

Copyright © by  
RUSSELL TIMKOVICH  
1974

THE STRUCTURE DETERMINATION OF M. DENITRIFICANS  
CYTOCHROME C<sub>550</sub>

Thesis by

Russell Timkovich

In Partial Fulfillment of the Requirements  
For the Degree of  
Doctor of Philosophy

California Institute of Technology

Pasadena, California

1975

(Submitted September 6, 1974)

## ACKNOWLEDGMENTS

I would like to thank my wife for her help and understanding, her sacrifices and patience, during a very busy four years. Professor A. Tulinsky of Michigan State University began my scientific training, and it was indeed an excellent beginning, by his advice and support. Professor R. E. Dickerson, my thesis advisor, has contributed invaluable guidance and inspiration to this work, to my education, and, perhaps most importantly, to my outlook on science. It is a joy and a pleasure to acknowledge a debt of gratitude to Professor E. Margoliash for the opportunity to work in his laboratory.

The research described in these covers would not have been possible without the assistance of George McLain and Dr. Lucile Smith of Dartmouth Medical School. During my pack-rat days, I relied heavily on equipment and facilities in the laboratories of Dr. R. Stroud and Professor L. Hood, for which I wish to express thanks. The author's graduate program was made possible through financial support from the National Science Foundation and the California Institute of Technology.

## ABSTRACT

The protein sequence and the three-dimensional folding structure for cytochrome  $c_{550}$  from Micrococcus denitrificans are reported. The tertiary structure resulted from an x-ray crystallographic determination. Standard protein chemistry techniques were applied to sequence all the fragment peptides from a tryptic digest of  $c_{550}$ . This information was combined with a 2.45 Å electron density map of the protein where identifying the tryptic fragments in the map led to the correct ordering of the peptides and the final complete sequence.

The crystallization of  $c_{550}$  is described. Three chemical derivatives of the protein were found,  $PtCl_4^{2-}$ ,  $UO_2^{2+}$ , and  $Pt(CN)_4^{2-}$ , and these aided in the isomorphous replacement solution to the crystallographic phase problem. The interpretation and refinement of the derivatives to produce an accurate protein map is discussed.

This bacterial  $c_{550}$  possesses a structure similar to the known cytochromes from eukaryotic sources and from a photosynthetic bacterium. The three are contrasted to conclude that for the cytochrome family, there exists a core structure of polypeptide folding. Current proposed mechanisms for the action of  $c$ -type cytochromes are rejected as being incompatible with the  $c_{550}$  structure. The view is advocated that the exposed heme edge in cytochrome may be the sole active site for the molecule.

TABLE OF CONTENTS

CHAPTER		PAGE
	ACKNOWLEDGMENTS	i
	ABSTRACT	ii
	LIST OF TABLES	v
	LIST OF FIGURES	vi
I.	INTRODUCTION	
	1. 1. General	1
	1. 2. Historical Background	2
	1. 3. The Cytochrome Family	3
	1. 4. <u>Micrococcus denitrificans</u> $\underline{c}_{550}$	5
	1. 5. References	10
II.	CRYSTAL STRUCTURE DETERMINATION	
	2. 1. Source of $\underline{c}_{550}$	12
	2. 2. Crystallization	13
	2. 3. Heavy Atom Derivatives	17
	2. 4. Data Collection	21
	2. 5. Data Reduction to Structure Factor Amplitudes	27
	2. 6. Solution of the Phase Problem	31
	2. 7. Refinement of the Heavy Atom Derivatives	49
	2. 8. Refinement Weighting Functions	69
	2. 9. Crystalline $\underline{c}_{550}$ as a Function of pH	75
	2. 10. Ammonium Selenate Binding to $\underline{c}_{550}$	77
	2. 11. References	80

III.	PRIMARY SEQUENCE OF $\underline{C}_{550}$	
3. 1.	Introduction	83
3. 2.	Materials and Methods	85
3. 3.	Results	89
3. 4.	Discussion	110
3. 5.	References	113
IV.	TERTIARY STRUCTURE OF $\underline{C}_{550}$	
4. 1.	Ordering the Tryptic Peptides	115
4. 2.	Molecular Structure – General	120
4. 3.	Molecular Structure – Details	127
4. 4.	Molecular Structure – Some Net Effects	134
4. 5.	Evolutionary Relationships	138
4. 6.	Current Proposed Mechanisms for the Reduction of Cytochrome $\underline{c}$	141
4. 7.	Alternate Mechanisms	142
4. 8.	References	152
	PROPOSITIONS	155

TABLE OF TABLES

TABLE	CONTENT	PAGE
1. 1.	Some Characteristics of Cytochromes	4
2. 1.	Molecular Weight Determination of $\underline{c}_{550}$ on Sephadex Gel	13
2. 2.	Crystal Forms of $\underline{c}_{550}$	16
2. 3.	Heavy Atom Derivatives	20
2. 4.	Collected Data	30
2. 5.	R-Factor Comparisons	32
2. 6.	Comparison of Centric and General Space Refinement	51
2. 7.	4 Å Parameters and Statistics	53
2. 8.	Refinement of Minor Sites	56
2. 9.	Final Parameters, Averaged Data	61
2. 10.	Final Parameters, Eight-Derivative Data	62
2. 11.	Final Refinement Statistics	63
2. 12.	Heavy Atom Binding Sites	70
2. 13.	Ammonium Selenate Binding Sites	79
3. 1.	Amino Acid Compositions	93
3. 2.	Amino Acid Composition of $\underline{c}_{550}$	100
3. 3.	Sequence of Tryptic Peptides	101
4. 1.	Assessment of Residues	117
4. 2.	$\underline{C}_{550}$ Sequence	118
4. 3.	Cytochrome Sequence Comparisons	128
4. 4.	Hydrogen Bond Comparisons	130
4. 5.	Bacteriophage Coat Proteins	133
4. 6.	Cytochrome Aromatic Ring Orientations	136
4. 7.	Sequence Similarity-Difference Matrix	140
4. 8.	Chemical Modifications of Cytochrome	147

TABLE OF FIGURES

FIGURE	CONTENT	PAGE
1. 1.	Comparative Reaction Kinetics	9
2. 1.	Crystal Habit of $\underline{c}_{550}$	18
2. 2.	Radial Distribution of Native Data	22
2. 3.	Centric $\text{PtCl}_4^{2-}$ Patterson Map	35
2. 4.	Patterson Harker Sections for $\text{PtCl}_4^{2-}$	36
2. 5.	Patterson Harker Sections for $\text{UO}_2^{2+}$	39
2. 6.	Patterson Harker Sections for $\text{Pt}(\text{CN})_4^{2-}$	42
2. 7.	Pseudo-Symmetry of $\text{PtCl}_4^{2-}$ Sites	47
2. 8.	Thermal Ellipsoids of the Heavy Atom Sites	57
2. 9.	Radial Distribution of the Figure of Merit	65
2. 10.	Heme Density Section, Three Derivative Phasing	66
2. 11.	Heme Density Section, Eight Derivative Phasing	68
2. 12.	Weighting Scheme Comparison	73
3. 1.	Purification of Tryptic Peptides	90
3. 2.	$\underline{C}_{550}$ Fingerprint of Peptides	92
4. 1.	$\alpha$ -Carbon Diagram of $\underline{c}_{550}$ , Front View	121
4. 2.	$\underline{C}_{550}$ , Side View	122
4. 3.	$\underline{C}_{550}$ , Top View	123
4. 4.	$\alpha$ -Carbon Comparison of $\underline{c}_{550}$ and Tuna $\underline{c}$	124
4. 5.	Schematic $\underline{c}_{550}$	125
4. 6.	Crystal Packing Diagram	132
4. 7.	Lysine Residues About the Heme Crevice	144
4. 8.	Heme Edge Binding of Transport Enzymes	150



## CHAPTER 1. INTRODUCTION

### 1.1. General

Imagine plotting a graph of biochemical information content versus the number of protein crystal structures which have been solved. If the information content was evaluated in terms of knowledge gained on the general principles of molecular architecture, the resultant curve might show a dramatic initial rise followed by a gradual turn toward a smaller slope. This might follow from the bursts of knowledge which the first few structures provided on the elements of protein folding—hydrophobic and entropic factors, salt bridges, hydrogen bonds, oil drop packing and so on. Later work has provided fresh examples without necessarily revealing astounding new principles. That statement is not quite accurate, for recent advances in the structural basis for allostery may give the hypothetical curve delta-function jumps.

On the other hand, if the information content were gauged in terms of the specific knowledge of how a protein or enzyme functions or what are the relations among a family of proteins, the curve might still appear steeply rising. Crystallographic results continue to be carefully compared with proposals for mechanisms of protein function, as the results continue to present the most stringent constraints against which models are judged. Structural studies on related proteins, such as the serine protease family, have revealed the common architectural features required to build a

catalytic active site. In a sense, the relevance of protein crystallography to biochemistry has shifted toward the solution of specific biological problems.

The motivating force for the research to be described in these pages is the idea that the second type curve, that of increasing specific knowledge, will continue. At issue here is the family of cytochrome c proteins. What is there to be learned about structural variations within the family? What are the limitations imposed by structure on the workings of the molecule?

## 1.2. Historical Background

The early stages of the work to be described here were set against a background of previous structural work on cytochrome c. The 2.8 Å crystal structure of horse cytochrome c had been submitted for publication in 1970 (Dickerson, et al., 1971), although the main results had been available since 1969. A structure determination was in active progress on the reduced form of the tuna protein with the species change brought about by the availability of x-ray quality crystals. The data on hand were provoking interest in the biochemical community on the mode of action of cytochrome and it was possible to detect an increase in the number of research papers appearing on the properties and action of cytochrome.

Already at this early date, there existed a definite promise for comparative structural biochemistry in the cytochrome family. In Dr. J. Kraut's laboratory at the University of California, San Diego, F. R. Salemme and co-workers were investigating the crystallographic

structure of a bacterial c-type cytochrome from Rhodospirillum rubrum. This molecule, called cytochrome c<sub>2</sub>, showed a typical c-type spectrum and molecular weight, but a redox potential some 60 mv higher and an isoelectric point some 4 pH units lower than the mitochondrial protein. The physiological role of c<sub>2</sub> was strikingly different as it participated in a photosynthetic electron transport chain, in contrast to the eukaryotic cytochrome in a respiration chain. Some sequence homology with mitochondrial cytochrome led to predictions concerning possible structural congruity (Dickerson, 1971), but it was not at all clear how similar the two molecular versions might be.

The structural analysis on c<sub>2</sub> was completed during the early stages of work on c<sub>550</sub> (Salemme, et al., 1973) and the results, which will be discussed in detail later, increased interest in c<sub>550</sub> for purposes of comparison. It began to look as if there would be a characteristic folding pattern within the cytochrome family. To recognize the pattern, a high resolution structure for three representative family members would be necessary.

### 1. 3. The Cytochrome Family

The ubiquity of c-type cytochromes is the ubiquity of electron transport chains. As such it extends to all known organisms with the exception of certain primitive bacteria which survive on a pure fermentative metabolism (e. g., Clostridia). The most widely known cytochrome role is in oxygen respiration as it occurs in the mitochondria of eukaryotes. As sequence studies show extensive

homologies among the eukaryotic cytochromes, and physiochemical characterizations reveal common properties, the class is taken as nearly identical in structure and function.

On the prokaryotic side of life, the uniformity which has been seen in the eukaryotes is obscured by a wealth of diversity. As a class, the prokaryotic cytochromes exhibit remarkable extremes in molecular weight, redox potential, spectra, isoelectric point, and, most significantly, biological function (see Table 1.1). It remains to be seen what structural variation can produce such a wide range of properties.

---

TABLE 1. 1

SOME CHARACTERISTICS OF CYTOCHROMES

Molecular weight	<u>P. aeruginosa</u> <u>c</u> <sub>551</sub> 82 residues	<u>M. denitrificans</u> <u>c</u> <sub>550</sub> 135 residues
Reduction potential	<u>Desulfovibrio</u> <u>desulfuricans</u> <u>c</u> <sub>3</sub> -205 mv	<u>Rhodospirillum</u> <u>rubrum</u> <u>c</u> <sub>2</sub> +320 mv
Isoelectric point	horse <u>c</u> 10.5	<u>M. denitrificans</u> <u>c</u> <sub>550</sub> 4.5
Spectra	<u>R. spheriodes</u> <u>cc'</u> $\alpha$ -band @ 548 nm	<u>R. palustris</u> <u>c</u> <sub>2</sub> $\alpha$ -band @ 555 nm
Function	horse <u>c</u> in O <sub>2</sub> respiration	<u>R. rubrum</u> <u>c</u> <sub>2</sub> in photosynthesis

---

As far as cytochrome lore goes, the situation is like a man from some previous century attempting to understand transistors. He may study some one transistor in some one radio in great detail

and still lack the knowledge of what is "transistor" and what is "transistor being in radio". By studying components of televisions or computers or what not, this level of understanding could be reached. Hence one looks at a variety of cytochromes c to see how structural variations can lead to the known diversity of function and physio-chemical properties.

#### 1.4. Micrococcus denitrificans c<sub>550</sub>

Granted that there is validity in looking at structure in a bacterial cytochrome, the next point is which one? Micrococcus denitrificans c<sub>550</sub> was an attractive candidate and a compelling reason was that its characterization, isolation, and purification to a high degree had been worked out by Scholes, McLain, and Smith (1971). Besides the practical factors, the possible contributions to the previously mentioned information content curve were considered.

An early realization of the cytochrome evolutionary implications was due to the thoughts and influence of Professor R. E. Dickerson. By comparing the known eukaryotic cytochrome structure with that of the prokaryotic c<sub>550</sub>, one could obtain some idea of the evolution which had taken place since the prokaryotic-eukaryotic divergence, an event dated some 2-3 billion years in the past (Barghoorn, 1971). With the forthcoming structure of photosynthetic c<sub>2</sub>, it would be possible to ask questions concerning the history of respiratory and photosynthetic transport proteins. It might be within reason to test ideas such as those of Gaffron (1962) which hold that photosynthesis

predates respiration and that components of the latter may be evolution products of the former. At issue was the thought that by comparing the two primitive bacterial cytochromes (photosynthetic  $c_2$  and respiratory  $c_{550}$ ) with the modern mitochondrial counterpart, some inkling of the structural folding of the ancestor protein might be gained.

To explain the differences in physiochemical characteristics between  $c_{550}$  and other cytochromes was another goal. At a molecular weight of circa 15,000,  $c_{550}$  differed markedly from  $c$  (12,500) and bacterial cytochromes such as *Pseudomonas aeruginosa*  $c_{551}$  (8,000). With an acid isoelectric point, 4.5, it contrasted with the strongly basic (10.5) mitochondrial protein. This was held to be a key point, for it has been well-documented that positive charges are involved in cytochrome recognition by oxidase and reductase (Smith and Minnaert, 1965; Davies, Smith and Wasserman, 1964). So even at this early date the possibility arose for identifying these charges as an island of positive lysines on a molecular surface of negative charge. Subtle spectral and redox potential differences are seen in comparing  $c_{550}$ ,  $c_2$ , and eukaryotic  $c$ . For example, the respective redox potentials range from +260 mv for  $c$  and  $c_{550}$  to +320 mv for  $c_2$ .

For the first few points, the body of this thesis can provide at least partial answers. For explaining the delicate adjustments in spectra and reduction potential, protein crystallography may be inadequate. Consider the crystallographically similar structures for

ferredoxin and the HIPIP iron-sulfur protein (Carter, et al., 1972) which differ in potential by 750 mv.

M. denitrificans is not a photosynthetic organism so  $c_{550}$  must differ in function from  $c_2$ . The protein has been implicated in the classical form of oxygen respiration and so it resembles eukaryotic  $c$  in that respect (Vernon, 1956). Are there any differences in the physiological function of  $c$  and  $c_{550}$ ? Some role has been ascribed to  $c_{550}$  in the nitrate reduction step of anaerobic respiration in M. denitrificans (Sato, 1956), but this function remains to be conclusively proved, in light of growing evidence for a  $b$ -type cytochrome mediating nitrate reduction in bacteria (Knook, Van t'Riet, and Planta, 1973). It has been more firmly established that  $c_{550}$  participates in the nitrite reduction step (Newton, 1969).

The most intriguing aspect of the comparative biochemistry lies in the realm of relative kinetics of electron transport. As purified and isolated components, the bacterial and eukaryotic cytochromes may be interchanged with their respective oxidases and reductases. As a general rule among the eukaryotic systems, if you take any eukaryotic cytochrome  $c$  (say from horse), it will react with any other eukaryotic oxidase (e. g., cow oxidase) or reductase at the same rate as it would with its own natural oxidase or reductase. When one starts to mix bacterial prokaryotic components with eukaryotic components, the general trend is that of very slow or no reaction (e. g., see Yamanaka and Okunuki, 1964).

Even reduced cytochrome  $c_{550}$ , reacting in a eukaryotic oxidase system, undergoes essentially no oxidation (2.6% specific activity taking eukaryotic cytochrome  $c$  as 100%). The same sort of experiment with eukaryotic reductase and oxidized  $c_{550}$  gives about 28% to 50% activity. The true surprise comes when the converse experiment is performed, reacting eukaryotic cytochrome  $c$  with M. denitrificans oxidase and reductase. With oxidase the specific rate is 410% and with reductase 220%! (For references see Smith, Newton, and Scholes, 1966; the data actually quoted here were the result of recent experiments to update the first reference given and will be the subject of a future L. Smith publication.)

The general picture is that the eukaryotic oxidase and reductase are quite specific in their acceptance of transport partners, suggesting a binding or recognition site for cytochrome  $c$ . By inference it can be assumed that bacterial oxidase and reductase have a recognition site tailored along the same lines. But here, in the process of evolving binding to bacterial  $c_{550}$ , the site has somehow become compatible with an unnatural redox component – eukaryotic cytochrome  $c$ . A schematic way of thinking about this is shown in Figure 1.1 for oxidase kinetics which show the most striking differences.

If the structure of  $c_{550}$  were known, it could be contrasted with the already known eukaryotic structure taken as a substrate analog of  $c_{550}$ . By looking for similarities and differences some idea of the oxidase/reductase binding site may be gained. In point of fact, the bacterial system electron transport protein interface would



# HYPOTHETICAL REACTIONS

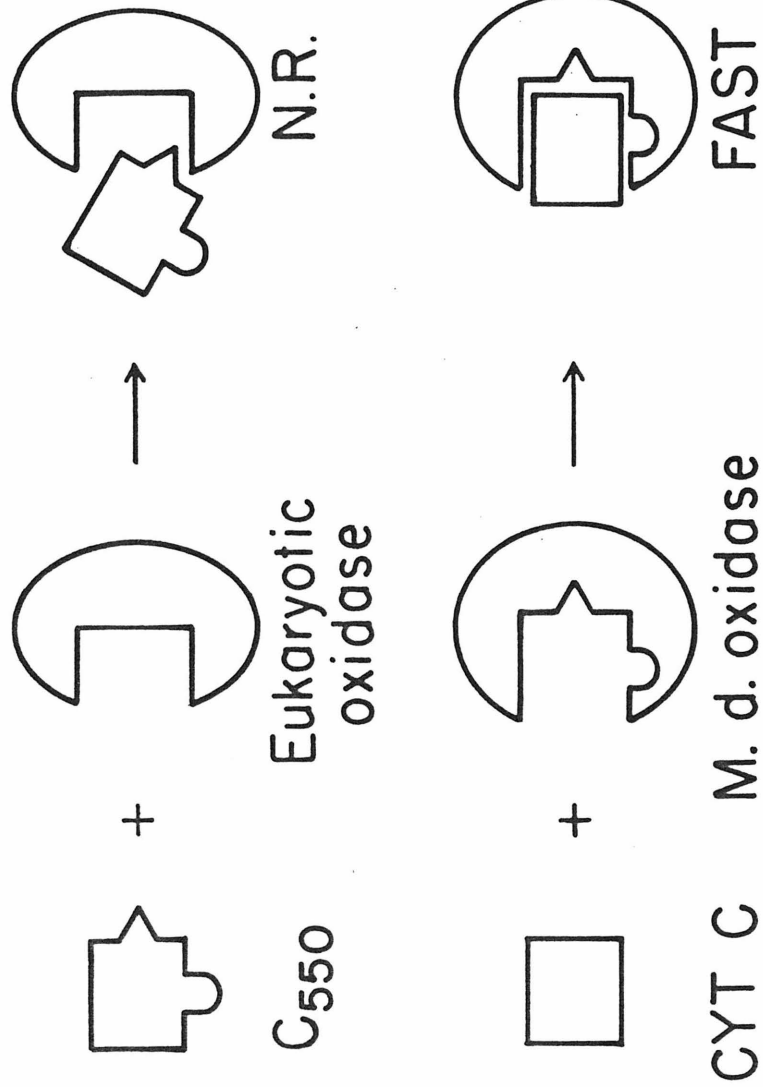


Figure 1.1. Schematic of the comparative reaction kinetics.

be sought at the level of cytochrome c by looking at kinetically acceptable structural variations between the substrates c and c<sub>550</sub>. It should be a simple step to say that the principles which govern the interface for the bacterial system also apply to the eukaryotic system, but with some changes in specificity. Broadly speaking then, the c<sub>550</sub> structure may say something about how electrons are passed in transport chains.

The next chapter deals with the x-ray crystallographic details for the solution of the crystal structure. Chapter 3 recounts the protein chemistry applied to sequence the cytochrome. Chapter 4 returns to biochemistry and discusses the results and implications of the complete structure.

#### 1.5. References

- Barghoorn, E. S., (1971). Scientific American, 225, no. 5, 30.
- Carter, C. W., Jr., Kraut, J., Freer, S. T., Alden, R. A., Sieker, L. C., Adman, E., and Jensen, L. H., (1972). Proc. Nat. Acad. Sci., 69, 3526.
- Davies, H. C., Smith, L., and Wasserman, A. R., (1964). Biochem. Biophys. Acta, 85, 238.
- Dickerson, R. E., (1970). J. Mol. Biol., 57, 1.
- Dickerson, R. E., Takano, T., Eisenberg, D., Kallai, O. B., Samson, L., Cooper, A., and Margoliash, E., (1971). J. Biol. Chem., 246, 1511.
- Gaffron, H., (1962). In Horizons in Biochemistry, p. 59-89, Academic Press, New York.
- Knook, D. L., Van t'Riet, J., and Planta, R. J., (1973). Biochim. Biophys. Acta, 292, 237.
- Newton, N., (1969). Biochim. Biophys. Acta, 185, 316.

- Salemme, F.R., Freer, S.T., Xuong, Ng. H., Alden, R.A. and Kraut, J., (1973). J. Biol. Chem., 248, 3910.
- Sato, R., (1956). In Inorganic Nitrogen Metabolism, eds. W.D. McElroy and B. Glass, Johns Hopkins University Press, Baltimore.
- Scholes, P.B., McLain, G., and Smith, L., (1971). Biochemistry, 10, 2072.
- Smith, L., and Minnaert, K., (1965). Biochim. Biophys. Acta, 105, 1.
- Smith, L., Newton, N., and Scholes, P., (1966). In Hemes and Hemo-proteins, B. Chance, R.W. Estabrook, and T. Yonetani, eds., Academic Press, New York.
- Vernon, L.P., (1956). J. Biol. Chem., 222, 1035.
- Yamanaka, T., and Okunuki, K., (1964). J. Biol. Chem., 239, 1813.

## CHAPTER 2. CRYSTAL STRUCTURE DETERMINATION

### 2.1. Source of c<sub>550</sub>

A small scale preparation of the purified cytochrome was attempted at Caltech according to the procedure of Scholes (Scholes, McLain, and Smith, 1971). The yield, less than 1 mg per 20 l of M. denitrificans culture, was so low that large fermentation batches (300 l) were subsequently necessary. The cultivation of the organism and subsequent extraction and purification of the cytochrome were performed in the laboratory of Dr. Lucile Smith, Dartmouth Medical School. The author is indebted to Dr. Smith and George McLain for gifts of the protein, which were used without further purification.

The small scale Caltech preparation confirmed qualitatively the spectral properties reported previously (Scholes, McLain, and Smith, 1971). An anomaly of the protein was noted in agreement with unpublished data of McLain and Smith. In molecular weight determination by gel filtration on Sephadex G-75 (Andrews, 1964), the protein shows a spuriously high molecular weight, indicative of a dimer (Smith data) or a tetramer (Table 2.1). Smith and McLain have analyzed their preparation by sedimentation techniques and have found the monomeric weight exclusively (unpublished results). Thus the aberration appears to be due to cytochrome-gel interaction and not due to the actual presence of polymeric forms, as has been observed for mammalian cytochromes (Margoliash and Schejter, 1966).

---

TABLE 2.1

GEL FILTRATION DATA

Protein	Molecular Weight	V(elution)-V(void)
Hemoglobin (marker)	67,000	7.0 ml
Candida krusei cyt <u>c</u> (marker)	12,500	49.4 ml
<u>c</u> <sub>550</sub> preparation	~60,000	7.4 ml

---

2.2. Crystallization

As a preliminary to the x-ray work, it was necessary to grow large protein crystals with sufficient internal order to produce x-ray scattering. Special crystallization experiments were designed to conform to the constraints imposed by the small amount of material, 20 mgs, available initially. Variations on a technique first introduced by Zeppezauer were used exclusively (Zeppezauer, Ekland, and Zeppezauer, 1968). A small bore glass capillary tube (length 30 mm, O.D. 6.5 mm, I.D. 1 mm) was fitted at one end with a small square of dialysis tubing, held over the opening by an O-ring cut from Tygon tubing with an inner diameter just smaller than the outer diameter of the tube. This cell was filled from a syringe with a small volume, 0.01 ml, of an appropriate protein solution, and sealed with parafilm covering the other end. The entire cell was then placed into solutions designed to slowly precipitate, i. e., crystallize, the protein. Best results were obtained when the cell was placed membrane side up in an upright

test tube of precipitant. A simple rationale may exist for the observation. Since all precipitant solutions were of higher density than the initial protein solution, the dense precipitant diffused across the membrane, and gravity helped to mix it with the dilute protein solution, thereby gradually increasing the concentration of precipitant throughout the cell. With the membrane down, a layering of light protein solution on heavy precipitant would lead to a high local concentration and rapid precipitation rather than slow crystallization.

Each cell was treated as a single crystallization experiment and, since each experiment required so little protein, extensive sets of conditions were explored. The nature of the precipitant varied from organic solvents to ammonium sulfate solutions. Among the latter trials, variable parameters consisted of protein concentration, salt concentration, pH, and temperature. Other experiments also explored the effect of buffers and salts added, in addition to the ammonium sulfate.

A general strategy was adopted for these trials. A small number of cells were employed with all conditions constant but one, which was varied over a large range. For example, at constant pH, temperature, and protein concentration, the concentration of ammonium sulfate was varied from 10% to 100% in intervals of 10%. The results of these trials were examined by microscope for any indication of crystallinity. Around the values of the varied parameter giving the most promising results, further trials were attempted on a finer scale. By permuting all parameters in this manner, conditions suitable for the growth of x-ray quality crystals were found.

Optimum conditions for crystallization are given in the following procedure. The stock protein, as a precipitate in ammonium sulfate, was collected by gentle centrifugation and redissolved in 20% saturated ammonium sulfate at pH 7.5 to give a 2% (w/v) protein solution, oxidized form. Microdiffusion cells containing this were placed in 95% saturated ammonium sulfate adjusted to pH 7.5 with concentrated aqueous ammonia, and containing 1.0 M NaCl. Crystal growth was complete in one week at room temperature, care being taken to avoid disturbing the cells during this time.

The requirement for NaCl was reminiscent of the crystallization of horse heart cytochrome c (Margoliash and Walasek, 1967). However, when the above procedure was duplicated with CsCl replacing the NaCl, crystals giving a diffraction pattern identical to the NaCl grown crystals were obtained. When NaI replaced NaCl only an oily, amorphous precipitate was formed. Thus it is the Cl<sup>-</sup> which is necessary for crystallization.

A variant crystal form, type II, was grown if the above procedure was followed, but with the addition of maleate at a concentration of 0.02 M and pH 6.0. This result was attributable solely to the maleate, for ammonium sulfate solution adjusted to pH 6.0 still produced type I crystals.

Precession photographs were taken with crystals sealed in capillary tubes, with a small amount of mother liquor to preserve humidity. Crystallographic parameters observed are listed in Table 2.2. Measurement of crystal density was by the flotation method (Stout and Jensen, 1968). For c<sub>550</sub>, advantage was taken of the fact that for

both crystal forms, the density was only slightly less than that of the precipitant solution. Saturated CsCl was added to a mixture of crystals in precipitant until the crystals attained neutral buoyancy, and then the solution was measured pycnometrically. Using this method, the reproducibility for Type I crystals was  $\pm 0.003$  g/ml. The density quoted for Type II crystals reflects difficulty in measurement due to the small number and size of these crystals. The resolutions quoted were visual estimations from photographic films of short exposure.

---

TABLE 2.2  
CRYSTAL FORMS OF  
MICROCOCCUS DENITRIFICANS CYTOCHROME C

	Type I	Type II
Class	Orthorhombic	Tetragonal
Space group	$P2_12_12_1$	$P4_1, P4_3$
Cell dimensions	$42.7 \times 82.1 \times 31.6 \text{ \AA}$	$37.5 \times 37.5 \times 96.7 \text{ \AA}$
Cell volume	$110,799 \text{ \AA}^3$	$135,722 \text{ \AA}^3$
Molecules/asymmetric unit	1	1
Crystal density	1.275 g/ml.	1.3 g/ml.
Mass (%) of protein in the unit cell	70%	56%
Maximum resolution of observable reflections	2.7 $\text{\AA}$	2.7 $\text{\AA}$

---

Type I crystals, oxidized cytochrome, were used for all subsequent work. This choice was based on the presence of three centrosymmetric projections in space group  $P2_12_12_1$ , which are of



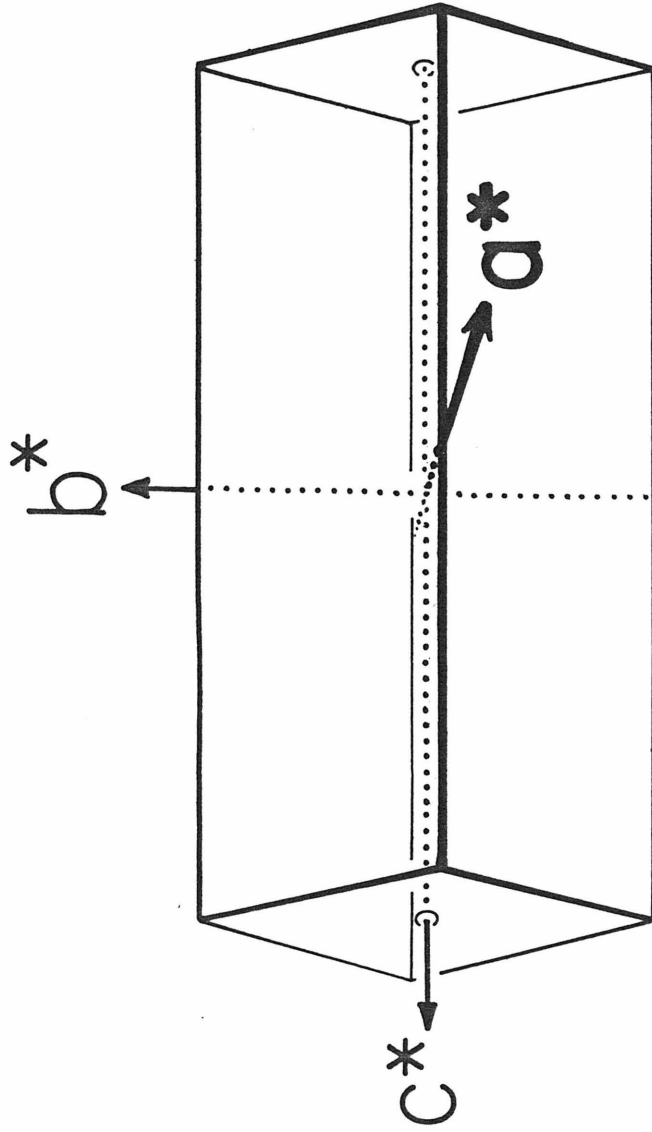
value in solving the phase problem, and the extreme stability of this crystal form to x-ray irradiation. One crystal was exposed for more than 250 hours to nickel-filtered copper radiation with no photographically observable decrease in intensities.

Figure 2.1 shows an idealized crystal morphology, Type I, in relation to the crystallographic axes. This crystal habit was rarely attained, since growth against the inner capillary walls gave highly distorted crystals. Usable crystals possessed dimensions in the range  $(0.02 \rightarrow 0.10 \text{ mm}) \times (0.05 \rightarrow 0.25 \text{ mm}) \times (0.2 \rightarrow 0.5 \text{ mm})$ .

### 2.3. Heavy Atom Derivatives

For M. denitrificans  $c_{550}$ , a de novo solution of the structure was desired and hence an isomorphous replacement solution to the crystallographic structure was sought (Phillips, 1966). It was realized that an alternative approach existed, that of the rotation function (for such an application, see Greer, 1971), but since the final  $c_{550}$  structure was to be compared with eukaryotic  $c$ , it seemed best to follow as independent a crystal analysis as was possible.

The survey for derivatives followed a general experimental scheme. An appropriate derivatizing reagent at high concentration was added to solutions of protein precipitant, 95% saturated ammonium sulfate, pH 7.5, 1.0 M in NaCl. Several crystals (5-10) were added to 0.5 ml of this solution and allowed to stand. At various time intervals, from one day to one month, crystals were removed and centric zone precession photographs taken, to compare the resultant diffraction patterns with the native pattern. Many compounds



## CRYSTAL HABIT

Figure 2.1. An idealized  $C_{550}$  crystal.

produced no changes, even in saturated solutions, or at the other extreme destroyed the internal order of the crystal. More hopeful candidates were examined as a function of concentration and soaking time, in an attempt (1) to produce maximum intensity changes consistent with isomorphous cell parameters, (2) to preserve the native "mosaic spread" (James, 1965), and (3) to preserve high resolution scattering data.

Conditions for derivative production and the properties of the modified crystals are given in Table 2.3. It will be noted that there were only three chemically distinct derivatives. Variations of concentration led to derivatives differing mainly in the number of sites of binding per protein molecule and the occupancies of the sites. Reference to the refinement section will illustrate that even under identical conditions, the binding produced different effective atomic numbers at the sites. This may have been due to a certain inherent lack of reproducibility in the liquid-solid reactions as has been frequently noted (Tulinsky, et al., 1973). Alternatively, the differences in calculated occupancies may reflect lack of precision in the measurements or a certain degree of insensitivity and indeterminacy in the refinement procedure.

TABLE 2. 3

HEAVY ATOM DERIVATIVES

Derivative	Concentration	Soaking Time	Cell Parameters (Å)		
			a	b	c
$K_2PtCl_4$	1. 3 mM	4 days	42. 81	82. 45	31. 64
$(NH_4)_2Pt(CN)_4$	5. 6 mM	7 days	42. 90	82. 36	31. 11
$(NH_4)_2Pt(CN)_4$	11. 2 mM	7 days	42. 80	82. 30	31. 17
$UO_2(NO_3)_2$	saturated	3 days	42. 76	82. 76	31. 62
$UO_2(NO_3)_2$	saturated (see text)	8 days	42. 67	82. 03	31. 65
Native	-	-	42. 70	82. 17	31. 56

For all compounds reported, binding was complete by the indicated time, and no further reaction was observed up to one month later. The two uranyl derivatives produced from saturated uranyl nitrate arose from a side reaction peculiar to the uranyl ion. Initially soluble in mother liquor, the uranyl ion precipitated from concentrated ammonium sulfate solutions, if the pH were above approximately 5. Although the product was not characterized, it may have arisen from the replacement of an aquo-coordination sphere with sulfate to form the sparingly soluble  $(UO_2)_2(SO_4)_2 \cdot 7H_2O$ .

The difference in preparation of the two forms of uranyl protein was based on the assumption that even at low concentration of uranyl nitrate the precipitation reaction proceeded and the product would not bind to the protein. A solution of 5 mg of  $UO_2(NO_3)_2 \cdot 6H_2O$  in 1 ml of precipitant buffer (95% ammonium sulfate, pH 7. 5, 1 M NaCl) was allowed to stand for 24 hours. The resultant precipitate was

removed by centrifugation and 0.5 ml of the supernatant used for derivatizing as described previously. For the low occupancy form, simple standing for the necessary time affected reaction. For the high occupancy form, every two days for a period of eight days, fresh supernatant was prepared and it replaced the old soaking media. If protein crystals were placed in the uranyl solution before the precipitate had formed, the crystals accelerated the formation of the uranyl solid and in turn became coated by a fine layer of the yellow precipitate. While an alchemist's delight, this produced powder diffraction lines which obscured the protein diffraction measurements.

#### 2.4. Data Collection

The bulk of the diffraction data was collected in two stages. The first included several redundant measurements of native crystals and all the derivatives to 4.0 Å resolution. This approach, with the resolution cut-off being dictated by consideration of a convenient number of reflections to process, was adopted so that the known chemical derivatives could be analyzed as to interpretability before exhaustive data collection was initiated. Once a derivative had been successfully interpreted, data were recollected, including remeasurement of the 4.0 Å shell, to provide a complete set from a single crystal.

