STRUCTURE DETERMINATION AND REFINEMENT OF *PSEUDOMONAS AERUGINOSA* CYTOCHROME c_{551} AT 2.0 Å RESOLUTION

Thesis by

Robert J. Almassy

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1978

(Submitted 25 May 1978)

ACKNOWLEDGMENTS

I would like to thank my friends and associates, especially Olga Kallai, Mary Kopka, Rosemarie Swanson, Benes Trus, and Tsune Takano, for their help and suggestions during the course of this work. Special thanks are extended to Dee Barr for her extraordinary assistance in preparing this thesis.

I would also like to thank John Chambers for the use of his refinement software and many helpful discussions. The friendship of John, Steve Spencer, Dick Vandlen, and Ralph Staley was particularly valued.

I am extremely fortunate in having wonderful parents and family. I am very grateful for their support and encouragement.

Finally, I would like to extend my sincere thanks to Professor Richard Dickerson for his efforts, guidance, and enthusiasm during the course of this work.

Financial support of the National Institutes of Health and the California Institute of Technology made this work possible.

ABSTRACT

The crystal structure of the protein cytochrome c_{551} from Pseudomonas aeruginosa has been solved by the method of multiple isomorphous replacement and refined by constrained difference Fourier methods to an R-factor of 16.2% at 2.0 Å resolution.

Crystallization conditions were found at pH 5.6. Three heavy atom derivatives (K_2PtCl_4, $UO_2(NO_3)_2$, and $NaAu(CN)_2$) were used in the multiple isomorphous replacement x-ray diffraction phase analysis. A low resolution protein electron density map was calculated. The protein main chain was continuous and easily interpreted. The location of the heme and a few side chains were also determined. The resolution was extended to 2.4 Å and a wire model was built to the electron density. The map quality was good and almost all atoms fit into strong density. These coordinates were improved by refinement at 2.0 Å resolution to an R-factor of 16.2%. Additional structural information was obtained, including individual isotropic temperature factors, estimates of coordinate errors, and solvent positions.

c_{551} has major changes, compared to eukaryotic c, in the number (82 compared to 103) and the sequence of amino acids. Residues which have been highly conserved in c-type cytochrome structures are altered or deleted in c_{551}. Homology alignment predictions based on sequence are discussed. Despite these differences, the overall main chain fold is conserved, with only one major deletion and some minor deletions and additions.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

INTRODUCTION

Eukaryotic cytochrome c is a water soluble heme protein containing 103 amino acids. It is located on the inner mitochondrial membrane and is one of the proteins involved in extracting energy from electron transport. Amino acid sequences of cytochromes c have been determined for 67 eukaryotic species. From sequence comparisons alone, the homology was clear (1). The determination of the crystal structures of horse and bonito cytochrome c proved the tertiary structures were identical (2).

Sequence determinations from c-type cytochromes of prokaryotic species found large differences in the protein chain lengths and the amino acid compositions. One of these, c_2 from Rhodospirillum rubrum, had 112 amino acids and a sequence similar enough to eukaryotic c for the homology to be clear (3). The subsequent x-ray structure determination proved this homology (4). Another prokaryotic cytochrome c from Paracoccus denitrificans, was also found to be homologous despite the numerous insertions of amino acids to produce the chain length of 134 residues (5).

Another class of cytochromes were found to have the typical Cys - x - y - Cys - His sequence and a Met axial heme ligand. These cytochromes c_{551} contained only 82 residues, with many of the previously conserved sites apparently altered significantly. Partial or full sequence homology had been proposed (6-8), but these predictions differed significantly and were far from conclusive. In

order to determine the relationship between these $\underline{c}_{551}$ proteins and eukaryotic cytochromes, the x-ray crystal structure determination of Pseudomonas aeruginosa $\underline{c}_{551}$ was started. These results would precisely answer the question of which residues in $\underline{c}_{551}$ were homologous to eukaryotic c residues. This would provide information about the location of important conservative and invariant residues. The differences in structures could also be correlated with differences in chemical and biological properties, which is valuable in the determination of the mechanism of electron transport for cytochrome c.

The proper crystallization conditions were discovered and three heavy atom derivatives were prepared. The crystal structure was first solved at low resolution (Chapter 2). From this electron density map, the main chain fold was correctly interpreted (9). The location of some key side chains were also seen. The evolutionary relationship to eukaryotic cytochrome c was clearly established, and an approximate sequence alignment was made.

After the crystal structure was solved at low resolution, data were collected to 2.0 Å for the native protein and to 2.4 Å for three derivatives. After refinement of heavy atom parameters, the MIR phases and the 2.4 Å resolution electron density were calculated. In this electron density map, the main chain was continuous and well defined. The locations of all side chains were clearly seen extending from the main chain. This permitted the accurate sequence alignment to be made between this bacterial cytochrome and those from eukaryotic species. The alignment, based on tertiary structure, displayed the

location of significant changes in sequence of residues previously thought to be conserved or invariant in c-type cytochromes.

The accuracy of the wire model coordinates obtained from the 2.4 Å resolution map was improved by refinement to an R-factor of 16.2% at 2.0 Å resolution. In addition to increasing coordinate accuracy, refinement provided additional structure information. Temperature factors, solvent binding locations, and estimated of coordinate errors were determined.

REFERENCES

1. Dickerson, R.E. (1972) Scientific American 226, 58-72.
2. Dickerson, R.E., Tanako, T., Eisenberg, D., Kallai, O.B., Samson, L., Cooper, A. and Margoliash, E. (1971) J. Biol. Chem. 246, 1511-1535.
3. Dus, K., Sletten, K. and Kamen, M.D. (1968) J. Biol. Chem. 243, 5507-5518.
4. Salemme, F.R., Freer, S.T., Xuong, Ng.-H, Alden, R.A. and Kraut, J. (1973) J. Biol. Chem. 248, 3910-3921.
5. Timkovich, R. and Dickerson, R.E. (1973) J. Mol. Biol. 79, 39-56.
6. Needleman, S.B. and Blair, T.T. (1969) Proc. Nat. Acad. Sci. USA 63, 1227-1233.
7. Dickerson, R.E. (1971) J. Mol. Biol. 57, 1-15.
8. McLachland, A.D. (1971) J. Mol. Biol. 61, 409-424.
9. Dickerson, R.E., Timkovich, R. and Almassy, R.J. (1976) J. Mol. Biol. 100, 473-491.

Chapter 2

LOW RESOLUTION STRUCTURE ANALYSIS

Purification and Crystallization

Pseudomonas aeruginosa cytochrome c_{551} was prepared from 100 liter cultures grown under the conditions of Harbury (1) and Ambler (2-4). The cells were grown for eighteen to twenty hours at 37°C in a medium containing sodium citrate, KH_2PO_4, $MgSO_4$, $NaNO_3$, and yeast extract. Anaerobic conditions, with no agitation of the culture, and nitrate as the terminal electron acceptor produced the highest yields of cytochrome c_{551}. Cells were harvested into a dry acetone cell powder, which was stored under vacuum in a desiccator at -20°C. The cytochromes were then extracted from the cells with alumina in 0.10 M $NH_4(CH_3COO)$ following the procedures of Harbury (1).

After the final purification step, which separated cytochrome c_{551}, c_{554}, and azurin by elution through a column of Sephadex SP C-25, the cytochrome c_{551} was lyophilized for concentration and storage.

From previous work on horse cytochrome c, it was thought that lyophilization might impede cytochrome crystallization. Accordingly, two modifications replaced the final lyophilization step in order to eliminate this as a possible source of trouble in crystallization: (a) the cytochrome c_{551} was stored on Sephadex SP C-25 resin at -20°C instead of as a lyophilized powder; (b) pressure dialysis of 1 to 2 ml of cytochrome c_{551} through Amicon Diaflow UM-2 membranes at 100 psi of $N_2(g)$ was employed to concentrate the protein after elution from the storage resin. Protein concentrations up to 5% were obtainable without the lyophilization procedure.

Horio et al (5,6) were the first to crystallize c_{551}. They found that crystals could be grown from cytochrome after oxidation with potassium ferricyanide, or reduction with sodium dithionite. Needle-shaped microcrystals, with approximate dimensions of 0.025 mm by 0.001 mm by 0.001 mm, were formed in a 30 to 40% $(NH_4)_2SO_4$ solution at pH 7 to 8 in the presence of sodium dithionite. Cytochrome c_{551} could also be crystallized into needle and octahedral morphologies from $(NH_4)_2SO_4$ without being reduced, but with a lower yield.

In this study, three methods of crystallization were used in attempts to produce single crystals of cytochrome c_{551} of suitable size for high resolution x-ray diffraction analysis. Suitable crystallization conditions were tested by using the micro membrane diffusion method of Zeppezauer (7) and a vapor diffusion method (8). Miniature batch crystallizations, with approximately 100 µl of protein were also used after the appropriate conditions were found.

The vapor diffusion method never produced crystals and was soon discarded in favor of other methods. It suffered from the disadvantage that the pH could be adjusted only by adjusting the pH of the stock protein solution prepared before vapor diffusion. With small volumes of stock solution (60 to 100 µl), this was extremely difficult. In the Zeppezauer procedure, the protein solution rapidly came to pH equilibrium with the readily adjustable reservoir solution.

The membrane diffusion cell was constructed with an inside diameter of 1 mm and a 30 mm length, containing a volume of approximately 20 µl. It was prepared for use by cleaning in a H_2SO_4/HNO_3 acid

solution for approximately 24 hours and then in saturated NaOH for one hour. After the clean cell was filled with protein solution, one end was sealed with Parafilm, held in place with Tygon tubing. The other end was sealed with Union Carbide cellulose casing membrane. Cellulose casing, which reportedly has a pore size smaller than standard dialysis tubing (9), was used to prevent leakage of the 8000 dalton molecular weight cytochrome $\underline{c}_{551}$. The Zeppezauer cell, with the membrane positioned upward, then was placed in a vertical test tube containing the reservoir solution.

The search for the appropriate crystallization conditions began using a 2% ferricytochrome $\underline{c}_{551}$ solution and a reservoir solution containing 1.0 M NaCl, 50% saturated $(NH_4)_2SO_4$, and 0.01 M $(NH_4)_2HPO_4$. The stock protein solution generally contained the same salt concentrations as the initial reservoir solution. This particular initial $(NH_4)_2SO_4$ concentration was chosen because of the success of Horio et al. (5,6) in obtaining microcrystals at this concentration. The NaCl was used because of its importance in the crystallization of c-type cytochromes from horse (10), tuna (11,12), and Paracoccus denitrificans (13). The $(NH_4)_2SO_4$ concentration was incremented by 2½% to 5% of saturation at various time intervals. The pH range of 4.8 to 7.5 was surveyed initially.

Under the above conditions, crystallization cells within the pH range of 5.6 to 5.9 produced small crystals. These crystals were found to belong to space group $P2_12_12_1$ with cell constants $\underline{a}$ = 29.43 Å, $\underline{b}$ = 49.00 Å, and $\underline{c}$ = 49.66 Å, with one molecule per asymmetric unit.

(0,k,ℓ) and (h,k,0) diffraction patterns recorded with a precession camera are shown in Figure 1.

These crystals had two major problems. First, the crystals were either too small to be used for high resolution x-ray analysis or they were severely intertwined without the possibility of being separated. Second, the conditions and procedure described above for crystal growth could not consistently produce crystals and, in fact, yielded them in only a small percentage of the trials.

The crystals have rectangular morphology with one long axis, parallel to the crystallographic a axis, and two near equal and relatively short axes, parallel to the b and c unit cell vectors. In single crystals, the long crystal length would generally be around 0.2 to 0.3 mm and the other dimensions approximately a tenth as long. Very rarely a single crystal of dimensions up to 0.8 x 0.15 x 0.15 mm would be formed.

In attempts to improve the crystal quality, the effect of NaCl and $(NH_4)_2HPO_4$ concentrations were examined. In all crystallization cells which did not contain NaCl, crystals were not produced. However, it cannot be stated unequivocally that NaCl is necessary for crystal growth, since crystals also failed to grow in many cells under the same conditions, but which also contained NaCl. NaCl concentrations from 0.50 M to 3.0 M all produced crystals, but the average quality seemed independent of the NaCl concentration within this range. As with the NaCl concentration, the $(NH_4)_2HPO_4$ concentration, on the average, did not appear to have any effect on the crystal quality

within the range of 0.0 M to 1.0 M.

The protein concentration did have a significant effect on crystal growth. As the concentration was increased to 5%, crystal size increased. However, this increase in size was also accompanied by increased precipitation of disordered protein and extensive crystal twinning. Occasionally, a single crystal of usable size was produced but barely often enough for a structure analysis.

Of the many procedures used in attempting to alleviate this problem, one proved to be particularly useful. Initially, the stock protein solutions were not Millipore filtered because of the small volumes (0.1 ml) usually handled. Instead, they were centrifuged at 7000 rpm with a Sorvall centrifuge (rotor SM 24) in a 2 ml centrifuge tube in an adaptor. This removed any large particulate contamination but apparently did not remove the submicroscopic particles from which amorphous precipitation was produced. A method was developed which allowed volumes of protein as small as 0.1 ml to be Millipore filtered with more than 85% recovery. A Millipore swinny syringe filter holder with a 0.5 μm pore size and 13 mm diameter was filled with 0.1 ml of stock protein solution. A 2 ml centrifuge tube and adaptor, cut off so as to hold approximately 0.4 ml, and the filter holder assembly were inserted into the rotor hole. The protein was forced through the filter by centrifugation at the minimum speed of a Sorvall RC-2 centrifuge. The filtrate collected in the tapered centrifuge tube could then be transferred quantitatively to the crystallization cells. The high recovery rate and efficient removal of particles

of sizes 0.5 μm and larger from very small volumes of protein solutions proved very useful in this study. This procedure should be useful in other protein crystallographic studies that are limited to small amounts of protein.

In addition to Millipore filtering the protein solution, gentle seeding was also done. This was accomplished by touching a finely drawn glass needle point to a $\underline{c}_{551}$ crystal and then touching this tip to the top of the protein solution in the crystallization cell.

The combination of Millipore filtration and seeding relieved the problem of reproducible crystal growth. With this procedure crystals grew from a clean solution as opposed to growing out of amorphous precipitate. This occured even at higher ranges of protein concentration, which helped in producing larger crystals. The crystal twinning problem also was somewhat improved. All these observations are consistent with the removal of excell microscopic and submicroscopic nucleation sites, permitting crystal growth to occur from only a very limited number of sites.

In addition to the experiments to produce crystals of ferricytochrome $\underline{c}_{551}$, crystallization trials of ferrocytochrome $\underline{c}_{551}$ were set up. The crystallization conditions were the same as for the oxidized protein except for a lower initial $(NH_4)_2SO_4$ concentration (30%), due to the lower solubility of the reduced protein (6), and sodium dithionite as a reducing agent. No crystals were produced in a limited number of trials. However, considering the reproducibility

problem discussed above, future crystallization attempts might be successful under conditions similar to these.

Success was attained in producing ferrocytochrome $\underline{c}_{551}$ crystals by reduction of ferricytochrome $\underline{c}_{551}$ crystals with an excess of $Na_2S_2O_4$. The color change upon addition of $Na_2S_2O_4$ indicated the crystal was reduced. The reduced crystal showed no macroscopic deterioration and diffracted as well as the oxidized crystals of comparable quality. A diffraction pattern is presented in Figure 1c. The intensity changes from the oxidized to the reduced diffraction pattern are substantial. With the exception of tuna ferrocytochrome $\underline{c}$ (12), this was the first $\underline{c}$-type cytochrome capable of changing oxidation state without either destroying the crystal or having no significant intensity changes. For tuna, no significant conformational changes could be detected. With the size of the intensity changes seen for $\underline{c}_{551}$, the existence of conformational changes is likely. This structure may provide a significant comparison after the structure of ferricytochrome $\underline{c}_{551}$ is solved.

Heavy Atom Derivatives

The general procedure of screening for heavy atom derivatives was to soak one or two pre-grown crystals in a 2-ml solution of 75% $(NH_4)_2SO_4$, 1.0 $\underline{M}$ NaCl, and the heavy atom complex. The phosphate buffer used during crystal growth was ordinarily omitted from the soaking solution to prevent undesirable reactions with the heavy atom complex.

Twenty-three difference compounds, containing eight different heavy atoms were studied to find suitable derivatives of cytochrome c_{551}. The types of heavy atom compounds, crystal-soaking conditions, and crystal-soaking times were selected so as to have the best possible chance of success. The summary of Baumber et al. (14) of successful soaking conditions in other protein crystallographic studies proved very useful. The heavy atom compounds and their concentration ranges are listed in Table I. The soaking times varied but generally were several days or longer. The pH of the soaking solution was adjusted to the range of 5.6 to 5.9 before the crystal was added. In a few cases, the pH changed significantly during the soaking process and had to be readjusted.

To determine the success of the derivation reaction, the x-ray diffraction pattern was recorded photographically with a precession camera, with a crystal-to-film distance of 90 mm, and with Ni-filtered Cu x-rays produced by a standard Cu target G.E. CA-8S sealed tube operated at 45 KV and 15 ma.

The crystals were mounted in thin-walled lithium borate glass capillary tubes, which were then attached to solid brass pins with a low melting point wax. Most of the crystal-soaking solution was drawn away from the crystal, leaving the crystal with only enough liquid to help prevent dehydration of the crystal. A column of soaking solution was placed at the other end of the capillary tube to maintain a constant humidity within the tube. The tube was then sealed with wax and the brass pin placed in a standard eucentric goniometer head. The

TABLE I: Heavy Atom Derivative Survey

		Concentration Range (mM)	
I.	Gold		
	$NaAu(CN)_2$	0.50	5.00
	$NaAuCl_4 \cdot 2H_2O$	0.05	5.00
	$KAuBr_4$	0.10	5.00
	$KAuI_4$	0.50	5.00
II.	Iridium		
	Na_3IrCl_6	5.00	50.00
	$K_3Ir(NO_2)_6$	5.00	
	$K_3Ir(CN)_6$	5.00	
	$(NH_4)_2IrCl_6$	5.00	
III.	Mercury		
	K_2HgI_4	0.01	0.12
	Mersalyl Na	0.50	5.00
	PCMPS*	5.00	
IV.	Osmium		
	Na_2OsO_4	5.00	
	$OsCl_3$	0.50	5.00
V.	Platinum		
	K_2PtCl_4	0.05	5.00
	$K_2Pt(NO_2)_4$	0.05	5.00
	$K_2Pt(SCN)_4$	0.05	5.00
	$(NH_4)_2Pt(CN)_4$	0.03	1.50
VI.	Samarium		
	$SmCl_3$	15.00	
VII.	Silver		
	Ag_2SO_4	0.50	5.00
VIII.	Uranium		
	$UO_2(NO_3)_2 \cdot 6H_2O$	1.00	50.00
	$UO_2(CH_3CO_2) \cdot 2H_2O$	1.40	5.82
	$UO_2(SO_4) \cdot 3H_2O$	0.50	1.20
	$K_3UO_2F_5$	3.50	5.00

* PCMPS = p-chloromercuriphenylsulfonic acid Li salt.

crystals were aligned optically to within a couple of degrees, followed by alignment with an accuracy greater than ten minutes, by the standard procedure of still and 3° precession screenless x-ray exposures.

Only crystals which were small in size or poor in quality were available for heavy atom derivative screening. This required exposures from 75 to over 100 hours for a typical 17° precession photograph, giving high backgrounds. From these photographs, it was sometimes difficult to determine if intensity changes were in fact real or if they were a result of crystal decay and background effects. Alignment also was a serious problem with these crystals.

Data Collection

From the twenty-three different heavy atom complexes examined, K_2PtCl_4, $K_2Pt(SCN)_4$, $K_2Pt(NO_2)_4$, $UO_2(NO_3)_2 \cdot 6H_2O$, and $NaAu(CN)_2$ resulted in clear, significant changes in intensities. The pattern of changes produced by the three platinum derivatives indicated that they might possibly share the same heavy atom sites and therefore would not yield three useful derivatives for isomorphous replacement phase analysis. Due to these similar diffraction changes and the severe shortage of crystals of sufficient quality for diffraction data collection, only the K_2PtCl_4, $UO_2(NO_3)_2 \cdot 6H_2O$, and the $NaAu(CN)_2$ derivatives were used for intensity data collection.

The crystal dimensions, salt conditions, and other crystal data are given in Table II for the native and derivative crystals from

TABLE II: Crystal Data

	Crystal Size (mm)	Diffractometer	Salt Conditions	Heavy Atom Concentration	Soaking Time
Native	0.5 x 0.15 x 0.15	G.E.	75% $(NH_4)_2SO_4$ pH = 5.5	---	2 days
K_2PtCl_4	0.4 x 0.15 x 0.15	$P\bar{1}$	75% $(NH_4)_2SO_4$ 1.0 M NaCl pH = 5.5	0.2 mM	6 days
$UO_2(NO_3)_2 \cdot 6H_2O$	0.6 x 0.15 x 0.15	$P\bar{1}$	75% $(NH_4)_2SO_4$ 1.0 M NaCl pH = 5.5	25.0 mM	6 days
$NaAu(CN)_2$	0.3 x 0.05 x 0.05	$P\bar{1}$	75% $(NH_4)_2SO_4$ 1.0 M NaCl pH = 5.6	1.0 mM	10 days

which data were collected. Data for the native protein were collected in 75% $(NH_4)_2SO_4$ prior to finding the best solvent conditions for the derivative reactions. It was subsequently found that the K_2PtCl_4 derivative diffracted better after being soaked in 1.0 M NaCl in addition to the 75% ammonium sulfate salt solution. Without the 1.0 M NaCl, the intensity changes were much greater (approximately 37% in one case) and the unit cell parameters changed more from those of the native crystal. This indicated that the derivative was less isomorphous. The differences in solvent between the derivatives with NaCl and the native without NaCl might have caused problems in the heavy atom solution.

Two data collection systems were used, as indicated in Table III. A modified General Electric XRD-490 quarter-circle diffractometer, automated with a DEC PDP-8 computer, was used. CuK_α x-radiation was generated with a G.E. CA-8S copper target x-ray tube, operated at 45 KV and 15 ma, and a graphite plane monochromator operated in the perpendicular mode, in order to minimize intensity fluctuations in the 2θ plane. The other data collection system was a Syntex $P\bar{1}$ autodiffractometer, controlled by a Data General NOVA 1200 computer. X-radiation was produced and filtered as described above, except a Phillips fine focus Cu target tube was used.

Crystals were mounted and sealed in capillary tubes as described earlier, but always with their needle axis (a axis) parallel to the capillary axis. This allowed capillary tubes with a smaller diameter to be used. Although the crystals were more difficult to mount, the

TABLE III: Data Collection Parameters

	2θ Range	Scan Width(ω)	Scan Time(sec)	% <0	% Observed (>2σ)	% ΔF
Native	2.00-22.14	0.60°	62	6.2	77.0	--
	22.14-29.00		93	7.2	70.0	--
	29.00-31.76		93	--	--	--
K_2PtCl_4	2.00-22.50	0.50°	60	4.3	88.7	17.1
	22.50-29.00		60	6.5	69.5	16.4
$UO_2(NO_3)_2 \cdot 6H_2O$	2.00-22.50	0.45°	81	2.7	92.9	24.6
	22.50-29.00		81	3.5	85.3	24.3
$NaAu(CN)_2$	4.50-10.00	0.48°	288	13.4	81.4	33.0
	10.00-17.50		288	13.4	81.4	33.0
	17.50-20.00		288	0.0	95.9	33.0
	20.00-22.50		360	43.7	51.9	33.0

smaller capillaries resulted in less x-ray background scattering (15), which was especially important for the background correction method described below.

The crystals were aligned on the diffractometer first optically and then by using standard procedures of measuring x-ray reflections (16). The more automated system on the Syntex $P\bar{1}$ proved more convenient than the G.E. system. After the crystal was aligned, several reflections at representative positions in reciprocal space were scanned and the crystal unit cell parameters were determined by the method of least squares. The size of the counter collimator pinholes were adjusted so as to obtain the maximum signal-to-background ratio for peak scans.

Finally, before data collection was initiated, the (8,0,0) reflection at $\chi = 90°$ was scanned at ϕ angles of 0° through 360° at increments of 10°. This was used later for x-ray absorption corrections.

All data were collected by step scans through the ω angle. The scan width was determined by scanning some of the most intense reflections, determining the width of the reflections, and adding a safety margin. This resulted in typical scan widths of 0.4° to 0.6°. Reflection scan times were calculated for each particular shell and crystal so as to have counting statistics such that a satisfactory percentage of the reflections were observed, by the 3σ criteria. Having set the scan width and time, the number of ω steps was determined so as to optimize the ratio of the total counting time to the total scanning time and still have a sufficiently fine ω scan.

A set of three standard reflections for the G.E. diffractometer and five standard reflections for the $P\bar{1}$ diffractometer were measured every 75 to 100 reflections in order to test for crystal misalignment and decay. Backgrounds were measured on both sides of the peak, displaced by 0.2° in ω from the scan limites. Short background times of ten to twenty seconds were used. The actual background used in data reduction was then calculated from a nonlinear least-squares fit to all the measured backgrounds as described below.

Data Reduction

Observed structure factors (F_{obs}) were calculated from net intensities (I_{net}) and Lorentz and polarization factors. I_{net} was computed from the integrated peak scans and background corrections.

The backgrounds were calculated by a nonlinear least-squares fitting procedure which fit all the measured backgrounds to a function of the diffractometer angles (15). Fitting all the backgrounds, measured for relatively short periods of time, to the smooth least-squares function improved their accuracy to the quality that would have been obtained by measuring backgrounds for times equal to scan times. The standard deviation calculated from the least-squares fit was usually better than a tenth that based on the counting statistics for the short backgrounds. A summary of some statistics for backgrounds calculated by the least-squares method is given in Table IV.

TABLE IV: Summary of Background Statistics

Crystal/ 2θ Range[1]	Background Function Terms[2] T:1234 C:0000	1234 1111	1234 2222	1234 3333	AVEΔ[3]	RMSΔ[4]	AVEΔ/BG[5]	AVE|Δ|/BG[5]	G.O.F.[6]	s[7]	$\langle\Delta^2\rangle$[8]	$\langle\sigma^2\rangle$[9]
Native												
2-22	1110	1110	0100	1000	28.60	35.89	.0109	.0833	1.1050	13.76	1288	1098
2-22	1100	1110	0100	1000	30.77	38.80	.0120	.0885	1.1750	20.17	1505	1098
22-29	1110	1110	--	--	47.92	59.09	.0086	.0754	--	22.16	3492	3000
K_2PtCl_4												
2-22	1100	1110	0100	--	43.33	55.93	.0073	.0719	1.0996	26.22	3128	2440
2-22	1110	1110	0100	1000	42.62	55.86	.0077	.0705	1.0893	26.08	3120	2440
2-22	1100	1110	0100	1000	43.11	55.77	.0075	.0715	1.0964	25.89	3110	2440
22-32	1110	1110	--	--	33.09	41.66	.0103	.0814	1.0163	8.6173	1735	1661
$UO_2(NO_3)_2 \cdot 6H_2O$												
0-22½	1100	1110	0100	--	38.14	47.61	.0041	.0545	1.0750	17.13	2267	1973
0-22½	1100	1110	0100	1000	38.11	47.57	.0041	.0545	1.0750	17.03	2263	1973
0-22½	1110	1110	0100	1000	37.33	47.03	.0043	.0530	1.0519	15.43	2211	1973
22-29	1110	1110	--	--	29.87	37.82	.0057	.0602	1.0153	9.1915	1430	1346
$NaAu(CN)_2$												
10-22½	1000	1110	--	--	164.78	212.41	.0050	.0593	1.1247	129.06	45117	28458
					120.51	151.00	.0055	.0592	1.0382	47.46	22799	20547
					169.48	220.28	.0054	.0572	1.0691	107.47	48521	36969
10-22½	1110	1110	--	--	156.12	201.65	.0049	.0557	1.0938	110.47	40662	28458
					122.34	152.35	.0052	.0600	1.0458	51.59	23209	20547
					--	--	--	--	--	--		
10-22½	1000	0110	--	--	166.41	213.72	.0051	.0599	1.1270	131.21	45675	28458
					120.34	151.02	.0055	.0591	1.0361	47.53	22807	20547
					169.08	220.57	.0054	.0571	1.0688	108.08	48650	36969

[1] Underlined 2θ ranges indicate these backgrounds were used for calculating structure factors.

[2] Independent variables to be included in the background function. T is the power of 2θ, and C is the power of χ for each term. A 1 indicates the term is included in the function.

[3] The average deviation between the observed and calculated background.

[4] The root-mean-square deviation.

[5] The fractional deviation.

[6] Goodness-of-fit.

[7] $s = \sqrt{(8)-(9)}$

[8] The average square of the deviation between the observed and calculated background.

[9] The average square of the standard deviation.

Absorption corrections, based on the (8,0,0) reflection measured at $\chi = 90°$ and ϕ angles at 10° increments, were applied to I_{net} using the procedure of North et al. (17). The Lorentz factor, $L = (\sin 2\theta)^{-1}$, for diffractometer data collection geometry (16) was applied to I_{net}. The polarization factor, P, for x-ray monochromatized by a graphite plane monochromator crystal in the perpendicular geometry was also applied (18,19):

$$P = \frac{\cos^2 2\theta + \cos^2 2\theta_{mono}}{1 + \cos^2 2\theta_{mono}}$$

The time decay, determined by the standard reflections measured every 75 to 100 reflections, was found to be insignificant and no correction was made.

The standard deviations of the F_{obs} were calculated as follows (16):

$$\sigma_F = \frac{1}{2}\left(\frac{K}{LP}\right)^{\frac{1}{2}} \sqrt{\frac{N_{scan} + N_{bkg} + 0.02\, N_{net}}{N_{net}}}$$

where K is a scale constant and N's are x-ray counts.

The resulting F_{obs} for the three 2θ shells of native protein data were combined into a master file and put on an arbitrary scale. The derivative data shells were then scaled to the corresponding native

data shells, using the scaling equation:

$$F_P = K \times F_{PH} \times \exp^{(-Bs^2)}$$

where F_P and F_{PH} are the native and derivative structure factor amplitudes, K is the linear scaling constant, B is the exponential falloff scaling constant, and $s = 2\sin\theta$. K and B are determined by a linear least squares fitting to a Wilson type plot of $\ln \{\langle F_P \rangle / \langle F_{PH} \rangle\}$ versus s^2 (20).

Heavy Atom Solution

Difference Patterson maps (21) were used to determine the coordinates of the heavy atom binding sites. Three-dimensional difference Pattersons, using coefficients $A = (\Delta F)^2$ and $B = 0$, were calculated for the K_2PtCl_4 and $UO_2(NO_3)_2 \cdot 6H_2O$ derivatives. Data to a resolution of 4 Å ($2\theta = 22.5°$) were included.

The three Harker sections for both derivatives are shown in Figures 2 and 3. The Harker peaks, for space group $P2_12_12_1$ are:

1/2 ± 2x	± 2y	1/2
± 2x	1/2	1/2 ± 2z
1/2	1/2 ± 2y	± 2z

The K_2PtCl_4 difference Patterson Harker section at w = 1/2 indicated one major vector peak and two minor vector peaks. Peak A was

consistent with the major Harker peaks A' on the Harker section v = 1/2 and A'' on the Harker section u = 1/2. A'' was quite clear while A' was somewhat distorted. All other vector peaks did not produce peaks on the other Harker sections at appropriate coordinates. For example, vector peak B should have generated a peak on the indicated line on Harker sections v = 1/2 and u = 1/2. Vector peaks C and D seemed to be simply accumulation of noise at the special positions (0,0,1/2) and (1/2,0,0), while E was related to A.

The $UO_2(NO_3)_2 \cdot 6H_2O$ difference Patterson Harker section w = 1/2 had two major peaks, A (21 contours) and B (15 contours). Peak A was consistent with the major Harker peak A' on the Harker section v = 1/2 and A'' on the Harker section u = 1/2. Peak A was also consistent with the smaller A_1' Harker peak at the v = 1/2 section and the A_1'' Harker peak at the u = 1/2 section. A_1' and A_1'' were found to be noise by the subsequent single isomorphous replacement (SIR) analysis. The relatively large B vector peak on the w = 1/2 Harker section did not correspond to any vector peaks at symmetry consistent locations on the v = 1/2 and u = 1/2 Harker sections. In addition, the C peak at v = 1/2 did not have consistent peaks at w = 1/2 and u = 1/2. The D peaks did have the proper Harker symmetry and indicated a possible minor binding site.

Although the three-dimensional difference Patterson maps and the Harker sections in Figures 2 and 3 did indicate the locations of some possible heavy atom sites, they were generally too noisy to be used as is with any confidence. Some peaks on the Harker sections were

completely consistent with peaks on other Harker sections, while there were also other major peaks which were not consistent. These unexplained peaks could simply be noise resulting from errors in the data, series termination errors, lack of isomorphism, or errors inherent to the difference Patterson technique. In addition to the heavy atom contribution to the ΔF terms, there were also contributions from the difference in the solvent (1.0 M NaCl) between the native data and the derivatives. Some of these sources of error will be discussed later.

