INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of “sectioning” the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
Jordan, Steven Ralph

STRUCTURE REFINEMENT OF CYTOCHROME C555

The University of Arizona

University Microfilms International

300 N. Zeeb Road, Ann Arbor, MI 48106

PH.D. 1983
STRUCTURE REFINEMENT OF CYTOCHROME C<sub>555</sub>

by

Steven Ralph Jordan

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF CHEMISTRY

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1983
As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Steven Ralph Jordan entitled Structure Refinement of Cytochrome C555 and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Philosophy.

[Signatures and dates]

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

[Signature and date]
This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: [Signature]
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF ILLUSTRATIONS</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vii</td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Biochemical Considerations</td>
<td>2</td>
</tr>
<tr>
<td>Crystallographic Terms and Definitions</td>
<td>8</td>
</tr>
<tr>
<td>2. X-RAY DIFFRACTION STUDIES OF CYTOCHROME C555</td>
<td>13</td>
</tr>
<tr>
<td>Data Collection</td>
<td>13</td>
</tr>
<tr>
<td>Parent Data</td>
<td>15</td>
</tr>
<tr>
<td>Derivative Data</td>
<td>17</td>
</tr>
<tr>
<td>Data Reduction</td>
<td>18</td>
</tr>
<tr>
<td>Patterson Maps</td>
<td>23</td>
</tr>
<tr>
<td>3. INITIAL PHASING AND MAP CALCULATIONS</td>
<td>26</td>
</tr>
<tr>
<td>Phase Refinement</td>
<td>26</td>
</tr>
<tr>
<td>Scaling</td>
<td>31</td>
</tr>
<tr>
<td>Initial Map Interpretations</td>
<td>33</td>
</tr>
<tr>
<td>4. ROTATION FUNCTION STUDIES</td>
<td>38</td>
</tr>
<tr>
<td>5. MAP IMPROVEMENT BY FOURIER INVERSION</td>
<td>44</td>
</tr>
<tr>
<td>3.0 Angstrom Studies</td>
<td>49</td>
</tr>
<tr>
<td>2.0 Angstrom Studies</td>
<td>54</td>
</tr>
<tr>
<td>6. DESCRIPTION OF MOLECULE</td>
<td>60</td>
</tr>
<tr>
<td>Backbone Folding</td>
<td>60</td>
</tr>
<tr>
<td>Heavy Atom Binding Sites</td>
<td>67</td>
</tr>
<tr>
<td>Heme Environment</td>
<td>67</td>
</tr>
<tr>
<td>Intramolecular Bonding</td>
<td>68</td>
</tr>
<tr>
<td>Intermolecular Interactions</td>
<td>69</td>
</tr>
<tr>
<td>7. COMPARISONS TO OTHER CYTOCHROME STRUCTURES</td>
<td>70</td>
</tr>
<tr>
<td>8. SIGNIFICANCE OF STRUCTURE AND FUTURE WORK</td>
<td>74</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS—Continued

APPENDIX A. COMPUTER PROGRAMS USED IN THIS STUDY ............... 78
REFERENCES ........................................................................... 87
# LIST OF ILLUSTRATIONS

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Heme Group and Nomenclature</td>
<td>3</td>
</tr>
<tr>
<td>2. Electron Transport Chain for <em>Chlorobium thiosulfatophilum</em></td>
<td>7</td>
</tr>
<tr>
<td>3. Patterson Maps for HGI4 and U02N</td>
<td>24</td>
</tr>
<tr>
<td>4. HGI4 Difference Map Showing Iodine Position</td>
<td>28</td>
</tr>
<tr>
<td>5. Figure of Merit as a Function of Resolution</td>
<td>32</td>
</tr>
<tr>
<td>6. Anomalous Dispersion Difference Map Showing Heme Iron Position</td>
<td>35</td>
</tr>
<tr>
<td>7. Procedure for Fourier Inversion</td>
<td>45</td>
</tr>
<tr>
<td>8. Smoothing Function Applied to Electron Density</td>
<td>52</td>
</tr>
<tr>
<td>9. Progress of Fourier Inversion</td>
<td>59</td>
</tr>
<tr>
<td>10. Stereoscopic View of Cytochrome c555 Backbone Conformation</td>
<td>61</td>
</tr>
<tr>
<td>11. Stereoscopic View of Cytochrome c555 Side Chain Conformation</td>
<td>63</td>
</tr>
<tr>
<td>12. Alternative Models for Residues 19 to 26</td>
<td>66</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table                                                                 Page
1. Amino Acid Sequence of Cytochrome c555 .................................. 5
2. Unit Cell Parameters ................................................................... 16
3. Percentage of Reflections Observed ............................................ 20
4. Final Scaling R Values .................................................................. 22
5. Refined Heavy Atom Positions and Parameters ............................... 27
6. Refinement Statistics .................................................................... 30
7. Centroid Distribution for HGI4 Derivative ................................... 47
8. Fourier Inversion Refinement at 3.0 Angstroms ............................. 53
9. Fourier Inversion Refinement at 2.0 Angstroms ............................. 55
10. Alternative Approach to Fourier Inversion
    Refinement at 2.0 Angstroms ..................................................... 58
ABSTRACT

The structure of cytochrome c555 from the green sulfur bacterium Chlorobium thiosulfatophilum was determined by using a single isomorphous derivative, K$_2$HgI$_4$, in combination with its anomalous signal. The initial 2.25 angstrom map was modified by the technique of Fourier inversion. The smoothing function for the electron density map addressed three different features in the map, the solvent density, the protein density and the volume surrounding the heavy atom binding sites known to contain spurious peaks.

This structure determination was undertaken for three reasons. First, Chlorobium thiosulfatophilum is a very primitive sulfur metabolizing bacterium and so its cytochrome c555 structure is important for its evolutionary implications. Second, the oxidation-reduction potential of cytochrome c555 is significantly different from the oxidation-reduction potential of other cytochromes whose structures have been determined. Comparisons with the other structures would provide information concerning the factors that are important in regulating oxidation-reduction potentials. Finally, the three dimensional structure may aid in explaining the pattern of reactivity cytochrome c555 displays with mitochondrial cytochrome c oxidase and reductase, which is reversed when compared to other bacterial c-type cytochromes.

The resulting structure contains three alpha helices. These features are consistent with other c-type cytochrome molecules.
previously determined. Two regions of the map appear to be disordered and are difficult to interpret. Possible causes of this observation are discussed and related to the significance of the structure.
A protein crystallographic structure analysis attempts to answer two fundamentally different types of questions. The first class of questions is concerned with the biochemical properties of the protein under study. The protein structure is determined in order to understand the relationship between the protein's structure and function. The structure can provide valuable information concerning the protein's interactions with other compounds, the forces that stabilize its three dimensional structure, and which amino acid residues are important for the biochemical properties of the protein.

The other set of questions a crystallographer must answer arises from the very nature of the x-ray diffraction experiment. Each crystal structure analysis usually presents a unique set of problems. Much of the time and effort spent on solving a protein structure is spent wrestling with purely crystallographic problems before the biochemical implications can be explored.

Reflecting the dual set of problems in a structural analysis this introduction has been divided into two sections. The first section will be concerned with the biochemical properties and the importance of the cytochrome c\textsubscript{555} structure while the second section will define several crystallographic terms that will be useful in subsequent chapters of this work.
Biochemical Considerations

The term cytochrome includes a wide variety of molecules with several types of prosthetic groups. The c-type cytochromes have been the class most extensively studied and will be the focus of this study. C-type cytochromes have been isolated from a wide variety of organisms ranging from primitive sulfur metabolizing bacteria to man. This class of cytochromes contains a protoheme IX as the prosthetic group covalently linked to the polypeptide chain by thioether bonds between the vinyl groups of the heme and two cysteinyl side chains. The c-type cytochrome class can be further subdivided based upon the nature of the axial ligands of the heme iron. The term mitochondrial c-type cytochromes is used to distinguish cytochromes with a histidine nitrogen and a methionine sulfur in the fifth and sixth coordination positions. This is, perhaps, an unfortunate term since many prokaryotic c-type cytochromes are included in this class. A more descriptive designation would be high potential, low spin c-type cytochromes.

The heme structure and nomenclature are shown in Figure 1. It is interesting to note that three of the four heme ligands are always found near the N terminus of the polypeptide chain with the sequence of CYS-X-Y-CYS-HIS. The methionine is further along the chain which allows the protein to wrap around to the other side of the heme.

The reduced mitochondrial c-type cytochromes have characteristic visible absorption spectra with three main peaks of interest. There is an intense Soret peak that falls in the range of 400 to 450 nm. In addition to the Soret band there are two weaker bands, α and β, with the
Figure 1. Heme Group and Nomenclature. Nomenclature follows the Brookhaven Data Bank convention. Cysteine linkages are in the S configuration. The histidine linkage occurs behind the plane of the page, while the methionine is located above the page.
a band typically in the region from 550 to 605 and the \( \beta \) band around 520 to 545 nm. It is the \( \alpha \) band absorption that gives the identifying numbers of most prokaryotic cytochromes. Upon oxidation, the \( \alpha \) and \( \beta \) bands broaden considerably and shift to slightly longer wavelengths while the Soret band shifts to a slightly shorter wavelength.

The eukaryotic mitochondrial oxidative phosphorylating system has been extensively studied and the pathway for the electron transfer is fairly well defined. For the prokaryotic systems, however, there exists a greater diversity of both substrates and pathways of electron transport, so that in many systems the exact nature of the electron transport pathway has yet to be deduced. What has not been proved by experiment in these systems has been assumed by analogy to the eukaryotic system. An integral part of most all electron transport schemes are the c-type cytochromes which generally serve as isopotential electron carriers. In the process of carrying electrons, the formal charge on the heme iron cycles between +2 and +3.

The subject of this study is the cytochrome isolated from an obligate anaerobic green sulfur bacterium, \textit{Chlorobium thiosulfatophilum}. This cytochrome was first reported by Kamen and Vernon (1954) and isolated by Gibson (1961). The protein was sequenced (Van Beeuman, Ambler, Meyer, Kamen, Olsen and Shaw 1976) and found to have 86 amino acid residues and one heme, corresponding to a total molecular weight of 9398 atomic mass units. The sequence is given in Table 1. It is interesting to note the large number of methionine residues.

\textit{Chlorobium thiosulfatophilum} is a very primitive organism that gains its energy either by photophosphorylation or metabolizing
<table>
<thead>
<tr>
<th>Residue</th>
<th>Amino Acid</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-5</td>
<td>TYR-ASP-ALA-ALA-ALA-GLY-</td>
<td>5</td>
</tr>
<tr>
<td>5-10</td>
<td>LYS-ALA-THR-TYR-ASP-ALA-SER-CYS-ALA</td>
<td>10</td>
</tr>
<tr>
<td>21-25</td>
<td>31-35</td>
<td>31</td>
</tr>
<tr>
<td>41-45</td>
<td>LYS-ALA-ALA-TRP-ALA-PRO-HIS-ILE-ALA-LYS-GLY-MET-ASN-VAL-MET</td>
<td>40</td>
</tr>
<tr>
<td>46-50</td>
<td>VAL-ALA-ASN-SER-ILE-LYS-GLY-TYR-LYS-GLY-THR-LYS-GLY-MET-MET</td>
<td>45</td>
</tr>
<tr>
<td>51-55</td>
<td>VAL-ALA-ASN-SER-ILE-LYS-GLY-TYR-LYS-GLY-THR-LYS-GLY-MET-MET</td>
<td>50</td>
</tr>
<tr>
<td>56-60</td>
<td>VAL-ALA-ASN-SER-ILE-LYS-GLY-TYR-LYS-GLY-THR-LYS-GLY-MET-MET</td>
<td>55</td>
</tr>
<tr>
<td>61-65</td>
<td>VAL-ALA-ASN-SER-ILE-LYS-GLY-TYR-LYS-GLY-THR-LYS-GLY-MET-MET</td>
<td>60</td>
</tr>
<tr>
<td>66-70</td>
<td>VAL-ALA-ASN-SER-ILE-LYS-GLY-TYR-LYS-GLY-THR-LYS-GLY-MET-MET</td>
<td>65</td>
</tr>
<tr>
<td>71-75</td>
<td>VAL-ALA-ASN-SER-ILE-LYS-GLY-TYR-LYS-GLY-THR-LYS-GLY-MET-MET</td>
<td>70</td>
</tr>
<tr>
<td>76-80</td>
<td>VAL-ALA-ASN-SER-ILE-LYS-GLY-TYR-LYS-GLY-THR-LYS-GLY-MET-MET</td>
<td>75</td>
</tr>
<tr>
<td>81-85</td>
<td>ASN-ALA-VAL-ALA-TYR-MET-VAL-GLY-GLN-SER-LYS</td>
<td>80</td>
</tr>
</tbody>
</table>

Table 1. Amino Acid Sequence of Cytochrome c₅₅₅.
thiosulfate or hydrogen sulfide. The electron transport scheme as worked out by Kusai and Yamanaka (1973) for green sulfur bacteria is represented in Figure 2. Bacteriochlorophyll becomes photooxidized and reduces an undetermined molecule, X in Figure 2, perhaps bacteriopheophytin. The electron ultimately reduces NADP⁺ to NADPH (Jones and Whale 1970). It is also possible that ATP is formed in a cyclic photophosphorylating chain (Sykes and Gibbon 1967), as indicated by the dotted line in the figure. Reducing equivalents may also be provided by the oxidation of hydrogen sulfide or thiosulfate which reduce c₅₅₃ and c₅₅₁, respectively. These cytochromes then reduce cytochrome c₅₅₅ and ultimately reduce photooxidized bacteriochlorophyll.