The resolution limit of 2.45 Å (or a Bragg  $2\theta$  angle of  $36.00^\circ$ ) was imposed by the inherent scattering from the protein crystals, as shown in Figure 2.2. This graph, illustrating the behavior of the mean structure factor amplitude as a function of data resolution, reveals

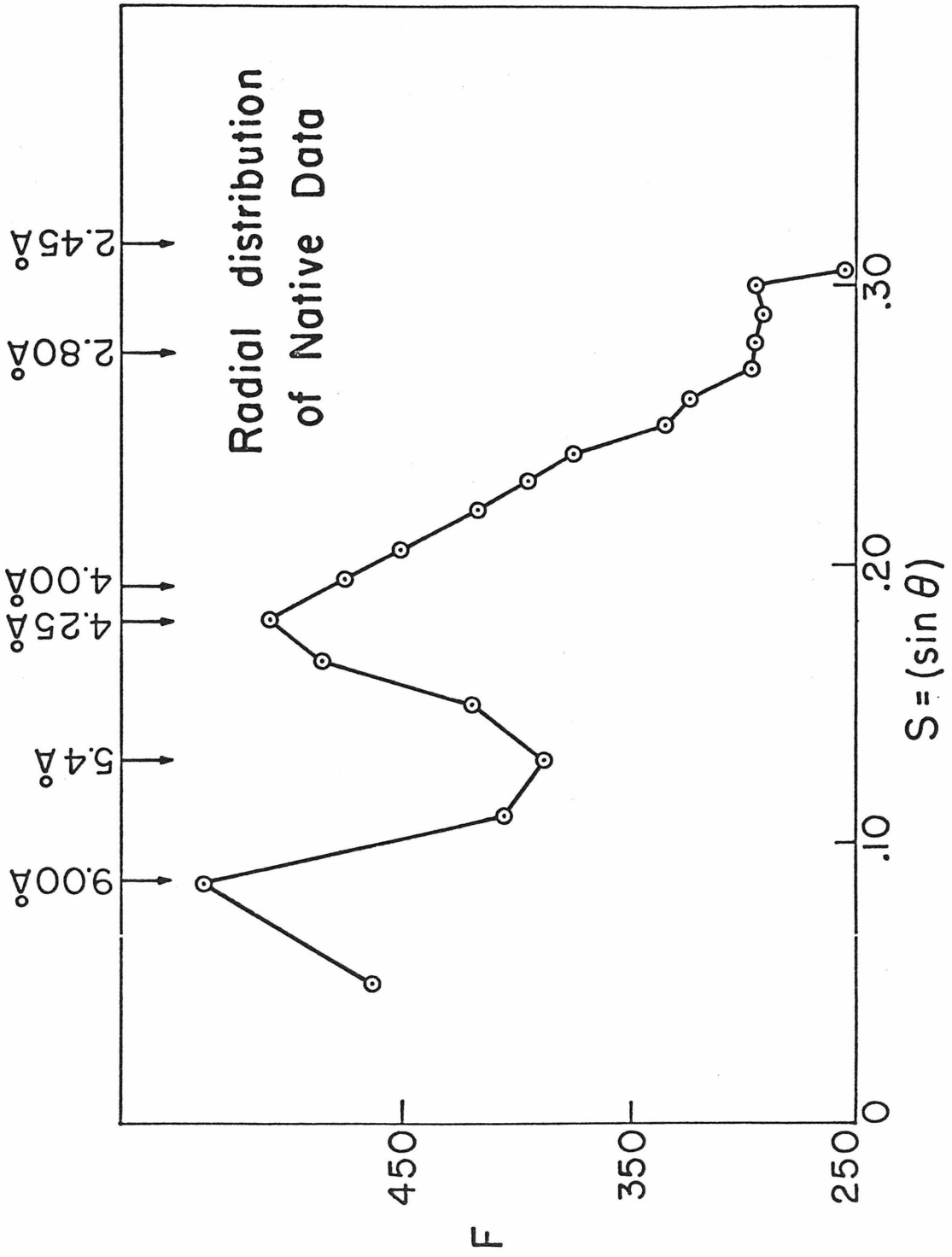


Figure 2. 2.

a rapid diminution beginning just beyond 2.8 Å. This was somewhat unexpected, since crystals of comparable size, containing protein of comparable molecular weight and even less mass percent protein per unit cell, have been reported as scattering to a much higher resolution (Wyckoff, et al., 1970; Blake, et al., 1965). Anticipating some comments from the structure results, the discrepancy may be explained by noting that certain parts of the molecule appear highly disordered between two or more possible configurations. Thus the number of scattering centers ordered to within the limits of the data resolution, relative to the total mass per cell, falls off as the resolution increases.

Unless otherwise noted the following comments are equally applicable to the 4 Å and the 2.45 Å data sets.

Precautions taken prior to intensity measurements improved accuracy. Crystals intended for diffractometry were carefully chosen to be single, without cracks or spurs, and of the largest available size. Heavy atom derivatives were prepared as noted previously, and at the end of the optimal soaking time, all were mounted in capillary tubes at the same time, thus insuring equal exposure time to the derivatizing agent.

The capillary tubes employed, of standard lithium borate glass (1 mm diameter), were pretreated by cleansing in fuming nitric acid and by coating with a water-repellant film. ("Desicote", Beckman Instruments, applied per manufacturer's directions.) This process enabled excess solvent to be removed from the crystal and capillary,

thereby lessening background x-ray scatter and improving signal-to-noise features.

Crystals were sealed in the capillaries with a drop of mother liquor to preserve humidity. They were invariably mounted with the  $c^*$  axis parallel to the spindle direction (azimuthal axis) of the diffractometer, since in this orientation the shape of the crystals most closely approximated a cylinder leading to a minimization of absorption effects.

Reflection intensities were measured on a General Electric quarter chi-circle XRD-490 diffractometer equipped with a graphite monochromator and controlled by a PDP-8 computer. The controlling programs were modifications by Dr. J. N. Brown of the standard GE control system, and are discussed in context below.

Preliminaries to a production mode data collection included a survey of the diffracted peak intensities as a function of the machine  $\omega$ -angle (equals the  $\theta$  Bragg angle). This revealed that incorporation of heavy atom derivatives led to the appearance of asymmetry and a broadening of the peak profiles, suggesting some disorder. For the native crystals and the  $\text{PtCl}_4^{2-}$  derivatives, peak width at half-peak height was circa  $0.13^\circ$  in  $\omega$ , relatively independent of  $2\theta$  over the angular range of  $2.45 \text{ \AA}$  resolution data, while for the  $\text{Pt}(\text{CN})_4^{2-}$  and  $\text{UO}_2^{2+}$  derivatives, this number could be as high as  $0.45^\circ$ . This necessitated an independent choice of peak scan widths for each crystal observed.

Monochromator adjustments (monochromator diffracting plane perpendicular to  $2\theta:\theta$  plane of the diffractometer) and incident beam



collimation (1.0 mm) were chosen to bathe the crystal in a beam of uniform cross-sectional intensity. The detector collimator, located 20 cm from the crystal, consisted of a system of 1.0 mm and 2.0 mm pinholes, chosen to minimize reception of background scatter while containing the entire diffracted ray.

Unit cell parameters were determined from a least squares fit of calculated and observed Bragg angles according to the  $\pm 2\theta$  method of Arndt and Willis (1966). The theoretical least squares standard deviations (Lipson and Cochran, 1966) were:  $\sigma(a) = 0.02 \text{ \AA}$ ,  $\sigma(b) = 0.03 \text{ \AA}$ ,  $\sigma(c) = 0.02 \text{ \AA}$ ; but a more realistic estimate is given by the statistical unbiased standard deviations of parameters from independent native crystals:  $\sigma(a) = 0.04 \text{ \AA}$ ,  $\sigma(b) = 0.05 \text{ \AA}$ ,  $\sigma(c) = 0.02 \text{ \AA}$ .

The intensities of  $(0, 0, \ell)$  reflections were measured at  $\chi = 90^\circ$  as a function of the azimuthal angle for future use in absorption correction. While the dependence of intensity upon azimuthal angle may in general change as a function of the index  $\ell$ , for  $\underline{c}_{550}$  crystals all  $(0, 0, \ell)$  reflections showed the same absorption profile and hence the most intense reflection,  $(0, 0, 4)$ , was employed to construct the empirical absorption behavior.

Integrated intensities were obtained by an omega step-scan procedure, always scanning through the peak from low to high omega to avoid gear backlash. The scan width was chosen as the average full peak width plus  $0.30^\circ$  to provide a margin for wider than average peaks and peak center mis-setting. Steps were generally  $0.02^\circ$  to  $0.04^\circ$ , with the larger steps used with the broadest peaks only. The average time spent scanning a reflection was 96 seconds. Background

was measured by stationary counting both at the beginning and end of the scan, each for one-half of the total scan time, with the total background taken as a sum of these counts.

To understand the order of data collection (the path taken by the instrument through reciprocal space) two cases must be considered. All the  $4\text{\AA}$  data sets were collected in a straightforward manner, with little attention paid to optimizing the task. In general, the reflections  $(h, k, \ell)$  were obtained by holding  $h$  and  $k$  constant and progressing through reflections of increasing  $\ell$  up to the two-theta limit imposed by  $4\text{\AA}$  resolution. Then  $k$  and finally the  $\ell$  index were permuted, leading to a scheme of collection consisting of nested loops of reflections with  $\ell$  changing most rapidly,  $h$  least.

For the high resolution data, the decision was made not to collect anomalous pairs, in spite of their potential usefulness (Matthews, 1966a, b; North, 1965), so as to complete data collection within the decay life-time of a single crystal. It was, however, necessary to collect sufficient re-measurements of some reflections during the crystal life, so as to ascertain the course and extent of decay. The data were therefore collected in segments according to the following scheme:

1. A "before" data zone of  $(h, k, 0)$  reflections from  $\infty \rightarrow 2.45\text{\AA}$  with  $h, k$  obeying the relationship  $h + k = 3n$ ,  $n = \text{integer}$ .
2. A high resolution shell consisting of reflections between  $2.45\text{\AA}$  and  $2.80\text{\AA}$  resolution.
3. An intermediate shell,  $2.80\text{\AA}$  to  $4.0\text{\AA}$ .
4. A completion shell,  $4.0\text{\AA} \rightarrow \infty$ .

5. An "after" zone consisting of the same reflections as in step 1.
6. Throughout steps 1-5 the same three, high two-theta reflections were monitored every 75 data points.

A comparison of data observed during steps 1 and 5 established the over-all extent of crystal x-ray damage and any systematic pattern in the destruction process. The  $3n$  condition limited the amount of reflections to a convenient number while preserving, for this space group, a random sampling of reciprocal space. The most likely victims of irradiation damage are high order data, describing molecular fine structure. Therefore, these were collected in the beginning before x-ray exposure of any duration in the indicated shell order.

#### 2.5. Data Reduction to Structure Factor Amplitudes

To the net intensities, standard Lorentz and polarization corrections were applied (Levy and Ellison, 1960), as were approximate absorption corrections (North, Philips, and Matthews, 1968) obtained from the empirical absorption curve previously described. The transmission coefficient for x-rays in the least absorbing direction divided by that in the most absorbing gave an average ratio of 1.3.

The stated method of monitoring x-ray decay revealed that under the given experimental choice of scanning times, irradiation damage was minimal. The classical R factor, defined as

$$R \equiv \frac{\sum |F_1 - F_2|}{\sum |F_1|}$$

between "before" and "after" data zones averaged 0.03 over all sets and was never higher than 0.036. Perusal of these zones failed to

disclose any systematic effect, and the check reflections confirmed the stability. In consideration of this behavior, no time decay corrections were applied to the data.

Reflections processed were assigned a standard deviation in percentage units defined by

$$\sigma = 100\% \cdot \frac{3}{2} \cdot \left( I_{\text{obs}} + \text{BKG} + (0.02 \cdot I_{\text{obs}})^2 \right)^{\frac{1}{2}} / I_{\text{net}}$$

where

$$I_{\text{obs}} = \text{total scan counts}$$

$$\text{BKG} = \text{total background counts}$$

$$I_{\text{net}} = |I_{\text{obs}} - \text{BKG}|$$

The square root dependency upon  $I_{\text{obs}} + \text{BKG}$  follows from counting statistics and represents a lower estimate to the actual standard deviation, which may be higher due to (a) machine error, (b) data collection error (in crystal orientation, cell parameters, crystal slippage within the capillary), and (c) data reduction error (in absorption and decay). The term  $(0.02 \cdot I_{\text{obs}})^2$  attempts to alleviate somewhat any underestimation and dampen the steep descent of  $\sigma$  with increasing  $I_{\text{obs}}$ . The value of 0.02 was used in a similar scheme in the solution of the trypsin structure (R. Stroud, personal communication), and was employed without any test of validity. The factor of 3/2 arises from a consideration of the statistical unbiased standard deviation for the monitor reflections, which were measured 50 to 100 times per data collection. This 3/2 factor was required to equate the calculated  $\sigma$  with that statistically determined.

A somewhat critical use must be made of these standard deviations, for there is obviously a point below which net intensity counts are meaningless and a reflection is accordingly unobserved, within experimental limits. To arrive at some measure of the lower limit of detectable intensity, the space group systematic absences were measured redundantly and processed as normal observations. Most gave, of course,  $I_{\text{net}} \sim 0$  to within  $\pm 50$  counts while the highest, the limits of random fluctuation, were 100 to 150 net counts. These spurious reflections, at the level of total scan intensity and background measured, convert to percent standard deviations in the range of 33% to 50%. Hence 33% was taken as a realistic indication of an observable data point, with such criteria leading to 75% to 85% observables for all data sets. In the parlance of crystallography this scheme amounts to retaining all reflections  $3\sigma$  above noise.

Table 2.4 lists the data sets which were instrumental in the solution of the crystal structure, a subset of the data actually taken. The ammonium selenate, the high pH, and the low pH data pertain to some ancillary experiments on the protein structure and will be discussed later. The data sets listed amount to 32,400 reflection measurements.

TABLE 2. 4  
SOME DATA SETS COLLECTED

I. Native Data	pH	Resolution
14N	7.5	2.8 Å
17N	7.5	2.8 Å ((h, 0, l) only)
20N	7.5	2.45 Å
Ac5.1	5.1	2.8 Å ((h, 0, l) only)
NH <sub>4</sub> OH	8.9	2.8 Å ((h, k, 0) and (0, k, l))
II. Derivative Data	Reagent	Resolution
ASe2a	(NH <sub>4</sub> ) <sub>2</sub> SeO <sub>4</sub>	2.8 Å ((h, k, 0) and (0, k, l))
C17a	PtCl <sub>4</sub> <sup>2-</sup>	4.0 Å
C19a	PtCl <sub>4</sub> <sup>2-</sup>	2.45 Å
C19b	PtCl <sub>4</sub> <sup>2-</sup>	2.45 Å
CN7c	Pt(CN) <sub>4</sub> <sup>2-</sup> (high concentration)	4.0 Å
CN9a	Pt(CN) <sub>4</sub> <sup>2-</sup> (low concentration)	2.45 Å
UN11a	(UO <sub>2</sub> )(NO <sub>3</sub> ) <sub>2</sub> (low concentration)	4.0 Å
UN12IIa	(UO <sub>2</sub> )(NO <sub>3</sub> ) <sub>2</sub> (high concentration)	2.45 Å

Table 2.5 reports some representative R-factors, thus assessing the precision of the measurements. From Table 2.5 it can be seen that data set pairs (14N, 20N), (CN10b, CN9a), and (C17a, C19b) agreed well within the pair. Thus they were averaged together to give a master native set and a set of "average" data for the  $\text{PtCl}_4^{2-}$  and  $\text{Pt}(\text{CN})_4^{2-}$  derivatives. As will be seen, the various uranyl sets have extensively differing substitution and hence in this case averaging was not possible.

## 2.6. Solution of the Phase Problem

Data sets collected from various crystals were placed on a common scale by the method of Dickerson, et al. (1967a). An additional multiplicative scale factor was calculated and applied to the heavy atom derivative sets according to the method of Kraut, et al., (1962) to account for the theoretical increase of scattering matter per unit cell.

The solution to the phase problem was essentially completed with the  $4\text{\AA}$  data. As will be seen, the higher resolution data were processed in a similar manner, but this mainly confirmed the original solution, improved the phases somewhat, and revealed features of the heavy atom derivatives lacking from the earlier work. So, until noted, the ensuing discussion refers solely to the  $4\text{\AA}$  data.

Two  $4\text{\AA}$  sets of data were collected from crystals modified with  $\text{Pt}(\text{CN})_4^{2-}$  at two different concentrations. Raw intensity differences indicated somewhat different substitution between the two, and so they were treated as independent derivatives. Later results will indicate

---

TABLE 2.5  
SOME R-FACTOR COMPARISONS

Type	Sets	R-factor*
Redundant zones, same crystal	14N vs. 14N	0.02
	20N vs. 20N	0.03
	UN12IIa vs. UN12IIa	0.04
	UN11a vs. UN11a	0.04
	ASe2a vs. ASe2a	0.03
Equivalent data, different crystals	14N vs. 17N	0.03
	14N vs. Ac5.1	0.04
	14N vs. NH <sub>4</sub> OH	0.05
	14N vs. 20N	0.04
	CN10b vs. CN9a	0.04
	C17a vs. C19b	0.05

---

\*  

$$R \equiv \frac{\sum |F_1 - F_2|}{\sum |F_1|}$$
, the summation over all comparable terms in the set.

---



that although the sites of attachment were the same, the occupancies differed.

For each of the derivatives, centric and full three-dimensional Patterson syntheses were calculated using difference coefficients (Blow, 1958),

$$(\Delta F)^2 = (|F_{PH}| - |F_P|)^2$$

where  $F_{PH}$ ,  $F_P$  refer to the derivative and native structure factors.

Considerable experimentation was conducted in an effort to improve the quality, i. e., feature-to-noise ratio, of these syntheses. For instance, various low order terms were omitted from the calculation under the assumption that they might arise from general salt or solvent differences. Artificial temperature factors of the form  $e^{-Bs^2}$  were applied to eliminate series termination effects or lack of isomorphism in the high orders. While manipulations led to changes in peak shapes and heights, it was not clear that any improvement in signal-to-noise ratio was obtained and certainly the maps were not easier to interpret. Such editing was abandoned and subsequent discussions refer to standard syntheses.

Comparisons between centric  $(\Delta F)^2$  maps and those for the same derivative in three-dimensions revealed a common theme — three-dimensional maps were easier to interpret and showed a higher ratio of peak height to general noise. Figure 2.3 shows an  $(h, k, 0)$  centric difference Patterson synthesis calculated with centric reflections from the complete data set used to calculate the map shown in Figures 2.4.1 through 2.4.3. Figure 2.3 has been contoured so that

the projected Harker peaks, designated by "H", were on the same scale as the Harker peaks in the three-dimensional map. The centric map contained general Patterson cross-vectors and noise peaks projected onto its limited area, hence adding background and obscuring interpretation. It also represented a Fourier summation over 150 terms, compared to 1100 in three-dimensions, making it more susceptible to random error in a few terms.

From Patterson syntheses a single site was found for the  $\text{PtCl}_4^{2-}$  and  $\text{UO}_2^{2+}$  derivatives and three sites for each  $\text{Pt}(\text{CN})_4^{2-}$  derivative. Figures 2.4, 2.5 and 2.6 present Harker sections from these maps. Only one  $\text{Pt}(\text{CN})_4^{2-}$ -map is shown, since the other was virtually identical. In each case the three mutually perpendicular sections are contoured at equal, arbitrary intervals with zero and negative contours omitted. For a single site of substitution in  $\text{P}2_12_12_1$ , peaks are expected at (1)  $\frac{1}{2}, \frac{1}{2}-2y, 2z$ ; (2)  $2x, \frac{1}{2}, \frac{1}{2}-2z$ ; (3)  $\frac{1}{2}-2x, 2y, \frac{1}{2}$ . The identification of these vectors is indicated by the letter H for the single sites in  $\text{PtCl}_4^{2-}$  and  $\text{UO}_2^{2+}$  and by the set  $\text{H}_1, \text{H}_2, \text{H}_3$  (decreasing occupancy) for the multiple site  $\text{Pt}(\text{CN})_4^{2-}$  derivative.

Many strong features in the  $\text{Pt}(\text{CN})_4^{2-}$ -map were unexplained by the solution indicated. They may be definitively ruled out as random experimental error, for their heights and positions reappear in all the  $(\Delta F)^2$  maps calculated from independently measured data sets. They may arise in part from the relatively large cell changes of  $\text{Pt}(\text{CN})_4^{2-}$  crystals versus native crystals. Further, the large amount of substitution could have caused local protein deformations and lack of isomorphism.

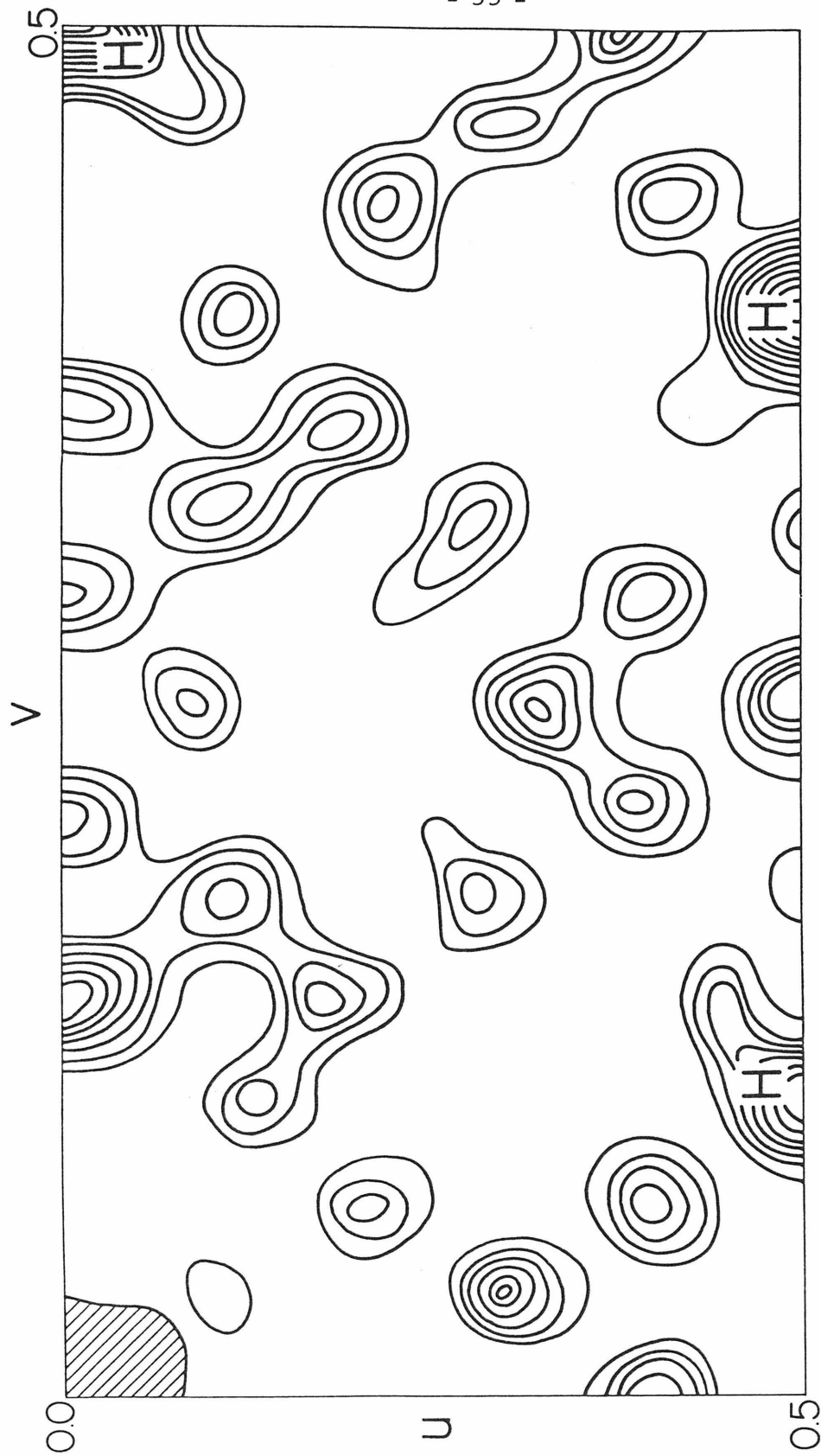


Figure 2.3.  $(\Delta F)^2$  Patterson synthesis of  $(h, k, 0)$   $\text{PtCl}_4^{2-}$  data.

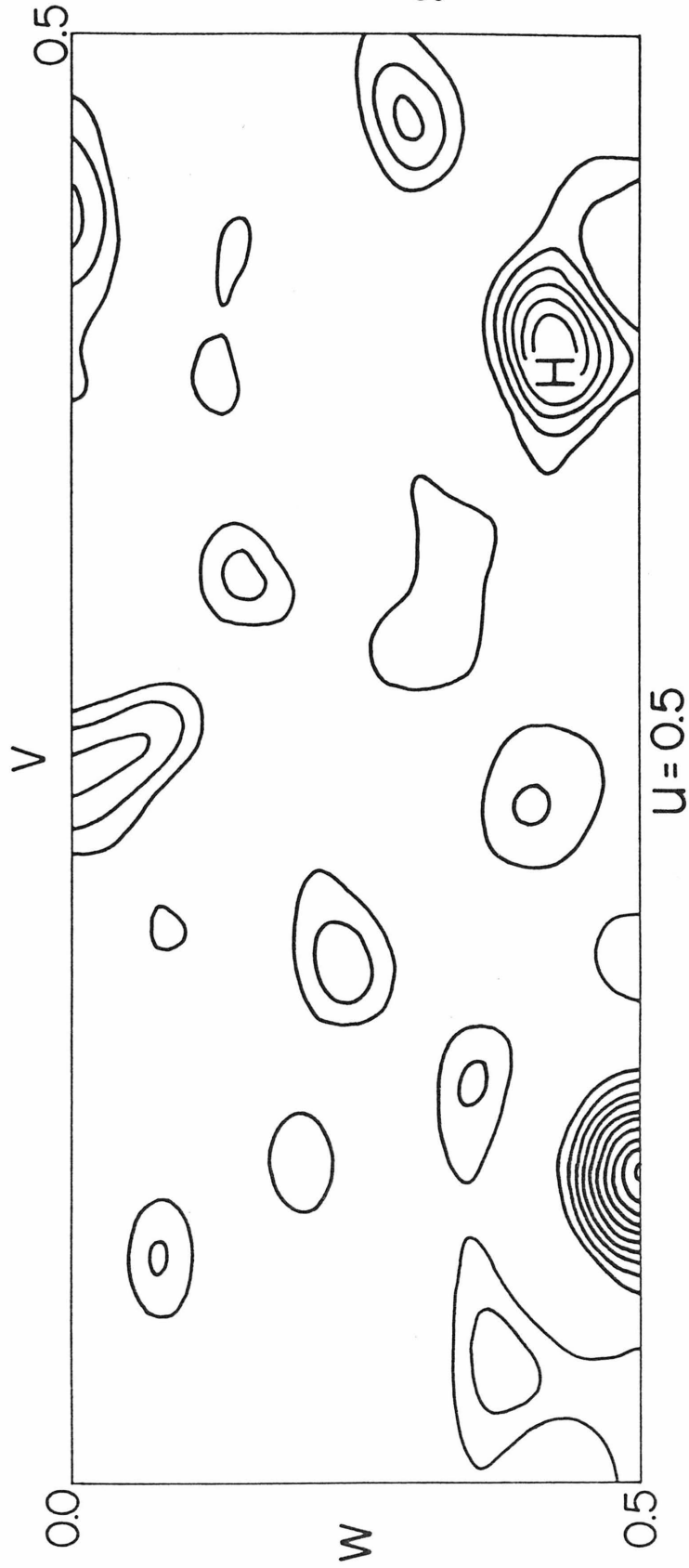


Figure 2.4.1. Harker section for  $(\Delta F)^2$  synthesis based on three-dimensional  $\text{PtCl}_4^{2-}$  data.

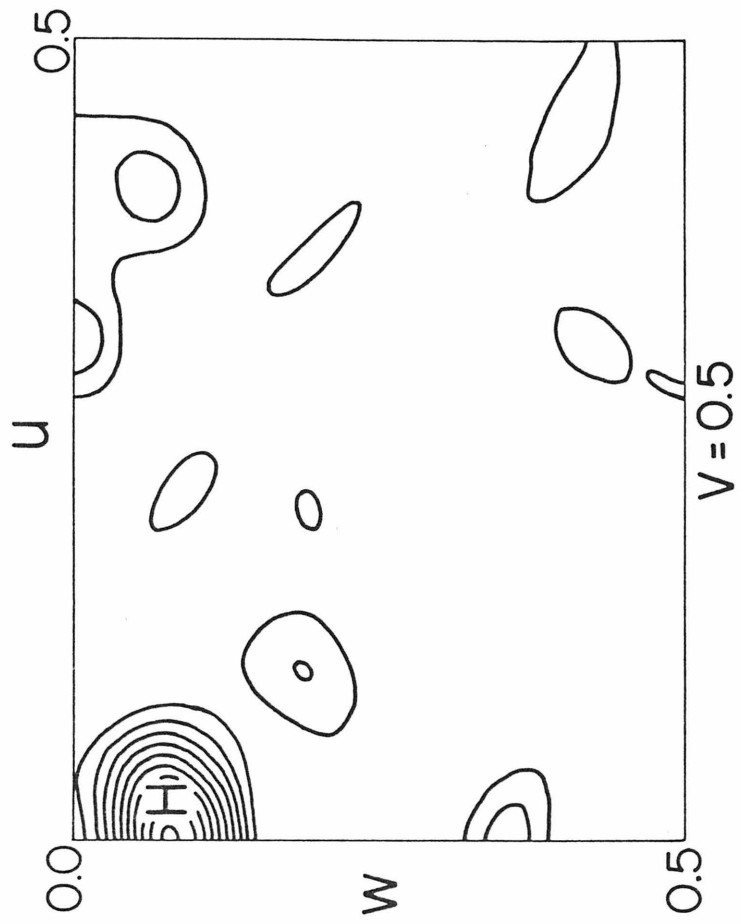


Figure 2.4.2. As in 2.4.1.

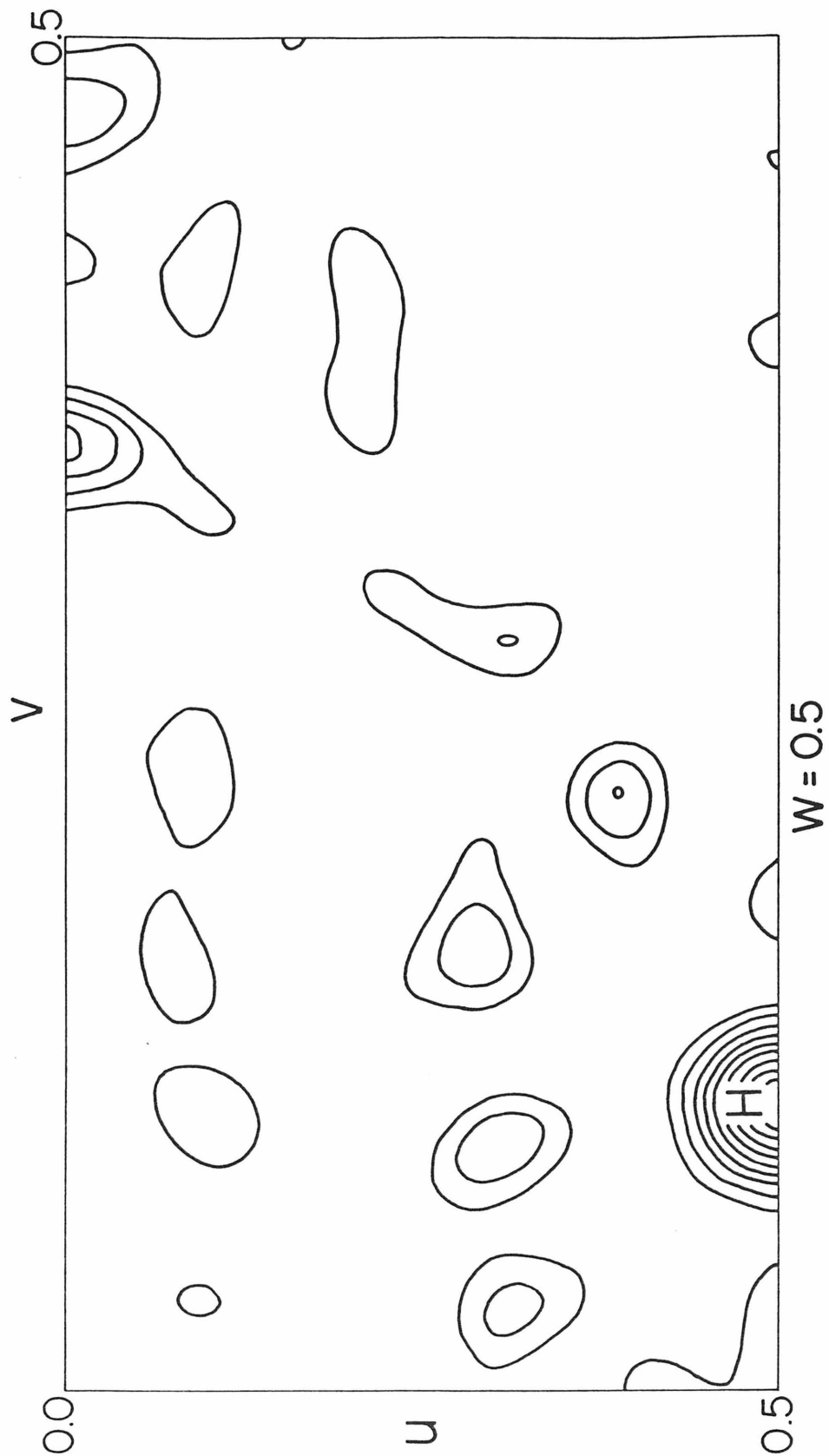


Figure 2.4.3. As in 2.4.1.

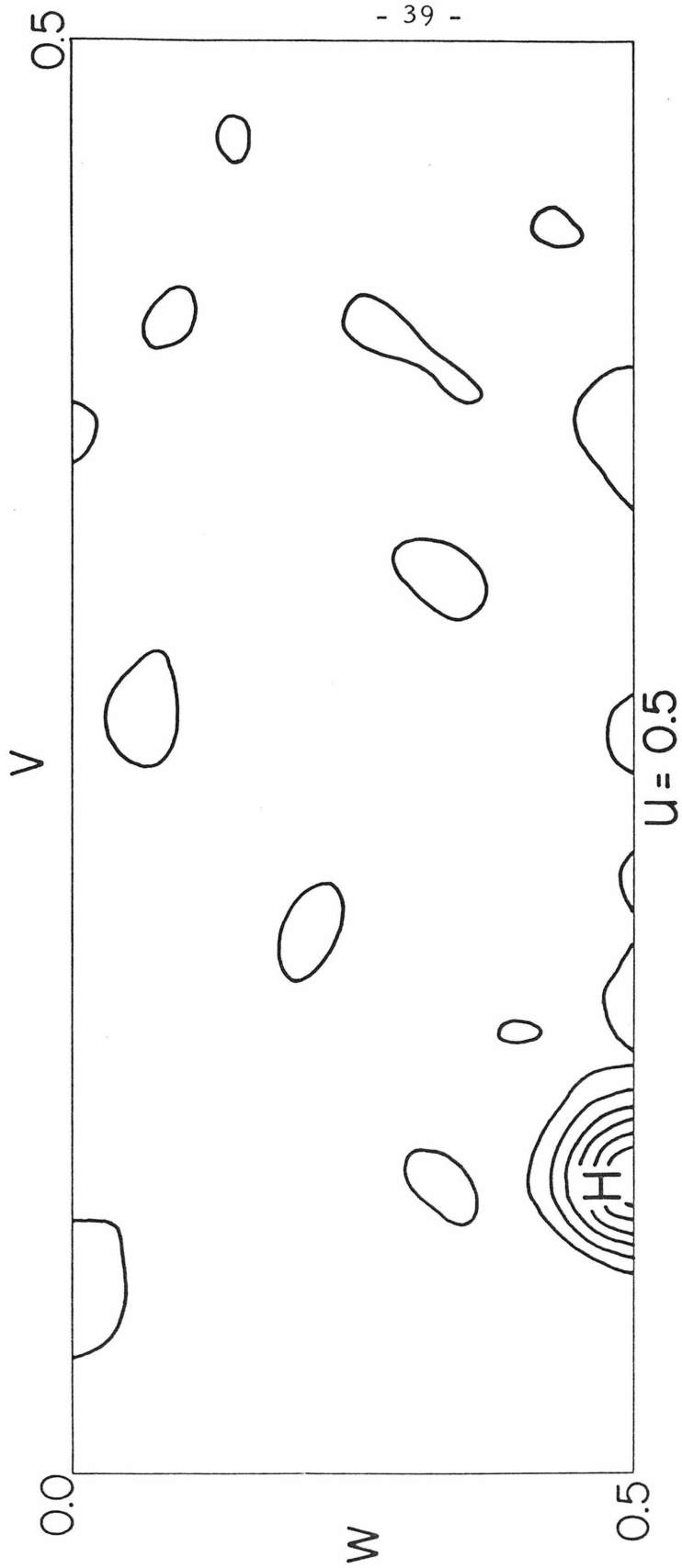


Figure 2.5.1. Harker section for  $(\Delta F)^2$  Patterson synthesis based on three-dimensional  $UO_2^{2+}$  data.