Difference Fourier maps were calculated in order to verify that the symmetry consistent large peaks in the difference Pattersons are really results of heavy atom vectors. These maps were also used to distinguish other peaks resulting from additional heavy atom sites from noise peaks. The Fourier phases used were the SIR phases from refined heavy atom sites which correspond to a consistent Harker vector set. Although SIR phases contain much larger errors than multiple isomorphous replacement (MIR) phases, they are very useful in solving for heavy atom positions (21).

The resulting SIR phased ΔF Fourier maps will also allow the atomic coordinates from the Pattersons to be placed on a common origin and will assist in locating minor sites.

The K_2PtCl_4 derivative produced cleaner Harker sections than the $UO_2(NO_3)_2 \cdot 6H_2O$ derivative, and sites A, A', and A'' formed the peaks with the largest relative peak heights. Coordinates based on this substitution site were subjected to six cycles of alternating least-squares refinement and SIR phase calculations. The results are shown

in Table V. A $UO_2(NO_3)_2 \cdot 6H_2O$ difference Fourier was then computed using these phases and the coefficients listed in Table V. A pseudo projection down the $\underline{z}$ axis of the resulting map is displayed in Figure 4, with the contours being displayed only for the $\underline{z}$ section of maximum density for each peak. The result was one major peak (A), several minor peaks only slightly above the general noise level of the map, and no peak higher than the general noise level at the coordinates of the K_2PtCl_4 phasing site. This major peak corresponded to the Harker vectors labeled A, A', and A'' in the $UO_2(NO_3)_2 \cdot 6H_2O$ difference Patterson in Figure 3. No other site was obvious and the inclusion of minor $UO_2(NO_3)_2 \cdot 6H_2O$ binding sites was postponed until the more accurate MIR phases were obtainable.

The same procedure was carried out using the $UO_2(NO_3)_2 \cdot 6H_2O$ SIR phases and K_2PtCl_4 ΔF terms to calculate a K_2PtCl_4 difference Fourier. Four cycles of $UO_2(NO_3)_2 \cdot 6H_2O$ SIR phase refinement were computed based on the major peak in the $UO_2(NO_3)_2 \cdot 6H_2O$ ΔF map (site A in Figure 4) and the corresponding vector peaks (A, A', and A'') in Figure 3. The refinement parameters are listed in Table V, and the resulting K_2PtCl_4 ΔF Fourier map is displayed in Figure 5. This pseudo projection ΔF map was contoured at a lower level and has a lower noise level than the corresponding $UO_2(NO_3)_2 \cdot 6H_2O$ ΔF map.

This smaller scale results from the smaller figure of merit from the $UO_2(NO_3)_2 \cdot 6H_2O$ phase analysis. The smaller figure of merit could indicate that there is either more error in the $UO_2(NO_3)_2 \cdot 6H_2O$ ΔF's or that the derivative is less isomorphous. This was consistent with

TABLE V: Single Isomorphous Replacement Phase Refinement

ΔF Map[1]	Phases	Atomic Parameters										No. of Cycles	FM / FM	R_K^3 / R_K^3	R_C^4 (centric) / R_C^4 (all)
		Initial					Refined								
		x	y	z	A	B^2	x	y	z	A	B^2				
$UO_2(NO_3)_2 \cdot 6H_2O$	K_2PtCl_4	0.0400	0.0300	0.1400	1000.0	0.00	0.0385	0.0288	0.1404	802.96	0.00	6	0.411	0.111	0.605
	Site A												0.379	0.092	0.615
K_2PtCl_4	$UO_2(NO_3)_2 \cdot 6H_2O$	0.1250	0.0240	0.8500	3000.0	0.00	0.1228	0.0253	0.8489	3013.29	1.40	4	0.258	0.213	0.790
	Site A												0.253	0.200	0.883
K_2PtCl_4	$UO_2(NO_3)_2 \cdot 6H_2O$	0.1300	0.0000	0.0000	1300.0	0.00	0.1308	0.0069	-0.0010	962.94	0.00	8	0.198	0.201	0.762
	Site A_1												0.182	0.170	0.759

[1] ΔF map coefficients: $A = 0.01 \times FM \times \Delta F \times \cos\phi_{SIR}$; $B = 0.01 \times FM \times \Delta F \times \sin\phi_{SIR}$

[2] Isotropic temperature factors were not refined, but were adjusted from Wilson-type plots.

[3] $R_K = \Sigma$ (lack of closure)$/\Sigma|F_{PH}|$ (22)

[4] $R_C = \Sigma$ (lack of closure)$/\Sigma|\Delta F|$ (23)

the larger percentage change in ΔF for the uranyl derivative, as compared to the platinum derivative, and the higher noise level in the $UO_2(NO_3)_2 \cdot 6H_2O$ difference Patterson. Again, there was a single major peak in the difference map, which was consistent with the best Harker peaks in the difference Pattersons. Also, there was no strong evidence of any minor sites and there was no feedback at the coordinates used for the phasing.

As a control, the coordinates from sites A_1, A_1', and A_1'' in the $UO_2(NO_3)_2 \cdot 6H_2O$ difference Patterson, which also form a consistent set of Harker peaks, were subjected to eight cycles of SIR phase calculations and atomic parameter least-squares refinement. These sites formed a consistent set of Harker peaks, in terms of their locations on the Harker sections, and they had significant intensity. However, they were not thought to be real because one peak was coincidental with the previously proven site A, and the other two were at special positions where errors tend to accumulate. As seen in Table V, the heavy atom occupancies refined to less than 1/3 that of site A and the figure of merit was over 5% lower than site A. The K_2PtCl_4 difference Fourier calculated from these phases is given in Figure 6. As was expected, there were no sites significantly above the average noise level.

The two heavy atom derivatives, solved very convincingly by difference Pattersons and SIR-phased difference Fouriers, were then combined for MIR phase analysis. With the improved MIR phases, the atomic parameters of the known sites were further refined, additional sites were located, the $NaAu(CN)_2$ derivative evaluated, and the

protein electron density map computed.

The general procedure used for completing the MIR refinement involved an alternating process. First, computer cycles of MIR phase calculations and least-squares heavy atom parameter refinement were calculated to convergence. This was followed by calculations of difference Fourier heavy atom maps and double difference Fourier error maps. Adjustments were made from these maps, when necessary, and the process was begun again. The double difference Fourier, or error synthesis, uses the derivative lack of closure, ε, as the Fourier amplitudes and the phases calculated for the derivative structure factors (24,25). This synthesis is used because it results in a lower error level, compared to a ΔF synthesis, and allows easy evaluations of errors in heavy atom occupancies and temperature factors.

The occupancy A and the isotropic temperature factor B would not successfully refine together because of very strong correlations between the two parameters at this low resolution. Rather than refine these individually, in alternating cycles, the temperature factors were adjusted by Wilson-type plots of $\ln \{<\Delta F>/<f_H>\}$ versus s (10) and by computed double difference maps.

The MIR refinement was begun using the SIR coordinates and refined for six cycles. The resulting phases were used to calculate the new K_2PtCl_4 and $UO_2(NO_3)_2 \cdot 6H_2O$ difference Fourier maps in Figures 7 and 8. These maps confirmed the sites found in the earlier SIR phased maps, with the peak-to-noise level greatly improved. Again, the K_2PtCl_4 difference map was cleaner, but this time both maps were

calculated from the same figure of merit.

The new MIR phases were then used to calculate the $NaAu(CN)_2$ ΔF Fourier map in Figure 9. A difference Patterson map was not previously calculated for this derivative in view of the large amount of noise in the K_2PtCl_4 and $UO_2(NO_3)_2 \cdot 6H_2O$ difference Pattersons and the poorer quality of the $NaAu(CN)_2$ data due to the extremely small crystal used for data collection.

The ΔF Fourier map was indeed quite noisy compared to the K_2PtCl_4 and $UO_2(NO_3)_2 \cdot 6H_2O$ ΔF Fouriers. However, a single site (site A) was located which was significantly above the background.

Refinement was continued with fifteen more cycles of three derivative MIR phase refinement. The $\ln \{<\Delta F>/<f_H>\}$ vs. s Wilson plot for the $NaAu(CN)_2$ derivative indicated that the isotropic temperature factor shou'd be less than zero (-3.03). This was indicative of the fact that there was something seriously wrong with this data set. The difference map calculated for the $NaAu(CN)_2$ derivative also indicated that there was a problem. The $NaAu(CN)_2$ peak was surrounded by a spherical diffraction ripple, indicating that the temperature factor was too low.

The $NaAu(CN)_2$ temperature factor was set arbitrarily at 2.5, a value judged as not too unreasonable from the K_2PtCl_4 and $UO_2(NO_3)_2 \cdot 6H_2O$ derivatives. Thirteen more cycles of refinement were done and new difference and double difference maps calculated. The new maps were improved slightly over the last set. However, the noise

level was still high, and the $(F_{PH} - F_P)$ and f_H were not scaled properly. The ratios of $\langle\Delta F\rangle/\langle f_H\rangle$ increased with increasing resolution.

At this point, the $NaAu(CN)_2$ derivative data were reexamined, the data between $2\theta = 20.0°$ and $22.5°$ were eliminated, and the remainder of the data to $2\theta = 20.0°$ were rescaled. The temperature factor was adjusted from a Wilson-type plot to 0.08.

The K_2PtCl_4 and $UO_2(NO_3)_2 \cdot 6H_2O$ derivative data between $2\theta = 22.5°$ and $29.0°$ were added to the lower resolution data and refinement was initiated again for nine cycles. The elimination of the shell of $NaAu(CN)_2$ data between $2\theta = 20.0°$ to $22.5°$ and rescaling of the remainder of the data resulted in improvements in the refinement statistics (i.e., the root mean square lack of closure and the R-factors). The refinement converged to a mean change in ϕ_{best} of $0.42°$, a mean relative error of 1.1745, and the figure of merit and R-factors given in Table VI. A display of the native structure factors and the refinement statistics as a function of resolution are given in Figure 10.

Reexamination of the Difference Pattersons

The SIR and MIR phased difference Fouriers in Figures 4 through 9 and subsequently calculated difference Fouriers and double difference Fouriers had very low noise levels compared to the difference Pattersons in Figures 2 and 3. The noise in the Pattersons could have been

TABLE VI: Multiple Isomorphous Replacement Phase Refinement

Refinement Cycle	Derivative	x	y	z	A	B[1]	ERMS	Centric Data R_K^2	Centric Data R_C^2	Centric Data FM	All Data R_K^2	All Data R_C^2	All Data FM
0	Pt	0.0385	0.0288	0.1404	273.0	3.18	127.4	0.129	0.579	0.432	0.104	0.604	0.444
	UO_2	0.1228	0.0253	0.8489	301.3	1.40	264.6	0.256	0.847		0.214	0.868	
6	Pt	0.0374	0.0300	0.1382	225.25	3.18	128.7	0.132	0.592	0.474	0.104	0.609	0.472
	UO_2	0.1216	0.0151	0.8478	188.94	1.40	201.6	0.208	0.688		0.164	0.666	
	K_2PtCl_4 ΔF and ΔΔF maps computed. $UO_2(NO_3)_2 \cdot 6H_2O$ ΔF and ΔΔF maps computed. $NaAu(CN)_2$ ΔF map computed and one major site located at x = 0.1790, y = 0.1900, z = 0.9400.												
21[3]	Pt	0.0376	0.0307	0.1409	1789.0	2.87	1073.8	0.143	0.634	0.584	0.110	0.641	0.621
	UO_2	0.1263	0.0154	0.8589	2252.9	2.52	1555.1	0.202	0.667		0.161	0.653	
	Au	0.1799	-.1915	0.9422	797.2	-3.03	2579.1	0.285	0.752		0.257	0.661	
	K_2PtCl_4 ΔF and ΔΔF maps computed. $UO_2(NO_3)_2 \cdot 6H_2O$ ΔF and ΔΔF maps computed. $NaAu(CN)_2$ ΔF and ΔΔF maps computed.												
34[3]	Pt	0.0381	0.0308	0.1420	1804.4	2.87	1048.2	0.141	0.625	0.559	0.108	0.627	0.584
	UO_2	0.1271	0.0153	0.8595	2265.8	2.52	1513.5	0.199	0.655		0.155	0.631	
	Au	0.1824	0.1906	0.9423	3313.9	2.50	2940.2	0.316	0.836		0.290	0.746	
	K_2PtCl_4 ΔF and ΔΔF maps computed. $UO_2(NO_3)_2 \cdot 6H_2O$ ΔF and ΔΔF maps computed and one minor site located at x = 0.2476, y = 0.4451, z = 0.7000. $NaAu(CN)_2$ ΔF and ΔΔF maps computed.												
	K_2PtCl_4 2θ = 22.5° to 29.0° added. $UO_2(NO_3)_2 \cdot 6H_2O$ 2θ = 22.5° to 29.0° added. $NaAu(CN)_2$ data terminated at 2θ = 20.0° and rescaled to native.												
43	Pt	0.0402	0.0304	0.1385	1706.0	2.87	1083.7	0.150	0.692	0.515	0.122	0.725	0.518
	U	0.1282	0.0153	0.8572	2213.6	2.52	1229.3	0.169	0.564		0.138	0.567	
	U	0.2509	0.4434	0.6985	1209.3	2.52							
	Au	0.1782	0.1888	0.9417	1699.7	0.08	1634.2	0.187	0.544		0.172	0.525	

[1] A and B did not refine satisfactorily together and B, therefore, was not refined by least squares. B was adjusted by Wilson-type plots and double difference maps.

[2] R_K = Σ (lack of closure)/Σ(F_{PH}) (22). R_C = Σ (lack of closure)/Σ|ΔF| (23).

[3] All structure factors were multiplied by 8.0.

a result of inherent errors in the method, lack of sufficient resolution, random or systematic errors in the data, or possibly a result of the native and derivative data being collected under different salt conditions. The first two explanations, inherent errors in the difference Patterson method and lack of resolution, are extremely unlikely since this method has frequently been used with success at this resolution, without the large noise levels found here. Also, a difference Patterson calculated from the subsequent independent high resolution analysis, with a resolution maximum of 4 Å, resulted in much lower noise levels.

In order to determine if the problem was an effect of different salt conditions, a difference Patterson was calculated for the K_2PtCl_4 and $UO_2(NO_3)_2 \cdot 6H_2O$ derivatives using only reflections with 2θ between 15° and 22.5°. The low order reflections, between 2θ = 2° and 15°, are affected more strongly by salt conditions and were eliminated to see if this would improve the difference Patterson. After rescaling the data, Harker sections were calculated and plotted in Figures 11 and 12. Some of the noise peaks were reduced significantly relative to the heavy atom vector sites. However, the new maps also contained some noise peaks with increased intensity. In general, the modified Harker sections are no better than the original ones.

Harker sections were also calculated for the low resolution K_2PtCl_4 derivative data minus the native data that were collected in the subsequent high resolution analysis. These native data were collected in 75% $(NH_4)_2SO_4$ and 1.0 M NaCl, as were the derivative data.

The K_2PtCl_4 data were scaled to the new native data and then they were both put back on a scale approximately equal to the other difference Pattersons. The computed Harker sections are displayed in Figure 13. These sections are very noise free and are a significant improvement over the other original difference Pattersons.

These results show conclusively that the high noise level in the difference Pattersons was not a result of random or systematic errors in the derivative data. It is also unlikely that the problem was a result of errors in the native data, since phase refinement, heavy atom ΔF and $\Delta\Delta F$ maps, and the protein Fourier maps were calculated and interpreted without difficulties. The most likely source of the high noise level in the difference Pattersons is the different salt conditions. When data were collected under the same salt conditions, the problem was eliminated (Figure 13).

In summary, it was seen that the difference Fourier method proved to be a powerful tool in the solution of the heavy atom derivatives, even when difference Patterson methods had major problems. Difference Fouriers, even with the relative inaccurate SIR phases, are much less susceptible to the noise problems seen in the difference Pattersons. Once a major site for one derivative is located, difference Fouriers with SIR phases can be used with confidence to solve other derivatives.

Low Resolution Electron Density Map

The MIR phases calculated using the refined heavy atom parameters in Table VI were used to calculate the electron density map. This map was contoured onto plexiglass. The strongest feature in the map was the heme. In addition to the heme, the N- and C-terminal alpha helices were quite obvious. The rods of main chain density in the extended chain structure between residues 50 and 64 were particularly dense. The main chain densities were continuous and well above the noise level with only one minor exception. Careful examination of this density and considerations of the various distances from known reference points allowed the chain path to be followed with confidence. A few side chains were located, including the Trp 56 residue that was found to be hydrogen bonded to the buried propionic acid.

A stereo view of the main chain density is shown in Figure 14. This front orientation was chosen to match the front view of other cytochromes. The observed tilt of the heme in Figure 14 relative to other c-type cytochromes was verified at high resolution.

Additional features with respect to the low resolution structure are presented in Appendix I. A more detailed discussion of c_{551} will be postponed until after the refined high resolution structure determination has been presented in Chapter 3.

REFERENCES

1. Harbury, H.A., Personal communication.
2. Ambler, R.P. (1963) Biochem. J. 89, 341-349.
3. Ambler, R.P. and Brown, L.H. (1967) Biochem. J. 104, 784-825.
4. Ambler, R.P. and Wynn, M. (1973) Biochem. J. 131, 485-498.
5. Horio, T., Higashi, T., Nakai, M., Kusai, K. and Okunuki, K. (1958) Nature 182, 1307-1308.
6. Horio, T., Hagashi, T., Sasagawa, M., Kusai, K., Nakai, M., and Okunuki, K. (1960) Biochem. J. 77, 194-201.
7. Zeppezauer, M., Eklund, H. and Zeppezauer, E. (1968) Arch. Bioch. 126, 564-573.
8. Hempel, A. and Bock, R. (1970) Biochemistry 9, 1873-1880.
9. Zeppezauer, M. (1971) in "Methods in Enzymology" (W.B. Jakoby, ed.), Vol. XXII. Academic Press, New York, pp. 253-266.
10. Dickerson, R.E., Kopka, M.L., Borders, C.L., Varnum, J., Weinzierl, J.E. and Margoliash, E. (1967) J. Mol. Biol. 29, 77-95.
11. Swanson, R., Trus, B.L., Mandel, N., Mandel, G., Kallai, O.B. and Dickerson, R.E. (1977) J. Biol. Chem. 252, 759-775.
12. Takano, T., Kallai, O.B., Swanson, R. and Dickerson, R.E. (1973) J. Biol. Chem. 248, 5234-5255.
13. Timkovich, R. and Dickerson, R.E. (1972) J. Mol. Biol. 72, 199-202.
14. Baumber, M.E., Moews, P.C. and Bunn, C.W., personal communication.

15. Krieger, M., Chambers, J.L., Christoph, G.G., Stroud, R.M. and Trus, B.L. (1974) Acta Crystallogr. A30, 740-748.

16. Stout, G.H. and Jensen, L.H. "X-Ray Structure Determination", Macmillan Co., London, 1968.

17. North, A.C.T., Phillips, D.C. and Mathews, F.S. (1968) Acta Crystallogr. A24, 351-359.

18. Azaroff, L.V. (1955) Acta Crystallogr. 8, 701-704.

19. Levy, H.A. and Ellison, R.D. (1960) Acta Crystallogr. 13, 270-271.

20. Dickerson, R.E. (1959) Acta Crystallogr. 12, 610-611.

21. Blundell, T.L. and Johnson, L.N. "Protein Crystallography", Academic Press, New York, 1976.

22. Kraut, J., Sieker, L.C., High, D.F. and Freer, S.T. (1962) Proc. Nat. Acad. Sci. USA 48, 1417-1424.

23. Cullis, A.F., Muirhead, H., Perutz, M.F., Rossmann, M.G. and North, A.C.T. (1961) Proc. Roy. Soc. Lond. A265, 15-38.

24. Black, C.C.F., Fenn, R.H., North, A.C.T., Phillips, D.C. and Poljak, R.J. (1963) Acta Crystallogr. 16, A77.

25. Stroud, R.M., Kay, L.M. and Dickerson, R.E. (1974) J. Mol. Biol. 83, 185-208.

FIGURE LEGENDS

Figure 1.

(a) (0,k,ℓ) precession photograph of oxidized P. aeruginosa cytochrome c_{551}. The crystal size was approximately 0.6 x 0.075 x 0.05 mm^3. A 20° precession angle was used with an exposure time of 65 hours.

(b) (h,k,0) precession photograph as described in (a), except from a crystal with dimensions 0.25 x 0.05 x 0.10 mm^3, a precession angle of 17° and an exposure time of 94.4 hours.

(c) (0,k,ℓ) precession photograph of c_{551} after crystal reduction with $Na_2S_2O_4$. A crystal size of 0.6 x 0.075 x 0.05 mm^3 was used for this 17° precession photograph, exposed for 81 hours.

Figure 2. K_2PtCl_4 $(\Delta F)^2$ Patterson sections at 4 Å resolution.

(a) (½,v,w) Harker section.

(b) (u,½,w) Harker section.

(c) (u,v,½) Harker section.

Figure 3. $UO_2(NO_3)_2 \cdot 6H_2O$ $(\Delta F)^2$ Patterson sections at 4 Å resolution.

(a) (½,v,w) Harker section.

(b) (u,½,w) Harker section.

(c) (u,v,½) Harker section.

Figure 4. $UO_2(NO_3)_2 \cdot 6H_2O$ projected ΔF Fourier with $PtCl_4$ SIR phases from site A of Figure 2. The Fourier coefficients are $A = (\Delta F_{UO_2} \times FM \times \cos\phi_{best})$ and $B = (\Delta F_{UO_2} \times FM \times \sin\phi_{best})$. The projection is down the z axis with the section of highest density being contoured for each peak. Contours begin at 400 and increment at 400. The z coordinates and peak heights are shown.

Figure 5. K_2PtCl_4 projected ΔF Fourier with $UO_2(NO_3)_2 \cdot 6H_2O$ SIR phases from site A of Figure 3. The Fourier coefficients are $A = (\Delta F_{Pt} \times FM \times \cos\phi_{best})$ and $B = (\Delta F_{Pt} \times FM \times \sin\phi_{best})$. The projection is down the z axis with the section of highest density being contoured for each peak. Contours begin at 200 and increment at 200. The z coordinates and peak heights are shown.

Figure 6. K_2PtCl_4 projected ΔF Fourier with $UO_2(NO_3)_2 \cdot 6H_2O$ SIR phases from site A of Figure 3. The Fourier coefficients are $A = (\Delta F_{Pt} \times FM \times \cos\phi_{best})$ and $B = (\Delta F_{Pt} \times FM \times \sin\phi_{best})$. The projection is down the z axis with the section of highest density being contoured for each peak. Contours begin at 200 and increment at 200. The z coordinates and peak heights are given.

Figure 7. K_2PtCl_4 projected ΔF Fourier with K_2PtCl_4 and $UO_2(NO_3)_2 \cdot 6H_2O$ MIR phases after six cycles of refinement of sites A from Figures 2 and 3. The projection is down the z axis with the section of highest density being contoured for each peak. Contours

begin and increment at 400. The z coordinates and peak heights are given.

Figure 8. $UO_2(NO_3)_2 \cdot 6H_2O$ projected ΔF Fourier with K_2PtCl_4 and $UO_2(NO_3)_2 \cdot 6H_2O$ MIR phases after six cycles of refinement of sites A from Figures 2 and 3. The projection is down the z axis with the section of highest density being contoured for each peak. Contours begin and increment at 400. The z coordinates and peak heights are given.

Figure 9. $NaAu(CN)_2$ projected ΔF Fourier with K_2PtCl_4 and $UO_2(NO_3)_2 \cdot 6H_2O$ MIR phases after six cycles of refinement for sites A from Figures 2 and 3. The projection is down the z axis with the section of highest density being contoured for each peak. The contours begin and increment at 800. The z coordinates and peak heights are given.

Figure 10. Summary of MIR refinement statistics for cytochrome c_{551}.

Figure 11. K_2PtCl_4 $(\Delta F)^2$ Patterson sections with a resolution range from 5.9 Å to 4.0 Å.

(a) $(\frac{1}{2},v,w)$ Harker section.

(b) $(u,\frac{1}{2},w)$ Harker section.

(c) $(u,v,\frac{1}{2})$ Harker section.

Figure 12. $UO_2(NO_3)_2 \cdot 6H_2O$ $(\Delta F)^2$ Patterson sections with a resolution range from 5.9 Å to 4.0 Å.

(a) $(\frac{1}{2},v,w)$ Harker section.

(b) $(u,\frac{1}{2},w)$ Harker section.

(c) $(u,v,\frac{1}{2})$ Harker section.

Figure 13. K_2PtCl_4 $(\Delta F)^2$ Patterson sections, calculated with 4 Å resolution native data obtained from the subsequent high resolution structure analysis.

(a) $(\frac{1}{2},v,w)$ Harker section.

(b) $(u,\frac{1}{2},w)$ Harker section.

(c) $(u,v,\frac{1}{2})$ Harker section.

Figure 14. Stereo view of the main chain fold of cytochrome c_{551} at low resolution.