There are several reasons why the cytochrome c₅₅₅ crystal structure was undertaken. Crystal structures of cytochromes from eukaryotes, cyanobacteria, nonphotosynthetic bacteria and purple nonsulfur bacteria have previously been reported. The structure from a primitive green sulfur bacterium should provide additional insights into the evolution of protein structures. Their oxidation-reduction potentials are much lower than those of the other cytochromes whose structures have been determined, so a comparison with the known structures should prove helpful in understanding the factors which regulate oxidation potentials. Finally, the structure may explain the pattern of reactivity with mitochondrial cytochrome c reductase and oxidase observed by Errede and Kamen (1978). Other prokaryotic cytochromes show reactivity with mitochondrial cytochrome c reductase but not with the oxidase. With cytochrome c₅₅₅ the reactivity pattern is reversed.
Figure 2. Electron Transport Chain for *Chlorobium thiosulfatophilum*. 
Crystallographic Terms and Definitions

The goal of this section is to present and define several terms commonly used in the field of crystallography. This section is not intended to be a rigorous explanation of x-ray diffraction. A more thorough description can be found in either Stout and Jensen (1968) or Blundell and Johnson (1976). Both of these references have proved invaluable in the course of this study, both as convenient references and as reliable educational resources.

X-ray diffraction is most easily visualized with an analogy to a microscope. A monochromatic beam of x-rays illuminates the crystalline object and the crystal diffracts the x-rays, acting somewhat like a lens. In a normal optical experiment a second lens is used to collect and focus the diffracted light into a coherent image. The nonexistence of a lens suitable in the x-ray wavelength range forces the crystallographer to resort to computational methods to reconstruct the image.

What can be observed and measured with x-rays is the diffraction pattern from the crystal. Due to the fact that the molecule of interest is arranged in a periodic array defined by the crystal, the continuous diffraction pattern of an individual molecule is sampled with a grid-like function giving rise to discrete reflections. The spacing between reflections is inversely related to the unit cell spacing of the crystal which is why diffraction space is commonly referred to as reciprocal space. The data that is collected in an x-ray diffraction experiment is a series of discrete reflections arranged in a three dimensional grid. In order to distinguish between reflections, each reflection is assigned a three number index, Miller index, depending upon its position in the
three dimensional array. The origin of the lattice is the point coincident with the undiffracted beam.

The three dimensional image can be reconstructed by calculating the Fourier summation given by:

$$\rho(x, y, z) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} F^{*} \exp(-2\pi i (hx + ky + lz))$$

where \(\rho\) is the value of the electron density for a particular value of \(x, y\) and \(z\). \(V\) is the volume of the unit cell; \(h, k\) and \(l\) are the Miller indices and \(F^*\) is a complex vector called the structure factor. The structure factor can be broken down into two components, the amplitude, \(|F^*|\), and the relative phase, \(\alpha\), so the structure factor can be written as:

$$F^* = |F^*| \exp(i\alpha).$$

The amplitude can be experimentally determined and is proportional to the square root of the diffraction intensity. The phase is not directly measurable which creates the basic problem in crystallography.

Introducing a heavy atom scatterer into the crystal (Green, Ingram and Perutz 1954) has been the most widely used and successful method in determining the phase. The method is known as MIR, for multiple isomorphous replacement, because in general at least two and usually more heavy atom derivative data sets must be collected. Blow and Crick (1959) described how to gain the most information and assess the effects of errors in this type of analysis. Others (North 1965 and Matthews 1966) have made use of the anomalous scattering of certain heavy atoms so that only a single isomorphous replacement, SIR, must be
made. To emphasize the importance of the anomalous scattering the abbreviation SIRAS is often used. The data set with no heavy atom added to it is called either the parent or native data, while the heavy atom data is generally referred to as the derivative data.

It is also possible to calculate the structure factors, both amplitudes and phases, for a known structure. The equation for this is given by:

\[ \hat{F}(h,k,l) = \sum_{i}^{N} f_i \exp(2\pi i(hx + ky + lz)) \exp(-B_i \sin^2\theta/\lambda^2) \]

where the sum is done over all \( N \) atoms. \( B_i \) is the temperature factor of the atom and is a rough measure of the thermal disorder for a given atom. The scattering factor for the \( i \)th atom at position \( x, y \) and \( z \) is given by \( f_i \), and \( \sin\theta/\lambda \) is a measure of the resolution for the particular \( hkl \). The resolution of a reflection is determined by its location in reciprocal space and is therefore a function of its Miller index and the unit cell parameters of the crystal. In general, the higher resolution reflections contain information about the fine details of the structure, while the low resolution reflections are responsible for the overall features in the electron density. It is important to note that every atom contributes to each reflection and that every reflection contributes to the value of the electron density for each real space grid point.

This structure factor calculation can be used in two ways. The first is that if a similar structure to the one of interest has been previously determined, it may be possible to calculate the diffraction
pattern for the known structure and compare it to the diffraction pattern for the unknown structure. A discussion of the results of an unsuccessful attempt at using the cytochrome c\textsubscript{551} structure to solve the c\textsubscript{555} structure is given in a later section. The other use for calculated structure factors is in the process of structure refinement. Once a reasonably correct model has been built, then phases and structure factors can be calculated. The calculated phases are then used in combination with the observed amplitudes to calculate a new electron density map. The new phases should provide an improved image.

Another useful synthesis that is commonly employed is the difference Fourier. In a difference Fourier synthesis the structure factor amplitudes are replaced with a difference between amplitudes for two different data sets. For example, an observed minus calculated difference Fourier is very useful in refining models. The amplitudes, \( (F_{\text{obs}} - F_{\text{calc}}) \), are used in combination with the calculated phases. The map produced by this synthesis should give positive peaks where atoms should be placed, and negative peaks where atoms have been incorrectly placed in the trial structure.

A useful quantity to measure the correctness of a structure is the R factor. The R factor is defined as:

\[
R = \frac{\sum |F_{\text{obs}} - F_{\text{calc}}|}{\sum |F_{\text{obs}}|}
\]

where \( F_{\text{obs}} \) and \( F_{\text{calc}} \) are the observed and calculated structure factor amplitudes. Typically the R factor for an unrefined starting model is
around 45% to 50%, while after refinement the R factor should drop to a value near 20%.

Difference Fouriers have a variety of uses. They are frequently used for finding heavy atom binding sites, where the differences between derivative and native amplitudes are used. To display the differences between oxidation states, a synthesis using $F_{ox} - F_{red}$ can be calculated where $F_{ox}$ and $F_{red}$ refer to the structure factor amplitudes from the oxidized and reduced data, respectively.

Additional terms and concepts will be introduced as the need arises in the thesis; however, it is hoped that this general introduction has defined the most commonly used terms and identified the concepts essential to this study.
CHAPTER 2

X-RAY DIFFRACTION STUDIES OF CYTOCHROME C₅₅₅

The cytochrome C₅₅₅ used in this study was isolated from Chlorobium thiosulfatophilum, strain Larsens, that was provided by Dr. Cusanovich. Crystals were grown from 65% saturated ammonium sulphate with 0.1 M Mg(NO₃)₂ as described by Korszun and Salemme (1977). Crystals grew in clusters necessitating that single crystals be carefully sliced from the larger clusters. The size of the crystals was on the order of 0.5 to 1.5 mm on a side and many of the crystals used were rather irregular in shape due to slicing from a larger cluster.

Data Collection

All data used in this structure determination was collected on a Syntex P3/F automated diffractometer, using Ni filtered Cu radiation from a sealed x-ray tube operated at a power of 45 KV and 30 mA. Crystals were mounted in sealed glass capillary tubes with the b axis along the length of the capillary. To prevent the crystals from dehydrating during data collection, a separate reservoir of crystallizing solution, or 'mother liquor', was included at one end of the sealed capillary tube. Reflections were measured using an ω-scan (Wyckoff, Doscher, Tsernoglou, Inagami, Johnson, Hardman, Allewell, Kelly and Richards 1967) with a typical scan width of 0.8 degrees. A variable scan rate was used, ranging from 0.6 to 15.0 degrees per minute depending upon the reflection's intensity. This allowed for more rapid
data collection, especially at low resolution. The crystal to detector distance was 40 cm. In order to reduce the background scatter due to air a helium beam path between the crystal and the detector was employed.

Data were collected in shells of reciprocal space in order to facilitate scaling for crystal decay. Due to symmetry considerations of space group $P2_12_12_1$, the unique data set in reciprocal space to be collected was for all positive Miller indices. To determine the anomalous scattering component, Friedel mates, $-h,-k,-l$, were collected for all parent and derivative crystals. In order to minimize scaling problems due to crystal decay, a row of reflections and their corresponding Friedel mates were collected in approximately the same time frame.

Three different types of empirical measurements must be made in order to properly scale the diffraction intensities from a single crystal. They are a measure of the background scatter, the crystal decay, and the absorption effects due to the nonuniform shape of the crystal. Background corrections were made by measuring on either side of the reflection for one fourth of the total time of data collection for the reflection. This meant that the total background was collected for one half of the time of the diffraction peak. Four standard reflections, widely separated in reciprocal space, were measured every 75 reflections to monitor crystal decay. An empirical absorption correction (North, Philips and Mathews 1968) was determined by monitoring two $0k0$ reflections. The $0k0$ reflections are chosen because, due to the manner in which the crystal was mounted, the crystal can be
rotated through 360 degrees about the b axis of the crystal and the 0k0 reflections remain in a diffracting position.

Parent Data

Two strongly diffracting native c₅₅₅ crystals were used to make up the complete 2.0 angstrom data set. The crystals diffracted well as indicated by the fact that overall 92% of the reflections were observed to 2.3 angstroms and in the range of 2.5 to 2.0 angstroms reflections were greater than 70% observed. The greatest concern during the data collection was that the native crystals, when initially mounted, are in a monoclinic space group with a beta angle of about 88 degrees. Table 2 gives unit cell parameters for the crystals used in data collection. During two days of exposure in the x-ray beam the diffraction pattern would gradually change to an orthorhombic pattern with the beta angle equal to 90 degrees.

Unit cell parameters given in Table 2 are the results of the orientation matrix calculated by the diffractometer during the alignment of the crystals. The parameters are based upon locating and centering from fifteen to twenty-two known reflections of moderate to strong intensity. No constraints were placed on any of the unit cell parameters.

Since it was a gradual and continuous transition from monoclinic to orthorhombic, it was impossible to collect a complete or meaningful data set in the monoclinic unit cell. After about a day and a half of irradiation the crystal stabilized in an orthorhombic unit cell and data collection was started in earnest.
Table 2. Unit Cell Parameters.

Unit cell lengths given in angstroms and angles in degrees.

<table>
<thead>
<tr>
<th>CRYSTAL ID</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>alpha</th>
<th>beta</th>
<th>gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td>PARENT 1</td>
<td>48.07</td>
<td>24.35</td>
<td>60.14</td>
<td>90.04</td>
<td>90.01</td>
<td>90.01</td>
</tr>
<tr>
<td>PARENT 2</td>
<td>48.02</td>
<td>24.37</td>
<td>60.22</td>
<td>90.09</td>
<td>90.07</td>
<td>90.02</td>
</tr>
<tr>
<td>MONOCLINIC</td>
<td>48.14</td>
<td>24.29</td>
<td>60.18</td>
<td>90.02</td>
<td>88.59</td>
<td>90.04</td>
</tr>
<tr>
<td>HGI4 1</td>
<td>48.06</td>
<td>24.39</td>
<td>59.71</td>
<td>90.06</td>
<td>90.03</td>
<td>90.00</td>
</tr>
<tr>
<td>HGI4 2</td>
<td>47.98</td>
<td>24.38</td>
<td>59.86</td>
<td>90.01</td>
<td>90.10</td>
<td>90.02</td>
</tr>
<tr>
<td>U02N 1</td>
<td>48.07</td>
<td>24.29</td>
<td>59.81</td>
<td>90.15</td>
<td>90.27</td>
<td>89.80</td>
</tr>
<tr>
<td>U022 1</td>
<td>48.52</td>
<td>24.53</td>
<td>60.36</td>
<td>90.05</td>
<td>89.92</td>
<td>90.10</td>
</tr>
</tbody>
</table>
Derivative Data

Native c555 crystals were soaked overnight in a crystallizing solution containing 10mM K₂HgI₄ and the resulting diffraction pattern was measured. Fortunately, the derivative crystals showed no signs of being monoclinic so data collection was initiated as soon as the crystal was properly aligned on the diffractometer. Our computing system requires an identifying code word for each derivative and so throughout this text the different derivatives will be referred to by the four letter code word of the derivative. The first HGI4 data set collected was marred by the fact that the crystal slipped in the capillary tube during data collection, so it was important to monitor the standard reflections carefully and the crystal had to be recentered periodically.

