymmetric Unit (Molecule I-
black, Molecule II-green).

from the copper. If these 4 ligands are taken to form the

co-ordination sphere, the co-ordination geometry resembles a very

distorted tetrahedron as found in plastocyanin (56). If the carbonyl

oxygen of Gly 45 is regarded as a fifth ligand, however, the geometry

is best described as trigonal bipyramidal, very similar to that

proposed by Miskowski et al. from spectroscopic results (35). It

must be stressed again, that at the present resolution it is

impossible to say exactly how this peptide is oriented, or to give

a precise Cu-O distance. All that can be said is that the best

interpretation of the electron density has the peptide oriented so

that the carbonyl oxygen points towards the copper, and is at a

distance from the copper atom smaller than would be expected if no

interaction (albeit weak) were occurring. Furthermore, when the

Table 4.4   The Bond Lengths and Angles of the Copper Co-ordination Sphere in Azurin

from Alc. denitrificans.

Ligand	Molecule I	Molecule II	Bond Angle	Molecule I	Molecule II
His 117 (ND1)	2.4 Å	2.3 Å	N_{117}-$\overset{\wedge}{Cu}$-S_{112}	125°	128°
			N_{117}-$\overset{\wedge}{Cu}$-N_{46}	126°	118°
His 46 (ND1)	2.4 Å	2.0 Å	N_{46} -$\overset{\wedge}{Cu}$-S_{112}	109°	114°
			N_{46} -$\overset{\wedge}{Cu}$-O_{45}	66°	77°
Cys 112 (SG)	2.3 Å	2.1 Å	S_{112}-$\overset{\wedge}{Cu}$-S_{121}	98°	99°
			S_{112}-$\overset{\wedge}{Cu}$-O_{45}	120°	110°
Met 121 (SD)	3.1 Å	3.5 Å	S_{121}-$\overset{\wedge}{Cu}$-N_{46}	63°	80°
			S_{121}-$\overset{\wedge}{Cu}$-N_{117}	107°	98°
Gly 45 (0)	2.7 Å	3.2 Å	O_{45} -$\overset{\wedge}{Cu}$-S_{121}	125°	149°
			O_{45} -$\overset{\wedge}{Cu}$-N_{117}	86°	75°

carbonyl oxygen is oriented this way, the amide nitrogen is in a position to form a hydrogen bond with the carbonyl oxygen of residue 87.

The $Cu-SD_{121}$ and $Cu-O_{45}$ (if it exists) bonds are both exceptionally long, and can only be regarded as very weak. As tetrahedral complexes of Cu(II) are rarely found, predictions of "normal" bond lengths have been obtained from small molecule complexes with distorted square planar and tetragonal stereochemistries. Cu(II)-S distances of between 2.28Å and 2.32Å are known for such complexes (182,183). Recently, the structure of a tetrahedral Cu(II) complex has been reported with Cu(II)-S bond lengths of 2.31 and 2.32Å (157). Thus a length of $\sim$ 2.3Å seems "normal" for Cu-S bonds. Cu-N and Cu-O bonds are normally 1.9-2.1Å in length (consistent with covalent radii of 1.35Å for Cu(II), 1.30Å for Cu(I), 0.66Å for O and 0.70Å for N (185,186).

Compounds with a geometry similar to that seen in azurin are rare. One which is somewhat similar is an organic $Cu(II)-N_2S_2O$ chromophore $\left(Cu(pma) SO_4,$ where pma = 2-pyridylmethylbis-(2-ethylthioethyl)amine (184)$\right)$. This complex was found to have a distorted trigonal bipyramidal geometry, the three equatorial ligands, 2 sulphurs and 1 nitrogen having bond distances of 2.39, 2.46 and 2.02Å respectively, while the 2 axial ligands, nitrogen and oxygen had bond distances of 2.03 and 1.91Å respectively. In addition, Cu-S bonds as long as 2.85Å have been reported for complexes with different stereochemistries (187,188), as has at least one very long Cu(I)-N bond of 2.92Å (26).

Such long range bonds may be regarded as ion dipole bonds arising from the attraction of the Cu(II) ion, which must have a residual positive charge, and the electron density of the lone pair of electrons on the oxygen or the sulphur. In this type of bond, appreciable mixing of the lone pair orbitals with those on the copper is not required, the overlap being small. The attraction arises mainly from an asymmetrical charge distribution within the separate

systems (189). The long bonds, if they are real, presumably arise because of the constraints of the protein structure, which do not allow the ligands to approach more closely. Since no small molecule complexes with a co-ordination number of three have yet been observed for Cu(II), these weak interactions may act to stabilize the Cu(II) state of the protein with respect to the Cu(I) state. Such findings only serve to highlight the difficulties in comparing the stereochemistry of the "blue" copper sites with that of well characterized low molecular weight complexes. They also point out the problems in making low molecular weight complexes which mimic the inorganic properties of the blue copper sites.

4.2.2   The Environment of the Copper Site.

The conformation of the protein around the copper binding site is severely restrained by the formation of a large number of hydrogen bonds. These are summarised in Figure 4.12 and are as follows.

(1)   The NE2 atom of His 46 appears to form a hydrogen bond with the carbonyl oxygen of Asn 10. The amide nitrogen and carbonyl oxygen of His 46 form hydrogen bonds with the mainchain atoms of Ile 87.

(2)   The thiolate sulphur of Cys 112 is in a position to form a hydrogen bond with the mainchain amide nitrogen of Asn 47. Its amide nitrogen is hydrogen bonded to the carbonyl oxygen of Met 121, and its carbonyl oxygen to the NH of Trp 118, thus stabilizing the Cu-binding loop.

(3)   The mainchain amide of His 117 hydrogen bonds to the carbonyl oxygen of Phe 114, while its carbonyl oxygen forms a bifurcated hydrogen bond with the amide nitrogens of both Met 120 and Met 121.
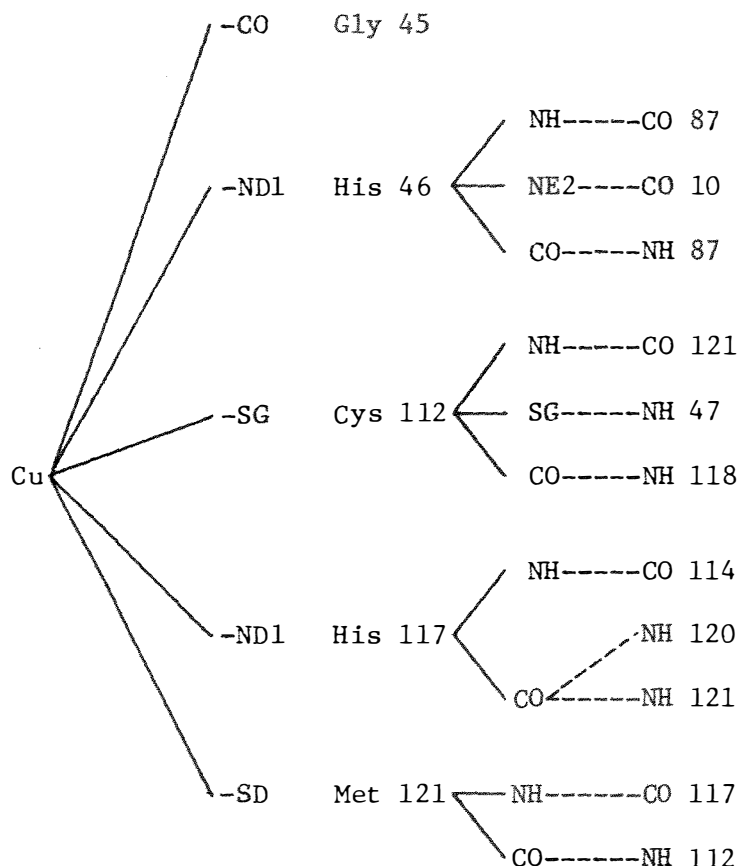
Figure 4.12   A Schematic Representation of the Hydrogen

Bonding Around the Copper Site.

The interconnecting system of hydrogen bonds between residues

providing ligands for the metal ion is extended by the hydrogen

bonding between the strands of the β-barrel, since His 46 lies at

one end of strand IV, Cys 112 at the end of strand VII and Met 121

at the beginning of strand VIII.

The nature of the copper site is quite hydrophobic and largely

conserved in all azurins that have been so far sequenced. Of the

nearest neighbours to the copper atom, 9 residues including the

5 possible ligands are invariant. These plus the next nearest

neighbours are listed in Table 4.5, and designated as invariant,

semi-conserved (i.e. replaced but retaining hydrophobic, aromatic or polar character), or different. The list includes the copper ligands, a group of non polar residues which form a hydrophobic patch on the protein surface near the sidechain of His 117, and a number of residues which contribute important hydrogen bonds to the structure. Together with other residues found in the structure which may or may not be functionally significant these are discussed more fully below.

Table 4.5   The Environment of the Copper Atom in Azurin from Alc. denitrificans.

Nearest Neighbours (a)			Next Nearest Neighbours (b)		
Invariant	Semi-Conserved	Variable	Invariant	Semi-Conserved	Variable
Gly 45			Leu 33	Asn 10	Ser 9
His 46			His 35	Gln 14	Trp 118
Cys 112			(Tyr 72)(c)	Tyr 15	
His 117			Gly 88	Leu 68	
Met 121			Pro 115	Val 86	
Met 13			Gly 116	Ile 87	
Met 44				Ser 113	
Asn 47				Ala 119	
Phe 114				Met 120	

(a)   Residues where the centroid of the sidechain is within $\sim 6\text{Å}$ of the copper atom.

(b)   Residues where the centroid of the sidechain is within $\sim 10\text{Å}$ of the copper atom.

(c)   in molecule II only.

The table shows that, of the 25 (26) residues with the centroids of their sidechains within approximately 10Å of the copper site, 14 (15) are invariant, while 9 are semi-conserved.

4.2.3   The Hydrophobic Patch around His 117.

One of the most notable features of the azurin molecule is the

hydrophobic patch on the "northern" surface of the molecular envelope

around the copper site and His 117 (Figure 4.13).
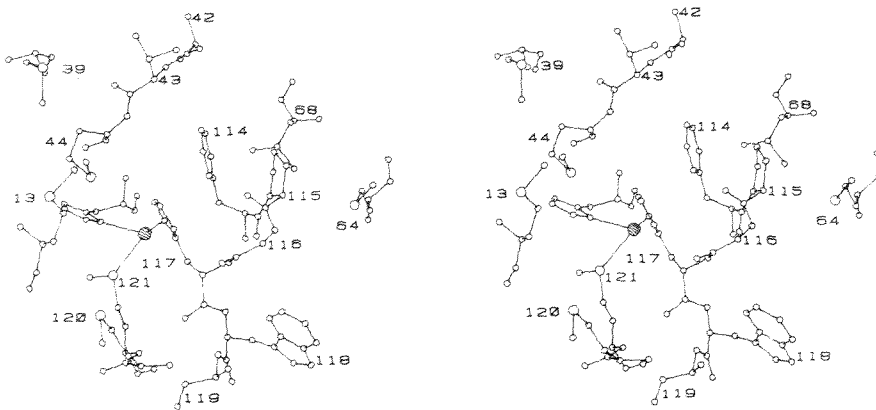
Figure 4.13   A Stereo Diagram of the Sidechains Making up the

Hydrophobic Patch Around the Sidechain of His 117.

Of the 11 sidechains making up this surface, 5 are invariant

(Met 13, Met 44, Phe 114, Pro 115 and Gly 116), and 5 are conservatively

substituted (Met 39, Ala 42, Val 43, Met 64 and Met 120). The area

is roughly circular with a radius in the vicinity of 7-8Å. The

tendency of large hydrophobic sidechains such as Val, Leu and Phe

to be buried rather than to be found on the surface of a protein

molecule is well established (105,171). The grouping together

of such a large number of these residues on the surface of the

azurin molecule may therefore have a functional significance,

as it might normally be expected to contribute to instability in the molecule as a whole (170). Such a view is encouraged by the fact that all that separates the copper from the solvent in this region is the sidechain of His 117, the edge of the imidazole ring being situated at the base of a hollow formed by the sidechains of Met 13, Met 44, Val 43, and Pro 115. Such a situation is reminiscent of the porphyrin ligand of cytochrome c which is wrapped in a hydrophobic pocket with one edge of the heme group exposed.

Unusual features of the patch are the large number of methionine residues (5) in such a small area of structure, and the orientation of the carbonyl oxygens of residues 42, 43, 115 and 116, all of which point towards the centre of the patch and His 117 (Figure 4.1). Pro 115 and Gly 116 are part of a short segment of mainchain involved in a reverse turn, and thus subject to conformational restraints. This means the carbonyl oxygens of these residues are not likely to change their configurations. Residues 42 and 43 on the other hand are part of the loop between strands III and IV, which was difficult to fit to the density. Although there is a possibility of a hydrogen bond between the carbonyl oxygen of Ala 40 and the amide nitrogen of Ala 42, (a hairpin bend of only 3 α-carbon atoms (190)),the loop is not otherwise constrained. This loop could be rather flexible, and the conformation of the carbonyl oxygens of Ala 42 and Val 43 may change on refinement. As charge interactions have previously been implicated in the binding of small molecule redox couples to azurin (191,192),and are thought to play a part in the reaction with its physiological partner, cytochrome c-551 (55,58), the hydrophobic patch could form a binding site,with recognition perhaps aided by a concentration of electronegative carbonyl oxygens.

An interaction of this sort is in fact seen in the packing of the two azurin molecules in the crystallographic asymmetric unit, which are arranged so that these "patches" on each molecule match,

as shown below (Figure 4.14).
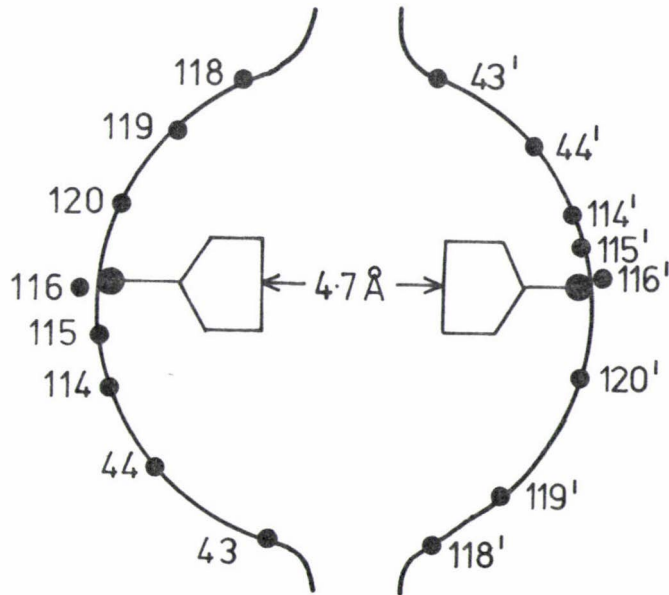
Figure 4.14   A Schematic Diagram of the Hydrophobic Patches on

the Two Azurin Molecules in the Crystallographic

Asymmetric Unit as they Pack in the Unit Cell.

Such packing presumably lowers the energy of the system by removing this hydrophobic patch on the surface of the protein molecules from contact with the solvent. Kauzmann (193) has estimated that for every non-polar hydrophobic sidechain of a protein that is removed from an aqueous to a non-polar environment, 4kcal of free energy stabilization is gained by the protein. It also suggests that azurin might dimerise in solution, although there have been no reports of this.

4.2.4   Individual Residues of Particular Interest.

(i) Histidine 35

His 35 is invariant in all azurins so far sequenced (32), and is of particular interest because it is oriented so that its

imidazole ring is almost parallel to the imidazole ring of His 46

(which provides a ligand for the Cu), although they do not quite

overlap. The sidechains of both these His residues are constrained

by hydrogen bonds, and, in the case of His 46, the Cu-ND1 bond.

Thus His 35 is hydrogen bonded through ND1 to the amide nitrogen

of Gly 37, and through NE2 to the carbonyl oxygen of Met 44. His 46

is bonded to Cu through ND1, and hydrogen bonded to the carbonyl

oxygen of Asn 10 through NE2. Figure 4.15 is a stereodiagram

showing the orientation of these two imidazole rings with respect

to each other.
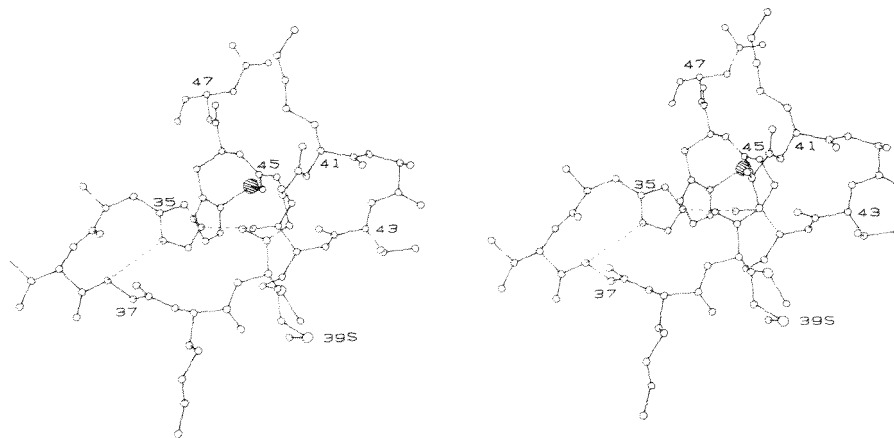
Figure 4.15   A Stereo Diagram Showing the Relative Orientations

of His 35 and His 46.

There have been suggestions that because of their relative

orientations, passage of an electron would be possible through a

conjugated system between the two His sidechains, and that His 35
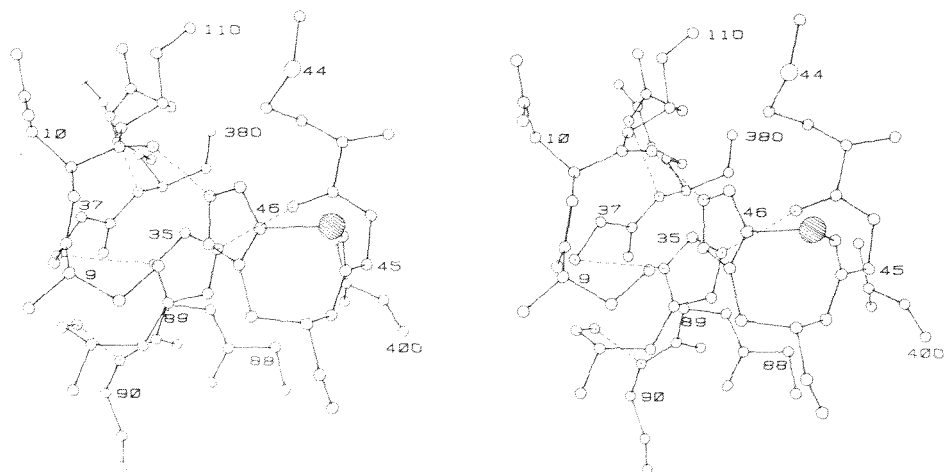
Figure 4.16   A Stereo Diagram of the Environment of His 35.

may be therefore involved in the mechanism of electron transfer

for azurin (194).   Such a hypothesis has received some supporting evidence

from b●th kinetic and affinity labelling experiments carried out on Ps.

aeruginosa azurin (192,194).   In the present structure it appears

that any interaction between these histidine sidechains through

overlap of the π orbitals of the two heterocyclic rings would be

inefficient, as not only is the distance between the centroids of

the rings approximately 4.2Å (the π electron density is expected to extend

1.7-1.85Å above the ring (195,196)), but the angle between the

planes containing them is approximately 26°, so that overlap is

restricted to one edge.   Both sidechains are very well resolved in

the electron density map, making it unlikely that their respective

orientations will change very much on refinement.   Also, His 35 is

not exposed to the solvent, being covered by the mainchain atoms of

residues 36-39 and 87-91. It is possible, however, that the binding of an electron transfer partner to azurin could induce small structural changes that would allow the delocalization of an electron on to the imidazole ring of His 35 and effective transfer of the electron through the proposed π relay system to the copper.

A further feature of the environment of His 35 is its highly conserved nature. Of its 12 nearest neighbours (whose sidechains are within approximately 7Å) 7 are invariant (Asp 11, Gly 37, Met 44, Gly 45, His 46, Gly 88, and Gly 90), 2 are conservatively substituted (Asn 10 and Met 39), and 3 vary (Ser 9, Val 36 and Gly 89). Of these three, Ser 9 and Gly 89, although not strictly conservatively substituted, are both always Gly, Ala or Ser. It is interesting to note that the only charged residue of this group, Asp 11, is one of two negatively charged residues on the surface of the molecular envelope which do not have a compensating positively charged side-chain nearby, in contrast to most of the other charged groups in the protein. Figure 4.16 is a stereo plot showing the relationship of His 35 to some of these residues around it.

(ii) Histidine 83

Another residue which has been implicated in the electron transfer mechanisms of Ps. aeruginosa azurin is His 83 (192). It also is in a fairly conserved environment, for it is one of the residues comprising strand V, one side of the "gap" in the β-structure already discussed in Section 4.1.1. Of its immediate neighbours, two out of three, Val 73 and Val 80, are only replaced by other non-polar residues. The third is either Tyr 72 (in the Molecule I interpretation) or, more probably, Asp 71 (the Molecule II interpretation). Tyr 72 is invariant, but Asp 71 can be replaced by Asn, Glu or Gln. Of the next nearest neighbours (whose sidechains are within a radius of approximately 7-8Å), Asn 47, Val 49, Ala 82 and Thr 84 are invariant, while the sidechain of Trp 48 is semi-conserved, in that it is only replaced by hydrophobic sidechains, and Lys 85 is variable,

being replaced by Ser in only one of the nine azurins so far
sequenced (32). Figure 4.17 is a stereo diagram of His 83 and its
immediate environment.
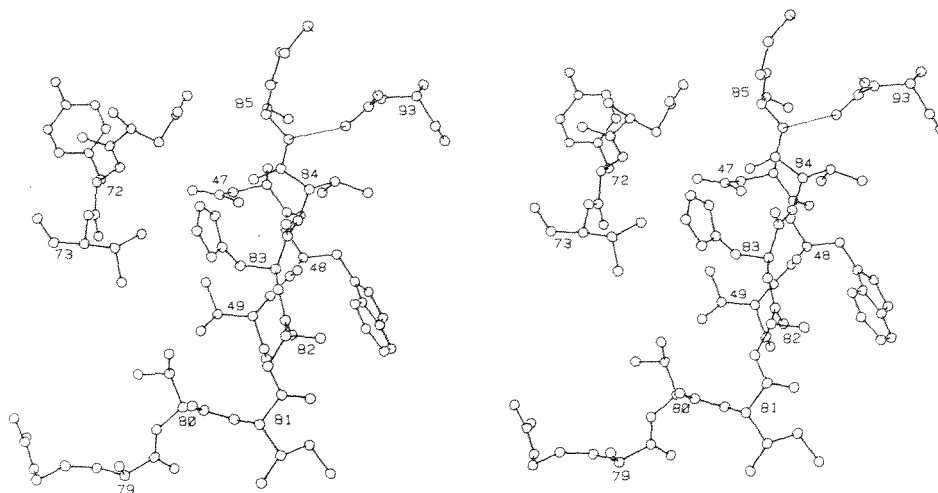
Figure 4.17   A Stereo Diagram of His 83 and its Immediate

Environment.

The orientation of the imidazole sidechain of His 83 is such
that the ND1 and NE2 atoms are not in a position to form hydrogen
bonds with any neighbouring groups. Thus it should be free to rotate
about its CB-CG bond. Although it will be largely uncharged at
physiological pH it has been shown that protonation of the His 83
sidechain has a dramatic effect on the rate of reaction of the
oxidant $[Co(phen)_3]^{3+}$ with the reduced protein (192). Whether this
effect is observed because the binding site is near to the sidechain,
or because there is a conformational change associated with its
protonation that is essential to the electron transfer mechanism,
remains to be seen.

(iii)   Asparagine 47

Asn 47 is invariant in all azurins and plastocyanins so far sequenced (32), and because Asn sidechains are rarely found buried in protein structures (171), it may have some functional significance. It is in a unique position in the molecule,as it links the immediate environment of the copper atom with the opposite side of the molecule (the face containing His 83 and the "gap" between the β strands V and VI) through a series of hydrogen bonds,including one to the thiolate sulphur of Cys 112. As well as this,hydrogen bonds between its sidechain atoms and the mainchain atoms of strands V and VII link these strands together, thus helping to maintain the tertiary structure of the molecule. A schematic representation of the hydrogen bonding of this residue is shown below.

(Ser 113) NH

(Ser 113) OH

O = C (Thr 84)

O

NH$_2$

C

CH$_2$

(Tyr 15) OH- - - - O == C —— CH —— NH - - - - - SG (Cys 112)
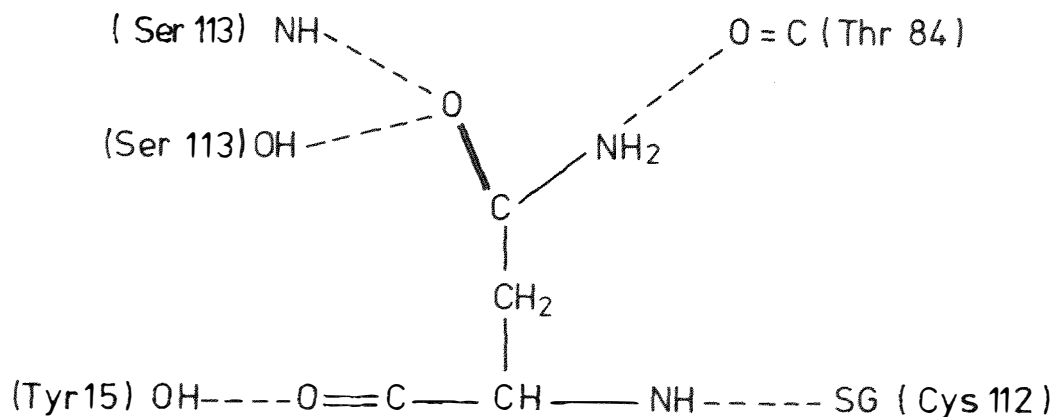
Figure 4.18   A Schematic Diagram of the Hydrogen Bonds Formed

by Asn 47.

Perhaps the most interesting of the hydrogen bonds is the

interaction between the mainchain amide nitrogen of Asn 47 and the thiolate sulphur of Cys 112. This part of the structure was represented by well resolved density in both molecules, so is unlikely to alter to any great extent on refinement. The peptide involved in this proposed hydrogen bond has its configuration further restrained by a hydrogen bond between the carbonyl oxygen of His 46 (one of the Cu ligands), and the amide nitrogen of Ile 87. The sidechain of Asn 47 is also fixed in position by hydrogen bonds. The amide oxygen (OD1) is in a position to form a bifurcated hydrogen bond with both the NH and OH groups of Ser 113, while the sidechain NH_2 is hydrogen bonded to the carbonyl oxygen of Thr 84. Ser 113 is Ser in all azurins so far sequenced except one, in which it is replaced by Thr (32). This means that the hydrogen bond between the OD1 of Asn 47 and the OH group of residue 113 will probably be maintained in all azurins.

Previous spectroscopic evidence has indicated that an amino acid residue interacts with the cysteinyl sulphur (34). If the copper atom is in, or near, the transition state between a Cu(II) and Cu(I) complex because of the particular co-ordination geometry and ligands enforced on it by the protein, the presence of such a hydrogen bond between the thiolate sulphur (S^-) and NH may act to stabilize the Cu(II) state with respect to Cu(I). A similar inter-action, between a tyrosine hydroxyl and the thioether sulphur of Met 91, the sixth ligand to the iron atom of Rhodospirillum rubrum cytochrome c_2, is thought to stabilize the oxidised state of the protein with respect to the reduced state (197).

(iv) Tyrosine 108

Another rigorously conserved residue in all azurins (and plastocyanins) is Tyr 108. From the diagram of the mainchain folding of azurin (Figure 4.8), it can be seen that Tyr 108 is situated at the "southern" end of the molecule. It is oriented so that its free hydroxyl group is at the molecular surface, and lies between a lysine sidechain (Lys 52) and a glutamic acid sidechain (Glu 106). The free hydroxyl, however, does not appear to form a hydrogen bond to either of these groups in Molecule I, although in Molecule II it does appear to be hydrogen bonded to the carboxyl group of Glu 106. Of the two neighbouring sidechains, Lys 52 is invariant in 6 out of 9 sequences published (32), being replaced by Thr in the remaining 3. Similarly, Glu 106 is conservatively substituted (being replaced only by Asp) in 8 of the 9 azurins, but is replaced by Gly in one species. Of the other neighbouring residues, Lys 101 which projects from the surface, and Leu 102 which is directed inwards to the non-polar core, are invariant, while Leu 50 and Ile 81, which are both internal, are conservatively substituted. Figure 4.19 is a stereo plot of the environment of Tyr 108.

A hydrophobic "channel" exists between this tyrosine and the copper site, formed by the sidechains of residues 7, 15, 31, 33, 48, 50, 81, 95, 97, 102, 110, 112, 121, 125, which make up the hydrophobic core of the molecule. They are all invariant or conserved in their hydrophobic character in all azurins so far sequenced. The distance between the aromatic ring of Tyr 108 and the copper atom is, however, very long at approximately 20Å which makes it unlikely as an electron transfer locus. Despite this, it is interesting to note that Glu 106 appears to be one of the binding sites for Cr(II) ions, which bind to the protein in the crystal and cause its reduction (see chapter 6).
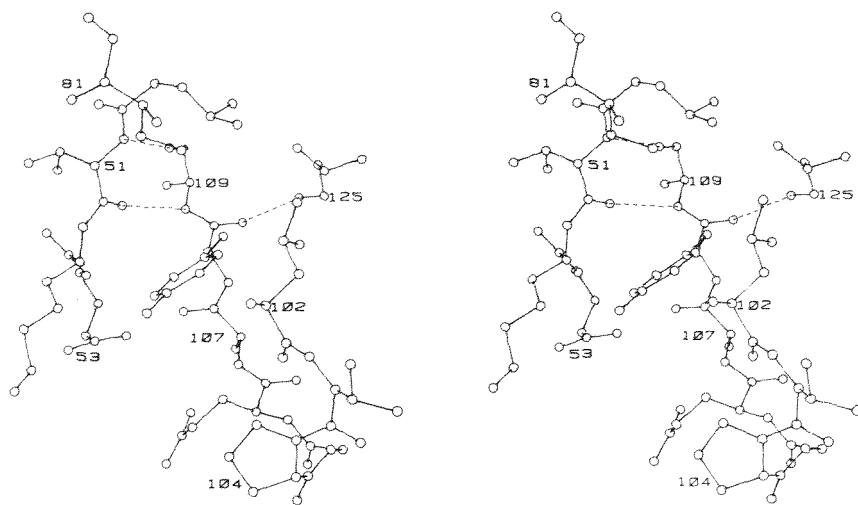
Figure 4.19   A Stereo Diagram of the Environment of Tyr 108.

4.2.5   The Non Polar Core

Sidechains with hydrophobic character, such as those of Ala,
Val, Leu, Ile, Phe, and Trp, are usually found in the interior of
protein structures (170,171). This is because they have little or
no attraction for water molecules, and can closely approach each
other with the exclusion of water to form a tightly bound structure.
The current view is that the water structure around the protein
molecule becomes less ordered when it is not disturbed by non-polar
groups. Removal of the non-polar groups therefore results in an
increase in entropy, owing to the increase in motion and the number
of arrangements possible for the water molecules in the immediate
vicinity of the protein molecule. Thus the stability of the system
is increased, the phenomenon being known as "hydrophobic bonding"
(105).

In the structure of azurin, all the residues in the interior of the β-barrel are invariant or conservatively replaced. All are hydrophobic, except for 2 Tyr and 6 neutral polar residues which are discussed in section 4.2.6. The residues making up the hydrophobic core are listed in Table 4.6 and are shown in a stereo diagram in Figure 4.20. This shows the arrangement of the sidechains within the hydrophobic core, and shows the hydrophobic channel consisting of the sidechains of residues 7, 31, 33, 48, 50, 81, 95, 97, 102, 121 and 125.
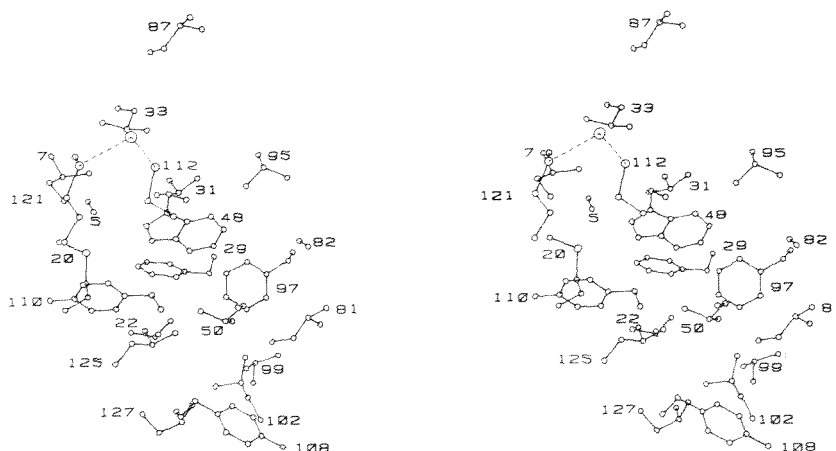
Figure 4.20   A Stereo Diagram of the Sidechains Making up the
              Hydrophobic Core.