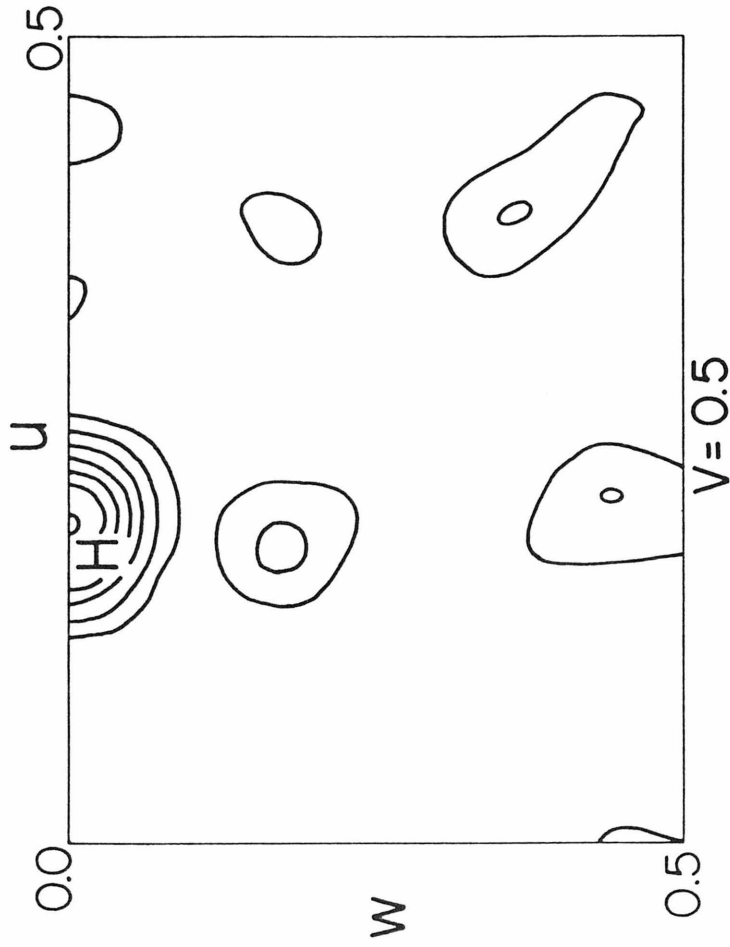


Figure 2.5.2. As in 2.5.1.



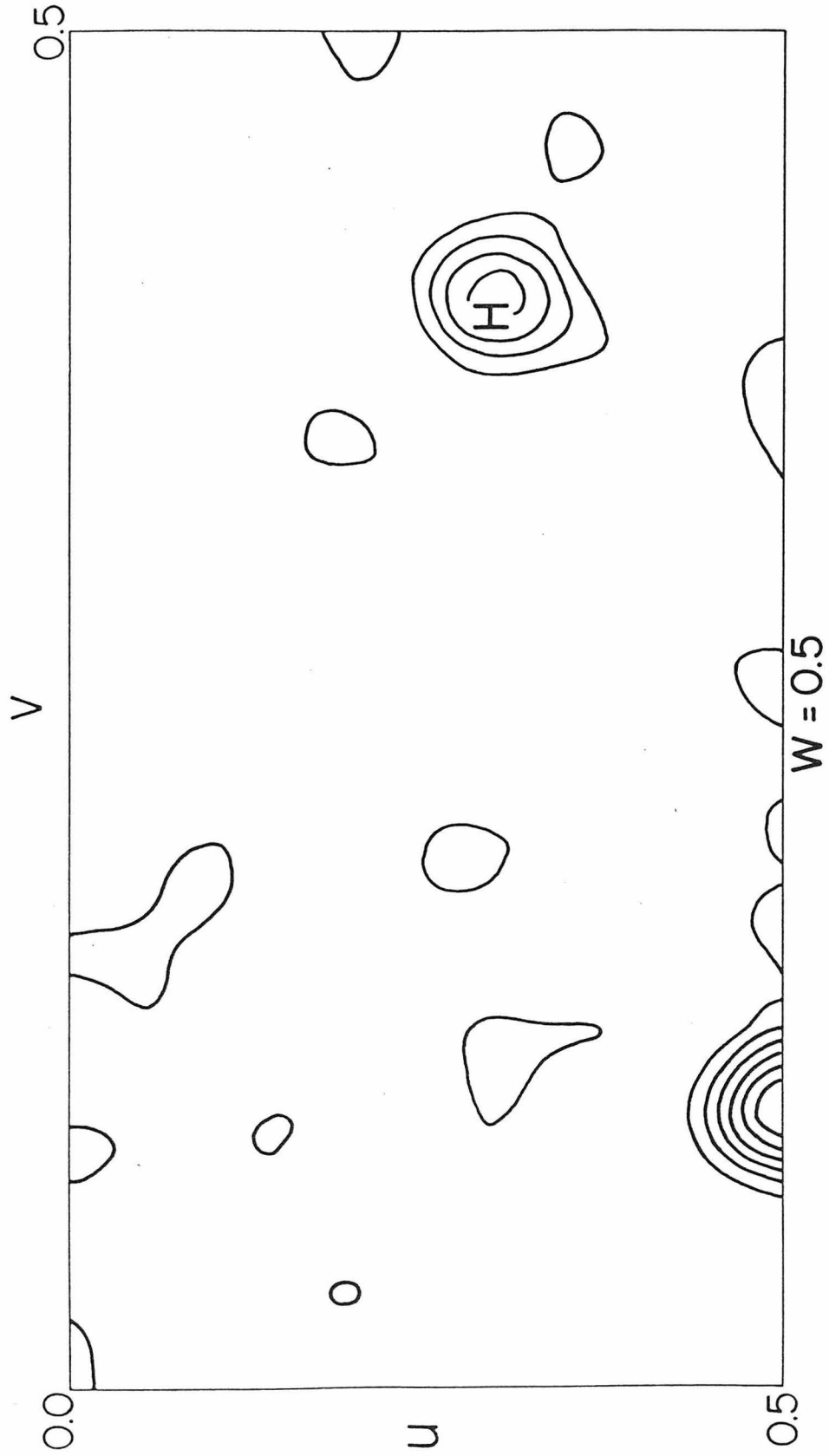


Figure 2.5.3. As in 2.5.1.

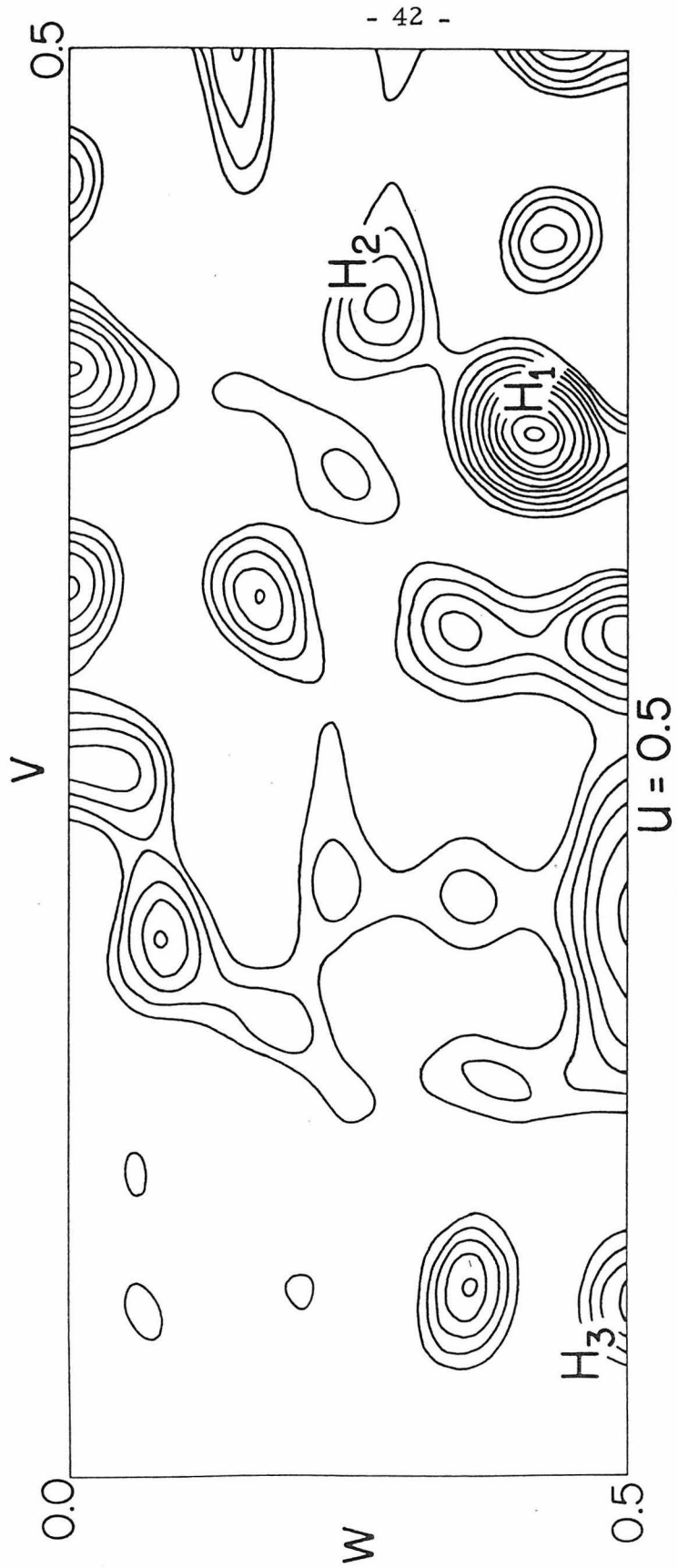


Figure 2.6.1. Harker section for  $(\Delta F)^2$  Patterson synthesis based on three-dimensional  $\text{Pt}(\text{CN})_4^{2-}$  data.

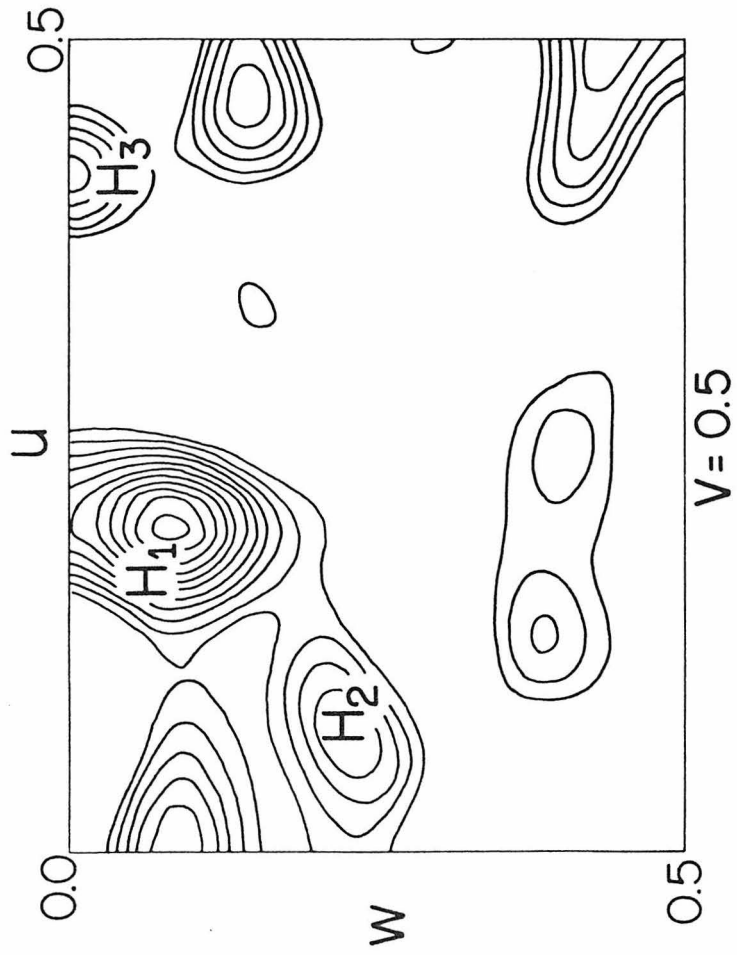


Figure 2.6.2. As in 2.6.1.

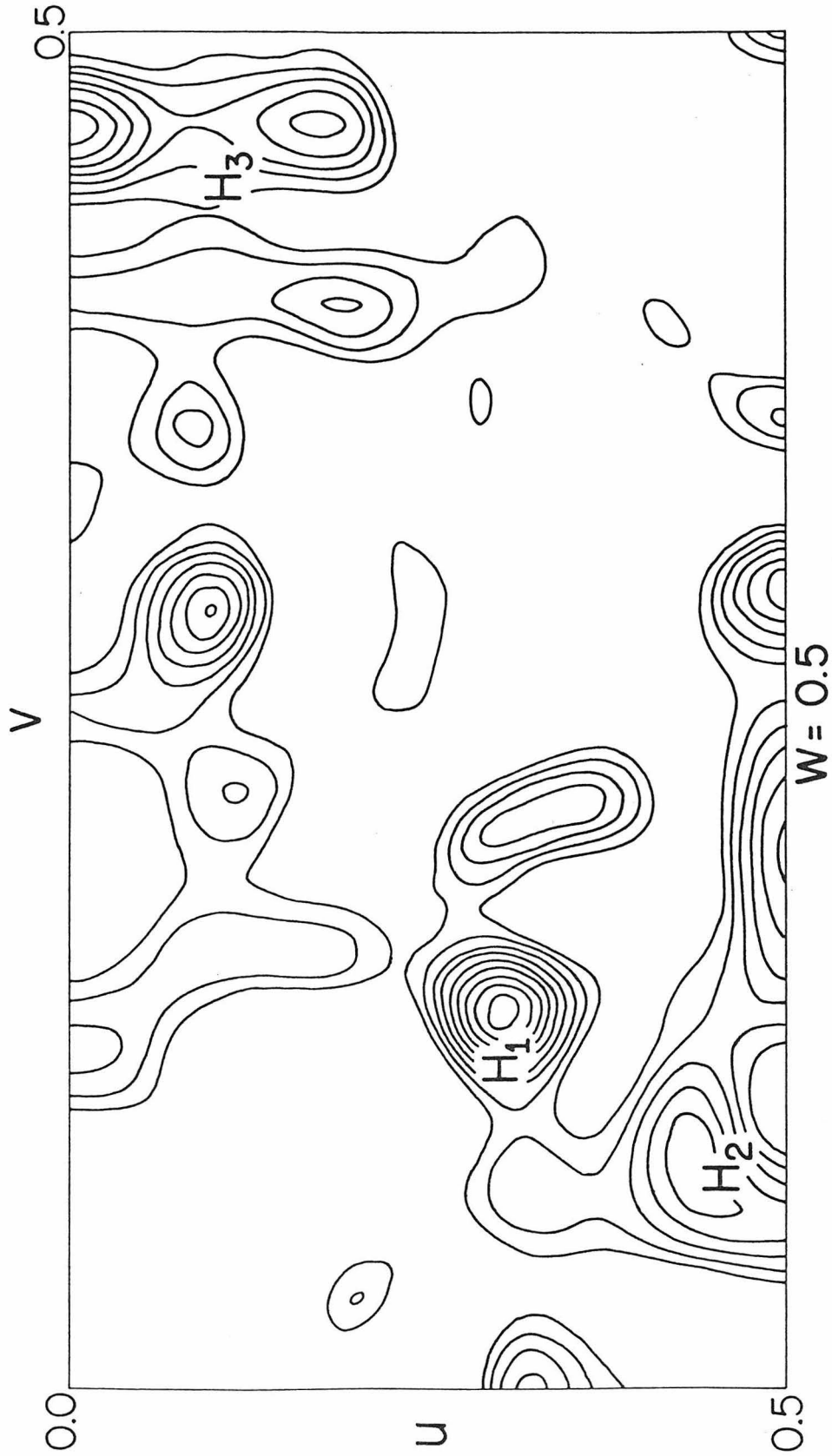


Figure 2.6.3. As in 2.6.1.

Up to this time, Wilson-type statistics (Wilson, 1942) had not been used to place the protein data on an absolute scale, because the method is unsuitable for protein crystals where the atomic scattering contribution of bound and interstitial solvent cannot be known a priori, and therefore cannot be included into a summation of squared atomic scattering factors over the crystal cell. Instead, the single site  $\text{PtCl}_4^{2-}$  derivative was used to scale approximately native and platinum data by determining the constant required to match observed  $|\Delta F|$  terms with terms from a centric difference structure factor calculation based on a site of 78 electrons. The atomic number of platinum was used rather than the sum of electrons in  $\text{PtCl}_4^{2-}$ , because there was evidence that the chloride ions had been replaced by interstitial solvent. This will be further substantiated in the discussion of anisotropic temperature effects.

Multiple solutions for atomic co-ordinates are possible from the  $\text{PtCl}_4^{2-}$  Harker vectors, and selection of one such set established a crystallographic origin for future work and fixed an enantiomorph. This arbitrarily selected "handedness" led to a calculated structure containing L-amino acids and right-handed helices and thus was fortuitously correct. The remaining task was to determine atomic co-ordinates for the other derivatives relative to the same origin.

First the single  $\text{PtCl}_4^{2-}$  site was refined over centric data for several cycles according to the method of Dickerson, Weinzierl, and Palmer (1968). These authors advocate using atomic scattering factors of the form  $Z e^{-Bs^2}$ , where Z is an effective occupancy in

electrons and B is a pseudo-temperature factor to include actual temperature-disorder effects and the scattering incoherence with increasing  $s = \sin \theta$ . Compared to Thomas-Fermi scattering centers, such a form, for typical values of B, was found to agree to better than 1.6% over the resolution range of the entire protein data. Initial values were obtained (Dickerson, et al., 1967a) and refinement adjusted atomic co-ordinates and occupancy. Attempts to refine B at this  $4\text{\AA}$  stage led to unrealistic and wildly fluctuating values, so for this and all other  $4\text{\AA}$  refinement, the initial value was held fixed. It is suggested that the scatter of values  $\langle \Delta F \rangle / \langle Z e^{-Bs^2} \rangle$  (averaged in zones of  $s^2$ ) about a line of zero slope is too great for the narrow resolution range to define B adequately.

For centric reflections with phase angles constrained to 0 or  $\pi$  (also  $\pi/2$  or  $3\pi/2$  in  $P2_12_12_1$ ), the ready knowledge of the magnitudes,  $|F_P|$  and  $|F_{PH}|$ , and just the phase of the heavy atom contribution allow unambiguous determination of the native protein phase. Therefore, the refined single  $\text{PtCl}_4^{2-}$  site permitted calculation of a set of native centric phases, which were in turn combined with appropriate terms  $\Delta F = (|F_{PH}| - |F_P|)$  from the  $\text{Pt}(\text{CN})_4^{2-}$  and  $\text{UO}_2^{2+}$  data to produce a series of Fourier density syntheses of these heavy atom sites, of necessity referred to the  $\text{PtCl}_4^{2-}$  origin.

The outstanding characteristic of the resultant calculations was the presence of a non-crystallographic pseudo-symmetry element at cell co-ordinates  $\frac{1}{4}, \frac{1}{4}, 0$  and  $\frac{3}{4}, \frac{3}{4}, \frac{1}{2}$ . Reference to Figure 2.7 reveals the source; the dark circles mark the  $\text{PtCl}_4^{2-}$  sites, four symmetry

# PSEUDO-SYMMETRY DIAGRAM

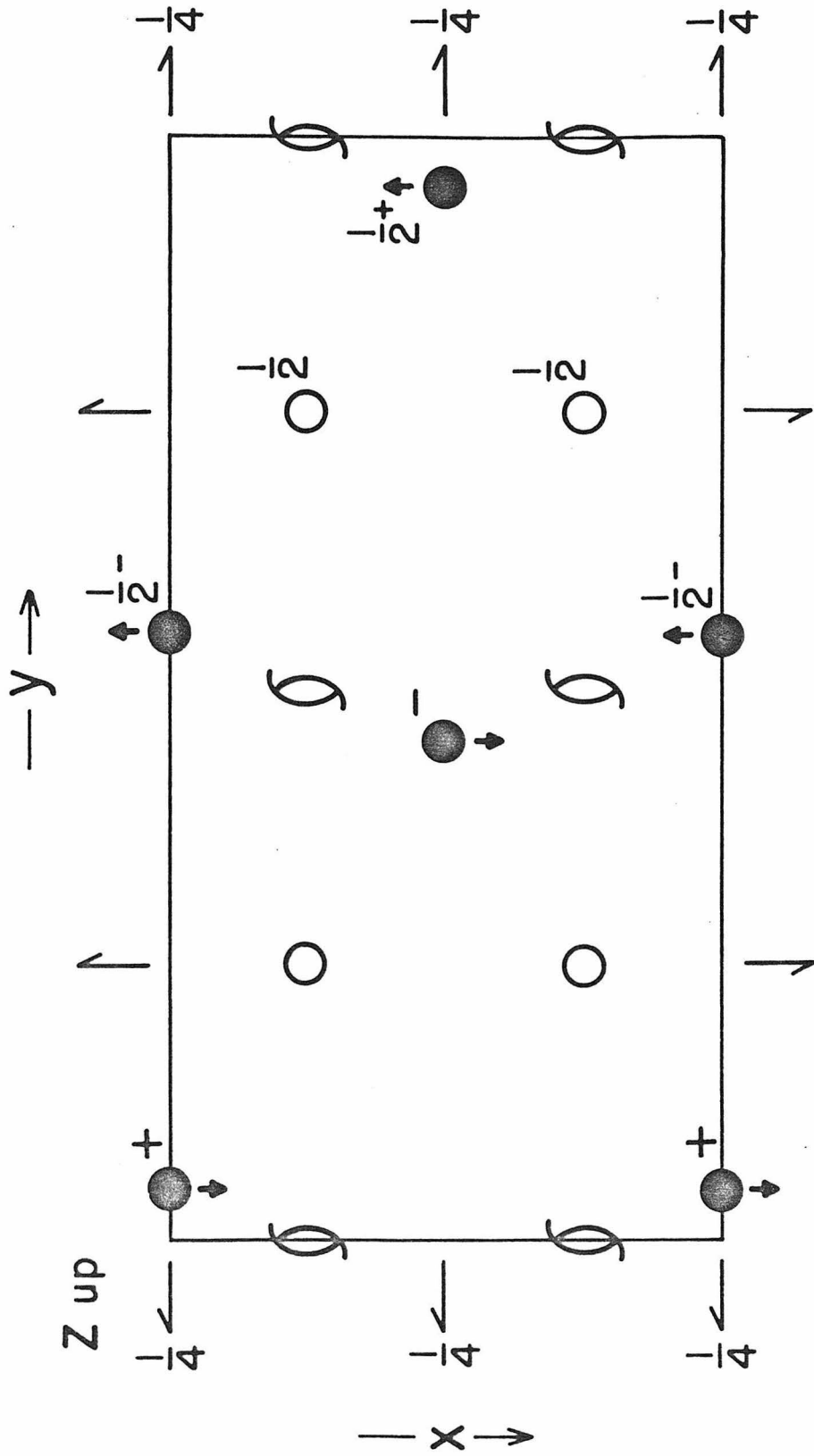


Figure 2.7.

equivalent sites per unit cell, which are seen to possess an x co-ordinate near zero (+0.01). This produced an accidental centrosymmetric distribution of the  $\text{PtCl}_4^{2-}$  sites about inversion centers at  $\frac{1}{4}, \frac{1}{4}, 0$  and  $\frac{3}{4}, \frac{3}{4}, \frac{1}{2}$ . The effect is also seen in the calculated heavy atom structure factors. If one takes the structure factor expressions collapsed over the four symmetry equivalent positions for  $P2_12_12_1$ ,

$$A = 4 \cos 2\pi \left( hx - \frac{h-k}{4} \right) \cos 2\pi \left( ky - \frac{k-l}{4} \right) \cos 2\pi \left( lz - \frac{l-h}{4} \right)$$

$$B = -4 \sin 2\pi \left( hx - \frac{h-k}{4} \right) \sin 2\pi \left( ky - \frac{k-l}{4} \right) \sin 2\pi \left( lz - \frac{l-h}{4} \right).$$

It can readily be seen that  $x \equiv 0$  leads to the conditions:

$$(0, k, l) \quad \text{no conditions}$$

$$(h, 0, l) \quad h+l \text{ odd, } A = B = 0$$

$$(h, k, 0) \quad k \text{ odd, } A = B = 0 .$$

Examination of heavy atom structure factors for the refined  $\text{PtCl}_4^{2-}$  site revealed such pseudo-extinctions were approximately obeyed for both calculated and observed  $\Delta F$  terms. Resultant  $\text{Pt}(\text{CN})_4^{2-}$  and  $\text{UO}_2^{2+}$   $\Delta F$  maps were constrained to show this pseudo-symmetry, one of the pseudo-symmetric pair being the correct site and the other spuriously introduced by the phasing abnormalities.

For the  $\text{Pt}(\text{CN})_4^{2-}$  maps, true peaks could not be differentiated from false images. For the  $\text{UO}_2^{2+}$  maps, the slight departure from perfect symmetry (x not identically zero) allowed the proper  $\text{UO}_2^{2+}$  single site to appear at a peak height four times the spurious image. That this choice was correct was confirmed by first refining the  $\text{UO}_2^{2+}$  data and site as a single derivative as had been done with the  $\text{PtCl}_4^{2-}$



data, then including the two into phase refinement concurrently and observing that the two derivatives refined congruently, i. e., predicting the same native protein centric phase.

The protein centric phases, now calculated from two derivatives, one not possessing pseudo-symmetry, were again combined with  $\text{Pt}(\text{CN})_4^{2-}$  differences to generate maps in which the three expected sites could be found with no evidence of the former spurious images. The higher concentration  $\text{Pt}(\text{CN})_4^{2-}$  data led to slightly higher Fourier peaks suggesting that the two derivatives differed solely in occupancies.

## 2. 7. Refinement of Heavy Atom Derivatives

With an origin-consistent set of atomic coordinates, refinement could be extended to the three-dimensional data using all derivatives. Weighting functions, a crucial aspect of the process, will be discussed in a separate section. Under refinement, the heavy atom parameters were varied in several combinations and permutations to determine the extent of correlation among the interdependent parameters, such as over-all scale and the isotropic temperature factor. As before, the thermal parameters could not be refined beyond their initial values. Refinement did not significantly alter the heavy atom data scale factors, yet caused fluctuations in occupancies, and so these scale factors were held fixed from the early stages onward. Refinement of the heavy atom sites in one derivative, holding the other derivative sites constant, invariably led to improvement in reliability indices (see Table 2. 7) for the adjusted derivative, at the expense of the same statistical index for the others. As the average figure of merit (Blow

and Crick, 1959) then suffered, it seemed obvious that all sites must be refined concurrently. The computational experiments in refinement were too numerous to be listed here. After the experiments, it was found that the same set of refined parameters were obtainable by five cycles of phase analysis-least squares adjusting concurrently co-ordinates and occupancies for all sites, starting with Patterson map initial co-ordinates and initial Z and B terms as discussed for  $\text{PtCl}_4^{2-}$ .

It is instructive to compare the results of three-dimensional refinement with refinement over centric data only. For purposes of comparison the  $\text{PtCl}_4^{2-}$  and  $\text{UO}_2^{2+}$  sites were refined in general reciprocal space and the results compared with the equivalent refinement, mentioned previously in connection with the relative origins problem, over centric data only. Table 2.6 lists the corresponding parameters and reveals the marked similarity between the two methods. Were general space refinement for some protein project to be prohibitively expensive in computation cost or time, it appears that restriction to the limited centric data would be justifiable.

At this stage, native protein phases were calculated from the refined parameters and a series of three-dimensional difference Fourier syntheses was produced using these calculated native phases and the observed  $\Delta F$ 's for each derivative in turn. Although inclusion of all derivatives leads to the possibility of some feedback in this Fourier approach, the number of derivatives used for phasing lessens this tendency (Dickerson, et al., 1967b).

TABLE 2.6

COMPARISON OF CENTRIC AND GENERAL SPACE REFINEMENT

Refined Co-ordinates				
	<u>x</u>	<u>y</u>	<u>z</u>	<u>Z</u>
PtCl <sub>4</sub> <sup>2-</sup> , centric method	0.0123	0.0519	0.2130	75.84
, general method	0.0135	0.0526	0.2135	74.66
UO <sub>2</sub> <sup>2+</sup> , centric method	0.0966	0.3011	0.7415	37.29
, general method	0.0975	0.3027	0.7378	33.16

Refinement Statistics*		
	<u>Centric Refinement Method</u>	<u>General Refinement Method</u>
Mean figure of merit, centric	0.822	0.826
Mean figure of merit, general	0.619	0.626

	<u>PtCl<sub>4</sub><sup>2-</sup></u>	<u>UO<sub>2</sub><sup>2+</sup></u>	<u>PtCl<sub>4</sub><sup>2-</sup></u>	<u>UO<sub>2</sub><sup>2+</sup></u>
R <sub>k</sub> , centric	0.090	0.058	0.091	0.056
R <sub>k</sub> , general	0.087	0.055	0.087	0.051
R <sub>c</sub>	0.481	0.656	0.485	0.629
ERMS	58.9	40.4	59.3	37.9

\*See Table 2.11 for definitions. These statistics follow from the final parameters and are independent of the refinement method. The statistics quoted as "centric" or "general" refer to their computation over respectively centric or general reflections. This presentation establishes that the refinement method does bias the statistics in favor of the data subset refined.

The maps were examined for minor sites not found from the Patterson, but now possibly revealed by the reasonably accurate protein phases coupled to the minor site contribution to  $\Delta F$ . From the frequency and distribution of features in each map, it appeared that the general noise level was approximately 8% of the peak height for the major site in that derivative. Therefore most attention was directed toward any features higher than this cut-off. For the  $\text{UO}_2^{2+}$  and  $\text{Pt}(\text{CN})_4^{2-}$  data there were none, and it was concluded that no minor sites were present in these derivatives.

For the  $\text{PtCl}_4^{2-}$  data, two minor sites were possible, each at 13% of the major site peak height. At their indicated occupancy, their Harker vectors would be obscured by noise in the  $(\Delta F)^2$  map, but, their cross vectors should appear, and indeed peaks were found, at about the Patterson noise level, at the calculated vector positions. The two minor sites were therefore included in phase refinement where they immediately (one cycle) refined to zero occupancy (1 and 4 electrons versus 69 for the main site), and hence were dropped from consideration.

Table 2.7 lists the final parameters at this  $4\text{\AA}$  stage. A  $4\text{\AA}$  electron density map was computed using figure of merit adjusted coefficients (Blow and Crick, 1959). A discussion of results will be postponed except for a comment on series termination errors. Since the data for the map are limited to a small subset, with significant terms omitted, there is the possibility that series termination has introduced false features into the synthesis. To test this possibility, artificial temperature factors of the form  $e^{(-Ds^2)}$  were applied to the

TABLE 2.7  
HEAVY-ATOM PARAMETERS AND  
REFINEMENT STATISTICS, 4 Å REFINEMENT

	$R_k^1$	$R_c^1$	Heavy-atom sites				
			x	y	z	$Z^2$	$B^2$
$\text{PtCl}_4^{2-}$	0.091	0.512	0.0087	0.0528	0.2147	68.6	4.0
$\text{UO}_2^{2+}$	0.053	0.619	0.0992	0.3018	0.7384	31.7	2.3
$\text{Pt}(\text{CN})_4^{2-}$ (5.6 mM)	0.108	0.458	0.1008	0.4307	0.7954	76.9	1.6
			0.4622	0.4664	0.8657	46.0	1.6
			0.3078	0.2810	0.7530	33.0	1.6
$\text{Pt}(\text{CN})_4^{2-}$ (11.2 mM)	0.123	0.499	0.1017	0.4312	0.7952	84.4	1.8
			0.4635	0.4672	0.8654	53.8	1.8
			0.3056	0.2812	0.7573	43.4	1.8

<sup>1</sup>See definitions accompanying Table 2.11.

<sup>2</sup>Isotropic atomic scattering factors =  $Z e^{-Bs^2}$ .

coefficients to generate a series with decreasing terms at the resolution limit. Let  $\langle F \rangle_0$  be the average native structure factor amplitude over all the  $4 \text{ \AA}$  data before application of the artificial factor and  $\langle F \rangle_{4 \text{ \AA}}$  be the average computed over a narrow range of  $s^2$  near the  $4 \text{ \AA}$  limit (the highest resolution 13% of the data); then the D factor was chosen such that  $\langle F \rangle_{4 \text{ \AA}}$  was 75%, 50%, or 25% of  $\langle F \rangle_0$ . Several sections of an electron density map were calculated in each case and compared, yielding the generalization that no change in features was observed, only a gradual broadening of peaks and troughs. It was held that series termination was not a serious error and the artificial factors were omitted.

For the high resolution refinement at  $2.45 \text{ \AA}$ , the  $4 \text{ \AA}$  process was essentially duplicated. In this instance, initial refinement used the averaged data sets for native,  $\text{PtCl}_4^{2-}$ , and  $\text{Pt}(\text{CN})_4^{2-}$  and the high concentration  $\text{UO}_2^{2+}$  set (Table 2.4, set UN12IIa). Several cycles were performed beginning with the  $4 \text{ \AA}$  parameters and it was now possible to refine temperature factors realistically for the individual heavy atoms.

The resulting protein phases were used to compute  $2.45 \text{ \AA}$   $\Delta F$  maps again matching the various derivatives in turn with the same master phase set. As in the previous case, these same-phase-cross-difference maps were searched for minor sites now with a minimum Fourier peak height of 6% of the main heavy atom site in the case of the  $\text{PtCl}_4^{2-}$  and  $\text{UO}_2^{2+}$  derivatives and 8% of the main site in  $\text{Pt}(\text{CN})_4^{2-}$ . Two possible minor sites for  $\text{PtCl}_4^{2-}$ , four for  $\text{UO}_2^{2+}$ , and six for

$\text{Pt}(\text{CN})_4^{2-}$  satisfied calculated Patterson cross-vector predictions and were entered into least squares. Table 2.8 gives the result of occupancy shifts and indicates that only one minor site in  $\text{UO}_2^{2+}$  (site b) and one in  $\text{Pt}(\text{CN})_4^{2-}$  (site h) behaved consistent with the data. These were therefore accepted as valid.

The now complete list of heavy atom sites was converted to an anisotropic temperature format and  $\beta_{ij}$  parameters adjusted with the phase program. The main  $\text{PtCl}_4^{2-}$  and  $\text{UO}_2^{2+}$  sites and the two strongest  $\text{Pt}(\text{CN})_4^{2-}$  sites showed reasonable, i. e. small, shifts. The light  $\text{UO}_2^{2+}$  and  $\text{Pt}(\text{CN})_4^{2-}$  sites yielded non-positive definite tensors, were reset to their isotropic equivalent, and subsequently were held there. This failure raised the question of whether the main sites were meaningfully represented by the anisotropic parameters or had they been shifted as extra degrees of freedom to accommodate random error in the data? To satisfy this dilemma an ORTEP plot was prepared graphing the heavy atom sites as directed thermal, vibrational ellipsoids specified by the  $\beta_{ij}$ 's and given in Figure 2.8. Below and to the left of each ellipsoid is presented an equivalently orientated section of the heavy atom as it appeared in the cross-difference Fourier synthesis. Recalling that these syntheses were computed from protein phases in turn based upon isotropic heavy atoms, the general correspondence in shapes lends credence to the anisotropic refinement.

The over-all billiard ball or egg-shape to the heavy atoms is at variance with their ligand arrangements. At 2.45 Å the ligands

TABLE 2. 8

OCCUPANCY REFINEMENT OF POTENTIAL MINOR SITES<sup>1</sup>

	Derivative Site	Initial Occupancy <sup>2</sup> (e <sup>-</sup> )	Final Occupancy (e <sup>-</sup> )
PtCl <sub>4</sub> <sup>2-</sup>	a	65.1 <sup>3</sup>	64.9
	b	5	- 0.3
	c	5	2.8
UO <sub>2</sub> <sup>2+</sup>	a	82.9 <sup>3</sup>	86.0
	b	6.9	8.4
	c	9.0	3.6
	d	6.5	1.3
	e	5.3	1.9
Pt(CN) <sub>4</sub> <sup>2-</sup>	a	81.7 <sup>3</sup>	82.4
	b	50.0 <sup>3</sup>	52.0
	c	36.4 <sup>3</sup>	36.2
	d	10.4	1.7
	e	7.5	2.2
	f	7.0	- 1.3
	g	7.7	- 4.8
	h	13.8	16.9
	i	7.9	1.3

<sup>1</sup> Test sites were entered with the same isotropic temperature factor as the main site, and this was also refined.

<sup>2</sup> Computed from the ratio of Fourier peak heights.

<sup>3</sup> Main, established sites.