Initial analysis of the data indicated that despite the problems arising from the crystal slippage this derivative looked quite promising. In order to get a better data set, another crystal was soaked in the same conditions and the data recollected. By combining the data sets and careful editing of observations a useful derivative was made. The unit cell parameters indicated that the heavy atoms caused no major change in the crystal packing. The largest change occurred in the c axis, decreasing it by 0.4 angstroms. This change borders upon the acceptable range of isomorphism (Crick and Magdoff 1956).

Crystals soaked for two weeks in a 10mM UO₂(NO₃)₂ crystallizing solution displayed the same problem with slippage in the capillary tube, but fortunately the crystal intensities remained strong enough to
collect multiple observations of most reflections so the erroneous observations could be edited out. The analysis of the U02N data set indicated that there were multiple derivative sites. In hopes of reducing the number of substitution sites another crystal was soaked in a similar UO$_2$(NO$_3$)$_2$ solution, however this time for a shorter time, 18 hours compared to two weeks. The substitution pattern for UO22 looked to be identical to that of U02N so the second data set, UO22, was not used.

**Data Reduction**

All data sets, native and derivative, were processed in the same manner. The raw intensities were corrected for absorption, background, polarization and the Lorentz factor (Blundell and Johnson 1976) by a program written on the Syntex P3/F diffractometer system. The overall corrected intensities term is given by:

$$I_{\text{cor}} = \left[ I_{\text{tot}} - 2*(I_{\text{br}} + I_{\text{bl}}) \right] \times (\text{ABS}) \times (\text{SR}) \times (LP)^{-1}$$

where $I_{\text{cor}}$ is the corrected intensity, $I_{\text{tot}}$ is the total unadjusted intensity, $I_{\text{br}}$ and $I_{\text{bl}}$ are the right and left background corrections, ABS is the empirically determined absorption correction and SR is the scan rate in degrees per minute. $(LP)^{-1}$ is the Lorentz polarization correction given by:

$$(LP)^{-1} = \frac{2 \sin 2\theta}{(1 + \cos^2 2\theta)}$$

where $2\theta$ is the angle between the undiffracted x-ray beam and the detector. These corrected intensities were then written onto a magnetic
tape and taken to the University Computing Center where all further manipulations and processing were performed on the Cyber 175.

Standard counting statistics are used to judge the significance of a reflection. This means that the standard deviation, \( \sigma \), of the corrected intensity is given by (Syntex 1978):

\[
\sigma_I = (I_{\text{tot}} + 2^2 \cdot (I_{\text{br}} + I_{\text{b1}}))^{1/2} \cdot (SR)
\]

where terms have the same meaning as given above. The factor of two squared is a consequence of the fact that backgrounds were measured for half the time of the main peak. Reflections were classified as unobserved if the corrected intensity was less than \( \sigma \). Table 3 gives the percentage observed for the native and derivative crystals as a function of distance in reciprocal space.

The final step in data reduction is to correct for crystal decay as determined by monitoring the reference reflections and to edit out the erroneous reflections. Due to the various problems that arose during data collection careful scrutiny of the data was required. By editing out the erroneous observations, in some cases entire shells of data had to be removed, the scaling R values dropped from 40% to 5%.

The scaling R is defined as:

\[
R = \frac{\sum_{h} \sum_{i} |I(h)_{i} - I(h)_{i}|}{\sum_{i} I(h)_{i}}
\]

where \( h \) is the index of the reflection, \( N \) is the number of observations,
Table 3. Percentage of Reflections Observed.

<table>
<thead>
<tr>
<th>RESOLUTION sinθ/λ</th>
<th>ANGSTROMS</th>
<th>C555</th>
<th>HGI4</th>
<th>UO2N</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000 - 0.0868</td>
<td>0.0868</td>
<td>5.76</td>
<td>98.64</td>
<td>98.22</td>
</tr>
<tr>
<td>0.0868 - 0.1093</td>
<td>0.1093</td>
<td>4.57</td>
<td>98.13</td>
<td>98.97</td>
</tr>
<tr>
<td>0.1093 - 0.1251</td>
<td>0.1251</td>
<td>4.00</td>
<td>97.11</td>
<td>97.66</td>
</tr>
<tr>
<td>0.1251 - 0.1377</td>
<td>0.1377</td>
<td>3.63</td>
<td>96.33</td>
<td>94.79</td>
</tr>
<tr>
<td>0.1377 - 0.1484</td>
<td>0.1484</td>
<td>3.37</td>
<td>94.18</td>
<td>93.92</td>
</tr>
<tr>
<td>0.1484 - 0.1577</td>
<td>0.1577</td>
<td>3.17</td>
<td>94.86</td>
<td>94.44</td>
</tr>
<tr>
<td>0.1577 - 0.1660</td>
<td>0.1660</td>
<td>3.01</td>
<td>95.60</td>
<td>92.42</td>
</tr>
<tr>
<td>0.1660 - 0.1735</td>
<td>0.1735</td>
<td>2.88</td>
<td>91.18</td>
<td>91.80</td>
</tr>
<tr>
<td>0.1735 - 0.1805</td>
<td>0.1805</td>
<td>2.77</td>
<td>89.10</td>
<td>90.51</td>
</tr>
<tr>
<td>0.1805 - 0.1869</td>
<td>0.1869</td>
<td>2.67</td>
<td>93.51</td>
<td>91.14</td>
</tr>
<tr>
<td>0.1869 - 0.1930</td>
<td>0.1930</td>
<td>2.59</td>
<td>86.58</td>
<td>88.27</td>
</tr>
<tr>
<td>0.1930 - 0.1987</td>
<td>0.1987</td>
<td>2.52</td>
<td>88.55</td>
<td>88.57</td>
</tr>
<tr>
<td>0.1987 - 0.2040</td>
<td>0.2040</td>
<td>2.45</td>
<td>87.93</td>
<td>86.86</td>
</tr>
<tr>
<td>0.2040 - 0.2091</td>
<td>0.2091</td>
<td>2.39</td>
<td>83.87</td>
<td>78.61</td>
</tr>
<tr>
<td>0.2091 - 0.2140</td>
<td>0.2140</td>
<td>2.35</td>
<td>82.35</td>
<td></td>
</tr>
<tr>
<td>0.2140 - 0.2193</td>
<td>0.2193</td>
<td>2.28</td>
<td>77.16</td>
<td></td>
</tr>
<tr>
<td>0.2193 - 0.2271</td>
<td>0.2271</td>
<td>2.20</td>
<td>68.60</td>
<td></td>
</tr>
<tr>
<td>0.2271 - 0.2345</td>
<td>0.2345</td>
<td>2.13</td>
<td>68.36</td>
<td></td>
</tr>
<tr>
<td>0.2345 - 0.2414</td>
<td>0.2414</td>
<td>2.07</td>
<td>54.53</td>
<td></td>
</tr>
<tr>
<td>0.2414 - 0.2479</td>
<td>0.2479</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OVERALL 88.02 90.04 96.37
$I(h)_i$ is the intensity for an individual observation and $\bar{I}(h)$ is the average value of the reflection. Table 4 gives a list of the scaling R values for the native and derivative crystals.

In order to correct for crystal decay, the reflections were divided into scaling groups based on the sequence of when they were collected. Then a scale factor was calculated and refined for each scale group (Hamilton, Rollett and Sparks 1965). This method was also used to scale data from different crystals. Centric reflections have no anomalous signal, so that differences in Friedel related centric reflections arise solely from errors in the data collection. The Friedel related centric reflections were averaged at this point so that differences would not be attributed to a meaningful signal.

Initially the HGI4 and UO2N derivatives were scaled to the parent data set using the program FINDKC. This program scales two data sets together by calculating a linear and an exponential scale factor to be applied to the derivative data so that the derivative data is scaled to the parent data. Mathematically the scaling is as

$$I = k \cdot I \cdot \text{EXP}[-c \cdot \sin^2 \Theta / \lambda^2]$$

where $k$ and $c$ are the constants to be determined to give the best fit of the derivative to the parent data. The intensity is given by $I$, $\lambda$ is the wavelength and $\Theta$ is the angle between the diffracted beam and the diffracting plane in the crystal.
Table 4. Final Scaling R Values.

<table>
<thead>
<tr>
<th>CRYSTAL</th>
<th>NUMBER OF OBSERVATIONS</th>
<th>NUMBER OF REFLECTIONS</th>
<th>R VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>C555 (NATIVE)</td>
<td>7,644</td>
<td>5,357</td>
<td>0.038</td>
</tr>
<tr>
<td>C555 (HIGH RES)</td>
<td>7,509</td>
<td>5,143</td>
<td>0.050</td>
</tr>
<tr>
<td>HGI4</td>
<td>14,610</td>
<td>6,886</td>
<td>0.074</td>
</tr>
<tr>
<td>U02N</td>
<td>10,123</td>
<td>4,661</td>
<td>0.044</td>
</tr>
</tbody>
</table>
Patterson Maps

The first step in the analysis of the heavy atom derivative data is to determine the number and location of the heavy atom binding sites. This is done by calculating what is known as a difference Patterson map. A difference Patterson is a phaseless synthesis using the square of the difference in scattering intensities between the heavy atom soaked crystal and the native crystal.

The Patterson map is best visualized as a map of all the interatomic vectors in the unit cell. When there are symmetry related molecules in the unit cell, the interatomic vectors between the crystallographic symmetry related atoms occur at special places called Harker peaks or sections. In space group P2₁2₁2₁ there are three Harker sections corresponding to the three unique 2₁ screw axes. The Harker planes for the difference Pattersons for the HGI4 and UO₂N derivatives are shown in Figure 3.

From these Harker sections and difference Fouriers it was determined that there were three binding sites for the UO₂N and one for the HGI4 derivative. These sites were confirmed and correlated to a unique origin by calculating difference Fouriers using both the SIRAS phases from the other derivative and the previously determined phases (Korszun 1979). The relative peak heights should be the same for all Harker sections, however, the intensity for the W section is considerably less for the HGI4 derivative. This may be due to systematic absorption in this direction or a problem related to the shape of the crystal. Another possible explanation is the lack of isomorphism along the c axis.
Figure 3. Patterson Maps for HGI4 and UO2N.
In Figure 3a the Harker sections for the HGI4 derivative are shown with the mercury position indicated.
Figure 3. Patterson Maps for HGI4 and U02N.
In Figure 3b the Harker sections for the U02N derivative are shown indicating three heavy atom binding sites.
CHAPTER 3

INITIAL PHASING AND MAP CALCULATIONS

After the heavy atom derivative difference Patterson has been interpreted and there are internally consistent density peaks on the Harker sections, the next step is to determine the phase for the native structure factors. Phases are refined to convergence by adjusting the positional and occupancy parameters of the heavy atom, and then an electron density map is calculated. Once a reliable starting model has been built, phases can be calculated from the starting model and used for further refinement. The most critical step, however, is obtaining a good initial phase set that produces an interpretable electron density map.

Phase Refinement

Using the method of Matthews (1966) to incorporate the anomalous scattering of the heavy atom derivative, phases were calculated and refined separately for both the HGI4 and U02N derivatives. Table 5 gives the final refined heavy atom positions and parameters for the HGI4 and U02N derivatives. As can be seen from the table the individual iodine positions could be located in the difference Fourier map calculated to 2.5 angstroms using phases derived from a single mercury atom in the heavy atom binding site. The difference map is shown in Figure 4.
Table 5. Refined Heavy Atom Positions and Parameters.

<table>
<thead>
<tr>
<th>DERIVATIVE</th>
<th>SITE</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>Q</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>HGI4</td>
<td>HG</td>
<td>.3006</td>
<td>.0557</td>
<td>.2464</td>
<td>4.056</td>
<td>13.67</td>
</tr>
<tr>
<td></td>
<td>I1</td>
<td>.3195</td>
<td>.1097</td>
<td>.2418</td>
<td>2.237</td>
<td>16.13</td>
</tr>
<tr>
<td></td>
<td>I2</td>
<td>.2992</td>
<td>-.0145</td>
<td>.2153</td>
<td>3.570</td>
<td>15.13</td>
</tr>
<tr>
<td></td>
<td>I3</td>
<td>.3579</td>
<td>-.0005</td>
<td>.2570</td>
<td>2.405</td>
<td>11.75</td>
</tr>
<tr>
<td></td>
<td>I4</td>
<td>.2707</td>
<td>.0566</td>
<td>.2720</td>
<td>2.119</td>
<td>14.28</td>
</tr>
<tr>
<td>UO2N</td>
<td>A</td>
<td>.4832</td>
<td>.1027</td>
<td>.1820</td>
<td>2.648</td>
<td>15.76</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>.0537</td>
<td>.3573</td>
<td>.3881</td>
<td>2.241</td>
<td>17.07</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>.8896</td>
<td>.0640</td>
<td>.4635</td>
<td>1.415</td>
<td>15.11</td>
</tr>
</tbody>
</table>

X, Y and Z are in fractions of a unit cell's edge.