Figure 1a

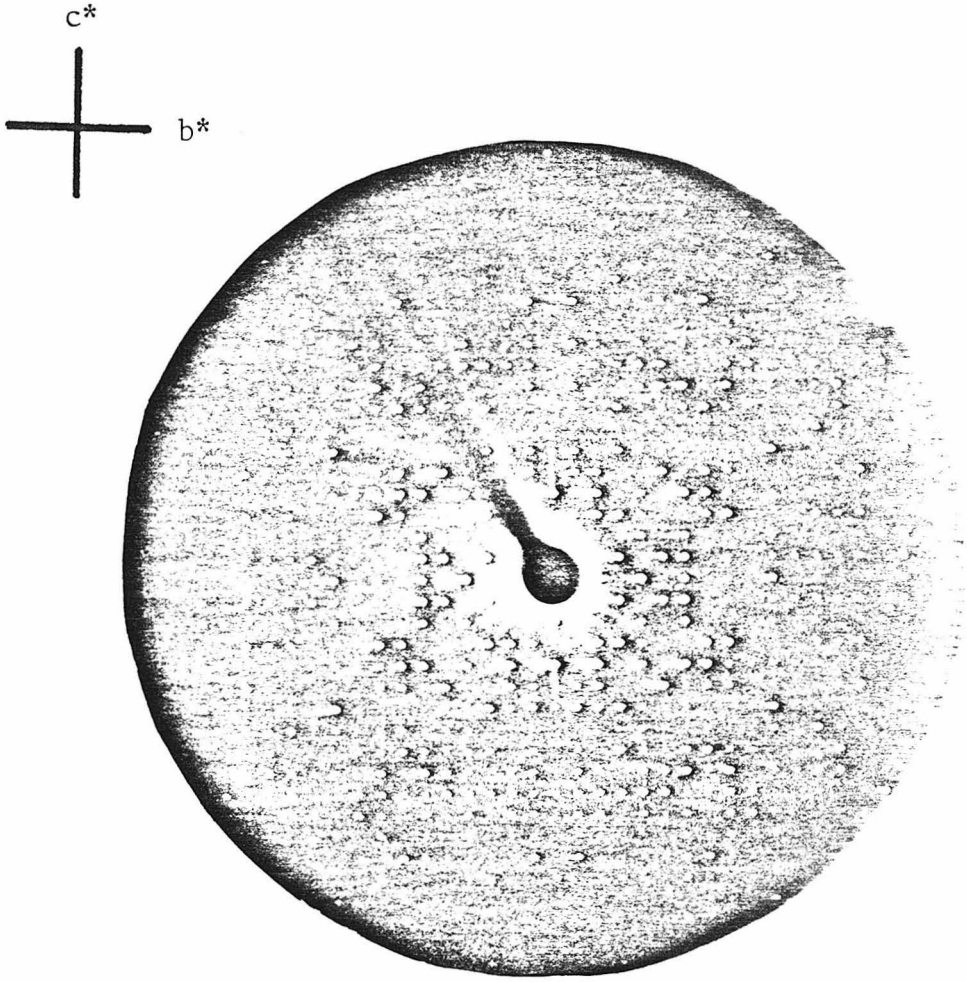


Figure 1b

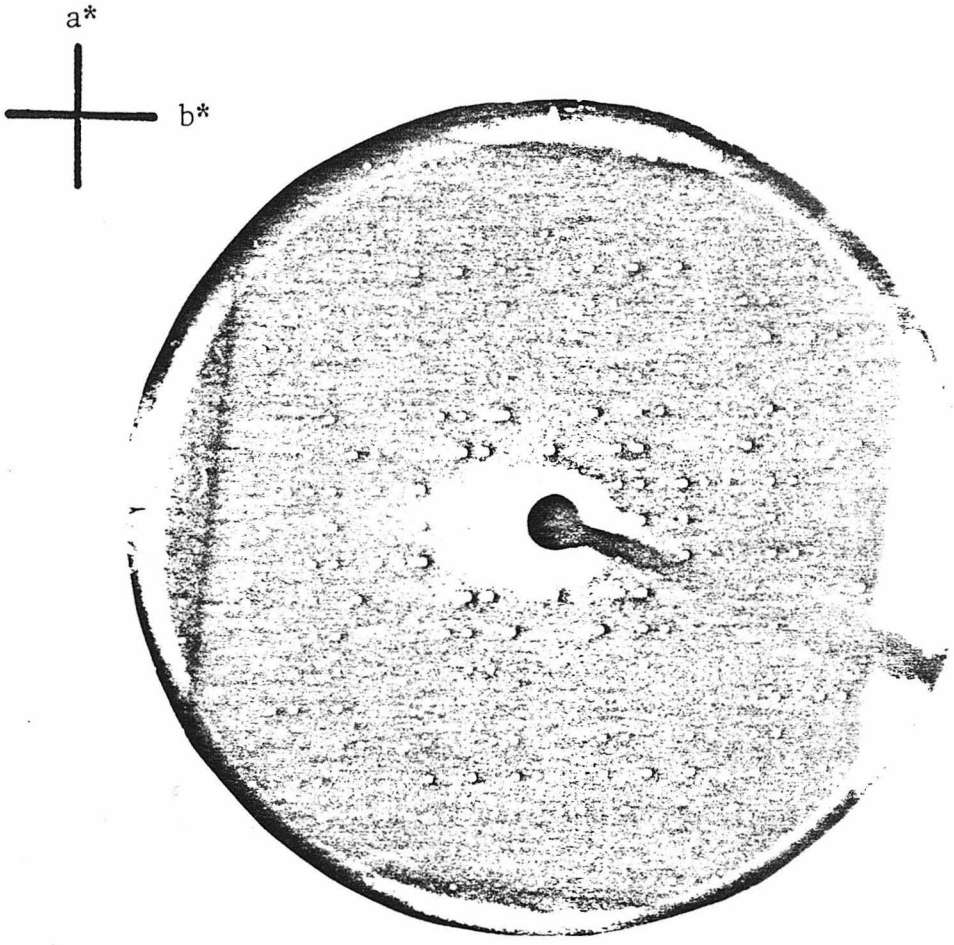


Figure 1c

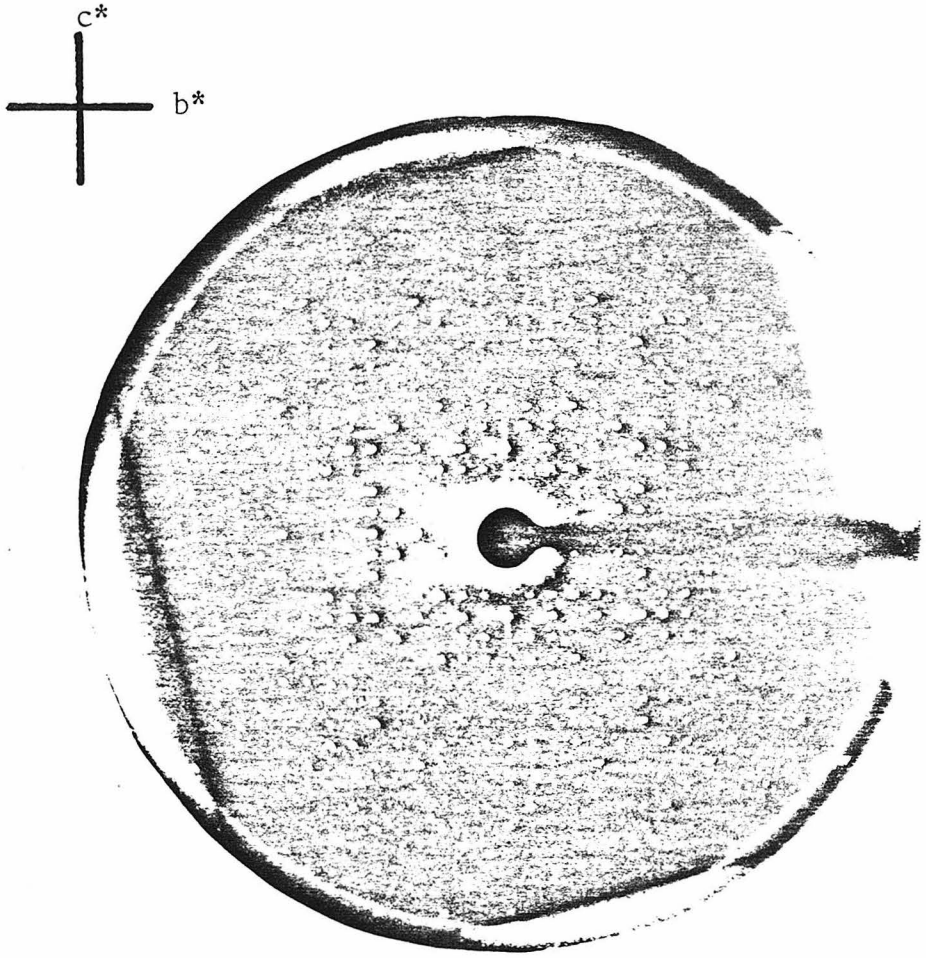


Figure 2a

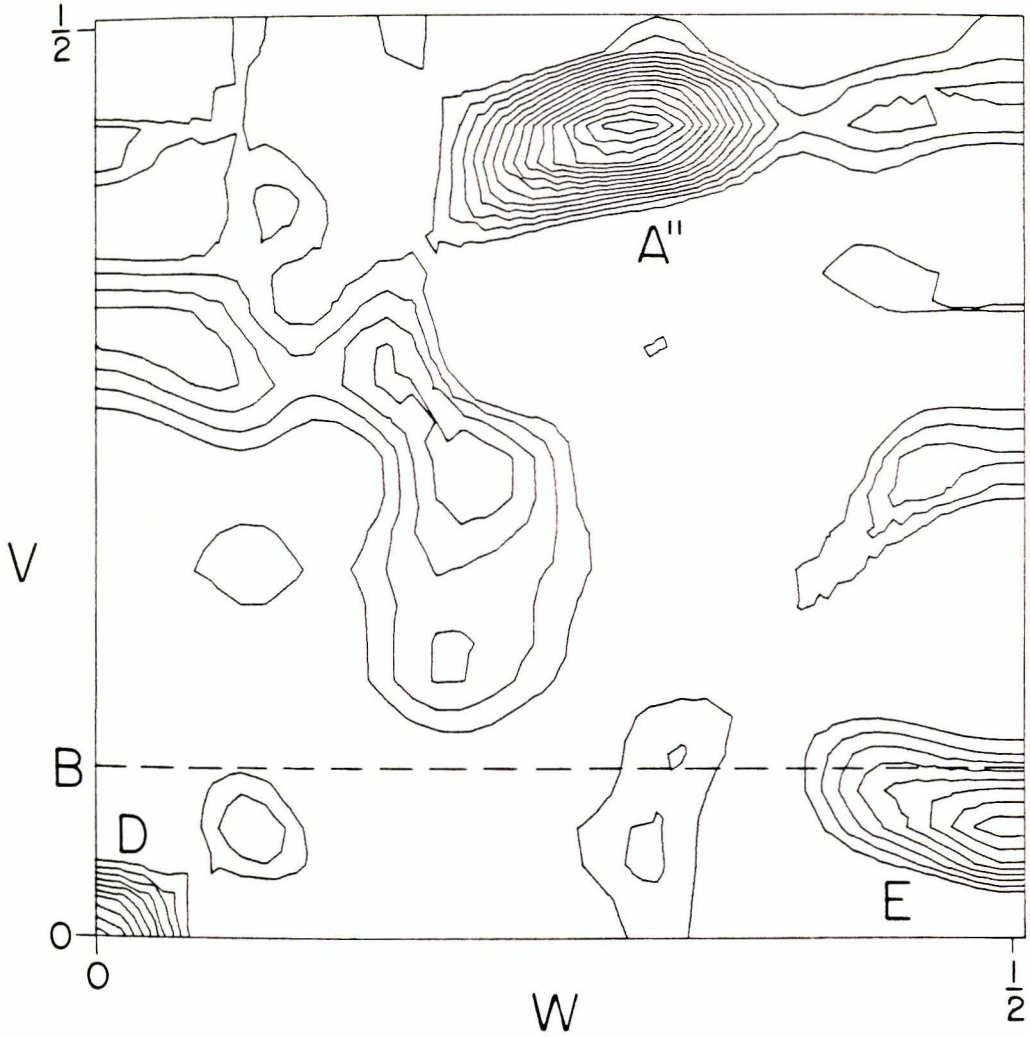


Figure 2b

Figure 2c

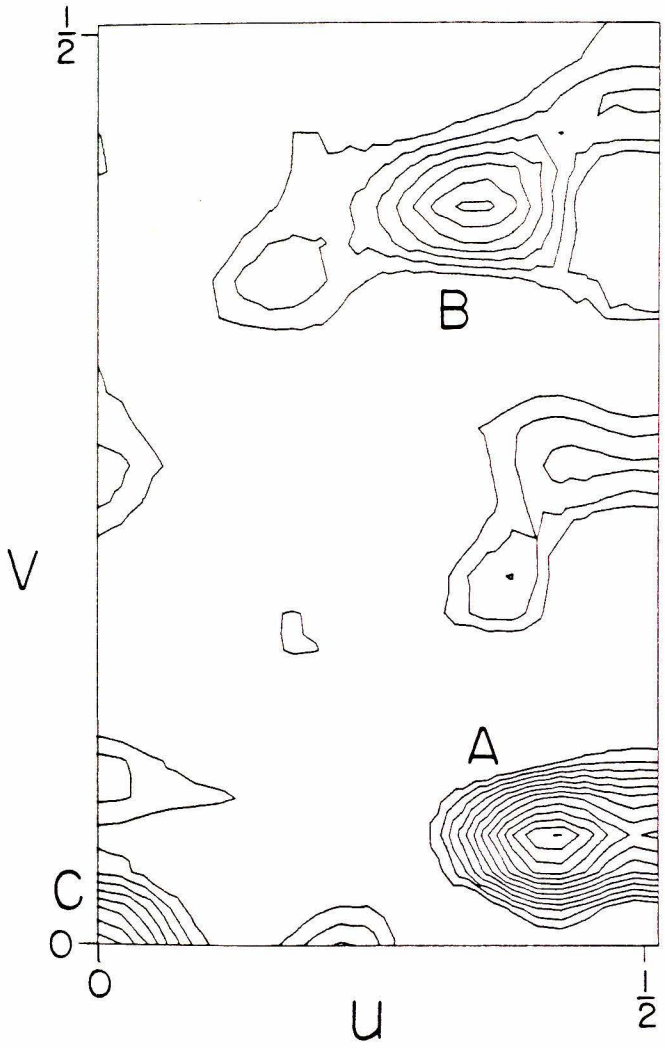


Figure 3a

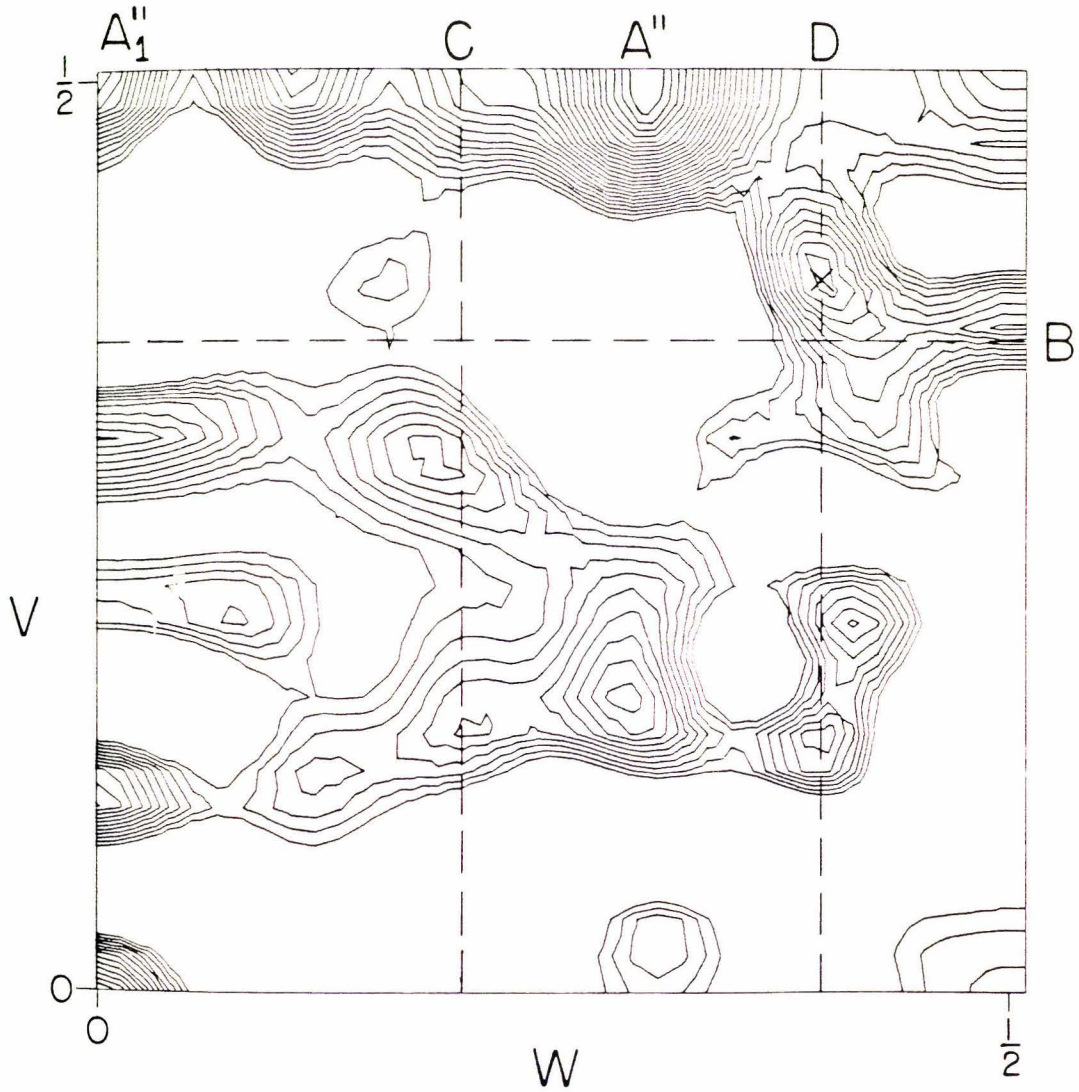


Figure 3b

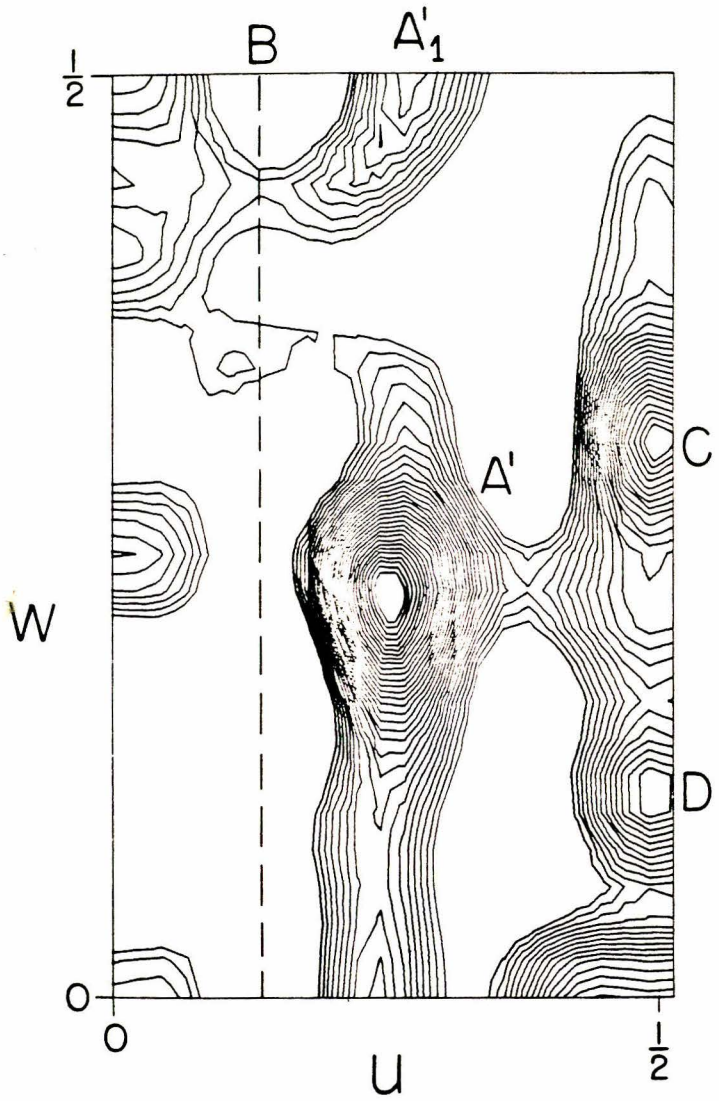


Figure 3c

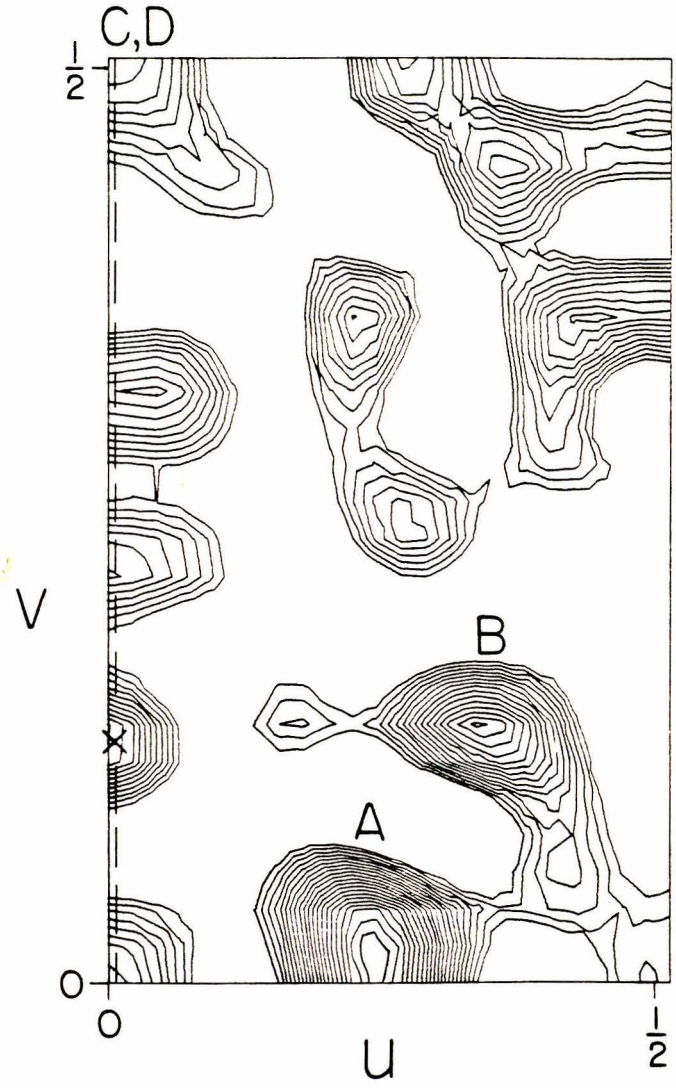


Figure 4

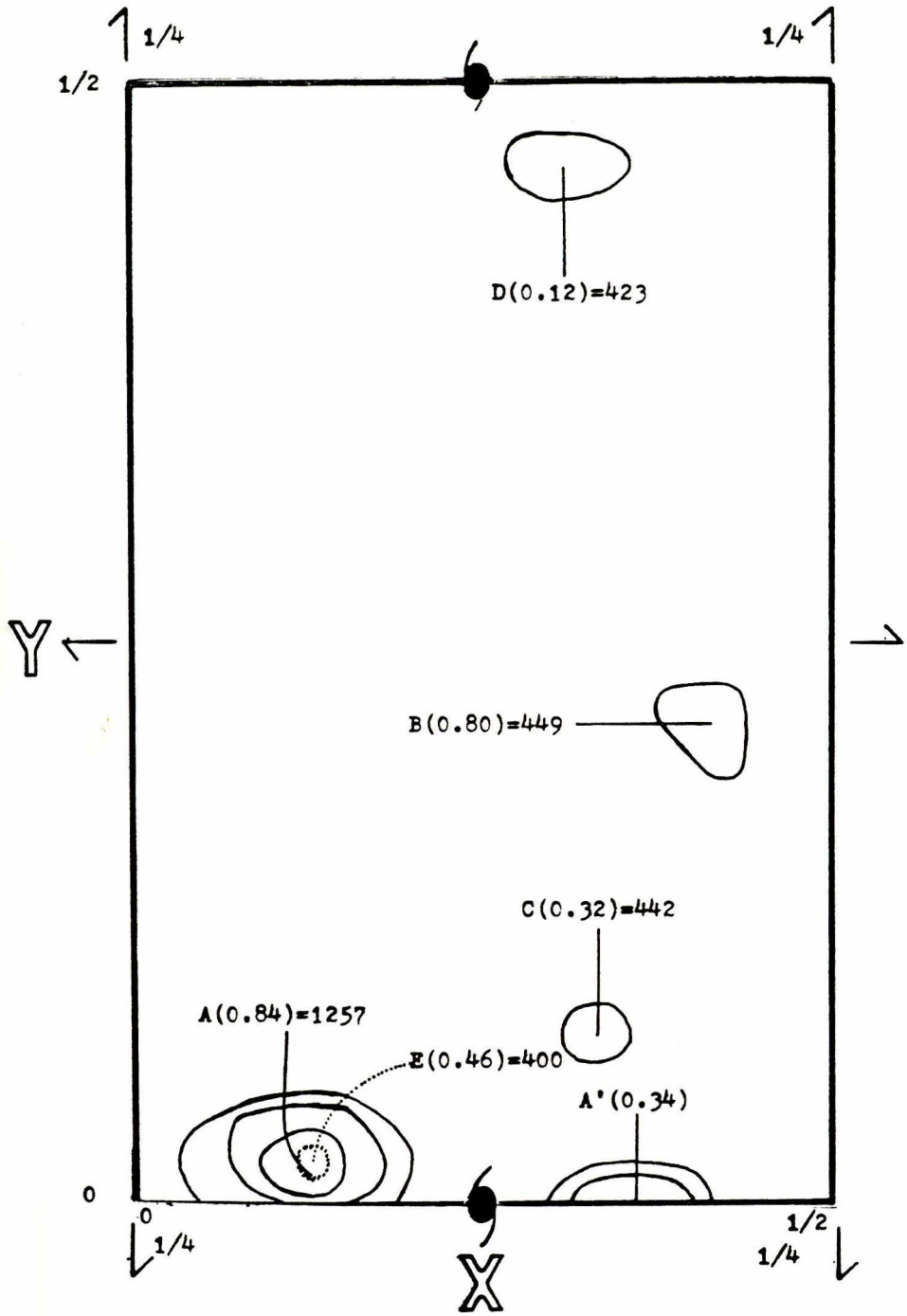

1/4
1/4
1/2
D(0.12)=423
Y
B(0.80)=449
C(0.32)=442
A(0.84)=1257
E(0.46)=400
A'(0.34)
0
0
1/4
1/2
1/4
X

Figure 5

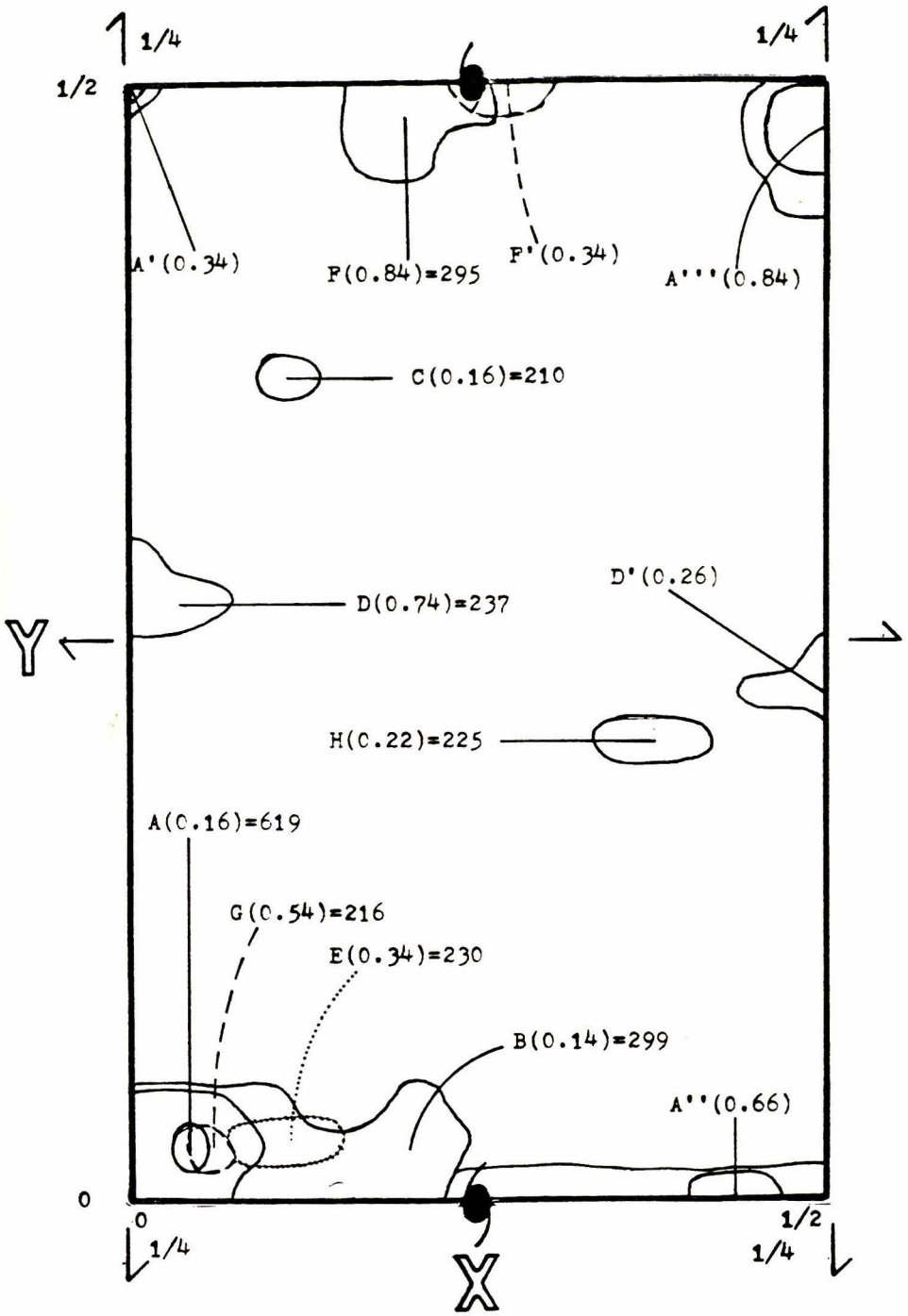

1/4
1/4
1/2
A'(0.34)
F(0.84)=295
F'(0.34)
A'''(0.84)
C(0.16)=210
D(0.74)=237
D'(0.26)
Y
H(0.22)=225
A(0.16)=619
G(0.54)=216
E(0.34)=230
B(0.14)=299
A''(0.66)
0
0
1/2
1/4
1/4
X

Figure 6

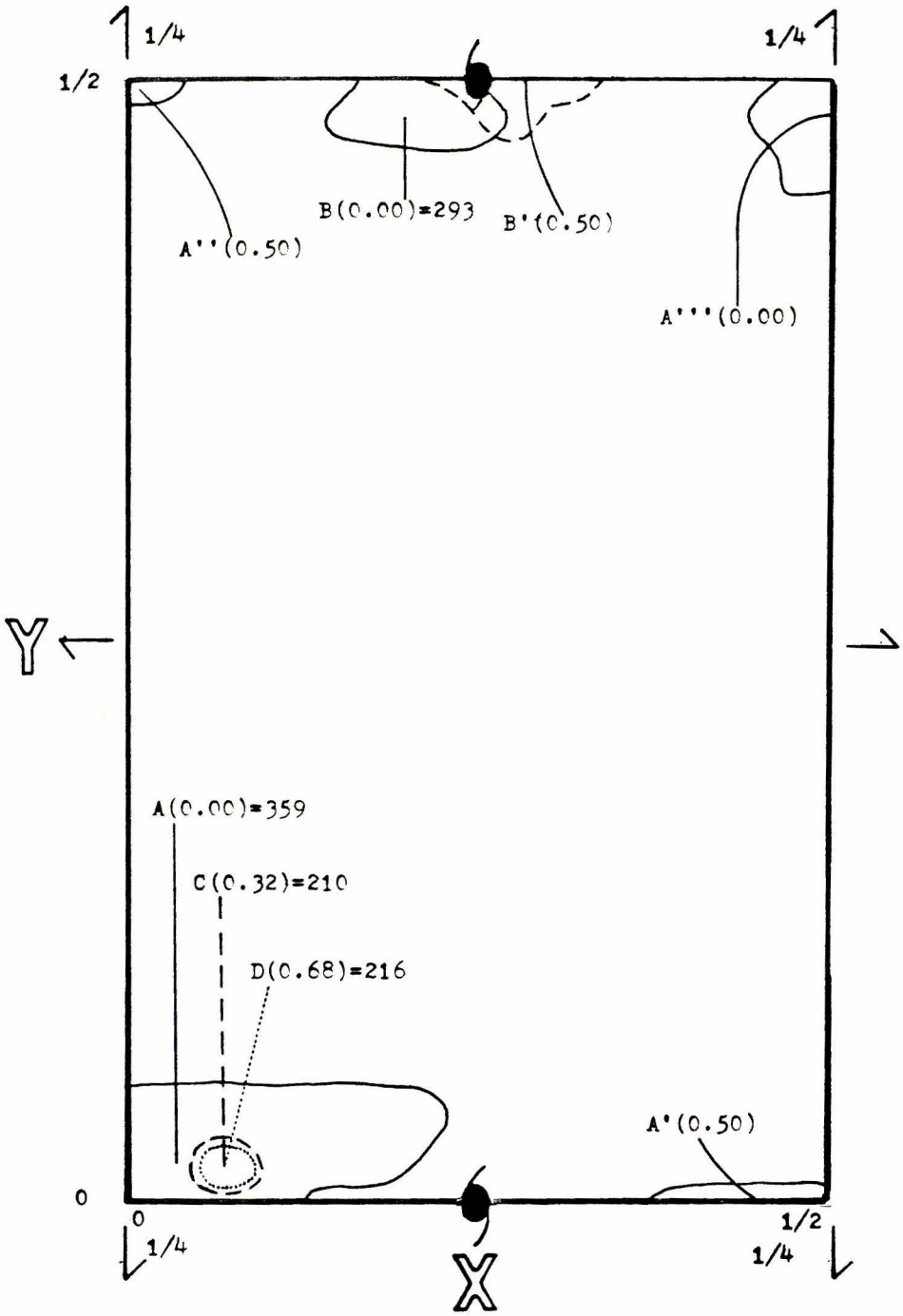

1/4
1/4
1/2
B(0.00)=293
B'(0.50)
A''(0.50)
A'''(0.00)
Y
A(0.00)=359
C(0.32)=210
D(0.68)=216
A'(0.50)
0
0
1/2
1/4
1/4
X

Figure 7

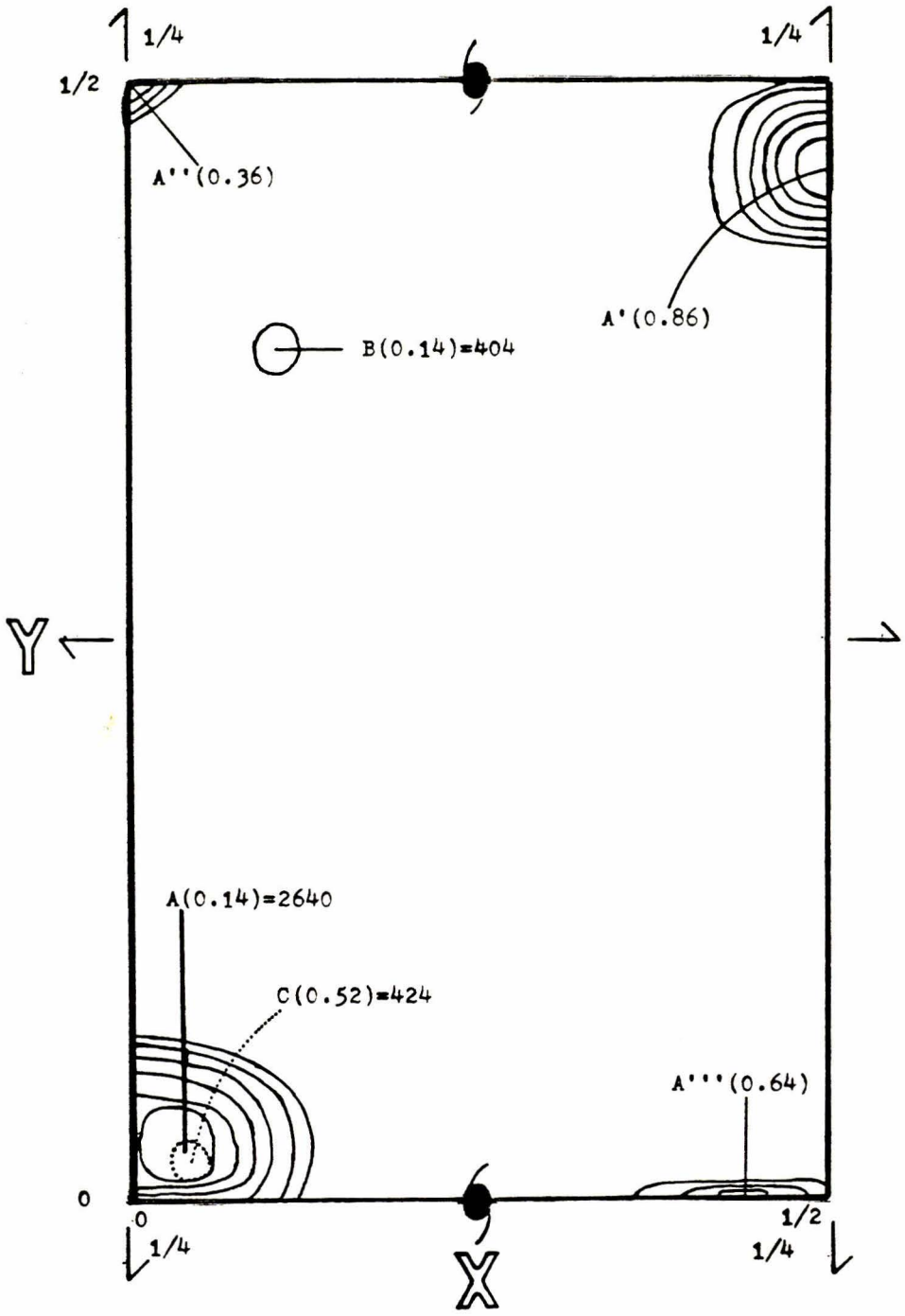

1/4
1/4
1/2
A''(0.36)
A'(0.86)
B(0.14)=404
Y
A(0.14)=2640
C(0.52)=424
A'''(0.64)
0
0
1/2
1/4
1/4
X

Figure 8

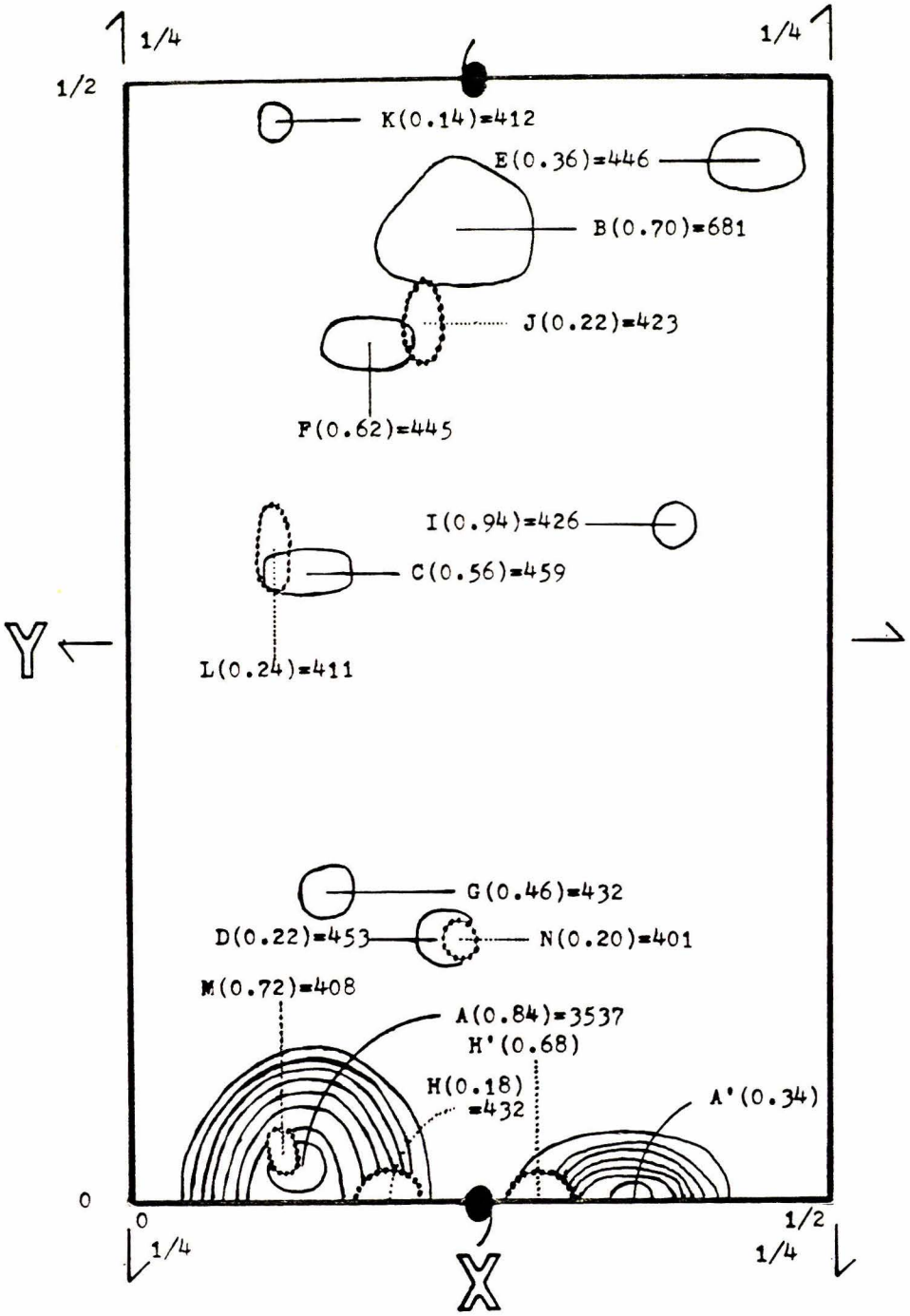

1/4
1/4
1/2
K(0.14)=412
E(0.36)=446
B(0.70)=681
J(0.22)=423
P(0.62)=445
I(0.94)=426
C(0.56)=459
L(0.24)=411
Y
G(0.46)=432
D(0.22)=453
N(0.20)=401
M(0.72)=408
A(0.84)=3537
H'(0.68)
H(0.18)
=432
A'(0.34)
0
0
1/2
1/4
1/4
X