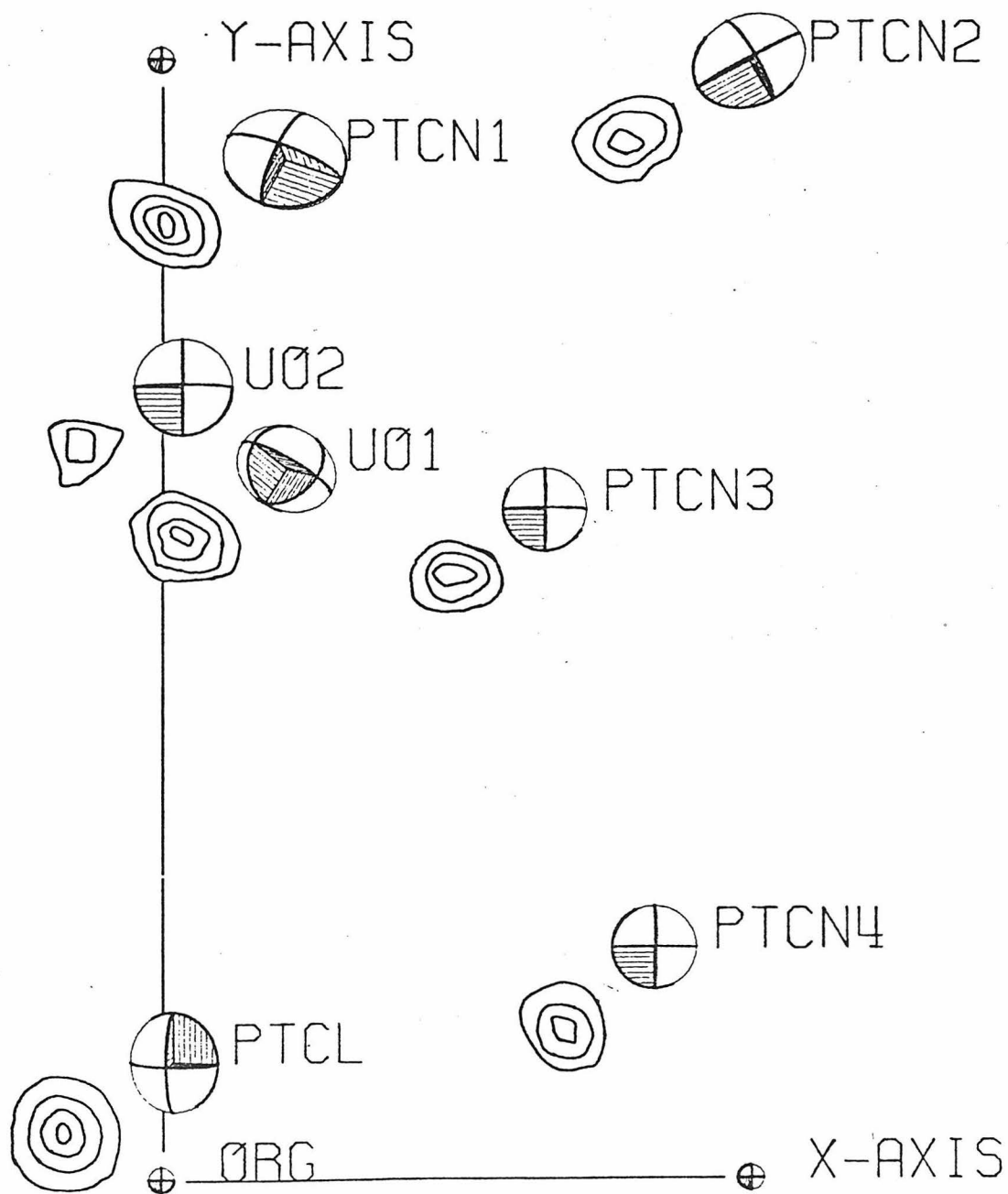


Figure 2.8. Thermal ellipsoids of the heavy atom sites.

would be expected as at least protrubances on the spherical central atom. The failure to detect the ligands has been attributed to the following: (1) there may be disorder at the heavy atom binding site on the protein surface where, for example, rotation about the protein-heavy atom bond could obscure substituent atoms; (2) the oxygens and cyanide ions are comparable in electron density to the interstitial sulfate solution and in a difference map could be subtracted out along with the solvent; and (3) ligand substitution is quite possible for  $\text{PtCl}_4^{2-}$  (Langford and Gray, 1965), so this derivative may also have substituents at the same electron density as the surrounding solvent.

Use of the averaged data sets for refinement at this stage was clearly a compromise. While averaging the various derivative sets to give one master native set and a master set for each chemical modification gives statistically more accurate  $F_{\text{PH}}$ 's, it may be that there are slight but physically real differences in at least occupancies among the non-averaged derivative sets. In the extreme there are the high and low  $4 \text{ \AA}$   $\text{Pt}(\text{CN})_4^{2-}$  sets and the  $4 \text{ \AA}$  versus the  $2.45 \text{ \AA}$   $\text{UO}_2^{2+}$  set where occupancies can change by a factor of two. Thus averaging may have obscured some useful differences between almost redundant data sets, differences that could be exploited to improve the native phases.

If two derivatives have the same sites of chemical substitution and differ only by a common ratio between all the occupancies of the identical sites, no new native phase information is provided by using both derivatives over that provided by either one. This is best seen by adopting a Blow-Crick (1959) formulation for the single isomorphous

replacement (SIR) calculated native protein phase. Then the native phase satisfies the condition

$$\left(\frac{d}{d\varphi} P(\varphi)\right)_{\varphi_b} = 0 = \left(\frac{d\epsilon}{d\varphi}\right)_{\varphi_b}$$

where  $P(\varphi)$  is the Blow-Crick phase probability distribution,  $\varphi$  the formal native phase,  $\varphi_b$  the best calculatable phase, and  $\epsilon$  the lack of closure error. Letting  $F_{PH}$  be the observed derivative scattering factor magnitude,  $F_P$  the same for the native structure,  $f$  the magnitude contributed by the heavy atoms,  $\alpha$  the phase of  $f$ , and  $D$  the calculated derivative scattering factor magnitude, the above condition implies

$$\frac{|F_{PH} - D(\varphi)| F_P f \sin(\varphi_b - \alpha)}{D(\varphi)} = 0$$

which for non-trivial cases says

$$\varphi_b = \alpha .$$

Thus a single derivative predicts a protein phase independent of  $f$ . Derivatives differing only by a common ratio of occupancies will have different values of  $f$ , but the same  $\alpha$ , and hence give the same  $\varphi_b$ .

Nevertheless, a joint Blow-Crick phase probability distribution function for the two combined is much sharper than for either single derivative and this sharpening influences the final "best" protein phase (see Proposition I).

Setting aside the final parameters from the work with averaged data sets, the process was duplicated treating all data sets as independent derivatives — one 4 Å and two 2.45 Å  $\text{PtCl}_4^{2-}$  sets, two

4 Å and one 2.45 Å  $\text{Pt}(\text{CN})_4^{2-}$  set, and one 4 Å and one 2.45 Å  $\text{UO}_2^{2+}$  set. The course of refinement followed ground rules established earlier, such as only constant isotropic B-values at 4 Å. Tables 2.9 through 2.11 record pertinent results from which it can immediately be seen that an improvement resulted in the figures of merit although co-ordinates and occupancies were consistent across the sets.

A theory too complex to be recounted here (but see Proposition I) predicts increases in figure of merit in the eight derivative refinement over the three derivative refinement due to purely statistical effects and not an improvement in phases. As noted, the mean figure of merit has risen, and in an attempt to calibrate this rise and assess the pure statistical contribution, a pseudo-eight derivative phasing model was prepared. This consisted of the averaged data from the three derivative model copied a number of times to give the same redundancy for a given chemical derivative as in the true expanded eight derivative approach. The eight derivative model contained three  $\text{PtCl}_4^{2-}$ , three  $\text{Pt}(\text{CN})_4^{2-}$ , and two  $\text{UO}_2^{2+}$  sets, so the pseudo-model contained  $\langle F \rangle_{\text{PtCl}_4^{2-}}$  three times,  $\langle F \rangle_{\text{Pt}(\text{CN})_4^{2-}}$  three times and  $\langle F \rangle_{\text{UO}_2^{2+}}$  twice. This copying, equivalent to exponential multiplicative factors in the Blow-Crick phase probability distribution of 3, 3, and 2 for the three derivative model, was done for all reflections, but only the results to 4 Å should be compared, as the actual non-averaged derivative model contained mostly 4 Å remeasurements of the same chemical derivative.

TABLE 2.9

THREE AVERAGED DERIVATIVES  
FINAL PARAMETERS

Derivative	Co-ordinates			Effective		Anisotropic <sup>1</sup> thermal parameters × 10 <sup>3</sup>					
	x	y	z	Z		β <sub>11</sub>	β <sub>22</sub>	β <sub>33</sub>	β <sub>12</sub>	β <sub>23</sub>	β <sub>13</sub>
PtCl <sub>4</sub> <sup>2-</sup>	0.0107	0.0520	0.2095	67.4		2.97	1.08	2.84	-0.10	0.69	1.24
Pt(CN) <sub>4</sub> <sup>2-</sup>	0.0979	0.4302	0.7928	89.4		5.20	1.02	4.64	-0.43	-1.20	2.26
	0.4684	0.4692	0.8628	55.2		4.54	0.93	2.04	0.85	-0.95	-2.45
	0.3072	0.2833	0.7651	39.6		2.52	0.68	4.60	0.00	0.00	0.00
	0.3951	0.0979	0.7758	20.4		4.20	1.13	7.67	0.00	0.00	0.00
UO <sub>2</sub> <sup>2+</sup>	0.0997	0.3013	0.7388	84.4		3.35	0.75	6.68	-0.56	0.23	0.30
	0.0179	0.3367	0.7034	8.7		3.42	0.92	6.24	0.00	0.00	0.00

<sup>1</sup> Atomic scattering factor =  $Z e^{-(\beta_{11} h^2 + \beta_{22} k^2 + \beta_{33} l^2 + \beta_{12} hk + \beta_{23} kl + \beta_{13} hl)}$ .

TABLE 2. 10

EIGHT DERIVATIVE  
FINAL HEAVY ATOM SITES

Derivative	Co-ordinates			Effective	Isotropic	Anisotropic <sup>1</sup>			thermal parameters × 10 <sup>3</sup>		
	x	y	z	Z	B	β <sub>11</sub>	β <sub>22</sub>	β <sub>33</sub>	β <sub>12</sub>	β <sub>13</sub>	β <sub>23</sub>
PtCl <sub>4</sub> <sup>2-</sup> (2.5Å)	.0110	.0524	.2097	84.2	-	3.02	1.05	2.55	0.24	1.34	2.54
PtCl <sub>4</sub> <sup>2-</sup> (2.5Å)	.0107	.0522	.2104	71.5	-	2.76	1.09	2.52	0.04	1.39	2.32
PtCl <sub>4</sub> <sup>2-</sup> (4.0Å)	.0104	.0522	.2122	71.4	16.0	-	-	-	-	-	-
Pt(CN) <sub>4</sub> <sup>2-</sup> (2.5Å)	.0979	.4307	.7934	86.8	-	5.87	0.97	5.09	-0.04	-1.35	4.65
	.4690	.4690	.8628	53.9	-	5.35	1.01	1.82	0.10	-1.17	-2.67
	.3074	.2834	.7649	39.3	-	3.71	0.94	1.46	0.66	0.99	1.34
	.3999	.0976	.7727	17.4	-	4.20	1.13	7.67	0.00	0.00	0.00
Pt(CN) <sub>4</sub> <sup>2-</sup> (4Å)	.0996	.4304	.7941	81.4	6.3	-	-	-	-	-	-
	.4680	.4681	.8635	52.6	6.3	-	-	-	-	-	-
	.3050	.2829	.7598	36.0	6.3	-	-	-	-	-	-
	.3862	.0980	.7828	16.7	12.8	-	-	-	-	-	-
Pt(CN) <sub>4</sub> <sup>2-</sup> (4Å)	.0991	.4309	.7943	85.9	7.0	-	-	-	-	-	-
	.4681	.4681	.8637	61.5	7.0	-	-	-	-	-	-
	.3049	.2835	.7623	45.4	7.0	-	-	-	-	-	-
	.3879	.0985	.7908	25.8	12.8	-	-	-	-	-	-
UO <sub>2</sub> <sup>2+</sup> (2.5Å)	.0995	.3012	.7396	80.3	-	3.22	0.69	7.93	-0.80	0.28	5.50
	.0169	.3368	.7014	7.1	-	3.42	0.92	6.24	0.00	0.00	0.00
UO <sub>2</sub> <sup>2+</sup> (4Å)	.0994	.3008	.7405	31.9	9.0	-	-	-	-	-	-

<sup>1</sup> See definition in Table 9.

TABLE 2. 11  
FINAL REFINEMENT STATISTICS

Derivative	$R_c^1$	$R_k^2$ (centric)	$R_k^2$	$E^3$	$\langle m \rangle_c^4$	$\langle m \rangle^4$
EIGHT DERIVATIVE METHOD					0.92	0.67
PtCl <sub>4</sub> <sup>2-</sup> (2.5Å)	0.643	0.153	0.118	89.6		
PtCl <sub>4</sub> <sup>2-</sup> (2.5Å)	0.595	0.115	0.092	69.8		
PtCl <sub>4</sub> <sup>2-</sup> (4.0Å)	0.497	0.080	0.071	58.9		
Pt(CN) <sub>4</sub> <sup>2-</sup> (2.5Å)	0.587	0.149	0.127	93.8		
Pt(CN) <sub>4</sub> <sup>2-</sup> (4.0Å)	0.421	0.106	0.094	80.5		
Pt(CN) <sub>4</sub> <sup>2-</sup> (4.0Å)	0.420	0.113	0.100	86.7		
UO <sub>2</sub> <sup>2+</sup> (2.5Å)	0.542	0.110	0.094	65.4		
UO <sub>2</sub> <sup>2+</sup> (4.0Å)	0.574	0.049	0.045	34.7		
THREE DERIVATIVE METHOD					0.87	0.61
PtCl <sub>4</sub> <sup>2-</sup>	0.606	0.114	0.096	69.6		
Pt(CN) <sub>4</sub> <sup>2-</sup>	0.550	0.140	0.125	90.2		
UO <sub>2</sub> <sup>2+</sup>	0.541	0.110	0.094	64.8		

$$^1 R_c = \frac{\Sigma(\text{lack of closure error})}{\Sigma |\Delta F|} \text{ over centric data (Cullis, et. al., 1961).}$$

$$^2 R_k = \frac{\Sigma(\text{lack of closure error})}{\Sigma |F_{PH}|} \text{ (Kraut, et. al., 1962).}$$

<sup>3</sup>Root mean square lack of closure (Blow and Crick, 1959).

<sup>4</sup>Respectively, centric and all data figure of merit (Blow and Crick, 1959).

A number of criteria are available to examine these models. Let  $\varphi_3^b, \varphi_8^b, \varphi_p^b$ , stand for the Blow-Crick "best" protein phase for the three, eight, and pseudo-eight derivative models, then

$$\langle |\varphi_3^b - \varphi_8^b| \rangle = 16.9^\circ$$

$$\langle |\varphi_3^b - \varphi_p^b| \rangle = 12.0^\circ$$

$$\langle |\varphi_8^b - \varphi_p^b| \rangle = 12.9^\circ$$

and further, those reflections for which  $\varphi^b$  differed by large amounts ( $\Delta\varphi^b > 2\sigma$ ) correlated with very low figures of merit, and so would be eliminated from a Blow-Crick "best" Fourier calculation. Increases in average figure of merit are revealed as purely statistical by Figure 2.9. The most striking comparison of the models is given by Figures 2.10 and 2.11 which show the same protein electron density section calculated with the different phasing. The Blow-Crick coefficients  $mF_P$  have in each case been divided by the over-all average figure of merit for that particular phasing model to compensate for the decrease in term magnitude as  $\langle m \rangle$  drops, thus placing all sections on the same density scale. The virtual identity of the sections establishes the net insensitivity of the final map to manipulations in the choice of derivatives.

Phase refinement thus beaten to death, the final protein density map was calculated using Blow-Crick coefficients  $mF_P$  with phases from the averaged derivative refinement. Two separate maps were calculated with sectioning along the x and y crystallographic directions for use in a Richards optical comparator which gives two



## Radial Distribution of Figure of Merit

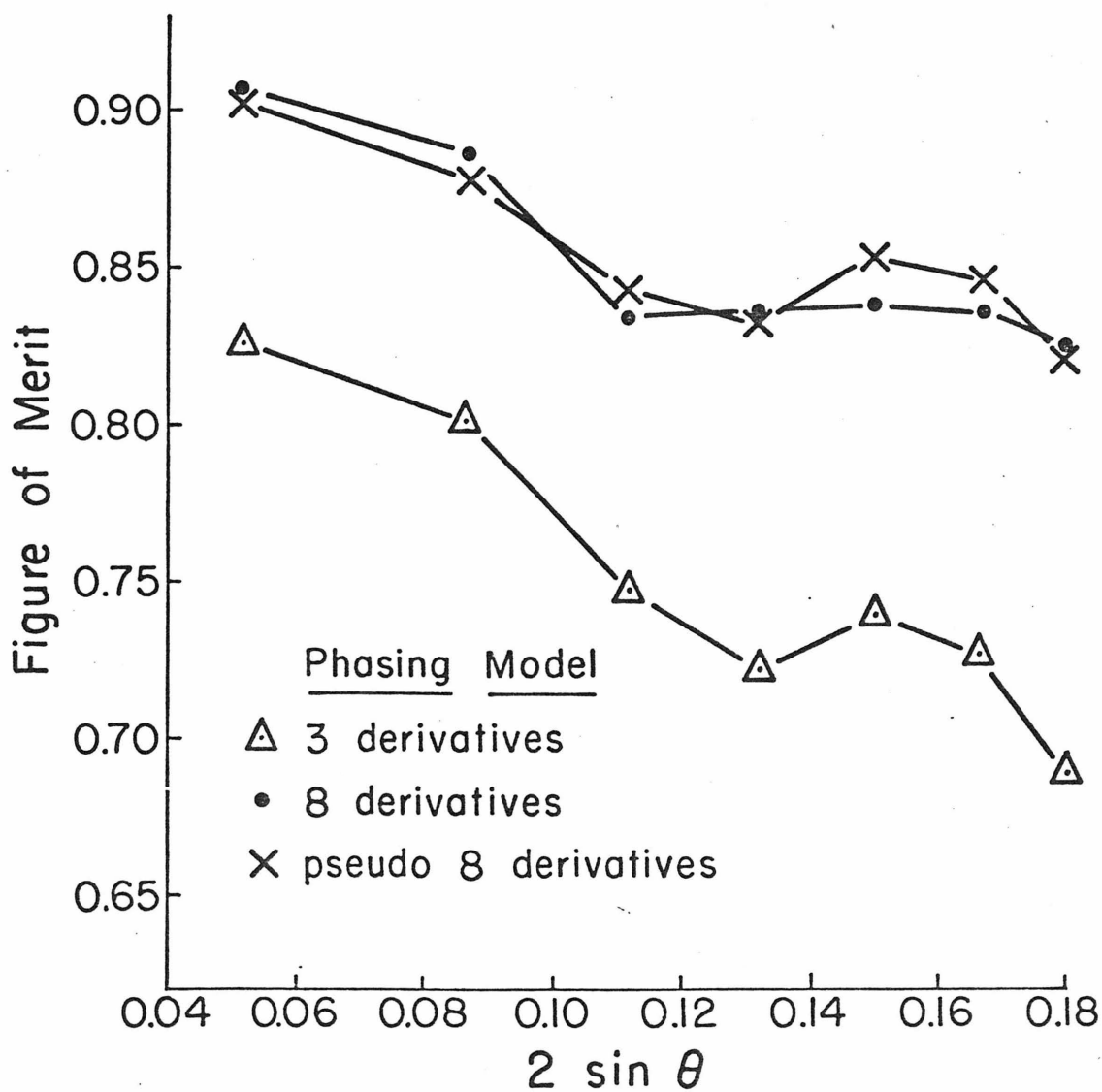
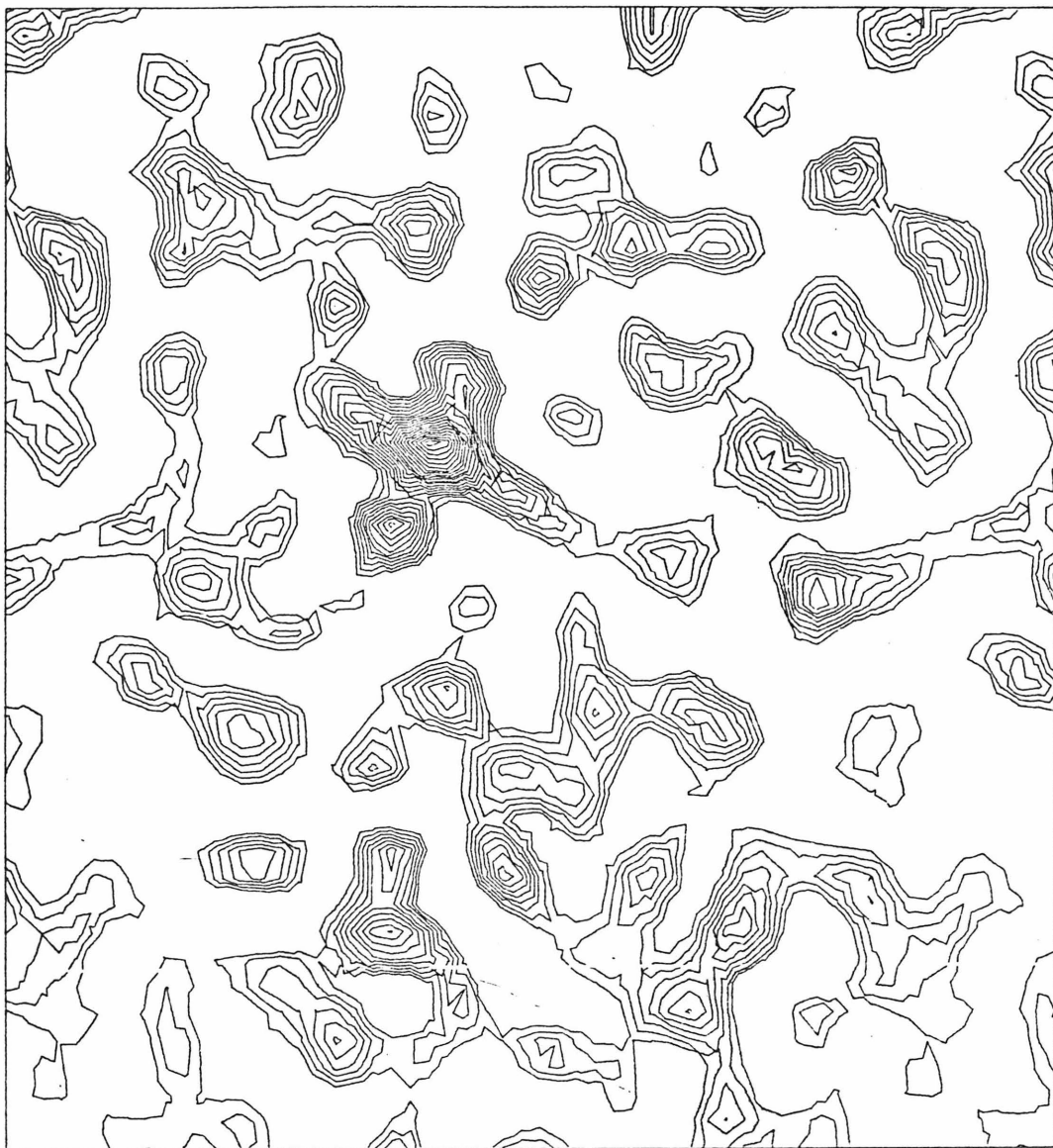
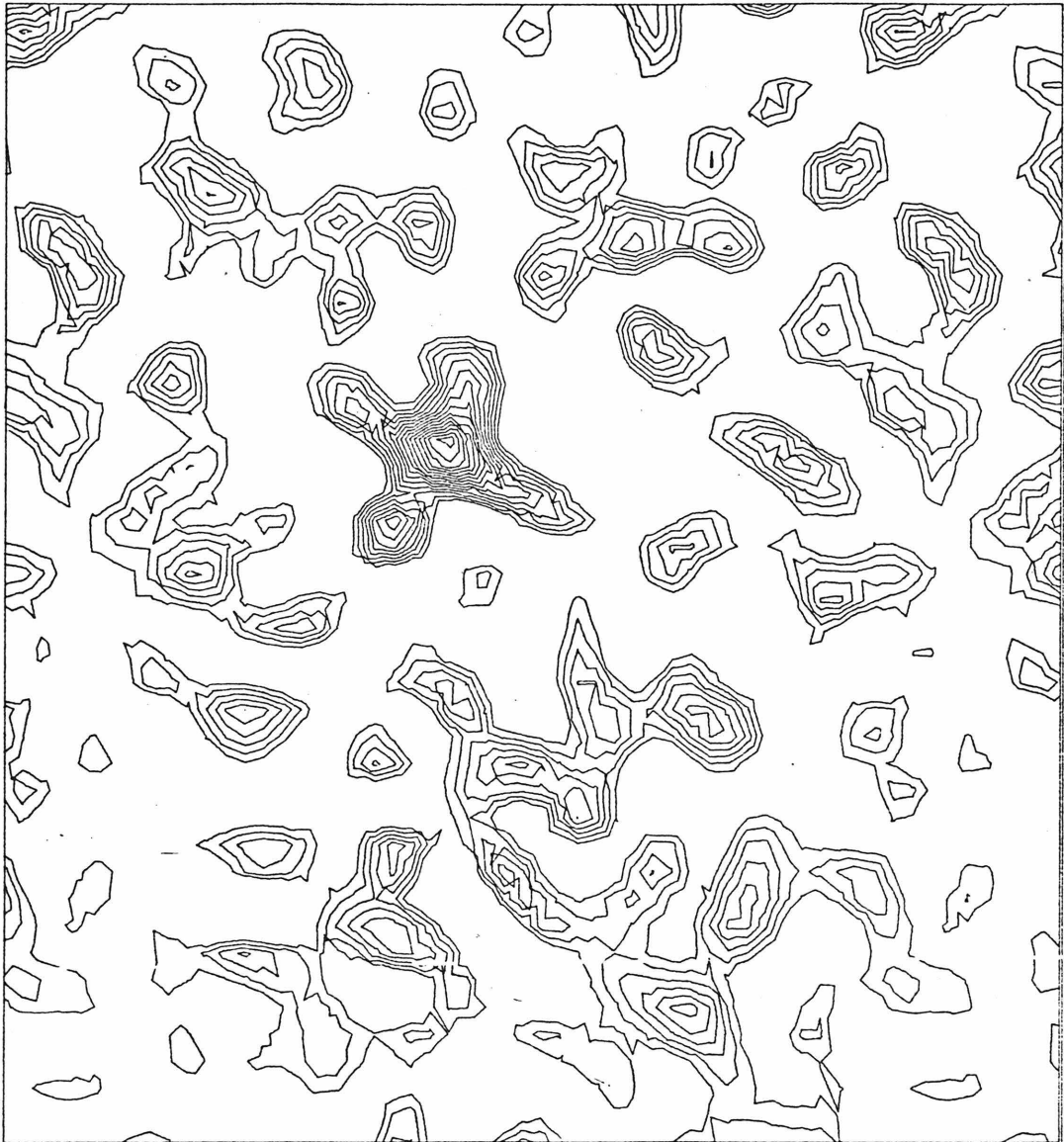


Figure 2. 9.



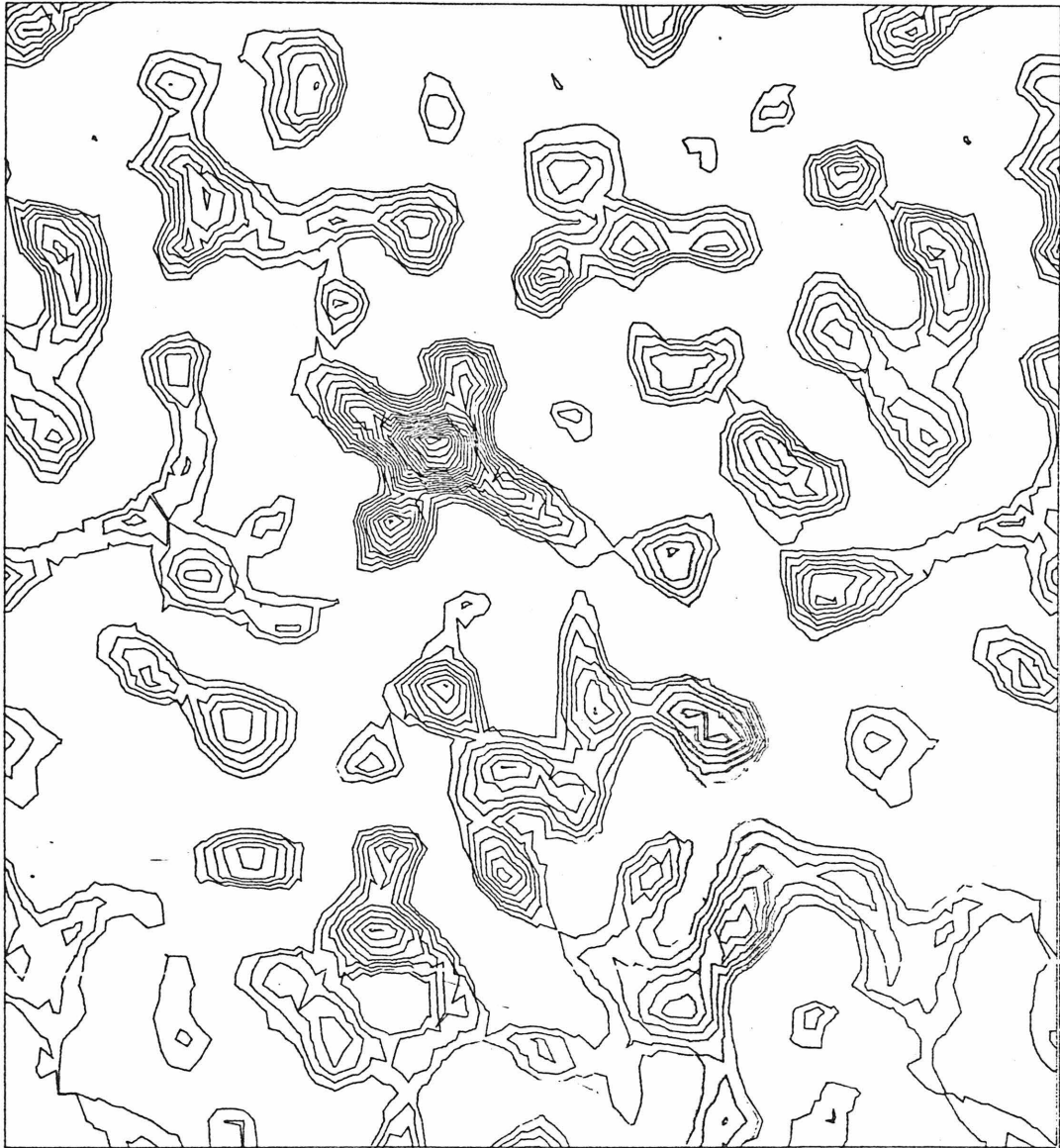
3 DERIVATIVE COEFFICIENTS : mF/.607,  $\phi$  (best)

Figure 2.10.1. Density section through the heme plane.



3 DERIVATIVE COEFFICIENTS : F,  $\phi$  (max)

Figure 2.10.2. As in 2.10.1 but with Blow-Crick  $\phi$ (max) phasing.



"8" DERIVATIVE COEFFICIENTS :  $mF/575, \phi$  (best)

Figure 2.11. Comparison with Figures 2.10.1 and 2.10.2 shows there is no improvement in the map when the expanded data set is used.

different views of the model and density simultaneously. Contouring was performed in an arbitrary manner, but with the goal of maximizing internal continuity without creating excessive cross-boundary connections.

The results of the structure determination are discussed in depth in the fourth chapter. Table 2.12 lists the identification of the heavy atom binding sites with respect to chemical functional groups on the protein. The binding of  $\text{PtCl}_4^{2-}$  to the methionine has been documented (Dickerson, et al., 1969). The binding of the uranyl ions suggests a chelation by oxygen atoms involved in unsaturated bonds. This same type of binding has been noted in chymotrypsin (Vandlen and Tulinsky, 1973), where a uranyl ion is poised between two acidic side chains of a glutamic and aspartic acid. The binding chemistry of  $\text{Pt}(\text{CN})_4^{2-}$  is obscure, with carboxylic oxygen, carbonyl oxygen, and amine groups involved.

## 2.8. Refinement Weighting Functions

The inclusion in this thesis of some few words on the weighting functions of least squares is justified by a colloquialism – you find what you seek. The experimenter can control to a remarkably fine degree the final set of refined parameters, the least squares solution, by choice of a weighting scheme applied to the observations, such a choice stemming from an astute knowledge of the reliability of the data or from a desire to obscure poor measurements or a lack of the proper solution.

The basic difficulty and the distinction from the normal crystallographic least squares problem lies in that the observation to

TABLE 2. 12  
HEAVY ATOM BINDING SITES

<u>Derivative</u>	<u>Co-ordinates</u>			<u>Binding Site</u>
PtCl <sub>4</sub> <sup>2-</sup>	0. 011	0. 052	0. 210	Adjacent to sulfur of Met 90
UO <sub>2</sub> <sup>2+</sup>	0. 010	0. 301	0. 739	Between carbonyl oxygen of Asn 1 side chain and carbonyl oxygen of main chain peptide bond 2
	0. 018	0. 337	0. 703	Between carbonyl oxygens of main chain peptide bonds 2 and 3
Pt(CN) <sub>4</sub> <sup>2-</sup>	0. 098	0. 430	0. 792	Near carbonyl oxygen and hydroxyl OH of Thr 98
	0. 468	0. 469	0. 863	Near ε-amine of Lys 89
	0. 307	0. 283	0. 765	Near side chain carboxyl oxygens of Glu 124
	0. 395	0. 098	0. 776	Near ε-amine of Lys 10

be fitted is actually a difference,  $|F(\text{derivative}) - F(\text{native})|$ , the magnitude of which is uncorrelated with the magnitude of the parental terms.

The lack of an obviously rational weighting scheme is reflected by the proliferation of schemes employed in the field. Dickerson, Weinzierl and Palmer (1968) have suggested assigning weights proportional to  $1/\langle \epsilon_{hkl}^2 \rangle_j^{\frac{1}{2}}$  where  $\epsilon_{hkl}$  is the lack of closure of reflection  $(h, k, l)$  for a particular derivative (Dickerson, Kendrew and Strandberg, 1961), and the root mean square average is computed over all  $j$  derivatives. Lipscomb, et al., (1966) have used

$$\text{weight} = n_{hkl} / \sum_j [\epsilon_j^2 + 9(E_j / |FPH_j|)^2]$$

where  $n_{hkl}$  is the number of derivatives on which  $(h, k, l)$  was measured;  $E_j$  is mean lack of closure; and  $FPH_j$  is the observed structure factor of heavy atom derivative  $j$ . Dreuth, Jansonius, and Wolthers (1967) proposed the use of a weight equal to the square of the figure of merit. Adams, et al., (1969) used

$$\text{weight}_{hkl,j} = q_j / (E_{j,\theta}^2 + E_F^2)$$

where  $q_j$  is the inverse of the scale factor required to place derivative  $j$  on the same scale as the native data;  $E_{j,\theta}^2$  is the root mean square lack of closure for derivative  $j$  computed over a narrow range of  $\sin \theta / \lambda$  appropriate to  $(h, k, l)$ ; and  $E_F^2$  is a resolution independent root mean square lack of closure averaged over all derivatives. Cohen, et al. (1969) and Ludwig, et al. (1971), among others, have avoided the issue by using unit weights.

In the early stages of the  $c_{550}$  phase refinement, three possible weighting schemes were employed: 1) the lack of closure scheme of Dickerson, et al. mentioned above; 2) a counting statistics based scheme where

$$\text{weight} = 1/[\sigma^2(F_P) + \sigma^2(F_{PH})] = 1/\sigma^2(\Delta F)$$

and  $\sigma(F_P)$ ,  $\sigma(F_{PH})$  are the standard deviations as defined earlier of the native and the derivative being refined; and 3) unit weights. A test was devised wherein each of these schemes was applied for four cycles of refinement of the major site of the  $\text{PtCl}_4^{2-}$  derivative. For computational ease, only the centric  $\Delta F(\text{PtCl}_4^{2-})$  data was refined. Figure 2.12 demonstrates the course of refinement for the co-ordinates and occupancy from which some generalizations may be made. For most of the parameters (x, y, z) all schemes are tending toward a common solution, while for some parameters (Z) the results may differ considerably. The lack of closure scheme is drastically slower to converge, which is understandable, as after every cycle the  $\epsilon$  for the  $\text{PtCl}_4^{2-}$  data changes, and hence to some extent so does the weight. This latter point is a serious problem in the early stages of refinement when many tests must be performed.