Q is the occupancy, in arbitrary units.

T is the temperature factor.
Figure 4. HGI4 Difference Map Showing the Iodine Positions.
The resulting refinement statistics for the individual derivatives, shown in Table 6, looked reasonable compared to the usual results reported in the literature from this type of analysis. When the two different heavy atom derivative data sets were combined to calculate phases, however, there was no indication that the phases improved. Instead, phases deteriorated due to the apparent domination of the HGI4 derivative.

Many different strategies were employed to combine the phase information from the different derivatives, such as fixing the sites of one derivative and refining the sites of the other derivative; or ignoring the anomalous scattering for one or both derivatives; but ultimately the phasing was dominated by the HGI4 derivative in all cases. Also, combinations with heavy atom derivative data sets taken earlier on the Picker diffractometer (Korszun 1979) did not improve the figure of merit.

Another approach to combining the derivative data was to do the combination in real space. This was accomplished by refining each derivative separately and calculating two electron density maps. The resulting real space maps were then averaged, grid point by grid point. Although several features of the averaged map looked better, the overall quality of the map was not markedly better than the SIR map using the HGI4 data alone.

The statistics for the SIRAS refinement looked encouraging enough for a careful inspection of the resulting map. The figure of merit is a measure of the reliability of a phase that can range from 0 to 1. For a perfectly determined phase, the figure of merit is 1.
Table 6. Refinement Statistics.

<table>
<thead>
<tr>
<th>HEAVY ATOM</th>
<th>RESOLUTION</th>
<th># OF REFLEX</th>
<th>FOM</th>
<th>K</th>
<th>C</th>
<th>KRAUT R</th>
<th>CENTRIC R</th>
<th>E/F</th>
</tr>
</thead>
<tbody>
<tr>
<td>HGI4</td>
<td>2.25</td>
<td>3274</td>
<td>.70</td>
<td>1.05</td>
<td>-1.63</td>
<td>5.04</td>
<td>53.78</td>
<td>.30</td>
</tr>
<tr>
<td>U02N</td>
<td>2.50</td>
<td>2432</td>
<td>.71</td>
<td>.87</td>
<td>-5.63</td>
<td>5.65</td>
<td>66.04</td>
<td>.52</td>
</tr>
</tbody>
</table>

RESOLUTION is in angstroms.

FOM is the Figure Of Merit.

K and C are scale constants as described in the text.

KRAUT R is defined as

\[
\frac{\sum |F_{PH}^{obs} - F_{PH}^{calc}|}{\sum F_{PH}^{obs}}
\]

CENTRIC R is defined as

\[
\frac{\sum |F_{PH}^{obs} + F_{P}^{obs} - F_{H}^{calc}|}{\sum |F_{PH}^{obs} - F_{P}^{obs}|}
\]

E/F is the error divided by the magnitude of the scattering due to the heavy atom alone.
Typically, the overall figure of merit should be in the range of 0.7 to 0.9. Figure 5 shows the figure of merit as a function of resolution for both derivatives. It is clear that the phases are reasonably well behaved to high resolution, 2.25 angstroms for the HGI4 derivative.

Scaling

The maps calculated by using the scale factors determined by FINDKC produced electron density maps with huge holes at the heavy atom binding site and large ripples of positive and negative electron density surrounding the site. This is a common observation in heavy atom phased electron density maps (Jack, Ladner and Klug 1976; Wright 1977 and Remington, Wiegand and Huber 1982).

There are many possible explanations for this effect, chiefly lack of isomorphism, but in practice it is very difficult to correct. One contributing factor that can be compensated for is the fact that in the heavy atom derivative data there are additional scatterers in the unit cell as compared to the native data. This means that the derivative data should be scaled not to the parent data, but to a value greater than the parent data to account for the proportional increase in scattering power of the heavy atom derivative.

Normally it is assumed (Blundell and Johnson 1976) that

\[
\frac{K_{PH}\langle I_{PH} \rangle}{K_p\langle I_p \rangle} = 1.0
\]

where \( K_{PH} \) and \( K_p \) are the scale factors for the derivatized crystal scattering and the native crystal scattering, respectively. Similarly, \( \langle I_{PH} \rangle \) and \( \langle I_p \rangle \) are the average scattering for the derivatized and native
Figure 5. Figure of Merit as a Function of Resolution.
crystal. In order to account for the additional scattering for the heavy atoms the relationship should be:

\[
\frac{K_{PH} \langle I_{PH} \rangle}{K_p \langle I_p \rangle} = 1.0 + \frac{\sum f_H^2}{\sum f_p^2}
\]

where \(f_H\) is the scattering due to the additional heavy atoms and \(f_p\) is the scattering of all the atoms in the protein. This ratio can be viewed as a correction term. Assuming one fully occupied \(\text{HgI}_4^{-2}\) site in the asymmetric unit for the \(\text{HGI}_4\) data the correction term is 0.5602. By using this scale factor the holes and ripples were lessened but not eliminated.

Another, more empirical, approach to determine the best scale factor was to calculate a series of electron density maps using different scale factors in the phase refinement and see which value best eliminated the holes. The scale factor was chosen so that the heavy atom was in neither strongly positive or negative density. No scale factor eliminated the large ripples surrounding the heavy atom sites. All methods of determining scale factors provided numbers that were relatively close to each other, but did give subtle differences in the resulting electron density maps.

**Initial Map Interpretation**

All maps were calculated on the Cyber 175 and contoured maps drawn on the Printronix raster plotter. Maps were contoured on somewhat arbitrary levels. The first and second contours were chosen to represent 30% and 15%, respectively, of the map volume as determined by the program FOSTAT. Succeeding contours were chosen with the same increment.
between the first and second contour. The initial map investigations were done at a variety of resolutions on a 1 angstrom grid. Maps were calculated on a scale of 0.6 cm per angstrom and photocopied to transparencies.

The goals of the initial examination of the map were to 1) determine the correct chirality of the map 2) locate the heme plane 3) outline the molecular envelope and 4) determine the position of the cysteiny1 linkages to the heme. If successful in accomplishing these goals, then many restraints would be placed on the possible orientations of the correct structure, given that several c-type cytochrome structures have already been solved.

Determining the heme iron position and the correct chirality of the map was accomplished by calculating an anomalous dispersion difference Fourier map using the HGI4 SIRAS phases. This type of synthesis uses the differences between the Friedel related reflections as the amplitudes for the structure factors. The resulting map displays the imaginary component of the electron density function. If the chirality of the map is correct then a large positive peak should appear at the position of the anomalous scatterers, whereas if the chirality is incorrect a large negative peak should occur at the inverse of the position of the anomalous scatterer (Strahs and Kraut 1968). The resulting map, Figure 6, shows both the chirality and the initial placement of the heme iron to be correct. Being assured of the heme iron position, the orientation of the entire heme plane in the electron density became clear. The iron position was not the largest peak in the electron density map, as larger peaks were found around the heavy atom
Figure 6. Anomalous Dispersion Difference Map Showing Heme Iron Position. Positive contours are drawn with the solid line, while negative contours are dashed.
binding site, but by knowing the iron position fitting the heme plane to the electron density was possible.

Outlining the molecular envelope at 2.25 angstrom resolution, unfortunately, proved to be a much more challenging task. There were regions which were clearly ascribable to solvent channels and by eliminating regions of symmetry related electron density much of the molecular boundary could be assigned. There remained about ten different regions of ambiguity where there were segments of continuous density that connected portions of two obviously different symmetry related molecules. One contributing factor to the difficulty in determining the molecular boundary is that the protein is tightly packed in the unit cell of the crystal. Assuming a typical protein density and using the volume of the unit cell and the molecular weight of the protein, only 34% of the cell volume is occupied by solvent. In most protein crystals, the solvent occupies approximately half of the unit cell volume (Matthews 1968).

A variety of maps at low and intermediate resolution, 5.0, 4.0, 3.0, 3.5 and 2.7 angstroms, were calculated and examined reasoning that it may be easier to determine the gross structural features at lower resolution. However, they proved no more helpful than the 2.25 angstrom map in resolving the ambiguous regions.

It was not possible to determine the cysteinyl linkages at this scale but the map looked encouraging enough to plot it on an expanded scale. An electron density map was calculated on a 0.75 angstrom grid at a scale of 2.0 cm/angstrom. The portion of the unit cell used was defined by $-10 < x < 40/64$, $-3 < y < 30/32$ and $-4 < z < 44/80$. This portion
of the unit cell was chosen on the basis of the earlier map studies defining the molecular envelope. Contour maps were traced by hand onto acetate sheets stretched over metal frames. The sheets were then put in an optical comparator, or Richards' box (Richards 1968), so that a molecular model could be built.

The connections to the heme were more discernable in this map than in the small scale maps. It was not possible to distinguish the methionine side of the heme from the histidine side, but because of the pattern of the heme side chains, as shown in Figure 1, it was possible to make a unique positioning of the heme in the electron density. This orientation gave good density for the propionic side chains and the cysteiny1 linkage at CYS14. The density near CYS17 was present, however, it was not as convincing as at CYS14. This can be explained by the fact that the heavy atom binding site is close to CYS17 introduced noise into this region. Two regions looked like strong candidates for the alpha helices that would be consistent with structural features seen in other c-type cytochromes. It was not possible to unambiguously trace the polypeptide backbone chain in this map although many stretches of continuous density did exist. It was also difficult to pick out very many side chains. The map, despite the problems, looked encouraging enough to make attempts at improving the electron density.
CHAPTER 4

ROTATION FUNCTION STUDIES

There are many examples in small molecule crystallography in which knowledge of a structure, or an appreciable fraction of a structure, can be used to provide a useful set of initial phases (Rossmann 1972). This type of phase determination has been done in both Patterson space and reciprocal space. These approaches are called either molecular replacement or the rotation function. To a lesser extent, these approaches have been applied to protein crystallography. Here the major obstacle is obtaining a suitable trial structure of the protein. The method has been most successful in cases where the crystal structure of a homologous protein has previously been determined by conventional MIR methods.

The rotation function results can also be used in combination with a single isomorphous derivative to remove the phase ambiguity (Hendrickson and Lattman 1970). In a variation of the rotation function, phase information can also be obtained by examining reciprocal space and determining the location of noncrystallographic symmetry elements (Rossmann and Blow 1962, 1963 and 1964). This is accomplished by rotating and comparing the diffraction pattern relative to itself.

Conceptually the rotation function is relatively simple. Using the trial structure, the theoretical scattering pattern is calculated. Next, the observed diffraction pattern of the crystal is rotated to
maximize the fit with the theoretical scattering pattern. After the proper rotational parameters for the trial structure have been determined, the model must be translated to the correct position in the unit cell in relation to the crystallographic symmetry operators. Knowing the appropriate rotation and translation operations to apply to the trial structure we can calculate phases for the observed structure factors. Using this phase information, an electron density map is constructed and the rotated starting model is modified to fit the resulting electron density and to correspond to the appropriate biochemical sequence. The actual performance of this calculation involves a considerable amount of computer time and memory, so the computation must be done as efficiently as possible.

Mathematically, the function to be evaluated is

\[ R(\theta_1, \theta_2, \theta_3) = \sum F^2(C[\theta_1, \theta_2, \theta_3], h) \cdot I(h) \]

where \( F \) is the calculated diffraction pattern for the trial structure that has had the rotation matrix \( C \) applied to it for a given set of Euler angles \( \theta_1, \theta_2 \) and \( \theta_3 \). \( I(h) \) is the observed intensity for the unknown structure. The angles are varied in a manner such that all possible orientations for the trial structure are explored. Ideally, the continuous diffraction pattern of the trial structure would be used; however, computationally the continuous transform is adequately approximated by placing the known structure in a large unit cell so that the transform is sampled on a very fine grid.

The rotation function must be evaluated over the appropriate range of Eulerian angles which is defined by the space group of the
crystal. Lattman (1972) has observed that a more efficient and symmetrical search can be achieved by using a linear combination of Euler angles such that \( \theta_+ = \theta_1 + \theta_3 \) and \( \theta_- = \theta_1 - \theta_3 \).

The resolution of the data used can have a large influence on the success of the calculation. Usually the resolution cutoff limit is such that only reflections within a shell of about 15 to 5 angstroms are used. The lower resolution data is excluded so that contributions from intermolecular vectors will not be important and the high resolution data is ignored because the trial structure is less likely to resemble the true structure at higher resolution. To minimize costs, the data set is further restricted so that only reflections with the largest deviation from the mean diffraction intensity are used.

The last and perhaps most troublesome problem, is relating the molecule to the proper origin of the unit cell. In space group P1 the choice of origin is arbitrary; however, in \( P2_12_12_1 \) the origin is defined by the position of the twofold screw axes. The most often used solution is the translation function as described by Tollin (1966). The translation function is a correlation of the Patterson function for the trial structure and the observed diffraction pattern that displays the intermolecular vectors.