Figure 9

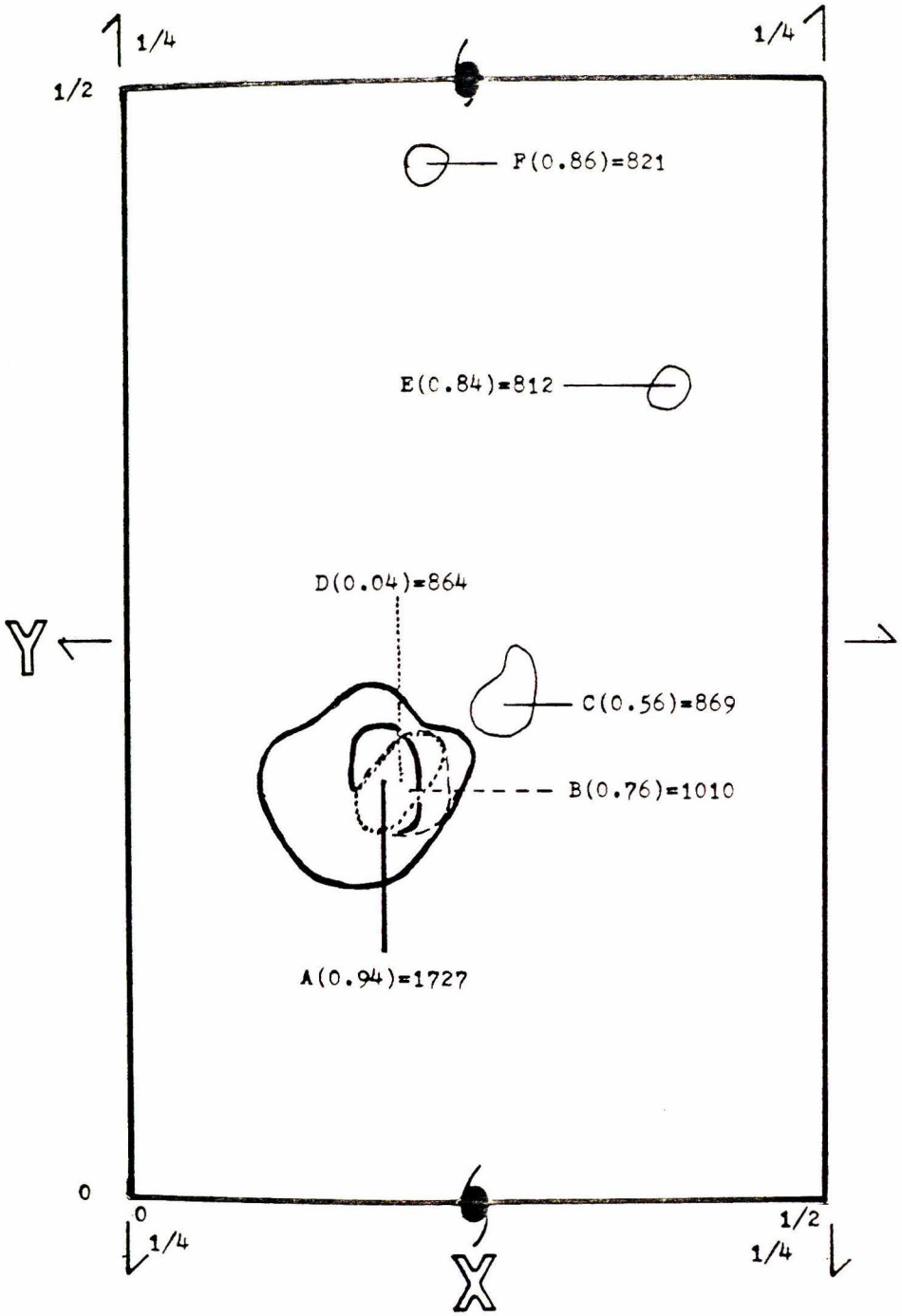

1/4
1/4
1/2
F(0.86)=821
E(0.84)=812
D(0.04)=864
Y
C(0.56)=869
B(0.76)=1010
A(0.94)=1727
0
0
1/4
1/2
1/4
X

Figure 10

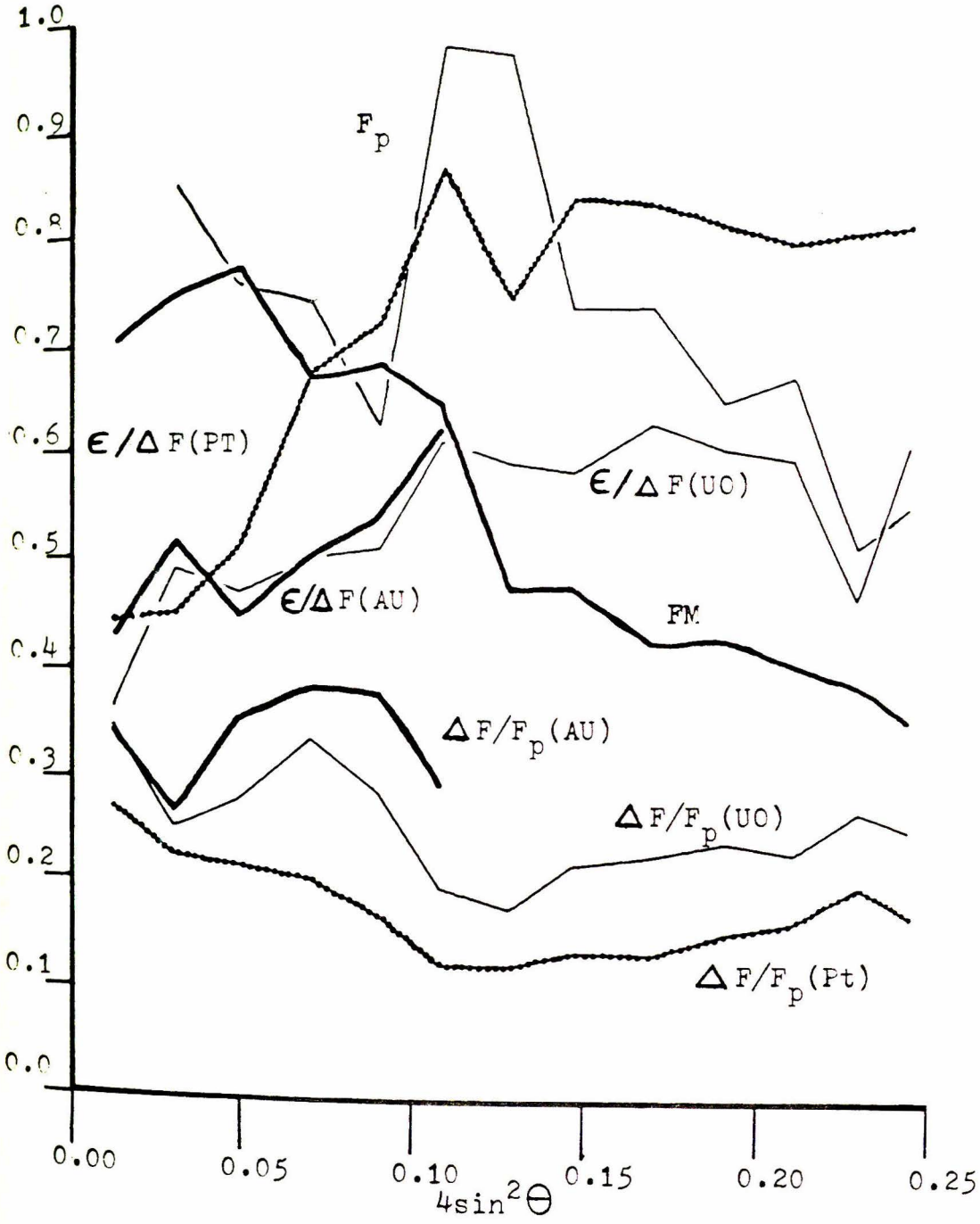

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
F_p
$\epsilon/\Delta F(PT)$
$\epsilon/\Delta F(UO)$
$\epsilon/\Delta F(AU)$
FM
$\Delta F/F_p(AU)$
$\Delta F/F_p(UO)$
$\Delta F/F_p(Pt)$
0.00
0.05
0.10
0.15
0.20
0.25
$4\sin^2\Theta$

Figure 11a

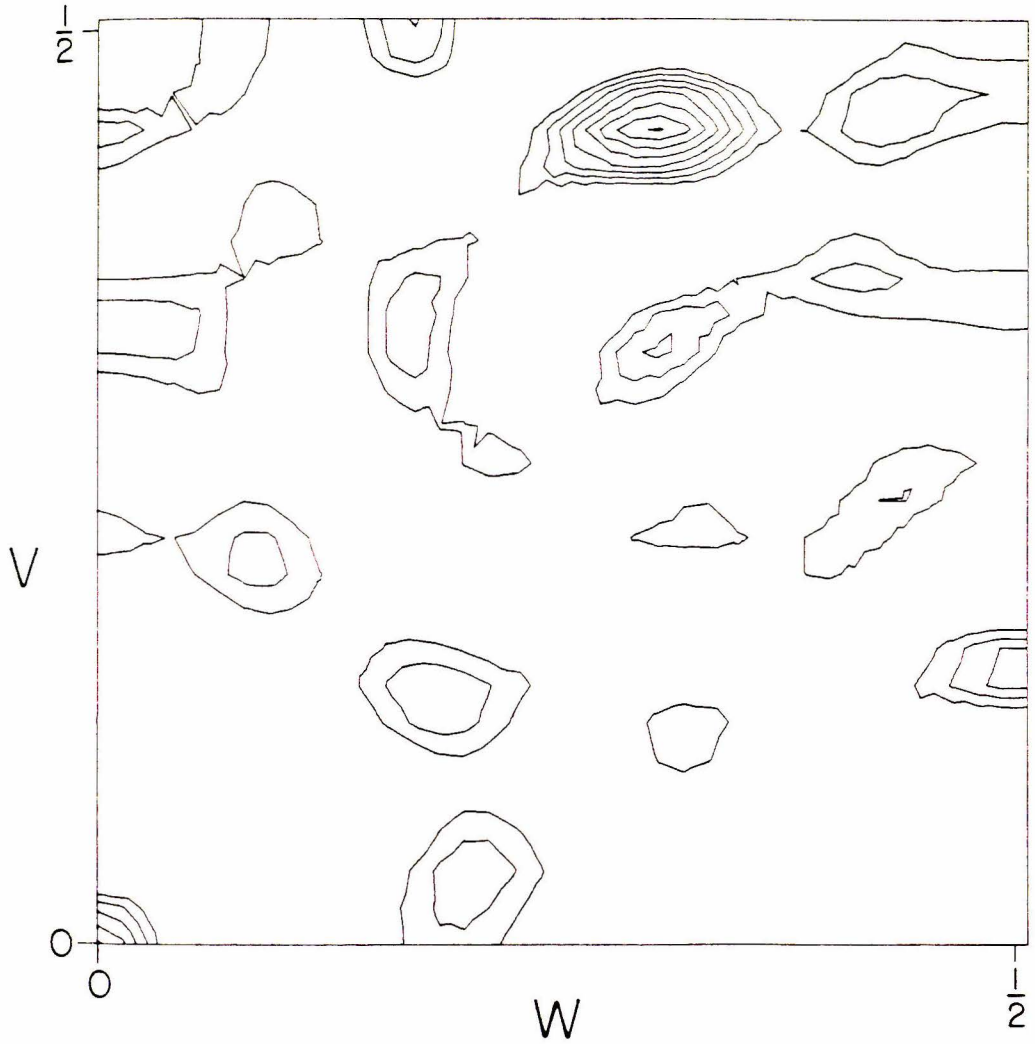


Figure 11b

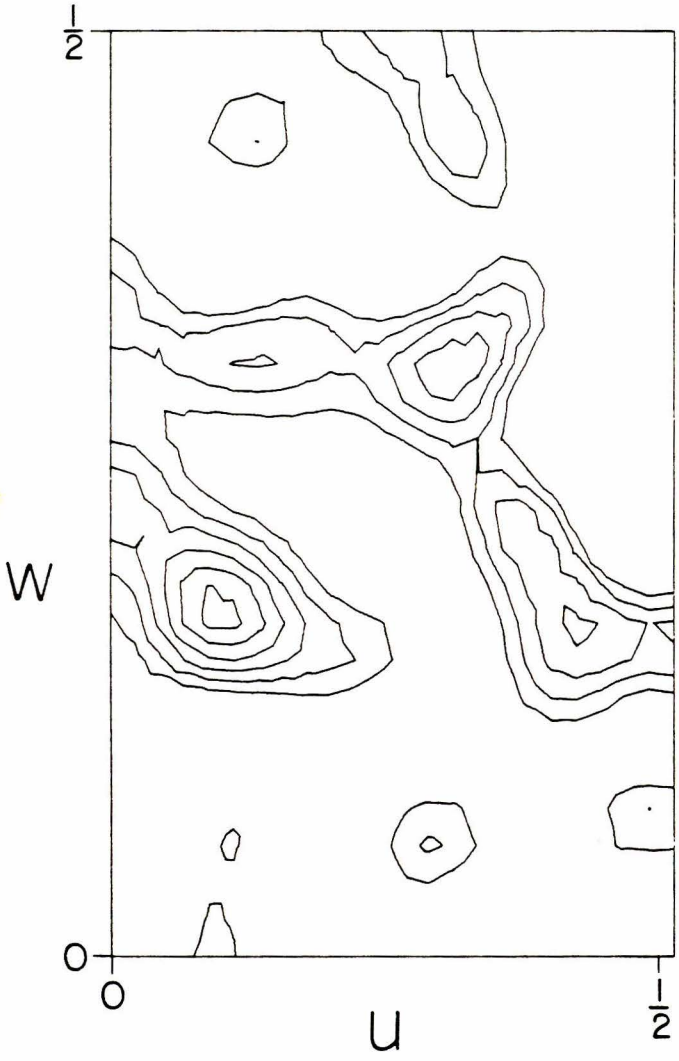


Figure 11c

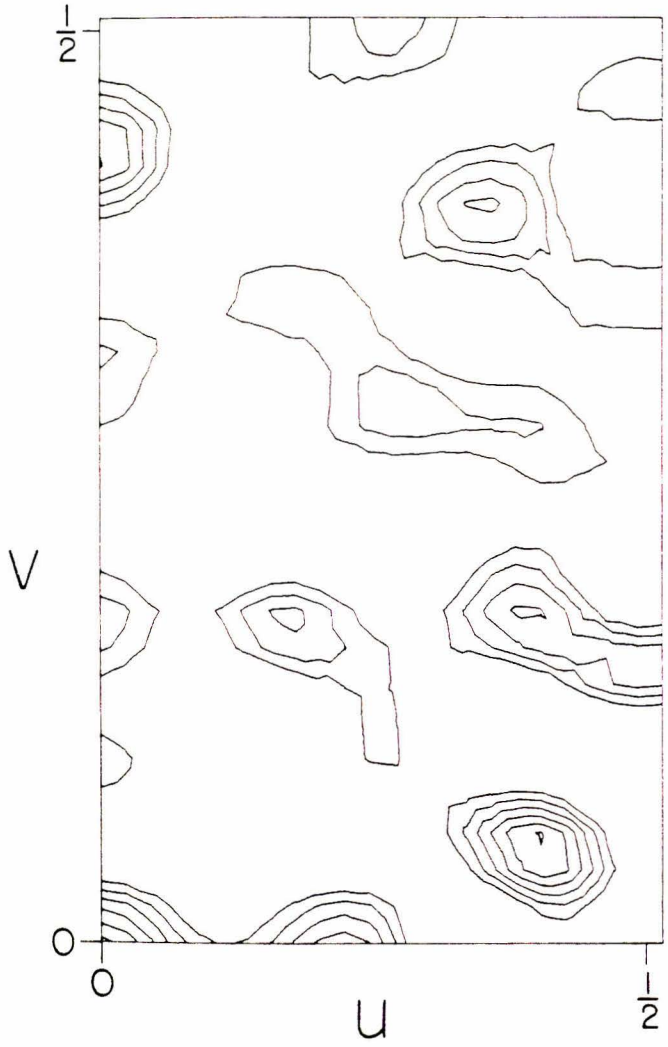


Figure 12a

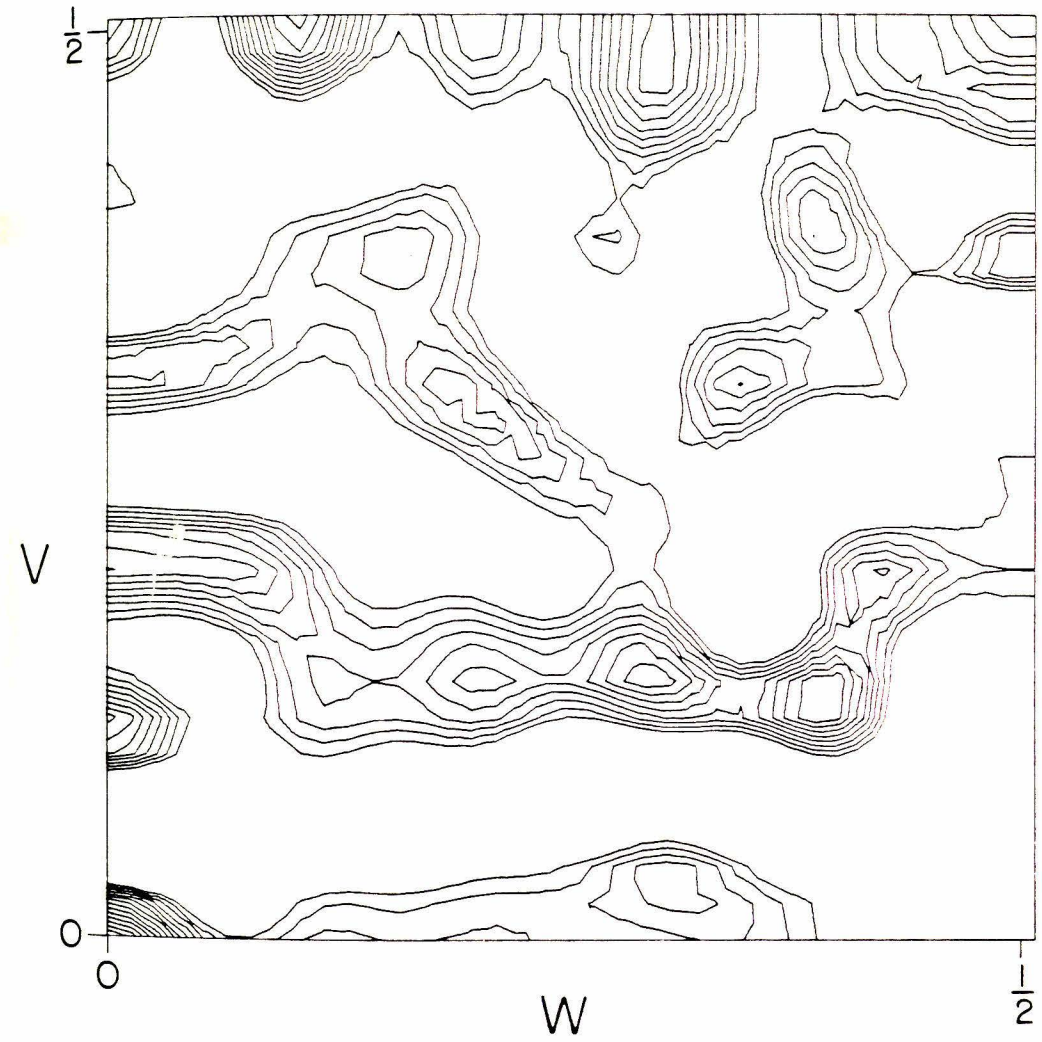


Figure 12b

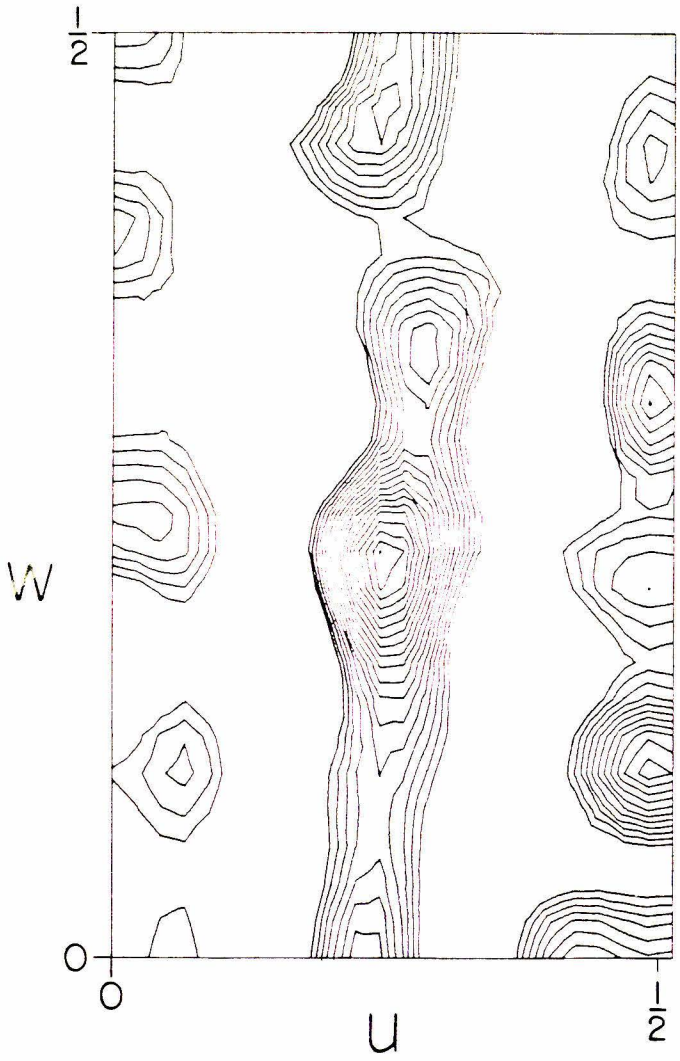


Figure 12c

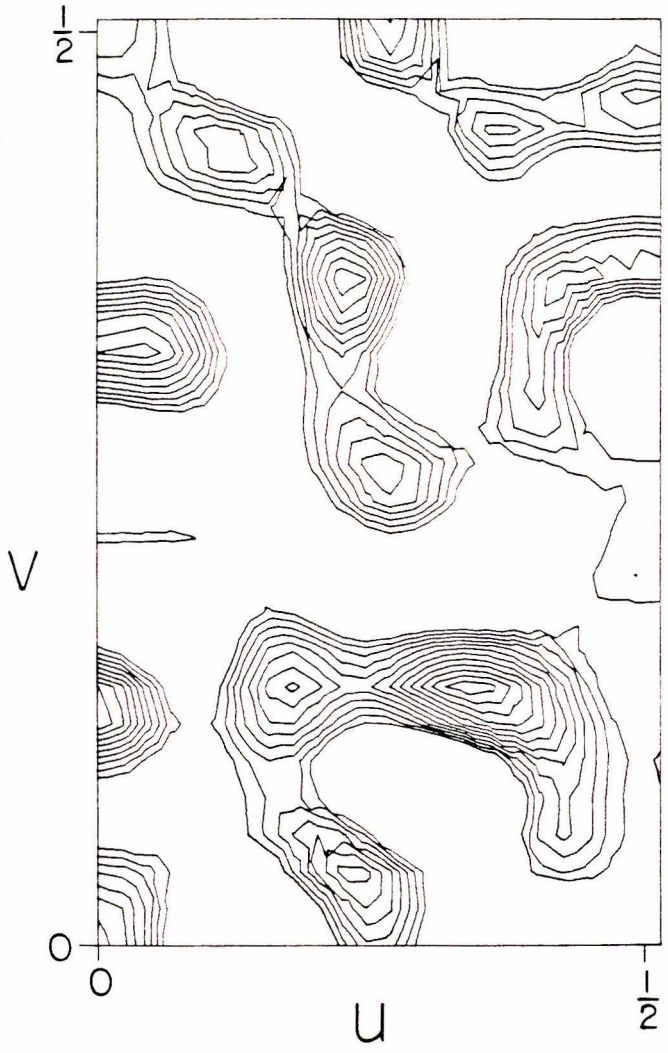


Figure 13a

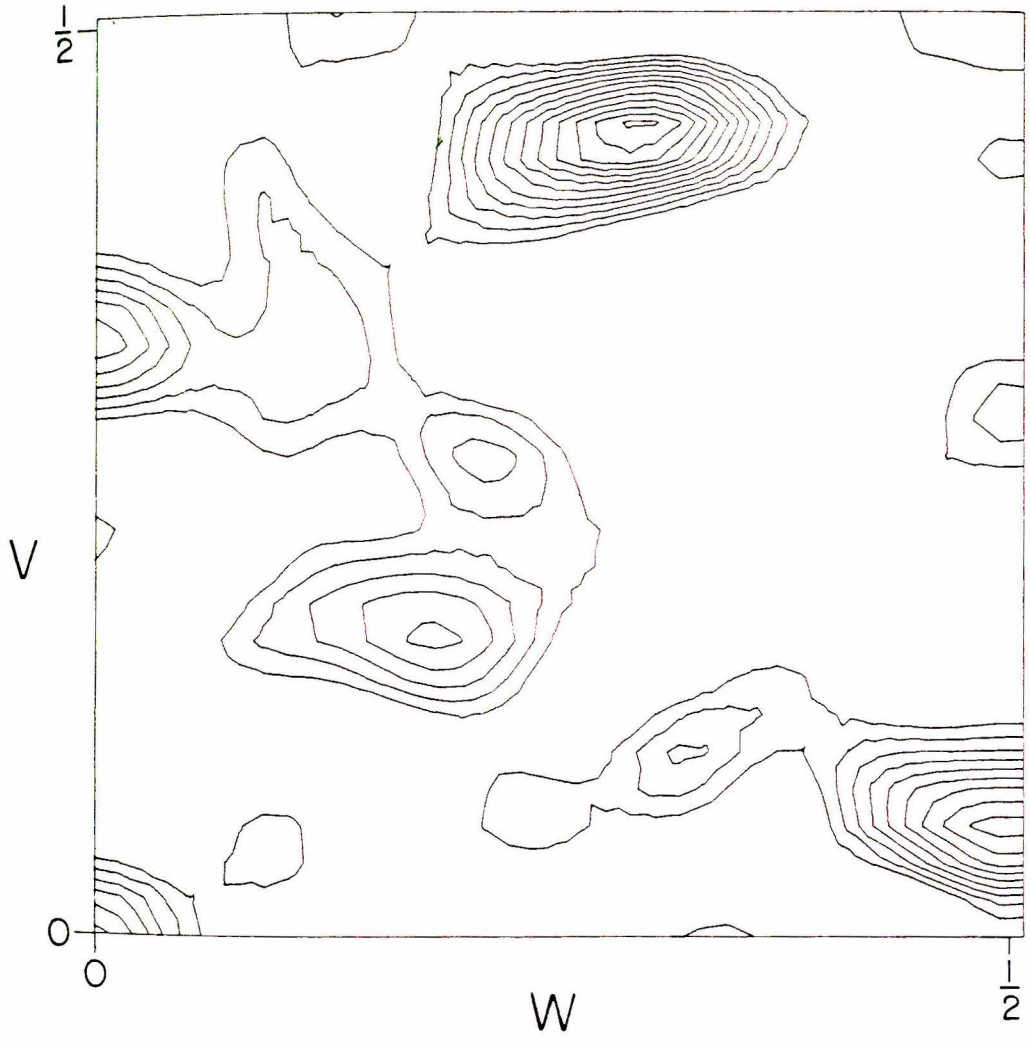
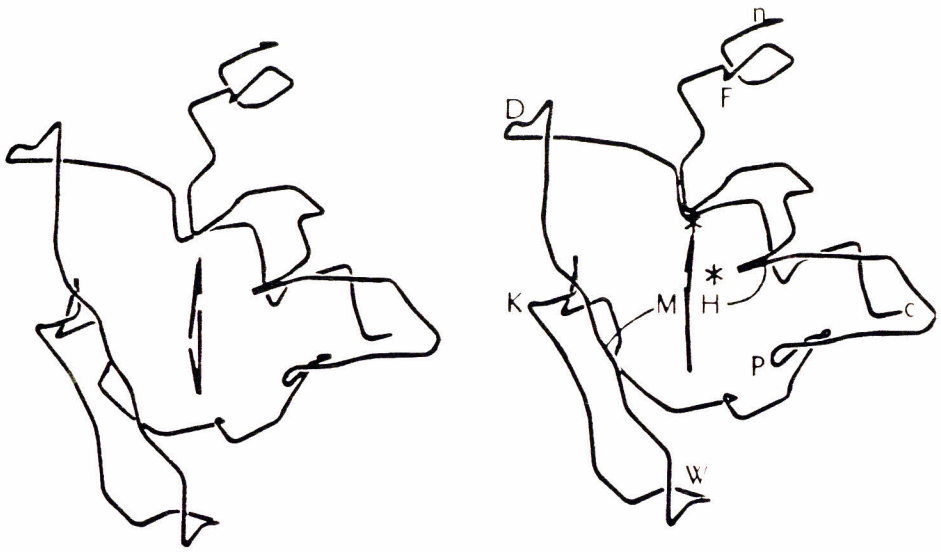


Figure 13b

Figure 13c

Figure 14



Chapter 3

STRUCTURE DETERMINATION AND REFINEMENT AT 2.0 Å RESOLUTION

MATERIALS AND METHODS

Extraction, Purification, and Crystallization

A dry powder of Pseudomonas aeruginosa (strain P6009) bacterial cells, and purified ferricytochrome c_{551} were the generous gift of Dr. Henry Harbury. P. aeruginosa cells were grown and the cytochromes isolated and purified by the method of Harbury (1) and Ambler (2). The purified protein was stored until used for crystallization on Sephadex SP-25 or as a lyophilized powder at -20°C.

Crystallization trials were initiated using the micromembrane diffusion technique of Zeppezauer (3,4). Capillary cells 1 mm in diameter and 30 mm long were filled with approximately 20 µl of 2.0% protein. One end of the capillary tube was covered with Union Carbide cellulose casing. Dialysis tubing was not used because it permitted leakage of the relatively small protein from the cell. The cell was placed in a reservoir containing the crystallizing medium.

Small rectangular crystals of cytochrome c_{551} grew in diffusion cells from reservoir solutions containing 40-50% saturated $(NH_4)_2SO_4$, 1.0 M NaCl, and 0.01 M $(NH_4)_2HPO_4$ in the pH range 5.6 to 5.9. The crystal length along the a axis was generally 0.2 to 0.3 mm, while the lengths in the b and c directions were approximately one-tenth as great. In extremely rare cases, a single crystal would grow with dimensions up to 0.8 x 0.15 x 0.15 mm. The crystal space group was $P2_12_12_1$ with unit cell dimensions a = 29.43 Å, b = 49.00 Å, c = 49.66 Å, and one molecule per asymmetric unit. There were two

difficulties involved in using these crystals for high resolution single crystal structure analysis. First, the crystals tended to be either too small, or else highly intertwined multiple crystals that could not be separated. Second, crystals could not always be grown from these conditions and many months often went by without successful growth of new crystals. Both these problems were alleviated to some extent by increasing the protein concentration to 5%, passing the protein solution through a Millipore filter to remove micronuclei, and seeding. Centrifugation at approximately 7,000 rpm in a 2-ml centrifuge tube was originally the final purification step prior to crystallization. Although this procedure removed some particulate impurities and crystallization nuclei, it was fairly ineffective. Relatively large amounts of precipitation still appeared in the cells before crystals began to grow. Millipore filtration was more successful and volumes of protein as small as 0.1 ml could be filtered with better than 85% recovery. In this procedure, a Millipore Swinny filter holder for syringes (13 mm diameter) was filled with approximately 0.1 ml of protein solution. A 2-ml centrifuge tube and adaptor with their top halves cut off, and the filter holder assembly were placed inside a Sorvall SM24 rotor tube. The protein was forced through the Millipore membrane by centrifugation at the minimum speed of the Sorvall RC-2 centrifuge. In addition to being able to Millipore filter very small volumes of protein with high recovery rates, the procedure avoided the foaming that usually occurs when protein solutions are passed through Millipore filters with syringes using

pressure. This procedure eliminated noncrystalline precipitation, reduced crystal twinning, and allowed higher protein concentrations to be used, which increased crystal size somewhat.