4.2.6   Internal Polar Sidechains

The tendency of polar residues to be found on the outside of the protein molcule is more pronounced than the tendency of

hydrophobic residues to be found only on the inside, but there are

frequent exceptions, especially in the case of serine and threonine,

where about one residue in five is found buried (171). Asn and Gln

are found less frequently in protein interiors, while buried

charged sidechains (Asp, Glu, Lys and Arg) are very rare. Where

residues such as Ser, Thr, Asn, and Gln are found in the interior,

Table 4.6   Internal Residues in the β-barrel of Azurin.

	Invariant		Conservatively Substituted		Variable
Phe 29	Thr 84	Ala 5	Leu 50	Ser 9
Leu 33	Val 99	Ile 7	Ile 81	
His 35	Leu 102	Tyr 15	Ile 87	
His 46	Met 121	Met 20	Val 95	
Asn 47	Leu 127	Val 22	Phe 97	
Ala 82		Val 31	Tyr 110	
		Trp 48	Ser 113	
			Leu 125	

they always appear to be hydrogen bonded (170) so that, together

with Tyr, His, and Trp sidechains, they may augment the internal

hydrogen bonding scheme by helping to saturate the hydrogen bonding

potential of the mainchain carbonyl oxygen and amide nitrogen atoms.

There are no buried charged sidechains in the azurin molecule, but

there are 8 uncharged polar sidechains in the internal region.

(i) Serine 9

The hydroxyl group of this residue does not appear to be in a

position to form a hydrogen bond with any mainchain or sidechain

atoms, although there are close approaches to both the imidazole

ring of His 46 and the terminal carbon atom of the Met 121 sidechain.

This is unusual, but it should be noted that Ser 9 is not invariant, and can be replaced by Gly in other azurins (32).

(ii) Tyrosine 15

The sidechain of Tyr 15 points towards the centre of the molecule, being hydrogen bonded through its hydroxyl group to the carbonyl oxygen of Asn 47.

(iii) Histidine 35

His 35 is hydrogen bonded to the amide nitrogen of Gly 37 through ND1 and to the carbonyl oxygen of Met 44 through NE2 (see previous discussion in Section 4.2.4).

(iv) Histidine 46

His 46 provides a ligand for the copper atom (through ND1) and is hydrogen bonded through NE2 to the carbonyl oxygen of Asn 10 (see also the previous discussions in sections 4.2.1 and 4.2.2).

(v)   Asparagine 47

Asn 47 is hydrogen bonded through its OD1 to both the NH and OH groups of Ser 113, and through its ND2 to the carbonyl oxygen of Thr 84 (see previous discussion in section 4.2.4).

(vi) Threonine 84

This residue is invariant in all azurins and is part of the highly conserved "gap" in the β pleated sheet. The hydroxyl group appears to be in a position to form a hydrogen bond with the OD1 of Asp 93 which also is invariant in all azurins.

(vii) Tyrosine 110

The hydroxyl group points towards the outside of the molecule, but is not exposed to the solvent, and is in fact hydrogen bonded to the peptide carbonyl of Lys 18.

(viii) Serine 113

Ser 113 is invariant in 8 out of 9 azurin sequences, being replaced by Thr in the remaining one (32). The hydroxyl group forms a hydrogen bond with the sidechain amide group of Asn 47.

4.2.7 The Interface Between the β-barrel and the "Flap"

Most of the residues lying between the flap and the β-barrel are hydrophobic. Furthermore, out of 18 residues, 15 are either invariant or semi-conserved, these being listed in Table 4.7. The dotted lines indicate which residues make contact across the interface. Figure 4.21 is a stereo diagram of this region.

Table 4.7   Conserved Residues Between the β-barrel and the Flap.

INVARIANT		CONSERVATIVELY SUBSTITUTED	
From the Barrel	From the Flap	From the Barrel	From the Flap
Val 49 –––			Val 59
His 83 –––			Val 73
	Asp 55 –––––––––	Thr 51	
		Val 80	
		Leu 68 –––––––––	Val 86
Phe 111 ––––––––––––––––––––––––––––––––––––––			Ala 60
Phe 114 ––––––––– Tyr 72			
Pro 115 ––––––––––––––––––––––––––––––––––––			Met 64
		Ser 113	

In addition, Asp 71 and Lys 85, both external, face one another across the interface, but both these residues are quite variable in different species.

Because several of the loops in the flap region were difficult to build, being fitted with different conformations in the two molecules (i.e. residues 68-72 and 77-79), it is difficult to describe with certainty links between the flap and the barrel apart from
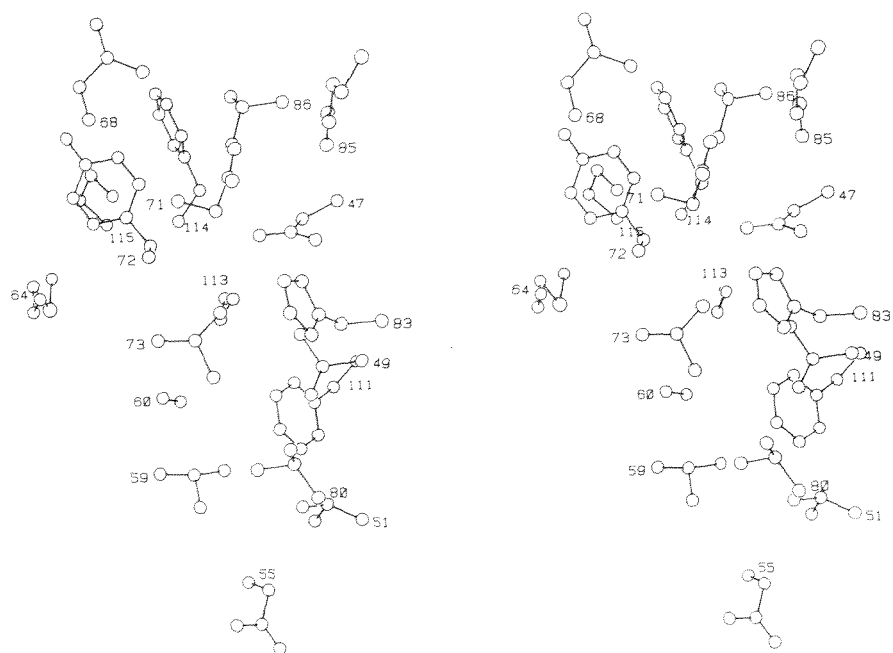
Figure 4.21   A Stereo Diagram of the Residues Making up the

Interface Between the β-Barrel and the Flap.

general hydrophobic interactions. At this stage, all that can be

said is that two possible hydrogen bonds can be seen between the

carbonyl oxygen of Tyr 72 and the hydroxyl group of Ser 113, and

between the hydroxyl group of Thr 51 and the carbonyl oxygen of

Arg 79.

4.2.8   The Disulphide Bridge.

There is one disulphide bridge in the structure, formed on the

molecular surface between Cys 3 and Cys 26 at the ends of strands I

and III of the β-barrel. These two residues are conserved in all

azurins (32), the bridge probably acting to stabilize the rather wide

loop between strands I and III   (see Figure 4.8).   The exact

conformation of the disulphide bridge is not clear at this stage, since

the density in this region was somewhat distorted and difficult to fit.

4.2.9 Charged Sidechains.

There are a large number of charged sidechains in azurin, 15 acidic and 14 basic, out of a total of only 129 residues. All are found on the surface of the molecule where they can interact with the solvent. Although the exact orientation of many of these side-chains is by no means certain (particularly in the case of some lysine sidechains), a few of them appear to interact with each other to form salt bridges. Such contacts, which stabilize the conformation of the protein, are indicated in Table 4.8 by solid lines linking the residues. Dotted lines between residues indicate that they are too far apart for a hydrogen bond, but are in reasonable proximity (4-6Å apart).

Table 4.8   The Surface Charged Residues

Invariant or Semi-Conserved		Variable	
Acidic	Basic	Acidic	Basic
		Glu 4 ---- His 32	
		Glu 8 ---- Lys 34	
Asp 11			
		Asp 16	
		Glu 19 ——— Lys 18	
			Lys 38
		Glu 53 ---- Lys 56	
		Glu 57	
Asp 55 ————————————————————————— Lys 52			
		Glu 106	
Asp 62 ——— Lys 74 and Arg 79			
		Asp 71 ---- His 83	
Asp 77 -------- Lys 74 and Arg 79			
Glu 91 -------- Lys 41			
Asp 93 ——————————————————————————— Lys 85			

Table 4.8   Continued

Invariant or Semi-Conserved		Variable	
Acidic	Basic	Acidic	Basic
Asp 98 -------- Lys 101			
	Lys 27		
C terminus ---- Lys 24			
			Lys 122
			Lys 126

From the table it can be seen that, apart from two very small patches of charge (those created by Asp 11 and Asp 16, and Lys 122 and Lys 126), there are no large concentrations of positive or negative surface charge as found in some other proteins (e.g. plastocyanin (56) and cytochrome c (197)). With the exception of Asp 11, and perhaps Asp 55 and Asp 93, all the invariant or semi-conserved residues occur in pairs. Of the residues that may form ion pairs with Asp 55 and Asp 93, Lys 52 is replaced by Thr in three other azurins, while Lys 85 is replaced by Ser in only one of the nine azurins sequenced (32). Of the unpaired charged residues, only Asp 11 is rigorously conserved, although Lys 122 is replaced by Thr in only one other azurin (32). Overall there does not seem to be any evidence for binding sites involving charged sidechains; their main function is presumably to enhance the solubility of the protein (counteracting the effect of the large non-polar surface patch around His 117).

4.2.10   The Distribution of Neutral Polar Sidechains on the Surface
          of the Molecule.

Most of the neutral polar sidechains found on the molecular surface of azurin may be substituted by other types of sidechains (non-polar or charged) in other azurins. Some, however, are invariant

and may be important to the structure of the protein. Table 4.9
lists the uncharged polar sidechains of azurin found on the molecular
surface.

Table 4.9   The Uncharged Polar Sidechains on the Molecular Surface
              of Azurin.

Invariant or Conservatively Substituted		Variable	
Ser 25	Ser 94	Thr 6	Ser 92
Thr 30	Thr 96	Gln 28	Ser 100
		Thr 61	Thr 103
Asn 10	Gln 14	Asn 65	Thr 124
		Gln 70	Ser 128
	Thr 51	Thr 78	

Asn 10 and Gln 14 are found to one side of the "hydrophobic
patch" (section 4.2.3), in close proximity to each other on opposite
sides of the loop between strands I and II. Hydrogen bonding between
the sidechain groups of these residues and mainchain atoms of the
strands are important in maintaining the conformation of the loop.
Hence ND2 of Asn 10 points towards the carbonyl oxygen of Gln 14,
while NE2 of Gln 14 can form a hydrogen bond with the carbonyl oxygen
of Tyr 15. This latter hydrogen bond may be important to the
maintenance of the kink observed in strand II of the β-structure
near Tyr 15.

        Thr 30, Ser 94 and Thr 96 are all residues involved in the
regular β-structure between strands III and VI. There is a hydrogen
bond between the hydroxyl groups of Thr 30 and Thr 96, and also
between the hydroxyl group of Ser 94 and NE2 of the adjacent His 32
(which is always polar in other azurins). Such hydrogen bonds may
augment or protect the hydrogen bonding of the β-structure. Ser 25

is adjacent to the disulphide bridge (Cys 3-Cys 26). Its hydroxyl

oxygen appears to be hydrogen bonded to the carboxyl group of Asp 23,

possibly helping to stabilize the surface conformation. Thr 51 is

at the interface between the β-barrel and the flap and discussed

in section 4.2.9.

4.3   Intermolecular Contacts.

Once the structure had been completed, and all the residues

identified on the electron density map, it was possible to identify

the points of close contact between the molecules in the unit cell.

This was done from inspection of the "mini-maps", and by calculating

distances to other molecules. Figure 4.22 shows molecule I and its

eight nearest neighbours. These are identified by the number of their

equivalent position in the crystallographic unit cell, i.e.

(1)	x, y, z	(5)	$\frac{1}{2}+x, \frac{1}{2}+y, z$
(2)	$x, \bar{y}, \bar{z}$	(6)	$\frac{1}{2}+x, \frac{1}{2}-y, \bar{z}$
(3)	$\bar{x}, \bar{y}, \frac{1}{2}+z$	(7)	$\frac{1}{2}-x, \frac{1}{2}-y, \frac{1}{2}+z$
(4)	$\bar{x}, y, \frac{1}{2}-z$	(8)	$\frac{1}{2}-x, \frac{1}{2}+y, \frac{1}{2}-z$

and the translation of this equivalent position along the cell axes.

For example, (4, -1 1 -1) would indicate a molecule at $\bar{x}, y, \frac{1}{2}-z$,

shifted to the next unit cell in the negative x direction, the

positive y direction and the negative z direction. All molecules

of Type I are cross hatched, while Type II molecules are outlined

only. The range of z co-ordinates for each of these positions is

as follows: For molecule I:

(1) $-.125 \rightarrow .25$; (2) $.125 \rightarrow -.25$; (3) $.375 \rightarrow .75$; (4) $.625 \rightarrow .25$.

For molecule II:

(1') $.125 \rightarrow .5$; (2') $-.125 \rightarrow -.5$; (3') $.625 \rightarrow 1.0$; (4') $.375 \rightarrow 0$.

The diagram shows that each molecule of one kind is closely

surrounded by six molecules of a different kind (non-crystallographically
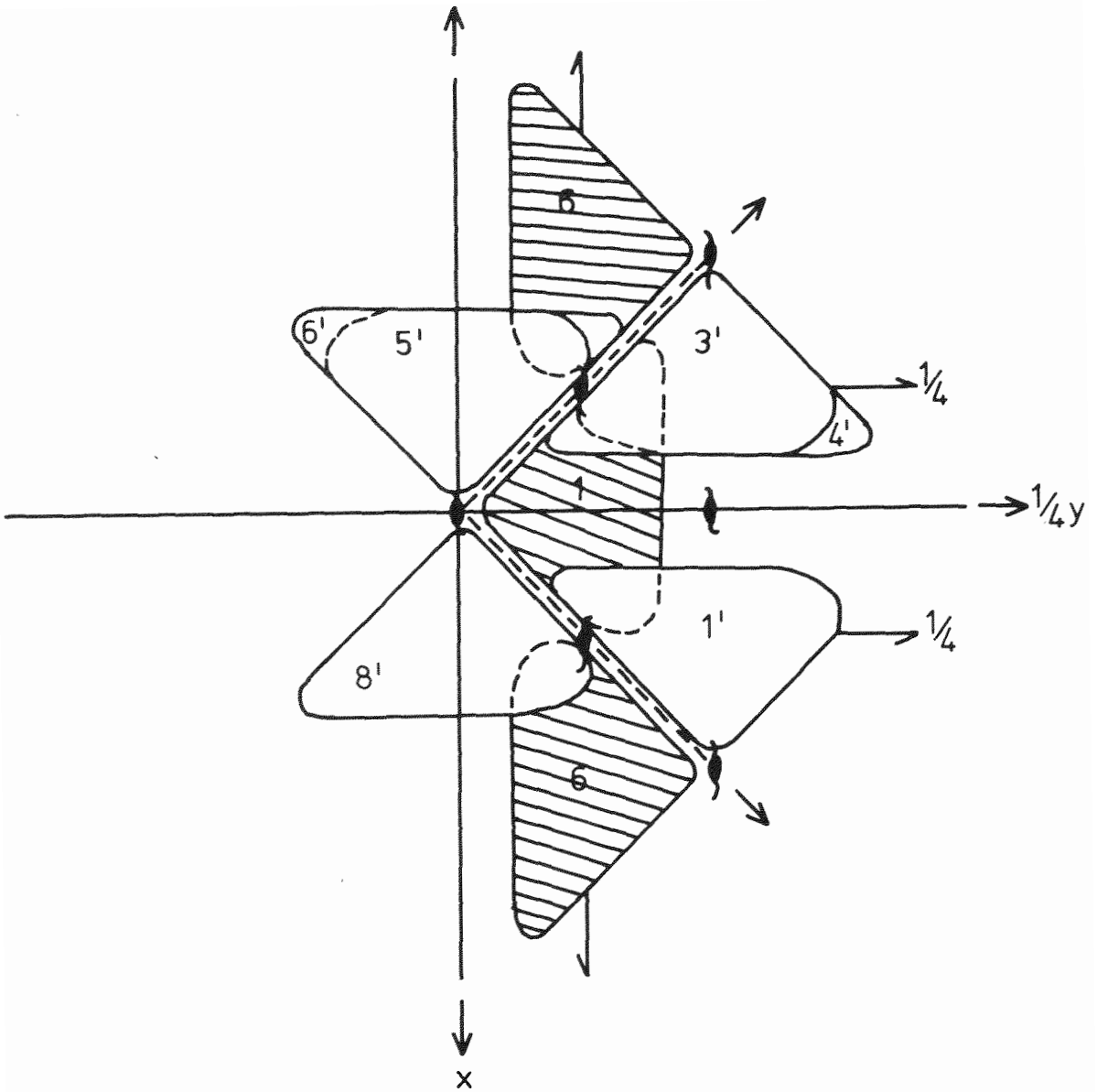
Figure 4·22 A schematic diagram of Molecule I
(cross-hatched) and its eight nearest
neighbours. These are 6 Type II
molecules (outlined) and 2 Type I
molecules. The numbers refer to the
equivalent positions in the unit cell,
and are defined as follows:

1'(1',0 0 0)   5'(5', -1 -1 0)

3'(3',0 1 -1)   6'(6', -1 0 0)

4'(4',0 0 0)   8'(8', 0 -1 0)

6 (6,0 0 0)   6 (6, -1 0 0)

related), and two of the same kind (crystallographically related),
so that the crystal is made up of groups of closely packed molecules
separated by larger solvent spaces. There are two main areas of
contact between the molecules.

The first involves the hydrophobic patch surrounding the edge
of the imidazole ring of His 117. Molecule I faces a type II
molecule (8', 0 -1 0) across the non-crystallographic 2-fold axis
in such a way that the two hydrophobic patches are in contact. Contacts
are both heterologous (i.e. involve residues of differing sequence
number such as : Met 44, Met 13/Met 120' ; Val 43/Trp 118' ;
Met 120/Met 13' ; Trp 118/Val 43') and homologous (i.e. involve
residues of the same sequence number such as : Met 64/Met 64' ;
Pro 115/Pro 115'). These contacts are all between 3.0 and 3.5Å. This
packing of the two molecules leads to a distance of $\sim$ 4.7Å between
the two His 117 ring edges.

The second involves the flap region of the molecule. This is the
region which was involved in the overlapping density between two
non-crystallographically related molecules at low resolution, and
where a cut had to be made to separate the two. Residues 65-70 of
Molecule I lie adjacent to the loops between strands II' and III',
and VI' and VII' (involving residues 23'-25' and 99'-103') of a type II
molecule (1', 0 0 0), as well as Asp 23, Ser 25 and Asn 129 of another
type I molecule (6, 0 0 0). Similarly, residues 23-25 and 99-103
of Molecule I are adjacent to residues 65'-70' of the flap of a
type II molecule (3', 0 1 -1) while residues 23, 25 and 129 are next
to residues 65-68 of another type I molecule (6, -1 0 0).

Other regions of contact between the molecules involve fairly
small local areas with few contacts of less than 3.5Å. Thus the
sidechains of residues 34, 91, 92 and 94 of the β-strands III and VI
of Molecule I, which are found at the "northern" end of the structure
(Figures 4.4 and 4.8) are adjacent to residues 2', 4', 27' and 32'
of strands I' and III' near the "southern" end of a non-crystallographic-

ally related molecule (4', 0 0 0). Ser 128 of Molecule I is adjacent
to Asn 129'of (6', -1 0 0), both being carboxy-terminus residues,
and Lys 38, situated at the top of the molecule, is near Ala 119'
and Lys 122' of (5', -1 -1 0) which are found immediately below
the hydrophobic patch of the type II molecule.

The contacts involving Molecule II are essentially the same
as those for Molecule I although there are differences in detail
resulting from the inexactness of the non-crystallographic diad,which
results in the two molecules in the asymmetric unit being slightly
differently oriented with respect to the crystallographic symmetry
elements. For example, residues 38-40 approach residues 40-38 of a
molecule of the same type across a crystallographic two-fold axis.
For molecule I, this distance is greater than 5Å, whereas in Molecule
II, the shortest contact is of the order of 3Å. Similarly the approach
of the sidechains of Glu 91 residues of molecules of the same type
across a crystallographic diad is different in the two molecules. Thus
the shortest contact between these sidechains for Molecule I is
greater than 8Å, while for Molecule II it is ∿ 6Å.

Intermolecular Salt Bridges.

There appears to be one salt bridge linking molecule I with a
type II molecule, viz. that between the carboxyl group of Glu 91
of Molecule I and the amino group of Lys 27' from (4', 0 0 0). There
is not an equivalent link between Glu 91' of Molecule II and a Type I
molecule, as again the inexactness of the non crystallographic
symmetry results in the environment of the two molecules being slightly
different. For Molecule II, the distance between the carboxyl group
of Glu 91' and the amino group of Lys 27 from a diad-related type I
molecule is greater than 4.5Å.

Intermolecular Hydrogen Bonds.

Possible intermolecular hydrogen bonds are listed in Table 4.10. These are only tentative, as they are based on criteria of distance alone, without regard for the geometry of the bond.

Table 4.10   Possible Intermolecular Hydrogen Bonds

Molecule I	Molecule II	N	Molecule I	Molecule I	N
Lys 24 (NZ)	– (C=O) Asn 65'	(3')	Asp 23 (OD1) –	(ND2) Asn 65	(6)
Lys 34 (NZ)	– (NE2) Gln 2'	(4')			
Lys 38 (NZ)	– (C=O) Ala 119'	(5')			
Gln 70 (C=O)	– (OG) Ser 100'	(1')			
Asn 65 (C=O)	– (NZ) Lys 24'	(1')			
Ser 92 (OG)	– (OE1) Glu 2'	(4')			
Ser 92 (OG)	– (OE1) Glu 4'	(4')			
Ser 94 (OG)	– (NE2) His 32'	(4')			
Gly 116 (NH)	– (C=O) Pro 115'	(8')			
Ser 128 (C=O)	– (ND2) Asn 129'	(6')			

Molecule II	Molecule I	N	Molecule II	Molecule II	N
Gln 2' (OE1)	– (OG) Ser 92	(4)	Asp 23' (OD1) –	(ND2) Asn 65'	(8')
Gln 2 (NE2)	– (NZ) Lys 34	(4)			
Gln 4' (OE1)	– (OG) Ser 92	(4)	Lys 38' (C=O) –	(NH) Ala 40'	(2')
Lys 24' (NZ)	– (C=O) Asn 65	(1)	Ala 40' (NH) –	(C=O) Lys 38'	(2')
His 32' (NE2)	– (OG) Ser 94	(4)	Asn 65' (OD1) –	(OE) Asn 129'	(8')
Asn 65' (C=O)	– (NZ) Lys 24	(3)	Asn 129' (C=O) –	(ND2) Asn 65'	(8')
Ser 100' (OG)	– (C=O) Gln 70	(1)			
Pro 115' (C=O)	– (NH) Gly 116	(8)			
Ala 119' (C=O)	– (NZ) Lys 38	(5)			
Asn 129' (ND2)	– (C=O) Ser 128	(6)			

N refers to the equivalent position in the unit cell, as follows:

1'	(1', 0 0 0)		5'	(5', -1 -1 0)
2'	(2', 0 1 1)		6'	(6', -1 0 0)
3'	(3', 0 1 -1)		8'	(8', 0 -1 0)
4'	(4', 0 0 0)			
1	(1, 0 0 0)		5	(5, 0 0 0)
3	(3, 0 1 0)		6	(6, 0 0 0)
4	(4, 0 0 0)		8	(8, 0 0 0)

CHAPTER 5

A COMPARISON OF AZURIN AND PLASTOCYANIN

Apart from the azurins, the plastocyanins (molecular weight

10,500 (56)) are the best characterised group of blue Type I

copper proteins. Plastocyanins are found in all higher plants, in

many green algae, and in some blue-green algae (cyano-bacteria).

They act as oxidants for a membrane-bound cytochrome (cytochrome f)

and as reductants for P700, a double chlorophyll pigment in photo-

system I (for a review see reference 198).

5.1 Sequence Homology.

Amino acid sequence homologies between the two blue proteins,

azurin and plastocyanin, have previously been noted by several

workers (32,56,74). Most comparisons have, however, been made on

purely statistical grounds, by aligning the sequences so that a

significant number of identical or similar residues come into

correspondence. As the azurin molecule is larger than that of

plastocyanin (129 residues cf. 99 residues respectively), it is

obvious that substantial deletions will have to be made in its

sequence to align it with that of plastocyanin.

Now that the three dimensional structures of both azurin and

plastocyanin are available, the sequences of the two proteins can

be aligned on a proper structural basis. Comparison of the two

three dimensional structures reveals that although the topologies of

the two molecules are essentially the same, there are differences

in the size and conformation of the loops between β-strands. These

differences include the "flap" in azurin which connects strands IV

and V, and which is deleted in the plastocyanin structure. They are

clearly seen in Figure 5.1, which is a stereo diagram of the two

molecules superimposed.
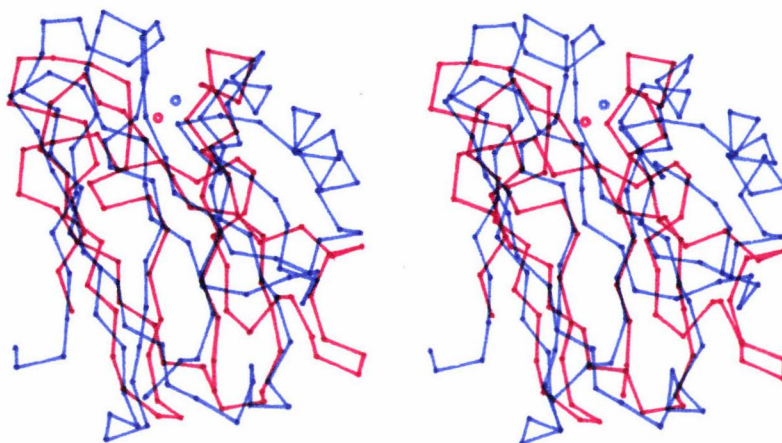
Figure 5.1   A Stereo Diagram of the Polypeptide Chains of Azurin

(blue) and Plastocyanin (red).

To obtain the best alignment, the correspondence of residues

around the copper site was first sought, since if the proteins are

indeed monophyletic this region should be highly conserved in both

proteins. In azurin, three of the ligands to the copper are provided

by a loop formed between residues 112-121, with the sequence

-Cys-x-x-x-x-His-x-x-x-Met- . Plastocyanin has a similar loop with

the sequence -Cys-x-x-His-x-x-x-x-Met-, which,despite the differences

of length and sequence,must maintain a similar stereochemistry about

the copper site to account for the similarities in absorption spectra,

redox potential and ESR parameters (13). Other residues in the

vicinity of the copper site, for example Tyr 15 (Phe 14), Asn 47 (Asn 38)

Tyr 110 (Phe 82), Phe 111 (Tyr 83) and Pro 115 (Pro 86) are highly

conserved in both structures and hence are likely to be found in
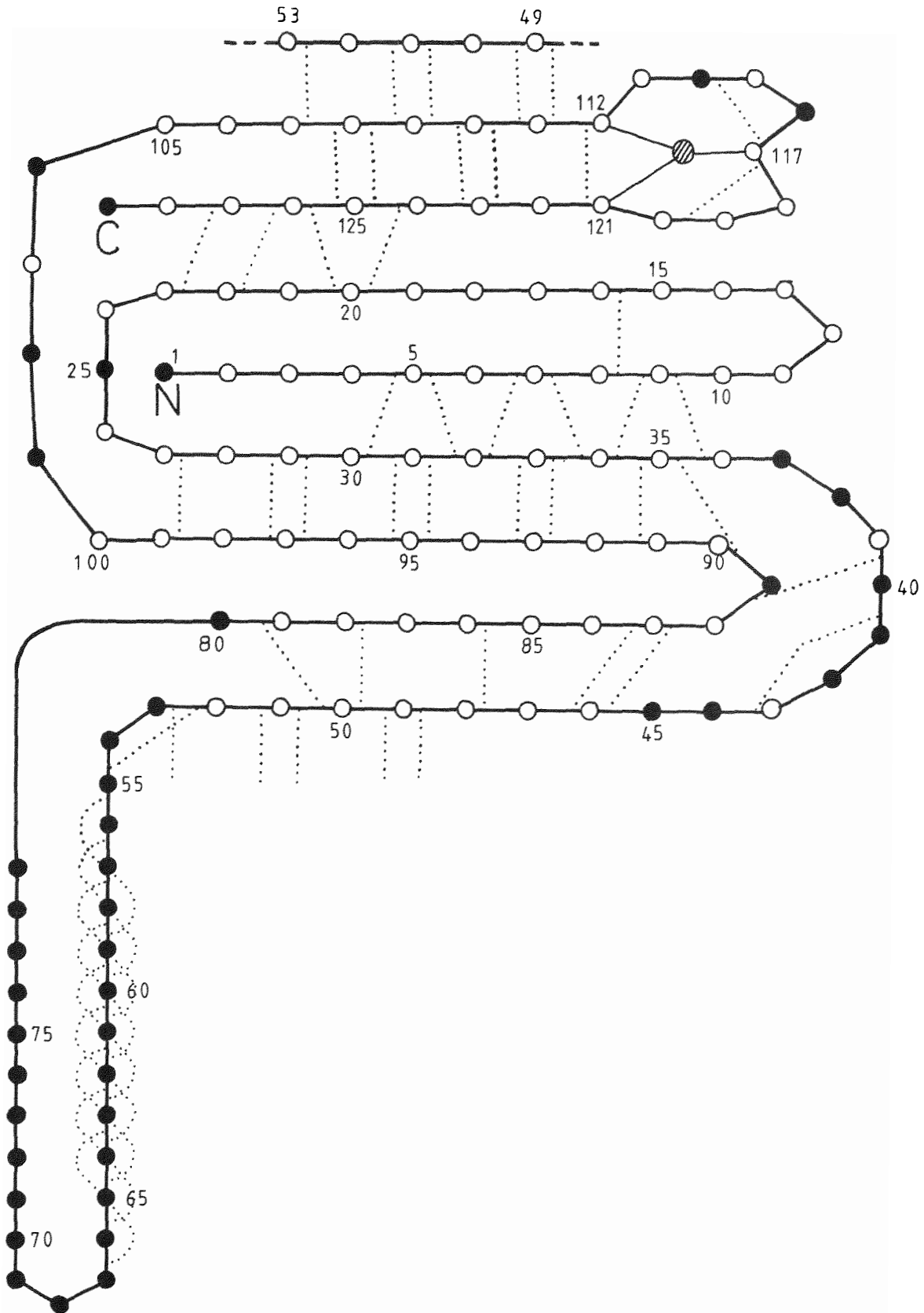
Figure 5-2. The secondary structure of azurin showing the deletions in the sequence required to give the best alignment with plastocyanin.

similar positions in both structures. Secondly, the integrity of
the β-strands was taken into consideration, as, for example, it would
be most unlikely for deletions to occur in the middle of a strand.
Finally, the hydrophobic nature of the buried residues was taken
into account, i.e. buried hydrophobic residues in azurin were
aligned with buried hydrophobic residues in plastocyanin. Apart
from the loops between the strands, the β-structure of both proteins
could be superimposed unambiguously. The loops could not be
superimposed, and as well as differing in size and conformation,
they had little structural or sequence homology. Figure 5.2 is a
diagram of the secondary structure of azurin,with the deletions in
the sequence required to bring it into alignment with the plastocyanin
structure shown as black dots.

The full sequence alignment is shown in Figure 5.3. It can be
seen that the major deletion required in the azurin sequence to bring
it structurally into alignment with plastocyanin is between residues
53 and 80, which make up the flap between strands IV and V of the
β-barrel. For the best alignment,however, a similar deletion is
required in the plastocyanin sequence between residues 45 and 53 to
account for the greater length of the fifth strand compared to that
in azurin. Other major deletions occur between residues 36 and 46,
which form an external loop between strands III and IV (7 residues),
and between residues 100 and 105, which are part of an irregular loop link-
ing strands VI and VII (3 residues). Overall, 9 residues are
invariant between all azurins and all plastocyanins. These include
Asn 47 (Asn 38) and Tyr 108 (Tyr 80),as well as the four residues
providing ligands to the copper (His 46 (His 37), Cys 112 (Cys 84),
His 117 (His 87),Met 121 (Met 92)). A further 19 are conservatively
substituted, retaining their hydrophobic or aromatic character.
The alignment is different to those of Ryden and Lundgren (32),
Dayhoff (74) and Colman et al. (56),although all of these have
a similar number of invariant residues.