The rotation function was calculated for the cytochrome \( c_{555} \) structure in order to provide additional assurances of the correctness of the structure and to see if the current set of phases could be improved with the additional information. The trial structure used was based upon the cytochrome \( c_{551} \) structure (Almassy and Dickerson 1978).
whose coordinates have been deposited at the Brookhaven Data Bank. The trial structure was composed of the main chain atoms, the beta carbons of the side chains and the atoms in heme plane extending to the first carbon beyond the pyrrole rings. The iron atom was not included in the calculation in order to provide a control to see if the solution to the rotation function would provide sufficient phasing power to bring back the iron in the resulting electron density maps.

The trial structure was centered at the origin of a 128 angstrom cube and structure factors were calculated to a resolution of 5.0 angstroms. The c555 data, in the resolution range of 12 to 5.5 angstroms, were expanded to a hemisphere and the 250 reflections with the greatest deviation from the mean were selected.

The rotation function was then calculated over the appropriate angles, taking into consideration the symmetry of space group P2₁2₁2₁. The initial search was done with an angular increment of 11 degrees. The rotation function results were normalized so that values ranged from 0 to 100. The average value of the normalized result was 40.0 with sigma equal to 14.6. There were three nonsymmetry related peaks with values greater than 90. The highest peak, A, had a value of 100 and the other two peaks, B and C, had a value of 91. The regions around these peaks were sampled with a finer angular increment, 4 degrees, to more accurately determine the rotation angles.

The refined angles of the three different solutions were input to the translation function to determine the correct positioning of the trial structure. The results of all of the translation function experiments were uninterpretable, which is a very common result. The
absence of a definitive result from the translation function required the use of additional information. The iron position of cytochrome c\textsubscript{555} was known from the anomalous dispersion signal as discussed earlier. Due to the fact that there are four cytochrome molecules in the unit cell, there were four possible iron positions. This meant each orientation of the trial structure had to be translated to the four symmetry related iron positions. For the three solutions to the rotation function, twelve different possible orientations of the trial structure had to be investigated.

Structure factors for the twelve different models were calculated in the 10.0 to 5.0 angstrom resolution range. The resulting R factors were all unacceptably large, ranging from 60% to 54%. The twelve possible trial structures were also examined with the graphic display system and compared to the model built to fit the SIR map. In most cases the heme planes of the trial structure and the model did not coincide, nor did the backbone tracing.

Due to the poor initial results and the cost of the preliminary studies, further experiments were not attempted. Recently, a more efficient rotation program (Crowther 1971) has been acquired and it now should be possible to experiment to find the optimal conditions for the rotation function. The atoms comprising the trial structure could be expanded to see if the addition of side chains thought to be homologous could improve the results. More importantly, varying the resolution may have profound effects on the results. It is conceivable that due to the tight packing within the c\textsubscript{555} crystal that the intermolecular vectors
are biasing the results. By eliminating more of the low resolution reflections this problem could be overcome. With the more efficient rotation program these studies should become computationally feasible.
CHAPTER 5

MAP IMPROVEMENT BY FOURIER INVERSION

Interpretation of electron density maps and subsequent model building is perhaps the most difficult and time consuming step in protein crystallography. The overall quality of the map dictates how time consuming and frustrating the model building will be. Anything that significantly improves the quality of the map greatly aids in its interpretation. Building a model in a high quality map is an exciting and rewarding experience, whereas interpreting a poorly resolved map is a frustrating and uncertain experience.

In the last several years a lot of work and effort has been spent on trying to improve electron density maps by a technique generally referred to as Fourier inversion. There are four basic steps in this approach. First, the calculation of an electron density map using the observed structure factor amplitudes and an initial set of phases, usually MIR or SIRAS. The next step is to improve the resulting electron density map in a manner that gives a better representation of the true electron density. This 'improved' map is then Fourier inverted, generating a new set of structure factor amplitudes and phases. The final step is to combine the calculated structure factors with the observed data to generate Fourier terms for a new electron density map. The process is then repeated until convergence, usually judged by the average change in the phase angle. This process is shown in Figure 7.
Figure 7. Procedure for Fourier Inversion.
Since the starting map will strongly bias the final set of phases, it is important to use the best starting map possible, devoid of any false peaks. By analyzing the centroid distribution, Table 6, and visually inspecting maps at various centroid cutoffs, it was decided to use only reflections with a centroid of 0.6 or greater. This value kept 67% of the reflections yet eliminated the poorly phased and generally weaker reflections. This rejection criterion seemed to reduce the noise in the map. The relationship between the centroid and the average phase error (Blundell and Johnson 1976) is given by:

\[ m = \langle \cos \Delta \alpha \rangle \]

where \( m \) is the centroid, or figure of merit, and \( \Delta \alpha \) is the difference between the experimentally determined and the true phase. Therefore, reflections with a centroid of 0.6 have an average phase error of 53 degrees.

Various map improvement schemes have been suggested in the literature. One common feature has been to define the molecular envelope and thereby define the solvent region. The solvent region is then replaced with a uniformly dense or level background (Hendrickson, Klippenstein and Ward 1975; Schevitz, Podjarny, Zwick, Hughes and Sigler 1981), eliminating large peaks and deep valleys in the regions known to contain solvent.

Another technique used to alter electron density maps is to enhance the strongly positive regions, most likely to be regions of protein density, and to diminish the weaker density (Collins, Cotton,
Table 7. Centroid Distribution for HGI4 Derivative.

This table gives the number of reflections within the specified centroid range as a function of resolution. Only reflections with a centroid greater than 0.6 were used to calculate the initial map.

<table>
<thead>
<tr>
<th>RESOLUTION ANGSTROMS</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.43</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>5</td>
<td>11</td>
<td>15</td>
<td>29</td>
<td>40</td>
<td>127</td>
</tr>
<tr>
<td>4.31</td>
<td>41</td>
<td>7</td>
<td>4</td>
<td>14</td>
<td>16</td>
<td>12</td>
<td>29</td>
<td>30</td>
<td>46</td>
<td>104</td>
</tr>
<tr>
<td>3.76</td>
<td>11</td>
<td>5</td>
<td>14</td>
<td>12</td>
<td>13</td>
<td>12</td>
<td>23</td>
<td>43</td>
<td>54</td>
<td>87</td>
</tr>
<tr>
<td>3.42</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>20</td>
<td>18</td>
<td>24</td>
<td>32</td>
<td>49</td>
<td>74</td>
</tr>
<tr>
<td>3.18</td>
<td>8</td>
<td>10</td>
<td>7</td>
<td>7</td>
<td>21</td>
<td>26</td>
<td>21</td>
<td>43</td>
<td>53</td>
<td>57</td>
</tr>
<tr>
<td>2.99</td>
<td>6</td>
<td>6</td>
<td>12</td>
<td>17</td>
<td>22</td>
<td>32</td>
<td>25</td>
<td>31</td>
<td>41</td>
<td>54</td>
</tr>
<tr>
<td>2.84</td>
<td>5</td>
<td>13</td>
<td>14</td>
<td>14</td>
<td>12</td>
<td>26</td>
<td>28</td>
<td>33</td>
<td>47</td>
<td>42</td>
</tr>
<tr>
<td>2.71</td>
<td>4</td>
<td>10</td>
<td>11</td>
<td>19</td>
<td>25</td>
<td>17</td>
<td>27</td>
<td>33</td>
<td>27</td>
<td>47</td>
</tr>
<tr>
<td>2.61</td>
<td>3</td>
<td>6</td>
<td>17</td>
<td>20</td>
<td>31</td>
<td>29</td>
<td>23</td>
<td>34</td>
<td>28</td>
<td>44</td>
</tr>
<tr>
<td>2.52</td>
<td>9</td>
<td>9</td>
<td>17</td>
<td>13</td>
<td>26</td>
<td>31</td>
<td>27</td>
<td>33</td>
<td>28</td>
<td>48</td>
</tr>
<tr>
<td>2.44</td>
<td>4</td>
<td>7</td>
<td>14</td>
<td>12</td>
<td>9</td>
<td>19</td>
<td>21</td>
<td>34</td>
<td>13</td>
<td>24</td>
</tr>
<tr>
<td>2.37</td>
<td>5</td>
<td>16</td>
<td>15</td>
<td>13</td>
<td>23</td>
<td>25</td>
<td>31</td>
<td>36</td>
<td>27</td>
<td>37</td>
</tr>
<tr>
<td>2.31</td>
<td>7</td>
<td>12</td>
<td>7</td>
<td>18</td>
<td>23</td>
<td>21</td>
<td>27</td>
<td>26</td>
<td>28</td>
<td>30</td>
</tr>
<tr>
<td>2.18</td>
<td>4</td>
<td>4</td>
<td>13</td>
<td>19</td>
<td>8</td>
<td>25</td>
<td>25</td>
<td>30</td>
<td>32</td>
<td>31</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>122</strong></td>
<td><strong>114</strong></td>
<td><strong>155</strong></td>
<td><strong>193</strong></td>
<td><strong>254</strong></td>
<td><strong>304</strong></td>
<td><strong>346</strong></td>
<td><strong>467</strong></td>
<td><strong>513</strong></td>
<td><strong>806</strong></td>
</tr>
</tbody>
</table>
Hazen, Meyer and Morimoto 1975; Raghavan and Tulinsky 1979); this effectively makes the peaks sharper.

The inversion of the modified map is accomplished by the Fast Fourier Transform, FFT (Cooley and Tukey 1965), which is an exceedingly rapid algorithm for calculating discrete Fourier transforms. The speed and efficiency of this algorithm have prompted an increased number of Fourier inversion studies to be attempted.

Two slightly different approaches have been tried to combine the observed structure factor amplitudes with the new calculated structure factors. First, the observed structure factor amplitudes were used in combination with the calculated phases. This is the most straightforward approach to combine the experimentally determined amplitudes with the unobservable, but calculable, phases. However, this approach tends to propagate erroneous peaks. In an attempt to eliminate the spurious peaks, a $2F_o - F_c$ synthesis was used. $F_o$ and $F_c$ are the observed and the calculated structure factor amplitudes. The rationale behind this synthesis is that by subtracting out the contributions of the input calculated structure in the structure factor amplitudes, incorrect features in the newly calculated model will tend to be eliminated.

The biggest difficulty to the Fourier inversion approach is to assess the quality of the resulting map. With an almost infinite variety of map editing procedures and resolution cutoffs to vary, it is easy to generate a great number of maps and the only true test of the map quality is to spend four to six months building the model.
3.0 Angstrom Studies

Initial Fourier inversion studies were carried out at three angstroms resolution. The map smoothing or improving was done by considering three different regions of density: the solvent area, the protein density, and the spurious peaks around the heavy atom binding site. As noted above, it was not possible to define explicitly the solvent region of the SIRAS map, so solvent leveling could not be performed as reported by others. Instead the solvent was defined as those grid points whose density was below a specified threshold. This value was chosen so that a given percentage of map grid points were below the solvent cutoff. Initially, 65% of the total cell volume was used as a cutoff because this is generally the value chosen for the first contour level of the protein in the electron density maps. The solvent grid points were normalized to range from 0 to negative 10. The most negative value in the map was set to -10, and the solvent cutoff value was set to 0. This region was then modified with a square root function.

The square root smoothing function as opposed to a uniform solvent background was chosen for several reasons. First, since the cutoff value is rather arbitrary, there most certainly will be grid points inside the true molecular envelope that we are now considering to be part of the solvent, so it is best if not all 'solvent' grid points are treated identically. The other reason is that it is undesirable to introduce a sharp, discontinuous function into the electron density map so a more gradual function was devised.
The function used to modify the protein portion of the map was derived from Collins' (1975) original suggestion of using:

\[ \rho_{\text{new}} = 3\rho^2 - 2\rho^3 \]

where \( \rho \) is the electron density. This function is defined in such a manner that the dynamic range of the map must be normalized to range from 0 to 1. The zero of the map is the value of the solvent cutoff and the maximum value of the map is set to 1. As can be seen from the polynomial portion of the curve in Figure 8, the region from 0 to 1/2 in the map is slightly downweighted and the region from 1/2 to 1 in the map is enhanced. Unfortunately, but not surprisingly, by simply normalizing from the solvent cutoff to the largest value of the map to be 0 to 1, the vast majority of the grid points fell in the region to be downweighted. In order to get a better distribution of points, the midpoint was chosen such that an equal number of grid points would be downweighted as would be enhanced. This condition caused some of the normalized points to have values greater than 1. To solve this problem, the smoothing function was altered so that the polynomial portion of the curve was replaced with a straight line from the point of greatest enhancement, 0.677, to the maximum value of the map as shown in Figure 8.

The most intense peak did not always, in fact never in an initial map, correspond to the position of the heme iron as determined from the anomalous scattering. These spuriously high peaks were usually associated with ripples from the heavy atom sites. For normalization purposes, the density at the heme iron position was considered the
highest value of the map, and this meant that the spurious peaks when weighted by the linear function would actually be downweighted (see Figure 8).