Little distinction in centric refinement could be seen between results from unit weights and counting statistics weights, yet, since the counting statistics do reflect the accuracy of  $|\Delta F|$ , the latter scheme was taken as most valid. In three-dimensional refinement, however, this scheme led to poorer refinement than unit weights, as judged by the various reliability indices. This may be due to the



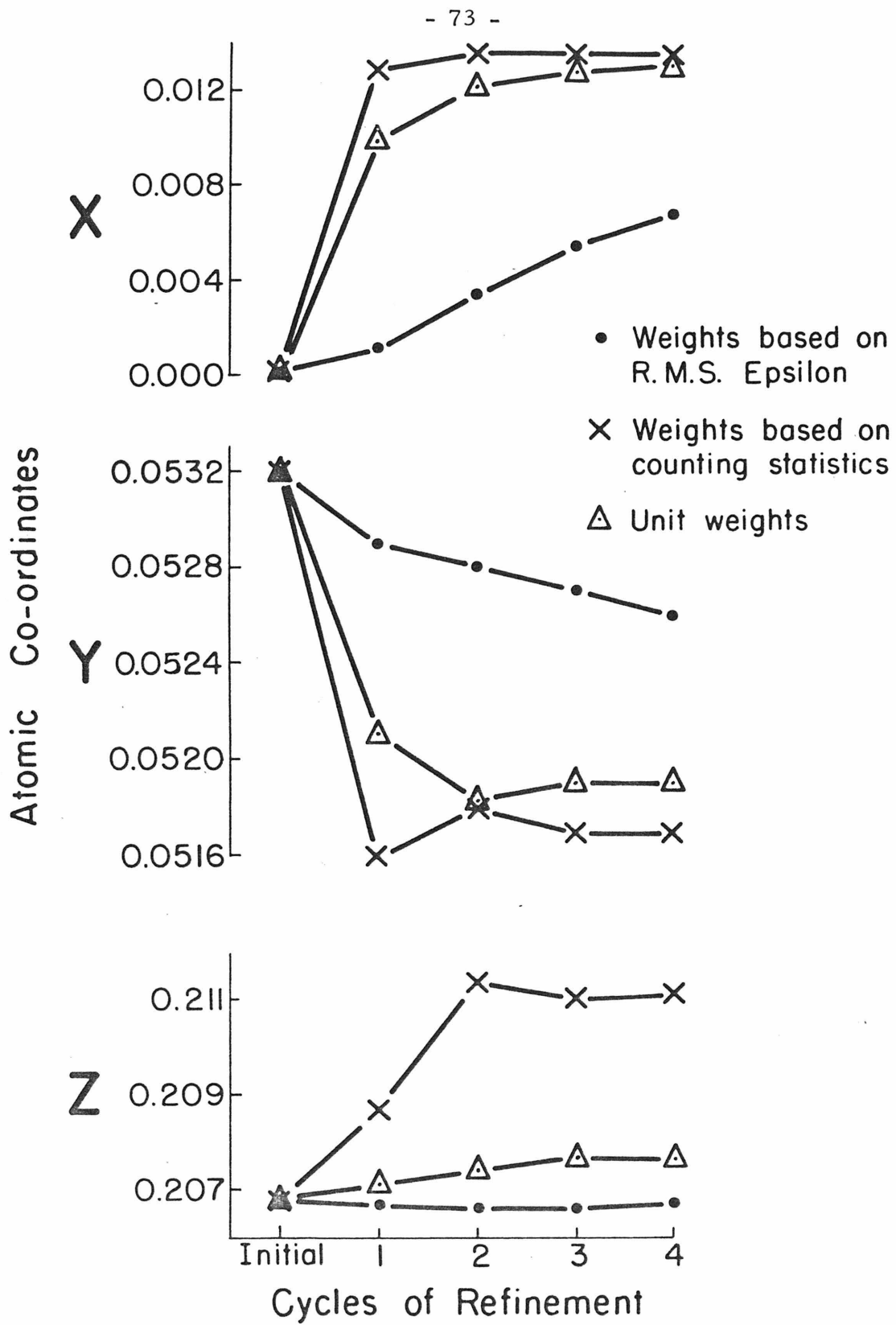


Figure 2.12. Weighting scheme comparison.

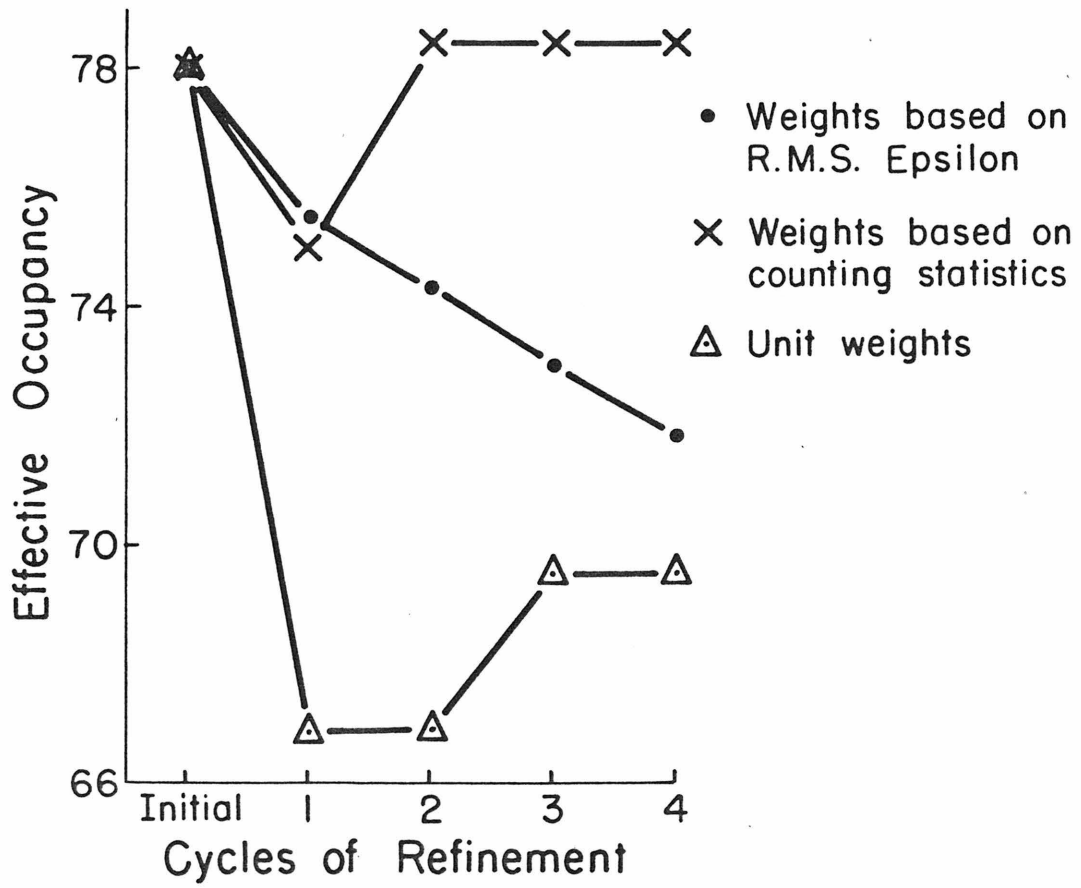


Figure 2.12. (continued)

fact that counting statistics results in a bias toward large  $\Delta F$  terms. For centric reflections, large  $\Delta F$  terms must be matched with large atomic  $f$  (calculated) values which are correspondingly sensitive to parameters, while in the general case  $f$  may be large, yet  $\Delta F = 0$  ( $f$  perpendicular in phase to  $F_P$ ). Hence a few large  $\Delta F$  terms dominate refinement and the majority of small  $\Delta F$  terms are not well matched.

Consideration of these points led to the following use of weighting schemes. Centric refinement was performed exclusively with counting statistics weights, and three-dimensional refinement with exclusively unit weights until a final convergence was reached. At that time, additional cycles of refinement were performed under the lack of closure scheme. In this manner the rapid convergence characteristics of unit weights could be merged with the inherent data evaluation of the  $\epsilon$ -scheme. Interestingly, when the jump between functions was so performed, the  $\epsilon$ -refinement converged very quickly (two cycles) with minimal readjustment of parameters. Fractional coordinates shifted by less than  $10^{-5}$  and occupancies by about 3%.

## 2.9. Crystalline $c_{550}$ as a Function of pH

Crystal structure experiments were undertaken to study the effect of pH on the conformation of the protein. In particular the question was asked — does  $c_{550}$  differ in structure at pH 5 and 9 from the native form at pH 7.5? The motivation for the experiments stems from recurrent observations on eukaryotic cytochrome suggesting some change at both high and low pH, (Margoliash and Schejter, 1966).

The high pH form of cytochrome c involves displacement of methionine 80 and is extensively invoked to explain a host of kinetic and structural data. Even though this high pH form is non-physiological, it has established that the integrity of the pH 7.5 form, which probably means a stable methionine-iron bond, is absolutely necessary for the in vivo redox mechanism. For example, it has been proposed that di-carboxymethylated methionine 80, methionine 65 cytochrome c, inactive toward cytochrome oxidase, mimics the high pH form in losing sulfur co-ordination (Schejter and Aviram, 1970).

For the present experiments, c<sub>550</sub> crystals at pH 7.5 still in the crystallization capillaries were transferred to soaking solutions at pH 5.1 (mother liquor adjusted with acetic acid) and pH 8.9 (mother liquor adjusted with concentrated aqueous ammonia), and as slow diffusion took place, the crystals equilibrated to the new pH. Centric zone data sets were collected and compared with the native data (see Tables 2.4 and 2.5).

The identity of the pH 5.1 and 7.5 data sets (R-factor = 0.4) immediately established that there is no structural change. Perhaps the result is not too surprising in that in the protein there is only one functional group with a  $pK_a$  in this range, the fifth ligand histidine, and its protonatable nitrogen is already involved in a stable hydrogen bond. The R-factor of 0.05 between native and pH 8.9 suggested a potential slight change. Difference Fourier syntheses, using observed  $\Delta F$ 's and native protein phases failed to reveal any specific feature of structural change, certainly nothing drastic in the Met 80 region. However, the pH of 8.9 is a sharp experimental limit, for attempts

to observe data at higher values saw complete disintegration of the crystals. The strongest statement that can be made by combining these two points is the claim that a sharp pH transition exists below which the native form predominates, above which some modified form exists, the modification being extensive enough to force the cytochrome molecules apart from their packing array in the crystal. The protonation of a lysine side chain, with a  $pK_a$  above 9, is certainly consistent with this view.

With hindsight, it can be seen that the above experiments really offer little conclusive evidence on the nature of the supposed structural change and border on the verge of worthlessness. The main problem is that the crystal lattice may be locking the cytochrome into its pH 7.5 conformation by crystal packing forces and thus preventing a change which would spontaneously occur in solution. Alternatively, the lattice might be isolating or sealing off from the external pH some group or groups which mediate the conformation change.

#### 2.10. Ammonium Selenate Binding to $\zeta_{550}$

Tulinsky and Wright (1973) have recently published results which may provoke profound repercussions in the field of protein crystallography. These authors made use of the remarkable similarity in chemical properties (size and  $pK_a$ ) between sulfate and selenate anions to study salt binding to chymotrypsin. By substituting selenate for sulfate in the mother liquor, measuring intensity changes, and calculating  $\Delta F$  maps, they were able to identify ordered binding sites for sulfate/selenate on the surface of the protein. The usefulness of

such information lies in interpreting the native protein map and being able to recognize density features as bound solvent rather than forcing something like a lysine side chain into the region. An additional potentiality lies in using the selenate-substituted crystals as derivatives for the phasing. This is a fascinating possibility in light of the fact that a large number of protein crystals are grown from ammonium sulfate solutions.

The substitution reaction was attempted on native  $c_{550}$  crystals already grown from sulfate solution. Commercial ammonium selenate is an extremely dirty preparation and must be repeatedly recrystallized from hot water-ethanol before use. By soaking experiments similar to those in the pH work, various solutions of differing selenate-sulfate concentrations were explored; the tightly packed protein crystals were highly inert to exchange and, eventually, solutions were used containing only selenate (95% saturated) and no sulfate.

Interestingly, no gross crystal damage occurred, although the mosaic spreads broadened considerably. A three-dimensional, 2.45 Å data set was collected (see Tables 2.4 and 2.5) and a full  $\Delta F$  Fourier map calculated. The resultant synthesis was very clean showing only a few peaks above the general noise level. This map was studied with the knowledge of the location of the molecular surface so that potential sites of selenate binding could be realistically deciphered from the general noise. Six sites were located and correlated with the molecular surface (see Table 2.13).

TABLE 2. 13

AMMONIUM SELENATE BINDING SITES

Co-ordinates			Molecular Location
0. 302	0. 363	0. 890	Near carbonyl oxygen of Glu 135
0. 254	0. 072	0. 072	Interstitial, greater than 4 Å from any protein group
0. 048	0. 041	0. 345	Near the ε-amine group of Lys 97
0. 309	0. 093	0. 633	Interstitial, greater than 4 Å from any protein group
0. 195	0. 184	0. 829	Near the hydroxyl OH of Thr 27
0. 273	0. 089	0. 371	Near the acidic side chain oxygens of Glu 64

The disappointingly weak appearance of the substitution sites may be attributed to several factors. (1) Close protein packing may lead to only a partial occupancy of the binding sites. (2) The binding sites may be disordered or possess unusually large temperature factors. (3) The change, sulfate to selenate, of 18 electrons may be inherently too small to be seen by this method. The small number of sites found was not surprising in light of point (1) above and a realization that most solvent sites must be so labile as to be essentially unordered and therefore non-existent in the x-ray analysis. It was anticipated that replacing the non-ordered interstitial sulfate with selenates of higher electron density would lead to greater  $\Delta F$  changes in the low order terms, those most sensitive to solvent effects.

However, a plot of R-factor calculated over ranges of  $\sin \theta / \lambda$  versus  $\sin \theta / \lambda$  demonstrated a line of zero slope. Possibly, lack of isomorphism at high resolution balances the loose solvent effect at low resolution to make  $\Delta F$  changes so uniform.

The two strongest sites were refined over the centric differences as a single derivative to respectively 18 and 6 electrons. However, the Cullis R-factor (0.75) and the figure of merit (0.36) clearly indicated that predicted protein phases were inaccurate. The final nail for the coffin was provided by a cross phase  $\Delta F$  map using  $\text{UO}_2^{2+}$   $\Delta F$ 's and phases predicted by the selenate data. This failed to show the  $\text{UO}_2^{2-}$  site, but did give a number of other nonsense sites.

In summary, the  $c_{550}$  selenate derivative, while indicating potential bound solvent, was not suitable for phasing. This latter failure may be attributable to the very low occupancies. This author feels that the potential for phasing via selenate still exists for other proteins in other circumstances. For instance, in some other case there may be a larger number of high occupancy sites, or, if one other strong, simple derivative is available, the selenate derivative may allow solution of the phase ambiguity by indicating which of the alternatives offered by the strong derivative is correct.

## 2.11. References

Arndt, U. W., and Willis, B. T. M., Single Crystal Diffractometry, University Press, Cambridge, 1966.

Adams, M. J., Haas, D. J., Jeffrey, B. A., McPherson, A., Mermall, H. L., Rossman, M. G., Schevitz, R. W., and Wonacrott, A. J., (1969). J. Mol. Biol., 41, 159.



- Andrews, P. (1964). Biochem. J., 91, 222.
- Blake, C. C. F., Mair, G. A., North, A. T. C., Phillips, D. C., and Sarma, V. R., (1965). Proc. Roy. Soc., B., 167, 365.
- Blow, D. M., (1958). Proc. Roy. Soc., 247, 302.
- Blow, D. M. and Crick, F. H. C., (1959). Acta Cryst., 12, 794.
- Cohen, G. H., Silverton, E. W., Matthews, B. W., Braxton, H., and Davies, D. R., (1969). J. Mol. Biol., 44, 129.
- Cullis, A. F., Muirhead, H., Perutz, M. G., Rossmann, M. G., and North, A. C. T., (1961). Proc. Roy. Soc., London, A265, 15.
- Dickerson, R. E., Kendrew, J. C., and Strandberg, B. E., (1961). Acta Cryst., 14, 1188.
- Dickerson, R. E., Kopka, M. L., Bordner, C. L., Varnum, J. C., Weinzierl, J. E., and Margoliash, E., (1967a), J. Mol. Biol., 29, 77.
- Dickerson, R. E., Kopka, M. L., Varnum, J. C., and Weinzierl, J. E., (1967b). Acta Cryst., 23, 511.
- Dickerson, R. E., Weinzierl, J. E., and Palmer, R. A., (1968). Acta Cryst., B24, 997.
- Dickerson, R. E., Eisenberg, D., Varnum, J., and Kopka, M. L., (1969). J. Mol. Biol., 45, 77.
- Dreuth, J., Jansonius, J. N., and Wolthers, B. G., (1967). J. Mol. Biol., 24, 449.
- Greer, J. (1971). J. Mol. Biol., 59, 107.
- James, R. W. (1965). The Optical Principles of the Diffraction of x-rays, Cornell University Press, Ithaca, New York, 1965, 43-46.
- Kraut, J., Sieker, L. C., High, D. F., Freer, S. T., (1962). Proc. Nat. Acad. Sci., 48, 1417.
- Langford, C. H., and Gray, H. B., (1965). Ligand Substitution Processes, W. A. Benjamin, Inc., New York.
- Levy, H. A., and Ellison, R. D., (1960). Acta Cryst., 13, 270.
- Lipscomb, W. N., Coppola, J. C., Hartsuck, J. A., Ludwig, M. L., Muirhead, H., Searl, J., and Steitz, T. A., (1966). J. Mol. Biol., 19, 423.

- Lipson, H. and Cochran, W., The Determination of Crystal Structures, Cornell University Press, Ithaca, New York, 1966.
- Ludwig, M. L., Andersen, R. D., Apgar, P. M., Burnett, R. M., LeQuesne, M. E., and Mayhew, S. G., (1971). In Cold Spring Harbor Symposium on Quantitative Biology, XXXVI, Cold Spring Harbor Laboratory, 1972.
- Margoliash, E., and Schejter, A., (1966). In Advances in Protein Chemistry, 21, (C. B. Anfinsen, M. L. Anson, J. T. Edsall, and F. M. Richards, Eds.), Academic Press, New York, 114.
- Margoliash, E., and Walasek, O. F., (1967). In Methods in Enzymology, X, Estabrook, R. W., and Pullman, M. E., (eds.), Academic Press, New York, 339-348.
- Matthews, B. W., (1966a). Acta Cryst., 20, 230.
- Matthews, B. W., (1966b). Acta Cryst., 20, 82.
- North, A. T. C., (1965). Acta Cryst., 18, 212.
- North, A. T. C., Phillips, D. C., and Matthews, F. S., (1968). Acta Cryst., A24, 351.
- Phillips, D. C., (1966). Advances in Structural Research; Diffraction Methods, 2, 75.
- Schejter, A., and Aviram, I., (1970). J. Biol. Chem., 245, 1552.
- Scholes, P. B., McLain, G., and Smith, L., (1971). Biochemistry, 10, 2072.
- Stout, J. H., and Jensen, L. H., (1968). X-ray Structure Determination, A Practical Guide, Macmillan Co., New York, 79-82.
- Tulinsky, A., Mani, N. V., Morimoto, C. N., and Vandlen, R. L., (1973). Acta Cryst., B29, 1309.
- Tulinsky, A., and Wright, L. H., (1973). J. Mol. Biol., 81, 47.
- Vandlen, R. L., and Tulinsky, A., (1973). Biochemistry, 12, 4193.
- Wilson, A. J. C., (1942). Nature, 150, 151.
- Wyckoff, H. W., Tsernoglou, D., Hanson, A. W., Knox, J. R., Lee, B., and Richards, F. M., (1970). J. Biol. Chem., 245, 305.
- Zeppezauer, M., Eklund, H., and Zeppezauer, E. S., (1968). Arch. Biochem. Biophys., 126, 564.

## CHAPTER 3. PRIMARY SEQUENCE OF $c_{550}$

### 3.1. Introduction

Even a cursory glance at such compilations of protein sequence data as the Dayhoff Atlas (Dayhoff, 1972) is sufficient to reveal that  $c$ -type cytochromes, especially from the higher organisms, are the proteins most extensively investigated in primary structure. It must be admitted that this is due in large part to their ubiquity and the relative ease, as biochemical substances go, with which they are isolated. At a somewhat higher plane, there are the contributions made by such exhaustive data toward understanding and reconstructing evolution at the precise molecular level, through sequence comparisons (Margoliash, Fitch and Dickerson, 1968). Among the closely related eukaryotes, the various cytochromes may be taken as a set of natural protein derivatives, showing the allowable modifications of side chain residues which still produce fully functional molecules. On the prokaryotic side of life, the available sequences reveal the astounding diversity of cytochromes, (for example, compare the four heme/protein cytochrome  $c_3$  of Desulfovibrio with horse heart cytochrome  $c$ ) and also the extent to which homology may be strained (for example, the comparison of the 82 amino acid Pseudomonas  $c_{551}$  with horse cytochrome  $c$  (Needleman and Blair, 1969; Dickerson, 1971)).

Since, of the cytochromes sequenced, the vast majority have been from eukaryotes, the primary order of a prokaryotic cytochrome such as M. denitrificans would be of obvious value in extending this comparative biochemistry. Since the denitrificans respiratory chain shows

cross-reactivity with that from eukaryotes at the level of cytochrome c, (Smith, Newton and Scholes, 1966) the information becomes of special interest. Potentially, a faltering step may be made toward unraveling a molecular phylogeny for microbes.

The sequence determination was accordingly undertaken for the c<sub>550</sub> protein. Of immediate concern was that this was necessary to compliment in a reliable manner the already existent 2.45 Å electron density map for the molecule available through an x-ray crystallographic determination. From this map, many of the side chain residues could be identified, although it was impossible to identify all with total confidence. A similar situation in other protein structure determinations has been encountered, in which a crystallographic electron density map has been used to deduce sequence information (Kendrew et al., 1961; Herriott et al., 1973). For the c<sub>550</sub> map, as in these other pure crystallographic approaches, a complete, reliable sequence could not be deduced, which generally would necessitate an independently determined sequence to compliment the x-ray data.

In this chapter, the preparation, isolation, and sequence characterization of the tryptic peptides from c<sub>550</sub> M. denitrificans is described. Subsequent chapters will deal with the combination of these peptide fragments into a full, ordered primary structure, this ordering being deduced from the identification of the unique peptides in the electron density map. This approach is novel in that it combines both chemical and crystal data. The actual sequence of residues determined as described in this paper is merged with the density map serving to align

the peptides, a function normally performed by overlap of peptides from a second and possibly a third limited hydrolysis of the protein or from sequence homology.

### 3.2. Materials and Methods

Purified  $c_{550}$  was the generous gift of Drs. G. McLain and L. Smith, prepared according to their procedure (Scholes, McLain and Smith, 1971). Since the isolation of the cytochrome requires large culture volumes not readily accessible on the laboratory scale, experimentation was designed to complete the tryptic peptide work on the available stock of 10  $\mu$ M.

Three times crystallized preparations of chymotrypsin and *l*-1-tosylamide-2-phenylethyl chloromethyl ketone (TPCK) - treated trypsin were obtained from Worthington Biochemical Corporation, as was a frozen solution of diisopropylphosphorofluoridate - treated carboxypeptidase B. Carboxypeptidase A (crystal suspension), leucine aminopeptidase (ammonium sulfate suspension), pepsin (crystalline), and thermolysin (lyophylyzed powder) were the highest grades available from the Sigma Chemical Company. Carboxypeptidase C (powder) from orange leaves was purchased from the Henley Company of New York.

Reagents for the Edman and dansyl chloride reactions were purchased from Pierce Chemical as "Sequenal" grade. Dowex resin was obtained from Bio-Rad Laboratories, gel filtration media from Pharmacia (Sephadex) and Bio-Rad Laboratories (Bio-gel), and cellulose and silica gel TLC plates from Eastman. Pyridine was

re-distilled before use and anhydrous hydrazine prepared by azeotropic vacuum co-distillation with toluene. All other chemicals were of the highest reagent grade.

Tryptic digestion of 10  $\mu$ M of desalted, oxidized native protein, 0.67  $\text{mM}$  in 0.05  $\text{M}$  ammonium bicarbonate, was affected by 1.2 mole % trypsin for 6.5 hr. at 25°C. The course of the digestion was monitored to assumed completion by cellulose TLC chromatography of 1  $\lambda$  aliquots of the mixture — the release of a high  $R_f$  red heme peptide and the disappearance of a zero  $R_f$  red protein indicating digestion. Chymotryptic digests were performed in 0.1  $\text{M}$  ammonium bicarbonate 0.5  $\text{mM}$  in peptide at mole percentages which will be discussed where pertinent. Thermolytic digestion followed the procedure of Matsubara and Sasaki (1968) at a mole % of 0.2 for 2 hours. All the above reactions were quenched by rapid freezing and lyophilization. Kinetic studies with leucine aminopeptidase (Delange and Smith, 1971), carboxypeptidases A and B (Ambler, 1967) and carboxypeptidase C (Tschesche and Kupfer, 1972) were patterned after established procedures. Partial acid hydrolysis in 0.25  $\text{M}$  acetic acid was performed at 105° *in vacuo* in sealed ampoules at a peptide concentration of 2.6  $\text{mM}$  for 2 hours.

The conditions for the initial separation of tryptic peptides on Dowex 50•X2 have been documented elsewhere (Matsubara and Smith, 1963). Subsequent use of Dowex 50 resin employed the "Macroporous AG-MP50" form in a short column (0.9 x 50 cm) at 40° C developed with (1) 60 ml of pyridine acetate (0.05  $\text{M}$  in pyridine, pH 3.1), (2) 100 ml of pyridine acetate (0.2  $\text{M}$  in pyridine, pH 3.1), (3) a linear

gradient formed by 150 ml of 0.2 M pyridine acetate plus 150 ml of the same buffer (2 M in pyridine, pH 5.1), (4) 100 ml of the latter buffer, and (5) 100 ml of pyridine acetate buffer (4 M in pyridine, pH 5.6). Fractions of 1.5 ml were collected under a pumped flow rate of 15 ml/hr.

Gel filtration columns (0.9 - 2.0 cm × 100 cm) were eluted with 0.1 M ammonium bicarbonate allowing the detection of peptide containing fractions from the absorbance at 220 nm. Preparative paper chromatography and electrophoresis on Whatman 3MM paper have been described (Chan and Margoliash, 1966). Amino acid compositions of peptides were determined from samples run on a Durrum Model D500 or Beckman Model 121B analyzer after in vacuo 24 hour hydrolysis with 6 N HCl. Peptide fingerprinting was performed on cellulose TLC plates with the electrophoresis (first dimension) and chromatography (second dimension) using the same buffers as for preparative work. Appropriate staining reagents, in addition to ninhydrin, as given by Easley (1965) were applied to visualize peptides containing tyrosine, histidine, cysteine, methionine, typtophan or a blocked N-terminus.

Amino-terminal residues were identified by the dansyl method of Gray and Hartley (Hartley, 1970). The combination Edman-dansylation sequencing procedure (Gray, 1972) was reserved for sequential degradation of peptides available in small amounts, as will be noted. The majority of sequencing operations were performed by a manual Edman reaction with direct product identification (Bitar, et. al., 1972). One peptide, derived from the carboxyl terminus of the protein (29 residues long) was sequenced automatically on a Beckman Sequenator Model 890B employing a peptide program. Direct phenylthiohydantoin

amino acids present in sufficient yield (> 50 nM) were identified via published chromatographic methods (Jeppsson and Sjoquist, 1967), and those in low yield on the Durrum analyzer after HI hydrolysis (Smithies et al., 1971).

For the more hydrophobic peptides the terminal-COOH derivatization of Foster et al. (1973) or the lysine derivatization of Braunitzer et al. (1970) were invaluable in attaching a hydrophilic group onto the peptide thus anchoring it into the aqueous phase during organic solvent extraction. No side reactions were evident and the free acid side chains of aspartate and glutamate were still detectable. For example, the hydrophobic peptide T6 (see Results) could be sequenced only four steps before total loss of the remaining material whereas the same starting amount after treatment according to Foster et al. gave high yields on nine of the twelve residues.

As will be noted later, certain atypical standards were required for identification of products encountered. The amino acid 3,5-dichlorotyrosine was synthesized according to Bouchilloux (1955) and characterized via chromatographic properties. The synthesis of 1-acetyl, 2-dansyl hydrazide followed Schmer and Kreil (1969) and the product was confirmed by elemental analysis (Calcd.: 54.50% C, 5.40% H, 13.83% N. Found: 54.30% C, 5.44% H, 13.27% N.)

For complete characterization of the heme containing peptide, the heme was removed by the method of Ambler (1963) and the cysteine residues carboxymethylated as described by Crestfield, Moore, and Stein (1963).



### 3.3. Results

Figure 3.1 outlines the purification of the tryptic peptides and other peptides derived from these by the indicated means, emphasizing, of course, that these were the procedures that ultimately worked, a subset of those attempted.

#### 3.3.1. Nomenclature

For the clarity of reporting results, a consistent nomenclature is adopted for the peptide fractions. Major peptide-containing fractions from the Dowex column on the whole tryptic digest are numbered T1 through T14 in the order of elution. Peptides further purified from these with no further digestion are labeled according to the manner of purification, "ch" designating preparative paper chromatography, "el" paper electrophoresis, and "S" Sephadex gel filtration. Peptides derived from tryptic peptides are indicated by the method of subsequent digestion; "C" for chymotrypsin, "Ac" for acetic acid partial hydrolysis, and "Th" for thermolysin. The final symbol refers to a fraction number of the preceding purification column (eg. C3 - third main fraction off the column used to separate the chymotryptic peptides) or the relative  $R_f$  in chromatography (eg. ch2 has second highest  $R_f$  of the mixture), or the net charge in electrophoresis (+, -, or N= neutral). Thus T11ch1C2 is a peptide from a chymotryptic digest of another peptide purified by paper chromatography after Dowex fractionation. Numbering is not necessarily complete as there were some fractions isolated in insufficient yield for further work or as intermediates between ill-resolved major components.

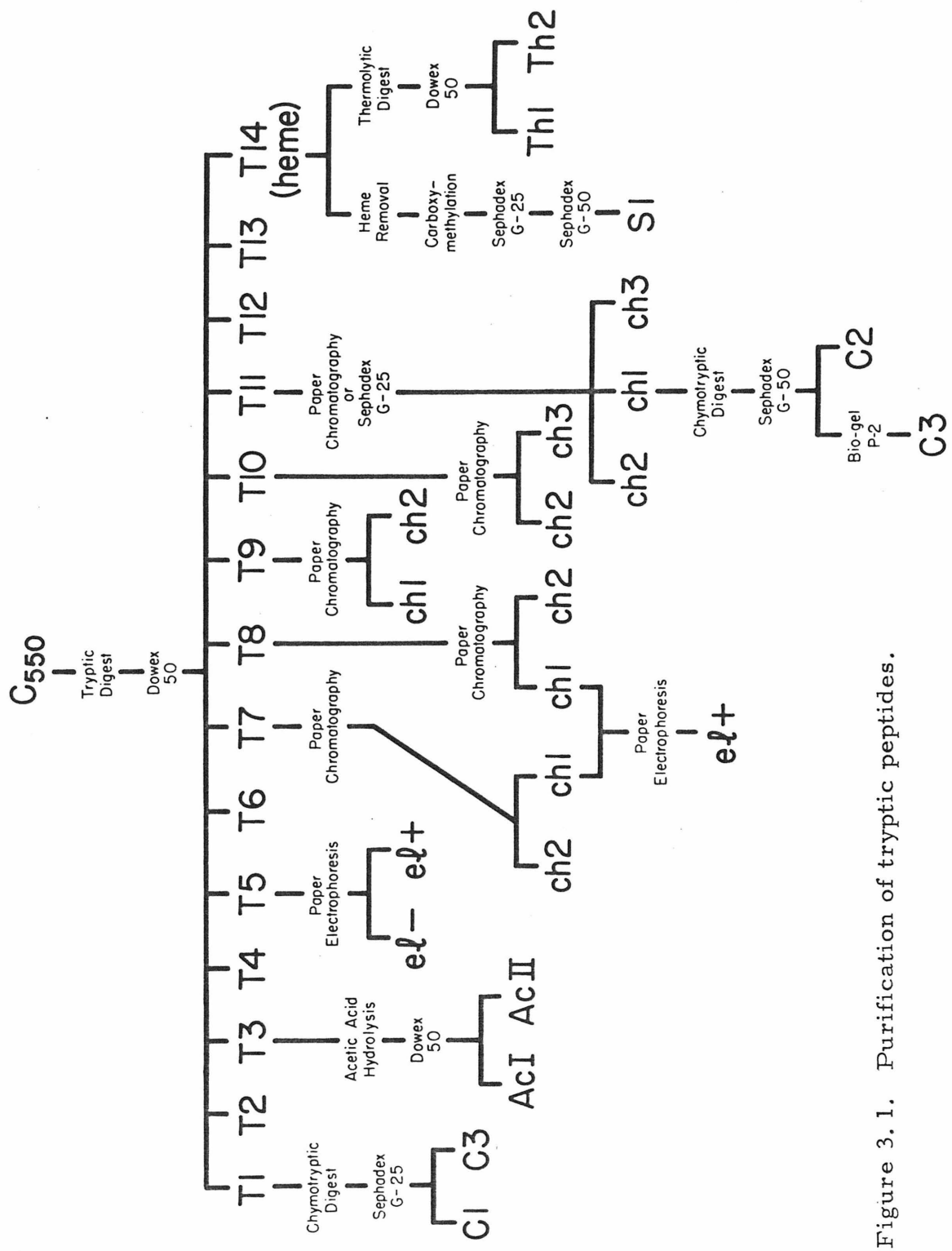


Figure 3. 1. Purification of tryptic peptides.

### 3.3.2. Analytical Results

Figure 3.2 presents a simulated peptide fingerprint map of the tryptic products. Table 3.1 gives the amino acid composition of relevant peptides. The purity of peptide products was judged by the stoichiometry of the residues in the compositions, by peptide fingerprints, and by dansyl end-group analyses. The sum composition of the residues in the unique peptides agrees well with the reported amino acid composition (see Table 3.2) and establishes 135 residues per mole of  $\zeta_{550}$ . The sequence determination of residues within the tryptic peptides is recounted below and the tryptic peptides are listed in Table 3.3.

### 3.3.3. Peptide T1

As this 29 residue peptide is the only major fragment not containing a trypsin specific Lys or Arg, it may be assigned to the carboxyl-terminal portion of the protein. Automatic Edman degradation establishes the sequence Asn-Glu-Ala-Asp-Val-Val-Ala-Phe-Leu-Ala-Gln-Asx-Asx-Pro-Asx-Ala-Gly-Glx-Gly, leaving ten unsequenced residues. Chymotryptic digestion at 0.6 mole% for 22 hours produced only two fragments through specific cleavage of the Phe-Leu bond. Sequence analysis of T1C3 gave Asn-Glu-Ala-Asp-Val-Val-(Ala, Phe). Further sequence analysis of the fragment T1C1 would not yield usable results beyond the tenth residue although a variety of Edman procedures were attempted. Very severe overlap was encountered after each Asn or Asp residue with the dansyl-Edman technique. Gray (1972) has noted this as a diagnostic indicating an  $\alpha$ - $\beta$  peptide shift at Asx residues (see also Bornstein and Balian, 1970).

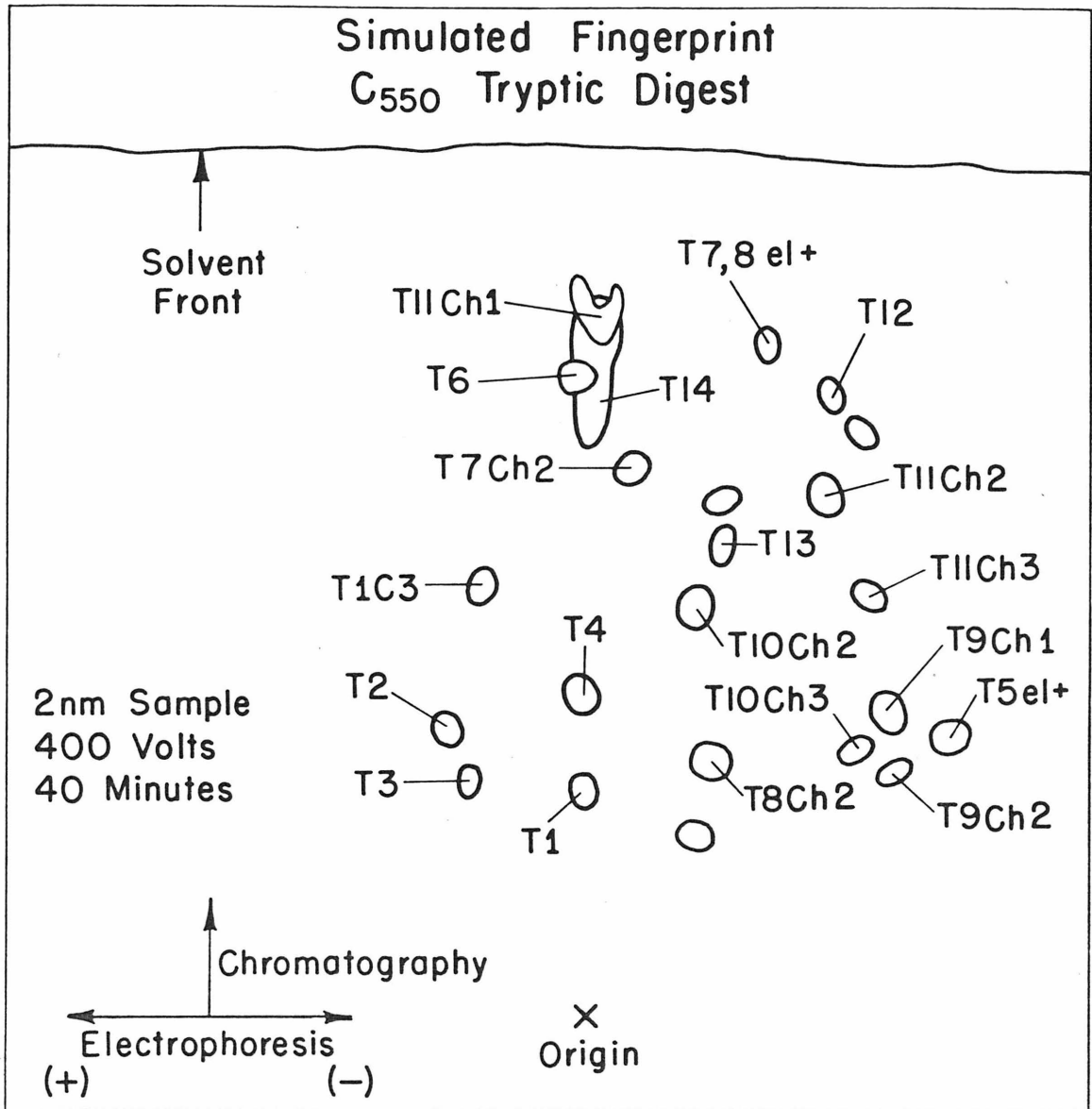


Figure 3.2. Simulated fingerprint. Spot shapes are only approximate. The true line of electroneutrality is displaced 1.7 cm toward the cathode by endosmosis and runs through T8ch2, T10ch2.