Azurin
H₂N-[ALA]-GLN-CYS-GLU-ALA-THR-ILE -GLU-SER-[ASN]-[ASP]- o -ALA-[MET]-GLN- TYR-ASP-LEU-LYS -GLU-MET-VAL-VAL-

Plast.
o -H₂N-ILE -ASP-VAL-LEU-[LEU]-[GLY]-ALA-ASP-ASP-[GLY]-SER-[LEU]-ALA-[PHE]-VAL-[PRO]-SER-GLU-PHE-SER-ILE -

Azurin
ASP-[LYS]-[SER]-[CYS]-[LYS]-GLN-[PHE]-[THR]-VAL-HIS -[LEU]-LYS -[HIS]-VAL-[GLY]-LYS -MET-[ALA]-LYS -[ALA]-VAL-[MET]-[GLY]-[HIS]-

Plast.
SER-[PRO]- o -GLY-GLU-LYS -[ILE]-VAL-PHE-LYS -[ASN]-[ASN]-ALA-GLY- o - o -PHE- o - o - o -[PRO]- o - o -[HIS]-

Azurin
[ASN]-[TRP]-[VAL]-[LEU]-[THR]-LYS -GLU-ALA-[ASP]-LYS -GLU-[GLY]-[VAL]-ALA- THR-[ASP]-[GLY]-MET-ASN-ALA-[GLY]-[LEU]-ALA-GLN-

Plast.
[ASN]-[ILE]-[VAL]-[PHE]-[ASP]-GLU- o - o - o - o - o - o - o - o - o - o - o - o - o - o - o - o - o - o -

Azurin
ASP-[TYR]-[VAL]-[LYS]-ALA-GLY-[ASP]-THR-[ARG]-[VAL]- o - o - o - o - o - o - o - o - o - o - o - o -[ILE]-[ALA]-[HIS]-

Plast.
o - o - o - o - o - o - o - o - o - o - o -ASP-SER-ILE -[PRO]-SER-GLY-VAL-ASP-ALA-SER-LYS -ILE -[SER]-

Azurin
[THR]- o -LYS- o - o -[VAL]-ILE -[GLY]-GLY-[GLY]-[GLU]-SER-[ASP]-[SER]-[VAL]-[THR]-PHE-[ASP]-[VAL]-SER-[LYS]-[LEU]-THR-PRO-

Plast.
[MET]-SER-GLU-GLU-ASP-[LEU]-LEU-ASN- o -ALA-LYS -[GLY]-GLU-THR-PHE-GLU-VAL -ALA-LEU-SER- o - o -ASN- o -

Azurin
GLY-GLU-ALA-[TYR]-ALA-TYR-[PHE]-[CYS]-SER-[PHE]-[PRO]-[GLY]-[HIS]- o -TRP-ALA-MET-[MET]-LYS -[GLY]-THR-LEU-LYS -[LEU]-

Plast.
LYS -[GLY]-GLU-[TYR]-SER-PHE-TYR-[CYS]-SER- o -[PRO]- o -[HIS]-GLN-[GLY]-ALA-[GLY]-[MET]-VAL-[GLY]-LYS-VAL-[THR]-[VAL]-

Azurin
SER-ASN-COOH

Plast.
ASN-COOH

[ ] UNVARIED RESIDUES

[ ] CONSERVATIVELY SUBSTITUTED RESIDUES

● CONSERVED AROMATIC OR HYDROPHOBIC CHARACTER

FIGURE 5-3. SEQUENCE HOMOLOGY — AZURIN AND PLASTOCYANIN

Despite the difference in size, and the low level of sequence homology between azurin and plastocyanin (9% invariant, 19% conservatively replaced, based on the 99 residues of plastocyanin), compared with other proteins thought to be related by common ancestry (199), extensive regions of the structure are very similar. This could be due to evolutionary convergence, i.e. the natural selection of a set of amino acids to provide a particular structure, or divergence from a common ancestor. As there are two aspects to the structural homology between the two proteins, viz. the mainchain conformation (topology) and the copper site (function), it is almost certain that azurin and plastocyanin are evolutionarily related by divergence from a common ancestor.

5.2 A Comparison of the General Structure.

The 129 amino acid residues of azurin from Alc. denitrificans form eight strands of polypeptide chain which are linked by loops of various size. The strands are arranged in the form of a barrel, most of the residues pointed inwards being hydrophobic, and most of those on the surface being polar. The strands have varying degrees of β character, the most regular being strands I, III and VI, which have a right-handed twist, and a fairly regular pattern of hydrogen bonds between them. They contain 8 residues each, and are a distinctive feature of the structure. The hydrogen bonding between the remaining strands is less regular, with none at all between strands V and VI. There is a "kink" in strand II between residues 15 and 18, and joining strands IV and V is a "flap" of approximately 28 residues, hanging outside the body of the molecule and containing a short piece of regular helix (approximately 3 turns).

The β-sheet structure of poplar leaf plastocyanin appears to be similar in topology and conformation to that of Alc. denitrificans

azurin, although, because no hydrogen bonding pattern for plastocyanin has yet been published, this cannot be verified. There appears to be a similar kink in strand II of plastocyanin (between residues 14 and 17), but there is no "flap" joining strands IV and V. Table 5.1 is a comparison of the number of residues found in each of the 8 pieces of chain making up the strands of the β-barrel in each of the proteins.

Table 5.1   A Comparison of the Number of Residues in the Eight
             Strands of the Azurin and Plastocyanin β-Barrel.

Strand	I	II	III	IV	V	VI	VII	VIII
Azurin (Alc. denitrificans)	8	8	8	7	8	8	6	8
Plastocyanin (poplar leaf)	6	8	8	9	13	8	6	8

The elongation of the fourth, and particularly the fifth, strands in plastocyanin can best be correlated to the so-called "flap" of azurin, although, obviously, the number of residues and conformation are completely different in each structure. The fifth strand of plasto-cyanin, it should be noted, is very irregular and has no β character (106). Other "kinks" in the plastocyanin structure, between residues 42 and 45 of strand IV, and 52 and 56, and 57 and 61 of strand V, are not observed in azurin because of the lack of structural homology in these regions.

In both molecules, the copper atom is bound at the top of the molecule, between strands IV, VII and VIII (in the orientation of Figure 5.1). Azurin contains a disulphide bridge between residues 3 and 26 which is not present in plastocyanin.

5.3   The Copper Site.

(i)   The Ligands

The copper atom in plastocyanin is co-ordinated by four residues, His 37, Cys 84, His 87 and Met 92 which   are the exact analogues of the liganded residues in azurin, His 46, Cys 112, His 117, and Met 121.   Unlike the situation in azurin, however, there is no possibility for any weak interaction between the carbonyl oxygen of Pro 36 (the analogue of Gly 45) in the plastocyanin structure, as it is further away from the copper atom (3.73Å), despite being oriented in the same way.   This can be seen in Figure 5.4, a stereo diagram showing the "best fit" of residues 30-48 of azurin with the corresponding region of the structure in plastocyanin.
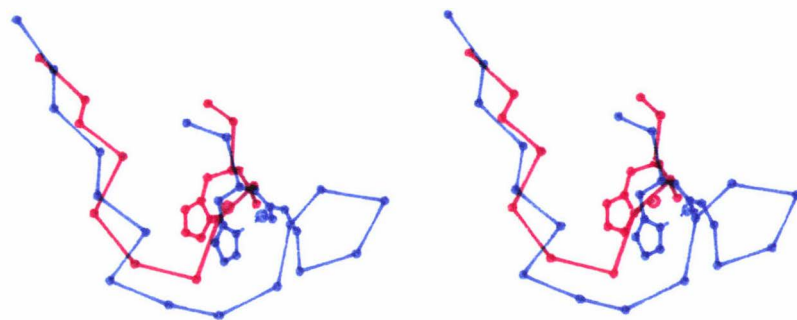
Figure 5.4   A Stereo Plot of the Polypeptide Chain Between Residues 30 and 48 of Azurin (blue) and the Corresponding Residues of Plastocyanin (red).

The Cu atoms are represented by large circles.

The sidechains of His 46, His 37 and the carbonyl groups of Gly 45 and Pro 36 are represented by filled circles.

Figure 5.4 shows that in plastocyanin the loop between strands III and IV is much shorter than that in azurin. In plastocyanin only 3 residues link His 37 to Asn 33, whereas in azurin, 10 residues link their analogues, His 46 and His 35. The obvious lack of structural and sequence homology in this region (Figure 5.3), makes it all the more remarkable that the peptide bonds linking residues 45 and 46 in azurin, and 36 and 37 in plastocyanin seem to maintain the same conformation. The conformation of this region of the structure in azurin, (the area around the copper site), is rigidly maintained by a network of mainchain-mainchain hydrogen bonds, including one between the peptide amide (46 NH) and the carbonyl oxygen of Ile 87. Although no hydrogen bonding details are available for plastocyanin, the presence of Pro at position 36 in plastocyanin probably directs the conformation of the loop between Asn 33 and His 37, resulting in this difference in the positions of the carbonyl oxygens in both structures. These small differences in the copper environment may be responsible for the differences in redox potentials and spectral properties noted for various blue copper proteins (see Table 1.2). It is known, for example, that replacing a sulphur ligand by oxygen or nitrogen reduces the redox potential of small molecule complexes (200).

(ii) The Co-ordination Sphere.

Figure 5.5 clearly shows that although the sidechain and ND1 atom of His 117 (azurin) and His 87 (plastocyanin), as well as SG of Cys 112 (azurin) and Cys 84 (plastocyanin), are virtually in the same positions with respect to the copper atom, Met 121 and His 46 in azurin appear to have been displaced around the co-ordination sphere (relative to Met 92 and His 37 in plastocyanin), to "make room" for the carbonyl oxygen of Gly 45. Although the resolution of the azurin map (3Å) means that the positions of these residues are not precise,
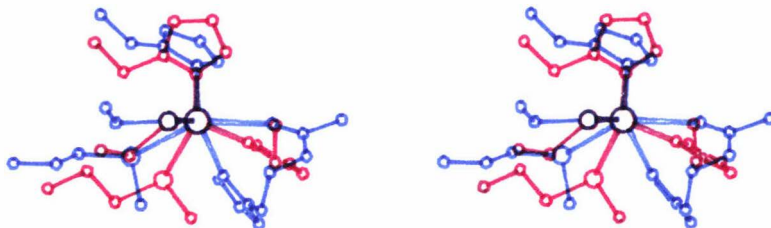
Figure 5.5   A Stereo Diagram of the Co-ordination Spheres of

Azurin (blue) and Plastocyanin (red).

it is unlikely that the geometry of the co-ordination sphere will

alter markedly on refinement, as the electron density around the

copper site in both molecules in the asymmetric unit was very well

resolved. Apart from this, it is obvious that the bond lengths of

the four common ligands are comparable within the accuracy of the

structure determination, and the geometry of the azurin site, if

the carbonyl oxygen is not considered as a ligand, although more

distorted than that of plastocyanin,could still be described as

distorted tetrahedral. The degree of distortion is indicated by the

fact that,for both azurin and plastocyanin, the copper atom is virtually

in the plane formed by the three shortest ligands, $N_{46(37)}$, $S_{112(84)}$,

$N_{117(97)}$ (0.003Å out of the plane in molecule I, 0.105Å in molecule II,

and 0.343Å in plastocyanin).

(iii)   The Environment of the Copper Atom.

As in azurin, the copper site in plastocyanin lies in a pocket between the three loops in the polypeptide backbone at the ends of strands IV, VII and VIII. The walls and rims of this pocket (in plastocyanin), are lined by the conserved hydrophobic sidechains of Phe 14, Met 92, Pro 36, Pro 86 and Leu 63. The copper atom of azurin is also found to be in a conserved hydrophobic environment (see section 4.2.2). At the molecular surface, all that separates the copper atom from the solvent in both molecules is the sidechain of His 117(His 87). In azurin, however, the sidechains of Met 13, Met 44 and Phe 114 shield the edge of His 117 from the solvent, whereas in plastocyanin the edge of His 87 is more exposed, as structural results show that the edge of the imidazole sidechain lies level with the molecular boundary (106). This is in good accord with kinetic results, which have predicted that for plastocyanin the redox centre is approximately 2.6$\overset{o}{A}$ below the surface of the molecule, while for azurin the distance is approximately 6.5$\overset{o}{A}$ (201).   A structural reason thus exists for the slower rates of reaction of azurin with most small molecule redox reagents compared to the corresponding reactions of plastocyanin (55,57,202).

5.4   The Hydrophobic Patch.

In both azurins and plastocyanins, one of the most distinctive regions in the structure is the hydrophobic patch on the "northern" surface of the molecule around the edge of His 117(His 87). This is made up, in plastocyanin, of the conserved sidechains of Leu 12, Leu 62 and Ala 90, and the less rigorously conserved Ala 13, Phe 35, and Ala 65, together with the exposed edges of Pro 36 and Pro 86. In azurin, the hydrophobic patch is somewhat more extensive, involving the invariant residues Met 13, Met 44, Phe 114, Pro 115 and Gly 116, the conservatively substituted Met 39, Ala 42, Val 43, Met 64 and

Met 120, and the variable residues Trp 118 and Ala 119. Because

of the conservation of this region in both azurins and plastocyanins,

and its relationship with the sidechains of His 117 and His 87

respectively, it is likely that this region is a locus for electron

transfer in these proteins. Thus, the hydrophobic patch may act

as a recognition site, used to orient a potential redox partner

in such a way as to allow the maximum potential overlap between the

orbitals of the semi-aromatic imidazole sidechain of both "blue"

proteins with the ligands of another metalloprotein, e.g. cytochrome

c-551 in the case of azurin, or cytochrome f in the case of

plastocyanin.

5.5   Asn 47 Asn(38).

Inspection of Figure 5.3, the alignment of the sequences, shows

that apart from residues 108-123 in azurin (80-94 in plastocyanin),

one of the most highly conserved regions of the structure in both of

the molecules is the sequence of residues 46-50 in azurin, (37-41

in plastocyanin). Both these peptides contain copper binding residues,

but an additional unusual feature of the latter sequence is the

invariance of the internal polar residue Asn 47 (Asn 38) in both

molecules. In azurin, the sidechain of Asn 47 forms hydrogen bonds

with mainchain atoms in strands V and VII, and with the conserved

hydroxyl group of residue 113 (Ser or Thr). Similar interactions

seem likely in plastocyanin, suggesting that this residue plays an

important part in stabilizing the three-dimensional structure by

forming a link between strands V and VII. The sidechain interactions

may also help to position the mainchain, both to enable His 46 (His 37)

to bind to the copper atom, and to orient the mainchain amide nitrogen

of residue 47 (38) towards the thiolate sulphur ligand of Cys 112

(Cys 84). The distance between NH of Asn 47 (Asn 38) and SG of

Cys 112 (Cys 84) appears to be the same in the two proteins ($\sim 3.4\text{\AA}$),

within the distance expected for N-H·····S hydrogen bonds found

in proteins (177). It is thus possible that in both proteins this

hydrogen bond provides an essential mechanism for stabilizing the

Cu(II) state with respect to the Cu(I) state, in a situation

reminiscent of that recently proposed for cytochrome c (197).

5.6   Conserved Aromatic Residues.

The interiors of both barrel-like molecules contain predominantly

hydrophobic sidechains. In plastocyanin, two of the aromatics,

Phe 35 and Tyr 83, have their sidechains pointing into the solvent.

Phe 35 is part of the hydrophobic patch in plastocyanin, as is its

counterpart, Met 39, in azurin. The analogue of Tyr 83 in azurin,

Phe 111, is no longer exposed, being part of the hydrophobic interface

between the β-barrel and the flap, as well as part of the hydrophobic

pocket for the copper atom.

In plastocyanin, a sequence of aromatics extends below the

copper ligand, Met 92, to the "southern" end of the molecule : Phe 14,

Phe 82, Phe 19, Phe 41 and Tyr 80. Residues 14, 41 and 80 are totally

invariant in all plastocyanins, Phe 82 is replaced by Tyr in one

algal plastocyanin, and Phe 19 is substituted by other hydrophobic

residues in several plastocyanins. In azurins the residues analogous

to Phe 14 and Phe 82 are always aromatic; those corresponding to

Phe 19 and Phe 41 are conservatively hydrophobic; and that

corresponding to Tyr 80 (Tyr 108) is totally invariant.

Phe 14 (Tyr 15) is the first residue of the "kink" or "bulge" in

strand II of both plastocyanin and azurin. The close resemblance

between the backbone conformations of azurin and plastocyanin in

this region (residues 15-18 in azurin, and 14-17 in plastocyanin)

is remarkable, since apart from this aromatic residue the sequences

are not homologous in the region of the bulge. This invariance clearly

suggests that the aromatic character of this residue is important

in some way to the structures, but until further information becomes available from the structures of other blue proteins, or from highly refined reduced and oxidised structures of both azurins (Alc. denitrificans and Ps. aeruginosa) which will provide a broader data base for comparison, no firm conclusions can be drawn.

The immediate environment of the sidechain of Tyr 80 in plastocyanin is conserved, and, as in azurin, its hydroxyl group points into the solvent. The invariance of this residue, Tyr 80 (Tyr 108), in the two structures, together with a similar conservation of its environment and location at the extremity of the molecule, might suggest some functional significance. Although involvement in electron transfer cannot be ruled out, the distances of the centroid of the sidechain from the copper in both molecules (approximately 15Å in plastocyanin and 20Å in azurin) would seem to preclude this.

5.7   The Distribution of Charged Sidechains.

In plastocyanin, the distribution of charged sidechains on the surface of the molecule is uneven, in striking contrast to that of azurin. Thus, in higher plant plastocyanins the three conserved basic residues (lysines) play a minor role compared to the ten conserved acidic ones (106). Of the latter, none lie in the top half of the molecule, and 6 are concentrated in two kinks in the backbone at residues 42-45 and 59-61. These two groups of residues have their negatively charged sidechains directed into the solvent, and form an elongated acidic patch. Near the middle of this patch a conserved tyrosine (Tyr 83) has its sidechain also directed into the solvent, and it has been postulated (56,106)   that this striking feature of the structure may be functionally important. Kinetic, NMR, and affinity labelling experiments using small molecule redox reagents (57,203,204,205) have supported such a theory. However, it should be noted that many algal plastocyanins do not have this acidic patch,

and in fact have only one conserved acidic residue, that of Asp 42
(32). In the closely related azurins, the region of structure
corresponding to this acidic patch is the part of the β-barrel
which faces the flap, so that the charged sidechains have been
largely replaced by invariant or conservatively substituted
hydrophobic residues. Since azurin and platocyanin are undoubtedly
monophyletic, a functional significance of this patch in plastocyanin
(e.g. the possibility that it is essential to the electron transfer
mechanisms) seems unlikely.

Azurin contains 5 invariant basic residues (4 Lys and 1 Arg),
and 7 invariant acidic residues (6 Asp and 1 Glu). With the exception
of Asp 11, these usually are paired with residues of opposite charge,
(section 4.2.11) in marked contrast to the distribution of charged
residues in poplar leaf plastocyanin. The remaining charged sidechains
in azurin are mostly well distributed over the surface of the molecule,
the resultant charge being + 1.5 compared to - 8 for plastocyanin (60).

5.8 Conclusions.

Comparison of proteins that are thought to have diverged from a
common ancestor, but which, although having a similar function, are
found in different biological systems, are different sizes and have
limited sequence homology, may allow discrimination between possible
electron transfer mechanisms. This was found to be the case for the
wide variety of cytochrome c-type molecules from different sources.
Comparison of the three dimensional structures of some of this
diverse class of proteins has allowed the prediction of a hypothetical
electron transfer complex, by the elimination of possibilities that
are not common to all the structures that have so far been solved (71).
This does, of course, assume that there are no major functional
differences, an assumption that is not necessarily true for azurin
and plastocyanin. While their electron transfer pathways should

presumably be the same, or similar, because of their monophyletic origin, the surface features of their binding sites may be different (since they interact with different partners).

For azurin and plastocyanin, comparison of the structures has revealed that one distinct region of the protein structures apart from the copper site is highly conserved in both structures. Thus a potential pathway for electron transfer exists at the "northern" end of the molecule, where the exposed edge of the His 117 imidazole ring seems to satisfy all the requirements of current electron transfer theory for contact between redox partners. The presence of the striking and highly conserved hydrophobic region on the surface of both molecules around His 117 (His 87) reinforces the impression that this part of the molecule is functionally important. This could act as a recognition patch used to specify a reaction with a particular redox partner by a process such as hydrophobic bonding, which will lower the energy of the system, providing a driving force for the formation of the complex. Such contact between the two redox partners would also reduce the dielectric constant between them, facilitating the transfer of an electron from one metal centre to another.

Although electrostatic interactions have been implicated in the enhancement of the rates of reaction of acidic cytochromes c (e.g. cytochrome c-551) with azurin, as mentioned in section 4.2.3 (58), they would seem unlikely when the overall charge on the azurin molecule is taken into account (-1 for azurin from Ps. aeruginosa, + 1.5 for this azurin, cf. - 8 for plastocyanin (60)). Although it is possible that local patches of conserved acidic sidechains (as found in plastocyanin), or the delocalization of positive charge from the Cu(II) ion to the imidazole sidechain of His 117 (His 87) could be responsible for such interactions (203), the activation parameters for both the reduced cytochrome c — oxidised azurin (55) and the cytochrome f - plastocyanin reaction (58) clearly indicate that

the positive entropy term (perhaps linked to hydrophobic bonding between the molecules, or breakage of ordered water structure), outweighs any contribution from charge-charge interactions (55). This means that specific recognition through electrostatic interaction is probably not involved in the reaction(60).

The remaining unanswered question is whether oxidation and reduction proceed via the same or different pathways. It is conceivable that one site is capable of making specific and productive collisions with both cytochrome c-551 and cytochrome oxidase. Such a hypothesis has recently been proposed for the redox interactions of cytochrome c (197   ).   The observation of more than one region of the molecule with distinctive and conserved structural features, however, may point to some functional significance for such regions. There are many of these in the azurin structure (see chapter 4), but apart from the hydrophobic patch, the only one with a counterpart in the plastocyanin structure is the rigorously conserved Tyr 108 (Tyr 80) at the "southern" end of the molecule. Within current understanding of electron transfer mechanisms, however, it would seem that, for azurin at least, the distance of this sidechain from the copper site ($\sim 20\text{Å}$) would preclude it from directly taking part in electron transfer processes.

CHAPTER 6

ELECTRON TRANSFER

6.1 Electron Transfer in Blue Proteins

In blue or type I copper proteins, the fundamental mechanism
of how electrons get to and from redox centres through the intervening
protein matrix is not clear, although several theories have been
put forward. A scheme that has gained general acceptance involves
the association of protein and metal complexes prior to electron
transfer (206).

$$P.Cu(I) \; + \; oxidant \; \overset{K}{\rightleftharpoons} \; P.Cu(I).oxidant$$

$$P.Cu(I).oxidant \; \overset{k_{et}}{\rightleftharpoons} \; Products$$
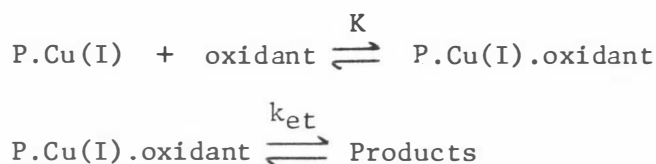
Electron transfer probably involves an outersphere mechanism, as NMR
(67,68)   and fluorescence (21) evidence, coupled with the structural
data from azurin and plastocyanin show that the metal atom lies
within a solvent-inaccessible environment 5-7Å below the surface
of the protein.

6.1.1 Experiments with Inorganic Redox Reagents.

The use of small inorganic complexes as redox partners for blue
copper proteins has been firmly established as a useful experimental
approach. By careful selection of such reagents, the redox potential,
and hence the free energy change of the reaction, may be controlled,
the size and the hydrophilic or hydrophobic nature of the reagent
may be varied by introducing changes in the structure and the charge
of the reagent may be altered. For example, use has been made of
the tris-complexes of cobalt(III)with phenanthroline,$[Co(phen)_3]^{3+}$
and its derivatives (1) 5-chloro- , (2) 5,6-dimethyl- , (3) 4,7-
dimethyl- and (4) 4,7-diphenyl-4-sulphonate-($[Co(4,7-DPS \; phen)_3]^{3-}$)
phenanthroline (57).
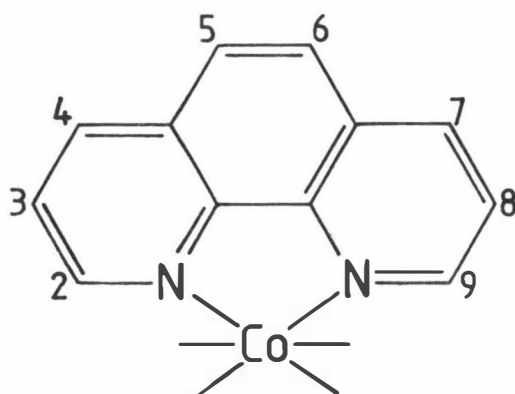
Figure 6.1   The numbering for cobalt–phenanthroline complexes.

Of these, the redox potentials of complexes (1) and (2) are higher, while those of (3) and (4) are lower than that of the parent complex. Substitution at the 4 and 7 positions blocks approach to the phenanthroline rings in the complexes, and this is likely to cause steric hindrance in the protein–complex interaction and hence affect the rate of reaction. Thus complexes (3) and (4) would be expected to have lower reactivities than complexes (1) and (2). Complex (4), as well as being the only negatively charged member of the set, is also by far the largest in size (57).

Generally, complexes with π-conducting ligands such as $[Co(phen)_3]^{3+}$ exhibit a much higher redox activity with metalloproteins than those like $Fe(EDTA)^{2-}$ or $Fe(CN)_6^{3-}$ whose hydrophilicity hinders close approach between the reagent and the more hydrophobic active centre of the protein (55). There is also a large amount of information available from the enthalpies and entropies of reaction and activation ($\Delta H^{\ominus}$, $\Delta S^{\ominus}$ and $\Delta H^{\ddagger}$, $\Delta S^{\ddagger}$ respectively). Thus the value of $\Delta S^{\ominus}$ for a reaction can be discussed in terms of electrostatic interactions between pairs of similar or opposite charges, and large values of $\Delta H^{\ominus}$ indicate that binding during complex formation is highly specific. The activation parameters $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$, on the other hand, are

independent of the charge on the redox complex (27). For example, the activation parameters for the Fe(EDTA)$^{2-}$-azurin(II) reaction provide an important clue as to the nature of this pathway. The very small $\Delta H^{\ddagger}$ value and large negative $\Delta S^{\ddagger}$ have been taken to imply a very weak interaction between the donor and acceptor redox orbitals, necessitating transfer of the electron over a large distance (> 3Å) to the buried copper site (207,208).

Many such kinetic investigations have been carried out on blue copper proteins, especially on azurin, plastocyanin and stellacyanin (for reviews see references 55 and 57). These show that the accessibility of the metalloprotein active centre to outer sphere contact with reagents increases in the order azurin < plastocyanin < stellacyanin. Such results fit in with the findings of the comparison of the three dimensional structures of azurin and plastocyanin (Chapter 5), for interactions involving His 117 (His 87), viz. that in azurin the edge of the imidazole ring is more shielded from the solvent than in plastocyanin.

6.1.2 Evidence for More than One Binding Site on the Azurin and Plastocyanin Molecules.

Redox reactions of azurin and plastocyanin with small molecule reagents are affected by pH, and the different pH dependences for different reagents suggest that more than one binding site or mode of interaction exists (192,208).

For example, the rate of oxidation of Cu(I)-Az with $[Fe(CN)_6]^{3-}$ alters from 0.5×10^4 ℓ mol^{-1}s^{-1} at pH 9.0 to 1.8×10^4 ℓ mol^{-1}s^{-1} at pH 5.0, the dependence of the second order rate constant on pH having a pKa of 7.1, which is shifted to 6.1 for the reduction of Cu(II)-Az with $[Fe(CN)_6]^{4-}$ (192). For plastocyanin, the rate of reaction alters from 7.9×10^4 ℓ mol^{-1}s^{-1} at pH 7.0 down to zero at pH ≤ 5.0, the dependence of the second order rate constant on pH

having a pK_a of approximately 6.0. The $[Fe(CN)_6]^{4-}$ reduction of Cu(II)-Pc does not, however, respond to pH in the same way, since the rate of reaction remains constant between pH 7 and 6, but increases rapidly with any further decrease in pH (209). As $[Fe(CN)_6]^{3-}$ and $[Fe(CN)_6]^{4-}$ probably use the same site on the protein, the protonation of the protein must be influenced by the oxidation state of the copper (27,209). Decreases in the rates of reaction for both proteins are thought to be due to the decreasing ability of the protein to form productive protein-oxidant or protein-reductant complexes. For azurin the decrease occurs as the pH is increased, while for plastocyanin it accompanies a decrease in pH (192,209).

For oxidation of Cu(I)-Az by $[Co(phen)_3]^{3+}$ the rate of reaction decreases with decreasing pH, the second order rate constant having a pK_a of 7.0 (6.0 for the corresponding oxidation of Cu(I)-Pc), while the reaction with $[Co(4,7-DPS\ phen)_3]^{3+}$ is independent of pH in the range 6.3 to 9.0 for both proteins (192,209). Kinetic parameters for the reaction of Cu(I)-Az with $[Co(phen)_3]^{3+}$ suggest that electrostatics play a minor role compared to the corresponding reactions of plastocyanin, and that a different pathway is operating compared to that proposed for the plastocyanin reaction (192). Possible explanations for these pH effects are mainly based on protonation rendering the protein inactive by

(i)   causing some modification in structure at the binding site which prevents complex formation

(ii) rendering the Cu ion redox inactive (e.g. by substantially stabilising or destabilising one redox state relative to the other)

(iii) preventing electron transfer by blocking the pathway between the bound reductant and the redox site of the protein.

6.2 Tentative Identification of the Binding Sites on Azurin and Plastocyanin.

The possibility of there being more than one binding site on azurin and plastocyanin, for small-molecule redox reagents at least, was suggested by the different pK_a values obtained for their pH dependent reactions with azurin and plastocyanin. Some evidence as to the nature of these sites comes from NMR spectroscopy (37,44, 64).

All azurins so far sequenced contain four invariant histidine residues (32). From NMR studies of Cu(I)-Az, one of the histidines has been shown to titrate with a pK_a of 7.57, while the second has a pK_a of approximately 7.0 (37,62,64). The other two histidines protonate at much lower pH values because they are co-ordinated to the copper atom (44). A small shift in the pK_a of the first histidine has been observed on the oxidation of Cu(I)-Az to Cu(II)-Az, suggesting it is far removed from the Cu site (64). The second histidine, which also has an unusually slow acid - base exchange rate, is near enough to the copper atom to be affected by a change in its oxidation state. These pK_a values for Cu(I)-Az determined by NMR are sufficiently close to those detected in kinetic studies to suggest involvement of the same histidines in the reaction mechanism. Thus $Co[phen]_3^{3+}$ may bind at a site on the protein close to the histidine with a pK_a of 7.6, probably His 83, as it is farthest from the Cu, while $[Fe(CN)_6]^{3-}$ may use a site influenced by the histidine with a pk_a of 7.1, probably His 35, since it is near the copper site ($\sim$ 7.7Å away). As the binding site for $[Co(4,7-DPS\ phen)_3]^{3-}$ is apparently pH independent, a different binding site is suggested for this reagent, which may be influenced by the hydrophobic nature of the complex. This site may be at the hydrophobic patch adjacent to His 117.

For plastocyanin, two sites, one in the vicinity of His 87, and

the other in the vicinity of Tyr 83 were suggested by Colman
et al. (56). A patch of acidic sidechains near Tyr 83, around Glu
43 and Glu 49, would account for the observed pK_a value of $\sim$ 6.0
for the oxidation of Cu(I)-Pc by $[Fe(CN)_6]^{3-}$ and $[Co(phen)_3]^{3+}$ (209).

6.2.1 Structural Interpretation of pH Effects.

The kinetic studies of Lappin et al. (192,209) have suggested
binding sites which involve, for different small molecule reagents,
(viz. $[Fe(CN)_6]^{3-}$, $[Co(phen)_3]^{3+}$ and $(Co(4,7-DPS\ phen)_3^{3-})$) His 35,
His 83 and the hydrophobic patch. The first two are predicted on the
basis of pK_a measurements. Particular residues responsible for a pH
transition with a given pK_a are, however, extremely difficult to assign
because the pK_as of individual sidechains are often shifted by their
environment, and because of the problems associated with multiple
equilibria.

This is clearly illustrated by plastocyanin. As discussed in
section 6.1.2, kinetic studies have been interpreted in terms of
modifications to potential binding sites in the region of Tyr 83 and
the acidic patch as a result of pH changes. Crystallographic work
has, however, characterised a low pH inactive form of plastocyanin
in which the copper ligand, His 87, is protonated and dissociates from
the copper. This is the binding site predicted for the oxidant
$[Co(4,7-DPS\ phen_3)]^{3-}$ from the same kinetic studies (209), on the
basis of the independence of its reaction with azurin to changes in
pH (in the range 5.2-7.5).

In azurin, there is no crystallographic evidence for similar
structural changes with pH, although experiments on the decolourisation
of azurin at very low pH (< 3.0) have been interpreted in terms of the
possible partial dissociation of the copper from its binding site,
or rearrangement of the ligands (84).

6.2.2 The Use of Chromous Ions as a Probe.

An elegant way to study electron transfer pathways between
metal complexes is to use chromous ions as a probe (210). This
is because while strongly reducing Cr(II) complexes are substitution-
labile, their single oxidation product, Cr(III) complexes, are
generally substitution-inert. Thus the co-ordination sphere of
the chromous ion in the transition state formed during the
electron transfer process should remain the same in the product.
One such example is the Cr(II) reduction of ferricytochrome c,
where the Cr(III) produced was shown to be very tightly bound to
the ferrocytochrome c (211). Further investigation by Grimes et al.
(212), using proteolytic cleavage of the Cr(III)-ferrocytochrome c
complex, resulted in the identification of the residues to which the
chromium was bound in the product, and most probably during electron
transfer.