The most disrupting influence in the map was the region near the heavy atom binding site. As described earlier, this problem was dealt with as best as possible by adjusting the scale factor, but with the Fourier inversion we have the opportunity to remove the disturbing features in real space and recalculate the phases. The ripples were removed in a manner that tried to avoid introducing any sharp, discontinuous features into the new map. To achieve this purpose, a ten angstrom radius sphere was placed around the mercury atom of the heavy atom binding site. The value of the electron density for each grid point within the sphere was weighted by one tenth the distance from the mercury atom. That means that grid points 1.0 angstrom from the heavy atom binding site were multiplied by the factor 0.1 and grid points nine angstroms away were multiplied by 0.9. This provided a gradual leveling function which eliminated the ripples due to the heavy atom. The normal procedure was to first level the heavy atom ripples, then to smooth the map with the functions shown in Figure 8 and then to invert. After two cycles using this combination of map modifications and inversion, the heavy atom ripples were gone and no subsequent adjustments were made in the region near the heavy atom binding site. An encouraging sign was that the heme iron then became and remained the strongest feature in the map.
Figure 8. Smoothing Function Applied to Electron Density. Points between the lowest value of the map (A) and the solvent cutoff point (B) are smoothed with a square root function. The map is normalized such that 50% of the grid points greater than the solvent cutoff fall in the range between B and C. D is the point of maximum enhancement and E is the value for the heme iron.
Table 8. Fourier Inversion Refinement at 3.0 Angstroms.

<table>
<thead>
<tr>
<th>CYCLE</th>
<th># REFS</th>
<th>R%</th>
<th>Δα_{SIR}</th>
<th>Δα_{LAST}</th>
<th>% USED</th>
<th>Fc/Fo</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1171</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>75.5</td>
<td>----</td>
<td>Starting Map</td>
</tr>
<tr>
<td>1</td>
<td>1339</td>
<td>31.36</td>
<td>26.4</td>
<td>13.3</td>
<td>86.5</td>
<td>.30</td>
<td>Heavy atom ripples leveled</td>
</tr>
<tr>
<td>2</td>
<td>1370</td>
<td>27.93</td>
<td>35.8</td>
<td>7.0</td>
<td>88.5</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1380</td>
<td>24.89</td>
<td>40.0</td>
<td>5.5</td>
<td>89.1</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1388</td>
<td>24.48</td>
<td>43.3</td>
<td>4.2</td>
<td>89.7</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>4'</td>
<td>1453</td>
<td>24.48</td>
<td>44.6</td>
<td>4.2</td>
<td>93.9</td>
<td>.15</td>
<td>Relax Fc/Fo ratio</td>
</tr>
<tr>
<td>5</td>
<td>1470</td>
<td>21.26</td>
<td>46.7</td>
<td>3.9</td>
<td>95.0</td>
<td>.15</td>
<td></td>
</tr>
</tbody>
</table>
The inversion phases were combined with the observed structure factor amplitudes to give the next cycle of maps. The ratio of $F_o$ to $F_c$ was used for as a cutoff criterion for the reflections. This rejection ratio was used to eliminate reflections with clearly incorrect calculated phases. For the first several cycles a cutoff of 0.30 was used and after the phases appeared to converge this condition was relaxed to 0.15. A summary of the progress of the 3.0 angstrom inversion is given in Table 8. The final map produced was encouraging, but not so clear that any new interpretations or conclusions could be drawn. Instead of building from the 3 angstrom map, considering the minimal amount of time and computer costs required to do the inversion, another inversion series was performed, this time starting at 2.25 angstroms and with the hope of extending the resolution to 2.0 angstroms.

**2.0 Angstrom Studies**

The Fourier inversions to 2.25 angstroms followed essentially the same steps as in the 3 angstrom study. The initial SIRAS map was calculated at 2.25 angstroms using all reflections greater than two sigma with a centroid greater than 0.6 on a 1 angstrom grid. The same map modification procedure was used. Once convergence was reached all reflections greater than one sigma were included in the calculation of a new map and several more cycles of inversion were performed. The heavy atom binding site required no special attention with the addition of the one sigma data. Once convergence had been achieved, reflections to 2.0 angstroms with a one sigma cutoff were used to calculate the next map. The grid spacing was changed to 0.75 angstroms to ensure proper sampling.
Table 9. Fourier Inversion Refinement at 2.0 Angstroms.

<table>
<thead>
<tr>
<th>CYCLE</th>
<th># REF</th>
<th>R%</th>
<th>Δa_SIR</th>
<th>Δa_LAST</th>
<th>% USED</th>
<th>Fc/Fo</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2187</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2661</td>
<td>34.87</td>
<td>28.9</td>
<td>12.4</td>
<td>81.3</td>
<td>.30</td>
<td>Starting Sir Map</td>
</tr>
<tr>
<td>2</td>
<td>2799</td>
<td>32.21</td>
<td>40.0</td>
<td>8.4</td>
<td>85.5</td>
<td>.30</td>
<td>Heavy atom ripples leveled</td>
</tr>
<tr>
<td>3</td>
<td>2850</td>
<td>27.66</td>
<td>44.8</td>
<td>6.1</td>
<td>87.0</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2858</td>
<td>26.63</td>
<td>47.8</td>
<td>4.6</td>
<td>87.3</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2868</td>
<td>25.93</td>
<td>50.2</td>
<td>3.6</td>
<td>87.6</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>5'</td>
<td>3036</td>
<td>25.93</td>
<td>51.9</td>
<td>3.6</td>
<td>92.7</td>
<td>.15</td>
<td>Relax Fc/Fo ratio</td>
</tr>
<tr>
<td>6</td>
<td>3088</td>
<td>21.82</td>
<td>54.2</td>
<td>3.8</td>
<td>94.3</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>6'</td>
<td>3361</td>
<td>22.79</td>
<td>84.2</td>
<td>---</td>
<td>91.7</td>
<td>.30</td>
<td>Include 1 sigma data 2.25 angst</td>
</tr>
<tr>
<td>7</td>
<td>3376</td>
<td>23.21</td>
<td>84.6</td>
<td>3.2</td>
<td>92.1</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3384</td>
<td>22.81</td>
<td>84.7</td>
<td>2.7</td>
<td>92.3</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3391</td>
<td>22.53</td>
<td>84.9</td>
<td>2.3</td>
<td>92.5</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>9'</td>
<td>3467</td>
<td>22.53</td>
<td>84.9</td>
<td>2.3</td>
<td>94.6</td>
<td>.15</td>
<td>Relax Fc/Fo ratio</td>
</tr>
<tr>
<td>10</td>
<td>3484</td>
<td>21.23</td>
<td>85.0</td>
<td>2.6</td>
<td>95.1</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>10'</td>
<td>4312</td>
<td>28.94</td>
<td>86.1</td>
<td>2.6</td>
<td>84.2</td>
<td>.30</td>
<td>Include 1 sigma data to 2.0 ang</td>
</tr>
<tr>
<td>11</td>
<td>4383</td>
<td>26.60</td>
<td>86.2</td>
<td>4.5</td>
<td>85.5</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>4400</td>
<td>25.90</td>
<td>86.4</td>
<td>2.7</td>
<td>85.9</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>12'</td>
<td>4684</td>
<td>25.73</td>
<td>86.4</td>
<td>2.0</td>
<td>91.4</td>
<td>.15</td>
<td>Relax Fc/Fo ratio</td>
</tr>
<tr>
<td>13</td>
<td>4737</td>
<td>22.59</td>
<td>86.4</td>
<td>2.9</td>
<td>92.4</td>
<td>.15</td>
<td></td>
</tr>
</tbody>
</table>
of real space for the higher resolution data. One important aspect of Fourier inversion is that higher resolution phases can be calculated without having to collect heavy atom derivative data (Barrett and Zwick 1971). The additional reflections were cycled to convergence. Table 9 gives a summary of the progress of this round of Fourier inversion.

A 2.0 cm/angstrom map was calculated and displayed as described earlier. The molecular envelope was defined as best possible on a smaller scale. The unique molecule was contoured in black, while the symmetry related densities were contoured in red. This made it easier to visualize the unique molecule and at the same time the symmetry related density was present so packing interactions could be studied. This also allowed changes in the interpretation of the tentative molecular boundary to be easily made. In the process of building the model, it seemed that the inversion process had made the peaks too sharp at the expense of losing the overall connectivity in the map. A model was built and coordinates were measured by use of a coordinate measuring machine (Salemme and Fehr 1972). After a model had been built another round of inversion studies was performed.

The same reflection data set was used and the same starting map was used as in the first 2.25 angstrom study. This time, however, a 0.75 angstrom grid was used throughout the entire study. Proper sampling of real space when performing the density improvement studies has become a point of contention and it is generally safest to sample as finely as economically feasible. The cutoff points for the smoothing function were changed so that the solvent value was taken to be 70% of the map volume and the midpoint of the curve was then set to 35%. The
lower solvent cutoff value was chosen to avoid placing all the electron density into the strongest peaks and losing the connectivity in the map. Also the maps for the next cycle of inversion were calculated using $2F_o - F_c$ synthesis with the calculated phases. Again this strategy was chosen because it was a more conservative approach. The results of this inversion are given in Table 10 and plotted in Figure 9. The overall impression of this map is that this approach produced a better map and future inversion studies should use this combination of techniques from the outset.
Table 10. Alternative Approach to Fourier Inversion Refinement at 2.0 Angstroms.

<table>
<thead>
<tr>
<th>CYCLE</th>
<th>#</th>
<th>R%</th>
<th>Δa_{SIR}</th>
<th>Δa_{LAST}</th>
<th>% USED</th>
<th>Fc/Fo</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2187</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>66.8</td>
<td>----</td>
<td>2.25A SIR Map</td>
</tr>
<tr>
<td>1</td>
<td>2545</td>
<td>33.77</td>
<td>21.8</td>
<td>9.3</td>
<td>77.7</td>
<td>.30</td>
<td>Heavy atom</td>
</tr>
<tr>
<td>2</td>
<td>2604</td>
<td>31.44</td>
<td>29.2</td>
<td>7.2</td>
<td>79.5</td>
<td>.30</td>
<td>Ripples leveled</td>
</tr>
<tr>
<td>3</td>
<td>2528</td>
<td>31.82</td>
<td>29.6</td>
<td>7.2</td>
<td>77.2</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2552</td>
<td>30.99</td>
<td>32.3</td>
<td>4.0</td>
<td>77.9</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2551</td>
<td>31.92</td>
<td>34.4</td>
<td>3.4</td>
<td>77.9</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2549</td>
<td>31.42</td>
<td>35.9</td>
<td>2.9</td>
<td>77.9</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>6'</td>
<td>2684</td>
<td>31.42</td>
<td>38.4</td>
<td>2.8</td>
<td>82.0</td>
<td>.15</td>
<td>Relax Fc/Fo ratio</td>
</tr>
<tr>
<td>7</td>
<td>2744</td>
<td>31.00</td>
<td>40.6</td>
<td>3.4</td>
<td>83.8</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2757</td>
<td>29.80</td>
<td>41.9</td>
<td>2.8</td>
<td>84.2</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2760</td>
<td>29.55</td>
<td>42.9</td>
<td>2.4</td>
<td>84.3</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2998</td>
<td>30.32</td>
<td>83.7</td>
<td>----</td>
<td>81.8</td>
<td>.30</td>
<td>Include 1 sigma</td>
</tr>
<tr>
<td>11</td>
<td>3041</td>
<td>29.63</td>
<td>84.2</td>
<td>3.5</td>
<td>83.0</td>
<td>.30</td>
<td>data to 2.25Ang</td>
</tr>
<tr>
<td>12</td>
<td>3042</td>
<td>29.29</td>
<td>83.9</td>
<td>4.0</td>
<td>83.0</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>3046</td>
<td>28.93</td>
<td>84.4</td>
<td>3.3</td>
<td>83.1</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>13'</td>
<td>3090</td>
<td>28.93</td>
<td>84.6</td>
<td>3.8</td>
<td>84.3</td>
<td>.15</td>
<td>Relax Fc/Fo ratio</td>
</tr>
<tr>
<td>14</td>
<td>3104</td>
<td>28.71</td>
<td>84.2</td>
<td>3.5</td>
<td>84.7</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>3106</td>
<td>28.47</td>
<td>84.6</td>
<td>3.2</td>
<td>84.8</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>3427</td>
<td>37.28</td>
<td>85.7</td>
<td>----</td>
<td>66.9</td>
<td>.30</td>
<td>Expand data to</td>
</tr>
<tr>
<td>17</td>
<td>3508</td>
<td>37.24</td>
<td>85.3</td>
<td>5.7</td>
<td>68.5</td>
<td>.30</td>
<td>2.0 Angstroms</td>
</tr>
<tr>
<td>18</td>
<td>3532</td>
<td>36.67</td>
<td>86.6</td>
<td>5.7</td>
<td>68.9</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>3544</td>
<td>36.82</td>
<td>85.5</td>
<td>5.3</td>
<td>69.2</td>
<td>.30</td>
<td></td>
</tr>
<tr>
<td>19'</td>
<td>3857</td>
<td>36.82</td>
<td>86.2</td>
<td>8.3</td>
<td>75.3</td>
<td>.15</td>
<td>Relax Fc/Fo ratio</td>
</tr>
<tr>
<td>20</td>
<td>3995</td>
<td>35.55</td>
<td>87.6</td>
<td>5.4</td>
<td>78.0</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>4043</td>
<td>34.77</td>
<td>87.3</td>
<td>5.1</td>
<td>78.9</td>
<td>.15</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>4062</td>
<td>33.93</td>
<td>87.7</td>
<td>4.8</td>
<td>79.3</td>
<td>.15</td>
<td></td>
</tr>
</tbody>
</table>
Figure 9. Progress of Fourier Inversion.
Closed circles indicate the phase change from the SIR phases.
Open circles indicate phase change from the previous cycle.
CHAPTER 6

DESCRIPTION OF THE MOLECULE

The overall folding of the cytochrome c\textsubscript{555} structure is very similar to that of the other previously determined c-type cytochromes. In the next chapter the structures of the other c-type cytochromes will be discussed and compared to the cytochrome c\textsubscript{555} presented here. There were several regions where there was little connectivity between segments of the polypeptide chain and many of the side chains were either ambiguous or ill-defined. The overall folding of the backbone chain appears to be reasonable; however, further refinement will be necessary to resolve the remaining ambiguities. Figures 10 and 11 show the folding of the cytochrome c\textsubscript{555} structure. Figure 10 shows just the backbone atoms while Figure 11 shows the side chains. In Figures 10a and 11a the c\textsubscript{555} structure is shown in what has become the standard cytochrome orientation. It is with reference to this orientation that parts of the chain will be described as being on the left or right side of the molecule.