TABLE 3. 1

Amino Acid Composition of Isolated Peptides<sup>2</sup>

	T1	T1C1	T1C3	T2
Asp	6.13 (6)	4.04 (4)	1.96 (2)	4.06 (4)
Thr				
Ser	1.83 (2)	1.76 (2)		1.68 (2)
Glu	5.22 (5)	4.24 (4)	1.11 (1)	3.94 (4)
Pro	1.00 (1)	ND (1) <sup>1</sup>		ND
Gly	4.16 (4)	4.11 (4)		4.00 (4)
Ala	7.22 (7)	5.07 (5)	2.05 (2)	4.99 (5)
Cys				
Val	1.57 (2)		1.84 (2)	
Met				
Ile				
Leu	1.02 (1)	0.99 (1)		0.69 (1)
Tyr				
Phe	1.04 (1)		1.04 (1)	
His				
Lys				
Arg				
Trp				
Totals	29	21	8	21
Net charge	(-)	(-)	(-)	(-)

<sup>1</sup>ND, not determined. Subsequent Edman degradation revealed one Pro.  
<sup>2</sup>Integral moles per mole peptide in parenthesis. Net charge was determined from electrophoretic mobility as discussed in the Methods section.

TABLE 3. 1 (continued)

	T3	T3AcI <sup>3</sup>	T3AcII	T4
Asp	1. 91 (2)	1. 18 (1)	0. 94 (1)	2. 13 (2)
Thr				1. 01 (1)
Ser				
Glu	0. 98 (1)	1. 00 (1)		
Pro				
Gly	0. 96 (1)	0. 54	0. 91 (1)	
Ala	2. 22 (2)	0. 23	2. 15 (2)	
Cys				
Val				
Met				0. 56 (1)
Ile				
Leu				
Tyr				
Phe				
His				
Lys	0. 92 (1)		0. 90 (1)	0. 86 (1)
Arg				
Trp				
Total	7	--	5	5
Net charge	(-)	(-)	(0)	(-)

---

<sup>3</sup>Known mixture of peptides. Actual moles of Glu defined as 1. 00 per mole per sample analyzed. See text.

TABLE 3.1 (continued)

	T5e1 <sup>-</sup>	T6	T7, 8ch1e1 <sup>+</sup>	T7ch2
Asp	1.95 (2)		1.06 (1)	
Thr	1.06 (1)		1.04 (1)	
Ser				0.84 (1)
Glu		4.02 (4)		2.13 (2)
Pro			0.97 (1)	
Gly		2.00 (2)	2.95 <sup>4</sup> (3)	1.04 (1)
Ala		1.04 (1)		0.96 (1)
Cys				
Val		0.97 (1)	1.65 <sup>5</sup> (2)	
Met	0.67 (1)			
Ile		0.98 (1)		1.02 (1)
Leu		1.06 (1)	0.97 (1)	
Tyr		1.02 (1)	1.00 (1)	
Phe				1.08 (1)
His				
Lys	0.99 (1)	0.92 (1)		0.93 (1)
Arg			1.24 (1)	
Trp				
Total	5	12	11	8
Net charge	(-)	(-)	(+)	(-)

---

<sup>4</sup>From a 72 hour hydrolysis. Shorter times erroneously indicate 4 gly/peptide. Other values agree between the 24 and 72 hour hydrozylates, except for Tyr which is degraded to 0.4 tyr/peptide.

<sup>5</sup>Edman degradation revealed a Val-Val bond which is resistant to acid hydrolysis.

TABLE 3.1 (continued)

	T8ch2	T9ch1	T9ch2	T10ch2
Asp				0.92 (1)
Thr				0.12
Ser				
Glu	1.00 (1)			0.94 (1)
Pro				
Gly	1.01 (1)	1.03 (1)	2.08 (2)	
Ala		0.96 (1)		
Cys				
Val				
Met				
Ile				
Leu				
Tyr				
Phe				1.01 (1)
His				
Lys	0.99 (1)	1.01 (1)	0.92 (1)	1.13 (1)
Arg				
Trp				
Total	3	3	3	4
Net charge	(0)	(+)	(+)	(0)



TABLE 3.1 (continued)

	T10ch3	T11ch1	T11ch1C3	T11ch2 <sup>7</sup>
Asp		3.88 (4)	0.86 (1)	0.21
Thr	1.06 (1)	3.04 (3)	0.76 (1)	1.13 (1)
Ser				
Glu		2.23 (2)		0.37
Pro		3.00 (3)	1.57 (2)	
Gly				0.82 (1)
Ala		1.09 (1)		
Cys				
Val		1.98 (2)	2.10 (2)	
Met				0.86 (1)
Ile		1.01 (1)		
Leu		2.91 (3)	1.17 (1)	
Tyr		1.03 (1)		
Phe				
His				
Lys	0.94 (1)	2.04 (2)	2.08 (2)	1.19 (1)
Arg				
Trp		1.36 <sup>6</sup> (1)		
Total	2	23	9	4
Net charge	(+)	(-)	(+)	(+)

---

<sup>6</sup>Determined by base hydrolysis according to Noltmann et al. (1962) and placed on the scale of the acid hydrozylates by equating the leucine detected in each method.

<sup>7</sup>Isolated in low yield. Probable met-sulfone form of peptide T12.

TABLE 3.1 (continued)

	T11ch3	T12	T13	T14 (heme) <sup>8</sup>
Asp		0.11	0.14	1.92 (2)
Thr		1.02 (1)	0.10	0.95 <sup>9</sup> (1)
Ser			0.72 (1)	
Glu		0.18	2.16 (2)	0.97 (1)
Pro				1.05 (1)
Gly	1.03 (1)	0.09	1.17 (1)	0.99 (1)
Ala		0.08	0.98 (1)	1.71 (2)
Cys				0.43 (2) <sup>10</sup>
Val			0.21	
Met	0.89 (1)	0.81 (1)		0.73 <sup>9</sup> (1)
Ile		0.07	0.93 (1)	2.07 (2)
Leu		0.08	0.14	
Tyr				
Phe		1.11 (1)	0.88 (1)	
His				0.73 <sup>9</sup> (1)
Lys	1.07 (1)	1.06 (1)	1.88 (2)	1.45 (2)
Arg				
Trp				
Total	3	4	9	16
Net charge	(+)	(+)	(0)	N. D. <sup>11</sup>

<sup>8</sup>96 hour hydrolysis. Subsequent work indicated 2 moles Lys/mole peptide, but only between 1.4 and 1.6 moles Lys could be liberated from the intact heme peptide by any given acid hydrolysis.

<sup>9</sup>Extrapolated to zero time.

<sup>10</sup>Not liberated quantitatively due to covalent attachment to the heme.

<sup>11</sup>Heme dominates electrophoretic behavior and precludes a clear charge determination.

TABLE 3.1 (continued)

	T14S1	T14Th1	T14Th2
Asp	2.13 (2)	1.98 (2)	1.90 (2)
Thr	1.05 (1)	0.95 (1)	0.94 (1)
Ser			
Glu	1.13 (1)	0.97 (1)	1.03 (1)
Pro	ND (1)	1.11 (1)	1.03 (1)
Gly	1.07 (1)	1.02 (1)	0.98 (1)
Ala	1.98 (2)	0.93 (1)	1.00 (1)
Cys	1.72 <sup>12</sup> (2)		
Val			
Met	0.56 (1)		
Ile	2.00 (2)	1.05 (1)	2.12 (2)
Leu			
Tyr			
Phe			
His	0.86 (1)		
Lys	1.78 (2)		0.68 (1)
Arg			
Trp			
Total	16	8	10
Net charge	(-) <sup>13</sup>	(-)	(-)

---

<sup>12</sup>Detected as carboxymethylcyteine after heme removal and treatment with iodoacetate.

<sup>13</sup>Carboxymethyl groups add two negative charges.

---

---

TABLE 3.2  
Amino Acid Composition of  $c_{550}$

	From Tryptic Peptides	From Intact Protein <sup>1</sup>
Asp	18	18
Thr	8	8
Ser	3	3
Glu	17	16
Pro	6	8
Gly	17	17
Ala	15	15
Cys	2	2
Val	7	7
Met	4	4
Ile	5	6
Leu	6	6
Tyr	3	3
Phe	4	4
His	1	1
Lys	17	16
Arg	1	1
Trp	1	ND
Totals	135 <sup>3</sup>	136 <sup>2</sup>

---

<sup>1</sup>Scholes, McLain, and Smith, 1971.

<sup>2</sup>Assuming 1 Trp.

<sup>3</sup>Assuming 2 moles free Lys per tryptic digest.

---

---

TABLE 3. 3

Sequence of Tryptic Peptides

T1:	Asn-Glu-Ala-Asp-Val-Val-Ala-Phe-Leu-Ala-Gln-Asx -Asx-Pro-Asx-Ala-Gly-Glx-Gly-(Asp <sub>1</sub> , Ser <sub>2</sub> , Glx <sub>2</sub> , Gly <sub>2</sub> , Ala <sub>3</sub> )
T3:	Ac <sup>1</sup> -Asn-Glu-Gly-Asp-Ala-Ala-Lys
T4:	Met-Thr-Asp-Asp-Lys
T5:	Lys <sup>2</sup>
T6:	Tyr-Gly-Glu-Gly-Ile-Leu-Glu-Val-Ala-Glu-Glu-Lys
T7, 8ch1el <sup>+</sup> :	Thr-Gly-Pro-Asn-Leu-Tyr-Gly-Val-Val-Gly-Arg
T7ch2:	Ile-Ala-Ser-Glu-Glu-Gly-Phe-Lys
T8ch2:	Gly-Glu-Lys
T9ch1:	Gly-Ala-Lys
T9ch2:	Gly-Gly-Lys
T10ch2:	Glu-Phe-Asn-Lys
T10ch3:	Thr-Lys
T11ch1:	Asn-Pro-Asp-Leu-Thr-Trp-Thr-Glu-Ala-Asn-Leu-Ile -Glu-Tyr-Val-Thr-Asp-Pro-Lys-Pro-Leu-Val-Lys
T11ch3:	Met-Gly-Lys
T12:	Met-Thr-Phe-Lys
T13:	Lys-Ile-Ala-Ser-Glu-Glu-Gly-Phe-Lys
T14:	Cys-Lys-Ala-Cys-His-Met-Ile-Gln-Ala-Pro-Asp-Gly -Thr-Asp-Ile-Lys

---

<sup>1</sup>Ac is the acetyl group.

<sup>2</sup>Two moles.

---

Severe decreases in yield were also noted at Gln or Glx residues, attributable to the formation of pyroglutamic acid. Finally, cleavage was noticeably slower at Gly residues in keeping with previous observations (Konigsberg and Hill, 1962). As these three problem amino acids account for over 50% of the peptide, lack of complete Edman degradation is not surprising.

Further attempts to generate usable shorter fragments were unsuccessful. The peptide T1C1 was resistant to digestion with pepsin. Papain at a variety of concentrations, times, and pH produced at least 14 major sub-fragments of similar chromatographic and electrophoretic mobilities. Partial acid hydrolysis for periods as short as 2 hours still produced 16 major sub-fragments, which, by a combination of Dowex 50 and Dowex 1 chromatography, could not be purified in sufficient yield for further characterization.

The successful completion of the sequence of this peptide must await the isolation of sufficient quantities of material for extensive purification of the nonspecific sub-fragments.

#### 3.3.4. Peptide T2

Direct Edman degradation established the initial sequence Leu-Ala-Gln-Asn and comparison with T1 indicated this peptide was derived from the carboxyl-terminal region of the larger fragment by cleavage between Phe-Leu. The yield of purified peptide was 10% and the amino-terminal fragment was not recovered. No other chymotryptic-like cleavages were observed. The supposed anomalous hydrolysis by TPCK-treated trypsin of a Try-Leu bond in c-type

cytochromes has been observed in several previous instances, (Nolan and Margoliash, 1966; Chan and Margoliash, 1966; and Matsubara and Smith, 1963) suggesting that an aromatic to hydrophobic linkage or some structural feature of the intact protein makes this bond susceptible to tryptic action. From the known three-dimensional structures for the eukaryotic cytochromes (Takano et al., 1973) and, as will be shown, the known structure of this protein, the Tyr-Leu (or Phe-Leu) bond occurs near the center of a long  $\alpha$ -helix, with other tryptic cleavage points at least 6-8 residues removed. Thus during enzymatic digestion of the whole protein, the segment may retain a helical conformation and this structural arrangement may in some way prime the bond toward tryptic hydrolysis.

### 3.3.5. Peptide T3

This pure peptide failed to give a ninhydrin reaction although it could be visualized with the peptide-bond stain of  $\text{Cl}_2$ /starch/KI. Edman degradation, dansylation, and incubation with leucine aminopeptidase failed to liberate a free amino-terminal residue. As the intact protein concurred in these negative results, it was concluded that this peptide represented the blocked amino-terminal segment of the protein.

The blocking agent was directly identified as an acetyl group through the hydrazinolysis method of Schmer and Kreil (1969). The 1-acetyl, 2-dansyl hydrazide produced in the reaction was identified through chromatography on aluminum oxide TLC sheets (neutral, 200 $\mu$ ; Merck Co.) in the Schmer and Kreil solvent systems versus an

authentic standard. Control experiments were performed on acetylated tuna cytochrome  $c$  and quantitation was assured through relative fluorescent intensities of product and standard.

The blocked peptide at 1  $\text{mM}$  was degraded in 0.25  $\text{N}$  acetic acid for 2 hours at  $105^\circ \text{C}$  in vacuo, conditions reputed to minimize loss of amide groups (Schroeder et al., 1963). The resultant mixture was purified on Dowex 50 to give two main fractions, AcI and II plus other minor mixtures. Ac II was a pure peptide and dansyl-Edman degradation established the first two residues as Gly-Asx. From the composition and known trypsin specificity, this sequences the peptide as Gly-Asx-Ala-Ala-Lys. The net charge on the peptide under electrophoresis at pH 6.5 was neutral implying an acid at the second position.

Fraction AcI demonstrated three ninhydrin-negative peptides under fingerprinting, two of which were identified as undigested T3 and acetyl-asparagine through comparison with standards. Amino acid composition studies indicated mainly Asx and Glx in roughly stoichiometric amounts. This, together with the composition difference between T3 and T3AcII implies acetyl-Asn-Glx as the main component of AcI. The parent peptide T3 carried a net negative charge, and in electrophoresis was more mobile than the -1 charged subfragments of fraction AcI. As its mobility was clearly greater than that of reference peptides of similar molecular weight and charge -1, it must carry the maximum charge which is -2 consistent with the directly determined acetyl-asparagine group. Then the complete sequence must be acetyl-Asn-Glu-Gly-Asp-Ala-Ala-Lys. Incubation of AcI with carboxypeptidase C did not liberate any amino acid, but this lack of reaction may be due to the short length of the peptide. Minor fragments of uncertain purity indicated free glycine, free alanine, (Ala<sub>2</sub>, Lys) and (Asx, Gly, Ala) in some support of the above assignments.



### 3.3.6. Peptide T4

Four direct Edman steps establish the sequence shown in Table 3.3. Sequence and compositions established this peptide as identical to T5e1<sup>-</sup>.

### 3.3.7. Peptide T5e1<sup>-</sup>

Fingerprint comparisons and amino acid analysis show this to be free lysine. Relative yields from the Dowex column are not helpful for establishing the yield of free lysine as fractions were selected for purity, not quantitation. A known molar amount of  $\underline{c}_{550}$  was digested with trypsin and the hydrozylate applied directly to the Beckman amino acid analyzer. Free lysine was detected at the level of 0.7 moles Lys/mole digest.

Subsequent results from the x-ray work will demonstrate the presence of both a Lys-Lys and an Arg-Lys bond in  $\underline{c}_{550}$ , which after maximum digestion would give 2 moles Lys/mole digest in disagreement with the above experiment. One grouping only was anticipated from the free lysine analyzed and the results of sequencing peptides T7ch2 and T13. The discrepancy may arise from the presence of extensive minor peptides Lys-x...x-Lys or x-...-Lys-Lys which were not recovered or an inaccuracy in the 0.7 mole determination. This latter point is quite possible as the peptide mixture applied to the analyzer gave a high background to the analysis and may have distorted the free lysine elution pattern.

### 3.3.8. Peptide T6

After the modification of Foster et al. (1973) direct Edman degradation gave Tyr-Gly-Glu-Gly-Ile-Leu-Glu-Val-Ala-(Glx<sub>2</sub>, Lys).

The net negative charge of the peptide is not sufficient to establish the side chain functions on the remaining two Glx residues as the already determined two Glu residues would account for the charge. Complete enzymatic hydrolysis by first papain then leucine aminopeptidase (Margoliash et al., 1962) released 4 moles Glu/mole peptide and no Gln. The placement of Lys follows from trypsin specificity and completes the sequence shown in Table 3. 3.

3. 3. 9. Peptide T7ch2

PTH-amino acids were observed for seven of the eight residues, all in agreement with the peptide net charge.

3. 3. 10. Peptide T8ch2

All three residues were observed as PTH amino acids.

3. 3. 11. Peptide T7, 8ch1e1<sup>+</sup>

As outlined in Figure 3.1 a major peptide had been isolated as an intermediate, split between two other consecutive major Dowex fractions. These were partially purified, recombined, then further purified to a homogeneous state. After Foster modification, ten of the eleven residues were identifiable as PTH-amino acids with Arg firmly fixed by specificity.

3. 3. 12. Peptides T9ch1, T9ch2 and T10ch2

In each case, complete sequencing was affected by Edman degradation.

3. 3. 13. Peptide T10ch3

Dansylation and composition complete this simple di-peptide.

3. 3. 14. Peptide T11ch1

By direct degradation, the following was established:

Asn-Pro-Asp-Leu-Thr-Trp-Thr-Glu-Ala-Asn-Leu-Ile-Glx-Tyr-Val-Thr-Asx. The triad Thr-Trp-Thr was detected in very low yields. Uphans, Grossweiner and Katz (1959) have noted that a combination of O<sub>2</sub> and trifluoroacetic acid leads to an X-Trp linkage rearranging to 3-carboxy- $\beta$ -carboline and this may explain the low yield.

When this peptide was purified by paper chromatography, the tyrosine present was found to be converted to 3,5 dichlorotyrosine which was detected as such by comparison with an authentic, synthetic standard. The formation of this amino acid in a side reaction involving chloride and traces of oxidizing agent has been reported (Sanger and Thompson, 1963). When the same peptide was purified via gel filtration, tyrosine was quantitatively recovered as the natural amino acid.

The long peptide was hydrolyzed in a pulse digestion experiment with 0.3 mole % chymotrypsin for 3 hours. One pure peptide, T11ch1C3, was isolated and analyzed to show a sequence Val-Thr-Asp-Pro-Lys-Pro-(Leu, Val, Lys). The sequence overlap with T11ch1 establishes this as derived from the carboxyl terminus of T11ch1. Kinetic digestion of T11ch1 with carboxypeptidases A and B indicated -Leu-Val-Lys.

Fingerprinting revealed another main fraction, T11ch1C2, to be composed of just undigested T11ch1 plus a new fragment. Dansylation indicated a single amino-terminus for the fraction suggesting the new fragment to be derived from the initial portion of T11ch1. Sequence analysis showed unique products through Asn-Pro-Asp-Leu-Thr-Trp. Carboxypeptidase A, which would not liberate the terminal Lys on T11ch1, indicated the new fragment to end in Tyr, establishing it as the complement to T11ch1C3.

Because of low yields, the Edman procedure could not definitively establish the charge of Glx 13 (peptide numbering). There are sufficient acid groups on peptides T11ch1 and T11ch1C2 to account for net peptide charge. The molecular weight of T11ch1 was so large that electrophoretic migration was not able to indicate clearly a net peptide charge of -1 or -2, corresponding to Gln 13 or Glu 13. There was insufficient remaining material for an accurate enzymatic digestion. The assignment of 13 as Glu was made solely on the basis of the failure of carboxypeptidase A to liberate Gln-Tyr in kinetic digestion experiments.

### 3. 3. 15. Peptides T11ch3, T12

All but the terminal Lys residues were directly detected in the Edman reaction.

### 3. 3. 16. Peptide T13

This peptide differed from T7ch2 only in having an amino-terminal Lys as well as the carboxyl-terminal Lys. The previously mentioned presence of free Lys and this peptide are consistent with an Arg-Lys and/or a Lys-Lys linkage. The lack of hydrolytic reactivity of an amino-terminal Lys has been documented (Kasper, 1970).

### 3.3.17. Peptide T14 (heme peptide)

The Edman reaction would not liberate any product until after application of the heme-removal reaction and carboxymethylation. The sequence then indicated was (Cys was detected as carboxymethyl cysteine): Cys-Lys-Ala-Cys-His-Met-Ile-Glx-Ala-Pro. The presence of the typical cytochrome heme linkage, -Cys-X-X-Cys-His-, and the applicability of the heme removal reaction are taken as indicating the normal manner of covalent heme-protein attachment, through thioether bonds on the two Cys residues.

Thermolytic digestion presented four major peptides. The highly insoluble heme peptide was not recovered from the ion exchange column. An Ile-Lys fragment and peptides T14Th1 and Th2, which differ in composition (Table 3.1) only by (Ile, Lys) were isolated as pure products. Direct or dansyl-Edman degradation of Th2 yielded Ile-Gln-Ala-Pro-Asx-Gly-Thr-Asx-(Ile, Lys). A dansyl-Edman procedure established T14Th1 as equivalent in sequence to T14Th2 but lacking the terminal Ile, Lys residues. The yield after the Pro step dropped markedly. A partial explanation may be that an  $\alpha \rightarrow \beta$  peptide bond shift is occurring at the succeeding Asx residue in a reaction similar to that discussed for peptide T1.

Treatment of intact peptide T14 with carboxypeptidases B and A established the terminal sequence as -Ile-Lys but failed to liberate any other amino acid in spite of high enzyme concentrations and long digestion times. Such a reactivity failure can be rationalized by the presence of Asp as the next residue in keeping with the final Edman product Asx of T14Th2. Both Th1 and Th2 carry net negative charges

and as Th2 adds one positive charge (Lys) onto Th1, the magnitude on Th1 must be -2 and hence both Asx residues are Asp.

In summation, the peptide T14 was sequenced as Cys-Lys-Ala-Cys-His-Met-Ile-Gln-Ala-Pro-Asp-Gly-Thr-Asp-Ile-Lys.

### 3.4. Discussion

The sequence of all tryptic peptides from c<sub>550</sub> (M. denitrificans) has been determined (Table 3.3) with the exception of the final ten residues at the protein carboxyl-terminus. The failure to complete these residues has been ascribed to side reactions inherent in a peptide of this composition, namely, an  $\alpha \rightarrow \beta$  peptide shift at aspartyl groups and extensive formation of pyroglutamyl blocking groups. The completion of the sequence is awaiting the isolation of sufficient quantities of c<sub>550</sub> protein to perform limited, non-specific hydrolysis and successful isolation of the fragments. The sequence of other tryptic peptides has been established by conventional techniques.

The placement of amide side chains has followed either direct identification as PTH products or simple arguments based upon net peptide charge. The number of amide groups so found raises an interesting point concerning the isoelectric point of the protein. Scholes, McLain, and Smith (1971) have reported a pI of approximately 4 and McLain (personal communication) has remeasured this number by gel and column isoelectric focusing as 4.5. From the amino acid composition of the tryptic peptides, considering the number of amide groups, an approximate isoelectric point may be calculated (Edsall and Wyman, 1958). The calculation is somewhat complicated by a

lack of knowledge concerning the amides in the unsequenced carboxyl terminal region. Nevertheless, a minimum pI can be calculated to be 4.6 (17 Lys + 1 Arg = 18 positive charges. The only His is involved in liganding. 13 Glu + 12 Asp + 2 propionic acids = 27 negative charges) by assuming that all the unknowns are acidic functions. The agreement would suggest that most of the remaining unsequenced groups are indeed acids.

The isoelectric point of this molecule may ultimately be a matter of some importance in view of the known cross-reactivity between the M. denitrificans electron transport chain and that of mitochondria (Scholes, McLain, and Smith, 1971). Mitochondrial cytochrome c carries a high isoelectric point, greater than 10, and it has been shown that the positive charges are involved in binding to either cytochrome oxidase or reductase (Smith and Minnaert, 1965). That c<sub>550</sub> cross-reacts with cytochrome c clearly emphasizes that the charge effect must be due to a grouping of charge, rather than a net charge effect. In subsequent work, it will be shown that c<sub>550</sub> contains a severe localization of its positive charges into a limited area on the surface of the molecule.

The reconstruction of the primary structure of the intact protein follows from a knowledge of the sequence of the tryptic peptides and the availability of a high resolution electron density map; this result will be the subject of a later chapter. The successful reconstruction follows from the constraints obtained through the traditional sequencing of the peptides reported here, for by knowing the tryptic peptides, the interpretation of the x-ray map requires only a matching of some

sixteen well characterized fragments to their unique location in the map and hence into sequence.



### 3.5. References

- Ambler, R. P., (1963). Biochem. J., 89, 349.
- Ambler, R. P., (1967). In Methods in Enzymology, XI, C. H. W. Hirs, ed., Academic Press, New York.
- Bitar, K., Vinogradov, S. N., Nolan, C., Weiss, L. J., and Margoliash, E., (1972). Biochem. J., 129, 561.
- Bornstein, P., and Balian, G., (1970). J. Biol. Chem., 245, 4854.
- Bouchilloux, S., (1955). Bull. Soc. Chim. Biol., 37, 255.
- Braunitzer, G., Schrank, B., and Ruhfus, A., (1970). Hoppe-Seyler's Z. Physiol. Chem., 351, 1589.
- Chan, S. K., and Margoliash, E., (1966). J. Biol. Chem., 241, 335.
- Crestfield, A. M., Moore, S., and Stein, W. H., (1963). J. Biol. Chem., 238, 622.
- Dayhoff, M. O. (1972). Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington, D. C.
- Delange, R. J., and Smith, E. L., (1971). In The Enzymes, v. III, 3rd ed., P. D. Boyer, ed., Academic Press, New York.
- Dickerson, R. E., (1971). J. Mol. Biol., 57, 1.
- Easley, C. W., (1965). Biochim. Biophys. Acta, 107, 386.
- Edsall, J. T., and Wyman, J., (1958). Biophysical Chemistry, 1, Academic Press, New York, p. 509-510.
- Foster, J. A., Bruenger, E., Hu, C. L., Albertson, K., and Franzblau, C., (1973). Biochem. Biophys. Res. Commun., 53, 70.
- Gray, W. R., (1972). In Methods in Enzymology, XXV, 333, C. H. W. Hirs, ed., Academic Press, New York.
- Hartley, B. S., (1970). Biochem. J., 119, 805.
- Herriott, J. R., Watenpaugh, K. D., Sieker, L. C., and Jensen, L. H., (1973). J. Mol. Biol., 80, 423.
- Jeppsson, J., and Sjoquist, J., (1967). Anal. Biochem., 18, 264.
- Kasper, C. B., (1970). In Protein Sequence Determination, S. B. Needleman, ed., Springer-Verlag, New York, p. 155.

- Kendrew, J. C., Watson, H. C., Strandberg, B. E., Dickerson, R. E., Philips, D. C., and Shore, V. C., (1961). Nature, 190, 663.
- Koningsberg, W., and Hill, R. J., (1962). J. Biol. Chem., 237, 2547.
- Margoliash, E., Kimmel, J. R., Hill, R. L., and Schmidt, W. R., (1962). J. Biol. Chem., 237, 2148.
- Margoliash, E., Fitch, W. M., and Dickerson, R. E., (1968). Structure, Function, and Evolution in Proteins; Brookhaven Symp. in Biol., 21, 259.
- Matsubara, H., and Smith, E. L., (1963). J. Biol. Chem., 238, 2732.
- Matsubara, H., and Sasaki, R. M., (1968). J. Biol. Chem., 243, 1732.
- Needleman, S. B., and Blair, T. T., (1969). Proc. Nat. Acad. Sci., 63, 1227
- Nolan, C., and Margoliash, E., (1966). J. Biol. Chem., 241, 1049.
- Noltman, E. A., Mahowald, T. A., and Kuby, S. A., (1962). J. Biol. Chem., 237, 1146.
- Sanger, F., and Thompson, E. O. P., (1963). Biochim. Biophys. Acta, 71, 468.
- Schmer, G., and Kreil, G., (1969). Anal. Biochem., 29, 186.
- Scholes, P. B., McLain, G., and Smith, L., (1971). Biochemistry, 10, 2072.
- Schroeder, W. A., Shelton, J. R., Shelton, J. B., Cormick, J., and Jones, R. T., (1963). Biochemistry, 2, 992.
- Smith, L., and Minnaert, K., (1965). Biochim. Biophys. Acta, 105, 1.
- Smith, L., Newton, N., and Scholes, P., (1966). In Hemes and Hemoproteins, B. Chance, R. W. Estabrook, and T. Yonetoni, eds., Academic Press, New York.
- Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G., and Ballantyne, D. L., (1971), Biochemistry, 10, 4812.
- Takano, T., Kallai, O. B., Swanson, R., and Dickerson, R. E., (1973). J. Biol. Chem., 248, 5234.
- Tschesche, H., and Kupfer, S., (1972). Eur. J. Biochem., 26, 33.
- Uphans, R. A., Grossweiner, L. I., and Katz, J. J., (1959). Science, 129, 641.

## CHAPTER 4. TERTIARY STRUCTURE OF C<sub>550</sub>

### 4.1. Ordering the Tryptic Peptides

With the tryptic peptides of known sequence, the available x-ray map could be interpreted to align the fragments into a complete primary structure. The ordering is normally obtained through a second or possibly third enzymatic digestion of the protein, or occasionally by sequence homology to some previously determined and highly similar protein. The merging of x-ray map and peptide sequence data is quite novel, as previous sequence assignments from crystal structure data have been based solely on a serial identification of the individual amino acids from the electron density (Kendrew, et.al., 1961; Herriott, et. al., 1973). While the pure x-ray approach produces correct assignments at a rate of 50% to 75%, a reliable, complete sequence has never been produced in this manner.

The method of aligning the peptides by electron density was quite straightforward, utilizing nothing more complex than the approach one adopts when building a jigsaw puzzle. At the resolution of the map used, 2.45 Å, the main chain could be easily differentiated from the side chain groups. When first considering the array of peptides and the map, concentration was focused on patterns in the size of the side chain groups and a similar pattern was sought in the density as it was presented in a Richards box. As with any puzzle, if the choice did not reveal a clear match, another was made. Once a candidate had been found, it was confirmed by carefully fitting a scaled Kendrew model to the box density.

In this process, it was not necessary to attempt to be linear, that is, to generate a sequence from the amino to the carboxyl terminus. Rather, a strong sense of checks and balances occurred when the fitted pieces, scattered about, finally were found to be joined by the remaining pieces.

Certain residues proved more helpful than others as guideposts (Table 4.1). Aromatics were especially clear and suggested many of the initial guesses. Glycine and alanine were conspicuous by an absence of density in their locale. Asparagine/aspartate, serine, valine and threonine were the most difficult to match; being of moderate size and shape, they offered no distinguishing characteristics. Histidine, cysteine, and arginine are ranked low in Table 4.1 simply because  $\underline{c}_{550}$  contains so few of these residues. In another protein, they should be quite easy to locate.

Table 4.2 reports the total primary structure assigned to  $\underline{c}_{550}$ . From Chapter 3 it will be recalled that the final ten residues could not be sequenced chemically. The arrangement reported in Table 4.2. represents the best guess possible from the map alone. Unfortunately, this is the poorest region of the map in which to make a pure x-ray sequence assignment. The terminal eight residues appear badly disordered in the map to the extent that several conformations appear possible. This conformation averaging smears the density and weakens it so that side chain identification is highly dubious.

The remainder of the sequence is presented with total confidence. The constraints imposed by knowing from the chemical work, the size

---

TABLE 4.1

ASSESSMENT OF RESIDUES

Trp

Tyr

Phe

Met

Pro

Lys

Gly

Ala

Ile

Glx

Leu

↓ decreasing clarity

Asx

Ser

Val, Thr

(His, Cys, Arg)

---

TABLE 4.2

M. DENITRIFICANS C<sub>550</sub> SEQUENCE

1	5	10
Ac - Asn - Glu - Gly - Asp - Ala - Ala - Lys - Gly - Glu - Lys		
←----- τ <sub>3</sub> -----→ ←----- τ <sub>8ch2</sub> -----→		
11	15	20
- Glu - Phe - Asn - Lys - Cys - Lys - Ala - Cys - His - Met		
←----- τ <sub>10ch2</sub> -----→ ←-----		
21	25	30
- Ile - Gln - Ala - Pro - Asp - Gly - Thr - Asp - Ile - Lys		
←----- τ <sub>14</sub> -----→		
31	35	40
- Gly - Gly - Lys - Thr - Gly - Pro - Asn - Leu - Tyr - Gly		
←----- τ <sub>9ch2</sub> -----→ ←----- τ <sub>7sch1*</sub> -----→		
41	45	50
- Val - Val - Gly - Arg - Lys - Ile - Ala - Ser - Glu - Glu		
←----- τ <sub>13</sub> -----→		
51	55	60
- Gly - Phe - Lys - Tyr - Gly - Glu - Gly - Ile - Leu - Glu		
←----- τ <sub>6</sub> -----→		
61	65	70
- Val - Ala - Glu - Glu - Lys - Asn - Pro - Asp - Leu - Thr		
←-----		
71	75	80
- Trp - Thr - Glu - Ala - Asn - Leu - Ile - Glu - Tyr - Val		
←----- τ <sub>11ch1</sub> -----→		
81	85	90
- Thr - Asp - Pro - Lys - Pro - Leu - Val - Lys - Lys - Met		
←-----		
91	95	100
- Thr - Asp - Asp - Lys - Gly - Ala - Lys - Thr - Lys - Met		
←----- τ <sub>4</sub> -----→ ←----- τ <sub>9ch1</sub> -----→ ←----- τ <sub>10ch3</sub> -----→		
101	105	110
- Thr - Phe - Lys - Met - Gly - Lys - Asn - Gln - Ala - Asp		
←----- τ <sub>12</sub> -----→ ←----- τ <sub>11ch3</sub> -----→ ←-----		
111	115	120
- Val - Val - Ala - Phe - Leu - Ala - Gln - Asx - Asx - Pro		
←----- τ <sub>1</sub> -----→		
121	125	130
- Asx - Ala - Gly - Glx - Gly - Glx - Ala - Ala - Gly - Ala		
←-----		
131	135	
- Gly - Ser - Asx - Ser - Glx		
←-----		

Amino-terminus is acetylated. 135 residues.

and shape of even short segments of the molecule are sufficient to generate a reliable structure.

Analogous to pure chemical sequencing, most information was provided by the longer peptides. The assignment of fragment Gly-Gly-Lys to positions 31-33 and Gly-Ala-Lys to positions 95-97 illustrates the most difficult decision which was made. The respective placements rest solely on the slight but definite methyl density seen at position 96. As long as peptides exhibit at least this amount of structural differentiation, and all others by far did, map density is sufficiently unique to determine structure unambiguously.

The multiple positive charge groupings -Lys<sub>88</sub> - Lys<sub>89</sub> - and -Arg<sub>44</sub> - Lys<sub>45</sub> - represent information derived more or less exclusively from the x-ray map. The presence of one such linkage was suggested (not proved) by finding free lysine in the tryptic digest and a tryptic peptide with an amino-terminal lysine. (See peptide T13, Chapter 3.) Ultimately, the suggestion was fulfilled by locating in the density the -Arg-Lys- linkage. The remaining group was inherently not detectable in the peptide chemistry initiated by trypsin hydrolysis. In fitting the map, the preceding peptide -Pro-Leu-Val-Lys and the succeeding peptide -Met-Thr-Asp-Asp-Lys- were matched quite well with the density, but, a gap remained, one residue long and with a side chain density indicative of Lys. Hence, a consideration of the map and the origin of the other peptides established the correct sequence as involving a double Lys bond.

#### 4.2. Molecular Structure – General

Perutz and co-workers have spent more than ten years on the hemoglobin molecule and yet, parts of the fine structure are still not clear. So this author finds no hesitancy in reporting that the one year old  $c_{550}$  structure is not complete. The answers are there; the problem has been to interpret correctly the less-than-atomic resolution map. The interpretive ambiguity and the inherent mechanical and optical inadequacies of Richard's Box techniques, referred to as model building limitations, are sufficiently severe that fine structure – placement of peptide amide planes, exact hydrogen bonding, salt bridge patterns, Van der Waals contacts, etc. – must be taken critically realizing that most are correct, but some potentially wrong.

What will be discussed here are only those results which are firmly established. The available information is rich in detail but is only effectively communicated in three dimensions. For the present purpose only some relevant highlights are offered, as a prelude to some conclusions which will be drawn concerning cytochrome evolution and mechanism.