Heavy Atom Derivatives

Heavy atom derivatives were prepared by transferring crystals to 75% $(NH_4)_2SO_4$ solutions with 1.0 M NaCl at pH 5.6 to 5.9, containing various metal complexes. The extensive experience summarized by Baumber et al. (5) was used as a guide in deciding on crystallizing conditions and variables.

In all, eight different heavy metals in the form of twenty-three different heavy atom compounds were tested under various soaking concentrations and times. Five complexes gave clear, unambiguous intensity changes, probably resulting from nearly identical heavy atom binding sites. Because of this, and the shortage of good crystals, diffractometer data were collected for only the K_2PtCl_4, the $UO_2(NO_3)_2 \cdot 6H_2O$, and the $NaAu(CN)_2$ derivatives listed in Table I.

Data Collection and Reduction

Data reduction, heavy atom refinement, multiple isomorphous replacement (MIR) phase calculations, and constrained difference Fourier refinement were done almost entirely on a Data General NOVA 800 minicomputer with a 32K, 16-bit-word core, floating point hardware,

TABLE I: Crystal Data

Crystal Type	Crystal Size (mm)	Soaking Time	Cell Constants (Å)	Heavy Atom	Change in ΔF	Resolution
Native	0.2 x 0.2 x 0.9	--	29.43 49.00 49.66	--	--	1.96 Å
K_2PtCl_4	0.2 x 0.2 x 0.8	240 hrs.	29.48 48.93 49.74	0.2 mM	21.6%	2.10 Å
$UO_2(NO_3)_2 \cdot 6H_2O$	0.2 x 0.2 x 0.6	18 days	29.51 49.24 49.42	25.0 mM	25.1%	2.43 Å
$NaAu(CN)_2$	0.3 x 0.3 x 0.8	7 days	29.45 49.11 49.64	1.0 mM	25.4%	2.25 Å

two magnetic tape drives, a 2.5 megabyte moving head disk, and matrix line printer and plotter. An IBM 370/158 computer was used for some Fourier calculations.

X-ray diffraction data were collected with a General Electric XRD-490 quarter circle diffractometer controlled with a PDP-8 computer. CuK_α radiation was generated with a GE CA-8S Cu target x-ray tube operated at 45 KV and 18 ma and a graphite plane monochromator in the perpendicular geometry.

Crystals were mounted in capillary tubes with diameters as small as possible (0.5 mm) in order to reduce background scattering (6). After crystal alignment, the (8,0,0) reflection at $\chi = 90°$ was measured at phi angles from 0° to 360° in increments of 10° to find the best phi zone for the data collection and for absorption corrections (7). Unit cell parameters were determined by scanning 15 to 20 reflections, spanning a representative sample of reciprocal space.

Data collection was started at $2\theta = 26°$. 2θ shells of increasing resolution, containing 800 to 1000 reflections, were collected to a maximum resolution permitted by counting statistics. The $2\theta = 2°$ to 26° shell was collected at the very end. This was done because the low resolution reflections are not affected as much by crystal decay (8). Also, backgrounds were counted for longer times in this shell and the additional crystal exposure then did not affect other data. The order in which the data shells were collected was the same for the native and all derivatives.

The background measurements were made on both sides of the peak at ω offsets of 0.10° to 0.20° from the limits of the peak scan. For 2θ greater than 26°, the background times were ten to fifteen seconds per side. These short backgrounds were used to calculate more accurate backgrounds by means of a linear least squares fitting routine (6). The backgrounds for the more difficult to fit 2θ = 2° - 26° shells were collected with the total background times equal to the scan times. Crystal decay of approximately 10% was corrected for by measuring three standard reflections at intervals of 100 reflections and applying a linear time correction to the intensities.

ω step scans were used. The scan width was determined by the size of the largest reflections for each particular crystal. The scan time was adjusted, depending on the intensities for the particular crystal and the 2θ shell, to obtain satisfactory counting statistics. All data for the native and each heavy atom derivative were collected on a single crystal each.

After correcting for the background, absorption, and crystal decay, Lorentz (9) and polarization (10,11) corrections for monochromatized x-rays were applied to the intensities and structure factors, $|F_{obs}|$, were calculated. The $|F_{obs}|$ of heavy atom derivatives were then scaled to the native data (12). A linear and exponential scaling constant were computed for each derivative 2θ shell to scale the derivative $|F_{obs}|$ to the native data.

Structure Solution

The distribution of native F_{obs} and the fractional changes in structure factors for the derivatives are displayed as a function of scattering angle in Figure 1. The average changes in structure factors for each of the derivatives were quite similar. The major heavy atom binding sites were located from $(\Delta F)^2$ difference Patterson maps calculated from each derivative to the resolution indicated in Table I. Harker sections for these Patterson maps are shown in Figures 2-4. The K_2PtCl_4 difference Patterson map indicated the presence of two major sites. These sites were consistent with the single K_2PtCl_4 site that was unresolved in the low resolution (4 Å) structure determination (13). Single sites were found in the $UO_2(NO_3)_2 \cdot 6H_2O$ and $NaAu(CN)_2$ difference Patterson maps. These sites were also the same as those found in the low resolution analysis.

Refined MIR phases were obtained by alternating cycles of MIR phase calculations (14,15) and least squares heavy atom parameter refinement (16,17). Additional minor sites were located and heavy atom parameters were adjusted from heavy atom difference Fouriers and error syntheses (18). The Patterson map vectors were converted into atomic coordinates, with the relative origin ambiuity having already been solved from low resolution single isomorphous replacement (SIR) phased ΔF Fouriers (13). Phase refinement was started with the Patterson map heavy atom coordinates and temperature factors calculated for each derivative from Wilson-type plots of $\ln \{<\Delta F>/<f_H>\}$ versus s^2. Ten

cycles of least squares refinement of coordinates and occupancies were carried out, with isotropic temperature factors being adjusted when necessary. At this point, double difference maps were calculated for each derivative. No additional sites were found for the K_2PtCl_4 derivative, one additional site was found for the $UO_2(NO_3)_2 \cdot 6H_2O$ derivative, and two additional sites were found for the $NaAu(CN)_2$ derivative. The additional $UO_2(NO_3)_2 \cdot 6H_2O$ derivative site had also been found in the low resolution analysis (13).

Based on these new sites, 26 cycles of refinement were completed. During this refinement, temperature factor refinement was begun which resulted in individual temperature factors for each site instead of the Wilson plot average temperature factors for all sites of the same derivative. New double difference maps were then calculated and two additional minor $UO_2(NO_3)_2 \cdot 6H_2O$ sites were added. From the $NaAu(CN)_2$ double difference map, it was apparent that this major site should be refined with anisotropic temperature factors. Final refinement cycles from here resulted in the atomic parameters and refinement statistics in Tables II and III.

The distribution of the figure of merit for all 2643 reflections with sin θ to a resolution of 2.4 Å is shown in Figure 1. The average figure of merit was 0.924 for centric reflections and 0.763 for all reflections. The ratios of the mean lack of closure error to the changes produced by the heavy atoms is also shown in Figure 1. The significantly higher values for the uranyl derivative along with other

TABLE II: Final Heavy Atom Parameters from Phase Refinement

Derivative	Site	x	y	z	A[1]	B[2] or β_{11}	β_{22}	β_{33}	β_{12}	β_{23}	β_{13}
K_2PtCl_4	1	0.0707	0.0308	-0.1655	15.967	2.14					
	2	0.0022	0.0219	-0.1420	7.098	4.22					
$UO_2(NO_3)_2 \cdot 6H_2O$	1	0.1166	0.0160	-0.8553	14.822	2.29					
	2	0.2581	0.4418	-0.6979	11.487	3.69					
	3	0.4071	0.0358	-0.3194	1.894	0.43					
	4	0.1064	0.0254	-0.9149	2.340	0.20					
$NaAu(CN)_2$	1	0.1818	0.1893	-0.9403	22.718	0.0070	0.0057	0.0026	0.0063	-0.0052	-0.0052
	2	0.1297	0.4830	-0.2935	4.809	2.58					
	3	0.1646	0.3889	-0.9740	4.041	2.10					

[1] Multiply the occupancy A by 2.65 to put it on an absolute scale.

[2] Temperature factor expression was $e^{-B(2\sin\theta)^2}$ for isotropic temperature factors and $e^{-(\beta_{11}h^2+\beta_{22}k^2+\beta_{33}\ell^2+\beta_{12}hk+\beta_{23}k\ell+\beta_{13}h\ell)}$ for anisotropic temperature factors.

TABLE III: Multiple Isomorphous Replacement Refinement Statistics

	All Data	Centric Data
Figure of Merit	0.763	0.924
Mean Relative Error[1]	2.35	
K_2PtCl_4 Derivative:		
RMS E[2]	8.75	--
R_{Kraut}[3]	0.092	0.114
R_{Cullis}[4]	--	0.425
$UO_2(NO_3)_2 \cdot 6H_2O$ Derivative:		
RMS E[2]	16.33	--
R_{Kraut}[3]	0.162	0.194
R_{Cullis}[4]	--	0.700
$NaAu(CN)_2$ Derivative:		
RMS E[2]	11.75	--
R_{Kraut}[3]	0.115	0.126
R_{Cullis}[4]	--	0.457

[1] MRE = mean relative error (17).

[2] RMS E = root mean square lack of closure (17).

[3] $R_{Kraut} = \Sigma$ (lack of closure)/$\Sigma |F_{PH}|$ (19).

[4] $R_{Cullis} = \Sigma$ (lack of closure)/$\Sigma|\Delta F|$ (20).

refinement statistics in Tables II and III indicate that this derivative was the poorest.

A map of the protein electron density to a resolution of 2.4 Å was calculated, using the centroid MIR phases and the native structure factors weighted with figures of merit. The density was plotted onto acetate sheets by computer and placed in a Richards box, where a Kendrew wire model was constructed to fit the density. The molecular boundary was unambiguous. The heme and alpha helices were especially prominent in this map. The polypeptide chain was continuous and could be followed quite easily. The only difficulty in main chain interpretation was around Ala 35, Gly 36, and Gln 37, where the chain is extended out from the molecular surface and the electron density is not as intense. Most of the main chain carbonyl oxygens were nicely defined, and were very useful in constructing the wire model. Certain aromatic and proline residues were the most dense side chains. However, other residues of this same type were not especially dense. The same is true for ionized side chains exposed to the solvent. Most of them were defined well enough to be confident of their location, but several were blurred and did not rise above the map noise level.

The root mean square error level (15) was computed to be 0.16 e/Å^3 from the final phase cycle. This value was obtained after placing the data on an absolute scale, following structure factor calculations during refinement (to be discussed later). On this scale, the iron atom was at a density of 3.12 e/Å^3, well-defined side chains were at approximately 1.0 e/Å^3, and the average main chain was around

0.75 e/Å^3. The electron density averaged at all atomic centers was calculated to be 0.62 e/Å^3.

Structure Refinement

The $\underline{c}_{551}$ structure was refined by a constrained difference Fourier technique, using programs developed by Chambers and Stroud (21). The details of this refinement procedure have been described by them, so only a brief summary will be presented here, including modifications to the earlier methods and aspects specific to $\underline{c}_{551}$. Constrained difference Fourier refinement is a cyclic process by which the relatively inaccurate MIR coordinates and phases are improved. After initial coordinates are obtained, (a) structure factors (F_{calc}) are calculated and scaled to the observed structure factors (F_{obs}); (b) a difference Fourier map is calculated using $|F_{obs}| - |F_{calc}|$ amplitudes and calculated phases, ϕ_{calc}; (c) coordinates and temperature factors are shifted using conventional equations (9,22); and (d) the coordinates are constrained to fit the difference map density and at the same time have bonding parameters that are consistent with those observed in amino acid and small peptide high resolution structures.

This automated procedure is cycled until no further improvement in the structure can be obtained. At that point, the difference map and atomic coordinates are plotted on plexiglass and visually inspected. Manual adjustments of coordinates are made based on

difference map densities, and the automated refinement process is started again.

Structure factors were calculated by the conventional method. Atomic scattering factors were calculated from the data of Forsyth and Wells (23). Isotropic temperature factors were applied using the expression $\exp(-B\sin^2\theta/\lambda^2)$, with B being allowed to assume any of seventeen possible values distributed between 6 Å^2 and 40 Å^2. The structure factor results were used to put the F_{obs} data on an absolute scale. A plot of $\ln\{\langle I\rangle/\langle f^2\rangle\}$ versus $\sin^2\theta/\lambda^2$ yielded the scale factor and an overall temperature factor of 12.4 Å^2. Throughout the refinement, the refined average temperature factor was approximately 17 Å^2. Four-parameter scaling of $|F_{calc}|$ to $|F_{obs}|$ was used, fitting in $\ln\{\langle F_{obs}\rangle/\langle F_{calc}\rangle\}$ to the function $A_1\exp(-A_2s^2) + A_3s^2 + A_4$. Despite the significant overall improvement obtained at low resolution in using this scaling over linear scaling (see Figure 6), reflections with d>7 Å were still omitted from difference maps. This was done in order to eliminate low order ripples resulting from scaling errors that will affect atomic parameter shifts adversely. Temperature factor shifts were especially vulnerable to this type of error. The program ran in one and three quarter hours for 5430 reflections to 2.0 Å resolution.

Difference maps were initially calculated on the IBM 370/158 using a conventional Fourier program and later using a fast Fourier transformation program (24) written for space group $P2_12_12_1$ and adapted for the NOVA 800 (25). The difference maps were calculated

with a grid interval of approximately 0.85 Å. In the final stages of refinement, significant improvement was obtained for the higher resolution data by decreasing the grid intervals to 0.5 Å. The FFT program ran in one-half hour to two hours, depending on the grid size.

Computations of shifts to coordinates and temperature factors, based on difference map densities, were made using conventional equations (9,22). The shift to atomic coordinates is $\Delta X = -2Ng/C$ for a noncentrosymmetric structure, where ΔX is the coordinate shift, N is the shift factor, $\underline{g}$ is the difference map density gradient at the atomic center, and C is the curvature at the atomic center in the corresponding F_{obs} synthesis. The gradient $\underline{g}$ was evaluated by fitting the difference map densities to a three-dimensional quadratic function (21). During early stages of refinement, curvatures proportional to the atomic number were used, with the shift factor proportionality constant being chosen to produce reasonable shifts (21). In later stages of refinement, significant improvements were found by using shifts calculated from real curvatures, computed in a separate program. Curvatures were calculated for all combinations of atom types and temperature factor groups. The curvature along the $\underline{x}$ axis is calculated as (22,26):

$$\frac{-4\pi^2}{Va^2} \sum_{hk\ell} mh^2 f \exp(-Bs^2)$$

where m is the multiplicity, and the summation is over the appropriate

resolution of data. The curvature was taken as the average of the values calculated along the three principle axes. Using real curvatures and a shift factor of 1.0 speeded up the rate of convergence and eliminated the need to guess at a proportionality constant.

The difference map density was evaluated at the new atomic center. If the atom had been moved into lower density, the shift was cut in half. If it was necessary to cut the shift in half more than three times, the shift was set to zero. The average shift during the initial refinement was as large as 0.4 Å, but as the R-factor was reduced, the average shift was also reduced. At R = 34%, the average shifts decreased to approximately 0.2 Å and then down to 0.1 Å during the final cycles of refinement.

Temperature factors initially were computed from difference map densities at atomic centers (21). However, throughout most of the refinement, temperature factor shifts were calculated from the equation (22):

$$\Delta B = 2.0\ N(\partial^2\rho/\partial r^2)/(\partial C/\partial B)\ \exp(-2\ G_{rad}/s_{max})$$

$(\partial^2\rho/\partial r^2)$ is the difference map curvature, averaged from the values calculated along the three crystallographic axes. $(\partial C/\partial B)$ is the derivative of the atomic curvature with respect to B and is evaluated as (22):

$$\frac{4\pi^2}{V} \sum_{hk\ell} mf \exp(-Bs^2)\ s^4$$

G_{rad} is the root mean square coordinate shift and S_{max} is the maximum shift permitted. The exponential term partially decouples the coordinate and temperature factor shifts as described previously (21). The shifting program ran in approximately one-half hour.

During the calculation of curvatures for use in the shifting program, estimated errors in the atomic coordinates were calculated by the method of Cruickshank (27). The standard deviations of peak positions were calculated from the quotient of the standard deviation of the electron density slope and the curvature. The standard deviation of the electron density in the $\underline{x}$ direction was calculated as:

$$\sigma\ (\partial\rho/\partial x) = \frac{4\pi}{Va} \left\{ \sum_{hk\ell} mh^2\ (F_{obs}-F_{calc})^2 \right\}^{\frac{1}{2}}$$

where V, a, h, and m are the cell volume, the axis length, the reflection index, and the reflection multiplicity. The total estimated error was calculated from the σ's along each axis. This calculation is based on the assumption that the error in the structure factors can be approximated by the distribution of $\Delta F = F_{obs}-F_{calc}$ and that this distribution is normal with no systematic errors. The curvature was calculated for each atom type and temperature factor group as described above. Estimated errors in the electron density were also calculated (27).

The constrained coordinates were obtained from the unconstrained difference map coordinates by a least squares conjugate-gradient minimization routine (21,28-31). The constraining shifts were applied to the coordinates so as to minimize deviations between calculated and idealized (32,33) bonding parameters. The residual contained terms for bond lengths (L), bond angles (A), bond torsion angles for planar atoms (T), and terms for atoms with chirality (C). In addition, there was a term that minimized the distances the atoms had been moved from the original position to the constrained position. The residual was calculated as follows:

$$R = W_L \Sigma w_L (L-L_s)^2 + W_A \Sigma w_A (A-A_s)^2 + W_T \Sigma w_T (T-T_s)^2 + W_C \Sigma w_C (C-C_s)^2 + \Sigma\, 1/B(r-r_o)^2$$

where W's were global weights for each type of constraint and w's were local weights, assigned to each specific constraint. r_o was the starting coordinate and $\underline{r}$ was the constrained position. How closely the structure approached the ideal structure was determined by the weightings. This minimization procedure was applied to a molten zone of residues, usually containing three amino acids. After a sufficient number of cycles had been calculated for a particular zone to be well constrained, the program then moved to the next zone. In general, better results were obtained by using relatively small W's and a large number of cycles. All bond lengths and angles in the structure were included in the constraining process. Torsion angles of 0° to 180°

were used to planarize groups of atoms. The heme was treated like the amino acid residues, but without a connecting amide bond. Residues attached to the heme had pseudo-atoms which constrained the geometry of the connection. These atomic coordinates were frequently equated with the corresponding real atom coordinates. The constraining program required three-quarter hour to execute.

An important observation with regard to the constraints was that the best results were not always obtained by constraining the structure after each cycle of shifts in the refinement cycle outlined above. At certain points, the one-shift and one-constraint process did not result in improvements in the structure. Computing several cycles consisting of shifts, structure factors, and difference maps without any constraints followed by the application of constraints would generate an improvement that was inefficient or not possible from the one-shift and one-constraint cycle. On the other hand, letting the structure deviate too far from ideality by applying too many shifting cycles without constraints would lead to a poorer structure after the final constraints were applied.

Difference maps on a scale of 0.25 inch per angstrom and the latest coordinates were plotted onto plexiglass sheets for visual inspection when the automated procedures failed to produce any further improvements. Careful evaluation of these minimaps was critical to successful refinement. These maps were frequently difficult to interpret, especially at the beginning of the refinement when the noise level was relatively high. At this point, not all changes were made

with a high degree of confidence. Despite making some incorrect movements, the overall changes improved the structure so that the automated procedure could transcend the local minimum. The occasionally incorrect interpretations from the minimaps were discovered in subsequent minimaps. As the R-factor approached 20%, these maps became much more noise free and the interpretations were made with confidence. In $\underline{c}_{551}$, the manual adjustments were almost always applied solely to the side chains, although movement of a main chain carbonyl oxygen was occasionally encountered. The magnitudes of the shifts were almost always larger than could be made by the automated program. In addition to simple translations, the other types of common movements made were rotations applied to side chains of Val, Thr, Asp, and Glu residues. On several occasions, groups of atoms were cycled between two different conformations in alternate minimaps. This occurred for ionized side chains on the molecular surface.

Each time a difference map was plotted on plexiglass for visual inspection, an automatic peak searching program was executed to locate possible solvent. The prospective solvent peaks obtained from the screening program were examined in the plexiglass map and evaluated as to whether they were solvent or density arising from other sources. The criteria used for this evaluation were the difference map density, the peak shape, and the proximity to protein atoms. Care was used so as not to place solvent in density arising from protein coordinate errors. New solvent peaks were added conservatively in an attempt to include only real solvent. Difference map densities and refined

isotropic temperature factors from previously included solvent were also examined to eliminate solvent molecules which no longer seemed reasonable. At the end of refinement, forty-two solvents remained in the structure, out of ninety-four inserted during refinement. Solvent peaks were given full oxygen scattering factors. Initial temperature factors were based on the difference map density.

After making the manual adjustments to the appropriate coordinates, including additional solvent and deleting previously defined solvent that did not refine, new structure factors were calculated and the automated refinement was started again. The manual adjustments and solvent changes always increased the R-factor 2½% to 5%, but it then fell quite rapidly to below the value where it hung up prior to visual inspection and manual adjustments.

The refinement progress is displayed in Figure 5, in which the R-factor is plotted at each structure factor calculation. Refinement steps that were discarded for one reason or another and were not directly involved in getting to the refined set of coordinates are not shown. A detailed description of all calculations is presented in Appendix III.

Refinement was begun by running nine cycles of atomic shifts and constraints in the MIR electron density map prior to the calculation of the first difference maps. Shifts were obtained by the same procedures as in difference map refinement except the gradients were obtained from the MIR map. This procedure resulted in an improvement

in the R-factor* from 48.6% to 43.3%. Attempts to obtain further improvements from shifting coordinates in the MIR map were unsuccessful. After converging to 43.4%, the MIR map was plotted onto plexiglass minimaps, along with the initial wire model coordinates and the 43.3% R-factor coordinates. In general, there was a significant improvement in the fit of the model to the MIR electron density. The electron density at atomic centers increased from 0.62 e/Å^3 for the wire model coordinates to 0.73 e/Å^3. The structure was also significantly more constrained. The improvement in the model fit to the density can be attributed to the reduction of errors that resulted in trying to fit the Kendrew model to the electron density in the Richards box. Although most of these errors are not obvious in the Richards box, they become very apparent in the minimaps. After model fitting, there were a small number of residues that did not fit the MIR electron density as well as before. This is probably a result of the new structure being more tightly constrained than the wire model coordinates. Although the fit did not produce the best possible set of MIR coordinates, the accuracy was more than enough to begin refinement.

Using the complete set of 5430 reflections to 1.96 Å resolution. with an R-factor of 43.3%, the first difference map was computed. After three cycles of difference map refinement, the R-factor dropped rapidly to 35.7%. At this point, it was realized that the heme atoms

---

* $R = \sum (|F_{obs}| - |F_{calc}|)/\sum|F_{obs}|$

were being overshifted. To alleviate this problem, their temperature factors were set to 7 $Å^2$, while the protein temperature factors remained at 14 $Å^2$. This resulted in increased heme atom curvatures and smaller heme atom shifts, relative to the protein atoms. This, along with the more accurate temperature factors in the structure factor calculation, resulted in a decreased R-factor to 33.7% after one cycle.

Attempts to improve these coordinates further using the automated shifting program failed, and minimap I was contoured on plexiglass with the atomic positions marked. The overall noise level was quite high and there were not too many obvious changes to make. The important feature was that the temperature factors for a great number of residues were obviously incorrect and using an average temperature factor of 14 $Å^2$ was not acceptable.

After making a few manual adjustments, the structure factors were calculated. The R-factor was found to increase to 36.2%, and automated difference map refinement was continued. Two cycles lowered the R-factor back to 33.6%. At this point, refinement of isotropic temperature factors for individual atoms was begun. In addition, 380 reflections with assigned $|F_{obs}| = 0$ (which included reflections with measured $|F_{obs}|<0$) were eliminated from the structure factor calculations and difference maps. Three more cycles of refinement resulted in the R-factor dropping from 33.6% to 25.8%.

Since the R-factor could not be induced to fall below 25.8%, minimap II was contoured. The difference densities were much smaller in this map, both in terms of density arising from noise and from

temperature factor errors. As a result, shifts that needed to be applied to coordinates were more obvious. In addition to the manual shifts, 35 solvents were added from the map. As usual, these adjustments resulted in an increased R-factor (R = 30.0%).

Automated shifts in the difference maps were again started, but this time real curvatures were used in place of the artificial curvatures set according to atomic number. Ten cycles reduced the R-factor from 30.0% to 21.5%. During the course of this refinement, reflections with (a) d>7 Å resolution and (b) $|F_{obs}|<2\sigma$ were eliminated from the structure factor calculations and difference maps. Reflections with d>7 Å were eliminated to reduce scaling errors that would introduce errors into atomic parameters, especially temperature factors. Reflections with $|F_{obs}|<2\sigma$ were eliminated because of their unreliability. The dramatic improvement in the subsequent difference maps indicated that these modifications were necessary. The estimated effect on the R-factor was -0.7% and -1.3%, respectively.

Minimap III, calculated at R = 21.5%, had a much lower noise level than did previous maps. As a consequence, coordinate shifts were much more obvious. The relatively large, positive, and diffuse noise levels observed in previous maps in alpha-helical regions of the protein had disappeared in this map. The general background noise level around the heme decreased. Although these improvements could have resulted from the better coordinates due to the additional refinement, they are probably a more direct result of the removal of the $|F_{obs}|<2\sigma$ reflections.

After coordinate shifts were made, 25 solvent molecules were added and eighteen were removed, structure factors were calculated to yield an R-factor of 26.4%. As seen in Figure 5, automated refinement was continued through two more minimaps calculated at 18.9% and 17.5% followed by a last set of automated refinement cycles to a final R-factor of 16.2%. The refined structure included 42 solvent molecules.

The R-factor is displayed in Figure 6 as a function of the resolution. The difference between the four-parameter and linear scaling of F_{calc} to F_{obs} is shown. The average change in phase is presented in Figure 7 as a function of resolution. The differences shown are between the refined phases and both the MIR phases to 2.4 Å resolution and the calculated phases of the wire model coordinates to 2.0 Å resolution. The average changes in phase angles to the indicated resolution were 40.9 Å and 54.2 Å, respectively, In addition, the average phase change was calculated to be 54.1° between the MIR phases and the initial wire model phases. The phase change between the refined phases and the MIR phases, weighted by the figure of merit, is also displayed in Figure 7. The smaller average change of 34.5° indicates the figure of merit is a good weighting function. MIR electron density maps calculated from a Fourier summation weighted by the figure of merit will increase the relative weights of terms with MIR phases closer to the refined phases.

RESULTS

Heavy Atom Binding Sites

As with other cytochrome structures (34), K_2PtCl_4 was found to bind in a double site to a methionine side chain, a consequence of statistically random binding to one or the other of the two electron lone pairs on the sulfur. In native cytochrome c_{551}, one of these two K_2PtCl_4 binding sites was relatively exposed and fully occupied, while the second site was within 2.1 Å of the epsilon oxygen atom of Gln 37 from a symmetry related molecule and only partially occupied. The major $UO_2(NO_3)_2 \cdot 6H_2O$ site was approximately 3.5 Å from the side chain carboxylate of Glu 4 while the third site was also near this residue. The other two sites were more than 5 Å from any native protein residues. The major $NaAu(CN)_2$ binding site was closest to the exposed heme pyrrole ring 2 (3.6 Å away), while one minor site was near Pro 60 and the other near arginine 47.

Protein Structure

The main chain conformation of Pseudomonas aeruginosa cytochrome c_{551} found at low resolution (35) has now been verified by this refined high resolution structure. The structure is illustrated by a ribbon drawing and stereo view of the alpha carbons in Figures 8 and 9. There was no difficulty in following the main chain path. The backbone conformation began with an extremely well-defined N-terminal helix, followed by attachments to the heme at residues Cys 12, Cys 15, and His 16. This was followed by a loop on the right side from His 16 to Val 23, similar to the 20's loop in eukaryotic cytochrome c (36,37). After this loop, a stretch of chain near the bottom of the heme led to the back of the molecule. Then, instead of looping below the heme to the front of the molecule and back again, as in other c-type cytochromes (36-39), the c_{551} chain cut across the bottom rear to start the 40's helix. The 40'x helix to the left side of the heme was followed by a loop of chain down to Trp 56 below the heme, and back up again to make the Met 61 axial heme attachment. This extended loop supplied a polypeptide cover for the heme at the bottom where the loop described earlier had been deleted, by comparison with other c-type cytochromes. The chain then traveled from Met 61 to the upper left rear, to the beginning of the final C-terminal helix.

The Ramachandran angles ϕ and ψ were calculated from the refined coordinates for each alpha carbon in the main chain. The results are listed in Table IV and plotted in Figure 10, along with the Pullman

TABLE IV. Ramachandran Angles

Residue	Phi	Psi	Residue	Phi	Psi
E 1	--	338	A31	126	138
D 2	85	304	A32	121	134
P 3	123	148	K33	115	127
E 4	110	147	F34	91	193
V 5	100	141	A35	231	60
L 6	123	134	G36	55	215
F 7	130	145	Q37	101	317
K 8	100	124	A38	116	314
N 9	127	130	G39	262	186
K10	98	186	A40	115	137
G11	258	191	E41	122	143
C12	130	137	A42	118	139
V13	126	161	E43	107	136
A14	97	167	V44	118	139
C15	53	172	A45	126	136
H16	43	330	Q46	116	145
A17	48	349	R47	101	144
I18	107	142	I48	124	133
D19	68	168	K49	105	132
T20	54	326	N50	85	176
K21	106	317	G51	279	337
M22	60	74	S52	30	328
V23	93	143	Q53	20	321
G24	57	357	G54	281	153
P25	133	331	V55	105	133
A26	116	330	W56	71	183
Y27	106	148	G57	331	358
K28	118	143	P58	94	170
D29	110	143	I59	80	325
V30	116	128	P60	100	336

Table IV (continued)

Residue	Phi	Psi
M61	92	304
P62	107	343
P63	120	323
N64	80	327
A65	93	244
V66	59	343
S67	90	356
D68	119	132
D69	110	150
E70	116	133
A71	120	142
Q72	118	132
T73	124	137
L74	106	142
A75	121	134
K76	108	155
W77	103	142
V78	124	132
L79	119	162
S80	84	191
Q81	102	147
K82	86	--

quantum energy boundaries at 3 kcal/mole above the global minimum of energy (40). Only a few residues had Ramachandran angles outside this 3 kcal/mole contour, Met 22 being one example. The residue homologous to Met 22 in eukaryotic cytochrome c (Lys 27, just before the sequence Val-Gly-Pro) also lies outside this barrier on the Ramachandran plot, and close to the c_{551} position (41). The only other c_{551} residues outside this contour level were Gly 51 and Ala 35, with small side chains.

A very unusual region of main chain conformation includes Pro 58 through Pro 63. This stretch of six residues contained three prolines. The structure in this region was very similar to the structure of polyproline (42) as seen in Figure 11. In this stereo figure, the threefold left-hand helix of polyproline is presented to the left and c_{551} is presented in a similar orientation on the right. The Ramachandran angles for residues Ile 59 through Asn 64 ranged from 80° to 120° for ϕ and 304° to 343° for ψ (Table IV). This is the region in the Ramachandran plot where the polyproline helix is located (43).

In both the unrefined and refined structures, most of the side chain atoms were well defined in electron density. Internal side chains were particularly good. As discussed under Methods, some terminal atoms for charged residues on the molecular surface extended into density at the noise level, and their coordinates were less precisely known. Data which can be used to assess coordinate reliability will be presented below. A detailed comparison of the side

chain structure with the other c-type cytochromes will be made in the Discussion section.

Figure 12 shows the effect of refinement on the atomic coordinates for main chain atoms, side chain atoms, and all atoms combined. Average distances from the MIR wire model to the final refined coordinates are plotted for each residue. For the entire structure, the average distances were 0.74 Å, 1.15 Å and 0.94 Å for main chain atoms, side chains, and all atoms. The largest deviations in the main chain atoms were in the regions of Ala 35 through Gly 39 and Gln 81 and Lys 82. These were the two regions in the MIR structure where the electron density was not very clear and the initial positions were less accurate. The first group of residues was a loop extended away from the molecular surface into solvent. Gln 81 and Lys 82 were at the C-terminus. Unfortunately, refinement did not improve either of these areas to the quality of the rest of the structure. From Figure 12, it is apparent that most of the large side chain movements are associated with charged residues on the surface of the molecule. An exception to this observation is the side chain of Ile 59. The large change during refinement resulted from a rotation about the bond between the alpha carbon and beta carbon.