Backbone Folding

The first eleven residues of the polypeptide chain form an alpha helix which has been observed for all other cytochrome structures. The helix is broken to make the thioether linkages to the vinyl groups of the heme with the sulfur atoms of cystiene residues 14 and 17. The chain continues to make the HIS18 linkage to the heme.
Figure 10. Stereoscopic View of Cytochrome c$_{555}$ Backbone Conformation. In A the heme is viewed from the edge.
Figure 10. Stereoscopic View of Cytochrome c₅₅₅ Backbone Conformation.
In A the heme is viewed from the top.
Figure 11. Stereoscopic View of Cytochrome c$_{555}$ Side Chain Conformation.
In A the heme is viewed from the edge.
Figure 11. Stereoscopic View of Cytochrome c555 Side Chain Conformation.
In B the heme is viewed from the top.
The density after the histidine for the next six to eight residues is very ambiguous; in fact, two radically different interpretations for the backbone chain have been built. One choice is to have the chain continue in a path that puts this segment of the chain folding back towards the center of the molecule and in the proximity of the heme. The other alternative is to build the chain so that this segment extends more towards the solvent at the bottom of the molecule. The problem with the latter interpretation is that a portion of the right side of the heme is left very much exposed to the solvent. Both alternatives are shown in Figure 12. Both models allow the carbonyl oxygen of PRO26 to be hydrogen bonded to the NH of HIS18 which is linked to the heme.

The polypeptide chain then continues along forming a short segment that appears to be a single turn of a helix. The chain then folds back on itself to the other side of the heme plane forming a long helix of twelve residues. This middle helix is on the opposite side of the heme plane from HIS18. The helix breaks to form a rather extended loop that covers and encloses much of the heme. A large portion of this loop lies in a plane roughly parallel to the heme, but the portion of the loop that forms the bend turns down around the heme to enclose the heme at the bottom of its pocket. As the backbone loop folds back upon itself the sulfur of MET60 makes a bond to the heme iron, filling in the sixth coordination site.

After MET60 there is a region of uncertainty and weak density. There are several possible interpretations for this region, none of which fit the electron density very well or are very satisfying from
Figure 12. Alternative Models for Residues 19 to 26.
Model A loops to the outside edge of the heme, while model B folds underneath the heme.
structural considerations. The final segment of polypeptide chain is a fairly well defined alpha helix from residues 72 to 86. This carboxy terminal alpha helix closes off the back side of the molecule.

**Heavy Atom Binding Sites**

The mercury tetra-iodide binds near the sulfur of CYS17 which is similar to the binding observed in tuna cytochrome c (Takano, Swanson, Kallai and Dickerson 1971). The A site of the uranyl derivative binds near the beginning of the amino terminal helix. This is consistent with the observation that there is a dipole associated with the alpha helix that amounts to half of a charge unit at either end of the helix (Hol, Van Duijnen, and Berendsen 1978). The B site binds near LYS40 of the middle helix and site C binds in the ill-defined region preceding the carboxy terminus. All sites occur at reasonable positions on the surface of the molecule which is confirmation that the molecular envelope was chosen properly in these regions of the map.

**Description of the Heme Environment**

As noted above, the heme environment on the right side of the cytochrome c₅₅₅ structure was subject to two different interpretations. One alternative places both MET22 and THR20 in close proximity to the heme. The other interpretation leaves much of the heme exposed to the solvent. The former explanation is a more probable and pleasing interpretation. Either interpretation is consistent with having the carbonyl oxygen of PRO26 pointing towards the hydrogen associated with the ring nitrogen of ND1 of HIS18. The possible significance of this will be discussed in the next chapter.
The left side of the heme crevice is formed predominately by the middle helix of residues 36 to 48 and the loop from 49 to 60. The side chain densities for the middle helix are not especially sharp so it is possible that the helix may be one turn out of register. However, based on the sequence in this region, the side chains that face the interior of the molecule and pack next to the heme are all hydrophobic in nature. The loop after this stretch of helix covers the remainder of the left side of the heme and encloses the bottom portion of the heme.

The electron density for the side chain that occupies the sixth ligand is very weak. This could be due to either thermal disorder in the region, or perhaps the inversion cycles built up the electron density of the iron at the expense of sulfur density. Whatever the reason, the precise orientation of the MET60 side chain was hard to assign.

**Intramolecular Bonding**

Any protein is stabilized by a variety of intramolecular bonds. The middle and carboxy terminal helices have regular geometry and reasonable hydrogen bonds that have been observed for alpha helices. The amino terminal helix appears to be distorted, but this is probably more a reflection of the quality of the map than an unusual type of helix. Upon refinement it is expected that the bond angles and hydrogen bond lengths will approach the standard values. The carboxy helix interacts with the amino helix through the side chains of ASP2 and THR9 to stabilize the structure.
Another interesting feature is that NEH2, the nitrogen of the imidazole ring, of TRP34 is hydrogen bonded to the inner propionate, OC22, of the heme. This hydrogen bond helps stabilize the negative charge in the interior of the protein.

**Intermolecular Contacts**

As noted earlier the protein is packed rather tightly in the unit cell and there are many intermolecular contacts between symmetry related molecules that caused a great deal of difficulty when trying to define the molecular envelope.

There is extensive contact between the phenyl rings of TYR10 and TYR80 of a symmetry related molecule. This gave rise to an extended region of strong electron density across the twofold screw axis that at first appeared to be a continuous segment of polypeptide chain, and yet it was illogical to consider it part of one molecule due to symmetry arguments. Another site of intermolecular interactions across a twofold screw is between LYS 40 and LYS 57. This interaction could be mediated by a bridging sulfate ion.

The unit cell parameters are such that along the short b axis, 24 angstroms, the molecules are tightly packed by a unit cell translation. This meant that the end of the molecule near PRO 36 and HIS 37 at the start of the middle helix, was packed against the heme edge around MET 60 and the CYS17 linkage of the next molecule.
CHAPTER 7

COMPARISONS TO OTHER CYTOCHROME STRUCTURES

There is a great wealth of knowledge regarding the c-type cytochromes in terms of both primary sequences and tertiary structures. Over one hundred primary sequences have been determined from a wide variety of organisms, and the database is constantly expanding. For the eukaryotic system a convincing model can be made to explain the pattern and rate of cytochrome c evolution that correlates well with other measures of evolution (Dickerson and Timkovich 1975).

The prokaryotic kingdom, with its greater diversity, has been more difficult to organize. Dickerson (1980a, 1980b) has classified the prokaryotic cytochromes into basically three classes based on length and structural homology. The longest class contain roughly 110 to 140 amino acids. The medium length class contain 95 to 110 residues and is most homologous with the eukaryotic cytochrome c. The cytochromes with 80 to 95 residues are considered in the short class, which is the classification of cytochrome c\textsubscript{555}.

These are not arbitrary distinctions; instead the classifications were made possible by an examination of the three-dimensional structure of several different prokaryotic cytochromes. The list of prokaryotic crystal structures include c\textsubscript{2} from \textit{Rhodospirillum rubrum} (Salemme, Freer, Xuong, Alden and Kraut 1973), cytochrome c\textsubscript{550} from \textit{Paracoccus denitrificans} (Timkovich, Dickerson and Margoliash 1976)
and cytochrome c551 from *Pseudomonas aeruginosa* (Almassy and Dickerson 1978 and Matsuura, Takano and Dickerson 1982) and c554 from *Anacystis nidulans* (Ludwig, Patridge, Powers, Dickerson and Takano 1982). These structures, in combination with the eukaryotic tuna mitochondrial cytochrome c (Swanson et al 1977), give a structural sampling covering a broad spectrum of organisms and a wide range of oxidation-reduction potentials.

Salemme (1977) has described the conserved structural elements found in all c-type cytochromes and noted how the structures have compensated for the insertion and deletions of various portions of the molecule. These points will be reviewed here briefly. Amino and carboxy helices have been observed in all mitochondrial c-type cytochrome structures to date. In some cases there is a stretch of extended chain that either precedes the amino helix or trails the carboxy helix, notably in cytochrome c550 (Timkovich, Dickerson and Margoliash 1976) which has extended chains at both termini. In almost all of the sequences determined, there is a glycine residue in a position which most probably corresponds to the middle of the amino terminus helix. This glycine plays an important structural role since it is found at the point of closest approach of the amino and carboxy helices. Another structurally conserved helix is the helix on the left side of the molecule. This helix comprises much of the left side of the cytochrome.

Two other amino acid residues, other than the ones involved with the heme linkages, that tend to be conserved are a proline and a tryptophan residue. The proline residue is important since the carbonyl
in most cases is hydrogen bonded to the ND1 of the histidine bound to the heme. The tryptophan appears to be important in stabilizing the negative charge on the heme propionate. In most structures the nitrogen of the tryptophan ring is hydrogen bonded to the propionate.

Additional helices and loops are seen in the medium and long class of cytochromes; however, the remainder of this discussion will be limited to the short class of molecules. The structures that have been solved in this class are the c₅₅₁ from *P. aeruginosa* to 1.6 angstrom resolution (Matsuura et al 1982), c₅₅₄ from *A. nidulans* to 3.0 angstrom resolution (Ludwig et al 1982) and the present study of c₅₅₅ from *C. thiosulfatophilum* to 2.0 angstroms. The results from the *A. nidulans* structure are quite tentative, but it is interesting to note that many of the regions of uncertainty correspond to ill-defined regions in the c₅₅₅ structure, notably the region on the right side of the molecule after the histidine linkage.

Both cytochromes c₅₅₅ and c₅₅₄ possess a glycine in the middle of the amino terminal helix at the point of closest approach of the amino and carboxy helicities as found in most cytochromes. The c₅₅₁ structure, however, lacks a glycine at the amino terminus but compensates with a proline in the corresponding position.

In both the c₅₅₅ and the c₅₅₁ structure it appears that a proline carbonyl is hydrogen bonded to the histidine ligand of the heme as has been observed in most all other cytochromes. Although there is a proline in a homologous position in the c₅₅₄ structure, in the initial model the carbonyl oxygen appears to be pointing away from the
histidine. There has been no other side chain or carbonyl group mentioned which could be hydrogen bonded to the histidine.

The tryptophan that bonds to the inner propionate of the heme is not always in a homologous region of the backbone chain. In the c555 structure TRP34 is on the segment before the start of the middle helix while in c551 the bond is formed with TRP56 which forms part of the loop at the bottom of the molecule that encloses the heme. This is a case where there is structural homology but not sequence homology. Interestingly, the only tryptophan in the c554 structure is at the carboxy terminus and appears not to bind to the inner propionate.

The region between the methionine heme linkage and the carboxy helix is ill-defined in both the c554 and the c555 structures. In the c551 structure, this region is four or five amino acid residues shorter and serves as a direct connection between the methionine and the carboxy helix. This is also part of a proline rich stretch of c551 which may impart a high degree of rigidity to the structure.
CHAPTER 8

SIGNIFICANCE OF STRUCTURE AND FUTURE WORK

There are a variety of experiments that suggest the oxidized structure of cytochromes is conformationally different from the reduced form. This conclusion is supported by proteolytic digestion studies (Nozaki, Mizushima, Horio and Okunuki 1988), D$_2$O exchange (Ulmer and Kagi 1968) and compressibility studies (Eden, Matthew, Rosa and Richard 1982). The crystal structures of the two oxidation states, however, show very little conformational change. There were very few differences in the oxidized and reduced structures in either the cytochrome c$_{551}$ or the tuna cytochrome c structures. In the c$_{551}$ studies (Matsuura et al. 1982), the reduction was carried out on the crystalline protein with no reported alteration in the unit cell parameters. The tuna cytochrome c structure (Takano and Dickerson 1980) was solved in a different space group for each oxidation state with small differences between the two structures.