Figures 4.1, 4.2 and 4.3 present three contrasting  $\alpha$ -carbon diagrams of the  $c_{550}$  structure and Figure 4.4 presents a comparison between reduced tuna  $c$  (Takano, et al., 1973) and  $c_{550}$ . Figure 4.5 summarizes the basic result of the entire structure determination. In brief,  $c_{550}$  is a molecule composed of a core structure virtually identical with the smaller eukaryotic cytochrome but with major insertions and a few minor deletions. These changes obey a consistent rule of occurring at



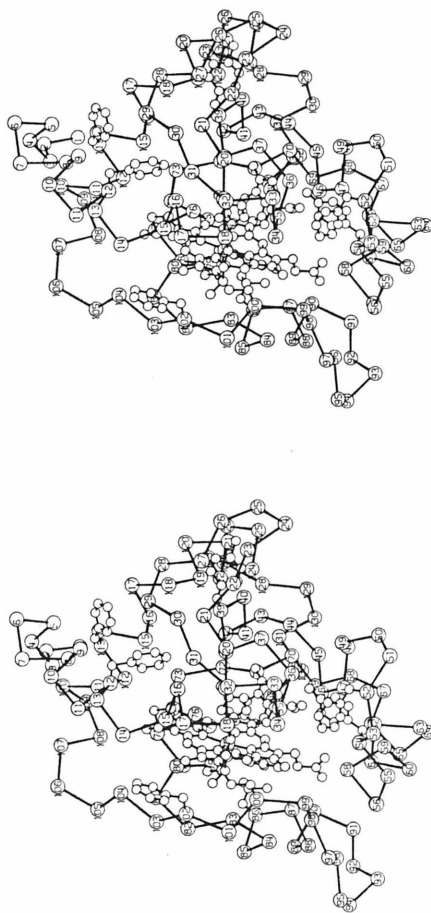


Figure 4. 1. An  $\alpha$ -carbon stereo diagram of  $C_{550}$ . Heavy bonds represent the distances between consecutive  $\alpha$ -carbons. The light line bonds represent true chemical bonding in the side chains and the heme. On a scale of 0.0559 inches/ $\text{\AA}$ .

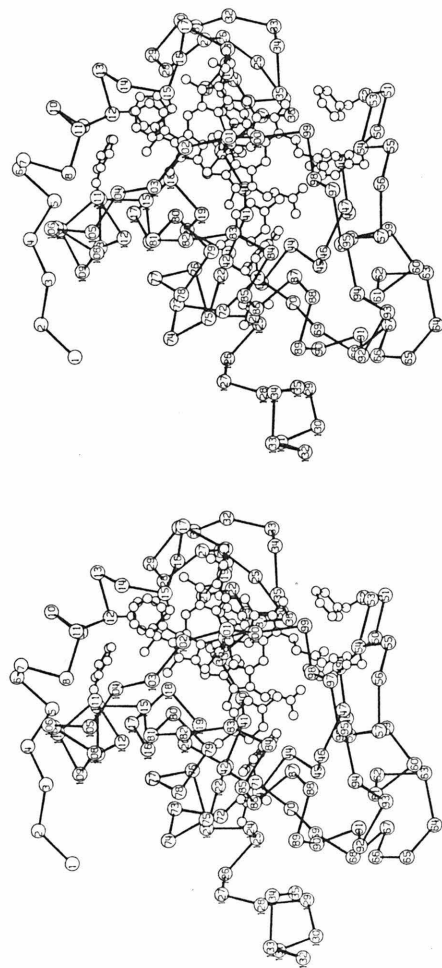


Figure 4.2. As in 4.1, rotated 90° about the vertical axis.

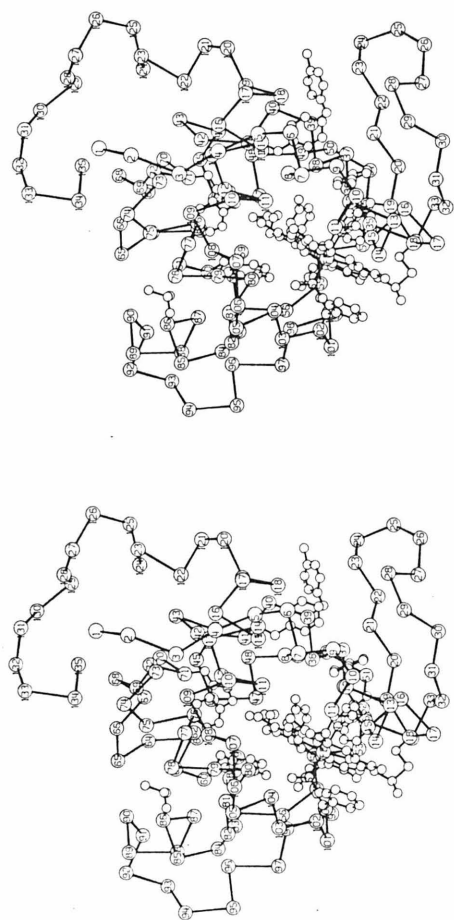


Figure 4.3. As in 4.1, rotated 90° about the horizontal axis.

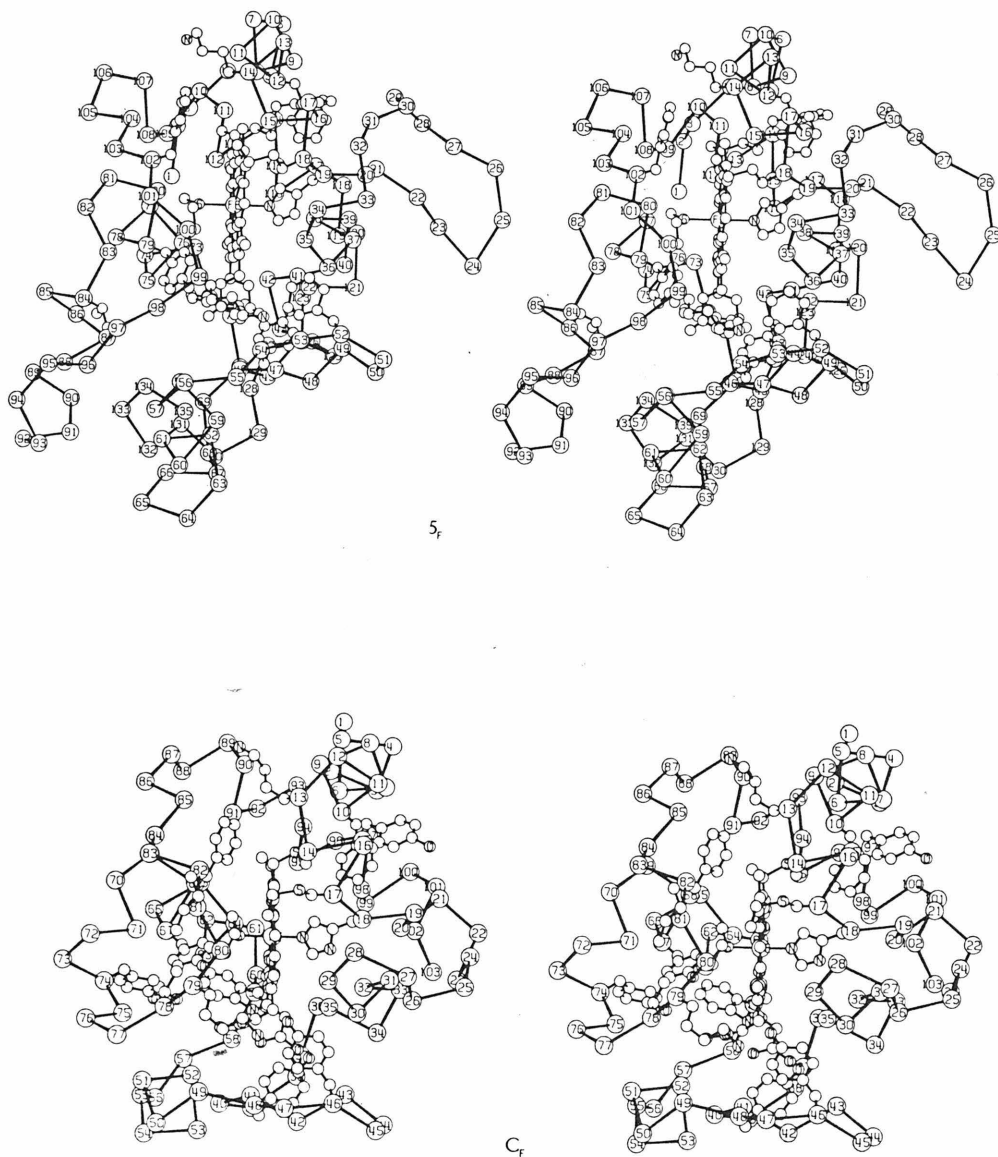


Figure 4.4.  $c_{55b0}$  ( $5_F$ ) and  $tuna\ c$  ( $C_F$ ) in equivalent orientation.

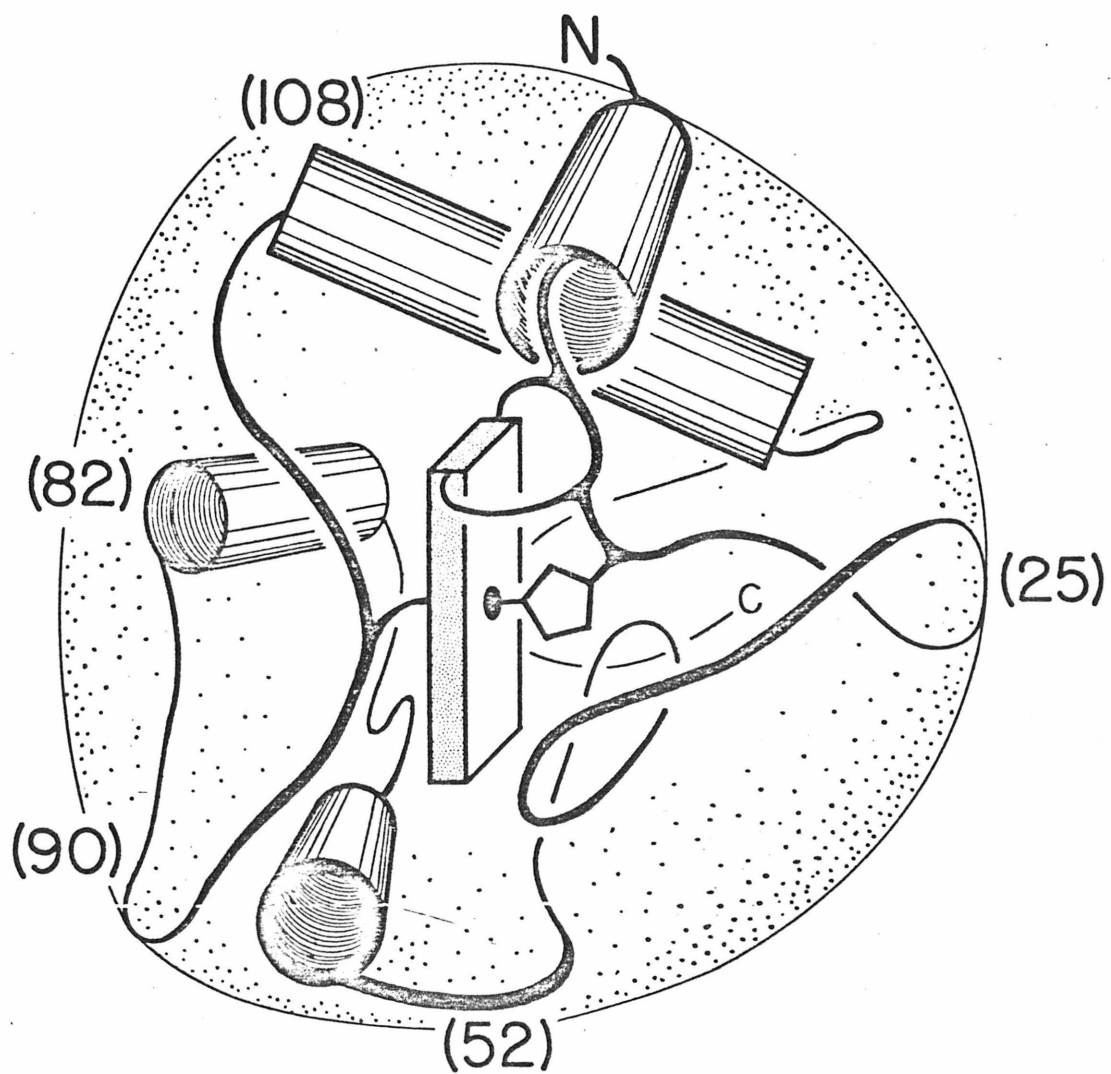


Figure 4.5. A stylized  $C_{550}$  molecule.

locales in the eukaryotic fold at a turn or change of direction, somewhat lengthening or abbreviating the turn, but ultimately returning to a common path. It is as if nature has been capricious with those regions of the molecule least likely to interfere with a central design for the protein.

A detailed description of the  $c_{550}$  polypeptide folding and its comparison with eukaryotic cytochrome  $c$  has been published (Timkovich and Dickerson, 1973). The structure comparison becomes more striking when R. rubrum cytochrome  $c_2$  (Salemme, et. al., 1973), another bacterial type cytochrome, is included, for  $c_2$  shows three of the minor insertions (61→68, 88→93, and 96→97;  $c_{550}$  numbering) and three of the minor deletions (12→13, 103→104, and 106→109) present in  $c_{550}$  with respect to  $c$ . To a high degree of accuracy, eukaryotic  $c$  stands as a core structure,  $c_2$  adds a few insertions onto the core, then  $c_{550}$  makes the same additions plus some other major insertions.

The striking similarities can be over emphasized leading to a false impression, in the same vein as calling myoglobin and hemoglobin identical proteins. The structural differences are sufficient to lead to the remarkable reactivity patterns outlined in Chapter 1, that  $c_{550}$  and  $c$  can be cross-reactive in a natural electron transport chain while  $c_2$  is inactive. Yet recall the intriguing fact that  $c_2$  is intermediate in structure between  $c$  and  $c_{550}$ !

The most fruitful approach to this comparative cytochrome biochemistry seems to begin with the reasonable assumption that all three cytochromes employ the same mode of action. For cytochrome function this means that for all three the reductase binds, however

fleetingly, to the same vicinity; the reducing electron enters by a highly analogous path, pre-or post-reduction conformation changes are at least somewhat similar, oxidase binding and electron donation are similar, and finally the change back to the oxidized conformation follows somewhat parallel lines. With this framework in mind, the structure of  $\underline{c}_{550}$  may be examined with an eye toward similarities and differences.

#### 4. 3. Molecular Structure — Details

In Table 4.3, the sequences of the three cytochromes have been aligned to demonstrate both sequence homology (all three proteins) and structural homology (only  $\underline{c}_{550}$  and  $\underline{c}$ ). In the latter, homology was defined rigorously as existing only when the  $\alpha$ -carbons and the peptide amide planes of the respective residues agreed in orientation versus neighboring groups and the molecule as a whole.

As would certainly be expected, sequence homologous residues are most often also structurally homologous, but some exceptions do arise (positions, in  $\underline{c}_{550}$  numbering 3, 4, 26, 30, 65, 84, 95, and 98). Generally, these are noticed to be near insertions in  $\underline{c}_{550}$ , and so may be the result of necessary changes forced upon  $\underline{c}_{550}$  to accommodate the new details. Potentially, these are accidental sequence homologs. To underscore a fundamental tenet of protein folding and protein architecture, one should notice the large number of structure homologs which are not equivalent in sequence nor compatible in sequence, where by compatible it is meant that the residues would agree in general chemical properties, such as Ile-Leu, or Glu-Gln. Examples

TABLE 4.3

Cytochrome Sequence Comparisons

$c$	:		1		5		10		15														
$c_2$	:		a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C			
$c_{550}$	:	a	N	E	G	D	A	A	A	G	E	K	V	S	-	K	K	C	L	A	C		
	:		1		5		10		15														
	:			*	*	=	=	*	=	*	=				=	=	*	*	=				
$c$	:		20						25				30										
$c_2$	:		H	T	V	E	-	-	K	G	G	-	-	K	-	H	K	T	G	P	N	L	
$c_{550}$	:		H	T	F	D	-	-	Q	G	G	-	-	A	-	N	K	V	G	P	N	L	
	:		H	M	I	Q	A	P	D	G	T	D	I	K	G	G	K	T	G	P	N	L	
	:		20					25					30					35					
	:		=	*	*													=	=	=	=	=	
$c$	:		35					40					45					50					
$c_2$	:		H	G	L	F	G	R	K	T	G	Q	A	P	G	F	T	Y	T	D	A	N	
$c_{550}$	:		F	G	V	F	E	N	T	A	A	H	K	D	N	Y	A	Y	S	E	S	Y	
	:		Y	G	V	V	G	R	K	I	A	S	E	E	G	F	K	Y	G	E	G	I	
	:		40					45					50					55					
	:		*	=	*	*	*	*	=			*		=	=	*	=				*	*	
$c$	:		55										60					65					
$c_2$	:		K	N	-	-	-	-	K	G	-	-	I	T	W	K	E	E	T	L	M	E	
$c_{550}$	:		T	E	M	K	A	-	K	G	-	-	L	T	W	T	E	A	N	L	A	A	
	:		L	E	V	A	E	E	K	N	P	D	L	T	W	T	E	A	N	L	I	E	
	:		60					65					70					75					
	:		*	*								*	=	=	*	=	*	*	=	*	*	*	
$c$	:		70					75															
$c_2$	:		Y	L	E	N	P	K	K	Y	I	-	-	-	-	-	P	G	-	-	T		
$c_{550}$	:		Y	V	K	N	P	K	A	F	V	L	E	K	S	G	D	P	K	A	K	S	
	:		Y	V	T	D	P	K	P	L	V	K	K	M	T	D	D	K	G	A	K	T	
	:		80					85					90					95					
	:		=	*	*	*	*		*	*													
$c$	:		80					85					90					95					
$c_2$	:		K	M	I	F	A	G	I	K	K	K	T	-	E	R	E	D	L	I	A	Y	
$c_{550}$	:		K	M	T	F	K	-	L	T	K	-	D	D	E	I	E	N	V	I	A	Y	
	:		K	M	T	F	K	-	M	G	K	-	N	-	Q	-	A	D	V	V	A	F	
	:		100					105									110						
	:		=	=	*	=	*	*	*	=						*	*	*	*	=	=		
$c$	:		100					104															
$c_2$	:		L	K	K	A	T	N	E														
$c_{550}$	:		L	K	T	L	K																
	:		L	A	Q	B	B	P	B	A	G	Z	G	Z	A	A	G	A	G	S	B	S	Z
	:		115				120					125					130					135	
	:		=	*	*																		

Cytochrome  $c$  is from horse (Margoliash, 1962) and  $c_2$  from *R. rubrum* (Dus, Sletten, and Kamen, 1968). Residues in blocks indicate sequence homology between  $c_{550}$  and either  $c_2$  or  $c$ . The small symbols beneath the  $c_{550}$  sequence indicate structural homology between  $c_{550}$  and residues in  $c$ . \* indicates that the  $\alpha$ -carbon and peptide are in equivalent conformation; = indicates that the side chains are also spatially equivalent relative to the main chain.



are provided by the stretches 20→21, 57→60, 86→89, 104→106, and 116→117. The general point is that the over-all conformation is determined by some obscure energetic-entropic equation of state which has as separate inputs, the individual side groups. Some may be radically changed, yet complimentary changes elsewhere eliminate any net effect, or the particular changes are low-sensitivity parameters in directing folding. To paraphrase, the whole is surely the sum of the parts, but a few of the parts may be played with willfully, for there are alternatives at arriving at the same sum.

Table 4.4 completes the equating of  $\underline{c}_{550}$  three-dimensionally with eukaryotic  $\underline{c}$ . This list of hydrogen bonds has been conservatively constructed, with the  $\underline{c}_{550}$  list considerably shortened by a lack of total confidence in some nonetheless reasonable bonds.

The added molecular weight of  $\underline{c}_{550}$  offers the most readily discernible differences toward the eukaryotic folding.

1. The insertion at the amino-terminus (2 residues) is trivial, as it is affecting no other part of the molecule.
2. The first eleven residues form an  $\alpha$ -helical segment of cytochrome. In  $\underline{c}_{550}$  the "best" model building, i. e. using density maxima, led to a distorted helix, but one not far from the ideal structures found in both  $\underline{c}$  and  $\underline{c}_2$ .
3.  $\underline{c}_{550}$  and  $\underline{c}_2$  share a common deletion relative to  $\underline{c}$  of a half-turn of helix just prior to the covalent heme.
4. Residues 24→31 in  $\underline{c}_{550}$  constitute a major new feature. Reference to Figure 4.4. shows that this is derived by a net insertion of five residues in a resultant conformation which is a clear cousin to the loop in  $\underline{c}$ .
5. The folding in all three cytochromes is constant until  $\underline{c}_2$  adds three residues and  $\underline{c}_{550}$  six at position 61 opposite the exposed

TABLE 4. 4

HYDROGEN BOND COMPARISONS

C<sub>550</sub> AND C

<u>Bonds in <u>c</u><sub>550</sub></u>	<u>Bonds in <u>c</u></u>
Cys 15 peptide O → Cys 18 peptide N	Cys 14 peptide O → Cys 17 peptide N
His 19 imidazole N → Pro 36 peptide O	His 18 imidazole N → Pro 30 peptide O
Lys 45 peptide O → Ser 48 peptide N	<u>c</u> polypeptide folds in a different manner
Glu 56 carboxylic O → Front propionic acid	Thr 49 hydroxyl OH → front propionic. Note that Thr 49 in <u>c</u> is structurally equivalent to Gly 55 in <u>c</u> <sub>550</sub>
Leu 59 peptide O → Glu 63 peptide N	Asp 50 peptide O → Ser 54 peptide N
Glu 60 peptide O → Glu 64 peptide N	Ala 51 peptide O → Lys 55 peptide N
Brief turn of $\alpha$ -helix	Brief turn of $\alpha$ -helix
Tyr 54 hydroxyl OH → back propionic acid	Tyr 48 hydroxyl OH → back propionic
Trp 71 indole N → back propionic	Trp 59 indole N → back propionic
Trp 71 peptide O → Gly 43 peptide N	Trp 59 peptide O → Gly 37 peptide N
$\alpha$ -helical hydrogen bonding pattern 72 → 81	$\alpha$ -helical hydrogen bonding pattern 60 → 68
Cys 18 peptide O → Gly 35 peptide N	Cys 17 peptide O → Gly 29 peptide N
Sequence analogous Thr 98 is rotated out toward solvent	Thr 78 hydroxyl OH → front propionic acid
$\alpha$ -helical hydrogen bonding pattern 107 → 118	$\alpha$ -helical hydrogen bonding pattern 89 → 101

heme edge.

6. Again  $\underline{c}_2$  and  $\underline{c}_{550}$  share six residues spanning 88→93 corresponding to an insertion into  $\underline{c}$  between positions 75 and 76. This extra coil of protein is in intimate proximity to the exposed heme edge and constitutes a new collar of polypeptide surrounding this edge. Attention should be directed toward this feature, as in the eukaryotic cytochromes the segment 70→80 is sequence conserved over sixty species, therefore presumably playing some mandatory role in structure-function. Slightly further along,  $\underline{c}_{550}$  adds another two residues which  $\underline{c}_2$  does not, leading to a net inserted coil slightly larger than shown by  $\underline{c}_2$ .
7. A minor deletion of three net residues in the region 107→109 shortens the turn  $\underline{c}_{550}$  makes before beginning the carboxyl terminal  $\alpha$ -helix.
8. A major insertion of 14 residues lengthens the tail of  $\underline{c}_{550}$  and adds an extra layer of protein onto the surface of the molecule diagonally opposite the exposed heme edge. The electron density map is extremely weak in this area suggesting high disorder. Indeed, the final five residues appear to be averaged between two possible conformations with the major form as shown in the Figures. Since these 14-residues protrude rather drastically into the solvent and do not appear to be locked into a fixed array by hydrophobic packing with the rest of the molecule or by hydrogen bonds, one would be justified in assuming that in solution they wave about as a random coil tail on a globular protein.

A few additional words of some entertaining, if not enlightening, speculation on the properties of this peptide tail seem appropriate. In the crystal, the protein molecules pack in one direction in a linear chain in which the insertion on the His side (point 4 above) and the tail insertion form a complimentary jigsaw puzzle fit. This is shown schematically in Figure 4.6. The accidental (?) compatibility explains the extreme close packing of molecules in this crystal form; but it is doubtful that the fit has any relevancy in solution. Next, Table 4.5 shows the amino acid composition of the tail

# Crystal Packing along $c^*$

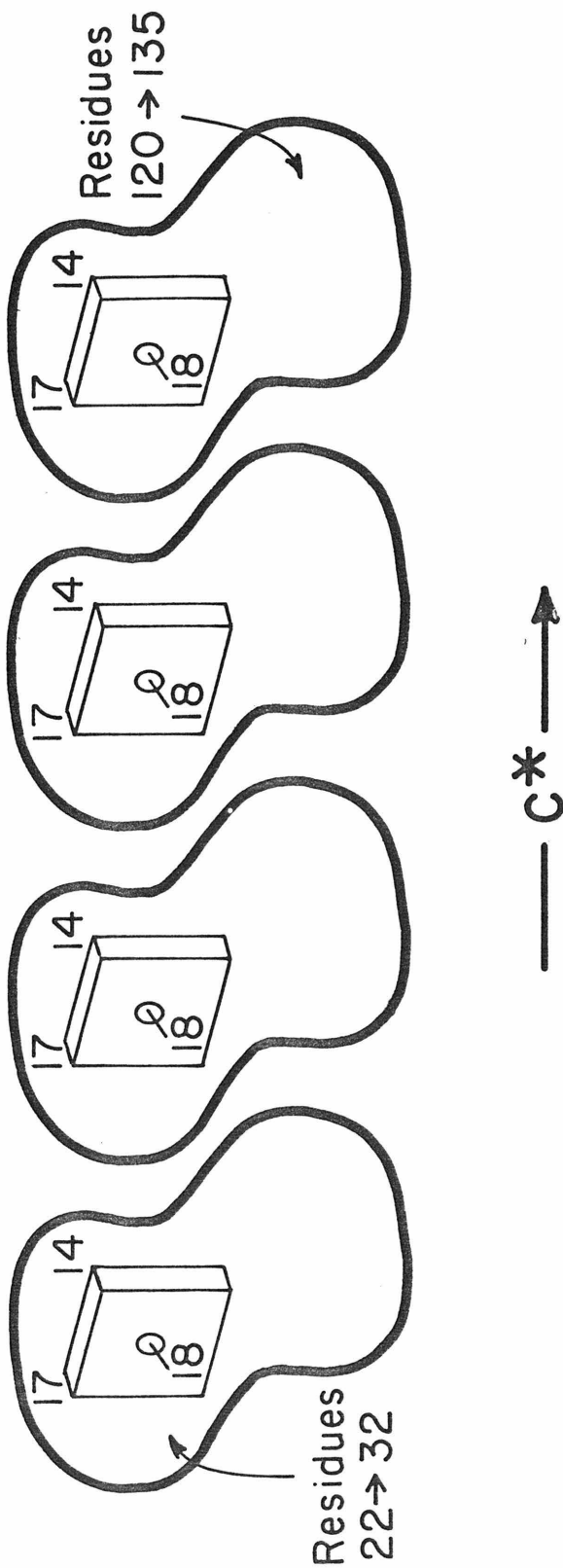


Figure 4. 6.

---

TABLE 4.5  
AMINO ACID COMPOSITIONS<sup>1</sup>  
BACTERIOPHAGE COAT PROTEINS

Residues	Bacteriophage			c <sub>550</sub> tail (115-134)
	FR	F2, R17	Q Bcta	
Asx	15	14	15	4
Thr	9	9	12	-
Ser	11	13	9	2
Glx	11	11	13	4
Pro	5	6	8	1
Gly	9	9	7	4
Ala	16	14	15	5
Cys	2	2	2	-
Val	16	14	13	-
Met	2	1	-	-
Ile	6	8	4	-
Leu	5	8	12	-
Tyr	4	4	4	-
Phe	5	4	3	-
His	-	-	-	-
Lys	7	6	7	-
Arg	4	4	7	-
Trp	2	2	-	-

---

<sup>1</sup>From Dayhoff, 1972.

---

contrasted with the amino acid compositions of some bacteriophage coat proteins. As can be seen, there is some correspondence in the popularity of certain amino acids across the table. The following model, then, is proposed without further justification to reconstruct an evolutionary mutational event of some billions of years in the past. It is well known that lysogenic viral particles, by becoming passive prophage, may infect a bacterium and insert their DNA into the bacterial chromosome without causing host demise. It is also known that extra genetic material, such as this or such as episomes or prokaryotic sex factors, can be excised from the chromosome either by taking a small piece of host DNA with it or by leaving some behind. One can, by straining, envision a phage infecting a M. denitrificans bacterium and then passively leaving behind DNA which has obliterated the synthesis termination signal for a pro-cytochrome c protein. Our ancestor organism then goes on to take over the genus as the result of some other survival trait, all the time carrying about a bit of junk peptide on its cytochrome c.

#### 4.4. Molecular Structure – Some Net Effects

In an early paper (Timkovich and Dickerson, 1973) the author stressed a preservation in the packing of aromatic and hydrophobic residues around the prosthetic group heme during the time eukaryotic c and bacterial c<sub>550</sub> have evolved from some presumed common ancestor. With the exception of incorrectly identifying amino acid number 86, a leucine, as a tyrosine or phenylalanine, and connecting the side chain phenyl group of Phe 114 to the wrong spot on the main

chain, the interpretations and structural details of this  $4\text{Å}$  analysis have been successfully borne out at high resolution. In point of fact, the comparison becomes even more striking, for now the orientation of hydrocarbons and aromatic rings can also be seen to be conserved (Table 4.6). For the eukaryotic cytochromes we have a choice of ring orientations to talk about – the oxidized or reduced structures, which show clear differences in ring alignments. Of the two, oxidized  $c_{550}$  appears to mimic the oxidized horse cytochrome most closely, but it is somewhat intermediate between oxidized horse and reduced tuna cytochrome. Some of the discrepancy is attributable to errors and a general lack of resolution in the original horse map, and this situation is being currently remedied by a high resolution redetermination of the oxidized structure, now using high quality oxidized tuna cytochrome crystals (Swanson, Trus, and Dickerson, personal communication). Preliminary results are consistent with the view that the ring structures between oxidized and reduced forms are more similar than was originally supposed and the final, fine differences will be more nearly equivalent to those now seen between reduced tuna and oxidized bacterial cytochrome.

The similarities here are worth belaboring, for it has been proposed (Kassner, 1972) that hydrophobic packing about the heme plays a major role in adjusting the redox potential of the iron center to its physiologically functional value. It would have been informative had  $c_2$ , with its  $E'_0$  60 mv higher than  $c$  or  $c_{550}$ , shown some striking difference in heme packing compared to  $c$  or  $c_{550}$ , but unfortunately

TABLE 4. 6

AROMATIC RING ORIENTATIONS

<u>c<sub>550</sub></u>	<u>Oxidized c<sup>1</sup></u>	<u>Reduced c<sup>2</sup></u>
Phe 12, adjacent to heme and tipped perpendicular to it.	Phe 10, as in <u>c<sub>550</sub></u> .	Phe 10, as in <u>c<sub>550</sub></u> .
Tyr 39, lies in a groove on the surface of the molecule with only its edge exposed to solvent, like a mini-buried heme.	His 33, although a different ring system, lies in the identical orientation.	Trp 33, another different group, but the same orientation.
No analog in <u>c<sub>550</sub></u> .	Phe 36, lies in a groove on the surface as did 33.	Phe 36, same locale but rotated by 90° to lie flush with the surface.
Phe 52 - Tyr 54, adjacent and perpendicular to the heme plane. With respect to each other, their planes are rotated by 60-80°.	Phe 46 - Tyr 48. Same.	Tyr 46 - Tyr 48. Same.
Trp 71. Hydrogen bonding to propionic acid. π-system 90° with respect to heme.	Trp 59. Same as <u>c<sub>550</sub></u> .	Trp 59. Similar but rotated slightly about a line perpendicular to the heme plane. Tends to place a flat face against the heme while the others place an edge.
Tyr 67. π-cloud mingles with heme π-cloud. At about 30° with respect to the heme.	Tyr 67. As in <u>c<sub>550</sub></u> but almost parallel with the heme.	Tyr 67. Same locale but rotated 90° to be perpendicular to the heme, parallel to the Trp.
The analog in <u>c<sub>550</sub></u> is a leucine.	Tyr 74. Parallel and adjacent to Trp 59. Flush with the molecular surface.	Tyr 74. Rotated by 90°. Perpendicular to the surface.



(Table 4. 6 - continued)

<u>c<sub>550</sub></u>	<u>Oxidized c<sub>1</sub></u>	<u>Reduced c<sub>2</sub></u>
Phe 102. Adjacent and parallel to the heme. A methionine-ligand-side equivalent to Phe 12.	Uncertain.	Phe 82. As in <u>c<sub>550</sub></u> .
Phe 114. Adjacent to Phe 12 on the side away from the heme. Roughly perpendicular.	Tyr 97. As in <u>c<sub>550</sub></u> but rotated to be parallel to Phe 10.	Tyr 97. As in <u>c<sub>550</sub></u> but rotated to be at about 45°.

---

<sup>1</sup>Dickerson, et. al., 1971.

<sup>2</sup>Takano, et. al., 1973.

---

this simplistic hope has not been realized. Things instead look uniform. The redox potential tuning must then lie in some fine structural detail;  $\pi$ -electron interactions or ligand bond strain offer two intriguing variables.

A most unusual and asymmetric charge distribution can be seen on the surface of  $c_{550}$  from an inspection of the distribution of acidic and basic side functions. Almost all of the lysine residues (15 out of 17) are grouped into a collar about the exposed edge of the heme in what shall be called the front of the molecule. This density of positive charges falls-off with the distance from the edge. There are 11 lysines within  $10 \text{ \AA}$  of the crevice opening, another 4 within  $17 \text{ \AA}$  forming a second more or less concentric ring, and only two which seem to be outside the collar. All the lysines are grouped into the front hemisphere of surface. The acidics show a more uniform distribution with perhaps what could be called a slight grouping on the back surface near residues 55-65, 120-135. Through default, the second hemisphere carries solely negative charges and thus  $c_{550}$  exhibits properties of a macro-dipole, being a rough sphere of  $35 \text{ \AA}$  diameter with a positive half and a negative half. The charge separation may have mechanistic implication and so will be referred to a later section.

#### 4.5. Evolutionary Relationships

Extensive comments have been previously published (Timkovich and Dickerson, 1973) on the phylogenetic relations between  $c$ ,  $c_2$  and  $c_{550}$  and so, only a brief review shall be given here. It seems most reasonable to suppose that a common ancestral protein existed for

the three and that divergence along three lines has produced the present contemporary versions. If the ancestor is taken as most resembling  $c_2$  in structure, folding, and size, then additions account for  $c_{550}$  and deletions have led to  $c$ , with this reconstruction being the simplest which avoids postulating parallel mutational events. By looking at in vivo function, it appears most plausible to suppose  $c_{550}$  and  $c$  to be more closely related to each other than to  $c_2$ , using the argument that respiratory proteins share more evolution history than would any respiration-photosynthesis pair.

The  $c_{550}$  sequence could have provided more of a base on which to view these suppositions. Yet, a similarity/difference census between sequences (Table 4.7) cannot clearly answer the question "Which two are most closely related?". It is possible for the insertions in  $c_{550}$  to be the product of a local gene duplication or an unequal cross-over event, but an examination of the added peptide regions shows no definite evidence of repetition of a neighboring sequence. It may simply be that time since the mutational event has allowed point mutations to scramble sequences sufficiently to obliterate local homology.

The general result of sequence comparisons is neutrality — things are so equidistant that the data can explain anything. It is to be expected, although disappointingly so, that time has obscured a reconstruction of phylogeny on the basis of sequence. At a very broad level, it may be stated that the respiratory mitochondrial cytochrome has diverged as far from the bacterial respiratory cytochrome as it has from a photosynthetic cytochrome.

---

TABLE 4.7  
SEQUENCE COMPARISON  
SIMILARITY - DIFFERENCE MATRIX

	<u>c</u>	<u>c</u> <sub>2</sub>	<u>c</u> <sub>550</sub>	
<u>c</u>	-	43	46	
<u>c</u> <sub>2</sub>	57	-	53	Same
<u>c</u> <sub>550</sub>	54	55	-	

Different

The Table was constructed from the sequence alignment of Table 4.3 and thus it represents a comparison only over those positions which are structurally equivalent in the proteins. Residues aligned with deletions in the contrasted proteins are not counted in either category. The upper right triangle lists the number of residues matching, the lower right, the number different. The amides and acids Asn/Asp and Gln/Glu were counted as equivalent because in both c<sub>2</sub> (Bartsch, et. al., 1971) and c<sub>550</sub> (G. McLain, personal communication) there exists the possibility of deamidation during isolation or in the in vivo cytochrome complement.