Figure 13 shows the antiparallel sheet structure between Gly 51 and Asn 64, including the polyproline-like stretch around the heme methionine attachment. Figure 13a was prepared with the initial MIR electron density and coordinates. The second difference map at R = 25.8% is plotted in Figure 13b to illustrate the information one

can obtain from such a map. In this map, Ile 59 had not yet been rotated about its alpha-carbon/beta-carbon bond as described above. Figure 13c illustrates the final difference map at R = 16.2%, and 13d the final $2F_{obs}-F_{calc}$ electron density map.

Several sections of the $\underline{c}_{551}$ structure approximately perpendicular to the N-terminal alpha helix are shown in the stereo Figure 14. The residues included in this figure are Val 5, Leu 6, Phe 7, and Lys 8. The improvement in the electron density in going from the MIR map (14a) through difference maps (14b,c) to the final $2F_{obs}-F_{calc}$ electron density map (14d) is clear. In some cases, the refined coordinates fit the MIR electron density even better than did the measured MIR coordinates themselves. This is seen in the fit of the alpha and beta carbons of Phe 7 in Figure 14a. Val 5, Leu 6, and Phe 7 were residues that were defined well in the MIR map. Lys 8 is an example of an external ionized residue with less well-defined electron density.

Temperature Factors, Coordinate Errors, and Constraints

Difference map refinement has the advantage of being able to determine individual isotropic temperature factors very accurately. In the early refinement stages of cytochrome $\underline{c}_{551}$, it was apparent that much of the difference map density was a result of errors introduced by using an overall temperature factor. As an example, many external side chains were located in strong negative density with no positive density within the vicinity. Allowing temperature factors

to refine significantly improved the quality of the difference map and permitted errors in coordinates to be seen more easily. In contrast to the results observed for the refinement of bovine beta-trypsin (44) and bovine trypsin and bovine trypsin inhibitor (45), it was found that the temperature factors could be obtained accurately and were a significant parameter in the description of the $\underline{c}_{551}$ structure. A small percentage of the refined temperature factors were significantly different from the values of temperature factors for adjacent atoms. This may have resulted from the influence of atoms in close proximity at this resolution. However, the general temperature factor distribution was very consistent and physically meaningful.

Compared to small molecule crystal structures, the average temperature factor of 17 $Å^2$ for the refined $\underline{c}_{551}$ structure was quite large. This is a general result seen in many protein crystal structures. Large temperature factors may be a result of thermal motion or crystal disorder, with the latter effect probably being more significant (46). The high percentage of solvent in the unit cell contributes to this effect.

Refinement of individual isotropic temperature factors was begun after minimap I in Figure 5. The effect of refinement on the average temperature factor for each residue is seen in Figure 15. The average temperature factor is given for each residue in minimaps II, III, IV, V, and in the final structure. The average temperature factor for the entire structure at each of these points is given in Table V. Main chain atoms and neutral side chains, which were usually located

TABLE V

Average Temperature Factors at Different Stages of Refinement

Minimap	Main Chain Atoms	Side Chain Atoms	All Atoms
II	13.3	19.5	16.5
III	14.1	20.9	17.6
IV	13.8	19.9	17.0
V	12.9	20.2	16.8
Final Structure	13.2	19.7	16.6

in the interior, had consistently lower temperature factors than ionized side chains that were located on the molecular surface and exposed to solvent. It is apparent from Figure 15 that the average main chain temperature factors (a) were lower than those for the side chains, (b) generally were more uniform from one residue to another, and (c) did not change as much during refinement. These observations are all consistent with their well-defined electron density. Exceptions included main chain atoms in residues Ala 35, Gly 36, Gln 37, and Lys 82. As discussed above, the first three residues were exposed to solvents and Lys 82 was the terminal residue. The higher temperature factors for these residues indicated disorder. In addition to main chain atoms, heme, aromatic, hydrophobic, polar, and even some charged residues in Figure 15 refined to low temperature factors. For these residues, the electron density was well defined and the accuracy of the coordinates was high.

The ionized surface side chains behaved less well, with a few exceptions. Residues with average refined temperature factors greater than 30 $\mathring{A}^2$ included Glu 4, Lys 8, Asp 19, Thr 20, Lys 28, Glu 41, Glu 43, Arg 47, Asp 69, Gln 72, Lys 76, and Lys 82. For many of these residues, only the terminal atoms were ill defined. However, all side chain atoms had large temperature factors and undefined electron densities in the case of Lys 82. Exceptions to this rule included the exterior residues Lys 10 and Glu 70. Although charged and on the molecular surface, these residues have lower temperature factors and

well-defined electron densities. Their ordered structure was a result of hydrogen and ionic bond stabilization.

In addition to temperature factors indicating thermal motion, statistical disorder, and the quality of the electron density map in the final structure, changes in temperature factors during refinement can be correlated with the accuracy of the changes in coordinates. Residues where coordinate fit to the electron density is significantly improved during refinement should show a relative decrease in the associated temperature factor. Alternately, a residue moved into an incorrect position as a result of either an error or misinterpretation of the difference map will have an increased temperature factor. A residue which had properly converged might show a small increase in temperature factor while other residues are refined to lower temperature factors and the overall temperature factor is maintained constant.

Temperature factors for aromatic residues illustrate some important points. These were residues that were well defined in the MIR electron density map. They did not need large manual shifts, and the automated shifts during refinement were small (for example, see Figure 12). In Figure 15, Phe 7, Phe 34, and Trp 56 all show small decreases in their average temperature factors during refinement. This is the expected behavior for residues whose positions improve only slightly because of an initially good fit. Tyr 27 and Trp 77 started out with relatively high temperature factors for this type of residue, but decreased nicely during refinement. There are many other examples of this type of improvement for residues in $\underline{C}_{551}$.

Some residues showed large decreases in temperature factors because of large coordinate shifts during refinement. Val 23 is the most striking example. Figure 16 shows the initial MIR map (16a), the first difference map (16b), the final difference map (16c), and the final $2F_{obs}-F_{calc}$ map (16d). The improvement in difference maps and coordinate fit coincided with the marked decrease in the temperature factors from 28 $Å^2$ to just over 10 $Å^2$ in Figure 15b. A very similar improvement is also seen for Ile 59 in Figure 13. Rotation about the $C_\alpha-C_\beta$ bond improved the difference map and lowered the temperature factors significantly.

Residue Leu 44 provides a particularly nice example of the importance of temperature factors in particular and the ability to detect errors from difference map refinement in general. During the measurement of coordinates from the wire model, a paper tape punching error occurred. The coordinate conversion program, which converts the voltage output from a coordinate measuring device to atomic coordinates, converted the wire model Leu 44 to a Val 44. This error went undetected through initial refinement cycles. As the R-factor dropped into the 20's, visual inspection of the minimaps revealed something was in error. Adjustments were attempted by rotations about the alpha-carbon to beta-carbon bond and extension of the Val 44 gamma-carbons into the positive density just beyond their reach. Neither of these corrections resulted in the removal of the positive density. As refinement approached an R-factor of 20%, the temperature factor for the $\gamma 2$-carbon atom increased to $40Å^2$, while the β-carbon and

γ1-carbon atoms had temperature factors of 14 Å^2 and 16 Å^2, respectively. These atoms were fitted into the Leu 44 density but resulted in the γ2-carbon being placed outside the electron density. Figure 17 shows a difference map (17a) and a $2F_{obs}-F_{calc}$ map (17b) with the new Leu 44 coordinates positioned into the maps calculated from the incorrect Val 44 atoms. Figure 17c and 17d shows the coordinates with the new maps. For this well-ordered residue, the difference map density in Figure 17a clearly indicates this residue should be extended to the indicated position. The beta-carbon and γ1-carbon atoms managed to fit the Leu density, but this resulted in the γ2 carbon being placed in an incorrect position with an isotropic temperature factor of 40 Å^2.

In summary, the isotropic temperature factors obtained from difference Fourier refinement are important parameters in describing the protein structure. They are related to the quality of the electron density map at the atomic sites and are therefore a good measure of the confidence one can place in the associated coordinates. Special caution should be exercised in using coordinates of atoms with particularly high temperature factors.

In addition to assessing the reliability of the coordinates from isotropic temperature factors, estimates of the errors in the coordinates were calculated using the procedure of Cruickshank (27) described under Methods. The estimated accuracies of the coordinates, assuming random errors, are listed in Table VI for each type of atom and each temperature factor group. The values ranged from 0.02 Å through 0.47 Å. As noted before, the accuracy placed on coordinates

TABLE VI. Estimated Errors (Å)

B	C	N	O	S	Fe
6	.13	.10	.08	.04	
7	.13	.11	.09		.02
8	.14	.11	.09		
9	.15	.12	.09		
10	.15	.12	.10		
11	.16	.13	.10		
12	.17	.13	.11		
13	.17	.14	.11		
14	.18	.14	.12		
16	.19	.15	.13		
18	.21	.17	.14		
20	.23	.18	.15		
24	.27	.21	.17		
28	.31	.24	.20		
32	.36	.28	.23	.13	
36	.41		.27		
40	.47	.37	.31		

is very dependent on the temperature factor. Carbon atoms with temperature factors from 30 Å^2 to 40 Å^2 had estimated errors from 0.36 Å to 0.47 Å. The values in Table VI in general, and the high temperature factor errors in particular, may be underestimated. Although the assumption concerning the estimated errors in the structure factors (27) was an overestimate, only the observed reflections ($F_{obs}>2\sigma$) were included in the calculation. This and the systematic errors probably led to an underestimate of coordinate errors. Calculation of estimated errors in the electron density (27) resulted in a value of 0.081 e/Å^3.

The final refined $\underline{c}_{551}$ structure was constrained to have an average bond distance deviation of 0.042 Å from standard values. The average deviations from idealized bond angles and torsion angles were 2.5° and 4.6°, respectively. Figure 18a,b shows the distribution of bond length deviations from idealized values for specific bonds. The main chain N-C_α distribution in Figure 18a has an average deviation of 0.039 Å and a root mean square deviation of 0.051 Å. The distribution of deviations for all main chain bond lengths in Figure 18b has average and root mean square values of 0.038 Å and 0.050 Å. Figure 18c shows similar deviations for all C_α-C_β-C_γ bond angles, and have average and root mean square deviations of 2.8° and 3.7°. Main chain torsion angle distributions are presented in Figure 18d with larger values of 6.3° and 8.0° for average and root mean square deviations.

Allowing the bonding parameters to be fairly flexible permitted the structure to obtain a good fit to the electron density. At certain

stages during refinement, it was advantageous to not constrain the structure at all between shifting cycles. A limited number of these cycles without constraints resulted in a significantly improved fit when the constraints were eventually applied. For the final constraints pass, the number of constraining cycles and the weights were chosen to result in a structure slightly more constrained than usual. Based on high resolution crystal structures of small molecules and peptides, the average bond distance deviation of 0.042 Å is probably only slightly larger than one would expect from the true structure. The bond angles and torsion angles are very well constrained. The standard deviations of the coordinates and the local quality of the electron density should always be taken into account when attaching significance to calculated distances and angles (27).

Hydrogen Bonding

Intermolecular main chain to main chain hydrogen bonding in the refined structure was examined by calculating all nitrogen to nitrogen, nitrogen to oxygen, and oxygen to oxygen distances up to 3.6 Å, exclusive of those within the same residue or to adjacent residues. Within this list of 129 contacts, there were only four oxygen to oxygen contacts and no nitrogen to nitrogen contacts. For those 125 contacts, hydrogen bonding parameters were calculated. The parameters included the O - N distance, the hydrogen bond angle O - H - N, the acceptor angle C - O - N, and the donor angle C - N - O. The

hydrogen coordinates were calculated from the C, N, and C_α coordinates and standard bonding geometry (47). These data were examined for a correlation between reasonable hydrogen bonds in the protein structure and the calculated parameters. Reasonable hydrogen bonds had at least one calculated angle greater than 130°. In fact, there were only three exceptions to this correlation, and these involve residues where the hydrogen bond determination was borderline. A summary of the acceptable hydrogen bonds and the rejected ones is given in Tables VII and VIII. Hydrogen bonds that were separated by large distances in terms of sequence but which were important in terms of the protein conformation are noted. The statistics for hydrogen bonding, based on the 3.6 Å maximum distance and an angle greater than 130°, resulted in an average bond length of 2.98 Å and the bond angles shown in Table IX. The donor and acceptor angle averages are comparable to the values 123° and 140° found in methemoglobin (48). Hydrogen bonds between side chains and from side chains to the main chain are listed in Table X for distances up to 3.6 Å. The average length for these bonds was 3.09 Å, with a minimum of 2.67 Å and a maximum of 3.48 Å.

Heme Structure

The c_{551} MIR heme electron density is displayed in Figure 19 for the planes parallel and perpendicular to the heme. Refinement was started with these coordinates. Figure 12b indicates that the average shift

TABLE VII. Intramolecular Main Chain Hydrogen Bonds

Hydrogen Bond Acceptor	Donor	θ^1_A	θ^2_D	θ^3_H	Distance (Å)	
D 20	L 6N	149.5	117.5	160.5	2.78	
P 30	F 7N	155.1	126.6	162.1	2.63	
V 50	N 9N	166.6	125.2	172.1	2.95	
L 60	K 10N	149.6	127.7	176.1	2.80	
F 70	C 12N	158.7	119.8	109.1	3.32	
F 70	V 13N	149.4	114.6	143.3	3.23	
K 80	G 11N	108.9	138.7	151.5	3.18	
G 110	A 14N	116.3	133.1	170.8	2.99	
C 120	C 15N	127.8	128.8	172.7	2.95	
C 120	H 16N	162.6	113.1	122.7	3.02	
C 150	G 24N	156.4	121.3	170.8	2.78	*
A 170	Y 27N	149.9	107.6	152.7	3.02	*
I 180	K 28N	149.5	132.1	168.6	2.81	*
G 240	P 22N	162.8	122.5	158.2	3.31	*
A 260	V 30N	160.4	119.3	156.1	3.07	
Y 270	A 31N	165.9	122.5	171.2	2.80	
K 280	A 32N	150.5	125.0	156.0	2.96	
D 290	K 33N	144.9	121.0	166.6	3.09	
V 300	F 34N	165.3	118.5	162.6	2.95	
A 310	A 35N	144.1	122.2	122.3	2.92	
A 320	A 35N	108.3	143.8	130.6	3.42	
F 340	Q 37N	145.4	123.3	173.5	2.92	
Q 370	A 40N	135.8	130.3	158.0	3.31	
G 390	E 43N	154.7	122.0	151.8	3.18	
A 400	L 44N	157.7	127.5	169.4	2.59	
E 410	A 45N	157.7	127.4	166.2	2.84	
A 420	Q 46N	142.3	117.9	154.1	3.18	
E 430	R 47N	153.3	116.4	148.4	3.26	
L 440	I 48N	164.2	124.1	167.6	2.61	
A 450	K 49N	157.2	114.5	157.7	3.11	
Q 460	N 50N	160.4	131.7	169.0	2.79	
R 470	G 51N	134.7	142.6	140.2	2.88	
K 490	N 64N	124.6	108.4	148.9	3.17	*
O 610	S 52N	143.0	107.9	154.7	2.92	*
S 520	O 61N	163.0	113.6	142.1	2.65	*
G 570	G 54N	174.6	154.4	138.8	3.10	
N 640	V 66N	78.4	92.9	134.7	3.09	
S 670	A 71N	176.8	119.8	170.8	2.91	
D 680	Q 72N	148.5	121.3	160.3	3.21	
D 690	T 73N	146.6	123.0	163.7	3.19	
E 700	L 74N	159.1	119.9	154.2	2.85	
A 710	A 75N	154.6	126.2	163.7	2.73	

Q 72O	K 76N	148.5	118.6	155.9	3.01
T 73O	W 77N	144.8	118.0	148.4	3.26
L 74O	V 78N	166.1	119.3	170.9	2.90
A 75O	L 79N	161.4	123.8	169.2	3.03
K 76O	S 80N	143.3	104.4	109.6	3.23
W 77O	S 80N	110.0	136.2	156.7	2.88
V 78O	Q 81N	128.0	129.4	173.9	2.95

---

[1] Acceptor angle C-O-N.

[2] Donor angle C-N-O.

[3] Hydrogen bond angle O-H-N.

* Hydrogen bonds which are separated by large distances in the sequence.

TABLE VIII. Rejected Intramolecular Main Chain Hydrogen Bonds

Hydrogen Bond Acceptor	Donor	θ^1_A	θ^2_D	θ^3_H	Distance (Å)
D 20	E 4N	68.9	70.8	79.5	3.32
D 20	V 5N	109.1	129.5	129.6	2.98
P 30	V 5N	71.4	60.1	72.1	3.34
P 30	L 6N	97.0	112.1	97.0	3.21
V 50	F 7N	76.1	55.6	69.6	3.36
V 50	K 8N	115.0	125.1	127.5	3.26
L 60	K 8N	75.9	65.7	71.6	3.15
L 60	N 9N	102.1	97.4	91.5	3.31
K 80	K 10N	71.6	57.6	71.3	3.47
N 90	G 11N	73.0	78.7	100.7	3.24
K 100	C 12N	73.8	76.3	102.3	3.47
G 110	V 13N	70.6	62.6	72.3	3.16
C 120	A 14N	77.6	78.6	92.0	2.90
V 130	C 15N	75.6	67.1	89.2	3.18
C 150	V 23N	117.5	126.2	127.7	3.14
A 170	D 19N	78.0	54.0	72.3	3.56
I 180	Y 27N	102.9	111.5	106.8	3.46
P 250	Y 27N	100.4	49.0	67.1	3.56
A 260	K 28N	78.7	60.6	70.6	3.42
A 260	D 29N	116.3	122.7	117.9	3.30
Y 270	D 29N	77.4	59.5	68.4	3.25
Y 270	V 30N	108.2	117.0	109.9	3.09
K 280	V 30N	70.7	57.5	71.8	3.49
K 280	A 31N	101.7	109.1	98.9	3.54
D 290	A 31N	69.9	46.4	62.3	3.56
D 290	A 32N	93.4	115.4	107.9	3.39
V 300	A 32N	74.2	57.9	69.1	3.24
V 300	K 33N	110.3	116.1	111.0	3.21
A 310	K 33N	71.1	56.3	67.1	3.46
A 310	F 34N	99.6	104.0	109.4	3.41
A 320	F 34N	79.8	48.9	71.4	3.52
K 330	A 35N	68.1	85.9	88.8	3.09
F 340	G 36N	95.2	68.6	63.8	3.24
G 360	A 38N	92.0	64.9	86.2	3.59
Q 370	G 39N	103.9	56.8	72.9	3.33
A 380	A 40N	71.6	75.0	98.5	3.38
G 390	E 41N	79.4	59.5	75.6	3.47
G 390	A 42N	116.8	119.1	114.1	3.54
A 400	A 42N	72.3	56.5	65.8	3.36
A 400	E 43N	101.0	118.8	111.6	3.07
E 410	E 43N	73.8	59.7	72.2	3.33
E 410	L 44N	104.9	110.8	102.9	3.35

A 42C	A 45N	93.4	114.9	106.2	3.51
E 43C	A 45N	71.7	55.6	69.1	3.36
E 43C	Q 46N	108.5	117.4	117.5	3.25
L 44C	Q 46N	75.1	57.7	67.7	3.27
L 44C	R 47N	111.1	123.4	121.4	3.01
A 45C	R 47N	74.6	63.5	77.9	3.37
A 45C	I 48N	101.9	119.6	105.5	3.24
Q 46C	I 48N	70.2	57.0	66.8	3.43
Q 46C	K 49N	107.0	114.2	112.9	3.40
R 47C	K 49N	76.0	54.4	68.7	3.46
Q 53C	V 55N	77.9	79.3	110.7	3.44
G 54C	W 56N	74.8	51.6	71.0	3.60
G 57C	I 59N	67.8	70.1	96.2	3.58
P 62C	N 64N	103.7	52.6	64.1	3.56
S 67C	D 69N	81.1	54.4	67.0	3.09
S 67C	E 70N	119.5	114.5	108.4	3.03
D 68C	E 70N	70.0	56.5	69.9	3.40
D 68C	A 71N	99.2	112.1	109.4	3.49
D 69C	C 72N	100.7	120.4	117.8	3.49
E 70C	Q 72N	73.4	57.7	69.5	3.28
E 70C	T 73N	106.0	112.3	104.6	3.13
A 71C	T 73N	70.0	51.0	65.6	3.43
A 71C	L 74N	100.5	117.3	114.0	3.22
Q 72C	L 74N	67.2	58.6	73.8	3.43
Q 72C	A 75N	96.8	116.3	108.7	3.37
T 73C	K 76N	98.2	116.4	116.9	3.48
L 74C	K 76N	74.7	55.6	68.6	3.38
L 74C	W 77N	110.9	118.7	120.0	3.19
A 75C	W 77N	78.9	60.3	74.1	3.27
A 75C	V 78N	110.1	110.9	104.2	3.34
K 76C	L 79N	101.4	119.0	107.9	3.44
W 77C	L 79N	71.9	56.1	68.1	3.38
V 78C	S 80N	82.5	73.5	89.4	2.85
L 79C	Q 81N	70.0	73.2	92.3	3.16

---

[1] Acceptor angle C-O-N.

[2] Donor angle C-N-O.

[3] Hydrogen bond angle O-H-N.

TABLE IX

Intramolecular Main Chain Hydrogen Bonding Statistics

	Average	Root Mean Square	Minimum	Maximum
Distance (Å)	2.98	3.00	2.59	3.42
Hydrogen Bond Angle	155.6°	156.5°	109.1°	176.1°
Acceptor Angle	147.7°	148.6°	78.4°	166.6°
Donor Angle	122.8°	123.2°	92.9°	154.4°

TABLE X. Intramolecular Side Chain Hydrogen Bonds

Hydrogen Bonds		Distance (Å)
D 20D2	E 4N	2.67
D 20D2	V 5N	3.39
K 10NZ	A 65O	3.05
K 10NZ	E 70OE1	3.42
H 16ND1	P 25O	2.85
K 33NZ	V 55O	3.20
E 41OE2	Q 72NE2	2.80
Q 46OE1	N 50ND2	3.24
I 48O	N 64ND2	2.75
K 49NZ	V 66O	2.74
Q 53OE1	P 58O	3.27
S 67N	E 70OE1	2.88
S 67OG	D 69OD1	3.48
S 67OG	E 70N	3.23
S 67OG	E 70OE1	3.15
D 69O	T 73OG1	3.17
W 77NE1	Q 81OE1	3.31
W 77O	S 80OG	2.96
W 77O	Q 81OE1	3.16

in coordinates during refinement was comparatively small (1/2 Å). The relatively low average temperature factor of approximately 10 $Å^2$ was roughly constant during refinement and indicated that the heme density was well defined.

The difference map electron density around the heme was at the noise level and except for one clear feature showed no systematic distribution. With the heme iron atom constrained to lie within the heme plane, the difference map at the iron position had large adjacent peaks of positive and negative electron density. This unambiguous difference map density indicated the iron position was incorrectly located and should be shifted. It was the largest feature in the entire map, and was consistently present in the same position in all maps with a planar iron atom. The direction of the indicated shift was toward the histidine side of the heme.

To illustrate the heme structure, difference and $2F_{obs}-F_{calc}$ electron density maps are displayed in Figures 20 through 22 at various points in the refinement. Up to minimap IV, the iron position was constrained to lie within the heme plane. The difference in R-factor between an in-plane and an unconstrained iron was 0.6%. Figures 20a and 20b present typical difference maps, calculated from an R = 19.5% structure factor set. The iron position was constrained at the heme center. These maps clearly indicated the iron should be shifted. The indicated direction of movement was perpendicular to the heme plane toward His 16 and parallel to the plane toward the heme nitrogen 2. This observation was not an artifact due to biased

phases since the iron position had been constrained to be at the heme center up until this point. The shift information, whether real or not, is in the F_{obs} or some aspect of the model other than bias from a previously uncentered iron position. The electron density calculated on the same planes, but for the $2F_{obs}-F_{calc}$ electron density maps are also shown in Figures 20c and 20d.

At this stage, it was thought that the difference electron density map was good enough to permit an atom as heavy as an iron to refine without constraints. All bond distance, bond angle, and torsion angle constrains were removed from the iron atom. This resulted in the very clean difference map shown in Figures 21a and 21b at an R-factor of 16.2%. The iron position was approximately 0.23 Å from the heme center, with the out-of-plane component being 0.14 Å toward the histidine. The in-plane shift was entirely along the 2N to 4N direction. This resulted in the iron to 2N bond distance being 0.18 Å short of the normal value. It appears from the $2F_{obs}-F_{calc}$ maps in Figures 21c and 21d that the peak electron density is at the heme center and the iron is being shifted off center due to an asymmetrical distribution of electron density. That is, the electron density weighted center appears to be difference from the peak position. However, this is not the case, since a three-dimensional map does not have the peak located at the heme center.

A final constraints cycle was run on these coordinates with the iron to heme nitrogen bond distance included in the constraints. This eliminated the in-plane component of the iron displacement, as

seen in Figure 22. The iron to heme nitrogen distances became 1.98 Å, 1.98 Å, 2.00 Å, and 2.05 Å for the nitrogens 1 to 4. The R-factor increased from 16.2% to 16.6%. As expected, this constraining movement resulted in a recurrence of the difference map density. Since only the in-plane iron shift component had been constrained, the difference map gradient was entirely in the heme plane.

As mentioned above, the $2F_{obs}$-F_{calc} heme plane sections indicated the electron density peaks were at the heme center. For example, in Figures 21c and 21d, it appears that the $2F_{obs}$-F_{calc} peak and iron positions are significantly different. The electron density peak appears to be very close to the heme center. However, a three-dimensional $2F_{obs}$-F_{calc} map has the iron electron density only 0.03 Å away from the unconstrained iron position. The $2F_{obs}$-F_{calc} peak and coordinates obtained from the difference map are essentially identical.

After refinement to an R-factor of 17.3%, the possibility of a nonplanar heme was examined. This possibility was more likely in view of the surprising result described above for the heme iron position. Refined high-spin protein prophyrins, with an out-of-plane iron atom, have been found to have nonplanar hemes (49,50). Prophyrins in general are quite flexible and are influenced by their surrounding environment (51).

At a 17.3% R-factor, all heme constraints were removed except for bond distances. The R-factor refined down to 15.8%, with the protein being constrained as usual. Figure 23a shows a difference map after refinement under these conditions. The heme is also shown

in Figure 23b after application of bond angle and torsion angle constraints and at an R-factor of 16%. This heme had much more realistic geometry, but the difference map contained more noise. After studying Figure 23 and previously calculated unconstrained heme difference maps, it was concluded that at 2.0 Å resolution non-planarity of the size suggested by the difference maps could not be determined accurately. Although there might be some puckering of the heme, refinement at higher resolution with a larger ratio of observations to parameters would be necessary to prove any heme non-planarity. From this point on, constraints were again applied to planarize the heme atoms, with the exception of the iron atom which continued to remain unconstrained.

The final $2F_{obs}-F_{calc}$ electron density map and coordinates for the heme at an R-Factor of 16.2% are shown in Figure 24. The electron density at the iron center was 6.7 e/Å^3. Some bond distances for this structure and the structure with constrained heme nitrogen to iron bond distance are given in Table XI.

TABLE XI. Heme Bond Distances

	16.2% Coordinates		16.6% Coordinates*	
	Fe	Center	Fe	Center
1 N	1.97	2.01	1.98	2.02
2 N	1.82	1.99	1.98	2.03
3 N	2.01	1.95	2.00	1.93
4 N	2.17	1.99	2.05	1.97
NE2 (His 16)	1.90	2.05	1.90	2.11
SD (Met 61)	2.39	2.25	2.40	2.19

* The angle between the histidine plane normal vector projected onto the heme plane, and the vector bisecting the 1 N and 2 N bonds is 89.8°. The heme center to ε-nitrogen atom of His 16 is 3.7° from the heme normal vector.

Solvent Structure

Protein-solvent interactions are recognized as being important in the structure and biological function of protein and enzymes. The solvent is usually classified into bulk solvent and bound solvent, with the distinction between the two being inexact and very dependent on the method employed in the determination. The x-ray crystallographic method is capable of locating ordered solvent, which may be in either classification but is usually bound very closely to the protein. This ordered solvent may be only a fraction of the bound solvent found by other methods. However, it contains detailed positional, occupancy, and thermal information that is not generally obtainable by other methods.

Improved phases from the refinement of a protein crystal structure have the distinct advantage of being able to obtain solvent structure information with an accuracy unobtainable from the unrefined MIR crystal structure. Some solvent could be located in the MIR map of cytochrome $\underline{c}_{551}$. However, many of the possible solvent electron density peaks in the MIR map were eliminated during the refinement and must have been a result of phase errors. Additional solvents were added during refinement that were not in the MIR-phased map.

The solvent was included fairly conservatively during the refinement of cytochrome $\underline{c}_{551}$ in order to obtain only genuine solvent and not solvent which was simply fit to errors in the difference electron density map. Each time a solvent was considered for inclusion into

the structure, it was evaluated by the following criteria: (a) the difference map electron density magnitude, (b) the shape of the peak, and (c) its location relative to the protein in order to prevent insertion of solvent into positive density that is really density arising from protein coordinate errors. The types of protein atoms within the vicinity of the solvent were not taken into account. After refinement, this can be used to evaluate the probability that the refined solvent is genuine and evaluate the overall ability to recognize solvent in difference maps at this resolution and stage of refinement. Even with the above precautions, approximately half of the total solvent included during the refinement was eventually discarded. At the time this incorrect solvent was included, it satisfied the above criteria, but further refinement found it unsuitable because of exorbitant temperature factors and negative difference map electron density at the solvent site. A final set of 42 solvents have been included, all with full oxygen scattering factors and refined temperature factors which can be used to approximate occupancy and order.

Forty-two solvent molecules represents 0.09 gm of solvent per gm of unsolvated protein and is only an extremely rough estimate. It was calculated assuming all ordered solvents were H_2O with an occupancy of 1.0. This relatively low hydration value, compared to values measured by other methods for other globular proteins (52,53) probably represents only a part of the hydration. In general, hydration from x-ray crystallographic studies is lower than those obtained

from hydrodynamic, NMR, isopiestic, or calorimetric methods (53). The rest of the solvent may not be seen due to the lack of resolution, errors in intensities or phases, or because solvation typically measured by other methods includes disordered solvent.

The refined solvent in the $\underline{c}_{551}$ crystal structure is listed in Table XII, along with the protein atoms within the specified ranges. All solvent protein distances of less than 3.0 Å are compiled in Column 1. It is interesting to note that all protein atoms within this list are polar in character and capable of hydrogen bonding, with only one exception. One half of the 42 solvents are within 3.0 Å of these polar protein atoms. Table XII also summarizes the hydrogen-bondable protein atoms within a range of 3.0 to 3.5 Å of the solvent. In addition to these atoms, there are also some hydrophobic protein atoms not listed. Thirty-four out of forty-two solvents are included in these two classifications and are within 3.5 Å of the protein. The remainder of the solvent, with protein contacts up to 5.0 Å, is listed in Table XII.

A stereo view of the location of the solvent is shown in Figure 24a and 24b. Most of the ordered solvent was found to be within 3.5 Å of the protein and near hydrophilic protein atoms. A majority of that solvent was on the molecular surface, with protein main chain carbonyl oxygens and amide nitrogens being very important binding sites. Additional external sites were near charged and polar amino acid side chains. Binding of solvent here did not seem as prevalent, and this

TABLE XII. Protein-Solvent Interactions*

Solvent Number and Temperature Factor		All Protein Atoms within 3.0 Å of Solvent		Protein Atoms Capable of H-Bonding within 3.5 Å of Solvent		Protein Atoms within 3.5 Å<d<5.0 Å Range for Solvent with no Contact <3.5 Å
1	18	K490 N640	2.86 2.70			
2	8	J0402 I590 G57C	2.85 2.80 2.98	G570 I59N	3.07 3.32	
3	9	A17N T200 M22N G240	2.84 2.72 2.92 2.88	P250	3.15	
4	14	--		--		--
5	16	A26N D290D1	2.81 2.83			
6	13	T730G1	2.81	E1N D690D2	3.22 3.49	
7	28			D690 T730G1	3.14 3.38	
8	16	N500	2.85	A38N**	3.03	
9	32			Q530E1	3.14	

Table XII (continued)

10	18	M610	2.75			
11	11	K100 N640D1 A65N	3.00 2.86 2.94			
12	16			D68N	3.07	
13	18	E10E1 D690D1	2.78 2.53			
14	24			E430E2	3.10	
15	16					Q720E1
16	28			A380	3.01	
17	24	--		--		--
18	28	V550	2.85	D69N**	3.14	
19	20	E10E2	2.67			
20	18					K21NZ P62CB
21	32					K49NZ P58CD
22	14	V230 P630**	2.99 2.80			

Table XII (continued)

23	20	P620 N640D1	2.66 2.83			
24	36			S52N	3.14	
25	36	--		--		--
26	32			N500	3.04	
27	28	G510	2.44			
28	24	W560 S670G**	2.47 2.62	D690D1**	3.22	
29	36			Q460E1	3.17	
30	32	V550	2.69	G540	3.40	
31	20	E430E2	2.32			
32	32			A380	3.24	
33	28	K28NZ	2.46	E1N** E1D**	3.36 3.49	
34	32	--		--		--
35	32			D290D2	3.30	
36	20					G39N Q53NEZ
37	24	K760	2.91			

Table XII (continued)

38	24			K49NZ V660	3.18 3.01
39	36			T730G1	3.34
40	32	K28NZ	2.93		
41	28	D290	2.69		
42	28			K21NZ W560	3.40 3.47

---

* Standard one-letter amino acid codes are used; the heme denoted by the code J.