Similar experiments have been recently carried out on azurin
from Ps. aeruginosa (194) and plastocyanin from French Bean by
Farver and Pecht (205). Both proteins, in solution, were stoi-
chiometrically reduced by Cr(II) ions to yield relatively inert
colourless Cr(III) species. When these were proteolytically digested,
with both trypsin and chymotrypsin in the case of azurin, or
thermolysin in the case of plastocyanin, peptides labelled with
Cr(III) could be separated by chromatographic methods. Amino acid
analyses showed that for azurin, the tryptic peptide consisted of
residues 80-92, while the chymotryptic peptide went from Thr 84 to
Val 95. When the chemical nature of the amino acid sidechains of

these peptides and the crystallographic structure of azurin were taken into account, it looked probable that the chromium was co-ordinated to Lys 85 and Glu 91. The proposed electron transfer pathway involves an opening in the peptide sheath between Gly 88 and Glu 91 which exposes ND1 of the imidazole sidechain of His 35. The hexaquochromium(II) ion approaches the opening and is bound to the carboxylate of Glu-91 and probably concomitantly to the NZ of Lys 85. From this position, it is proposed that the electron is transferred from Cr(II), most likely via a co-ordinated water molecule hydrogen bonded to ND1 of His 35, and flows through a conjugated pathway formed by the aligned imidazole sidechains of His 35 and His 46 to the copper atom (194).

For plastocyanin, the Cr-binding peptide included Asp 42, Glu 43, Asp 44, and Glu 45, some of the residues which make up the "acidic patch" on the molecule. Near the middle of this patch, the conserved Tyr 83 also has its mainchain directed into the solvent, and as the fluorescence emission intensity and pH dependence usually observed for tyrosines in proteins differed markedly in the Cr(III) adduct, some sort of interaction between the tyrosine and chromium was suggested. The proposed electron transfer pathway was through the interacting π-systems of a highly invariant array of aromatic residues between the copper and Tyr 83, the distance being approximately 10Å (205). High resolution NMR studies also point to this region as a binding site for the chromium analogue of $[Co(phen)_3]^{3+}$, $[Cr(phen)_3]^{3+}$ (203,204). The effect of $Cr(CN)_6^{3-}$ is clearly different. No effect is noted upon the tyrosine residue, but the resonances of both His 87 and Met 92, both Cu ligands, are broadened, indicating that the complex may bind to the hydrophobic patch over the copper site (203,204).

Reduction with Cr(II) has also been used to demonstrate that there may be two binding sites on the protein molecule for the

physiological electron transfer reactions of azurin (194). Thus it was found that while the cytochrome oxidase catalyzed oxidation of Cr(III)-Cu(I)-Az by O_2 is unaffected by the presence of the bound Cr(III), the electron transfer reaction between Cr-labelled azurin and cytochrome c-551 is markedly slower than with the native azurin (194). These results suggest there may be two different sites for redox interactions, one of which is blocked by Cr(III).

6.3 Protein-Protein Interactions

Although the relevance of small-molecule studies can be questioned in terms of their relation to the interaction between macromolecular redox partners, pH effects similar to those found in the small molecule oxidations of Cu(I)-Az (by $[Fe(CN)_6]^{3-}$ and $[Co(phen)_3]^{3+}$) also influence the oxidation of Ps. aeruginosa azurin by cytochrome c-551. Thus two conformers of azurin have been identified, one inactive to oxidation, the other active (213,214). The active species was determined to be a protonated species, and as the group involved had a positive enthalpy of ionization, and a pK_a of approximately 7, it was thought to be the histidine, probably His 35, implicated in the oxidation of azurin by $[Fe(CN)_6]^{3-}$ (192,214). The very slow proton exchange rate of this histidine ($1 < k < 35 \ sec^{-1}$) (62,64), is consistent with the reciprocal relaxation time reported for the interconversion between HA(r) and A(r) (where A(r) is reduced azurin), in the reaction between oxidised azurin (A(o)) and reduced cytochrome c-551 (C(r)).

$$HC(r) \ + \ A(o) \rightleftharpoons C(o) \ + \ HA(r)$$
$$\Updownarrow \qquad\qquad\qquad\qquad \Updownarrow$$
$$C(r)^* + H^+ \qquad\qquad A(r)^* + H^+$$
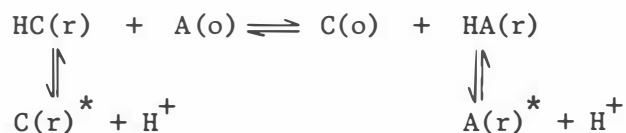
(where $*$ indicates an inactive conformer).

The pH independence of the overall equilibrium constant for the above reaction (214), indicates that redox inactive and active forms of cytochrome c-551 may also exist. This has been confirmed by two studies (214,216). The nature of the transition in reduced cytochrome c-551 ($pK_a \approx 7.2$) is thought to be linked to the protonation of one of the heme propionates, which are more exposed in cytochrome c-551 than in cytochrome c. This may explain the absence of a similar pH transition in the reduced form of the mitochondrial protein. The transition is reported to be coupled to a redox potential change ($300 \rightarrow 240$ mV (216)) in going from pH $5 \rightarrow 9$. A complementary pH dependence of azurin has been reported by Lappin et al. over the same pH range ($360\rightarrow300$ mV) (192).

The above results clearly implicate the protonated form of His 35 in electron transfer between Ps. aeruginosa azurin and Ps. aeruginosa cytochrome c-551. The situation is less clearcut for other azurins. Investigation of the electron transfer equilibrium and kinetics between azurin from Alc. faecalis, and cytochrome c-551 from Ps. aeruginosa,has failed to detect a conformational change in this azurin (85). The specific rates for electron transfer, however, are still very fast, although slightly attenuated, and the activation parameters are quite different compared to those between azurin and cytochrome c-551 from the same organism, particularly for the azurin oxidation (85). These differences may be due to species specific interactions between the proteins, originating perhaps from either areas of complementary charge on the two proteins at neutral pH, or from hydrophobic interactions (213). There are sequence differences around His 35 in Alc. faecalis and Alc. denitrificans azurins compared with Ps. aeruginosa azurin. Thus Pro 36 and Ser 34 of the Ps. azurin are replaced by a threonine and lysine respectively in the case of Alc. faecalis, and by a valine and lysine in the case of Alc. denitrificans. Such

substitutions could cause conformational differences around the imidazole of His 35, or solvent accessibility could be altered, affecting its reactivity and thus the protonation-dependent conformational transition. On the other hand, it may be that such a transition is only seen when cytochrome c-551 of the correct species is used.

To try and resolve this question, affinity labelling procedures using Cr(II) were applied to Alc. denitrificans azurin in the crystal form in order to try to ascertain the binding site by crystallographic methods.

6.4 Crystallographic Studies on Reduction of Azurin

Because protein crystals dissolve or deteriorate out of their mother liquor, all work was undertaken in 75% saturated solutions of ammonium sulphate made up with 0.05M ammonium acetate buffer pH 5.0. Furthermore, because of the ease with which chromous ions are oxidised in air, the entire procedure from the reduction of the crystals to their subsequent mounting and sealing in glass capillaries had to be undertaken in an oxygen-free atmosphere. To achieve this, a glove box continuously flushed with oxygen-free nitrogen was used. Great care was taken to ensure that all solutions were degassed, then flushed with nitrogen before being introduced into the system. The azurin crystals themselves, and all equipment used, were left to equilibrate in the nitrogen atmosphere overnight before the experiment was begun.

A stock solution of chromous material was prepared by reducing a 0.1M solution of chromic sulphate made up in 75% saturated ammonium sulphate solution using a Jones reductor column (194,217, 218   ). A few drops of the resulting pale blue solution were added to the mother liquor containing the crystals, and mixed by gently swirling the solution. The crystals, usually a very dark blue,

became colourless in just under 30 minutes. Two reductions were carried out simultaneously, the first on crystals suitable for X-ray analysis, the second on a large quantity of small crystals to be used for atomic absorption analysis. The larger crystals were then mounted in glass capillary tubes, and small amounts of the chromous solution drawn up into the bases of the capillaries before they were sealed with wax. They were then placed in screw cap jars which had also been equilibrated in nitrogen before being removed from the glove box and stored.

6.5 The Nature of the Binding Site

Precession photographs were taken of one of these crystals as already described. These showed that the crystals diffracted well, and that there appeared to be no obvious change in the axial lengths, and no large changes in intensities. However, it could be seen that under irradiation, the crystals gradually regained their blue colour. Farver and Pecht (194), in their work on azurin from Ps. aeruginosa, reported that the cytochrome oxidase catalysed oxidation of $Cr(III)-A(r)$ by O_2 is unaffected by the presence of bound $Cr(III)$. As azurin has been reported to be auto-oxidisable in the presence of molecular oxygen (51) it seems logical to expect that such oxidation of the chromium-reduced crystals may occur. It was thus assumed that some oxygen was present in the system at the time the crystals were mounted (despite the care taken to ensure oxygen-free conditions), or was somehow entering the capillary tubes. As the mother liquor in the bottom of these tubes showed no sign of evaporation or crystallization (it contained 75% saturated ammonium sulphate solution), the latter was considered unlikely. It was observed later, however, that the crystals not subjected to irradiation by X-rays gradually regained their blue colour, i.e. became oxidised, over a period of approximately 8 weeks.

To try to establish whether chromium was in fact bound to the protein, analysis of the second batch of crystals (the smaller ones) was carried out by atomic absorption. To ensure that excess chromium was removed from the colourless crystals, they were first filtered, then washed with approximately 100 mls of 7.5% ammonium sulphate solution (made up with deionized water). This involved suspending the crystals in repeated amounts of the solution, waiting for them to settle (1-2 hours) and then decanting the supernatant. After several washings, approximately 10 ml of distilled water was added to the pellet in order to dissolve the crystals. It became apparent, however, that the solubility of the crystals had decreased, as they only partially dissolved, contrary to expectations (azurin crystals usually being very soluble in water). On standing, the supernatant gradually turned blue (overnight) although the remaining undissolved crystals did not regain their colour. Atomic absorption analysis of this blue solution showed that it contained approximately $2\frac{1}{2}$ times as much chromium as copper, and assuming that all of the copper in the solution was associated with protein, this is a clear indication that there were some chromium atoms bound to the protein.

Three dimensional data were collected in the manner already described, to low resolution (5.0Å), using one of the larger crystals. During exposure to the X-rays, the colourless crystals gradually turned blue, as they had done previously. Because this change in colour did not seem to be associated with any obvious changes in cell dimensions, and as it had been shown that chromium was still bound to the reoxidised protein, the data was processed in the usual way. In order to determine the possible binding sites of the chromium species, both a difference Fourier map, phased by the $KAu(CN)_2$, uranyl acetate, and $Hg(NH_3)_2Cl_2$ derivatives, and a difference Patterson map using $(F_{PH} - F_P)^2$ coefficients were calculated. The

resulting difference Fourier map had 5 main peaks. These 5, which were compatible with the difference Patterson map, had co-ordinates as follows:-

A (.281, .497, .127)   A' (.492, .280, .119)

B (.152, .088, .157)   B' (.072, .189, .100)

C (.342, .417, .059)

When these positions were plotted on to the electron density map, sidechains providing the binding sites could be identified. The sites are:

A.   A carboxyl oxygen of Glu 106 (a site very similar to one occupied by the uranyl derivative). This site appears in both molecules

B.   Near the carbonyl oxygen of Phe 114, about 3.5Å from CB of His 117 and 4Å from the thiolate sulphur (approximately 5Å from the copper atom). Peaks appear at this position for both molecules, but binding at this site would involve some disturbance of the structure. Whether chromium is bound at this site, or whether, the peaks arise from some movement in the copper site (e.g. in the copper binding loop 112-117) is not clear.

C.   Near the carboxyl oxygens of Glu 19.

6.6   The Significance of the Binding Sites

Because of the conditions under which the experiment was conducted, with SO_4^{2-}, NH_4^+ and acetate ions in solution, the species binding to the protein crystals is unknown. As the chromous solution gradually changed colour (from a pale sky-blue to a darker grey-blue) during the reaction, it is possible that different species, some of

which are probably not responsible for reducing the protein, may be binding at other sites (105). Inspection of the binding sites, however, show that the species binding at A and C are most likely to be positively charged as they are bound to negatively charged carboxyl groups. The species binding at the second site (near residues 117 and 112) may be different although it may still have some positive charge allowing a polar interaction to occur between it and the carbonyl oxygen of Phe 114. It may be possible that there is some interaction between co-ordinated groups on the chromium species and either the imidazole sidechain of His 117 or the thiolate sulphur of Cys 112.

Of the three sites, site B would seem to be the most likely to effect the reduction of the protein, as it is nearest to the copper, and in the vicinity of two copper ligands, His 117 and Cys 112. Although the other two sites cannot be completely dismissed, as electron transfer mechanisms are not yet fully understood, their distance from the copper atom should preclude them ($\sim$ 20$\overset{o}{A}$ for site A, and $\sim$ 19$\overset{o}{A}$ for site C). It is of interest, however, that the carboxyl oxygens of Glu 106 are adjacent to the free hydroxyl group of Tyr 108 which is invariant in all azurins (32). In plastocyanins, the corresponding residue is the invariant Tyr 80, part of the highly conserved aromatic channel running from the base of the copper site down the front of the molecule (see Figure 1.3), suggested to be functionally important (106,204). In contrast to the results of Farver and Pecht (194), there appears to be no binding of chromium at the carboxyl oxygens of Glu 91, the proposed electron transfer locus in azurin from Ps. aeruginosa, despite this being a major uranyl site in these crystals and therefore freely accessible to reaction.

Finally, the action of thiomersal on the protein merits comment. In binding to azurin, at a site approximately 6Å from the chromium site B discussed above, but within 2Å of the thiolate sulphur of Cys 112, it also effects a reduction of the copper site. As the derivative crystals show some non-isomorphism, indicating that the peaks representing this binding site may not be due entirely to the presence of an extra heavy atom, but also to some conformational change, the exact binding site is not certain. It is possible that there may be direct interaction of the sulphur orbitals of this complex with those of the copper atom (the mercury is situated approximately 3.4Å from the Cu). However, the reduction seems to be permanent, unlike that effected by chromium, suggesting that a different mechanism may be operating. Although there is no certainty that inorganic redox reagents will use the same electron transfer sites, pathways or mechanisms as the biological partners of a metallo-protein, it may be significant that two quite different reagents that reduce the protein bind in a region of the protein, that, rather than being near to His 35, or His 83, is nearer to the copper site itself, and within the conserved hydrophobic patch.

CHAPTER 7

CONCLUSIONS

Many aspects of this work have been discussed in previous chapters. A few points, however, are worthy of restatement, or of further comment.

The structure analysis of azurin from Alc. denitrificans was not straightforward because of the difficulty in preparing high quality heavy atom derivatives. This was because, generally, heavy atom compounds that produced good changes in the intensity pattern of the crystals also caused changes in the axial lengths, so that the derivatives were not isomorphous. In retrospect, however, more could have been done in this area, such as attempting to replace the copper atom with heavier metals (see appendix III) and possibly in further refinement of the derivatives. For example, knowledge of the crystal packing obtained from the structure determination suggested that the uranyl sites close to the two-fold axis could have been better described. Nevertheless, the resulting electron density map had good solvent/protein contrast, and was interpretable. The presence of two copies of the structure in the asymmetric unit was of particular help in coming to a detailed interpretation, as areas of the structure that were ambiguous in one molecule were usually clear in the other. Several small differences in interpretation remained between the two molecules, but the structure determination showed unequivocally that the folding pattern of the polypeptide chain was the same as that of plastocyanin, rather than that of the alternative conformation originally proposed for Ps. aeruginosa azurin (82). It was also consistent with the theory that azurin and plastocyanin had diverged from a common ancestor, since both conformational (i.e. polypeptide chain folding) and functional (i.e. the copper site) elements are homologous.

A precise description of the copper site is not yet possible because of the need to refine the structure. It is clear that, as in plastocyanin, the copper has three ligands that are closely bound ($\sim 2\text{Å}$), 2 His and 1 Cys, the copper atom being approximately coplanar with the ND1 atoms of the two histidine sidechains and SG of the cysteine sidechain. There is also no doubt that, in both structures, Met 121 provides a fourth ligand at a rather long bond distance ($\sim 3\text{Å}$). A totally unexpected result, however, was the closeness of the carbonyl oxygen, belonging to the peptide bond between residues 45 and 46, to the copper. (At 3Å resolution, the density of this group merged with that of the copper in both molecules). It is still not possible to say whether this oxygen should be regarded as a fifth ligand, but in any case it is probably close enough to the copper to affect the redox potential. The possibility also exists that it may be the carbonyl oxygen responsible for the unassigned peak "x" in the NMR work of Ugurbil et al. (44). The latter felt the peak might be due to a peptide carbonyl co-ordinated to the copper atom in the reduced form of the protein. Although in plastocyanin this peptide carbonyl is too far away ($\sim 3.7\text{Å}$) to be considered bound to the copper, the report on the structure of <u>Ps. aeruginosa</u> azurin (83) notes that the plane of peptide 45 could be oriented satisfactorily with either the NH or the C=O pointing toward the copper at a distance of 3.3Å, leaving open the question of any interaction.

Of all the predictions of stereochemistry for the "blue" copper sites in proteins, that of Miskowski et al., which was based on resonance Raman spectra (35), agrees best with the findings of this structure determination, despite the conflict between such stereo-chemistry and ESR data. Structurally, it is unlikely that the Cu-S (Met 121) and the Cu-O (Gly 45) (if it exists) bonds contribute much individually to the copper site. Together, however, these weak bonds, together with the hydrogen bond between the thiolate sulphur

of Cys 112 and the amide nitrogen of Asn 47, may provide a mechanism

by which the Cu(II) state of the protein is stabilised relative to

the Cu(I) state, and hence they may be functionally important.

A recently reported copper co-ordination complex, with a thio-ether

donor at the apex of a square pyramid whose base is the donor set

N_2O_2, has ESR parameters typical of normal square planar geometry

(185). It would thus appear that, for square planar geometry at

least, the $g_{\parallel}$ and $A_{\parallel}$ values are relatively insensitive to even

quite strong axial perturbations (219). If square pyramidal stereo-

chemistry is considered a possibility for blue copper proteins, the

basal plane should contain nitrogen and sulphur donors on the basis

of the $g_{\parallel}$ and $A_{\parallel}$ values reported for "blue" centres. In azurin from

Alc. denitrificans (this work), the three stronger bonds formed by

ligands from His 46, His 117 and Cys 112 can be regarded as forming

such a basal plane, except that the fourth ligand, Met 121, is

distorted from the plane so that the geometry becomes more trigonal

bipyramidal. In model complexes, such a distortion of the basal

plane of square pyramidal complexes produces small hyperfine splitting

constants in the E.S.R. parameters ($0.0065-0.0100$ cm^{-1}), comparable

to those in blue copper proteins (47). All such complexes, however,

including a Cu(II) N_2S_2O chromophore with a distorted trigonal

bipyramidal stereochemistry similar to that found for the copper site

in Alc. denitrificans azurin, have $g_{\parallel} < g_{\perp}$, in contrast to the

blue copper centres in proteins, for which $g_{\parallel} > g_{\perp}$ (47). Unlike

the azurin site, however, the Cu(II)-N_2S_2O complex (184) has compressed

rather than elongated axial bonds, and as no model complex with a

stereochemistry and ligand set like those of azurin has yet been

reported, the possibility of 5-co-ordinate geometry cannot be completely

discounted on ESR grounds.

The trigonal bipyramidal model itself is consistent with the

suggestion that the peculiar spectroscopic properties of "blue"

copper centres reflect a co-ordination sphere which is intermediate between the normal co-ordination geometry of Cu(II) and Cu(I), thereby allowing efficient electron transfer. Although Cu(I) rarely shows tetragonal co-ordination, it is otherwise quite variable. Tetrahedral co-ordination is the most common, but trigonal co-ordination is also known (220, 221, 222). In the trigonal bipyramidal model all that is required to generate a stable trigonal $Cu(I)-SN_2$ complex is the further weakening of the $Cu-S_{121}$ and $Cu-O_{45}$ bonds, which would require little reorganisation (in line with the Franck-Condon Principle). If the carbonyl oxygen is not in fact bound to the copper atom, the remaining four ligands form a distorted tetrahedron, a geometry which is again intermediate between normal geometries found for copper(I) and copper(II). (7)

The chemical character of the ligands bound to copper is also intermediate between the requirements of copper in its two oxidation states. Thus in terms of the simple "hard-soft-acid-base" theory (1,2), the ligands in azurin include two groups which can be classified as "soft" (RS^- and R_2S) and prefer to bind to Cu(I), and two groups which are intermediate between "hard" and "soft" (imidazole), and prefer to bind to Cu(II). (Main chain carbonyl oxygens also prefer to bind to Cu(II) rather than Cu(I)). Such a combination of ligands should be advantageous for outer sphere electron transfer, which requires the first co-ordination sphere to remain intact.

The changes accompanying reduction can only be guessed at. Structure determinations of ascorbate-reduced plastocyanin at pH 3.8, 4.4, 5.1, 5.9 and 7.0, and resolutions between 1.7 and 2.1Å, have been carried out (106). These show scarcely any significant differences between the five structures of the reduced protein, or between them and the structure of the oxidised protein, **except** at the copper site. A comparison of the reduced and oxidised structures

at pH 6.0 shows that the reduction causes the copper atom to move away from His 87 toward Met 92, so that the $Cu-N_{87}$ bond increases from 2.10Å to 2.82Å while the $Cu-S_{92}$ bond decreases in length from 2.90Å to 2.68Å. The bond distances of the other two ligands remain essentially the same. As the pH decreases, the $Cu-N_{87}$ bond length gradually increases with a concomitant reduction of the $Cu-S_{121}$ bond length until the Cu-N (His 87) distance of $\sim 3.4\text{Å}$ is sufficient to accommodate a Van der Waals contact between the Cu atom and a proton on the imidazole ND1 atom. The Cu atom is, in this situation, co-planar with the N_{37}, S_{84} and S_{92} atoms, and under such conditions is assumed to be the redox inactive form seen at low pH by Lappin et al. (192).

A similar change to the co-ordination sphere for azurin would require more reorganisation, as any movement of the His 117 imidazole ring toward the molecular surface would have to be accompanied by a change in conformation of the Met 13 and Met 44 sidechains, as these shield the edge of the His sidechain from the solvent, in contrast to the more open site in plastocyanin. As the response to a decrease in pH for both oxidations and reductions of azurin by some small molecule reagents is the opposite to that for plastocyanin, a different mechanism may be acting in azurins. Although the results of inorganic and kinetic experiments may be interpreted in terms of possible protein-protein interactions and electron transfer mechanisms, only a comparison of the refined oxidised and reduced structures, in conjunction with such results, is likely to provide a proper base for understanding such processes. The limited studies on reduction of the protein in this investigation, however, have shown that effecting a permanent reduction of the protein, especially under X-irradiation, may be difficult, and more work needs to be done in this area. Similar problems have been noted for plastocyanin (106) and dithionite-reduced rubredoxin (223).

The observations discussed in Chapter 6 suggest a number of possible sites for electron transfer. These are:-

(i)   His 117 and the hydrophobic patch around its sidechain. This is strongly suggested by comparison of the structures of azurin and plastocyanin, and the fact that the residues making up this "patch" are highly conserved in all azurins and plastocyanins so far sequenced (32,74). NMR (203,204) and kinetic (192,209) experiments also implicate this area in electron transfer between certain small molecule reagents and both azurin and plastocyanin. Some support also comes from the Cr(II) reduction experiments on crystals of azurin described in section 6.4.

(ii)   His 35 and its neighbouring sidechain His 46. These residues have been implicated in the electron transfer processes of Ps. aeruginosa azurin by the affinity labelling experiments of Farver and Pecht (194), and the kinetic experiments of Lappin et al. (192). His 35 is also thought to be the group involved in the transition between the active and inactive forms of reduced azurin observed in the electron transfer reaction between azurin and cytochrome c-551 from Ps. aeruginosa (214).

(iii)   His 83 is implicated in electron transfer mechanisms by the pH dependence of kinetic experiments involving small molecule inorganic redox reagents (192).

(iv)   Tyr 108. Suggestions that Tyr 108 may be functionally important to the protein have arisen mainly because of the rigorous conservation of both it and its analogue in plastocyanin, Tyr 80, in all azurins and plastocyanins so far sequenced (32,74,106).

These results focus on two main areas thought to be involved in electron transfer processes in azurin, viz. the hydrophobic patch and His 35. Of these, there is little doubt about the involvement of the hydrophobic patch and His 117, both from the point of view of the high degree of conservation and from its compatibility with current models for electron transfer. Both these points have been fully discussed in Chapters 4 and 5. It is interesting to speculate, however, on its role in the possible electron transfer pathway between azurin and cytochrome c-551, and/or cytochrome oxidase. Because the interactions of azurin with cytochromes c-551, c-553 and f are three orders of magnitude faster than that with mammalian cytochrome c, (in contrast to expectations based on the consideration of overall charge), it has been proposed that azurin reacts specifically with acidic cytochromes (58). Thus specific electrostatic interactions between the two proteins (azurin and cytochrome c-551) are implicated in the rate enhancement. A comparison of the rates of reduction of azurin from Ps. aeruginosa with both cytochrome c-551 and cytochrome c-553 shows they differ only by a factor of about two (viz. 6.1×10^6 and 1.43×10^7 $M^{-1}s^{-1}$ respectively) (60), yet the difference in overall charge on the two cytochrome molecules is large, viz. -13 for cytochrome c-553, -2 for cytochrome c-551. For the reduction of azurin from Alc. faecalis with the same cytochromes, the rates were much lower (viz. 5.23×10^5 and 2.10×10^6 $M^{-1}s^{-1}$ respectively) even though the overall charge on the two azurin molecules is the same, viz. -1 (60). Such results suggest that the role of electrostatics in the interaction is minor, that if they do play a part, only small localized areas of charge may be involved, and that species-specific interactions may be important.

The increasing number of three dimensional structure analyses of cytochromes c from a number of diverse sources, including cytochrome c-551 from Ps. aeruginosa (238), has enabled comparisons to be made which

suggest a structural reason for the rate enhancements observed above.
Thus, cytochromes c-551, c-553, and f, although they come from
bacterial as well as plant and eukaryotic algal sources, may all be
classed as "small" compared to other cytochromes (79). The
extensive deletions in their sequences compared to that of horse
heart cytochrome c, for instance, result in the heme, which is
situated in a hydrophobic pocket (similar to the copper in azurin),
being more exposed. Hence its exposed edge, the one containing small
non-polar groups, may be able to penetrate the hydrophobic patch
around His 117 of the azurin molecule to a greater extent than that
of a larger cytochrome c, allowing a more efficient overlap of the π
orbitals of the heme and the imidazole ring. This would allow the
facile passage of an electron from one metal centre to the other
through a conjugated system of liganded groups.

It is possible, however, that some of the resultant negative
charge on the heme (Fe, 3+/2+; two deprotonated heme nitrogens, -2;
two propionate sidechains, -2), may be delocalised over the conjugated
system of the group. This may be sufficient to bring about an
electrostatic interaction with the slight positive charge on the
imidazole ring of His 117, possibly through delocalisation of some of
the positive charge from the copper atom. Electrostatics, therefore,
may add to the already favourable protein-protein interaction which
is characterised by a positive activation entropy (55). The
hydrophobic patch may thus be essential to the electron transfer
mechanism of the protein through:

(a) allowing the close approach, and penetration of the heme

(b) orienting the metalloprotein in such a way as to allow maximum
overlap of the π orbitals of the heme with those of the His 117
sidechain

(c) possibly isolating any partial charge on the imidazole ring

from the variable charge distributed over the rest of the
molecule.

The rates of reaction of plastocyanin with various metalloprotein
reductants also do not appear to depend on the net charge carried
by such a reagent (58) despite the large net charge on the
molecule ($\sim$ -9). A similar mechanism may thus be operating. It
would be interesting to see whether there is a similar "hydrophobic"
patch on the physiological oxidant reputed to interact with azurin,
cytochrome oxidase, especially as the subunit containing copper
is very similar to azurin in sequence, possessing residues analogous
to many of those which make up the hydrophobic patch in azurin. For
example, Met 39, Met 44, Phe 114 and Met 120 are analogous to Leu 95,
Met 100, Ile 199 and Phe 206 in cytochrome c oxidase (81).

The involvement of His 35 is a little more difficult to
rationalize. As well as suggestions that it may be in the vicinity
of the electron transfer locus for small molecule redox reagents
(192,194), it is implicated in the pH dependent conformational change
observed in reduced azurin (214). The possible nature of the change
with pH can be guessed at from the structure since the imidazole
ring can be oriented with its ND2 and NE2 atoms adjacent either
to two proton acceptors (87 O and 11 COO^-), or to one proton donor
(37 NH) and one proton acceptor (44 O). Thus in the deprotonated
inactive form, the sidechain may be oriented so that hydrogen bonds
are formed between NE2 and the C=O of Met 44, and ND1 and the NH
of Gly 37. On the other hand, in its protonated or active form it may
be oriented so that hydrogen bonds are formed between ND1 and the
carboxyl sidechain of the invariant Asp 11, and NE2 and the C=O
of Ile 87. It is more difficult to see, however, how any direct
interaction of the sidechain of His 35 with a macromolecular redox
agent can occur, as in this structure the sidechain appears quite
inaccessible. Although it is close to the surface, it is covered

by two adjacent mainchain loops, 88-90 and 36-39, which would have to move apart to allow access to the imidazole ring. Whether this occurs when a macromolecular complex is formed, or when the hydrogen bonding pattern is changed by a pH change,is a matter for speculation. The unusually slow acid-base exchange rate found for this residue in Ps. aeruginosa azurin (62) would suggest that the sidechain is normally inaccessible to solvent in that structure as well.

In plastocyanin, however, His 35 is replaced by Ala 33, and there is no suggestion that this residue is involved in electron transfer. It seems more likely,therefore, that the structure around His 35 is not a binding site for any redox partner, but that the conformational change involving His 35 has an indirect effect on the actual binding site, perhaps the hydrophobic patch, some 7-8Å away. This could occur for example through movement of Asp 11, or, more probably, through movement of the structure around Met 44 (part of the hydrophobic patch) when the hydrogen bond between 35 NE2 and 44 0 is broken.

Control of the electron transfer mechanism, i.e. the signal that the protein is ready to donate or accept an electron may therefore be related to a pH dependent conformational change, either in the copper co-ordination sphere (as has been shown crystallographically to be the case for plastocyanin (106)), or in His 35 (as suggested here for azurin). As a similar pH dependent change has been observed for cytochrome c-551 (214,216), such equilibria may be acting as sensitive negative feed-back mechanisms for controlling the flux of elctrons down the electron transfer chain. In order to confirm the exact nature of such changes, however, a series of accurate three dimensional structure analyses of both oxidised and reduced azurin between pH 3.0 and pH 9.0 is necessary. Not only would the results of such a study have direct relevance to possible electron transfer

pathways,but they may elucidate the conformational changes required to bring about the decolourisation of the protein at the extremes of this range (84,224).

It is possible that the electron transfer pathways of azurin and plastocyanin may have diverged somewhat, so that although they may, for instance, have similar binding sites and electron transfer mechanisms for oxidation (say the hydrophobic patch and His 117 or His 87),those for reduction may differ (for example,involve the acidic patch in plastocyanin, and His 35 in azurin).On evolutionary grounds, however, it is more likely that the binding sites and electron transfer mechanisms will be the same in both proteins. Thus comparison of the three dimensional structures of azurin and plastocyanin suggests that oxidation and reduction may proceed via the same electron transfer pathway, this being the exposed edge of His 117 (or His 87 for plastocyanin) with the hydrophobic patch acting as a binding site. The pH effects observed by other workers (192) are therefore more likely to be related to conformational changes remote from the active site, but in some way, as yet unknown, essential to its function.

Finally, it is interesting to extend the homologies between azurin and plastocyanin to other, larger, blue proteins such as laccase, ascorbate oxidase, and perhaps cytochrome oxidase (81). Although the latter is not usually considered as a blue protein, it is known to contain a Type I copper in subunit II, and the amino acid sequence of this subunit is very similar to that of azurin, especially about the copper binding site. However, although the ESR spectrum of this copper (one of the two per molecule) is typical of a Type I copper, it has even narrower hyperfine structure (81), and its optical spectra cannot be studied due to the presence of the heme. The "best" alignment of the sequences shows that the extended loop between residues 37 and 46 of the azurin structure, which is

deleted in plastocyanin, is present in subunit II of cytochrome c oxidase, and the sequence around Gly 45 and His 46 in azurin, Met-Gly-His-Asn-Trp, is matched by Met-Gly-His-Gln-Trp in cytochrome c oxidase(see Figure 1.2.) Since the structural differences between plastocyanin and azurin in this region may be responsible for the observed difference in the co-ordination sphere of the copper (viz. the nearness of the carbonyl oxygen of Gly 45 to the copper atom in azurin), it may be that an analogous close contact between copper and a carbonyl oxygen also occurs in cytochrome oxidase.