If the differences between oxidation states are not to be found in the static structures as deduced by x-ray crystallography, then the differences must lie in the dynamic structures as suggested by Salemme (1977) and Cooper (1980). Interpreted from this viewpoint, the reduced structure must be a tighter or closed structure compared to the more open or floppy oxidized structure. This interpretation is consistent with preliminary NMR results (Wuthrich, unpublished observations) from
solution studies of cytochrome c555. There appear to several conformational states seen in the hydrogen NMR of the oxidized molecule that coalesce to one state upon reduction of the protein.

This implies that the regions of ambiguity may not be due solely to errors in the electron density map, but perhaps the molecule is partially disordered. For this reason, it may be more profitable to attempt to solve the reduced c555 structure in order to get a better starting model for the ambiguous regions in the oxidized structure. It will be interesting to analyze the temperature factors of the refined oxidized and reduced structures to see if there is an overall difference or if there are differences localized at specific segments of the molecule.

The differences in the dynamic structure may also be a contributing factor in explaining the differences in oxidation-reduction potentials. Several proposals have been presented to explain the observed differences in oxidation potentials for cytochromes and other heme proteins and compounds.

Moore and Williams (1977) have proposed that the differences can be attributed to a difference in the bond length between the methionine sulfur and the heme iron. From the x-ray structures thus far completed there are no obvious differences in bond length. However, the errors inherent in a protein crystal structure are large enough to leave room for uncertainty. Recently, EXAFS studies have shown no significant deviations in the axial ligand bond lengths for several cytochromes in either oxidation state (Korszun, Moffat, Frank and Cusanovich 1982).
Instead the latter authors propose it is the orientation of the histidine ring relative to the heme plane that is responsible for the control of oxidation potentials. This is based on the observation that a model compound with a relatively low oxidation potential has its imidazole ring only four degrees from being coplanar with the iron atom and two opposing pyrrole nitrogens (Mashiko, Reed, Haller, Kastner and Scheidt 1981). In contrast, high potential tuna cytochrome c has the histidine rotated by 45 degrees from coplanarity (Takano and Dickerson 1980). The current model of cytochrome c₅₅₅ has the histidine rotated 25 degrees. This is consistent with the proposal of Korszun et al. in the sense that the cytochrome c₅₅₅ oxidation potential is between the tuna cytochrome c and the model compound; however, a more quantitative prediction relating the amount of rotation necessary to alter the oxidation potential has not been made.

Other theories are more concerned with the overall nature of the heme environment. Kassner (1972 and 1973) proposes that the hydrophobicity of the heme pocket dictates the oxidation-reduction potential. This proposal is based on heme analogs in water/benzene solvent systems. Stellwagen (1979) claims that the extent of solvent exposure is a controlling factor in the oxidation potential. The comparison of cytochrome c₅₅₁ and c₅₅₅ ultimately will be a good model because of the very similar size and structure, but with a large difference, 145mV, in potential. Unfortunately, due to the ambiguity in the right side of the heme for the c₅₅₅ structure, any detailed calculation concerning solvent
exposure or hydrophobicity must await further structure refinement and resolution of the backbone folding.

The current model of electron transfer assumes that the transfer occurs at the exposed front edge of the heme (Salemme et al. 1973). The front side of the heme is the site of electron transfer both to and from the cytochrome. This model is based upon comparisons of the various crystal structures and a wide range of chemical modification and kinetic data. It has been observed that there is a ring of at least four positive charges from lysines surrounding the exposed edge of cytochromes c, c₂ and c₅₅₀. In c₅₅₅ there is structural homology for three of the four lysines; however, to the upper right of the molecule there is a serine where almost all other c-type cytochromes have a lysine. This absent lysine may explain why c₅₅₅ exhibits no reactivity with mitochondrial cytochrome c reductase, in contrast to the other bacterial cytochromes. The resolution of many of these questions depends upon the refinement of the current model and the structure determination of the reduced protein.
APPENDIX A

COMPUTER PROGRAMS USED IN THIS STUDY

This is a brief summary of the computer programs used in this study. Most programs are written for the CYBER 175 at the University of Arizona’s Computing Center. The location of the remaining programs is indicated in parentheses.

DATA REDUCTION PROGRAMS

DATRED (Diffractometer system): This program reads the data collection tape and applies the necessary corrections. If desired, an empirical absorption correction is applied using one or more reference reflections. Background corrections are assumed to be empirical. Lorentz and polarization corrections are also calculated and applied for each reflection.

Two magnetic tape files written by the data collection program are used as input. One file contains the reflections used for the absorption correction and the other file contains the data collection information. The user is prompted for additional input at the terminal.

Two disk files are written. The first, D9user.DA, is a binary file containing the Miller indices, $F^2$ and sigma. The second file, D7user.DA, is a listing of the standard reflections measured during data collection.
IBMOUT9 (Diffractometer system): Writes the disk file D9user.DA to a magnetic tape in a form that can be read by another computer. This program is the main link between the diffractometer and other computers.

DIFMR: Reads input reflections from diffractometer. (NOTE: Reflection file first must be decoded from the magnetic tape by the CYBER's system program FORM.) User specifies unit cell parameters and size of scale groups. Data is divided into scale groups based upon when a reflection was collected. The reflections are then sorted so that multiple observations of a reflection can be compared easily. The program then writes a MULTIREF file, which is compatible with the X-RAY system of programs.

SCALE1: This program reads a MULTIREF file and calculates scale factors for the scale groups to correct for crystal decay. User specifies magnitude and number of observations a reflection must have in order to be included in the calculation of scale factors. User may treat Friedel reflections as multiple observations. Scaling R-factors are calculated.

CRITIQUE: A MULTIREF file is examined, identifying one hundred observations with the largest discrepancies in four different categories. This program is useful in deciding to remove a particular observation of a reflection.

STACKUP: User specifies unreliable observations of a reflection and program will remove these observations from the MULTIREF tape. This
program is also used to combine two MULTIREF tapes if, for example, reflections from two different crystals must be combined.

EDITMR: MULTIREF tapes can be edited rapidly by removing observations with F less than a specified sigma ratio. Also individual observations may be deleted from data set.

REDUCE: After the MULTIREF data set has been scaled and edited, this program is used to write a SUBDATA tape, where all observations have been averaged to give a single value for the reflection. A reflection is coded unobserved if F is less than sigma. Data is analyzed for percentage observed and average size of F as a function of resolution. A Wilson plot is also calculated and plotted.

PHASE REFINEMENT AND FOURIER CALCULATION PROGRAMS

FINDKC: Reads a SUBDATA tape and scales two data sets on the tape together, using both a linear and an exponential scale factor. The program also calculates an R factor using the new scale factors. This is useful for scaling both derivative and calculated structure factor amplitudes to native data.

PHREF: A space group specific phase refinement program. Program reads and writes a SUBDATA tape. Program will calculate phases based upon input heavy atom parameters and then refine new heavy atom parameters using either MIR or SIR data. User may fix any or all parameters.
EDIT: Reflections may be deleted from a SUBDATA tape based upon either resolution, sigma ratio or magnitude of F.

PERCENT: A SUBDATA tape is read and the reflections are sorted into shells based upon resolution. Within each shell, reflections are then sorted by magnitude of F and only the largest percent are kept. The user specifies what percent of each shell to accept, the number of shells and the resolution range.

SHELL: This program reads a SUBDATA tape and calculates the average value of F for a given shell of reciprocal space. User specifies resolution range and number of shells to use.

FFTSET: Reads a SUBDATA tape and prepares an input file for the FFT. User specifies the resolution range, type of phases, combination of structure factor amplitudes and type of weighting. The centroid distribution as a function of resolution is also displayed.

HEXPAND: Reads a SUBDATA tape and generates a hemisphere of data for space group P6_2 for input to the FFT.

FFT (P21, P1, KERNAL): Space group specific Fast Fourier Transform programs. KERNAL, which contains the basic subroutines for the FFT, is used in conjunction with P21 for space group P2_12_12_1, and P1 for space group P1.

FASTCON: Produces a contour map on a line printer after reading a binary map produced by the FFT. User can specify up to eight contour levels.
FOUR: The original Fourier program in the X-RAY system. This program is still useful because electron density sections can be calculated along any axis.

CONCON: Converts a binary electron density map, produced by FOUR, into a format resembling the FFT binary map. This program will also generate a number map on a line printer reading a FOUR binary map.

FOSTAT: Reads a FFT format electron density map and calculates the distribution of electron density values. This program is useful in choosing contour levels and the solvent boundary.

CONTOUR: Reads a FFT produced map and generates a file containing instructions to operate the Calcomp plotter. User may specify scale factor, pen size and up to fourteen contour levels. Negative contours are drawn in red.

FOURIER INVERSION PROGRAMS

MAPGEN: Generates an electron density map of the entire unit cell using an asymmetric unit in space group P2₁2₁2₁.

CUT: Reads the map generated by MAPGEN to slice out the desired portion of real space. This program generates both a map in FFT binary format, which can be contoured, and an ASCII format map that is compatible with the MMSX graphics system.

RFAC: This program reads two FFT format reflection files, generally a set of observed structure factors and calculated structure
factors, and generates a new set of structure factors, either $F_{\text{obs}}$, $(2F_{\text{obs}} - F_{\text{calc}})$ or $(F_{\text{obs}} - F_{\text{calc}})$ using calculated phases. The program calculates the average difference in phase, a linear scale factor between data sets and the $R$ factor.

**SMOOTH:** Reads a FFT format map and applies a smoothing function to generate a new electron density map. User specifies solvent cutoff value, mid-point of smoothing function, minimum and maximum values of the map.

**LEVEL:** This program attempts to remove the heavy atom ripples around the heavy atom binding site. A linear weighting function, that is proportional to the distance from the heavy atom, is applied to all map grid points within a ten angstrom radius of the heavy atom.

**INVER (KERNEL):** Calculates structure factors by Fourier inverting an electron density map. The inversion is done in space group $P1$, so a complete unit cell must be present and an entire hemisphere of data is generated. The subroutines associated with the FFT, KERNEL, are also used.

**STRUCTURE REFINEMENT PROGRAMS**

**SCATIN:** Reads a SUBDATA tape and prepares a list of reflections to be used in the structure refinement. Reflections are sorted into shells of reciprocal space and the user can specify both the upper and lower fraction of each shell to use. User also specifies the resolution and the number of shells to use.
SCATT (DEC): Reads the ASCII file generated by SCATIN and creates a binary reflection file containing the reflection information and the scattering factors for each type of atom in the structure.

PROTIN (DEC): This program generates a binary file of distance restraints to be used in the refinement program. The user inputs a dictionary of standard amino acids, protein coordinates, and the weighting factors for various type of distance and bond angle restraints.

PROLSQ (DEC): In addition to the files generated by the programs SCATT and PROTIN, this program reads two other input files. One file contains weighting parameters concerning the relative distance restraints. The other file contains atomic shift information calculated from previous cycles of refinement. The program then calculates the least squares refinement using Hendrickson and Konnert's (1979) conjugate gradient algorithm. A new file of atomic shifts is generated. R factors are calculated for a subset of the reflections using different damping factors for this cycle of shifts, in order to give an indication of the optimal damping factor for the next cycle. A listing of atom shifts and differences between ideal and current distances is also produced.

APPLY (DEC): This program reads the atomic shift file generated by PROLSQ and applies the shifts, weighted by the appropriate damping factor, to the atomic coordinates to generate a new set of coordinates to be shipped to the CYBER.
ASU: Reads coordinates in Hendrickson and Konnert format and converts to standard ATOM format of the X-RAY system. If desired coordinates can be placed in the unique asymmetric unit of the cell.

ROTATION FUNCTION PROGRAMS

MAKDK2: Selects reflections from the observed diffraction pattern for use in the rotation function program. Reflections are sorted by magnitude and the user can vary how many large and small reflections are chosen.

MAINFC: Calculates a hemisphere of structure factors in space group P1 from a known structure, which should be centered at the origin of the unit cell. Input atom coordinates must be sorted according to atom type.

PACKI: Reads the structure factor file generated by MAINFC and packs reflections and their intensities using the subroutine PAC.

ROTMAT: This program reads the reflection files created by PACKI and MAKDK2 and calculates the value of the rotation function over the angles specified by the user. The results of this calculation are generally normalized to 100. It is possible to specify the scale constants, usually obtained from an earlier run, to ensure that all calculations are scaled to the same value.

PACKAB: Similar to PACKI, however, both the real and imaginary parts of the reflection are stored. This creates the input file for
TRANF reading the output from MAINFC.

TRANF: Performs the translation function calculation using the rotation angles determined from ROTMAT. This program generates structure factors that are then used as input to a Pl Fourier program.
REFERENCES


Kassner, R. J. (1973) J. Am. Chem. Soc. 95, 2674.