** Symmetry-related solvent.

may reflect the fact that a good fraction of these side chains in $\underline{c}_{551}$ were disordered to some extent, as discussed previously.

There were several internal solvents that are possibly important to the $\underline{c}_{551}$ structure or function. One interesting example, solvent 3, was tetra-coordinated to two amide nitrogens and two carbonyl oxygens of the backbone loop from His 16 to Pro 25. All distances were within 3.0 Å. Solvent 2, another internal solvent, was hydrogen bonded between the extended loop containing residues Gly 57 through Ile 59 and the exposed propionic acid.

Solvents located near the heme and heme crevice had been found. These included solvents 11, 13, 20, 23, 25, 28, and 41. Solvents 11 and 23 were of particular interest. Solvent 11 was hydrogen bonded between the main chain nitrogen of Ala 65, the main chain carbonyl oxygen of Lys 10, and the delta oxygen of Asn 64. Its electron density was well defined and may play an important role in stabilizing both the polypeptide backbone and Asn 64 side chain conformations. Solvent 23 was deeply imbedded within the heme crevice and hydrogen bonded to the carbonyl oxygen of Pro 62 and the delta oxygen of Asn 64. This solvent was also quite well defined in the more refined difference maps and the final electron density map. Its existence may be important in terms of the mechanism of electron transfer to and from cytochrome $\underline{c}$. This region of cytochrome $\underline{c}$ has been implicated by a variety of experimental evidence as a possible location for electron transfer. Not only is the very edge of the heme exposed in $\underline{c}_{551}$, but it is apparent now that solvent actually extends inside the molecular

surface to a position adjacent to the heme. If this solvent is located at the actual site of electron transfer to the $\underline{c}_{551}$ heme, then this is only another example of solvent participation at the active site of biological reactions (54).

DISCUSSION

Sequence Homology

Prior to the low resolution crystal structure determination of cytochrome $\underline{c}_{551}$, the evolutionary relationship between this protein and eukaryotic cytochrome $\underline{c}$ was unclear. Although $\underline{c}_{551}$ was known to have the $\underline{c}$-type cytochrome heme attachment Cys-x-y-Cys-His and a Met near the C-terminal end, the number (82 compared to 102) and sequence of amino acids were so different from eukaryotic cytochrome $\underline{c}$ that homology alignments were difficult. Along with the overall differences observed in the $\underline{c}_{551}$ sequence, there were specific changes in the sequence that made alignments speculative because they resulted in changes to residues thought to be invariant from all other $\underline{c}$-type cytochromes.

From the low resolution structure (35), it was found that $\underline{c}_{551}$ was clearly related to eukaryotic cytochrome $\underline{c}$ despite the numerous sequence differences. The main chain fold was identical with the exceptions of one major deletion and some minor deletions and insertions. The approximate sequence alignment with $\underline{c}$ was determined from the few known side chain positions and other approximate side chain positions obtained from an analysis of the distances along the main chain from known reference points. The solution of the structure at high resolution has now determined the accurate sequence homologies with regard to side chain positions. The homology can be seen in

terms of the distances between the alpha carbons for $\underline{c}_{551}$ and tuna $\underline{c}^{III}$ in Figure 26. The distances were calculated after minimizing the deviations between 44 structurally similar alpha carbons. From this figure, it is apparent that the $\underline{c}_{551}$ backbone structure is most similar to eukaryotic cytochromes $\underline{c}$ in the regions of the N-terminal alpha helix, the heme attachment residues, the stretches of chain next to the heme (residues 21-30 and 40-51), and the C-terminal alpha helix. The sequence alignment is shown in Table XIII, along with several predicted alignments. This $\underline{c}_{551}$ comparison to the 103 residues in tuna cytochrome $\underline{c}$ shows that there were only sixteen homology sites where the residue remained unchanged. The predictions of Needleman and Blair (55) were the farthest from being correct, with only 32 and 29 correct residue alignments out of 82. Dickerson (56) with 58 and McLachlan (57) with 74 and 69 matches were closer to being correct. However, none of the predictions correctly placed the major nineteen-residue deletion in the bottom of the molecule.

The key to the Needleman and Blair alignment was the equivalence of Trp 59(H)* and Trp 56(P), based on the assumed invariance of a hydrogen-bonded tryptophan to the buried heme propionic acid. Although they are both hydrogen bonded as assumed, the Trp comes from an entirely different structural position in the molecule. The

---

* The letter in parentheses is used to denote the species of cytochrome. H = horse, T = tuna, P = Pseudomonas.

TABLE XIII. Observed and Predicted c_{551} Homologies with Tuna Cytochrome c

	10	20	30
Tuna c	G D V A K G K K - T F V Q K - C A Q C H T V E N G G K H K V G P - N - - L W G L F		
c_{551} [1]	- - - E D P E V - L F K N K G C V A C H A I D T - - - K M V G P - A - - Y K D V A		
N.B. I [2]	E D P E - V - - - L F K N K G C V A C H A I D - - - T K M V G P A Y K D V A A K F		
N.B. II [2]			P - A - - Y K D V A
D [3]	E - D P E - - - - V L F K N K G C V A C H A I D - - - T K M V G P - - - - - - - - -		
M. I [4]	- - - E D P E V - L F K N K G C V A C H A I D - - - T K M V G P - A - - Y K D V A		
M. II [4]			P - - - - - - - - -

[1] Sequence homology determed from the high resolution crystal structure.

[2] Predicted sequence homology of Needleman and Blair (55).

[3] Predicted sequence homology of Dickerson (56).

[4] Predicted sequence homology of McLachlan (57).

TABLE XIII (continued)

	40	50	60	70
Tuna c	G R K T G Q A E G Y S Y T D A - N - K S K G - - I V W N N D T L M E Y L E N P K K			
c_{551}	A K F A - - - - - - - - - - - - - - - - - G - - Q A G A E A E L A Q R I K N G S Q			
N.B. I	A G Q A G A E A E L A Q R I K - N G S Q - G - - V - W G - - - - - - - - - - P - -			
N.B. II	A K F A G Q A G A E A E L A Q - R - I K N G S Q G V W G			
D	- - - - - - - - - - A Y K D V - A A K F A G - - Q A G A E A E L A Q R I K N G S Q			
M. I	A K F A G Q A G A - E A E L A Q R I K N G S Q			
M. II	- - - - - - - - - - A Y K D V A A - K F A G - - Q A G A E			

TABLE XIII (continued)

	80	90	100 103
Tuna c	Y - - - - I - P G T K M I F A G I K K K G E - R Q D L V A - Y - L - - K S A T S		
c_{551}	G - - - - V W G P I P M P P N A V - S D D E - A Q T L A K - W - V - - L S Q K		
N.B. I	- - - - - I - P - - - M P P N A V - S D D E A Q - T L - A K W V L S Q K - - -		
N.B. II			
D	G V W G P I - P - - - M P P N A V S D D - E - A Q T L A K - W - V - - L S Q K		
M. I	G - - - - V W G P I P M P P N - A V S D D E - A Q T L A K - W - V - - L S Q K		
M. II			

alignment of Trp 56(P) and Trp 59(H) in linear sequence was incorrect, although they were functionally and structurally (in three dimensions) homologous.

Dickerson correctly recognized that the sequence homology of W56(P) and W59(T) was in error and that W56(P) was part of an inserted loop of $\underline{c}_{551}$ which had no sequence homology to eukaryotic $\underline{c}$. This sequence prediction used as a basis the crystal structures of cytochromes $\underline{c}$ and $\underline{c}_2$ and the equivalence of Tyr 48(T) and Tyr 27(P). Although it was a much better fit than the Needleman and Blair alignment, its significant fault was at the start of the major deletion at the bottom of the molecule. Because of the assumed equivalence of Tyr 48(T) and Tyr 27(P), the deletion was started at Asn 31 (the molecular "front") instead of Gly 41 (the "back" of the molecule). Tyr 48(T) and Tyr 27(P) are unrelated and a residue equivalent to Tyr 48(T) is not present in $\underline{c}_{551}$.

McLachlan's procedure was based on substitution compatabilities of amino acids observed in proteins with known homology. A computer search compared all pairs of segments in the two proteins to be tested and calculated statistics which indicated the probabilities that the observed correlation could have occurred by chance. This method proved to be the most accurate and did not rely on any tertiary structural information for the unknown protein.

It was also of interest to note that the number of exact correlations between a given sequence and the predicted homology were not a good measure of the alignment accuracy. From Table XIII, there

there were sixteen exact matches of the real $\underline{c}_{551}$ alignment to the tuna cytochrome $\underline{c}$. The Needleman and Blair predictions, although the least accurate, had the greatest number of exact sequence matches with 24 and 25. This is more than were found experimentally. The Dickerson prediction had 21 matches, and the most accurate predictions of McLachlan had only eighteen and nineteen matches. It was apparent that this criterion was not suitable for assessing the reliability that should be placed in a prediction. It was also apparent from these results that an alignment based on an assumed single residue functional homology is likely to be incorrect and should be used only with extreme caution. If this particular case can be extrapolated to predictions of homologies of other proteins, it seems the best criterion to use is the overall fit of similar residues. The difficulty is to know which residues are similar. The calculations of McLachlan from substitutions seen in known homologous proteins appear to be moderately successful.

Despite finding the overall folding of cytochrome $\underline{c}_{551}$ to be very similar to other $\underline{c}$-type cytochromes, differences in sequence were found at positions previously thought to be invariant. A comparison of $\underline{c}_{551}$ with other cytochromes, especially in terms of invariant residues and the heme environment, has already been discussed (58; Appendix II). However, a few additional points will be mentioned here.

Prior to the structure determination of bacterial cytochromes, a great many eukaryotic cytochrome $\underline{c}$ invariant residues were known to

exist. Dickerson and Timkovich (59) summarized the sequences of 67 eukaryotic cytochromes c and found only the 28 totally conserved residues in Table XIV. With the crystal structures of cytochromes c_2 and c_{550} and the sequences of three more c_2 cytochromes, this list of 28 invariant residues was markedly shortened. Based on the 67 eukaryotic c's, four c_2's, and c_{550}, Dickerson and Timkovich listed a new set of twelve totally invariant residues. Cytochrome c_{551}, being a large step removed from these c, c_2, and c_{550} cytochromes in terms of sequence similarities, reduced this number even more dramatically. Of particular importance was the discovery that Tyr 48(T) and Phe 82(T) are no longer invariant. In addition, many residues that were thought to be important structurally or functionally because of their semi-invariance have been deleted or radically changed in the c_{551} protein. With the newer sequences in Table XV from the summary of Dickerson (60), the number of invariant residues is further cut to only a few. Based on previously published sequences and the data in Table XV, only Cys 17, His 18, and Met 80 are totally conserved. It may be only a matter of time until this list is shortened further, although finding substitutions at the axial heme positions does seem unlikely at this time.

Although the number of totally invariant side chains is very small, there are many important semi-invariant features in cytochromes. This results from the similarity in the main chain folding and conservative substitutions of side chains. A critical function could be satisfied by one of several residues, with each of these

TABLE XIV. Totally Invariant Residues

67 Eukaryotic Cytochromes	+ c_{550} and 4 c_2's	+ c_{551}	+ Sequences from Table XV
Gly 1			
Gly 6	Gly 6		
Phe 10			
Gln 16			
Cys 17	Cys 17	Cys 17	Cys 17
His 18	His 18	His 18	His 18
Gly 29	Gly 29	Gly 29	
Pro 30	Pro 30	Pro 30	
Leu 32	Leu 32		
Gly 34	Gly 34		
Arg 38			
Gly 41			
Tyr 48	Tyr 48		
Asn 52			
Trp 59	Trp 59		
Leu 68			
Asn 70			
Pro 71			
Lys 72			
Lys 73			
Pro 76			
Gly 77			
Thr 78			
Lys 79			
Met 80	Met 80	Met 80	Met 80
Phe 82	Phe 82		
Gly 84			
Arg 91			

TABLE XV. Prokaryotic Cytochrome c Sequences

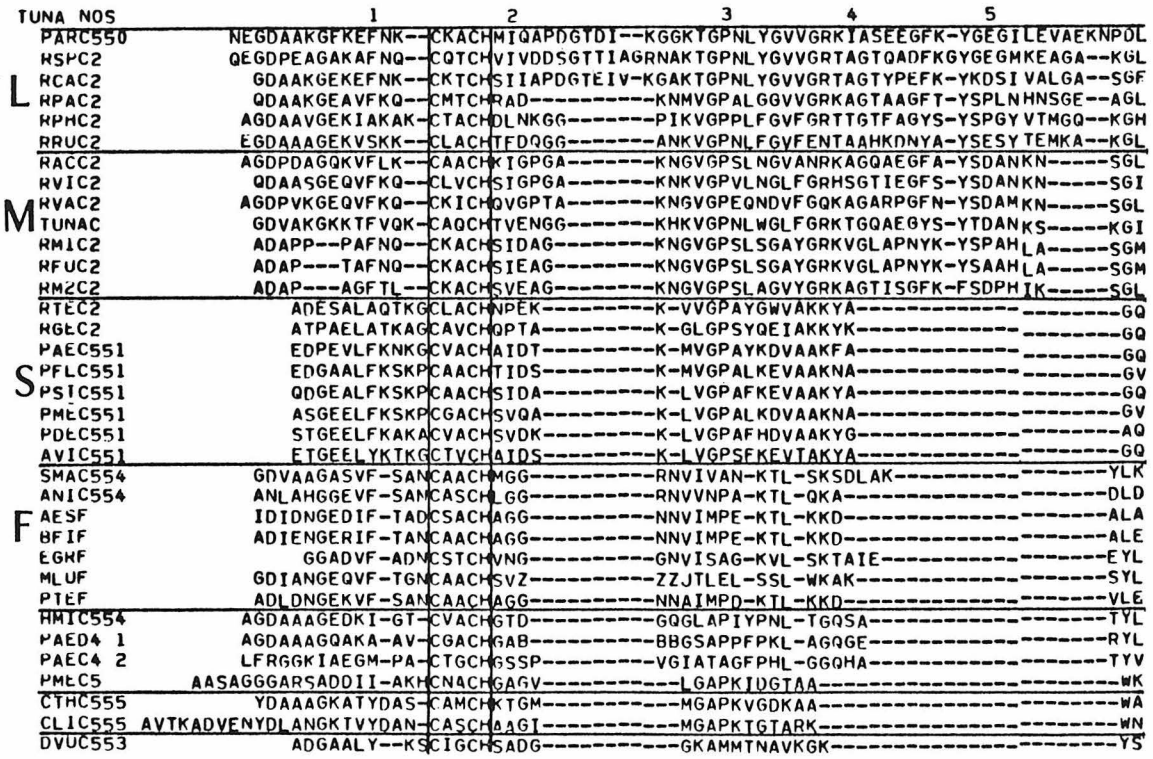


TABLE XV (continued)

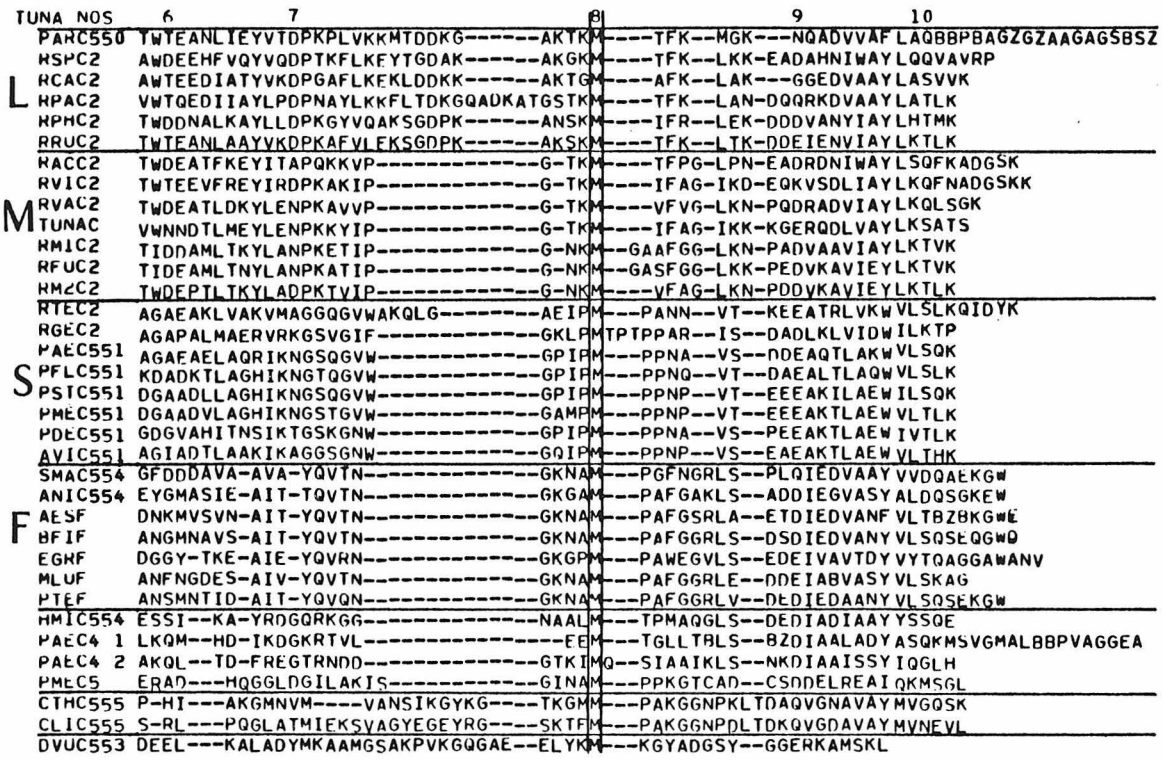


TABLE XV (continued)

PROKARYOTIC CYTOCHROME C SEQUENCE COMPARISONS

PARC550	Paracoccus denitrificans C550
RSPC2	Rhodopseudomonas spheroides C2 (ATCC 17023)
RCAC2	Rps. capsulata C2 (ATCC 23782)
RPAC2	Rps. palustris C2 (ATCC 17001)
RPHC2	Rhodospirillum photometricum C2
RRUC2	R. rubrum C2 (ATCC 11170)
RACC2	Rps. acidophila C2
RVIC2	Rps. viridis C2
RVAC2	Rhodomicrobium vannielii C2 (ATCC 17100)
TUNAC	Tuna C2
RM1C2	R. molischianum C2, iso-1
RM2C2	R. molischianum C2, iso-2
RFUC2	R. fulvum C2, iso-1
RTEC2	R. tenue C2
RGEC2	Rps. gelatinosa C2
PAEC551	Pseudomonas aeruginosa C551 (P6009)
PFLC551	Pseudomonas fluorescens C551 (ATCC 17588)
PSTC551	Pseudomonas stutzeri C551 (ATCC 17400)
PMEC551	Pseudomonas mendocina C551 (CH 110)
PDEC551	Pseudomonas denitrificans C551 (ATCC 13867)
AVIC551	Azotobacter vinelandii C551 (ATCC 12837)
SMAC554	Spirulina maxima C554 (cyanobacteria)
ANIC554	Anacystis nidulans C554 (cyanobacteria)
AESF	Alaria esculenta F (brown alga)
BFIF	Bumilleriopsis filiformis F
EGRF	Euglena gracilis F
MLUF	Monochrysis lutheri F
PTEF	Porphyra tenera F (red alga)
HMIC554	Halotolerant Micrococcus C554 (ATCC 12084)
PAED4 1	Pseudomonas aeruginosa C4, first half of chain
PAED4 2	Pseudomonas aeruginosa C4, second half of chain
PMEC5	Pseudomonas mendocina C5 (CH 110)
CTHC555	Chlorobium thiosulphatophilum C555 (NCIB 8346)
CLIC555	Chlorobium limicola C555
DVUC553	Desulfovibrio vulgaris C553 (NCIB 8303)

residues having a common structural or chemical property. This is what the success of McLachlan's predictive method is saying. Rather than specific, exact matches being the important criterion for successful alignments, high correlations between residues that are similar but not identical are more important. With this in mind, some of the common and dissimilar c-type cytochrome structural features will be discussed.

Heme Environment

The result of the iron atom refining to a position off the heme plane was quite unexpected and has not been seen in other high resolution unrefined cytochrome c structures (36-39). From the 2.4 Å resolution MIR maps of cytochrome c_{551} (Figure 19), the iron atom appears centered in the heme electron density. However, it is more difficult to assess the planarity of the iron atom from MIR maps because of the inaccuracies in fitting the lighter atoms to the electron density. The map densities themselves are also less accurate. The refined c_{551} structure clearly shows the iron position displaced from the heme center. Nonplanar iron atoms may also be found for other c-type cytochromes when refinement yields more accurate coordinates and electron densities.

The refined heme iron position was unexpected, not only because of the differences compared to other cytochromes c, but because all low-spin hemes have been predicted to have planar iron atoms (61).

This prediction has been found to be true for refined high resolution small molecule structures (62). Also, it has generally been observed for MIR low-spin iron porphyrins. CN-methemoglobin (63) (2.5 Å resolution, three-derivative structure) and Glycera dibronchiate hemoglobin (64) (2.5 Å resolution, three-derivative structure) have iron atoms at the heme center. In Erythrocruorin (65) (2.5 Å resolution with phases from the MIR met structure), the iron moves toward the heme plane in going from the met structure (high-spin) to the CO structure (low-spin). However, the shift is not entirely perpendicular to the heme. A concurrent shift in the position of the heme is necessary if the iron is to be located at the center. The only refined low-spin protein structure, other than c_{551}, is CO-myoglobin (66). At an R-factor of 37%, the neutron structure indicates the iron is out of the plane by approximately 0.1 Å.

In conclusion, although c_{551} is a low-spin heme protein (67), its iron position refined to a position away from the heme center. If this result is real, it could be a consequence of the special geometry of the protein ligands, which are not found in small molecules. This, in combination with the fact that the iron to histidine bond is stronger than the iron to methionine, could explain the results. Although it cannot be stated unequivocally that the out-of-plane position is real, the size of density produced from the difference maps when the iron was constrained to be at the heme center makes this possibility very likely. The estimated errors for heme atoms in

Table VI relative to the observed iron shift also indicate this is likely a real effect.

The alternate explanation is that the shifted iron position is not real. It could possibly result from a number of errors or misinterpretations. Small errors in the observed structure factors could produce this effect, even though the shift was not seen in the less sensitive MIR electron density. The apparent shifts could arise from errors in the lighter heme atom coordinates, but the heme plane must be well defined as a consequence of the large number of atoms it contains. Other possible sources of this unusual iron position include series termination errors and anomalous dispersion effects.

Another important feature of the cytochrome $\underline{c}_{551}$ heme structure is its orientation relative to the polypeptide backbone. As indicated previously (58), the $\underline{c}_{551}$ heme is rotated approximately 11° to the left and tilted forward 16° compared to the heme orientation in cytochromes $\underline{c}$, $\underline{c}_2$, and $\underline{c}_{550}$, and with respect to the basic cytochrome protein conformation. This change in orientation relative to such features as the alpha helices is undoubtedly a requirement in order to accommodate the substantial changes in the positions of certain protein side chains whose interactions with the heme are critical. The rotation forward and to the left is accompanied by an insertion of Gly 11 in the sequence of $\underline{c}_{551}$ relative to the sequences of cytochromes which do not have this rotation. This permits Lys 10 and the entire N-terminal alpha helix to be in nearly identical positions

in all cytochromes and at the same time allows Cys 12 in $\underline{c}_{551}$ to be extended to the new heme attachment site created by the heme rotation.

Another important feature resulting from the differences in side chain positions is the location of the buried propionic acid. In cytochromes $\underline{c}$, $\underline{c}_2$, and $\underline{c}_{550}$, this propionic acid is to the right side of the heme and hydrogen bonded to a tryptophan at the back of the molecule. In $\underline{c}_{551}$ this propionic is hydrogen bonded to Trp 56 at the bottom of the molecule. This dictates the placement of the buried propionic acid to be to the left of the heme plane, and on the same side as the exposed propionic acid.

Hydrogen bonding to the $\underline{c}_{551}$ heme is present to a lesser extent than in the other cytochromes which contain the bottom twenty residues. With respect to the buried propionic acid, tuna $\underline{c}$ is hydrogen bonded to Trp 59, Tyr 48, Asn 52, and the amide nitrogen of Gly 41 (36). In $\underline{c}_{551}$, the buried propionic is only hydrogen bonded to Trp 56 and Arg 47. Even the Arg 47 hydrogen bond is very doubtful. In one of its possible positions, this side chain can bond to the propionic. But its electron density is not well defined in either the MIR or refined map. Its position may also be in either a more solvated alternate position or completely disordered. If it were hydrogen bonded, its density would be expected to be well defined. With respect to the exposed propionic acid, tuna cytochrome $\underline{c}$ is hydrogen bonded to Thr 49, Thr 78, and the carbonyl oxygen of Lys 79. In $\underline{c}_{551}$, this propionic is hydrogen bonded only to Ser 52 and solvent.

An additional difference in the structure of $\underline{c}_{551}$ relative to the other cytochrome structures is the position of the ε-methyl from the heme ligand Met 61. In $\underline{c}$, $\underline{c}_2$, and $\underline{c}_{550}$, this methyl is extended approximately parallel to the iron-1N bond. (1N is the nitrogen in pyrrole ring 1, which is attached to the first Cys heme attachment.) In $\underline{c}_{551}$, this methyl is directed approximately 90° from this direction toward 4. (Ring 4 is attached to the buried propionic acid.) The difference between these two conformations is a consequence of a different set of sulfur lone pair electrons from the axial Met ligand coordinating to the heme iron. The important question is why this difference exists in view of the fact that the methionine side chain approaches the heme iron in a similar conformation. The answer is that tyrosine 67 in tuna, or the equivalent residues in $\underline{c}_2$ and $\underline{c}_{550}$, prevents the ε-methyl from assuming the $\underline{c}_{551}$ conformation for steric reasons. The hydroxyl group of this tyrosine is directed toward the expected position of the lone pair on the delta sulfur of the methionine. This would not allow enough room for the methyl to be placed there. In contrast, $\underline{c}_{551}$ has no such residue in this position and either conformation is as likely from this particular steric point of view. The fact that it does adopt its observed conformation must be a consequence of a more subtle energetic requirement. Despite this positional change, the ε-methyl proton NMR absorptions are very similar. For the oxidized proteins, the resonances are -24.2 for $\underline{c}$ (68), -15.2 for $\underline{c}_2$ (69), and -15.7 for $\underline{c}_{551}$ (70), in parts per million downfield from 2,2,dimethyl-2-silapentane-5-sulfonate (DSS). In the

reduced proteins, the values are -3.3, -2.9, and -2.9, respectively. It is apparent that despite the major difference between the ε-methyl of $\underline{c}_{551}$ and that of $\underline{c}$ and $\underline{c}_2$, $\underline{c}_{551}$ and $\underline{c}_2$ are more alike magnetically, and $\underline{c}$ is significantly different.

Side Chain Structural Homology

The relationship between Trp 59 in tuna cytochrome $\underline{c}$ and Trp 56 in $\underline{c}_{551}$ is very interesting. Tuna Trp 59 was deleted in $\underline{c}_{551}$ along with the molecule's bottom loop. The 70's loop in tuna $\underline{c}$ was extended in $\underline{c}_{551}$ to partially replace this loop and provide a new Trp (56) for hydrogen bonding to the buried heme propionic. This loop is also extended in $\underline{c}_2$ and $\underline{c}_{550}$, but it is not used for hydrogen bonding to the propionic acids in these proteins. Although these two Trp are functionally equivalent and are in very rough spatial equivalence, they are unrelated in terms of the linear amino acid sequence. This seems to be an example of evolutionary convergence where a tryptophan was placed in an approximately equivalent position from two independent sites in the amino acid sequence because of a functional requirement.

In fact, it appears from the sequences of the $\underline{f}$ cytochromes and $\underline{c}_{554}$ cytochromes from S. maxuma and A. nidulans in Table XV that there is even another evolutionary independent Trp that performs the same functional role. This third Trp site is at the very end of the C-terminal helix. One can easily imagine this slightly extended chain

coming around the back of the structure to hydrogen bond to the buried propionic acid. This possibility can be easily visualized from the rear stereo view of Figure 27, modified from Reference 71. The changes would involve (a) a short circuit between 36 and 60 and a deletion of the bottom loop as in $\underline{c}_{551}$, and (b) an extension from residue 104 to positions analogous to 37, 38, 39, and 40, where the terminal or near terminal Trp could easily hydrogen bond to the buried propionic acid. It is also likely that the residues 33 through 36 in Figure 27 would have to remain in their relatively high position rather than assume the lower helical conformation seen in $\underline{c}_{551}$. From the sequences in Table XV, M. lutheri cytochrome does not have this terminal Trp, but it does have a Trp fourteen residues past the axial histidine ligand, which could easily occupy a position homologous to Trp 59 in tuna $\underline{c}$.

A case similar to that of Trp 56 and 59 is seen in comparing tuna Thr 78 to $\underline{c}_{551}$ Ser 52. Again, these residues are performing the same function in hydrogen bonding to the exposed propionic acid. And a Thr or Ser appears to be present somewhere near these positions in all the sequences listed in Table XV. Therefore, although it appears to be a functionally critical residue, it is derived from a structurally different position in tuna $\underline{c}$ and $\underline{c}_{551}$. In tuna $\underline{c}$, Thr 78 is located following the 70's loop, while in $\underline{c}_{551}$ Ser 52 come prior to this loop.

Another very important feature of the $\underline{c}_{551}$ structure is the side chain packing at the left side of the heme. Just as the heme

attachment sequence at the right side of the heme is fundamental to the c-type cytochromes, so is the sequence A(1) - x - y - z - A(2) at the left of the heme. A(1) and A(2) can be the hydrophobic Leu, Val, or Ile, with only minor exceptions. x, y, and z can be any residue. A summary for this comparison is given in Table XVI, with exceptions and required shifts in alignments indicated. There are very few exceptions within this diverse list of cytochrome sequences. Phe is used in three sequences. This is a bulky hydrophobic group like the Leu, Ile, and Val residues, and is not too much of a change. Met is substituted in a few eukaryotic sequences and three prokaryotic sequences. The reason for this conservation of hydrophobic residues in this position is seen from the crystal structures. These two residues are located at helical positions n and n+4 on the 60's helix. Both residues have internal side chains packed very tightly against the heme. The particular spacing is a result of the main chain helical conformation. It will be of interest to see if this predicted homology is accurate for some of the more diverged cytochromes in Table XV.

The tyrosine residue at positions 27 in c_{551} is in a structurally homologous position to the conserved Leu 32 in c, c_2, and c_{550}. The explanation of this conservation is that it is very tightly packed against the heme. This substitution of the hydrophobic Leu 32 by a tyrosine in c_{551} with a polar hydroxyl group is quite unexpected, even more so than the proposed substitution of Phe for Leu, Ile, or Val discussed above.

TABLE XVI. Residues at Positions A(1) and A(2) in the Sequence A(1) - x - y - z - A(2)

Cytochrome[1]		Sequence Position[2] A(1)	A(2)	
Eukaryotic		Leu, Met	Leu	*
L	1	Leu	Val	
	2	Phe	Val	*
	3	Ile	Val	
	4	Ile	Leu	
	5	Leu	Leu	
	6	Leu	Val	
M	1	Phe	Ile	*
	2	Phe	Ile	*
	3	Leu	Leu	
	4	Leu	Leu	
	5	Leu	Leu	
	6	Leu	Leu	
	7	Leu	Leu	
S	1	Leu	Val	
	2	Met	Val	*
	3	Leu	Ile	
	4	Leu	Ile	
	5	Leu	Ile	
	6	Leu	Ile	
	7	Ile	Ile	
	8	Leu	Ile	
F	1	Val	Val	**
	2	Ile	Val	**
	3	Ile	Val	**
	4	Ile	Val	**
	5	Ile	Val	**
	6	Ile	Val	**

TABLE XVI (continued)

7	Ile	Val		**
1	Leu	Ile		**
2	Leu	Met	*	**
3	Val	Leu		**
4	Leu	Leu		**
1	Val	Ile		**
2	Ile	Val		**
1	Leu	Met	*	**

---

[1] Prokaryotic cytochrome c sequences from Table III.

[2] Positions A(1) and A(2) are the proposed prokaryotic homology sites with eukaryotic tuna sites 64 and 68.

* A residue other than Leu, Ile, or Val is at the site A(1) or A(2).

** A shift in the amino acid sequence in Table XV was required to make the alignment.

The aromatic residue Phe 82 in tuna has been thought to be important because of its invariance. Asn 64 in $\underline{c}_{551}$ was found to occupy the homologous position, which is quite surprising considering it is a residue very close to the heme and there is quite a large difference between the two types of side chains.