Because of the great difference in size between subunit II of cytochrome oxidase and azurin, (227 residues c.f. 129), large additions to the azurin sequence have to be made in order to achieve proper alignment. The largest of these are at the beginning and end of the polypeptide chain. The part of the sequence which is aligned with that of azurin shows a surprising degree of homology, considering the diverse nature of the proteins involved. Thus, 11 residues are invariant between subunit II and all azurins, with a further 23 conservatively substituted (81). With this alignment, most of the insertions required, in addition to those at the N- and C- terminal ends, are in loops at the ends of β strands, or in the "flap" region, and so are not important to the topology. It is thus reasonable to postulate a β-barrel-type structure for this subunit.

It will be interesting to see if further characterisation of the sequences of other multi-copper oxidases such as laccase and ascorbate oxidase,which contain at least one "Type I" copper, match those of either azurin or plastocyanin. If, when the full sequences are known, they do appear to have evolved from the same ancestral gene as azurin and plastocyanin, prediction of their three dimensional structures may be possible.

APPENDIX I

The Purification and Crystallization of Cytochromes c'
from Alcaligenes denitrificans and
Alcaligenes sp.

Cytochrome c', like azurin, is thought to be involved in the electron transport chain of photosynthetic and denitrifying bacteria (54,225,226,227). In cytochrome c' the iron atom is in a high spin state, like those in deoxy-haemoglobin and myoglobin, but unlike the iron in eukaryotic cytochrome c. Cytochrome c' also has very distinct magnetic properties,unlike any other haem protein (227, 228,229). All cytochromes c' so far investigated have shown unusual spectral properties, such as a split Soret band, and a dependence upon pH (54,91,225,228,230). They are water soluble, auto-oxidisable and exhibit intermediate oxidation-reduction potentials of between 0 and 150 mV, at or above pH 7.0 (54,91,225). The protein usually exists as a dimer (231), with identical subunits of molecular weight 14,000,each containing a heme group bound to the protein through cysteine-thioether bonds. In contrast to other cytochromes c, the latter connections are near the carboxy rather than the amino terminus, with His as a fifth ligand and no sixth bond to Met (86,91,232).Cytochrome c' can be involved in both aerobic and anaerobic metabolism in Pseudomonas species (91), but it appears to be mainly associated with anaerobic photosynthetic bacteria, where it has been shown to play an indirect role in the electron transport pathway (148).

Purification

The solutions of cytochromes from the carboxy-methyl cellulose columns (Section 2.2(iv)) were concentrated by ultra-filtration and

re-equilibrated in 25 mM tris-HCl buffer, pH 8.5, by gel filtration
(Sephadex G-25 medium). Chromatography on diethylaminoethyl-
cellulose (Whatman DEAE-52) was then carried out to remove any
nucleic acids remaining in solution. The cytochrome adsorbed
weakly to the column and could be eluted with 0.05 M tris-HCl buffer,
pH 8.5. The resulting solution was concentrated by ultra-filtration
before finely powdered ammonium sulphate was added slowly with
stirring. During the addition, small amounts of other cytochromes
were removed by centrifugation at approximately 38, 49 and 55%
saturation (the cytochromes being identified as c-type cytochromes
by their spectra). At 60% saturation, cytochrome c' was precipitated,
the resulting pellet being dissolved in a minimum amount of distilled
water and subjected to gel filtration (Sephadex G-75 fine)
under two different sets of conditions, as described by Ambler (86).
The best material obtained had a characteristic spectrum (Figure I-1),
and, as can be seen, could easily be distinguished from the c-type
cytochromes. It was homogeneous by S.D.S. acrylamide gel electro-
phoresis (93) and had a spectral ratio, Soret to 280 nm, of 0.33 -
0.35 (91). The yield was approximately 0.3 mg of protein per litre
of culture. Protein concentration was calculated using the extinction
coefficients given by Cusanovich (91).

Crystallization of the cytochromes c'.

Cytochromes c' from several bacterial sources have previously
been crystallised. These include Ps. denitrificans (Alc. sp. of
this work), Rhodospirillum rubrum, Rhodopseudomonas palustris,
Rhodospirillum molischianum and Rhodopseudomonas capsulata (232,
233,234). Of these, the first crystallised as hexagonal bipyramids
which were too small for X-ray work, the next three crystallised in
the orthorhombic space group $P2_12_12_1$ with 1 or 4 dimer(s) per
asymmetric unit, while the last crystallised in the hexagonal space
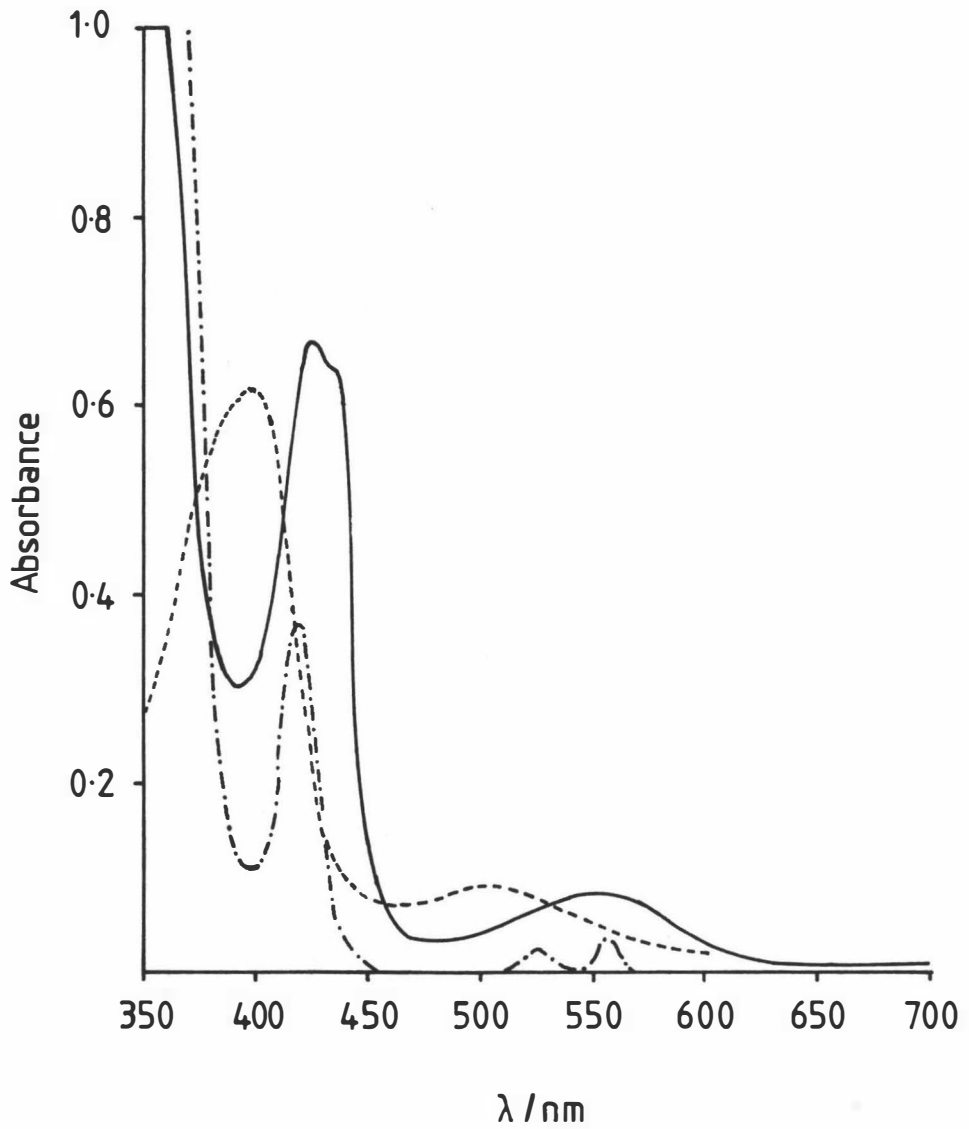
Figure I-1. Absorption spectra of cytochrome c'

——— cytochrome c' reduced

------ cytochrome c' oxidised

·—··—··— cytochrome c reduced

group P 6_2 (or P 6_4) with one subunit per asymmetric unit. While this work was in progress the structure of cytochrome c' from R. molischianum was reported at 2.5Å resolution (235).

Cytochrome c' from both Alc. denitrificans and Alc. sp. could be crystallised in three ways.

(a) By vapour diffusion: Finely powdered ammonium sulphate was added slowly with stirring to the protein solution (2-10 mg ml^{-1} in 0.1 M phosphate buffer pH 8.0) until the solution showed a faint turbidity (approximately 50% saturated). Buffer was then added drop by drop until the turbidity vanished, and the solution was centrifuged at high speed before being left to equilibrate against 95% saturated ammonium sulphate at room temperature.

(b) By microdialysis: The protein (approximately 25 mg) was precipitated with ammonium sulphate, centrifuged, and the pellet made up to 1 ml with 0.1 M phosphate buffer pH 8.0. This solution was then dialysed against a 95% saturated ammonium sulphate solution made up with the same buffer and made 1 M with sodium chloride.

(c) By free interface diffusion: 50 μl of protein solution made up as in (b) was carefully layered on 200 μl of saturated ammonium sulphate solution made using 0.1 M phosphate buffer pH 8.0.

The largest crystals were obtained using the first method and usually appeared within 3-4 weeks. Those from Alc. denitrificans were chunky hexagonal bipyramids, very dark brown in colour and up to 3 mm in length. (Figure I-2). Those from Alc. sp. were dark brown hexagonal needles up to 5 mm long with a cross section about 1/6 their length at one end, tapering to the other (Figure I-3).
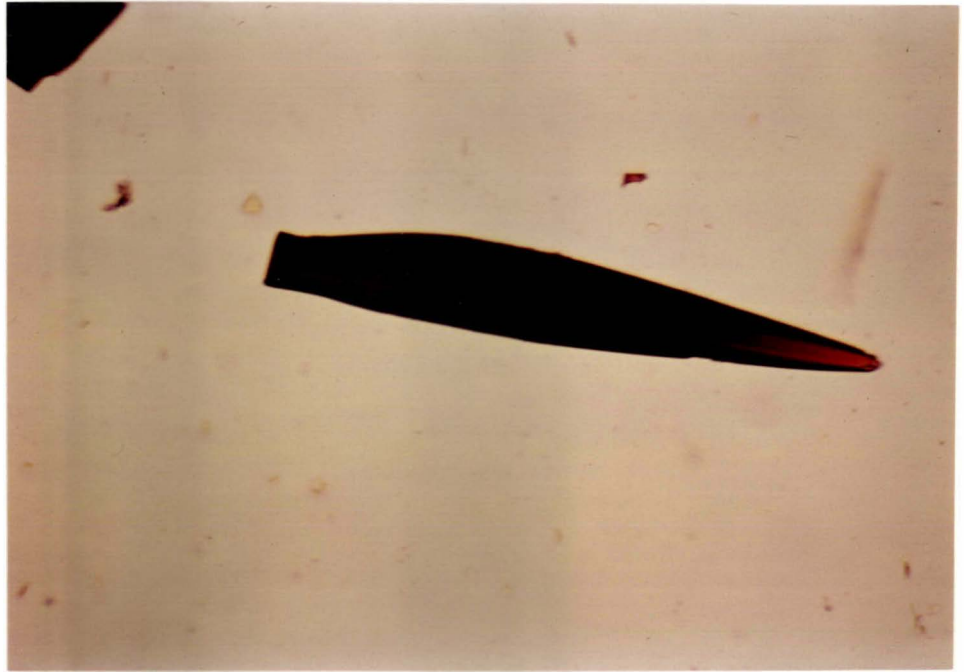
Figure I-2 Crystals of cytochrome c' from *Alcaligenes denitrificans* (× 63).
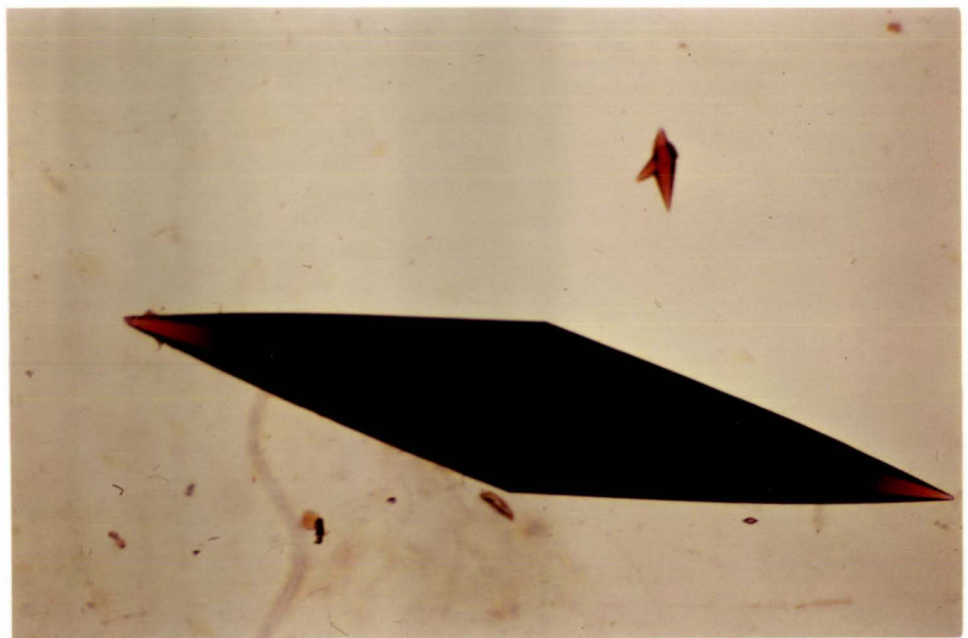
Figure I-3 Crystals of cytochrome c' from *Alcaligenes sp* (× 63).

Characterization of the Crystals.

Both cytochromes c' crystallized in the same hexagonal space group $6_1 22$ (or its enantiomorph P $6_5 22$). The length of the non-unique axes and the intensity distribution in the diffraction pattern seemed to be identical in both protein crystals, so that apart from a small difference in the c axis, they can be considered isomorphous. Table I-1 summarizes the unit cell parameters for both crystals.

Table I-1.   The Unit Cell Parameters for Cytochrome c' from Alc. denitrificans and Alc. sp.

Source	Alcaligenes denitrificans	Alcaligenes sp.
a(Å)	54.7	54.7
b(Å)	54.7	54.7
c(Å)	181.5	187.3
α^o	90	90
β^o	90	90
γ^o	120	120
V	$4.703 \times 10^5 \text{Å}^3$	$4.85 \times 10^5 \text{Å}^3$
ρ^a	1.22 g cm^{-3}	1.303 g cm^{-3}
V_m^b	2.8 Å^3 dalton^{-1}	2.89 Å^3 dalton^{-1}
X_p^c	44%	43%
n^d	1 subunit	1 subunit

a   Crystal densities were measured by a density gradient column (112) made up with water-saturated bromobenzene/kerosene with a range of 1.059 g cm^{-3} – 1.426 g cm^{-3}

b   crystal volume per unit of protein molecular weight (113)

c   fraction of the unit cell occupied by the protein (113)

d   molecules of protein per asymmetric unit.

The crystals diffracted well, the diffraction pattern extending strongly to at least 2Å. A value of V_m very similar to that found for R. capsulata (2.77 Å dalton^{-1}) suggests a similar packing of the symmetry-related subunits for both proteins.

APPENDIX II

Computer Programs Used

The authors whose names are given in parentheses have made important modifications to the programs, and/or have adapted the programs to run on the B6700 of the Massey University Computer Centre.

1. DENSFOR

    C.E.Nockolds and R.H. Kretsinger, University of Virginia, U.S.A.

    (B.F. Anderson and S.V. Rumball, Massey University, New Zealand).

This program extracts integrated, background-corrected intensities from density data output on magnetic tape by the Optronics rotating drum microdensitometer. A "fiducial hole" must be punched in each film at the approximate centre of the reciprocal lattice nett. It is assumed that one axis in the diffraction pattern lies within 0.5° of the vertical direction of the film. Thus an initial transformation matrix can be derived, and this is then refined using a row of reflections parallel to the densitometer x direction, and containing a reasonable number of strong reflections. This matrix is in turn used to predict the approximate positions of all reflections on the film. The intensity for each reflection is integrated over a box, the dimensions of which can be entered into the program to allow for variations in spot shape and size. The background is measured either side of the spot in the x or y direction and subtracted from the integrated intensity.

2. VENUS

    B.W. Matthews,C.E. Klopfenstein, and P.M. Colman. J. Phys.

    E5 (1972) 353.

    (B.F. Anderson, Massey University, New Zealand)

Converts the output from DENSFOR to structure factor amplitudes. In

doing this it averages all symmetry-related reflections, calculating

a merging R factor for the latter, and ensures that systematically

absent reflections are removed. Film-to-film scale factors and

Lorentz-polarisation corrections are applied.

3. WPSCALE

    B.F. Anderson, Massey University, New Zealand.

This program uses a Wilson-type plot to scale the derivative

intensities, F^2_{PH} to the native intensities F_P^2. Values of

$Log(\Sigma F_{PH}^2/\Sigma F_P^2)$ in blocks of $(\sin\theta/\lambda)^2$ are calculated using linear

regression procedures to obtain 2B (the slope) and the scale factor

K   (intercept). The derivatives are scaled using the following

equation.

$$F_{PH}^2(corr) = KF^2_{PH} \exp(2B\sin^2\theta/\lambda^2)$$

The output is a list of reflections with amplitudes F_P and F_{PH}, the

derivative intensity F_{PH}^2 and the isomorphous difference Δ_{iso}

(where $\Delta_{iso} = (F_{PH} - F_P)$).

4. FOURIER

    A. Zalkin, University of California, U.S.A.

    R.J. Dellaca, Canterbury University, New Zealand.

A rewritten and expanded version of the program FORDAP by A. Zalkin,

written by R.J. Dellaca, Canterbury University, New Zealand. It

calculates Fourier syntheses, and Patterson maps.

5. CUCLS

    W.R. Busing, K.O. Martin, and H.A. Levy. Oak Ridge National
    Laboratory, U.S.A.

    R.J. Dellaca, Canterbury University, New Zealand.

A modified version of ORFLS (by Busing, Martin and Levy) written by
R.J. Dellaca. A full matrix least squares program for refinement.

6. HILGNIX

    R.J. Dellaca, Canterbury University, New Zealand.

    P.E. Nixon,   Auckland University, New Zealand.

    (E.N. Baker, Massey University, New Zealand.)

HILGNIX is a processing program for data collected on a Hilger and
Watts diffractometer. Parts of it are based on the program DRED
(by J.F. Blount) and parts on the program PICKOUT (by R.J. Doedens).
Key subroutines include the following.

(a) Input: Reads the data from the paper tape. Mis-setting of the
     crystal is frequently revealed by unequal backgrounds. Although
     the peak count is often relatively unaffected, the uneven
     backgrounds adversely affect the background-averaging routine.
     To avoid this, if one background is significantly greater than
     the other, both are set equal to the lower of the two.

(b) Stdiv: Corrects for radiation damage to the crystal. A linear
     correction is made, where the corrected intensity

$$I_{corr} = I/K, \quad \text{where}$$

$$K = 1 - (N/Nt)P/100$$

P = % drop in intensity in a
block of data

N = reflection number

Nt = total number of reflections
in the block of data.

(c) <u>Backav</u>: Averages the backgrounds, enabling them to be counted for a shorter time than is usual. This procedure is discussed fully in Section 2.6.2 under the heading "Data Processing".

(d) <u>Bsublp</u>: Applies background correction and Lp corrections.

(e) <u>Empabs</u>: Corrects for absorption using an empirical absorption curve as described by North, Phillips and Matthews (130). See Section 2.5.2.

7. MERGEDATA

E.N. Baker, Massey University, New Zealand.

Shells of data are merged into one unique data set. A weighted mean is taken for reflections measured more than once, and for equivalent reflections. For heavy atom derivatives, Friedel pairs are kept separate. The average intensity is calculated in intervals of $\sin^2\theta$, as is an R index measuring the agreement between the Friedel pairs

$$R_F = \Sigma|F^+ - F^-|/\Sigma\tfrac{1}{2}|F^+ + F^-|$$

8. PHASIP

E.N. Baker, Massey University, New Zealand.

This program scales the native and derivative data sets together using procedures described by Wyckoff et al. (132). All derivative data is scaled by a factor k, where

$$k = \frac{\Sigma F_P^2}{\Sigma F_{PH}^2}\left(1 + \left(\frac{\Sigma|F_{PH} - F_P|}{\Sigma F_P}\right)^2\right)$$

The output is a file of reflections hkl with structure factor amplitudes for the native crystals and for each heavy atom derivative (with the anomalous data being retained for the derivatives).

9. PHASE

    B.W. Matthews, Institute for Molecular Biology, University of

    Oregon, Eugene, Oregon, U.S.A.

    (B.F. Anderson, E.N. Baker, Massey University, New Zealand).

The program determines phases from isomorphous replacement

measurements. Errors are taken into account using the method of

Blow and Crick (153,158). Both the "most probable" and "best" phases

are calculated. Anomalous scattering measurements may be included

in the phase determination as described by North (155) and Matthews

(147).

10. MODFIT

    N. Isaacs and E.J. Dodson, Chemistry Department, University

    of York, United Kingdom.

    (E.N. Baker, Massey University, New Zealand).

A program which adjusts a set of protein co-ordinates by least

squares to obtain an acceptable geometry. It uses an iterative

conjugent gradient algorithm for the minimisation. (For details see

Rollett "Computing Methods in Crystallography"). A complete

description of the method and program is given by Dodson et al. (237).

11. NOL1N6

    W.G.J. Hol, Department of Chemical Physics, University of Groningen,

    The Netherlands.

    (E.N. Baker, Massey University, New Zealand).

This program enables all or part of one protein molecule to be

superimposed on another. An iterative non-linear least squares method

is used, based on that of Rao and Rossman (163). See also section 3.12

12. ORTEP

K.C. Johnson, Oak Ridge National Laboratory Report, ORNL-5138 (1976) U.S.A.

(B.F. Anderson, Massey University, New Zealand).

This program produces diagrams of molecular structures, either in mono or in stereo, from supplied atomic co-ordinates.

13. CONTOUR

E.J. Dodson. University of York, U.K.

APPENDIX III

The Removal of Cu by KCN

In solution, the Cu atom can readily be removed by dialysis of the reduced protein against 0.5 M cyanide (236), and of the oxidised protein against 0.1 M cyanide (97). It was found that soaking the azurin crystals in 0.1 M KCN solution at room temperature caused the crystals to become colourless within 1 week. When this was originally discovered (when crystals were first obtained), it was observed, however, that after a period of time the blue colour of the crystals gradually returned, and it was thus assumed that the copper had been reduced rather than removed.

Near the end of the structure determination when interest was centred on finding an isomorphous reduced crystal, the effect of KCN was recalled, and the experiment repeated. In order to establish the fact that the copper had in fact been reduced and not removed, atomic absorption analysis for copper was carried out on some crystals that were treated in the following way: Some small crystals were soaked in a 0.1 M KCN solution, 75% saturated with ammonium sulphate, until they became colourless. They were then filtered, washed with a 75% saturated solution of ammonium sulphate, before being dissolved in deionized water and subjected to atomic absorption analysis. The results of this analysis showed the copper concentration to be approximately 10% of the protein concentration (as measured by the absorbance at 280 nm), indicating that most of the copper had been removed from the protein.

Some crystals, large enough for photography, had been treated in the same way, and as these had no surface cracks, suggesting no major conformational changes in the protein, a precession photograph was taken of the hoℓ zone of one of them. This showed that the

crystal diffracted well, although there appeared to be a slight
increase in both the a and c axial lengths (a = 75.3 c.f. 75.0Å ;
c = 99.6Å c.f. 99.5Å).

If this had been realized when the search for heavy atom
derivatives was being undertaken, metal replacement experiments
would have been carried out with the hope of isomorphism being
restored on the binding of another metal. Several such replacements
have been successfully carried out in the process of spectral
investigations into the blue copper site (17,18,42,43) and although
these have involved metals with atomic numbers smaller than that of
copper, i.e. nickel, cobalt and manganese,it is possible that
substitution of the metal site by mercury or cadmium might have been
possible, resulting in a useful change of scattering material.

APPENDIX 4.   The Atomic Co-ordinates

AZURIN   MOLECULE I

ALA   1
  N   -.317 0.223 0.024
  CA   -.301 0.231 0.018
  C   -.295 0.247 0.026
  O   -.287 0.245 0.037
  CB   -.304 0.235 0.003

GLN   2
  N   -.299 0.263 0.021
  CA   -.293 0.279 0.029
  C   -.274 0.284 0.026
  O   -.267 0.298 0.031
  CB   -.296 0.275 0.044
  CG   -.311 0.261 0.045
  CD   -.319 0.261 0.060
  OE1 -.317 0.274 0.067
  NE2 -.328 0.246 0.063

CYS   3
  N   -.266 0.273 0.017
  CA   -.247 0.277 0.013
  C   -.235 0.283 0.025
  O   -.234 0.299 0.028
  CB   -.245 0.288 0.000
  SG   -.251 0.311 0.001

GLU   4
  N   -.227 0.269 0.031
  CA   -.214 0.270 0.042
  C   -.212 0.250 0.047
  O   -.223 0.239 0.043
  CB   -.219 0.283 0.053
  CG   -.236 0.279 0.061
  CD   -.231 0.270 0.074
  OE1 -.227 0.254 0.074
  OE2 -.231 0.281 0.084

ALA   5
  N   -.197 0.247 0.054
  CA   -.193 0.228 0.057
  C   -.197 0.223 0.071
  O   -.212 0.226 0.077
  CB   -.174 0.223 0.053

THR   6
  N   -.184 0.214 0.077
  CA   -.182 0.207 0.090
  C   -.172 0.189 0.090
  O   -.180 0.175 0.086
  CB   -.198 0.207 0.100
  OG1 -.211 0.220 0.096
  CG2 -.192 0.210 0.114

ILE   7
  N   -.155 0.190 0.094
  CA   -.145 0.173 0.093
  C   -.137 0.168 0.107
  O   -.140 0.179 0.117
  CB   -.130 0.176 0.083
  CG1 -.118 0.159 0.082
  CD1 -.101 0.164 0.073
  CG2 -.120 0.193 0.085

GLU   8
  N   -.129 0.153 0.108
  CA   -.122 0.146 0.121
  C   -.102 0.141 0.120
  O   -.096 0.136 0.109
  CB   -.134 0.131 0.126
  CG   -.147 0.138 0.136
  CD   -.162 0.124 0.139
  OE1 -.177 0.130 0.142
  OE2 -.158 0.108 0.137

SER   9
  N   -.094 0.141 0.132
  CA   -.075 0.135 0.134
  C   -.075 0.123 0.146
  O   -.085 0.127 0.156
  CB   -.064 0.152 0.136
  OG   -.045 0.146 0.136

ASN   10
  N   -.065 0.109 0.145
  CA   -.064 0.096 0.157
  C   -.045 0.096 0.163
  O   -.032 0.102 0.156
  CB   -.071 0.077 0.154
  CG   -.065 0.067 0.142
  OD1 -.062 0.050 0.143
  ND2 -.064 0.076 0.131

ASP   11
  N   -.043 0.087 0.174
  CA   -.025 0.084 0.180
  C   -.018 0.066 0.174
  O   -.002 0.063 0.173
  CB   -.025 0.084 0.195
  CG   -.038 0.070 0.201
  OD1 -.051 0.075 0.208
  OD2 -.033 0.053 0.198

ALA   12
  N   -.032 0.056 0.170
  CA   -.029 0.038 0.164
  C   -.019 0.039 0.151
  O   -.016 0.026 0.144
  CB   -.047 0.029 0.162

MET   13
  N   -.013 0.056 0.148
  CA   -.002 0.059 0.136
  C   -.013 0.057 0.123
  O   -.013 0.043 0.117
  CB   0.014 0.046 0.136
  CG   0.026 0.050 0.149
  SD   0.031 0.031 0.159
  CE   0.051 0.036 0.168

GLN   14
  N   -.021 0.072 0.119
  CA   -.030 0.073 0.106
  C   -.043 0.089 0.104
  O   -.044 0.101 0.112
  CB   -.037 0.055 0.100
  CG   -.056 0.050 0.105
  CD   -.056 0.031 0.112
  OE1 -.042 0.022 0.113
  NE2 -.072 0.027 0.117

TYR   15
  N   -.052 0.088 0.092
  CA   -.064 0.103 0.088
  C   -.081 0.094 0.082
  O   -.080 0.081 0.075
  CB   -.057 0.113 0.076
  CG   -.053 0.134 0.078
  CD1 -.045 0.143 0.067
  CE1 -.041 0.161 0.069
  CZ   -.043 0.170 0.081
  CE2 -.051 0.160 0.092
  CD2 -.056 0.142 0.090
  OH   -.038 0.188 0.083

ASP   16
  N   -.095 0.103 0.085
  CA   -.114 0.100 0.081
  C   -.119 0.102 0.066
  O   -.117 0.117 0.060
  CB   -.125 0.113 0.089
  CG   -.144 0.105 0.092
  OD1 -.147 0.091 0.085
  OD2 -.153 0.112 0.101

LEU   17
  N   -.128 0.088 0.061
  CA   -.138 0.091 0.048
  C   -.128 0.084 0.036
  O   -.129 0.068 0.033
  CB   -.141 0.112 0.047
  CG   -.152 0.119 0.059
  CD1 -.170 0.110 0.059

CD2 -.154 0.140 0.057

LYS   18
  N   -.121 0.097 0.028
  CA   -.115 0.093 0.014
  C   -.110 0.111 0.008
  O   -.095 0.117 0.008
  CB   -.131 0.084 0.007
  CG   -.143 0.089 0.015
  CD   -.164 0.086 0.006
  CE   -.181 0.083 0.015
  NZ   -.178 0.070 0.025

GLU   19
  N   -.124 0.120 0.003
  CA   -.123 0.137 -.004
  C   -.137 0.151 0.000
  O   -.152 0.146 0.003
  CB   -.127 0.132 -.019
  CG   -.129 0.111 -.020
  CD   -.149 0.105 -.018
  OE1 -.152 0.088 -.020
  OE2 -.160 0.117 -.014

MET   20
  N   -.131 0.168 -.001
  CA   -.143 0.183 0.001
  C   -.149 0.191 -.013
  O   -.141 0.186 -.023
  CB   -.134 0.198 0.009
  CG   -.122 0.191 0.020
  SD   -.129 0.195 0.037
  CE   -.131 0.173 0.044

VAL   21
  N   -.162 0.202 -.013
  CA   -.170 0.211 -.025
  C   -.180 0.229 -.021
  O   -.191 0.229 -.012
  CB   -.184 0.198 -.031
  CG1 -.177 0.179 -.033
  CG2 -.191 0.205 -.044

VAL   22
  N   -.176 0.243 -.029
  CA   -.184 0.260 -.027
  C   -.200 0.266 -.035
  O   -.209 0.279 -.031
  CB   -.170 0.275 -.028
  CG1 -.152 0.269 -.023
  CG2 -.176 0.292 -.020