---

#### 4.6. Current Proposed Mechanisms for the Reduction of Cytochrome $c$

Takano, et. al. (1973) have recently proposed a mechanism for cytochrome reduction along the following lines. In the oxidized state an aromatic residue at position  $74_E$  (subscript E refers to sequence numbering according to the eukaryotic cytochromes, B will refer to bacterial  $c_{550}$ , and P the photosynthetic  $c_2$ ) lies parallel to the molecular surface and to the aromatic ring of Trp  $59_E$ . Tyr  $67_E$  meanwhile lies adjacent and parallel to the heme. An electron is seen as being donated to the  $\pi$ -orbital of  $74_E$  and by  $\pi$ -cloud overlap of proximal groups, it migrates to  $59_E$ . As this is going on, a stable  $\pi$ -electron from  $67_E$  migrates to the heme and reduces the iron center. The protein then undergoes a conformation change and Trp  $59_E$  and Tyr  $67_E$  become parallel. Through cancellation of an electron "hole" and the antibonding electron on  $59_E$ , the process is complete.

The  $c_{550}$  structure raises grave doubts concerning this proposal in so far as  $c_{550}$  and  $c$  share a common mechanism. It can be seen that the aromatic Tyr  $74_E$  is a crucial component, for it initiates the  $\pi$ -conduction process and provides the link to the surface and the cytochrome reductase. Yet in  $c_{550}$  this residue (structurally  $86_B = 74_E$ ) is an aliphatic leucine with no other possible aromatics to act as replacements. By knowing the sequence and structure of  $c_{550}$  it is thus possible to eliminate consideration of the Takano, et. al. mechanism.

In a similar fashion, a definitive evaluation may be made of a different scheme proposed by Salemme, Kamen, and Kraut (1973). Strictly speaking, the mechanism as discussed here is applicable only

to  $\underline{c}_2$  with some changes in detail when it is applied to  $\underline{c}$ . The slight differences in the two versions were mostly motivated by structural features on oxidized  $\underline{c}$  which have since been found to have been erroneously reported (Swanson, Trus, and Dickerson, personal communication). Oxidized  $\underline{c}$  now appears to be more like  $\underline{c}_2$  and so the  $\underline{c}_2$  version will be discussed as a unified mechanism. Here, the reduction of cytochrome is seen as the result of a simultaneous protonation and electron donation. The acceptance of the electron directly by the heme iron is seen as the result of the protonation of a nearby Ser 89<sub>P</sub>(= Thr 78<sub>E</sub> = Thr 98<sub>B</sub>) which changes an elaborate hydrogen bonding network Ser 89<sub>P</sub> - Tyr 52<sub>P</sub> - Tyr 70<sub>P</sub> (or Thr 78<sub>E</sub> - Asn 52<sub>E</sub> - Tyr 67<sub>E</sub>) which in turn has been stabilizing a bent iron-methionine sulfur bond. In  $\underline{c}_{550}$ , the pivotal hydrogen bond donor, 52<sub>P</sub> = 52<sub>E</sub> = 58<sub>B</sub>, is an isoleucine with no possibility of forming such bonds. Further, in  $\underline{c}_{550}$  approximately the same amount of bonding may be seen in the iron sulfur bond as is seen in  $\underline{c}_2$ , but it is bent in exactly the opposite direction. Although  $\underline{c}_{550}$  is not at as high a resolution, nor has it been refined crystallographically, the author feels that the small amount of bending seen (0.35 Å) is more the result of noise in the analysis than in structure.

#### 4.7. Alternative Mechanisms

Before a concrete proposal for a unified mechanism of action for cytochrome  $\underline{c}$  can be made, consistent with the three known structural examples of cytochrome, the oxidized  $\underline{c}_{550}$  co-ordinates

must be refined and the reduced structure investigated. However, some preliminary remarks are called for.

The surface structural feature which is most strongly conserved is the collar or ring of positive lysines residues around the exposed edge of the heme. Attention was first directed toward this by Salemme, Kamen, and Kraut (1973) while Dickerson and co-workers were fostering the idea that the similarity was between a patch of lysines on the "right" or histidine side and a patch on the "left" or methionine side. The severe localization of charge into a collar about the heme crevice in  $c_{550}$  favors the former view. Figure 4.7 demonstrates in a schematic way the relevant collar grouping.

The large insertions on the surface of  $c_{550}$  make certain regions on cytochrome  $c$  unlikely as binding sites for oxidase and reductase (bacterial or photosynthetic) in view of the partial cross-reactivity between  $c_{550}$  and  $c$ . The "right" side (residues  $20_B \rightarrow 40_B$ ), the "bottom" (residues  $85_B \rightarrow 98_B$  and  $55_B \rightarrow 70_B$ ), and the "back" (residues  $120_B \rightarrow 135_B$ ) all contain large perturbations on the eukaryotic cytochrome surface. The "top", the "left" side, and the heme crevice remain as possible choices.

The three known structures are consistent with the following mechanisms. The reductase and oxidase each utilize a different part of the lysine collar to recognize cytochrome  $c$  through protein-protein binding. However, each uses the same exposed heme edge as the actual pathway for electron migration, this being either through the heme  $\pi$ -orbital or directly to the iron. Conformation changes are the result, not the cause, of electron donation.

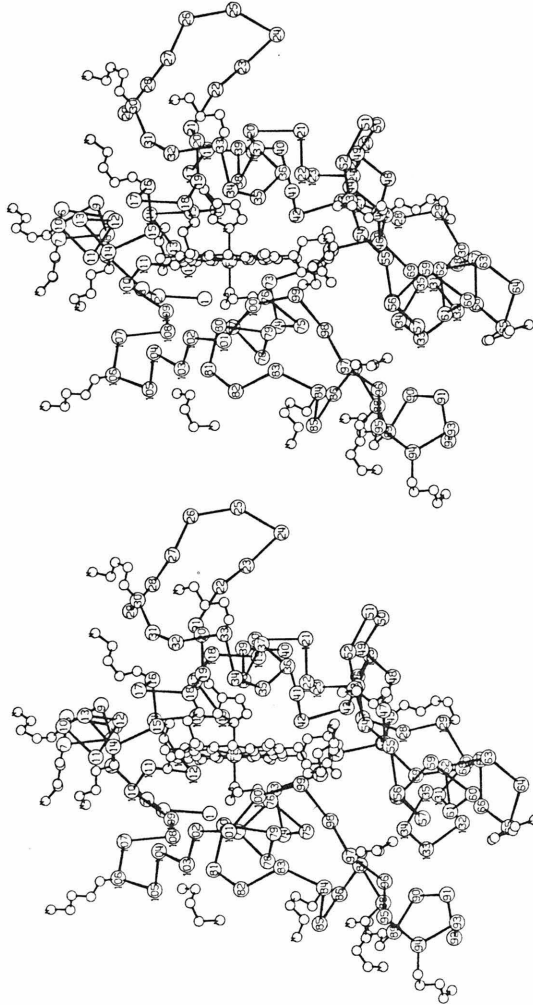


Figure 4.7. A front view of  $C_{550}$  with all the lysine side chains. Note the grouping about the heme crevice.



Certainly less elaborate than previous proposals, this is almost an "anti-mechanism" as far as most biochemists are concerned, for it says in essence that the cytochrome electron transfer is not terribly different than that of simple model system redox reactions. As such, mechanistic explanations must proceed to a higher level of detail involving orbital overlap, perturbed transition states, tunneling and/or other quantum mechanical considerations common to all redox reactions.

This is not to say that the protein about cytochrome c plays no role. Clearly, it performs certain unique, vital tasks. 1) It confers specificity upon cytochrome c to the reductase and oxidase — a vital control function for electron flow. 2) It adjusts the iron center redox potential to appropriate values. 3) It more or less isolates the iron center and this may prevent redox reactions with other physiological components which would lead to a short-circuit or by-pass of the normal chain. (4) The highly water-insoluble heme is made soluble by engulfing it in a hydrophilic matrix. Inevitably, but for some unknown reason, cytochrome c is the only truly soluble protein in an extensive transport chain involving several hemoproteins.

A total justification of these ideas would require at least as much space as the hundreds of papers which have appeared involving cytochrome c, so the obvious restriction to a few salient experiments is made.

Recently Margoliash and co-workers (Margoliash, et. al., 1973) have presented an investigation into the effects of the

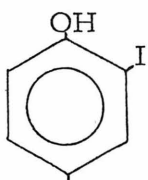
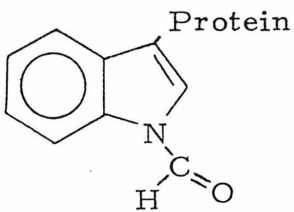
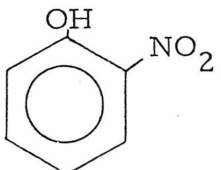
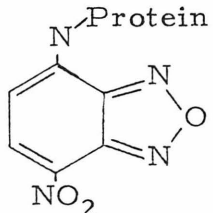
chemical modification of selected, single amino acid residues in cytochrome c which by their invariance among eukaryotic cytochromes has suggested a crucial function in cytochrome. In summary, these workers find that purified, single-site derivatives at positions indicated in Table 4.8 lead to a cytochrome which can function better with oxidase than with reductase, or the converse, better with reductase than with oxidase. Their adopted explanation is that the reductase and oxidase must utilize separate pathways through the cytochrome molecule, delineated by the surface locations of the crucial modified residues.

As mentioned before, c<sub>550</sub> possesses a "derivatived" group at the location corresponding to 74 in eukaryotic cytochrome, namely a leucine substituted for the otherwise invariant aromatic (tyrosine or phenylalanine). This suggests that it is not the aromaticity in itself at 74 which is mechanistically important but rather that the structural integrity of this region is important. The broad statement just made is best clarified by citing possibilities wherein 74 is directly recognized by reductase as part of a binding site on cytochrome, or perhaps 74 contributes in maintaining some general conformation used in binding.

The derivative experiments are consistent with the simpler view that they selectively interfere with separate oxidase or reductase binding sites without affecting the transfer event. Arguments using the reported  $V_{\max}$  and Michaelis constants  $K_m$  for the derivatives cannot be used to decide between interference with binding or transfer

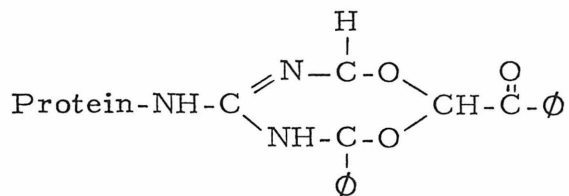
TABLE 4. 8

SOME CHEMICAL MODIFICATIONS OF CYTOCHROME C

Residue	Derivative	Oxidase Reaction	Reductase Reaction
Tyr 74	Mono-iodination at ring position 3 or 5, exactly which is unknown.	normal	defective
	 <p>Protein</p>		
Trp 59	The production of 1-formyl Tryptophan	normal <sup>1</sup>	auto-oxidizable
	 <p>Protein</p>		
Tyr 67	Mono-nitration on the aromatic ring	normal <sup>1</sup>	auto-oxidizable
	 <p>Protein</p>		
Met 80	Carboxymethylation Protein-CH <sub>2</sub> -CH <sub>2</sub> -S-CH <sub>2</sub> -COO <sup>-</sup>	normal <sup>1</sup>	auto-oxidizable
Lys 13	Reacted with 1-chloro-4-nitrobenzo 2'-oxa-1',3' diazole	defective	40% activity
	 <p>Protein</p>		

(Table 8 - continued)

<u>Residue</u>	<u>Derivative</u>	<u>Oxidase Reaction</u>	<u>Reductase Reaction</u>
Arg 13 ( <u>Candida</u> <u>Krusei</u> cytochrome <u>c</u> )	Reacted with phenylglyoxal ( $\phi$ -C-C-H) $\begin{array}{c} \parallel \quad \parallel \\ \text{O} \quad \text{O} \end{array}$	defective	50% activity




---

<sup>1</sup> Only at the limit of very high substrate (cytochrome c) concentration. See Smith, Davies, and Nava (1974) for the qualifications this introduces.

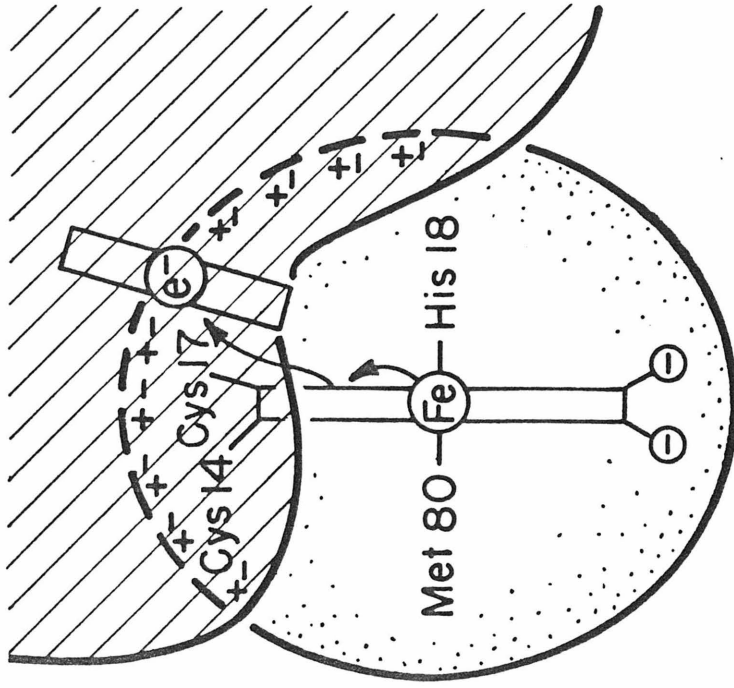
---

because it has been shown that oxidase/reductase reactions do not follow standard Michaelis-Menten kinetics (Smith, Davies, and Nava, 1974). The derivatives of "internal" residues (67 and 59) may still produce binding abnormalities through a domino effect whereby internal alterations are transmitted to the surface conformation.

A series of experiments related in principle to those of chemical modification are provided by Smith, et al. (1972) working with antibodies elicited against cytochrome c. Purified, partially purified, and heterogeneous populations of Fab fragments were found which selectively inhibited either the oxidase or reductase reaction, leaving the other normal, by binding to specific (but unknown) surface areas on cytochrome c. Here also it seems simplest to say that binding problems at or near the lysine collar produce inhibition.

As with the modification work, the antibody results have been taken as indicating separate electron entry and exit pathways although the idea of just separate binding sites was mentioned. The motivation for adopting this view as opposed to a simple binding interference has been the feeling that antibody fragments are so large that for them to block binding without covering the transfer site (e. g. the heme edge), the binding site and the transfer site must be widely separated. In rebuttal it must be pointed out that there are no data available on how much (or how little) of the Fab fragment is involved in immunological recognition and so a consideration of antibody size is an insufficient argument. A simple scheme as outlined in Figure 4.8 would make use of the exposed heme edge for transfer and the lysine collar for binding and leave respectively a hemisphere of surface for

Cytochrome Oxidase



Cytochrome Reductase

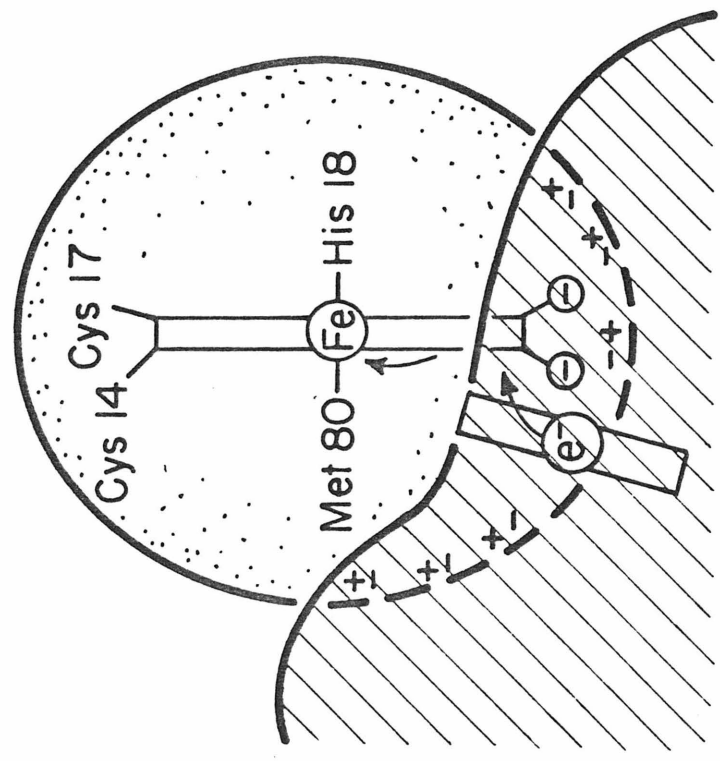


Figure 4.8. Binding of transport enzymes to cytochrome c.

the anti-reductase antibody and anti-oxidase antibody to sit upon.

To the extent of taking the heme edge as the electron path, this author concurs with Salemme, Kamen, and Kraut (1973) in citing the studies of Redfield and Kupta (1971) and Eaton and Hochstrasser (1967). These authors quote NMR and single crystal absorption spectra as indicating an asymmetric electron density distribution in the plane of the heme. Redfield and Kupta have further identified this as a concentration of electron density in the oxidized state such that the unpaired spin density is oriented toward the most exposed edge of the heme (toward pyrrole ring II, nearest Cys 17). The implication is that the oxidized structure placed the majority of density, to be paired with the incoming electron, at the site where donation occurs — the exposed heme edge.

A variety of model reaction systems have been studied in which cytochrome is oxidized or reduced by simple, inorganic reagents (Ewall and Bennett, 1974; Lambeth, et. al., 1973; Yandell, Fay, and Sutin, 1973). Of particular interest are the experiments of Hodges, Holwenda, and Gray(1974) on the kinetics of reduction of ferricytochrome c by ferrous EDTA ( $\text{Fe}(\text{EDTA})^{2-}$ ), which show that this model system is clearly in line with typical reaction rates for other inorganic systems in which the transfer is "outer sphere". By this it is meant that the reactant metal centers affect transfer without a transition state involving ligand substitution but by a simple proximity of centers with intact ligands. The point of relevancy is that the cytochrome heme edge could thus function by being in contact with an electron donor/acceptor.

Technically speaking, more elaborate proposals, such as the  $\pi$ -cloud conduction path of Dickerson, et. al., are also outer sphere.

If the simple inorganic reactants make use of an electron pathway, (for example, heme edge), which is shown to be energetically, kinetically favorable, then why should physiological reactants employ a more intricate path? At stake is finding a low activation barrier reaction course, then demanding that alternatives be shown to be clearly lower in barrier height.

The situation is far from clear concerning a mechanism of cytochrome action consistent with everybody's experiments on this most ubiquitous of proteins. Critics can certainly offer counter-arguments to the situation as presented here. The hope for the future is that the known structures of cytochromes plus perhaps a bit more information from the forthcoming structures of reduced  $c_{550}$  and the very small  $c_{551}$  of *Pseudomonas aeruginosa* (R. Almassey, personal communication) as well as further sequence studies, will provide sufficient constraints to limit the possibilities for cytochrome mechanism and suggest conclusive dynamic experiments. Model system experiments are underway (e. g. by Gray, Bennett and others) as are new attempts at chemical modification in several laboratories (by Margoliash and Harbury). Perhaps in the near future the books can be closed on the story of a protein first observed in 1886 and purified by 1930.

#### 4. 8. References

Bartsch, R. G., Kakuno, T., Horio, T., and Kamen, M., (1971).  
J. Biol. Chem., 246, 4489.



- Dayhoff, M. O., (1972). Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington, D. C.
- Dickerson, R. E., Takano, T., Eisenberg, D., Kallai, O. B., Samson, L., Cooper, A., and Margoliash, E., (1971). J. Biol. Chem., 246, 1511.
- Dus, K., Sletten, K., and Kamen, M. D., (1968). J. Biol. Chem., 243, 5507.
- Eaton, W. A., and Hochstrasser, R. M., (1967). J. Chem. Phys., 46, 2533.
- Ewall, R. X., and Bennett, L. E., (1974). J. Amer. Chem. Soc., 96, 940.
- Herriott, J. R., Watenpaugh, K. D., Sieker, L. C., and Jensen, L. H., (1973). J. Mol. Biol., 80, 423.
- Hodges, H. L., Holwenda, R. A., and Gray, H. B., (1974). J. Am. Chem. Soc., 96, 3132.
- Kassner, R. J., (1972). Proc. Nat. Acad. Sci., 69, 2263.
- Kendrew, J. C., Watson, H. C., Strandberg, B. E., Dickerson, R. E., Phillips, D. C., and Shore, V. C., (1961). Nature, 190, 666.
- Lambeth, D. O., Campbell, K. L., Zand, R., and Palmer, G., (1973). J. Biol. Chem., 248, 8130.
- Margoliash, E., (1962). J. Biol. Chem., 237, 2161.
- Margoliash, E., Ferguson-Miller, S., Tulloss, J., Kang, C. H., Feinberg, B. A., Brautigan, D. L., and Morrison, M., (1973). Proc. Nat. Acad. Sci., 70, 3245.
- Redfield, A. G., and Gupta, R. K., (1971). In Cold Spring Harbor Symposia on Quantitative Biology, XXXVI, 405.
- Salemme, F. R., Freer, S. T., Xuong, Ng. H., Alden, R. A., and Kraut, J., (1973). J. Biol. Chem., 248, 3910.
- Salemme, F. R., Kraut, J., and Kamen, M. D., (1973). J. Biol. Chem., 248, 7701.
- Smith, L., Davies, H. C., Reichlin, M., and Margoliash, E., (1972). J. Biol. Chem., 248, 237.
- Smith, L., Davies, H. C., and Nava, M., (1974). J. Biol. Chem., 249, 2904.

Takano, T., Kallai, O. B., Swanson, R., and Dickerson, R. E., (1973).  
J. Biol. Chem., 248, 5234.

Timkovich, R., and Dickerson, R. E., (1973). J. Mol. Biol., 79, 39.

Yandell, J. K., Fay, D. P., and Sutin, N., (1973). J. Amer. Chem. Soc., 95, 1131.

PROPOSITION 1  
SOME NOTES ON THE FIGURE OF MERIT  
AS A COMPARATIVE STATISTIC

The figure of merit is an often quoted statistic used to gauge the quality of a protein crystallographic structure determination based on isomorphous replacement. It derives its existence from the error analysis for isomorphous replacement by Blow and Crick (1959). It is empirically well known that there are correlations between the mean figure of merit and such parameters as the root mean square lack of closure and the number of isomorphs in the analysis. While the root mean square lack of closure certainly reflects the quality of the data and such critical features as the degree of isomorphism, the number of isomorphs can be a pure statistical effect independent of the actual quality of the phase determinations. In either case it is not obvious what will be the analytical result upon mean figure of merit as one changes the root mean square lack of closure or the number of heavy atom derivatives.

It is the working hypothesis of these notes that, as currently used, the figure of merit is a worthless comparative measure of the accuracy of a structure determination. It will be shown that statistical effects which arise from the number of isomorphs and from potential isomorphous relationships among the sites of heavy atom binding negate the usefulness of this statistic when comparing different structure determinations, although for a single structure, it continues to aid in assessing internal error. A modification in the calculation of the

mean figure of merit will be proposed to normalize the statistic so as to make it independent of the number of derivatives available in any particular analysis.

This treatment follows the theory of Blow and Crick (1959). Notation employed is consistent with that adopted by Dickerson, Kendrew, and Strandberg (1961). Definitions:

$F_P$	The native structure factor magnitude
$F_{PH}$	The derivative magnitude
$f$	The magnitude of the heavy atom atomic scattering vector
$P(\varphi)$	The probability that $\varphi$ is the correct native protein phase.
pdf	The probability density function for $P(\varphi)$
$\varphi$	The native protein phase angle for $F_P$
$\varphi_m$	Maximum likelihood phase angle
$\varphi_b$	The Blow-Crick "best" phase angle
$m$	The figure of merit
$\langle \rangle$	An arithmetic average or expectation value
$E$	Root mean square lack of closure
$j$	Running index for heavy atom derivatives
$\alpha$	Heavy atom phase
$n$	The number of derivatives
SIR	The single isomorphous replacement solution to the phase problem

Two hypothetical limiting cases will be adopted which allow a closed, analytical expression for the mean figure of merit. Real cases can be considered as some combination of these two opposite

extremes. This is not an attempt to deduce expressions for  $\langle m \rangle$  which would be usable for calculation, but the resultant approximations will indicate the relationships between  $m$  and various parameters.

It is a prerequisite to show

$$m = \langle \cos \Delta \rangle \quad \text{where} \quad \Delta = \varphi - \varphi_b \quad (1)$$

Let  $f(x)$  = pdf of any random variable  $x$ ;  $y = H(x)$  be some function defined on  $x$ . Then by elementary probability theory

$$\langle y \rangle = \int_x H(x)f(x)dx \quad (2)$$

For the present case, this says

$$\langle \cos \Delta \rangle = \int_{\varphi} \cos \Delta P(\varphi)d\varphi \quad (3)$$

But

$$\cos \Delta = \cos(\varphi - \varphi_b) = \cos \varphi \cos \varphi_b + \sin \varphi \sin \varphi_b \quad (4)$$

so

$$\begin{aligned} \langle \cos \Delta \rangle &= \cos \varphi_b \int_{\varphi} \cos \varphi P(\varphi)d\varphi + \sin \varphi_b \int_{\varphi} \sin \varphi P(\varphi)d\varphi \\ &= m \left( \cos^2 \varphi_b + \sin^2 \varphi_b \right) \\ &= m \end{aligned} \quad (5)$$

When  $P(\varphi)$  is a combined pdf determined over several derivatives, it empirically comes to resemble closely a Gaussian curve centered about  $\varphi_m$  and falling quickly to zero away from the centroid. It is then called a uni-modal pdf. Assume that  $P(\varphi)$  can be analytically written as a true Gaussian curve about  $\varphi_m$  given by

$$P(\varphi) = Ne^{-(\varphi - \varphi_m)^2/2\sigma^2}, \quad \varphi_m - 180^\circ \leq \varphi \leq \varphi_m + 180^\circ \quad (6)$$

Take  $\Delta = \varphi - \varphi_m$ . Now approximate  $P(\varphi)$  by

$$P(\varphi) \approx \hat{P}(\varphi) = \hat{N} e^{-\Delta^2/2\sigma^2} \quad \text{where} \quad -\infty < \varphi < +\infty \quad (7)$$

the true  $P(\varphi)$  falls off quickly from  $\varphi_m$  so both the analytical assumption and its approximation are realistic. The latter is introduced solely to make subsequent manipulations easy.  $\hat{N}$  can be readily evaluated to be  $\hat{N} = 1/(2\pi)^{\frac{1}{2}}\sigma$ . Then

$$\begin{aligned} m &= \langle \cos \Delta \rangle \\ &\approx \hat{N} \int_{-\infty}^{+\infty} \cos \Delta e^{-\Delta^2/2\sigma^2} d\varphi \\ &= 2\hat{N} \int_0^{\infty} \cos \Delta e^{-\Delta^2/2\sigma^2} d(\Delta) \end{aligned} \quad (8)$$

The integral can be evaluated in closed form to imply

$$m = \langle \cos \Delta \rangle \approx e^{-\sigma^2/2} \quad (9)$$

From this it can be seen that anything which decrease  $\sigma$  increases  $m$ . The empirically based approximation for the pdf can be roughly equated to the formal Blow-Crick definition of  $P(\varphi)$ .

$$P(\varphi) = N e^{-\sum(\epsilon_j^2/2E_j^2)} \approx \hat{N} e^{-\Delta^2/2\sigma^2} \quad (10)$$

Note that  $N \neq \hat{N}$ , since

$$P(\varphi_m) = N e^{-\sum[\epsilon_j^2(\varphi_m)/2E_j^2]} = \hat{N} \quad (11)$$

Now at values of  $\varphi \neq \varphi_m$

$$\ln N - \sum_j \left( \frac{\epsilon_j^2(\varphi)}{2E_j^2} \right) \approx \ln \hat{N} - \frac{\Delta^2}{2\sigma^2} \approx \ln N - \sum_j \left( \frac{\epsilon_j^2(\varphi_m)}{2E_j^2} \right) - \frac{\Delta^2}{2\sigma^2} \quad (12)$$

Assume equal "centric quality" derivatives.

$$E = E_j \quad \text{for all } j \quad (13)$$

This is a good approximation, for from the M. denitrificans analysis

$$\left. \begin{aligned} E(\text{PtCl}_4^{2-}) &= 74 \\ E(\text{Pt}(\text{CN})_4^{2-}) &= 80 \\ E(\text{UO}_2^{2+}) &= 73 \end{aligned} \right\} \quad (14)$$

then

$$\frac{\Delta^2}{\sigma^2} = \frac{\sum_j (\epsilon_j^2(\varphi) - \epsilon_j^2(\varphi_m))}{E^2} \quad (15)$$

Since

$$\langle \epsilon_j^2 \rangle = \frac{1}{n} \sum_j \epsilon_j^2 \quad (16)$$

$$\sigma^2 = \Delta^2 E^2 / n (\langle \epsilon_j^2 \rangle_\varphi - \langle \epsilon_j^2 \rangle_{\varphi_m}) \quad (17)$$

This allows a semi-quantitative prediction of various effects upon  $m$  and in turn upon  $\langle m \rangle$ . Now  $\langle \epsilon_j^2 \rangle$  is influenced by the quality of the data and the degree of isomorphism and it will not change drastically as the number of derivatives changes. Hence,  $m$  increases exponentially as  $1/n$  decreases.

$$m \simeq e^{-\sigma^2/2} = e^{-\frac{1}{2}(\Delta^2 E^2 / n) \cdot \text{constant}} \quad (18)$$

evaluated at some  $\varphi$ .

The dependence of  $m$  upon  $n$  suggests that the pdf should be rewritten to remove this purely statistical effect;  $\langle m \rangle$  should be normalized so that it depends only upon the phase agreement. The simplest approach would be to take

$$P(\varphi) = N e^{-\frac{1}{n} \sum_j (\epsilon_j^2 / 2E_j^2)} \quad (19)$$

then

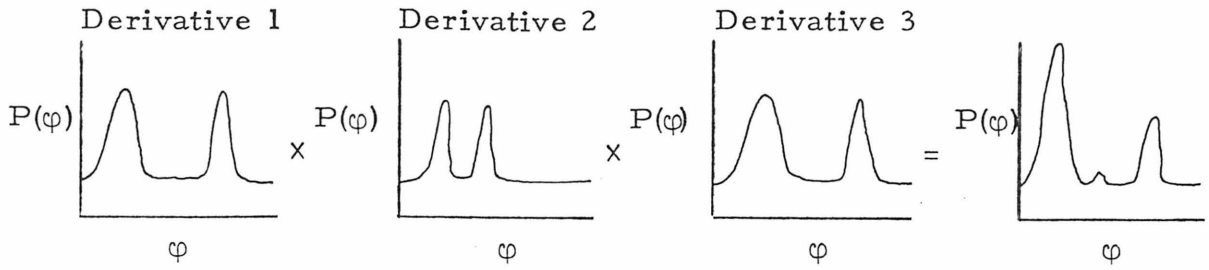
$$\langle m \rangle_{\text{normal}} = \langle m \rangle_{\text{old}}^n \quad (20)$$

That this would remove the  $n$  dependence is a testable result. In the  $\underline{c}_{550}$  analysis certain heavy atom derivatives were measured independently, but as later work indicated similar substitution, equivalent (but not identical) data sets were averaged. After refinement over the three averaged sets,  ${}_3\langle m \rangle = 0.59$ . A new refinement was performed whereby the redundant  $F_{\text{PH}}$  measurements were treated as independent derivatives. Such an approach has been frequently used on other structures. Then  ${}_7\langle m \rangle = 0.741$ . If the discussed dependency is correct, then

$$\langle m \rangle_{\text{norm}} = {}_3\langle m \rangle^3 = {}_7\langle m \rangle^7 \quad (21)$$

but  ${}_7\langle m \rangle^7 = 0.133$  and  ${}_3\langle m \rangle^3 = 0.205$ . In other words, the  $\langle m \rangle$  did not increase enough to be explained by a  $1/n$  dependency. The failure lies in one of the assumptions. For chemically distinct added derivatives, the form of  $P(\varphi)$  will more closely resemble a Gaussian distribution, since the ambiguous pair to the correct protein phase will not be correlated among distinct chemical derivatives and hence will average in the product to zero. For a redundant data set, where the added set is either a strict remeasurement or is similar to some previous set, the added set can correlate with an "ambiguous pair", leading to a bi-modal distribution.





This is carried to an extreme in the SIR case where a Blow-Crick treatment leads to a pdf containing two equal peaks separated by the difference between  $\varphi$  and the incorrect alternate angle. From a consideration of  $\varphi_b$  as the centroid angle, it is obvious that  $\varphi_b$  is the mid-angle of the two. Let  $\Delta'\varphi = 2(\varphi_b - \varphi_{\text{true}})$ . It will be shown that in the SIR case,  $m$  is largely determined by  $\Delta'\varphi$ . This suggests that for a pdf with a large bi-modal character,  $m$  is a function of  $\Delta\varphi$ . Let

$$\left. \begin{aligned} P(\varphi) &= p && \text{for } \varphi = \varphi_1 \\ &= 1 - p = q && \text{for } \varphi = \varphi_2 \\ &= 0 && \text{elsewhere} \end{aligned} \right\} \quad (22)$$

as the limiting case of a bi-modal distribution. Then, as in Dickerson, Kendrew, and Strandberg (1961)

$$\left. \begin{aligned} m \cos \varphi_b &= p \cos \varphi_1 + q \cos \varphi_2 \\ m \sin \varphi_b &= p \sin \varphi_1 + q \sin \varphi_2 \end{aligned} \right\} \quad (23)$$

Solution for  $m$  and taking  $\Delta\varphi = \varphi_1 - \varphi_2$  gives

$$m = \{p^2 + q^2 - 2pq \cos(180^\circ - \Delta\varphi)\}^{\frac{1}{2}} \quad (24)$$

If one calls  $p$  a vector with magnitude  $p$  and phase  $\varphi_1$  and does the same for  $q$ , then  $m = |\vec{p} + \vec{q}|$ . Note that if  $|\vec{p}| = \frac{1}{2}$  (the SIR case), then

this reduces to

$$m = \cos(\Delta\varphi/2) \quad (25)$$

Any bi-modal distribution can be treated as arising from a single SIR phase determination modulated by some not severe weighting function. Since, for SIR, it can be shown that

$$\varphi = \alpha \pm \cos^{-1} \left( \frac{F_{PH}^2 - (F_P^2 + f^2)}{2F_P f} \right) \quad (26)$$

so that

$$\frac{\Delta\varphi}{2} = 2 \cos^{-1} \left( \frac{F_{PH}^2 - (F_P^2 + f^2)}{2F_P f} \right) \quad (27)$$

It follows that in this limiting bi-modal case,  $m$  will be explicitly independent of the number of derivatives.

Take any actual  $m$  as approximately due to a combination of the limiting uni-modal and bi-modal cases,

$$m = k_1 e^{-\sigma^2/2} + k_2 \cos\left(\frac{\Delta\varphi}{2}\right) \quad (28)$$

where  $k_1, k_2$  are weighting terms describing the relative contributions. Attempts to normalize  $\langle m \rangle$  by raising to the  $n^{\text{th}}$  power are now seen as incorrect since only the first term should be so treated.

It is proposed that the most reasonable form for the normalized  $\langle m \rangle$  would be given by computing individual  $m$ 's from an adjusted pdf

$$P(\varphi) = e^{-\left(\sum_j k_j\right) \left[\sum_j (\epsilon_j^2 / 2E_j)\right]} \quad (29)$$

where the modifying constants  $k_j$  represent the fraction of electrons in derivative  $j$  which are unique to the entire set of heavy atom electrons. This is best explained by means of the following example.

Example Calculation of k

Derivative	Sites Occupied	Effective electrons	Unique Scattering electrons	k
Salt A	1	100	100+50+15=165	165/175
	2	50		
	3	25		
		<u>175</u>		
Salt B	1'	125	25+10+10=45	45/195
	2'	60		
	3'	10		
		<u>195</u>		
Salt C	4	75	75+15=90	90/150
	5	75		
		<u>150</u>		
Salt D	5'	100	25+60=85	85/160
	6	60		
		<u>160</u>		
n=4				

Under such a system, for example, a determination using six truly independent derivatives would have its pdf adjusted by a factor of  $\sum_j k_j = 1+1+1+1+1+1 = 6$  to counter the implicit 1/6 in the summation  $\sum_j e_j^2 / 2E_j^2$ . Treating two redundant measurements of the same derivative as pseudo-distinct derivatives does not have this 1/n factor and the summation  $\sum_j k_i = 1+0 = 1$  will account for this.

It is to be recalled that this is not necessarily a proposal to calculate m for use in figure of merit weighted Fourier analyses. It is at this stage intended as a way to re-calculate m prior to reporting

$\langle m \rangle$  so as to make the statistic comparable over different protein determinations.

Additional work is clearly necessary on this proposal. It should be examined theoretically to insure that it reasonably accounts for the uni-modal and bi-modal contribution to  $m$ . It would be most interesting to calculate  $\langle m \rangle$  according to this outline for a variety of protein structure determinations then compare the resulting predictions of quality with other factors which assess reliability, such as R-factors, percentage E's, and electron density background-to-noise ratios.

#### References

- Blow, D. M., and Crick, F. H. C., (1959). Acta Cryst., 12, 794.
- Dickerson, R. E., Kendrew, J. C., and Strandberg, B. E., (1961).  
in Computing Methods and the Phase Problem in X-ray  
Crystal Analysis, R. Pepinsky, J. M. Robertson, and  
J. C. Speakman, eds., Pergamon Press, New York, p. 236.