ASP   23
  N   -.205 0.256 -.046

```
     CA  -.220 0.261 -.054          OE1 -.191 0.405 0.037          CD2 -.061 0.213 0.115          SD   0.026 0.084 0.222
     C   -.217 0.280 -.060          NE2 -.191 0.431 0.024                                         CE   0.020 0.079 0.205
     O   -.204 0.290 -.057                                      LYS   34
     CB  -.237 0.261 -.046      PHE   29                             N   -.120 0.192 0.139      ALA   40
     CG  -.252 0.250 -.053          N   -.165 0.344 0.013          CA  -.123 0.186 0.153          N    0.010 0.160 0.234
     OD1 -.246 0.235 -.058          CA  -.164 0.326 0.019          C   -.109 0.172 0.158          CA   0.006 0.180 0.234
     OD2 -.267 0.256 -.053          C   -.160 0.327 0.035          O   -.105 0.159 0.151          C    0.013 0.193 0.226
                                    O   -.152 0.340 0.040          CB  -.142 0.179 0.155          O    0.020 0.209 0.230
LYS   24                           CB  -.150 0.314 0.012                                         CB   0.001 0.187 0.248
     N   -.226 0.284 -.071          CG  -.157 0.295 0.013      HIS   35
     CA  -.225 0.301 -.077          CD1 -.144 0.280 0.013          N   -.102 0.175 0.170      LYS   41
     C   -.238 0.316 -.072          CE1 -.150 0.263 0.015          CA  -.089 0.161 0.175          N    0.027 0.187 0.215
     O   -.236 0.332 -.076          CZ  -.168 0.259 0.018          C   -.099 0.149 0.185          CA   0.038 0.201 0.209
     CB  -.225 0.299 -.093          CE2 -.181 0.274 0.018          O   -.097 0.150 0.197          C    0.057 0.198 0.215
     CG  -.210 0.285 -.096          CD2 -.175 0.291 0.016          CB  -.072 0.170 0.180          O    0.058 0.195 0.228
     CD  -.202 0.288 -.110                                         CG  -.055 0.158 0.180          CB   0.032 0.219 0.216
     CE  -.199 0.270 -.117      THR   30                           ND1 -.056 0.141 0.186          CG   0.017 0.229 0.203
     NZ  -.181 0.270 -.124          N   -.165 0.312 0.041          CE1 -.039 0.135 0.185          CD   0.014 0.247 0.214
                                    CA  -.165 0.311 0.056          NE2 -.029 0.147 0.178          CE   0.012 0.262 0.203
SER   25                           C   -.160 0.292 0.062          CD2 -.039 0.162 0.175          NZ   0.022 0.278 0.207
     N   -.249 0.310 -.063          O   -.170 0.279 0.060
     CA  -.262 0.322 -.056          CB  -.186 0.313 0.060      VAL   36                          ALA   42
     C   -.252 0.335 -.046          OG1 -.192 0.328 0.052          N   -.111 0.138 0.179          N    0.070 0.198 0.207
     O   -.257 0.350 -.045          CG2 -.188 0.317 0.075          CA  -.121 0.126 0.188          CA   0.088 0.192 0.211
     CB  -.275 0.310 -.048                                         C   -.108 0.112 0.195          C    0.089 0.172 0.209
     OG  -.271 0.292 -.051      VAL   31                           O   -.112 0.096 0.196          O    0.103 0.164 0.208
                                    N   -.145 0.291 0.069          CB  -.137 0.117 0.182          CB   0.088 0.194 0.227
CYS   26                           CA  -.139 0.273 0.074          CG1 -.146 0.104 0.192
     N   -.240 0.326 -.040          C   -.139 0.269 0.089          CG2 -.151 0.131 0.178      VAL   43
     CA  -.230 0.334 -.028          O   -.133 0.280 0.097                                         N    0.073 0.166 0.204
     C   -.215 0.346 -.034          CB  -.120 0.270 0.069      GLY   37                           CA   0.071 0.147 0.200
     O   -.204 0.340 -.042          CG1 -.116 0.250 0.069          N   -.093 0.120 0.199          C    0.052 0.143 0.196
     CB  -.223 0.320 -.019          CG2 -.107 0.280 0.078          CA  -.078 0.109 0.204          O    0.039 0.149 0.203
     SG  -.226 0.323 -.001                                         C   -.072 0.115 0.218          CB   0.079 0.134 0.210
                                HIS   32                           O   -.031 0.125 0.226          CG1  0.071 0.140 0.224
LYS   27                           N   -.146 0.253 0.093                                         CG2  0.099 0.136 0.210
     N   -.214 0.362 -.028          CA  -.146 0.248 0.107      LYS   38
     CA  -.193 0.373 -.032          C   -.137 0.230 0.110          N   -.056 0.108 0.221      MET   44
     C   -.185 0.373 -.020          O   -.143 0.215 0.107          CA  -.048 0.108 0.234          N    0.050 0.133 0.185
     O   -.168 0.374 -.022          CB  -.167 0.246 0.110          C   -.027 0.109 0.233          CA   0.032 0.128 0.180
     CB  -.203 0.392 -.036          CG  -.175 0.264 0.108          O   -.019 0.095 0.231          C    0.021 0.145 0.177
     CG  -.188 0.405 -.030          ND1 -.194 0.266 0.108          CB  -.053 0.090 0.241          O    0.005 0.144 0.176
     CD  -.196 0.424 -.027          CE1 -.196 0.283 0.106          CG  -.051 0.092 0.256          CB   0.021 0.113 0.186
     CE  -.182 0.438 -.030          NE2 -.181 0.292 0.104          CD  -.053 0.077 0.263          CG   0.012 0.104 0.174
     NZ  -.190 0.456 -.029          CD2 -.167 0.280 0.105          CE  -.053 0.059 0.256          SD   0.028 0.097 0.161
                                                                   NZ  -.065 0.044 0.264          CE   0.038 0.073 0.170
GLN   28                        LEU   33
     N   -.192 0.372 -.008          N   -.123 0.231 0.118      MET   39                          GLY   45
     CA  -.181 0.371 0.004          CA  -.113 0.215 0.122          N   -.021 0.126 0.234          N    0.031 0.160 0.174
     C   -.181 0.351 0.010          C   -.118 0.210 0.137          CA  -.002 0.130 0.232          CA   0.022 0.177 0.171
     O   -.195 0.343 0.011          O   -.118 0.221 0.146          C   -.002 0.150 0.229          C    0.014 0.176 0.157
     CB  -.189 0.382 0.016          CB  -.092 0.219 0.123          O   -.015 0.156 0.222          O    0.019 0.164 0.149
     CG  -.187 0.403 0.013          CG  -.081 0.206 0.114          CB   0.006 0.117 0.222
     CD  -.190 0.413 0.026          CD1 -.081 0.186 0.119          CG   0.007 0.098 0.228      HIS   46
```

```
N    0.003 0.189 0.154        CD2 -.061 0.235 -.001        O   0.093 0.207 0.011        CB  0.161 0.288 0.041
CA  -.005 0.159 0.140                                      CB  0.068 0.187 -.015        CG  0.176 0.299 0.035
C   -.008 0.208 0.134        THR   51                      CG  0.052 0.175 -.021        OD1 0.172 0.308 0.025
O   -.010 0.221 0.142         N  -.002 0.259 -.005         CD  0.086 0.160 -.035        OD2 0.191 0.299 0.041
CB  -.023 0.178 0.141         CA  0.011 0.263 -.015        CE  0.106 0.177 -.038
CG  -.019 0.158 0.143         C   0.009 0.254 -.023        NZ  0.111 0.186 -.051        GLY   63
ND1 -.006 0.149 0.136         O   0.002 0.239 -.029                                      N   0.146 0.271 0.069
CE1 -.007 0.132 0.140         CB  0.030 0.257 -.009        GLU   57                      CA  0.137 0.275 0.081
NE2 -.019 0.129 0.149         OG1 0.043 0.270 -.014         N   0.104 0.190 -.006        C   0.146 0.266 0.093
CD2 -.027 0.146 0.151         CG2 0.035 0.239 -.016         CA  0.120 0.186 0.003        O   0.151 0.274 0.103
                                                            C   0.129 0.205 0.006
ASN   47                     LYS   52                       O   0.135 0.208 0.017        MET   64
 N  -.008 0.207 0.121         N   0.017 0.263 -.038         CB  0.133 0.174 -.005        N   0.149 0.248 0.092
 CA -.008 0.223 0.112         CA  0.020 0.254 -.051         CG  0.153 0.179 -.003        CA  0.158 0.238 0.103
 C  -.020 0.218 0.100         C   0.029 0.236 -.048         CD  0.160 0.187 -.016        C   0.173 0.250 0.109
 O  -.025 0.203 0.099         O   0.041 0.235 -.040         OE1 0.158 0.178 -.026        O   0.177 0.248 0.121
 CB  0.011 0.224 0.107        CB  0.035 0.266 -.059         OE2 0.168 0.202 -.014        CB  0.164 0.219 0.098
 CG  0.020 0.243 0.106        CG  0.026 0.280 -.068                                      CG  0.167 0.207 0.110
 OD1 0.037 0.243 0.104        CD  0.041 0.289 -.076        GLY   58                      SD  0.169 0.183 0.105
 ND2 0.010 0.257 0.109        CE  0.038 0.286 -.091         N   0.129 0.215 -.005        CE  0.189 0.175 0.113
                              NZ  0.043 0.303 -.098         CA  0.135 0.234 -.005
TRP   48                                                    C   0.132 0.245 0.007        ASN   65
 N  -.022 0.232 0.091        GLU   53                       O   0.145 0.252 0.013         N   0.180 0.261 0.100
 CA -.031 0.229 0.078         N   0.023 0.223 -.056                                      CA  0.194 0.274 0.105
 C  -.025 0.245 0.069         CA  0.030 0.204 -.056        VAL   59                       C   0.186 0.291 0.111
 O  -.031 0.260 0.071         C   0.051 0.203 -.056         N   0.115 0.248 0.011        O   0.192 0.299 0.120
 CB -.051 0.226 0.079         O   0.058 0.189 -.052         CA  0.112 0.259 0.023        CB  0.210 0.277 0.096
 CG -.061 0.229 0.066         CB  0.022 0.194 -.068         C   0.114 0.248 0.036        CG  0.207 0.271 0.081
 CD1 -.070 0.217 0.059        CG  0.027 0.204 -.081         O   0.120 0.255 0.046        OD1 0.211 0.282 0.072
 NE1 -.078 0.226 0.048        CD  0.019 0.223 -.082         CB  0.094 0.270 0.023        ND2 0.202 0.255 0.079
 CE2 -.074 0.243 0.049        OE1 0.002 0.225 -.080         CG1 0.094 0.286 0.014
 CZ2 -.079 0.257 0.040        OE2 0.030 0.236 -.086         CG2 0.092 0.276 0.038        ALA   66
 CH2 -.072 0.275 0.043                                                                   N   0.172 0.297 0.104
 CZ3 -.062 0.278 0.054       ALA   54                      ALA   60                       CA  0.162 0.313 0.109
 CE3 -.057 0.264 0.063        N   0.059 0.217 -.061         N   0.108 0.232 0.035        C   0.158 0.313 0.124
 CD2 -.063 0.246 0.060        CA  0.079 0.217 -.063         CA  0.108 0.220 0.047        O   0.160 0.326 0.131
                              C   0.088 0.226 -.051         C   0.126 0.222 0.054        CB  0.144 0.315 0.101
VAL   49                      O   0.104 0.225 -.049         O   0.127 0.222 0.067
 N  -.013 0.241 0.060         CB  0.084 0.227 -.076         CB  0.105 0.200 0.043        GLY   67
 CA -.006 0.256 0.052                                                                    N   0.153 0.297 0.128
 C  -.009 0.252 0.037        ASP   55                      THR   61                       CA  0.149 0.294 0.143
 O  -.008 0.237 0.032         N   0.076 0.235 -.043         N   0.140 0.223 0.046        C   0.137 0.278 0.145
 CB  0.015 0.257 0.053        CA  0.082 0.246 -.032         CA  0.158 0.223 0.050        O   0.140 0.263 0.141
 CG1 0.020 0.276 0.058        C   0.079 0.235 -.019         C   0.165 0.242 0.056
 CG2 0.024 0.253 0.039        O   0.077 0.242 -.008         O   0.173 0.242 0.066        LEU   68
                              CB  0.071 0.263 -.031         CB  0.170 0.218 0.038         N   0.122 0.283 0.152
LEU   50                      CG  0.080 0.279 -.038         OG1 0.162 0.204 0.031        CA  0.107 0.271 0.155
 N  -.013 0.267 0.030         OD1 0.076 0.281 -.050         CG2 0.188 0.211 0.043        C   0.100 0.278 0.169
 CA -.019 0.266 0.016         OD2 0.090 0.288 -.031                                      O   0.110 0.286 0.176
 C  -.003 0.270 0.006                                      ASP   62                      CB  0.113 0.251 0.155
 O   0.006 0.284 0.007       LYS   56                       N   0.162 0.256 0.048        CG  0.102 0.239 0.145
 CB -.034 0.279 0.013         N   0.080 0.217 -.020         CA  0.168 0.274 0.051        CD1 0.082 0.241 0.147
 CG -.047 0.271 0.003         CA  0.076 0.205 -.009         C   0.161 0.279 0.066        CD2 0.108 0.243 0.130
 CD1 -.056 0.254 0.008        C   0.092 0.201 -.000         O   0.171 0.287 0.074
```

```
ALA   69
  N    0.082 0.275 0.170
  CA   0.073 0.287 0.180
  C    0.073 0.307 0.176
  O    0.061 0.313 0.169
  CB   0.081 0.285 0.194

GLN   70
  N    0.087 0.316 0.179
  CA   0.092 0.335 0.178
  C    0.089 0.344 0.164
  O    0.089 0.361 0.162
  CB   0.112 0.338 0.181
  CG   0.122 0.323 0.173
  CD   0.141 0.329 0.170
  OE1  0.153 0.332 0.179
  NE2  0.144 0.332 0.157

ASP   71
  N    0.088 0.332 0.154
  CA   0.083 0.334 0.140
  C    0.086 0.313 0.136
  O    0.093 0.303 0.144
  CB   0.105 0.342 0.134
  CG   0.103 0.362 0.130
  OD1  0.087 0.36  0.130
  OD2  0.117 0.370 0.126

TYR   72
  N    0.075 0.309 0.127
  CA   0.072 0.289 0.126
  C    0.068 0.286 0.111
  O    0.052 0.286 0.107
  CB   0.059 0.283 0.136
  CG   0.045 0.298 0.139
  CD1  0.034 0.297 0.151
  CE1  0.023 0.312 0.153
  CZ   0.023 0.327 0.145
  CE2  0.034 0.328 0.134
  CD2  0.045 0.313 0.131
  OH   0.011 0.341 0.148

VAL   73
  N    0.083 0.286 0.104
  CA   0.082 0.287 0.089
  C    0.095 0.301 0.083
  O    0.104 0.298 0.073
  CB   0.063 0.290 0.084
  CG1  0.057 0.273 0.076
  CG2  0.062 0.306 0.075

LYS   74
  N    0.095 0.318 0.088
  CA   0.107 0.330 0.081


C    0.104 0.351 0.031
O    0.090 0.357 0.076
CB   0.108 0.324 0.066
CG   0.126 0.330 0.060
CD   0.134 0.346 0.067
CE   0.153 0.348 0.062
NZ   0.154 0.350 0.047

ALA   75
  N    0.119 0.360 0.033
  CA   0.121 0.379 0.081
  C    0.106 0.387 0.072
  O    0.098 0.378 0.063
  CB   0.139 0.383 0.074

GLY   76
  N    0.103 0.405 0.074
  CA   0.092 0.416 0.065
  C    0.103 0.419 0.052
  O    0.116 0.430 0.053

ASP   77
  N    0.099 0.410 0.042
  CA   0.109 0.409 0.029
  C    0.097 0.399 0.020
  O    0.098 0.383 0.018
  CB   0.125 0.397 0.033
  CG   0.134 0.388 0.020
  OD1  0.124 0.386 0.010
  OD2  0.151 0.385 0.021

THR   78
  N    0.083 0.409 0.016
  CA   0.068 0.399 0.010
  C    0.072 0.379 0.014
  O    0.065 0.371 0.023
  CB   0.070 0.400 -.006
  OG1  0.056 0.389 -.011
  CG2  0.088 0.391 -.009

ARG   79
  N    0.083 0.371 0.004
  CA   0.035 0.351 0.004
  C    0.066 0.346 -.000
  O    0.057 0.358 -.007
  CB   0.089 0.344 0.018
  CG   0.099 0.326 0.018
  CD   0.115 0.325 0.008
  NE   0.120 0.307 0.005
  CZ   0.129 0.302 -.006
  NH1  0.133 0.285 -.008
  NH2  0.134 0.315 -.014

VAL   80


N    0.059 0.331 0.004
CA   0.041 0.326 -.001
C    0.026 0.335 0.007
O    0.029 0.349 0.015
CB   0.039 0.306 -.002
CG1  0.042 0.300 -.016
CG2  0.053 0.297 0.007

ILE   81
  N    0.010 0.328 0.006
  CA  -.006 0.336 0.011
  C   -.010 0.334 0.027
  O    0.002 0.330 0.035
  CB  -.023 0.331 0.004
  CG1 -.020 0.315 -.006
  CD1 -.037 0.307 -.011
  CG2 -.032 0.347 -.003

ALA   82
  N   -.027 0.338 0.029
  CA  -.034 0.333 0.043
  C   -.031 0.320 0.050
  O   -.034 0.305 0.045
  CB  -.054 0.342 0.043

HIS   83
  N   -.026 0.322 0.063
  CA  -.022 0.307 0.072
  C   -.029 0.309 0.086
  O   -.030 0.324 0.091
  CB  -.002 0.304 0.073
  CG   0.006 0.320 0.080
  ND1  0.011 0.320 0.093
  CE1  0.018 0.336 0.096
  NE2  0.018 0.347 0.085
  CD2  0.011 0.337 0.074

THR   84
  N   -.032 0.294 0.092
  CA  -.039 0.293 0.106
  C   -.022 0.291 0.116
  O   -.007 0.290 0.111
  CB  -.050 0.276 0.107
  OG1 -.045 0.264 0.097
  CG2 -.070 0.281 0.105

LYS   85
  N   -.027 0.290 0.129
  CA  -.014 0.289 0.139
  C   -.010 0.269 0.143
  O   -.021 0.257 0.141
  CB  -.021 0.299 0.152
  CG  -.022 0.319 0.149
  CD  -.039 0.324 0.141


CE  -.043 0.345 0.143
NZ  -.062 0.343 0.141

VAL   86
  N    0.004 0.267 0.151
  CA   0.008 0.250 0.157
  C   -.005 0.249 0.169
  O   -.008 0.262 0.176
  CB   0.027 0.247 0.161
  CG1  0.031 0.227 0.164
  CG2  0.039 0.253 0.149

ILE   87
  N   -.014 0.233 0.170
  CA  -.028 0.230 0.180
  C   -.025 0.211 0.186
  O   -.017 0.200 0.180
  CB  -.047 0.233 0.174
  CG1 -.049 0.221 0.161
  CD1 -.070 0.220 0.158
  CG2 -.051 0.253 0.171

GLY   88
  N   -.031 0.210 0.199
  CA  -.029 0.192 0.206
  C   -.047 0.185 0.210
  O   -.060 0.195 0.212

GLY   89
  N   -.047 0.167 0.211
  CA  -.064 0.157 0.214
  C   -.076 0.168 0.223
  O   -.070 0.174 0.234

GLY   90
  N   -.092 0.170 0.219
  CA  -.105 0.180 0.227
  C   -.105 0.200 0.222
  O   -.112 0.212 0.228

GLU   91
  N   -.097 0.202 0.210
  CA  -.095 0.219 0.204
  C   -.102 0.220 0.189
  O   -.108 0.206 0.183
  CB  -.075 0.225 0.204
  CG  -.072 0.237 0.216
  CD  -.053 0.238 0.221
  OE1 -.044 0.223 0.230
  OE2 -.048 0.253 0.226

SER   92
  N   -.101 0.236 0.183
  CA  -.108 0.239 0.170
```

```
C    -.105 0.259 0.166          ASP   98              PRO   104              TYR   110
O    -.103 0.271 0.175            N   -.118 0.379 0.006   N   -.091 0.316 -.114    N   -.031 0.199 0.004
CB   -.128 0.236 0.170            CA  -.132 0.381 -.005   CA  -.090 0.309 -.128    CA  -.030 0.208 0.018
OG   -.134 0.230 0.183            C   -.123 0.378 -.018   C   -.094 0.288 -.129    C   -.018 0.196 0.027
                                  O   -.107 0.382 -.020   O   -.089 0.279 -.139    O   -.019 0.180 0.026
ASP   93                          CB  -.140 0.400 -.004   CB  -.071 0.312 -.133    CB  -.049 0.211 0.023
  N   -.105 0.263 0.153                                   CG  -.060 0.315 -.120    CG  -.062 0.194 0.022
  CA  -.102 0.262 0.149         VAL   99                  CD  -.073 0.319 -.109    CD1 -.075 0.194 0.012
  C   -.110 0.286 0.135            N   -.134 0.372 -.028                           CE1 -.086 0.178 0.010
  O   -.115 0.274 0.128            CA  -.128 0.366 -.041 GLY   105                 CZ  -.083 0.163 0.019
  CB  -.082 0.285 0.150            C   -.125 0.381 -.052   N   -.105 0.281 -.120    CE2 -.070 0.164 0.029
  CG  -.079 0.305 0.146            O   -.133 0.395 -.051   CA  -.110 0.261 -.122    CD2 -.059 0.179 0.030
  OD1 -.092 0.316 0.148            CB  -.136 0.349 -.046   C   -.093 0.249 -.121    OH  -.094 0.148 0.017
  OD2 -.064 0.309 0.142            CG1 -.123 0.333 -.045   O   -.091 0.237 -.129
                                   CG2 -.143 0.352 -.060                          PHE   111
SER   94                                                GLU   106                  N   -.007 0.205 0.034
  N   -.111 0.304 0.133         SER   100                 N   -.082 0.253 -.111    CA  0.007 0.195 0.041
  CA  -.120 0.309 0.120            N   -.114 0.375 -.061   CA  -.066 0.243 -.109    C   0.010 0.201 0.056
  C   -.106 0.319 0.111            CA  -.106 0.385 -.072   C   -.069 0.227 -.099    O   0.007 0.216 0.060
  O   -.091 0.323 0.114            C   -.087 0.376 -.072   O   -.084 0.220 -.093    CB  0.024 0.194 0.033
  CB  -.135 0.322 0.123            O   -.081 0.369 -.082   CB  -.051 0.255 -.103    CG  0.035 0.212 0.033
  OG  -.144 0.317 0.135            CB  -.101 0.404 -.066   CG  -.036 0.257 -.114    CD1 0.039 0.220 0.020
                                   OG  -.114 0.417 -.070   CD  -.028 0.238 -.118    CE1 0.049 0.235 0.020
VAL   95                                                  OE1 -.020 0.238 -.129    CZ  0.057 0.242 0.031
  N   -.114 0.323 0.099         LYS   101                 OE2 -.031 0.225 -.110    CE2 0.053 0.234 0.044
  CA  -.105 0.334 0.089            N   -.081 0.376 -.060                           CD2 0.043 0.219 0.044
  C   -.117 0.340 0.077            CA  -.065 0.367 -.055 ALA   107
  O   -.126 0.330 0.071            C   -.069 0.346 -.054   N   -.055 0.224 -.091  CYS   112
  CB  -.087 0.328 0.084            O   -.068 0.338 -.043   CA  -.055 0.210 -.081    N   0.017 0.188 0.064
  CG1 -.088 0.313 0.074            CB  -.061 0.375 -.040   C   -.042 0.215 -.070    CA  0.021 0.191 0.078
  CG2 -.077 0.345 0.078            CG  -.064 0.395 -.041   O   -.026 0.212 -.071    C   0.041 0.194 0.079
                                   CD  -.068 0.402 -.027   CB  -.050 0.191 -.087    O   0.052 0.182 0.077
THR   96                           CE  -.071 0.423 -.027                           CB  0.015 0.175 0.087
  N   -.114 0.358 0.075            NZ  -.082 0.428 -.039 TYR   108                  SG  0.017 0.180 0.105
  CA  -.123 0.366 0.063                                   N   -.050 0.223 -.060
  C   -.108 0.366 0.052         LEU   102                 CA  -.040 0.228 -.047  SER   113
  O   -.092 0.365 0.054            N   -.073 0.339 -.066   C   -.042 0.213 -.037    N   0.045 0.211 0.082
  CB  -.129 0.385 0.065            CA  -.078 0.320 -.067   O   -.055 0.204 -.036    CA  0.064 0.217 0.083
  OG1 -.146 0.387 0.059            C   -.087 0.316 -.080   CB  -.050 0.245 -.042    C   0.069 0.217 0.098
  CG2 -.117 0.398 0.057            O   -.084 0.301 -.086   CG  -.043 0.262 -.050    O   0.070 0.231 0.105
                                   CB  -.092 0.317 -.056   CD1 -.027 0.260 -.057    CB  0.067 0.236 0.078
PHE   97                           CG  -.085 0.306 -.044   CE1 -.020 0.275 -.064    OG  0.050 0.243 0.074
  N   -.116 0.367 0.040            CD1 -.091 0.286 -.045   CZ  -.028 0.292 -.063
  CA  -.107 0.368 0.027            CD2 -.092 0.314 -.030   CE2 -.044 0.294 -.055  PHE   114
  C   -.121 0.368 0.016                                   CD2 -.051 0.279 -.049    N   0.073 0.200 0.103
  O   -.135 0.358 0.018         THR   103                 OH  -.021 0.307 -.069    CA  0.081 0.197 0.116
  CB  -.091 0.355 0.025            N   -.098 0.328 -.084                           C   0.096 0.184 0.113
  CG  -.098 0.338 0.017            CA  -.110 0.328 -.096 ALA   109                 O   0.095 0.173 0.104
  CD1 -.105 0.323 0.025            C   -.107 0.318 -.109   N   -.027 0.210 -.030    CB  0.068 0.190 0.126
  CE1 -.110 0.307 0.018            O   -.121 0.313 -.115   CA  -.027 0.197 -.019    CG  0.072 0.183 0.140
  CZ  -.108 0.306 0.004            CB  -.119 0.347 -.098   C   -.025 0.208 -.006    CD1 0.079 0.166 0.142
  CE2 -.100 0.320 -.003            OG1 -.118 0.357 -.086   O   -.020 0.224 -.007    CE1 0.083 0.160 0.155
  CD2 -.095 0.336 0.004            CG2 -.138 0.344 -.102   CB  -.013 0.182 -.020    CZ  0.081 0.172 0.166
                                                                                  CE2 0.075 0.139 0.164
```

```
CD2 0.070 0.195 0.151          CA   0.029 0.079 0.061       LYS   126
                               C    0.012 0.090 0.064          N    -.114 0.213 -.045
PRO  115                       O   -.003 0.083 0.063          CA   -.117 0.220 -.059
   N   0.109 0.185 0.122       CB   0.031 0.062 0.070          C    -.134 0.231 -.058
   CA  0.124 0.172 0.121       CG   0.025 0.064 0.034          O    -.144 0.230 -.048
   C   0.118 0.153 0.118       SD   0.031 0.044 0.094          CB   -.122 0.203 -.067
   O   0.112 0.143 0.127       CE   0.036 0.027 0.082          CG   -.139 0.195 -.061
   CB  0.133 0.173 0.135                                       CD   -.147 0.180 -.069
   CG  0.122 0.186 0.144                                       CE   -.136 0.162 -.067
   CD  0.111 0.198 0.133    MET   121                          NZ   -.137 0.157 -.053
                               N    0.014 0.108 0.065
GLY  116                       CA  -.002 0.119 0.066        LEU   127
   N   0.122 0.147 0.106       C   -.004 0.129 0.053          N    -.136 0.241 -.069
   CA  0.119 0.129 0.102       O    0.000 0.144 0.051          CA   -.151 0.254 -.071
   C   0.100 0.122 0.099       CB  -.002 0.130 0.079          C    -.169 0.244 -.074
   O   0.096 0.106 0.099       CG   0.003 0.118 0.091          O    -.183 0.252 -.070
                               SD   0.002 0.128 0.107          CB   -.146 0.266 -.083
HIS  117                       CE  -.020 0.136 0.109          CG   -.147 0.286 -.079
   N   0.087 0.135 0.097                                       CD1  -.166 0.291 -.076
   CA  0.069 0.129 0.094    LYS   122                          CD2  -.136 0.289 -.065
   C   0.063 0.128 0.080       N   -.009 0.113 0.043
   O   0.052 0.117 0.076       CA  -.008 0.122 0.029        SER   128
   CB  0.055 0.140 0.103       C   -.027 0.122 0.022          N    -.169 0.229 -.081
   CG  0.055 0.133 0.117       O   -.038 0.110 0.023          CA   -.184 0.219 -.084
   ND1 0.046 0.143 0.127       CB   0.003 0.107 0.022          C    -.199 0.229 -.091
   CE1 0.049 0.133 0.138       CG  -.002 0.104 0.007          O    -.205 0.224 -.107
   NE2 0.060 0.119 0.136       CD   0.001 0.122 -.000          CB   -.191 0.209 -.072
   CD2 0.064 0.119 0.122       CE  -.002 0.119 -.015          OG   -.207 0.199 -.075
                               NZ  -.000 0.137 -.022
TRP  118                                                    ASN   129
   N   0.069 0.141 0.072    GLY   123                          N    -.216 0.227 -.084
   CA  0.063 0.144 0.058       N   -.030 0.137 0.015          CA   -.238 0.230 -.092
   C   0.065 0.128 0.048       CA  -.047 0.138 0.008          C    -.241 0.250 -.092
   O   0.063 0.130 0.036       C   -.051 0.155 -.000          O    -.256 0.255 -.089
   CB  0.073 0.160 0.052       O   -.047 0.170 0.004          CB   -.250 0.218 -.084
   CG  0.091 0.155 0.046                                       CG   -.245 0.198 -.084
   CD1 0.097 0.161 0.034    THR   124                          OD1  -.239 0.190 -.075
   NE1 0.114 0.153 0.032       N   -.061 0.152 -.011          ND2  -.249 0.191 -.096
   CE2 0.117 0.142 0.043       CA  -.066 0.165 -.021          OE   -.233 0.261 -.096
   CZ2 0.133 0.131 0.045       C   -.083 0.176 -.018
   CH2 0.134 0.121 0.057       O   -.094 0.173 -.000       MTL   130
   CZ3 0.119 0.122 0.066       CB  -.057 0.156 -.035          CU   0.019 0.160 0.122
   CE3 0.104 0.132 0.064       OG1 -.077 0.140 -.033
   CD2 0.103 0.143 0.052       CG2 -.048 0.151 -.039

ALA  119                    LEU   125
   N   0.068 0.112 0.054       N   -.084 0.190 -.026
   CA  0.070 0.097 0.044       CA  -.098 0.204 -.026
   C   0.053 0.086 0.046       C   -.098 0.213 -.039
   O   0.047 0.075 0.037       O   -.034 0.219 -.044
   CB  0.086 0.085 0.048       CB  -.095 0.217 -.014
                               CG  -.100 0.236 -.017
MET  120                       CD1 -.087 0.249 -.008
   N   0.044 0.088 0.057       CD2 -.119 0.240 -.013
```

AZURIN   MOLECULE II

```
ALA    1
  N    0.244 0.245 0.277
  CA   0.230 0.235 0.283
  C    0.212 0.241 0.278
  O    0.208 0.257 0.278
  CB   0.231 0.236 0.299

GLN    2
  N    0.200 0.223 0.275
  CA   0.182 0.232 0.270
  C    0.182 0.251 0.264
  O    0.182 0.253 0.252
  CB   0.169 0.230 0.282
  CG   0.169 0.211 0.237
  CD   0.171 0.210 0.302
  OE1  0.158 0.210 0.310
  NE2  0.188 0.212 0.306

CYS    3
  N    0.183 0.264 0.273
  CA   0.185 0.283 0.271
  C    0.191 0.292 0.284
  O    0.189 0.309 0.286
  CB   0.169 0.294 0.265
  SG   0.149 0.282 0.262

GLU    4
  N    0.198 0.281 0.293
  CA   0.205 0.288 0.305
  C    0.224 0.293 0.304
  O    0.235 0.283 0.299
  CB   0.204 0.274 0.317
  CG   0.194 0.282 0.329
  CD   0.175 0.274 0.329
  OE1  0.174 0.257 0.329
  OE2  0.162 0.285 0.328

ALA    5
  N    0.229 0.309 0.309
  CA   0.248 0.314 0.308
  C    0.252 0.326 0.321
  O    0.241 0.333 0.328
  CB   0.252 0.324 0.295

THR    6
  N    0.270 0.327 0.322
  CA   0.277 0.336 0.334
  C    0.296 0.344 0.331
  O    0.307 0.335 0.325
  CB   0.278 0.323 0.346
  OG1  0.261 0.314 0.347
  CG2  0.281 0.333 0.359

ILE    7
```

```
  N    0.297 0.360 0.335
  CA   0.314 0.371 0.335
  C    0.317 0.377 0.350
  O    0.304 0.382 0.356
  CB   0.310 0.388 0.326
  CG1  0.321 0.404 0.330
  CD1  0.323 0.417 0.318
  CG2  0.290 0.392 0.327

GLU    8
  N    0.333 0.376 0.354
  CA   0.337 0.380 0.369
  C    0.348 0.397 0.370
  O    0.357 0.403 0.361
  CB   0.347 0.363 0.375
  CG   0.334 0.351 0.383
  CD   0.345 0.335 0.388
  OE1  0.336 0.322 0.393
  OE2  0.362 0.336 0.387

SER    9
  N    0.348 0.404 0.382
  CA   0.357 0.421 0.385
  C    0.374 0.419 0.393
  O    0.383 0.405 0.393
  CB   0.345 0.435 0.391
  OG   0.328 0.434 0.384

ASN    10
  N    0.379 0.433 0.401
  CA   0.395 0.433 0.409
  C    0.400 0.452 0.413
  O    0.401 0.465 0.406
  CB   0.410 0.423 0.402
  CG   0.424 0.436 0.395
  OD1  0.420 0.450 0.390
  ND2  0.441 0.430 0.396

ASP    11
  N    0.403 0.454 0.427
  CA   0.407 0.472 0.432
  C    0.415 0.484 0.421
  O    0.403 0.499 0.419
  CB   0.421 0.470 0.444
  CG   0.416 0.453 0.453
  OD1  0.401 0.453 0.459
  OD2  0.428 0.441 0.453

ALA    12
  N    0.429 0.477 0.415
  CA   0.437 0.487 0.404
  C    0.425 0.489 0.392
  O    0.408 0.487 0.394
  CB   0.455 0.478 0.400
```

```
MET    13
  N    0.432 0.494 0.380
  CA   0.421 0.497 0.369
  C    0.424 0.483 0.358
  O    0.425 0.487 0.346
  CB   0.425 0.516 0.363
  CG   0.410 0.530 0.367
  SD   0.417 0.544 0.381
  CE   0.441 0.541 0.382

GLN    14
  N    0.426 0.466 0.362
  CA   0.429 0.451 0.353
  C    0.412 0.439 0.351
  O    0.405 0.432 0.361
  CB   0.444 0.440 0.358
  CG   0.462 0.446 0.352
  CD   0.477 0.446 0.363
  OE1  0.486 0.460 0.365
  NE2  0.480 0.430 0.369

TYR    15
  N    0.406 0.439 0.338
  CA   0.390 0.427 0.336
  C    0.396 0.408 0.334
  O    0.409 0.404 0.326
  CB   0.380 0.434 0.323
  CG   0.360 0.438 0.327
  CD1  0.352 0.431 0.338
  CE1  0.334 0.436 0.341
  CZ   0.325 0.449 0.333
  CE2  0.334 0.456 0.322
  CD2  0.351 0.451 0.319
  OH   0.308 0.454 0.336

ASP    16
  N    0.385 0.395 0.339
  CA   0.388 0.376 0.335
  C    0.378 0.372 0.322
  O    0.362 0.378 0.321
  CB   0.383 0.363 0.346
  CG   0.365 0.355 0.344
  OD1  0.363 0.339 0.348
  OD2  0.353 0.366 0.339

LEU    17
  N    0.387 0.363 0.313
  CA   0.378 0.360 0.300
  C    0.384 0.373 0.289
  O    0.400 0.374 0.285
  CB   0.358 0.361 0.302
  CG   0.349 0.342 0.300
  CD1  0.363 0.327 0.304
```

```
  CD2  0.333 0.340 0.309

LYS    18
  N    0.371 0.383 0.283
  CA   0.376 0.395 0.272
  C    0.364 0.391 0.260
  O    0.362 0.401 0.250
  CB   0.396 0.393 0.263
  CG   0.402 0.410 0.261
  CD   0.397 0.411 0.246
  CE   0.413 0.408 0.237
  NZ   0.414 0.424 0.228

GLU    19
  N    0.357 0.374 0.260
  CA   0.346 0.367 0.249
  C    0.328 0.360 0.256
  O    0.329 0.349 0.265
  CB   0.356 0.352 0.242
  CG   0.350 0.348 0.228
  CD   0.362 0.334 0.222
  OE1  0.377 0.331 0.228
  OE2  0.357 0.326 0.211

MET    20
  N    0.314 0.366 0.250
  CA   0.296 0.360 0.253
  C    0.288 0.354 0.239
  O    0.287 0.365 0.230
  CB   0.284 0.374 0.259
  CG   0.275 0.367 0.272
  SD   0.287 0.370 0.287
  CE   0.274 0.387 0.297

VAL    21
  N    0.287 0.336 0.237
  CA   0.284 0.329 0.224
  C    0.264 0.322 0.224
  O    0.257 0.318 0.235
  CB   0.296 0.311 0.224
  CG1  0.308 0.312 0.211
  CG2  0.309 0.313 0.237

VAL    22
  N    0.256 0.324 0.213
  CA   0.237 0.319 0.212
  C    0.234 0.305 0.201
  O    0.239 0.307 0.190
  CB   0.225 0.335 0.211
  CG1  0.235 0.353 0.215
  CG2  0.208 0.333 0.219

ASP    23
  N    0.226 0.289 0.206
```

```
CA  0.222 0.275 0.196      OE1 0.072 0.338 0.271      CD2 0.264 0.450 0.369      C   0.346 0.503 0.475
C   0.203 0.280 0.191      NE2 0.079 0.365 0.280                                 O   0.341 0.496 0.465
O   0.197 0.295 0.192                                 LYS  34                    CB  0.379 0.509 0.470
CB  0.223 0.257 0.203      PHE  29                    N   0.277 0.392 0.390      CG  0.398 0.508 0.475
CG  0.232 0.242 0.195      N   0.146 0.344 0.259      CA  0.282 0.390 0.404      SD  0.413 0.525 0.469
OD1 0.249 0.242 0.195      CA  0.163 0.345 0.266      C   0.300 0.400 0.406      CE  0.416 0.519 0.451
OD2 0.222 0.231 0.188      C   0.162 0.355 0.279      O   0.311 0.402 0.396
                           O   0.157 0.371 0.279      CB  0.284 0.371 0.409      ALA  40
LYS  24                    CB  0.176 0.355 0.256      CG  0.282 0.353 0.397      N   0.337 0.514 0.483
N   0.193 0.267 0.186      CG  0.196 0.348 0.259      CD  0.285 0.333 0.402      CA  0.313 0.517 0.481
CA  0.175 0.272 0.181      CD1 0.210 0.361 0.259      CE  0.268 0.331 0.409      C   0.313 0.529 0.469
C   0.162 0.257 0.180      CE1 0.227 0.355 0.262      NZ  0.273 0.321 0.421      O   0.324 0.534 0.461
O   0.159 0.247 0.170      CZ  0.230 0.336 0.263                                 CB  0.310 0.526 0.494
CB  0.177 0.281 0.167      CE2 0.216 0.324 0.263      HIS  35
CG  0.195 0.292 0.166      CD2 0.199 0.330 0.260      N   0.303 0.408 0.418      LYS  41
CD  0.195 0.304 0.154                                 CA  0.319 0.419 0.420      N   0.296 0.533 0.469
CE  0.200 0.293 0.141      THR  30                    C   0.330 0.409 0.431      CA  0.288 0.544 0.453
NZ  0.218 0.299 0.137      N   0.169 0.346 0.289      O   0.324 0.405 0.442      C   0.291 0.564 0.458
                           CA  0.171 0.355 0.303      CB  0.312 0.437 0.426      O   0.282 0.574 0.452
SER  25                    C   0.190 0.353 0.308      CG  0.326 0.450 0.429      CB  0.268 0.540 0.456
N   0.154 0.255 0.192      O   0.200 0.342 0.303      ND1 0.343 0.446 0.431      CG  0.263 0.521 0.460
CA  0.139 0.241 0.194      CB  0.158 0.345 0.313      CE1 0.352 0.461 0.433      CD  0.243 0.517 0.456
C   0.127 0.251 0.204      OG1 0.142 0.355 0.313      NE2 0.341 0.475 0.432      CE  0.242 0.522 0.441
O   0.112 0.257 0.201      CG2 0.166 0.347 0.327      CD2 0.324 0.469 0.429      NZ  0.223 0.520 0.436
CB  0.148 0.224 0.201
OG  0.166 0.228 0.204      VAL  31                    VAL  36                    ALA  42
                           N   0.194 0.366 0.317      N   0.347 0.404 0.427      N   0.304 0.569 0.466
CYS  26                    CA  0.213 0.367 0.322      CA  0.357 0.393 0.437      CA  0.308 0.589 0.465
N   0.133 0.252 0.217      C   0.212 0.369 0.338      C   0.373 0.402 0.444      C   0.318 0.590 0.451
CA  0.124 0.262 0.227      O   0.202 0.379 0.343      O   0.388 0.396 0.443      O   0.311 0.587 0.440
C   0.123 0.282 0.223      CB  0.219 0.386 0.317      CB  0.361 0.374 0.432      CB  0.320 0.595 0.476
O   0.136 0.289 0.218      CG1 0.240 0.388 0.318      CG1 0.377 0.374 0.422
CB  0.135 0.260 0.240      CG2 0.212 0.389 0.303      CG2 0.345 0.366 0.425      VAL  43
SG  0.153 0.277 0.241                                                            N   0.335 0.592 0.453
                           HIS  32                    GLY  37                    CA  0.349 0.590 0.442
LYS  27                    N   0.224 0.359 0.344      N   0.369 0.416 0.452      C   0.355 0.570 0.444
N   0.107 0.290 0.223      CA  0.226 0.362 0.359      CA  0.382 0.425 0.460      O   0.360 0.564 0.455
CA  0.105 0.308 0.218      C   0.244 0.370 0.363      C   0.371 0.434 0.472      CB  0.363 0.604 0.445
C   0.111 0.323 0.227      O   0.258 0.361 0.362      O   0.355 0.432 0.472      CG1 0.390 0.601 0.438
O   0.113 0.338 0.224      CB  0.219 0.349 0.369                                 CG2 0.355 0.623 0.442
CB  0.087 0.312 0.212      CG  0.213 0.331 0.363      LYS  38
CG  0.074 0.319 0.223      ND1 0.224 0.317 0.362      N   0.380 0.444 0.481      MET  44
CD  0.062 0.303 0.228      CE1 0.214 0.304 0.356      CA  0.370 0.452 0.492      N   0.355 0.561 0.432
CE  0.058 0.306 0.243      NE2 0.197 0.309 0.353      C   0.369 0.472 0.492      CA  0.358 0.542 0.431
NZ  0.060 0.288 0.250      CD2 0.196 0.327 0.357      O   0.369 0.480 0.503      C   0.341 0.532 0.427
                                                      CB  0.377 0.444 0.505      O   0.341 0.516 0.424
GLN  28                    LEU  33                    CG  0.373 0.424 0.507      CB  0.369 0.533 0.442
N   0.115 0.317 0.240      N   0.243 0.387 0.367      CD  0.391 0.414 0.506      CG  0.382 0.519 0.436
CA  0.119 0.331 0.250      CA  0.258 0.398 0.371      CE  0.387 0.393 0.506      SD  0.399 0.529 0.426
C   0.137 0.328 0.257      C   0.261 0.396 0.386      NZ  0.396 0.385 0.494      CE  0.390 0.526 0.409
O   0.142 0.313 0.261      O   0.248 0.397 0.394
CB  0.104 0.331 0.261      CB  0.253 0.418 0.367      MET  39                    GLY  45
CG  0.101 0.350 0.266      CG  0.270 0.430 0.367      N   0.369 0.480 0.480      N   0.326 0.542 0.428
CD  0.082 0.350 0.273      CD1 0.280 0.428 0.354      CA  0.366 0.500 0.480      CA  0.309 0.536 0.423
```

```
C    0.311 0.529 0.409
O    0.324 0.532 0.402

HIS   46
N    0.298 0.518 0.405
CA   0.301 0.507 0.392
C    0.283 0.503 0.386
O    0.271 0.498 0.392
CB   0.312 0.491 0.397
CG   0.331 0.491 0.394
ND1  0.338 0.502 0.384
CE1  0.356 0.498 0.384
NE2  0.360 0.485 0.393
CD2  0.344 0.480 0.399

ASN   47
N    0.282 0.507 0.373
CA   0.266 0.504 0.365
C    0.270 0.496 0.351
O    0.284 0.487 0.349
CB   0.254 0.521 0.365
CG   0.256 0.531 0.351
OD1  0.270 0.535 0.346
ND2  0.240 0.534 0.346

TRP   48
N    0.258 0.499 0.341
CA   0.262 0.489 0.328
C    0.247 0.492 0.318
O    0.233 0.482 0.319
CB   0.264 0.469 0.331
CG   0.267 0.456 0.320
CD1  0.282 0.449 0.316
NE1  0.279 0.437 0.305
CE2  0.262 0.437 0.303
CZ2  0.252 0.427 0.294
CH2  0.234 0.428 0.293
CZ3  0.225 0.440 0.302
CE3  0.235 0.450 0.311
CD2  0.253 0.448 0.312

VAL   49
N    0.249 0.505 0.309
CA   0.235 0.510 0.299
C    0.239 0.502 0.285
O    0.253 0.504 0.280
CB   0.234 0.531 0.298
CG1  0.235 0.539 0.312
CG2  0.250 0.537 0.290

LEU   50
N    0.225 0.495 0.279
CA   0.227 0.487 0.265
C    0.221 0.501 0.254


O    0.205 0.504 0.252
CD   0.216 0.469 0.264
CG   0.224 0.458 0.252
CD1  0.239 0.445 0.253
CD2  0.209 0.447 0.245

THR   51
N    0.234 0.508 0.247
CA   0.230 0.521 0.236
C    0.242 0.518 0.224
O    0.250 0.504 0.222
CB   0.234 0.541 0.242
OG1  0.213 0.547 0.249
CG2  0.238 0.554 0.230

LYS   52
N    0.241 0.531 0.215
CA   0.250 0.531 0.202
C    0.268 0.542 0.201
O    0.269 0.556 0.208
CB   0.238 0.537 0.191
CG   0.223 0.549 0.196
CD   0.207 0.550 0.186
CE   0.200 0.569 0.184
NZ   0.197 0.573 0.170

GLU   53
N    0.280 0.535 0.193
CA   0.297 0.543 0.190
C    0.295 0.563 0.187
O    0.285 0.568 0.177
CB   0.305 0.533 0.178
CG   0.291 0.531 0.167
CD   0.293 0.534 0.153
OE1  0.303 0.519 0.147
OE2  0.300 0.550 0.149

ALA   54
N    0.305 0.574 0.194
CA   0.304 0.594 0.192
C    0.288 0.602 0.200
O    0.286 0.619 0.201
CB   0.303 0.598 0.177

ASP   55
N    0.277 0.590 0.205
CA   0.262 0.595 0.213
C    0.269 0.595 0.228
O    0.264 0.606 0.236
CB   0.247 0.582 0.211
CG   0.231 0.592 0.204
OD1  0.236 0.602 0.194
OD2  0.216 0.590 0.209


LYS   56
N    0.282 0.583 0.229
CA   0.292 0.581 0.242
C    0.296 0.599 0.249
O    0.293 0.601 0.261
CB   0.308 0.569 0.240
CG   0.323 0.577 0.232
CD   0.341 0.567 0.234
CE   0.342 0.551 0.224
NZ   0.361 0.545 0.222

GLU   57
N    0.305 0.611 0.242
CA   0.310 0.629 0.248
C    0.293 0.637 0.254
O    0.291 0.640 0.265
CB   0.317 0.642 0.238
CG   0.333 0.635 0.230
CD   0.350 0.643 0.236
OE1  0.360 0.632 0.243
OE2  0.353 0.659 0.233

GLY   58
N    0.280 0.640 0.245
CA   0.263 0.648 0.249
C    0.256 0.640 0.262
O    0.258 0.648 0.273

VAL   59
N    0.249 0.623 0.261
CA   0.243 0.614 0.273
C    0.255 0.619 0.285
O    0.249 0.625 0.296
CB   0.242 0.594 0.271
CG1  0.234 0.589 0.257
CG2  0.231 0.585 0.282

ALA   60
N    0.272 0.616 0.283
CA   0.286 0.619 0.293
C    0.285 0.637 0.300
O    0.279 0.639 0.311
CB   0.305 0.617 0.286

THR   61
N    0.291 0.651 0.293
CA   0.290 0.670 0.298
C    0.274 0.673 0.307
O    0.275 0.681 0.318
CB   0.292 0.684 0.287
OG1  0.286 0.676 0.275
CG2  0.312 0.690 0.286

ASP   62


N    0.258 0.668 0.302
CA   0.242 0.671 0.310
C    0.242 0.658 0.322
O    0.236 0.663 0.333
CB   0.225 0.672 0.302
CG   0.229 0.681 0.239
OD1  0.230 0.670 0.279
OD2  0.230 0.693 0.288

GLY   63
N    0.248 0.641 0.320
CA   0.249 0.629 0.332
C    0.253 0.640 0.344
O    0.242 0.640 0.354

MET   64
N    0.263 0.650 0.344
CA   0.273 0.662 0.355
C    0.259 0.676 0.353
O    0.255 0.679 0.370
CB   0.291 0.671 0.352
CG   0.295 0.685 0.363
SD   0.317 0.694 0.362
CE   0.314 0.718 0.362

ASN   65
N    0.253 0.685 0.348
CA   0.240 0.699 0.349
C    0.224 0.696 0.359
O    0.218 0.708 0.365
CB   0.235 0.709 0.336
CG   0.247 0.726 0.334
OD1  0.253 0.735 0.343
ND2  0.250 0.729 0.321

ALA   66
N    0.219 0.679 0.360
CA   0.204 0.675 0.370
C    0.212 0.668 0.383
O    0.211 0.678 0.393
CB   0.189 0.663 0.364

GLY   67
N    0.220 0.652 0.383
CA   0.231 0.645 0.393
C    0.226 0.638 0.407
O    0.222 0.649 0.416

LEU   68
N    0.229 0.621 0.409
CA   0.224 0.614 0.423
C    0.204 0.618 0.424
O    0.198 0.632 0.419
CB   0.235 0.623 0.434
```

```
CG  0.253 0.614 0.436
CD1 0.250 0.597 0.445
CD2 0.265 0.627 0.444

ALA  69
 N  0.194 0.606 0.431
 CA 0.175 0.609 0.432
 C  0.168 0.614 0.417
 O  0.166 0.630 0.414
 CB 0.171 0.625 0.441

GLN  70
 N  0.166 0.599 0.410
 CA 0.161 0.600 0.396
 C  0.178 0.595 0.388
 O  0.185 0.605 0.379
 CB 0.153 0.618 0.391
 CG 0.155 0.622 0.376
 CD 0.138 0.625 0.369
 OE1 0.131 0.640 0.369
 NE2 0.130 0.611 0.363

ASP  71
 N  0.185 0.580 0.393
 CA 0.202 0.573 0.388
 C  0.211 0.586 0.379
 O  0.209 0.603 0.380
 CB 0.198 0.555 0.381
 CG 0.182 0.547 0.388
 OD1 0.185 0.542 0.400
 OD2 0.167 0.547 0.382

TYR  72
 N  0.221 0.579 0.369
 CA 0.233 0.591 0.362
 C  0.230 0.592 0.346
 O  0.243 0.592 0.338
 CB 0.252 0.535 0.365
 CG 0.259 0.599 0.376
 CD1 0.263 0.616 0.371
 CE1 0.270 0.630 0.379
 CZ 0.274 0.625 0.393
 CE2 0.270 0.608 0.398
 CD2 0.262 0.595 0.389
 OH 0.231 0.638 0.401

VAL  73
 N  0.213 0.595 0.343
 CA 0.208 0.596 0.329
 C  0.200 0.615 0.326
 O  0.207 0.626 0.319
 CB 0.195 0.581 0.326
 CG1 0.204 0.567 0.317
 CG2 0.178 0.589 0.319


LYS  74
 N  0.184 0.617 0.332
 CA 0.173 0.633 0.330
 C  0.154 0.626 0.331
 O  0.151 0.610 0.336
 CB 0.176 0.639 0.315
 CG 0.171 0.658 0.312
 CD 0.177 0.671 0.323
 CE 0.179 0.691 0.317
 NZ 0.186 0.703 0.328

ALA  75
 N  0.141 0.637 0.328
 CA 0.122 0.629 0.328
 C  0.122 0.618 0.315
 O  0.136 0.616 0.308
 CB 0.108 0.644 0.328

GLY  76
 N  0.106 0.611 0.311
 CA 0.104 0.599 0.300
 C  0.114 0.606 0.287
 O  0.106 0.617 0.280

ASP  77
 N  0.129 0.598 0.284
 CA 0.141 0.602 0.273
 C  0.137 0.590 0.261
 O  0.138 0.573 0.262
 CB 0.160 0.598 0.278
 CG 0.174 0.610 0.271
 OD1 0.170 0.615 0.259
 OD2 0.187 0.613 0.277

THR  78
 N  0.133 0.598 0.249
 CA 0.130 0.587 0.237
 C  0.148 0.580 0.232
 O  0.150 0.574 0.220
 CB 0.120 0.596 0.226
 OG1 0.128 0.613 0.222
 CG2 0.101 0.601 0.231

ARG  79
 N  0.162 0.582 0.241
 CA 0.180 0.577 0.238
 C  0.187 0.560 0.245
 O  0.202 0.555 0.243
 CB 0.192 0.593 0.241
 CG 0.188 0.609 0.231
 CD 0.206 0.616 0.226
 NE 0.214 0.627 0.237
 CZ 0.214 0.645 0.236


NH1 0.221 0.655 0.246
NH2 0.203 0.653 0.225

VAL  80
 N  0.174 0.552 0.253
 CA 0.173 0.535 0.260
 C  0.160 0.525 0.259
 O  0.147 0.531 0.264
 CB 0.183 0.539 0.274
 CG1 0.188 0.559 0.276
 CG2 0.167 0.534 0.283

ILE  81
 N  0.160 0.509 0.252
 CA 0.143 0.500 0.252
 C  0.138 0.495 0.266
 O  0.123 0.495 0.270
 CB 0.144 0.483 0.243
 CG1 0.153 0.467 0.250
 CD1 0.173 0.467 0.247
 CG2 0.153 0.487 0.229

ALA  82
 N  0.153 0.492 0.274
 CA 0.152 0.489 0.288
 C  0.169 0.495 0.295
 O  0.184 0.493 0.289
 CB 0.149 0.468 0.291

HIS  83
 N  0.167 0.502 0.307
 CA 0.183 0.508 0.315
 C  0.183 0.498 0.329
 O  0.172 0.503 0.337
 CB 0.183 0.523 0.317
 CG 0.166 0.534 0.323
 ND1 0.165 0.543 0.335
 CE1 0.147 0.547 0.337
 NE2 0.137 0.540 0.327
 CD2 0.149 0.533 0.318

THR  84
 N  0.195 0.486 0.330
 CA 0.196 0.476 0.343
 C  0.195 0.491 0.354
 O  0.193 0.507 0.352
 CB 0.214 0.466 0.345
 OG1 0.210 0.450 0.352
 CG2 0.227 0.478 0.353

LYS  85
 N  0.199 0.484 0.366
 CA 0.200 0.497 0.373
 C  0.220 0.498 0.382


 O  0.231 0.487 0.378
 CB 0.189 0.491 0.390
 CG 0.171 0.502 0.390
 CD 0.163 0.503 0.404
 CE 0.172 0.489 0.413
 NZ 0.160 0.472 0.413

VAL  86
 N  0.222 0.510 0.391
 CA 0.239 0.513 0.399
 C  0.237 0.502 0.412
 O  0.222 0.497 0.415
 CB 0.238 0.534 0.402
 CG1 0.221 0.541 0.395
 CG2 0.254 0.543 0.394

ILE  87
 N  0.252 0.498 0.418
 CA 0.251 0.487 0.430
 C  0.269 0.489 0.438
 O  0.282 0.494 0.432
 CB 0.248 0.467 0.426
 CG1 0.252 0.464 0.411
 CD1 0.260 0.445 0.409
 CG2 0.230 0.460 0.430

GLY  88
 N  0.269 0.484 0.450
 CA 0.285 0.484 0.458
 C  0.293 0.465 0.458
 O  0.283 0.452 0.456

GLY  89
 N  0.310 0.464 0.460
 CA 0.320 0.446 0.459
 C  0.312 0.433 0.470
 O  0.313 0.437 0.482

GLY  90
 N  0.306 0.418 0.465
 CA 0.293 0.404 0.474
 C  0.278 0.402 0.470
 O  0.270 0.390 0.476

GLU  91
 N  0.272 0.413 0.461
 CA 0.254 0.413 0.456
 C  0.253 0.404 0.442
 O  0.265 0.394 0.438
 CB 0.246 0.432 0.456
 CG 0.237 0.436 0.470
 CD 0.229 0.455 0.470
 OE1 0.227 0.461 0.482
 OE2 0.226 0.463 0.459
```

```
SER   92
   N    0.239 0.409 0.435
   CA   0.236 0.402 0.421
   C    0.218 0.404 0.415
   O    0.205 0.406 0.423
   CB   0.244 0.383 0.419
   OG   0.231 0.370 0.422

ASP   93
   N    0.216 0.406 0.402
   CA   0.199 0.410 0.396
   C    0.197 0.402 0.382
   O    0.210 0.395 0.376
   CB   0.198 0.431 0.394
   CG   0.178 0.438 0.394
   OD1  0.172 0.439 0.382
   OD2  0.171 0.441 0.405

SER   94
   N    0.181 0.403 0.377
   CA   0.177 0.394 0.364
   C    0.163 0.404 0.355
   O    0.148 0.408 0.360
   CB   0.169 0.375 0.367
   OG   0.183 0.364 0.372

VAL   95
   N    0.167 0.406 0.342
   CA   0.154 0.413 0.333
   C    0.151 0.400 0.321
   O    0.163 0.392 0.316
   CB   0.160 0.432 0.327
   CG1  0.175 0.430 0.317
   CG2  0.144 0.442 0.321

THR   96
   N    0.134 0.399 0.317
   CA   0.129 0.387 0.306
   C    0.123 0.399 0.293
   O    0.112 0.412 0.295
   CB   0.114 0.375 0.310
   OG1  0.116 0.371 0.324
   CG2  0.114 0.357 0.302

PHE   97
   N    0.130 0.395 0.282
   CA   0.125 0.405 0.270
   C    0.121 0.392 0.258
   O    0.124 0.376 0.258
   CB   0.139 0.419 0.266
   CG   0.156 0.411 0.260
   CD1  0.171 0.407 0.269
   CE1  0.187 0.401 0.263
```

```
   CZ   0.189 0.398 0.250
   CE2  0.175 0.403 0.241
   CD2  0.153 0.409 0.246

ASP   98
   N    0.113 0.401 0.247
   CA   0.107 0.392 0.235
   C    0.120 0.395 0.223
   O    0.126 0.410 0.221
   CB   0.089 0.400 0.231
   CG   0.075 0.399 0.242
   OD1  0.059 0.401 0.239
   OD2  0.081 0.396 0.254

VAL   99
   N    0.122 0.380 0.216
   CA   0.134 0.380 0.204
   C    0.125 0.385 0.191
   O    0.134 0.384 0.181
   CB   0.146 0.363 0.204
   CG1  0.144 0.354 0.218
   CG2  0.165 0.368 0.202

SER   100
   N    0.108 0.390 0.192
   CA   0.100 0.397 0.179
   C    0.109 0.414 0.175
   O    0.108 0.421 0.164
   CB   0.080 0.401 0.182
   OG   0.070 0.384 0.179

LYS   101
   N    0.118 0.421 0.186
   CA   0.127 0.439 0.185
   C    0.148 0.438 0.187
   O    0.157 0.451 0.189
   CB   0.119 0.452 0.195
   CG   0.120 0.443 0.209
   CD   0.120 0.458 0.220
   CE   0.102 0.458 0.227
   NZ   0.105 0.463 0.242

LEU   102
   N    0.154 0.421 0.186
   CA   0.174 0.419 0.188
   C    0.181 0.412 0.174
   O    0.185 0.397 0.172
   CB   0.179 0.405 0.199
   CG   0.194 0.413 0.207
   CD1  0.192 0.434 0.203
   CD2  0.193 0.406 0.222

THR   103
   N    0.179 0.426 0.165
```

```
   CA   0.163 0.423 0.150
   C    0.202 0.416 0.148
   O    0.215 0.423 0.154
   CB   0.179 0.441 0.143
   OG1  0.162 0.446 0.147
   CG2  0.180 0.438 0.128

PRO   104
   N    0.203 0.403 0.139
   CA   0.219 0.392 0.137
   C    0.235 0.401 0.130
   O    0.235 0.404 0.118
   CB   0.213 0.376 0.129
   CG   0.193 0.373 0.126
   CD   0.189 0.398 0.130

GLY   105
   N    0.249 0.405 0.138
   CA   0.265 0.410 0.132
   C    0.271 0.430 0.136
   O    0.286 0.435 0.132

GLU   106
   N    0.260 0.440 0.143
   CA   0.267 0.458 0.147
   C    0.278 0.453 0.160
   O    0.272 0.442 0.168
   CB   0.252 0.471 0.151
   CG   0.234 0.467 0.144
   CD   0.220 0.481 0.147
   OE1  0.214 0.431 0.159
   OE2  0.215 0.492 0.138

ALA   107
   N    0.293 0.463 0.162
   CA   0.302 0.459 0.174
   C    0.290 0.466 0.185
   O    0.285 0.482 0.186
   CB   0.320 0.469 0.175

TYR   108
   N    0.285 0.453 0.194
   CA   0.272 0.460 0.205
   C    0.282 0.463 0.217
   O    0.295 0.454 0.221
   CB   0.256 0.447 0.206
   CG   0.240 0.453 0.198
   CD1  0.231 0.469 0.202
   CE1  0.217 0.475 0.194
   CZ   0.211 0.466 0.183
   CE2  0.220 0.450 0.179
   CD2  0.234 0.444 0.187
   OH   0.197 0.472 0.175
```

```
ALA   109
   N    0.277 0.478 0.224
   CA   0.287 0.485 0.236
   C    0.277 0.483 0.249
   O    0.262 0.489 0.251
   CB   0.293 0.504 0.233

TYR   110
   N    0.286 0.474 0.258
   CA   0.282 0.473 0.272
   C    0.298 0.481 0.279
   O    0.313 0.475 0.276
   CB   0.276 0.454 0.276
   CG   0.293 0.441 0.276
   CD1  0.298 0.431 0.265
   CE1  0.314 0.421 0.266
   CZ   0.324 0.422 0.278
   CE2  0.319 0.432 0.289
   CD2  0.303 0.442 0.288
   OH   0.340 0.412 0.279

PHE   111
   N    0.296 0.495 0.287
   CA   0.310 0.505 0.294
   C    0.303 0.514 0.307
   O    0.288 0.512 0.310
   CB   0.319 0.520 0.285
   CG   0.304 0.533 0.279
   CD1  0.306 0.540 0.266
   CE1  0.294 0.553 0.262
   CZ   0.280 0.558 0.270
   CE2  0.278 0.551 0.283
   CD2  0.290 0.538 0.288

CYS   112
   N    0.316 0.522 0.314
   CA   0.311 0.528 0.328
   C    0.312 0.549 0.329
   O    0.326 0.558 0.328
   CB   0.325 0.520 0.338
   SG   0.321 0.528 0.355

SER   113
   N    0.296 0.557 0.331
   CA   0.295 0.576 0.333
   C    0.301 0.583 0.346
   O    0.304 0.599 0.348
   CB   0.276 0.583 0.329
   OG   0.263 0.570 0.332

PHE   114
   N    0.302 0.571 0.356
   CA   0.309 0.579 0.369
   C    0.326 0.583 0.366
```

```
O    0.339 0.560 0.362
CB   0.309 0.565 0.381
CG   0.315 0.573 0.394
CD1  0.333 0.580 0.396
CE1  0.338 0.588 0.407
CZ   0.326 0.591 0.418
CE2  0.308 0.584 0.416
CD2  0.303 0.576 0.404

PRO  115
N    0.325 0.606 0.368
CA   0.338 0.619 0.363
C    0.358 0.614 0.361
O    0.368 0.612 0.371
CB   0.336 0.636 0.371
CG   0.317 0.634 0.378
CD   0.311 0.615 0.376

GLY  116
N    0.363 0.613 0.348
CA   0.382 0.610 0.345
C    0.385 0.590 0.342
O    0.397 0.585 0.335

HIS  117
N    0.372 0.580 0.348
CA   0.372 0.560 0.346
C    0.370 0.555 0.331
O    0.378 0.542 0.327
CB   0.356 0.553 0.355
CG   0.363 0.552 0.369
ND1  0.354 0.543 0.379
CE1  0.365 0.545 0.390
NE2  0.380 0.554 0.388
CD2  0.379 0.559 0.374

TRP  118
N    0.359 0.566 0.325
CA   0.355 0.565 0.310
C    0.372 0.566 0.302
O    0.370 0.567 0.289
CB   0.342 0.580 0.307
CG   0.344 0.589 0.294
CD1  0.332 0.590 0.284
NE1  0.338 0.598 0.273
CE2  0.355 0.603 0.275
CZ2  0.367 0.611 0.267
CH2  0.385 0.615 0.272
CZ3  0.390 0.609 0.284
CE3  0.378 0.600 0.293
CD2  0.360 0.597 0.288

ALA  119
N    0.387 0.566 0.308

CA   0.404 0.563 0.300
C    0.414 0.550 0.300
O    0.425 0.546 0.292
CB   0.415 0.583 0.305

MET  120
N    0.409 0.539 0.310
CA   0.418 0.522 0.313
C    0.404 0.508 0.312
O    0.409 0.490 0.312
CB   0.426 0.522 0.327
CG   0.442 0.535 0.328
SD   0.452 0.536 0.344
CE   0.476 0.538 0.341

MET  121
N    0.387 0.511 0.311
CA   0.374 0.497 0.311
C    0.364 0.499 0.297
O    0.349 0.505 0.296
CB   0.361 0.499 0.323
CG   0.367 0.487 0.334
SD   0.367 0.498 0.351
CE   0.388 0.492 0.359

LYS  122
N    0.373 0.492 0.287
CA   0.367 0.493 0.273
C    0.369 0.475 0.266
O    0.383 0.470 0.262
CB   0.379 0.507 0.266
CG   0.397 0.509 0.274
CD   0.412 0.508 0.264
CE   0.415 0.489 0.258
NZ   0.433 0.487 0.253

GLY  123
N    0.353 0.467 0.264
CA   0.354 0.450 0.257
C    0.339 0.446 0.248
O    0.325 0.456 0.248

THR  124
N    0.341 0.432 0.240
CA   0.331 0.425 0.229
C    0.316 0.412 0.233
O    0.320 0.397 0.238
CB   0.345 0.416 0.219
OG1  0.355 0.430 0.213
CG2  0.358 0.405 0.228

LEU  125
N    0.301 0.417 0.228
CA   0.285 0.405 0.228

C    0.282 0.402 0.212
O    0.231 0.415 0.205
CB   0.269 0.414 0.234
CG   0.254 0.399 0.234
CD1  0.236 0.407 0.229
CD2  0.251 0.393 0.249

LYS  126
N    0.282 0.385 0.209
CA   0.281 0.380 0.194
C    0.266 0.366 0.192
O    0.254 0.354 0.200
CB   0.299 0.372 0.189
CG   0.315 0.382 0.195
CD   0.329 0.359 0.200
CE   0.343 0.367 0.189
NZ   0.360 0.361 0.195

LEU  127
N    0.257 0.368 0.180
CA   0.245 0.353 0.177
C    0.257 0.339 0.169
O    0.265 0.343 0.159
CB   0.230 0.359 0.168
CG   0.211 0.358 0.175
CD1  0.202 0.340 0.170
CD2  0.200 0.374 0.172

SER  128
N    0.257 0.323 0.175
CA   0.268 0.308 0.170
C    0.258 0.297 0.160
O    0.243 0.299 0.158
CB   0.275 0.296 0.182
OG   0.291 0.304 0.187

ASN  129
N    0.267 0.282 0.158
CA   0.260 0.269 0.147
C    0.267 0.252 0.151
O    0.253 0.267 0.150
CB   0.264 0.275 0.133
CG   0.275 0.291 0.133
OD1  0.269 0.307 0.132
ND2  0.292 0.289 0.134
OE   0.283 0.250 0.151

MTL  130
CU   0.336 0.525 0.371
```

BIBLIOGRAPHY

(1)   Pearson, R.G., J. Am. Chem. Soc., <u>85</u> (1963) 3533-3539

(2)   Sigel, H. and McCormick, D.B., Acc. Chem. Res., <u>3</u> (1970)
201-208

(3)   Hemmerich, P., in "The Biochemistry of Copper" (Eds J.
Peisach, P. Aisen and W.E. Blumberg). Academic Press,
N.Y., (1966) pp 15-34

(4)   Österberg, R., Co-ord. Chem. Rev., <u>12</u> (1974) 309-347

(5)   Gagné, R.R., J. Am. Chem. Soc., <u>98</u> (1976) 6709-6710

(6)   Gazo, J., Bersuker,I.B., Garaj, J., Kabešová, M., Kahout, J.,
Langfelderová, H., Melník, M., Serátor, M., and Valach, F.,
Co-ord. Chem. Rev. <u>19</u> (1976), 253-297

(7)   Jameson, R.F., in "Metal Ions in Biological Systems", Vol 12,
(Ed H. Sigel) Marcel Dekker, N.Y., (1981), pp 1-30

(8)   Orgel, L.E., "An Introduction to Transition Metal Chemistry",
Methuen, London, (1966) pp 59-64

(9)   Malkin, R., and Malmström, B.G., Adv. Enzymol., <u>33</u> (1970)
177-244

(10)   Beinert, H., Co-ord. Chem. Rev., <u>23</u> (1977) 119-129

(11)   Brill, A.S., "Transition Metals in Biochemistry", Vol 26
(Molecular Biology, Biochemistry and Biophysics Series),
Springer-Verlag Berlin, (1977), pp 40-80

(12)   Rotilio, G., in "Metalloproteins" (Ed U. Weser) Thieme-Verlag
Stuttgart, (1979), pp 1-28

(13)   Fee, J.A., Struct. Bond., <u>23</u> (1975) 1-60

(14)   Moore, G.R., and Williams, R.J.P., Co-ord. Chem. Rev. <u>18</u>
(1976), 125-197

(15)   Blumberg, W.E., in "The Biochemistry of Copper", (Eds J.
Peisach, P. Aisen and W.E. Blumberg), Academic Press N.Y.
(1966) pp 49-65

(16)   Brill, A.S., and Bryce, G.F., J. Chem. Phys., _48_ (1968)
         4398-4404

(17)   McMillin, D.R., Holwerda, R.A., and Gray, H.B., Proc. Natl.
         Acad. Sci., U.S.A., _71_ (1974) 1339-1341

(18)   McMillin, D.R., Rosenberg, R.C., and Gray, H.B., Proc. Natl.
         Acad. Sci., U.S.A., _71_ (1974) 4760-4762

(19)   Solomon, E.I., Hare, J.W., and Gray, H.B., Proc. Natl. Acad.
         Sci., U.S.A., _73_ (1976) 1389-1393

(20)   Solomon, E.I., Clendening, P.J., Gray, H.B., and Grunthaner,
         F.J., J. Am. Chem. Soc., _97_ (1975) 3878-3879

(21)   Finazzi-Agrò, A., Rotilio, G., Avigliano, L., Guerrieri, P.,
         Boffi, V., and Mondovi, B., Biochemistry, _9_ (1970) 2009-2014

(22)   Solomon, E.I., Hare, J.W., Dooley, D.M., Dawson, J.H.,
         Stephens, P.J., and Gray, H.B., J. Am. Chem. Soc., _102_,
         (1980) 168-178

(23)   Penfield, K.W., Gay, R.R., Himmelwright, R.S., Eickman, N.C.,
         Norris, V.A, Freeman, H.C., and Solomon, E.I., J. Am. Chem.
         Soc., _103_ (1981) 4382-4388

(24)   Bertini, I., and Scozzafava, A., in "Metal Ions in Biological
         Systems" Vol 12 (Ed. H. Sigel) Marcel Dekker N.Y. (1981)
         pp 34-35

(25)   Thompson, J.S., Marks, T.J., and Ibers, J.A., Proc. Natl.
         Acad. Sci., U.S.A., _74_ (1977) 3114-3118

(26)   Agnus, Y., Rémy, L., and Weiss, R., J. Chem. Soc. Chem. Comm.
         (1980) 867-869

(27)   Handford, P.M., and Lee, W.K. in "Inorganic Biochemistry",
         Vol 2, Royal Society of Chemistry (1981) pp 108-126

(28)   Frieden, E. in "Metal Ions in Biological Systems" Vol 13
         (Ed H. Sigel) Marcel Dekker, N.Y. (1982) in press

(29)   Richardson, J.S., Thomas, K.A., Rubin, B.H., and Richardson,
         D.C., Proc. Natl. Acad. Sci., U.S.A., _72_ (1975) 1349-1353

(30)   Deinum, J., Lerch, K., and Reinhammar, B. F.E.B.S. Lett.
       69 (1976) 161-164

(31)   Malkin, R., in "Inorganic Biochemistry" Vol 2 (Ed. G.L.
       Eichhorn) Elsevier Amsterdam (1973) Chap 21 689-709

(32)   Ryden, L., and Lundgren, J-O., Nature 261 (1976) 344-346

(33)   McLendon, G., and Martell, A.E., J. Inorg. Nucl. Chem. 39
       (1977) 191-193

(34)   Siiman, O., Young, N.M., and Carey, P.R., J. Am. Chem. Soc. 98
       (1976) 744-748

(35)   Miskowski, V., Tang, S-P.W., Spiro, T.G., Shapiro, E., and
       Moss, T.H., Biochemistry 14 (1975) 1244-1250

(36)   Vanngard, T., in "Biological Applications of Electron Spin
       Resonance" (Eds Swartz, H.M., Bolton, J.R., and Borg, D.C.)
       Wiley, N.Y. (1972) Chapt. 9. 411-447

(37)   Hill, H.A.O., Leer, J.C., Smith, B.E., Storm, C.B., and Ambler,
       R.P., Biochem. Biophys. Res. Comm. 70 (1976) 331-338

(38)   Markley, J.L., Ulrich, E.L., Berg, S.P., and Krogmann, D.W.
       Biochemistry 14 (1975) 4428-4433

(39)   Cass, A.E.G., Hill, H.A.O., Smith, B.E., Proc. Eur. Mol. Biol.
       Int. Workshop. 5th, 1976 (Ed J.V. Bannister) Springer Berlin
       (1977) pp 128-135

(40)   Wurzbach, J.A., Grunthaner, P.J., Dooley, D.M., Gray, H.B.,
       Grunthaner, F.J., Gay, R.R., and Solomon, E.I., J. Am. Chem.
       Soc. 99 (1977) 1257-1258

(41)   Hare, J.W., Solomon, E.I., and Gray, H.B., J. Am. Chem. Soc.
       98 (1976) 3205-3208

(42)   Solomon, E.I., Rawlings, J., McMillin, D.R., Stephens, P.J.,
       and Gray, H.B., J. Am. Chem. Soc. 98 (1976) 8046-8048

(43)   Tennent, D.L., and McMillin, D.R., J. Am. Chem. Soc., 101
       (1979) 2307-2311

(44) Ugurbil, K., Norton, R.S., Allerhand, A., and Bersohn, R.,
Biochemistry, 16 (1977) 886-894

(45) Ferris, N.S., Woodruff, W.H., Rorabacher, D.B., Jones, T.E.,
and Ochrymowycz, L.A., J. Am. Chem. Soc. 100 (1978) 5939-5942

(46) Morpurgo, L., Finazzi-Agrò, A., Rotilio, G., and Mondovi, B.,
Eur. J. Biochem. 64 453-457 (1976)

(47) Sakaguchi, U., and Addison, A.W., J.C.S. Dalton (1979) 600-608

(48) Larsen, N.G., Ph.D. Thesis Massey University (1980)

(49) Peisach, J., and Mims, W.B., Eur. J. Biochem. 84 (1978) 207-214

(50) Tullius, T.D., Frank, P., and Hodgson, K.O., Proc. Natl. Acad.
Sci., U.S.A., 75 (1978) 4069-4073

(51) Sutherland, I.W., and Wilkinson, J.F., J. Gen. Microbiol., 30
(1963) 105-112

(52) Doudoroff, M., in "Bergeys Manual of Determinative Bacteriology"
8th Ed. pp 438-439, Williams and Wilkins, Baltimore (1974)

(53) Pittman, M., ibid., pp 282-283

(54) Kamen, M.D., Dus, K.M., Flatmark, T., and Klerk, H. in "Electron
and Coupled Energy Transfer in Biological Systems" (Eds
T.E. King and M. Klingenberg) Marcel Dekker Inc. N.Y. (1971)
Ch.7

(55) Holwerda, R.A., Wherland, S., and Gray, H.B., Ann. Rev. Biophys.
Bioeng., 5 (1976) 363-396

(56) Colman, P.M., Freeman, H.C., Guss, J.M., Murata, M., Norris, V.A.,
Ramshaw, J.A.M., and Venkatappa, M.P., Nature, 272 (1978) 319-324

(57) Wherland, S., and Gray, H.B., in "Biological Aspects of Inorganic
Chemistry" (Eds Addison, A.W., Cullen, W.R., Dolphin, D., and
James, B.R.) Wiley Interscience, N.Y. (1977) pp 289-368

(58) Wood, P.M., Biochim. Biophys. Acta, 351 (1974) 370-379

(59) Johnson, M.K., Thomson, A.J., Walsh, T.A., Barber, D., and
Greenwood, C., Biochem. J., 189 (1980) 285-294

(60) Wherland, S., and Pecht, I., Biochemistry, 17 (1978) 2585-2591

(61)   Finazzi-Agrò, A., Giovagnoli, C., Avigliano, L., Rotilio, G., and Mondovi, B., Eur. J. Biochem., 34 (1973) 20-24

(62)   Hill, H.A.O., and Smith, B.E., J. Inorg. Biochem., 11 (1979) 79-93

(63)   Hill, H.A.O., Smith, B.E., Storm, C.B., and Ambler, R.P., Biochem. Biophys. Res. Comm., 70 (1976) 783-790

(64)   Ugurbil, K., and Bersohn, R., Biochemistry, 16 (1977) 3016-3023

(65)   Ugurbil, K., Maki, A.H., and Bersohn, R., Biochemistry, 16 (1977) 901-907

(66)   Ugurbil, K., and Bersohn, R., Biochemistry, 16 (1977) 895-900

(67)   Boden, N., Holmes, M.C., and Knowles, P.F., Biochem. Biophys. Res. Comm. 57 (1974) 845-848

(68)   Koenig, S.H., and Brown, R.D., Ann. N.Y. Acad. Sci., 222 (1973) 752-763

(69)   Vallee, B.L., and Williams, R.J.P., Proc. Natl. Acad. Sci., U.S.A. 59 (1968) 498-505

(70)   Hopfield, J., Proc. Natl. Acad. Sci., U.S.A., 71 (1974) 3640-3644

(71)   Salemme, F.R., Ann. Rev. Biochem., 46 (1977) 299-329

(72)   Takano, T., Kallai, O.B., Swanson, R., and Dickerson, R.E., J. Biol. Chem., 248 (1973) 5234-5255

(73)   Dickerson, R.E., Ann. N.Y. Acad. Sci. 227 (1974) 599-612

(74)   Dayhoff, M.O., ed., "Atlas of Protein Sequence and Structure" National Biomedical Research Foundation, Washington, D.C., Vol. 5., Suppl.1., (1978) pp S20-22

(75)   Bergman, C., Gandvik, E-K., Nyman, P.O., and Strid, L., Biochem. Biophys. Res. Comm., 77 (1977) 1052-1059

(76)   Kingston, I.B., Kingston, B.L., and Putnam, F.W., Proc. Natl. Acad. Sci., U.S.A., 76 (1979) 1668-1672

(77)   Kingston, I.B., Kingston, B.L., and Putnam, F.W., J. Biol. Chem., 255 (1980) 2886-2896

(78) Steffens, G.J., and Buse, G., H-S.Z. Physiol. Chem., <u>360</u> (1979) 613-619

(79) Dickerson, R.E., Scientific American, <u>242</u> (1980) 98-110

(80) Dwulet, F.E., and Putnam, F.W., Proc. Natl. Acad. Sci., U.S.A., <u>78</u> (1981) 2805-2809

(81) Ryden, L., and Lundgren, J-O., Biochimie, <u>61</u> (1979) 781-790

(82) Adman, E.T., Stenkamp, R.E., Sieker, L.C., and Jensen, L.H., J. Mol. Biol., <u>123</u> (1978) 35-47

(83) Adman, E.T., and Jensen, L.H., Isr. J. Chem., <u>21</u> (1981) 8-12

(84) Timkovich, R., J. Inorg. Biochem. <u>14</u> (1981) 245-252

(85) Rosen, P., Segal, M., and Pecht, I., Eur. J. Biochem., <u>120</u> (1981) 339-344

(86) Ambler, R.P., Biochem. J., <u>135</u> (1973) 751-758

(87) Sutherland, I.W., Archiv. fur Mikrobiol., <u>54</u> (1966) 350-357

(88) Parr, S.R., Barber, D., and Greenwood, C., Biochem. J., <u>157</u> (1976) 423-430

(89) Ambler, R.P., Biochem. J., <u>89</u> (1963) 341-349

(90) Ambler, R.P., and Wynn., M., Biochem. J., <u>131</u> (1973) 485-498

(91) Cusanovich, M.A., Tedro, S.M., and Kamen, M.D., Arch. Biochem. Biophys., <u>141</u> (1970) 557-570

(92) Ambler, R.P. and Brown, L.H., Biochem. J. <u>104</u> (1967) 784-825

(93) Davis, B.J., Ann. N.Y. Acad. Sci. <u>121</u> (1964) 404

(94) Goldberg, M., and Pecht, I., Biochemistry, A.C.S., <u>15</u> (1976) 4197-4208

(95) White, A., Handler, P. and Smith, E.L. "Principles of Biochemistry" McGraw-Hill, N.Y. (1968) p 90

(96) Brill, A.S., Bryce, G.F., and Maria, H.J., Biochim. Biophys. Acta, <u>154</u> (1968) 342-351

(97) Suzuki, H., and Iwasaki, H., J. Biochem. (Tokyo), <u>52</u> (1962) 193

(98)   Strahs, G., Science, 165 (1969) 60-61

(99)   Baker, E.N. Personal Communication

(100)   McPherson, A., Methods of Biochem. Anal., 23 (1976) 249-345

(101)   McPhie, P., Methods in Enzymology, 22 (Ed. Jakoby, W.B.)
         Academic Press N.Y. (1971) p 25

(102)   Zeppezauer, M., Methods in Enzymology 22, (Ed. Jakoby, W.B.)
         Academic Press, N.Y. (1971) pp 253

(103)   Zeppezauer, M., Eklund, H., and Zeppezauer, E.S., Arch.
         Biochem. Biophys., 126 (1968) 564-573

(104)   Camerman, N., Hofmann, T., Jones, S., and Nyburg, S.C., J. Mol.
         Biol. 44 (1969) 569-570

(105)   Blundell, T.L., and Johnson, L.N., "Protein Crystallography",
         Academic Press, London (1976)

(106)   Freeman, H.C., Co-ord. Chem. 1980 (pub 1981) 21 29-51

(107)   Fried, M., and Chun, P.W., "Methods in Enzymology, 22, (Ed.
         Jakoby, W.B.) Academic Press, N.Y. (1971) pp 242

(108)   Sigler, P.B., Jeffery, B.A., Matthews, B.W., and Blow, D.M.,
         J. Mol. Biol., 15 (1966) 175-192

(109)   Watson, H.C, Wendell, P.L., and Scopes, R.K., J. Mol. Biol.,
         57 (1971) 623-625

(110)   Hanson, A.W., Applebury, M.L., Coleman, J.E., and Wyckoff, H.W.,
         J. Biol. Chem., 245 (1970) 4975-4976

(111)   Timkovich, R., and Dickerson, R.E., J. Mol. Biol., 72 (1972)
         199-203

(112)   Low, B.W., and Richards, F.M., J. Am. Chem. Soc., 74 (1952)
         1660-1666

(113)   Matthews, B.W., J. Mol. Biol., 33 (1968) 491-497

(114)   Bokhoven, C., Schoone, J.C., and Bijvoet, J.M., Acta Cryst.,
         4 (1951) 275-280

(115)   Harker, D., Acta Cryst., 9 (1956) 1-9

(116)   Blow, D.M., Proc.Roy.Soc., London, A 247 (1958) 302-336

(117)   Bodo, G., Dintzis, H.M., Kendrew, J.C., and Wyckoff, H.W.,

         Proc.Roy.Soc., A 253 (1959) 70-102

(118)   Bijvoet, J.M., Peerdeman, A.F., and van Bommel, A.J., Nature,

         168 (1951) 271-272

(119)   Ramachandran, G.N., and Raman, S., Current Sci. (India), 25

         (1956) 348-351

(120)   Raman, S., Acta Cryst., 12 (1959) 957-964

(121)   Crick, F.H.C., and Magdoff, B.S., Acta Cryst., 9 (1956) 901-908

(122)   Dickerson, R.E., Eisenberg, D., Varnum, J., and Kopka, M.L.,

         J. Mol. Biol., 45 (1969) 77-84

(123)   Stout, G.H., and Jensen, L.H., "X-ray Structure Determination",

         Collier Macmillan Ltd., London, (1968) p 182

(124)   Blake, C.C.F. and Phillips, D.C. in "Biological Effects of

         Ionizing Radiation at the Molecular Level",I.A.E.A. Symposium

         Vienna, (1962) p 183

(125)   Traub, W., and Hirshfeld, F.L., Acta Cryst., 13 (1960) 753-760

(126)   Kraut, J., Sieker, L.C., High, D.F., and Freer, S.T., Proc. Natl.

         Acad. Sci., U.S.A., 48 (1962) 1417-1424

(127)   Arnone, A., Bier, C.J., Cotton, F.A., Day, V.W., Hazen Jr. E.E.,

         Richardson, D.C., Richardson, J.S., and, in part, Yonath, A.,

         J. Biol. Chem. 246 (1971) 2302-2316

(128)   Herriott, J.R., Sieker, L.C., Jensen, L.H., and Lovenberg, W.,

         J. Mol. Biol., 50 (1970) 391-406

(129)   Watenpaugh, K.D., Sieker, L.C., Herriott, J.R., and Jensen, L.H.,

         Acta Cryst., B 29 (1973) 943-956

(130)   North, A.C.T., Phillips, D.C., and Mathews, F.S., Acta Cryst.,

         A 24 (1968) 351-359

(131)   Krieger, M., Chambers, J.L., Christoph, G.G., Stroud, R.M.,

         and Trus, B.L., Acta Cryst., A 30 (1974) 740-748

(132)   Wyckoff, H.W., Tsernoglou, D., Hanson, A.W., Knox, J.R., Lee, B.,

         and Richards. F.M., J. Biol. Chem., 245 (1970) 305-328

(133)   Patterson, A.L., Phys. Rev., 46 (1934) 372

(134)   Phillips, D.C., Adv. Str. Res. Diffr. Methods, 2, (Eds.
         Brill, R., and Mason, R.) John Wiley and Sons Inc., N.Y.,
         (1966) p 75

(135)   Kartha, G., and Parthasarathy, R., Acta Cryst., 18 (1965)
         745-749

(136)   Blow, D.M., Proc.Roy.Soc., London, A 247 (1958) 302-336

(137)   Rossmann, M.G., Acta Cryst., 13 (1960) 221-226

(138)   Dodson, E.J., in "Crystallographic Computing Techniques",
         (Eds. F.R. Ahmed, K. Huml, and B. Sedláček), Munksgaard,
         Copenhagen, (1976) pp 259-268

(139)   Matthews, B.W., Acta Cryst., 20 (1966) 230-239

(140)   Singh, A.K. and Ramaseshan, S., Acta Cryst., 21 (1966) 279-280

(141)   Stryer, L., Kendrew, J.C., and Watson, H.C., J. Mol. Biol., 8
         (1964) 96-104

(142)   Petsko, G.A., Acta Cryst., A 32 (1976) 473-476

(143)   Brayer, G.D., Delbaere, L.T.J., and James, M.N.G., J. Mol.
         Biol., 124 (1978) 243-259

(144)   Matthews, B.W., Fenna, R.E., Bolognesi, M.C., Schmid, M.F.,
         and Olson, J.M., J. Mol. Biol., 131 (1979) 259-285

(145)   Scouloudi, H., and Baker, E.N., J. Mol. Biol., 126 (1978)
         637-660

(146)   Richardson, J.S., Thomas, K.A., Rubin, B.H., and Richardson,
         D.C., Proc. Natl. Acad. Sci., U.S.A., 72 (1975) 1349-1353

(147)   Matthews, B.W., Acta Cryst., 20 (1966) 82-86

(148)   Bartsch, R.G., in "The Photosynthetic Bacteria", (Eds.
         Clayton, R.K., and Sistrom, W.R.) Plenum, N.Y., (1978) 249-279

(149)   Dickerson, R.E., Kopka, M.L., Varnum, J.C., and Weinzierl, J.E.,
         Acta Cryst., 23 (1967) 511-522

(150) North, A.C.T., and Phillips, D.C., Prog. Biophys. Mol. Biol., 19 (1969) 1-132

(151) Green, D.W., Ingram, V.M., and Perutz, M.F., Proc. Roy. Soc., A 225 (1954) 287-307

(152) Blow, D.M., and Rossman, M.G., Acta Cryst., 14 (1961) 1195-1202

(153) Blow, D.M., and Crick, F.H.C., Acta Cryst., 12 (1959) 794-802

(154) Dickerson, R.E., Weinzierl, J.E., and Palmer, R.A., Acta Cryst., B 24 (1968) 997-1003

(155) North, A.C.T., Acta Cryst., 18 (1965) 212-216

(156) Blow, D.M., and Rossmann, M.G., Acta Cryst., 15 (1962) 1060

(157) Griffith, E.A.H., Spofford, W.A., and Amma, E.L., Inorg. Chem., 17 (1978) 1913-1917

(158) Dickerson, R.E., Kendrew, J.C., and Strandberg, B.E. Acta Cryst., 14 (1961) 1188-1195

(159) Matthews, B.W., in "Crystallographic Computing", (Eds. Ahmed, F.R., Hall, S.R., and Huber, C.P.) , Munksgaard, Copenhagen, (1969)   pp 146-159

(160) Richards, F.M., J. Mol. Biol., 37 (1968) 225-230

(161) Colman, P.M., Jansonius, J.N., and Matthews, B.W., J. Mol. Biol., 70 (1972) 701-724

(162) Kretsinger, R.H., and Nockolds, C.E., J. Biol. Chem., 248 (1973) 3313-3326

(163) Rao, S.T., and Rossmann, M.G., J. Mol. Biol., 76 (1973) 241-256

(164) Rossmann, M.G., and Blow, D.M., Acta Cryst., 15 (1962) 24-31

(165) Richardson, J.S., Nature, 268 (1977) 495-500

(167) Speakman, J.C., "The Hydrogen Bond and other Intermolecular Forces", The Chemical Society, (1975)

(168) Coulson, C.A., "Valence" Oxford University Press (1961) pp 344-356

(169) Adman, E., Watenpaugh, K.D., and Jensen, L.H., Proc. Natl. Acad. Sci., U.S.A., 72 (1975) 4854-4858

(170)   Dickerson, R.E., and Geis, I. "The Structure and Action of Proteins" Harper and Row, N.Y. (1969)

(171)   Matthews, B.W., in "The Proteins", Vol 3, (3rd ed), (Ed. H. Neurath), Academic Press, N.Y. (1975) pp 403-590

(172)   Birktoft, J.J., and Blow, D.M. (1972) J. Mol. Biol., 68 187-240

(173)   Chothia, C., J. Mol. Biol., 75 (1973) 295-302

(174)   Baker, E.N., J. Mol. Biol., 141 (1980) 441-484

(175)   Chandrasekhar, K., McPherson, A., Adams, M.J., and Rossmann, M.G., J. Mol. Biol., 76 (1973) 503-518

(176)   Schulz, G.E., Elzinga, M., Marx, F., and Schirmer, R.H., Nature, 250 (1974) 120-123

(177)   Adman, E.T., Sieker, L.C., and Jensen, L.H., J. Biol. Chem., 248 (1973) 3987-3996

(178)   Quiocho, F.A., and Lipscomb, W.N., Adv. Protein. Chem., 25 (1971) 1-78

(179)   Holmgren, A., Söderberg, B-0, Eklund, H., and Brändén, C-I., Proc. Natl. Acad. Sci., U.S.A., 72 (1975) 2305-2309

(180)   Venkatachalam, C.M., Biopolymers, 6 (1968) 1425-1436

(181)   Crawford, J.L., Lipscomb, W.N., Schellman, C.G., Proc. Natl. Acad. Sci., U.S.A., 70 (1973) 538-542

(182)   Villa, A.C., Manfredotti, A., Guastini, C., Crystal Structure Comm., 1 (1972) 207-210

(183)   Spofford, W.A., Griffith, E.A.H., Amma, E.L., J. Chem. Soc. D., (1970) 533-534

(184)   Karlin, K.D., Dahlstrom, P.L., Hyde, J.R., and Zubieta, J., J.C.S. Chem. Comm., (1980) 906-908

(185)   Pauling, L., "The Nature of the Chemical Bond", 3rd ed., Cornell University Press, N.Y., (1960) p 224

(202)   Holwerda, R.A., Knaff, D.B., Gray, H.B., Clemmer, J.D.,
         Crowley, R.C.,Smith, J.M., and Mauk, A.G., J.Am. Chem. Soc.,
         102(1980) 1142-1146

(203)   Cookson, D.J., Hayes, M.T., and Wright, P.E., Nature, 283
         (1980) 682-683

(204)   Handford, P.M., Hill, H.A.O., Lee, R.W-K., Henderson, R.A.,
         and Sykes, A.G., J. Inorg. Biochem. 13 (1980) 83-88

(205)   Farver, O., and Pecht, I., Proc. Natl. Acad. Sci., U.S.A., 78
         (1981) 4190-4193

(206)   Segal, M.G., and Sykes, A.G., J. Am. Chem. Soc., 100 (1978)
         4585-4592

(207)   Wherland, S., Holwerda, R.A., Rosenberg, R.C., and Gray, H.B.,
         J. Am. Chem. Soc., 97 (1975) 5260-5262

(208)   Reynolds, W.L., and Lumry, R.W., "Mechanisms of Electron
         Transfer", Ronald, N.Y. (1966).

(209)   Lappin, A.G., Segal, M.G., Weatherburn, D.C., and Sykes, A.G.,
         J. Am. Chem. Soc., 101 (1979) 2297-2301

(210)   Taube, H., Myers, H., and Rich, R., J. Am. Chem. Soc., 75
         (1953) 4118-4119

(211)   Kowalsky, A., J.Biol. Chem., 244 (1969) 6619-6625

(212)   Grimes, C.J., Piszkiewicz, D., and Fleischer, E.B., Proc.
         Natl. Acad. Sci., U.S.A., 71 (1974) 1408-1412

(213)   Rosen, P., and Pecht, I., Biochemistry, 15 (1976) 775-785

(214)   Silvestrini, M.C., Brunori, M., Wilson, M.T., and Darley-Usmar,
         V.M., J. Inorg. Biochem. 14 (1981) 327-338

(215)   Antonini, E., Finazzi-Agrò, A., Arigliano, P.G., Guerrieri,
         P., Rotilio, G., and Mondovi, B., J. Biol. Chem., 245 (1970)
         4847-4849

(216)   Moore, G.R., Pettigrew, G.W., Pitt, R.C., and Williams, R.J.P.,
         Biochim. Biophys. Acta, 590 (1980) 261-271

(186)   Hathaway, B.S., and Billing, D.F., Co-ord. Chem. Rev., 5
         (1970) 143-207

(187)   Freyberg, D.P., Mockler, G.M., and Sinn, E., Inorganic Chem.,
         16 (1977) 1660-1665

(188)   Davis, P.H., White, L.K., Belford, R.L., Inorganic Chem., 14
         (1975) 1753-1757

(189)   Heslop, R.B., and Robinson, P.L., in "Inorganic Chemistry",
         Elsevier (1961) p 115

(190)   Matthews, B.W., Macromolecules, 5 (1972) 818-819

(191)   Pecht, I., Farver, O., and Goldberg, M., Advances in Chemistry,
         162 Bioinorganic Chem. II, (Ed. Raymond, K.N.), Am. Chem. Soc.
         (1977), pp 179-206

(192)   Lappin, A.G., Segal, M.G., Weatherburn, D.C., Henderson, R.A.,
         and Sykes, A.G., J. Am. Chem. Soc., 101 (1979) 2302-2306

(193)   Kauzmann, W., Adv. Prot. Chem., 14 (1959) 1-63

(194)   Farver, O., and Pecht, I., Israel. J. Chem., 21 (1981) 13-17

(195)   Pauling, L., "The Nature of the Chemical Bond", 3rd ed. Cornell
         University Press, N.Y., (1960) p 262

(196)   Cotton, F.A., and Wilkinson, G., "Advanced Inorganic Chemistry",
         2nd Ed., Interscience, U.K., (1966) p 115

(197)   Swanson, R., Trus, B.L., Mandel, N., Mandel, G., Kallai, O.B.,
         Dickerson, R.E., J. Biol. Chem., 252 (1977) 759-775

(198)   Crofts, A.R., and Wood, P.M., in "Current Topics in Bioenergetics,"
         Vol 7, Photosynthesis Part A (eds.D.R. Sanadi and L.P. Vernon),
         Academic Press, N.Y. (1978) pp 175-244

(199)   Doolittle, R.F., Science, 214 (1981) 149-159

(200)   El-Ichiro Ochai "Bio-inorganic Chemistry" Allyn and Bacon Inc.
         Boston (1977) p 236

(201)   Mauk, A.G., Scott, R.A., Gray, H.B., J. Am. Chem. Soc., 102
         (1980) 4360-4363

(217)   Ocone, L.R., and Block, B.P., "Inorganic Syntheses", Vol 8.,
        (Ed. Holtzclaw, H.F.) McGraw-Hill pp 125-132

(218)   Huheey, J.E., "Inorganic Chemistry" Harper and Row, 2nd ed.,
        (1978) p 641

(219)   Addison, A.W., Carpenter, M., Lau, K.K-M., and Wicholas, M.,
        Inorganic Chem., 17 (1978) 1545-1552

(220)   Corfield, P.W.R., and Eller, R.G., J. Chem. Soc. D., (1971)
        105-106

(221)   Lewin, A.H., Michl, R.J., Ganis, P., and Lepore, U., J. Chem.
        Soc. Chem. Comm., (1972) 661-662

(222)   Weininger, M.S., Hunt, G.W., and Amma, E.L., J. Chem. Soc.
        Chem. Comm., (1972) 1140-1141

(223)   Sieker, L.C., Watenpaugh, K.D., and Jensen, L.H., Am. Cryst.
        Assoc. Spring Meeting Abstr., Berkeley, (1974) p 79

(224)   Maria, H.J., Nature, 209 (1966) 1023-1024

(225)   Bartsch, R.G., Ann. Rev. Micro., 22 (1968) 181-200

(226)   (a) Kamen, M.D., and Horio,T., Ann. Rev. Biochem., 39 (1970)
             673-700

        (b) Kamen, M.D., and Horio,T., Ann. Rev. Micro., 24 (1970)
             399

(227)   Meyer, T.E., Ambler, R.P., Bartsch, R.G., and Kamen, M.D., J.
        Biol. Chem., 250 (1975) 8416-8421

(228)   Maltempo, M.M., Moss, T.H., Cusanovich, M.A., Biochim.
        Biophys. Acta, 342 (1974) 290-305

(229)   Rawlings, J., Stephens, P.J., Nafie, L.A., and Kamen, M.D.,
        Biochemistry, 16 (1977) 1725-1729

(230)   Imai, Y., Imai, K., Sato, R., and Horio, T., J. Biochem. (Tokyo),
        65 (1969) 222-237

(231)   Kennel, S.J., Meyer, T.E., Kamen, M.D., and Bartsch, R.G.,
        Proc. Natl. Acad. Sci., U.S.A., 69 (1972) 3432-3435

(232)   Weber, P., and Salemme, F.R., J. Mol. Biol., $\underline{117}$ (1977) 815-820

(233)   Salemme, F.R., Arch. Biochem. Biophys. $\underline{163}$ (1974) 423-425

(234)   Langridge, R., Biochim. Biophys. Acta, $\underline{54}$ (1961) 585

(235)   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., Nguyen huu Xuong, and Salemme, F.R.,Nature
        $\underline{286}$ (1980) 302-304

(236)   Yamanaka, T., Kijimoto, S., Okunuki, K., J. Biochem. (Tokyo),
        $\underline{53}$ (1963) 256-9

(237)   Dodson, E.J., Isaacs, N.W., and Rollet, J.S., Acta Cryst.,
        A 32, 311-315

(238)   Almassy, R.J. and Dickerson, R.E.,Proc. Natl. Acad. Sci.
        U.S.A., $\underline{75}$ (1978) 2674-2678.