The aromatic pair of residues in the right channel of cytochrome $\underline{c}$ (Phe 10 and Tyr 97) are conserved in the structure of P. aeruginosa $\underline{c}_{551}$. However, this conservation cannot be extended to all cytochromes. The small cytochromes R. tenue $\underline{c}$ and R. gelatinosa $\underline{c}$ in Table XV have very high sequence homology with $\underline{c}_{551}$ and must also have very similar tertiary structures, However, neither of these sequences contain an aromatic residue in the first position.

The aromatic pair Phe 34 and Trp 56 in $\underline{c}_{551}$ form a large percentage of the molecular bottom and shield the heme from the solvent. A similar aromatic pair in tuna cytochrome $\underline{c}$, Trp 59 and Tyr 48, have similar orientations.

Structure and Chemical Properties

The major changes in the sequence and composition for $\underline{c}_{551}$ compared to other $\underline{c}$-type cytochromes has resulted in alterations of many of the physical, chemical, and biological properties. $\underline{c}_{551}$ is an acidic protein with an isoelectric point of 4.7, compared to 10.0 for eukaryotic $\underline{c}$. Some of the highly conserved positive Lys residues implicated in the mechanism of cytochrome $\underline{c}$ are not present in $\underline{c}_{551}$.

One apparently important Lys is conserved. This is Lys 10 in $\underline{c}_{551}$, which had the same orientation as Lys 13 in eukaryotic $\underline{c}$. It was an unusual Lys in the structure of $\underline{c}_{551}$ because despite its solvent-exposed location on the surface, its electron density was well defined and its temperature factors were low. This is probably a result of ionic interactions with Glu 70, which was another unusual ionic surface residue that was well defined. From the structures of $\underline{c}$, $\underline{c}_2$, and $\underline{c}_{550}$, it is apparent that this same type of ion pair can be formed, although the acidic residue may originate from one of several positions. Extension of this feature to all the prokaryotic sequences in Table XV cannot be made with certainty from sequence data alone. However, the possibility exists and with the solution of additional structures, this feature might be verified. The distribution of other charged residues has already been discussed (35; Appendix II).

In addition to charge difference, visible (72), proton NMR (70), and EPR (67) spectral differences have been found for $\underline{c}_{551}$. Similarities between the NMR spectra of $\underline{c}_{551}$ and $\underline{c}_2$ and the significant difference between these and $\underline{c}$ have been discussed in terms of the methionine resonances. These same characteristics are seen in the NMR resonances of other protons and in ESR spectra.

The pH-dependent conformational changes associated with cytochrome $\underline{c}$ are significantly altered for $\underline{c}_{551}$. Specifically, the transition monitored by the disappearance of the 695 nm band between pH 7 and 10 for cytochrome $\underline{c}$ is not present in $\underline{c}_{551}$ (73). This spectral

transition is associated with the replacement of Met as the sixth ligand, and a conformational change probably resulting in the ε-nitrogen of Lys 79 becoming the ligand. In $\underline{c}_{551}$, this transition does not occur, and only after pH 11 is the Met ligand replaced. There are two possible explanations for this observation in terms of the structure of $\underline{c}_{551}$. First, the Lys 79 in $\underline{c}$ is not present in $\underline{c}_{551}$. If this residue was responsible for the displacement of the Met ligand, then its nonexistence in $\underline{c}_{551}$ would explain the unobserved transition. A more likely explanation is the nonexistence of a Tyr residue homologous to the residue 67 in $\underline{c}$. This residue in cytochrome $\underline{c}$ is an integral part of the heme packing, as opposed to being solvent exposed as is Lys 79. Its ionization around the pH of the observed transition would certainly have a disruptive effect on the heme and protein structure in this region. This disruption would allow Lys 79 to assume the axial position. In $\underline{c}_{551}$ there is no homologous Tyr or Lys and thus no transition. Additional evidence is obtained for the importance of Tyr 67 in this transition from chemical modification studies. Modification of Tyr 67 by mononitration (74) and diiodination (75) lowered the pK_a of the tyrosine hydroxyl group and resulted in a lowered apparent pK_a of this transition. Salemme et al. have previously suggested the importance of Tyr 67 in this transition, but have proposed that Tyr 67 itself assumed the axial position (76).

REFERENCES

1. Harbury, H., personal communication.
2. Ambler, R.P. (1963) Biochem. J. 89, 341-349.
3. Zeppezauer, M., Eklund, H. and Zeppezauer, E. (1968) Arch. Bioch. 126, 564-573.
4. Zeppezauer, M. (1971) in "Methods in Enzymology" (W.B. Jakoby, ed.), Vol. XXII. Academic Press, New York, pp. 253-266.
5. Baumber, M.E., Moew, P.C. and Bunn, C.W., personal communication.
6. Krieger, M., Chambers, J.L., Christoph, G.G., Stroud, R.M. and Trus, B.L. (1974) Acta Crystallogr. A30, 740-748.
7. North, A.C.T., Phillips, D.C., and Mathews, F.S. (1968) Acta Crystallogr. A24, 351-359.
8. Blundell, T.L. and Johnson, L.N. "Protein Crystallography", Academic Press, New York, 1976.
9. Stout, G.H. and Jensen, L.H. "X-ray Structure Determination", McMillan Co., London, 1968.
10. Azaroff, L.V. (1955) Acta Crystallogr. 8, 701-704.
11. Levy, H.A. and Ellison, R.D. (1960) Acta Crystallogr. 13, 270-271.
12. Dickerson, R.E., Kopka, M.L., Borders, C.L., Varnum, J., Weinzierl, J.E., and Margoliash, E. (1967) J. Mol. Biol. 29, 77-95.
13. Almassy, R.J., Doctoral Thesis, Chapter 2.
14. Blow, D.M. and Crick, F.H.C. (1959) Acta Crystallogr. 12, 794-802.

15. Dickerson, R.E., Kendrew, J.C. and Strandberg, B.E. (1961) Acta Crystallogr. 14, 1188-1195.

16. Dickerson, R.E., Weinzierl, J.E. and Palmer R.A. (1968) Acta Crystallogr. B24, 997-1003.

17. Dickerson, R.E., Kendrew, J.C. and Strandberg, B.E. (1961) in "Computing Methods and the Phase Problem in X-ray Crystal Analysis" (P. Pepinsky, J. Robertson and J. Speakman, eds.), Pergamon Press, New York, pp. 236-251.

18. Henderson, R.C. and Moffat, J.K. (1971) Acta Crystallogr. B27, 1414-1420.

19. Kraut, J., Sieker, L.C., High, D.F., and Freer, S.T. (1962) Proc. Nat. Acad. Sci. USA 48, 1417-1424.

20. Cullis, A.F., Muirhead, H., Perutz, M.F., Rossmann, M.G. and North, A.C.T. (1961) Proc. Roy. Soc. Lond. A265, 15-38.

21. Chambers, J.C. and Stroud, R.M. (1977) Acta Crystallogr. B33, 1824-1837.

22. Cochran, W. (1951) Acta Crystallogr. 4, 408-411.

23. Forsyth, J.B. and Wells, M. (1959) Acta Crystallogr. 12, 412-415.

24. Ten Eyck, L.F. (1973) Acta Crystallogr. A29, 183-191.

25. Chambers, J.L., personal communication.

26. International Tables for X-ray Crystallography, Vol. II (J. Kasper and K. Lonsdale, eds.), The Kynoch Press, England, p. 331, 1959.

27. Cruickshank, D.W.J. (1949) Acta Crystallogr. 2, 65-82.

28. Freer, S.T., Alden, R.A., Carter, C.W. and Kraut, J. (1975) J. Biol. Chem. 250, 46-54.

29. Hermans, J. and McQueen, J.E. (1974) Acta Crystallogr. A30, 730-739.

30. Dodson, E.J., Isaacs, N.W. and Rollett, J.S. (1976) Acta Crystallogr. A32, 311-315.

31. Ten Eyck, L.F., Weaver, L.H. and Matthews, B.W. (1976) Acta Crystallogr. A32, 349-350.

32. Marsh, R.E. and Donohue, J. (1967) Adv. Protein Chem. 22, 235-256.

33. Handbook of Biochemistry and Molecular Biology. Physical and Chemical Data, 3rd Edition (G. Fasman, ed.), CRC Press, Inc., Cleveland, pp. 222-223, 1976.

34. Dickerson, R.E., Eisenberg, D., Varnum, J. and Kopka, M.L. (1969) J. Mol. Biol. 45, 77.

35. Dickerson, R.E., Timkovich, R. and Almassy, R.J. (1976) J. Mol. Biol. 100, 473-491.

36. Swanson, R., Trus, B.L., Mandel, N., Mandel, G., Kallai, O.B. and Dickerson, R.E. (1977) J. Biol. Chem. 252, 776-785.

37. Takano, T., Trus, B.L., Mandel, N., Mandel, G., Kallai, O.B., Swanson, R. and Dickerson, R.E. (1977) J. Biol. Chem. 252, 776-785.

38. Timkovich, R. and Dickerson, R.E. (1976) J. Biol. Chem. 251, 4033-4046.

39. Salemme, F.R., Freer, S.T., Xuong, Ng.-H., Alden, R.A. and Kraut, J. (1973) J. Biol. Chem. 248, 3910-3921.

40. Maigret, B., Pullman, B. and Perakia, D. (1971) J. Theoret. Biol. 31, 269-285.

41. Mandel, N., Mandel, G., Trus, B.L., Rosenberg, J., Carlson, G. and Dickerson, R.E. (1977) J. Biol. Chem. 252, 4619-4636.

42. Sasisekharan, V. (1959) Acta Crystallogr. 12, 897-903.

43. Dickerson, R.E. and Geis, I. (1969). "The Structure and Action of Proteins," W. A. Benjamin, Inc., Menlo Park.

44. Bode, W. and Schwager, P. (1975) J. Mol. Biol. 98, 693-717.

45. Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J. and Steigemann, W. (1974) J. Mol. Biol. 89, 73-101.

46. Matthews, B.W. (1976) Ann. Rev. Phys. Chem. 27, 493-523.

47. Bernhard, S.A. (1968) "The Structure and Function of Enzymes," W. A. Benjamin, Inc., Menlo Park.

48. Ladner, R.C., Heidner, E. and Perutz, M.F. (1972) J. Mol. Biol. 114, 385-414.

49. Takano, T. (1977) J. Mol. Biol. 110, 537-568.

50. Takano, T. (1977) J. Mol. Biol. 110, 569-584.

51. Fleischer, E.B. (1970) Accounts of Chem. Res. 3, 105-112.

52. Scanlon, W.J. and Eisenberg, D. (1975) J. Mol. Biol. 98, 485-502.

53. Kuntz, I.D. and Kauzmann, W. (1974) Adv. Protein Chem. 28, 239-345.

54. Tanford, C. (1968) Adv. Protein Chem. 23, 122-275.

55. Needleman, S.B. and Blair, T.T. (1969) Proc. Nat. Acad. Sci. USA 63, 1227-1233.

56. Dickerson, R.E. (1971) J. Mol. Biol. 57, 1-15.

57. McLachlan, A.D. (1971) J. Mol. Biol. 61, 409-424.

58. Almassy, R.J. and Dickerson, R.E. (1978) Proc. Nat. Acad. Sci. USA--in press.

59. Dickerson, R.E. and Timkovich, R. (1975). In "The Enzymes" (P. Boyer, ed.), Vol. 11, pp. 397-547. Academic Press, New York.

60. Dickerson, R.E. (1978) International Symposium on Biomolecular Structure, University of Madras.

61. Hoard, J.L. (1971) Science 174, 1295-1302.

62. Collman, J.P., Gagne, R.R., Reed, C.A., Robinson, W.T., and Rodley, G.A. (1974) Proc. Nat. Acad. Sci. USA 71, 1326-1329.

63. Hendrickson, W.A. and Love, W.E. (1971) Nature 232, 197-203.

64. Padlan, E.A. and Love, W.E. (1974) J. Biol. Chem. 249, 4067-4083.

65. Huber, R., Epp, O., and Formanek, H. (1970) J. Mol. Biol. 52, 349-354.

66. Norvell, J.C. and Schoenborn, B.P. (1976) Brookhaven Symposia in Biology 27, II-12-II23.

67. Bersohn, R., personal communication.

68. Kowalsky, A. (1965) Biochemistry 4, 2382-2388.

69. Smith, G.M. and Kamen, M.D. (1974) Proc. Nat. Acad. Sci. USA 71, 4303-4306.

70. Keller, R.M. and Wuthrich, K. (1976) FEBS Letters 70, 180-184.

71. Dickerson, R.E. (1971) J. Mol. Evolu. 1, 26-45.

72. Horio, T., Hagashi, T., Sasagawa, M., Kusai, K., Nakai, M., and Okunuki, K. (1960) Biochem. J. 77, 194-201.

73. Vinogradov, S.N. (1970) Biopolymers 9, 507-509.

74. Skov, K., Hofmann, T. and Williams, G.R. (1969) Can. J. Biochem. 47, 750-752.

75. Morton, R. (1973) Can. J. Biochem. 51, 472-475.

76. Salemme, F.R., Kraut, J. and Kamen, M.D. (1973) J.Biol. Chem. 248, 7701-7716.

FIGURE LEGENDS

Figure 1. Radial distribution of (a) the mean native structure factor amplitudes (F_p), (b) the mean figure of merit (FM), (c) the mean fractional change in structure factor amplitudes for the three derivatives ($\Delta F/F_p$), and (d) the mean ratio of the lack of closure error to the changes in structure factor amplitudes produced by each derivative ($\varepsilon/\Delta F$). PT = K_2PtCl_4, UO = $UO_2(NO_3)_2 \cdot 6H_2O$, and AU = $NaAu(CN)_2$.

Figure 2. Harker sections from the 2.1 Å resolution $(\Delta F)^2$ Patterson calculated for the K_2PtCl_4 derivative. The two heavy atom binding sites A and B were located. The sections are contoured at equal intervals above zero, with the zero contour level omitted.

Figure 3. Harker sections from the 2.43 Å resolution $(\Delta F)^2$ Patterson calculated for the $UO_2(NO_3)_2 \cdot 6H_2O$ derivative. The single heavy atom binding site A was located. The sections are contoured as in Figure 2.

Figure 4. Harker sections from the 2.43 Å resolution $(\Delta F)^2$ Patterson calculated for the $NaAu(CN)_2$ derivative. The single heavy atom binding site A was located. The sections are contoured as in Figure 2.

Figure 5. The progress of refinement is displayed in terms of the R-factor. The R-factor is shown before and after each of the five minimaps were inspected and manual adjustments made. Important details that occurred during computer refinement between minimaps are presented. The R-factors are plotted for all structure factor calculations used to obtain the refined coordinates. A more detailed summary of refinement is presented in Appendix III.

Figure 6. The R-factor calculated for all data to the resolution indicated on the abscissa. The R-factor is calculated for both linear (0) and four parameter scaling (+) of F_{calc} to F_{obs}.

Figure 7. The radial distribution of the calculated mean phase change between the refined phases and (a) the MIR phases, (b) the calculated phases from the initial wire model coordinates, and (c) the MIR phases with the differences weighted by the figures of merit.

Figure 8. A ribbon drawing illustrating the structure of cytochrome $\underline{c}_{551}$.

Figure 9. (a) A stereo front view of the structure of $\underline{c}_{551}$. The orientation was calculated from a least-squares fit to 44 homologous alpha carbons of tuna cytochrome $\underline{c}$ (36). The protein thus has an orientation similar to the cytochromes in Ref. 35. The main chain alpha carbons are shown along with several interesting side

chains. (b) Figure (a) rotated by 90 degrees.

Figure 10. A Ramachandran plot of the ϕ and ψ angles in Table IV. The contour level is drawn at 3 kcal/mole above the global minimum of energy (40). These results can be compared to the plot obtained from eukaryotic cytochrome c (41).

Figure 11. A stereo view of an idealized polyproline three-fold helix (left in each stereo pair) and the c_{551} chain from residues 58 to 63. The Pro-Ile-Met-Pro-Pro sequence apparently is a device for obtaining an extended chain configuration past the site for the heme attachment.

Figure 12. The average distance between the initial MIR wire model coordinates and the refined coordinates for each residue. (a) Main chain atoms. (b) Side chain atoms. (c) All atoms.

Figure 13. The structure of cytochrome c_{551} from residues Asn 50 through Gln 53 and Pro 58 through Ala 65. These residues form an antiparallel chain with a loop below the density map which reverses the direction. The heme pyrrole ring 2, part of ring 3, and the iron electron density are shown at the right-center. The Ile 59 side chain was rotated about the alpha to beta carbon bond during refinement. Other interesting features are the hydrogen bonds between the two chains and the solvent.

(a) The MIR electron density map and wire model coordinates. The contour levels start at 0.24 e/Å^3 and increment at 0.24 e/Å^3. The RMS electron density error is 0.16 e/Å^3.

(b) Difference map and coordinates at an R-factor of 25.8% (minimap II in Figure 5). The electron density contours start at 0.20 e/Å^3 and increment at 0.10 e/Å^3. The RMS error in the electron density is 0.14 e/Å^3.

(c) Difference map and coordinates at an R-factor of 16.2%. The electron density contours start at 0.20 e/Å^3 and increment at 0.10 e/Å^3.

(d) The $2F_{obs}$-F_{calc} electron density and coordinates at an R-factor of 16.2%. The electron density contours start at 0.60 e/Å^3 and increment at 0.60 e/Å^3. The RMS error in the electron density is 0.08 e/Å^3.

Figure 14. Sections of the electron density perpendicular to the N-terminal alpha helix. Residues Val 5, Leu 6, Phe 7, and Lys 8 are included. Lys 8 is the only one of these residues which is not well defined. The coordinates in (a) and (d) include both the initial wire model coordinates (open bonds) and the refined coordinates (solid bonds). Changes during refinement can be seen.

(a) MIR electron density map. The contour levels start at 0.45 e/Å^3 and increment at 0.45 e/Å^3. The RMS electron density error is 0.16 e/Å^3.

(b) The difference map and coordinates at an R-factor of 25.8%.

The contour levels start at 0.45 e/Å^3 and increment at 0.45 e/Å^3. The RMS electron density error is 0.14 e/Å^3.

(c) The difference map and coordinates at an R-factor of 16.2%. The electron density contours start at 0.20 e/Å^3 and increment at 0.10 e/Å^3.

(d) The $2F_{calc}-F_{obs}$ electron density and coordinates at an R-factor of 16.2%. The electron density contours start at 0.60 e/Å^3 and increment at 0.60 e/Å^3. The RMS error in the electron density is 0.08 e/Å^3.

Figure 15. Average temperature factors calculated at various points in the refinement. The average temperature factor for each residue is plotted as a histogram. The values were calcualted at minimaps II through V (see Figure 5) and for the final structure (from left to right). (a) Main chain atoms. (b) Side chain atoms. (c) All atoms.

Figure 16. Sections of electron density for Val 23. Both the initial wire model coordinates (open bonds) and the refined coordinates (solid bonds) are plotted. From minimap I (Figure 16b), a manual rotation about the alpha to beta carbon was made. The improvement was accompanied by a large drop in the residue's temperature factors and improved electron density maps.

(a) The MIR electron density map. Contours start at 0.24 e/Å^3 and increment at 0.24 e/Å^3.

(b) The difference electron density map at an R-factor of 33.6% (minimap I). Contours start at 0.30 e/Å^3 and increment at 0.10 e/Å^3.

(c) The refined difference electron density map at an R-factor of 16.2%. The orientation is rotated 180° about the vertical axis from that in (a). Contours start at 0.10 e/Å^3 and increment at 0.10 e/Å^3.

(d) The refined $2F_{obs}$-F_{calc} electron density map at an R-factor of 16.2%. The orientation is the same as in (c). The contours start at 0.40 e/Å^3 and increment at 0.40 e/Å^3.

Figure 17. Sections of the electron density for Leu 44. The Leu coordinates were obtained by a fitting to the difference and $2F_{obs}$-F_{calc} electron density maps calculated with Val coordinates at an R-factor of 16.2%. The incorrect Val atoms are shown as small circles. The gamma carbon directly beneath the beta carbon had a temperature factor of 40 Å . Difference map contours start at 0.075 e/Å^3 and increment at 0.075 e/Å^3. $2F_{obs}$-F_{calc} map contours start at 0.00 e/Å^3 and increment at 0.20 e/Å^3. (a) The difference map calculated from structure factors obtained with Val coordinates. (b) The $2F_{obs}$-F_{calc} map calculated from structure factors obtained with Val coordinates. (c) The difference map calculated from structure factors obtained with Leu coordinates. (d) The $2F_{obs}$-F_{calc} map calculated from structure factors obtained with Leu coordinates.

Figure 18. The distribution of deviations between the observed bonding parameters and standard bonding parameters used in constraints. (a) N-C_α bond distances. (b) All main chain bond distances. (c) C_α-C_β-C_γ bond angles. (d) Main chain torsion angles C_α-C-N-C_α and O-C-N-C_α.

Figure 19. The 2.4Å resolution MIR electron density map. (a) The heme plane section. (b) The plane perpendicular to the heme plane and the 2N - 4N direction. (c) The plane perpendicular to the heme plane and the 1N - 3N direction.

Figure 20. A 2.0Å resolution map calculated at an R-factor of 19.5%. The iron was constrained to be at the heme center. Sections were calculated perpendicular to the heme plane. (a) The difference map contoured at intervals of 0.20 e/Å^3 above zero. (b) Same as (a). (c) $2F_{obs}$-F_{calc} electron density map contoured at increments of 0.40 e/Å^3 above zero. (d) Same as (c).

Figure 21. A 2.0Å resolution map calculated at an R-factor of 16.2% with all iron constraints removed. (a) The difference map contoured at intervals of 0.20 e/Å^3 above zero. (b) Same as (a). (c) $2F_{obs}$-F_{calc} electron density map contoured at increments of 0.40 e/Å^3 above zero. (d) Same as (c). (e) $2F_{obs}$-F_{calc} heme plane electron density section.

Figure 22. A 2.0Å resolution map calculated at an R-factor of 16.6% after the iron-heme nitrogen bond constraints were added to the structure in Figure 21. (a) The difference map contoured at intervals of 0.20 e/Å^3 above zero. (b) Same as (a). (c) $2F_{obs}-F_{calc}$ electron density map contoured at increments of 0.40 e/Å^3 above zero. (d) Same as (c).

Figure 23. (a) The difference electron density map and heme structure after refinement with heme bond angle and torsion angle constraints removed.

(b) The difference electron density map and heme structure after applying heme bond angle and torsion angle constraints to the heme in (a).

Figure 24. The 2.0 Å resolution $2F_{obs}-F_{calc}$ electron density map and heme structure at an R-factor of 16.2%.

Figure 25. A stereo view of the solvent structure of cytochrome $\underline{c}_{551}$. Solvent with contacts up to 3.0 Å resolution are indicated with a heavy bond. Hydrophilic solvent contacts up to 3.5 Å resolution are indicated with a light bond. Symmetry related solvents have the same identification number. (a) Front orientation. (b) Left orientation.

Figure 26. A comparison of the distances between alpha carbons in cytochromes $\underline{c}$ and $\underline{c}_{551}$.

Figure 27. A stereo view of eukaryotic cytochrome c which has been modified to show a possible mechanism by which the terminal Trp of f-type cytochromes might hydrogen bond to the buried propionic acid. The bottom residues are deleted as in c_{551}, which allows the C-terminal helix to be extended into hydrogen bonding position.

Figure 1

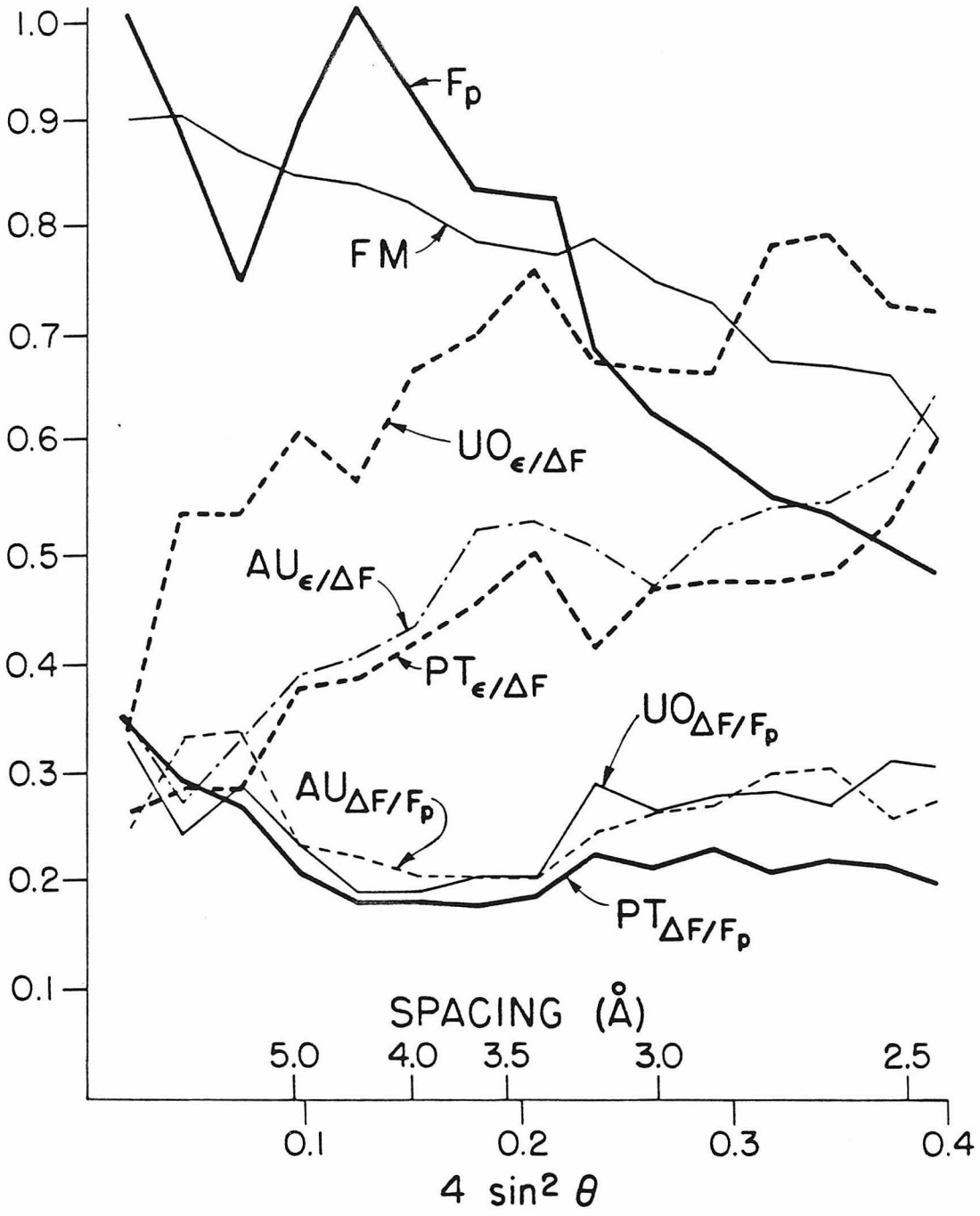

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
F_p
FM
$UO_{\epsilon/\Delta F}$
$AU_{\epsilon/\Delta F}$
$PT_{\epsilon/\Delta F}$
$UO_{\Delta F/F_p}$
$AU_{\Delta F/F_p}$
$PT_{\Delta F/F_p}$
SPACING (Å)
5.0
4.0
3.5
3.0
2.5
0.1
0.2
0.3
0.4
4 sin² θ

Figure 2a

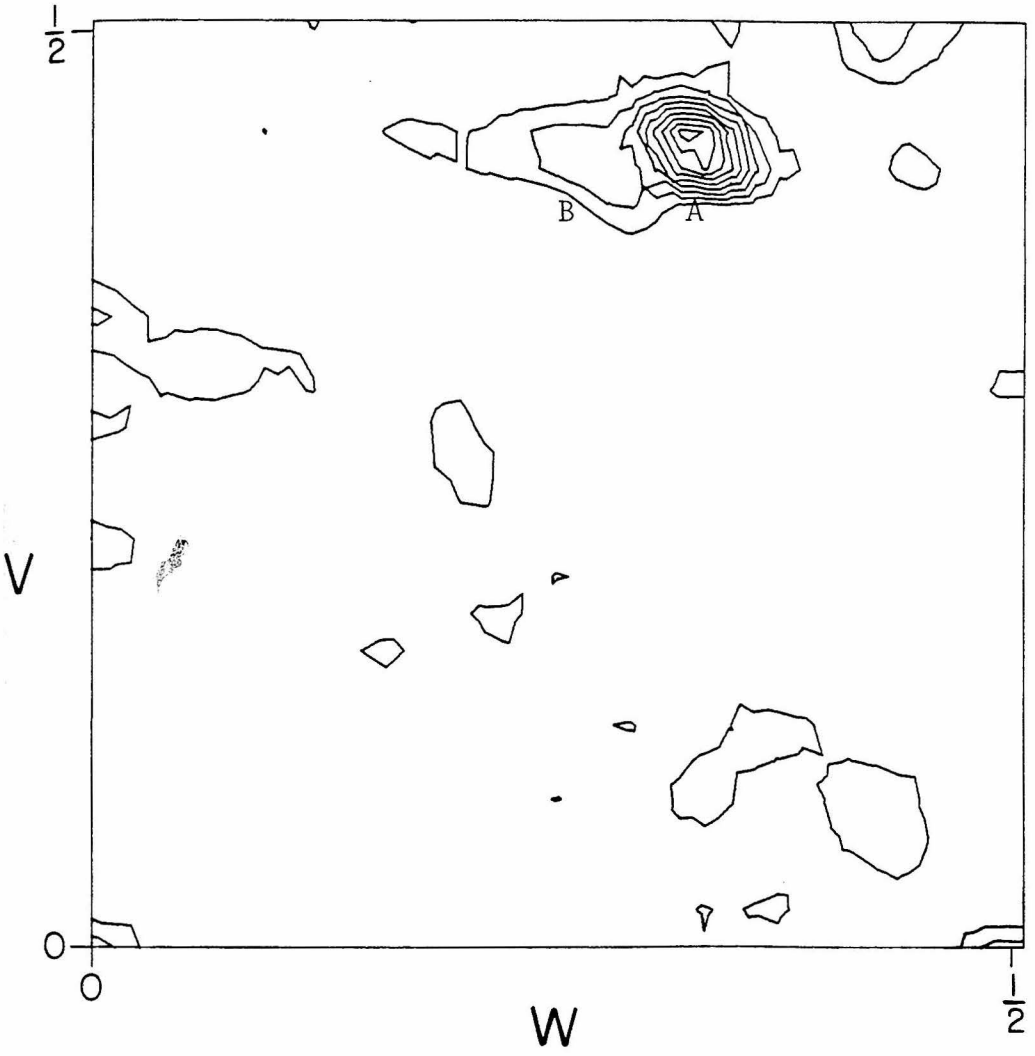


Figure 2b

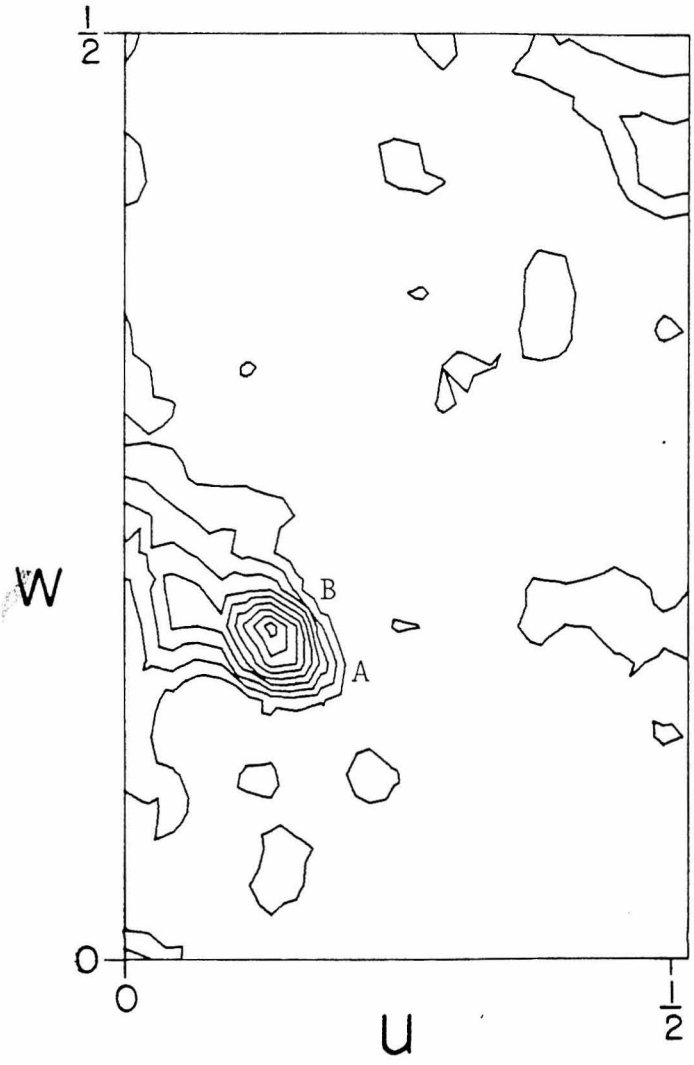


Figure 2c

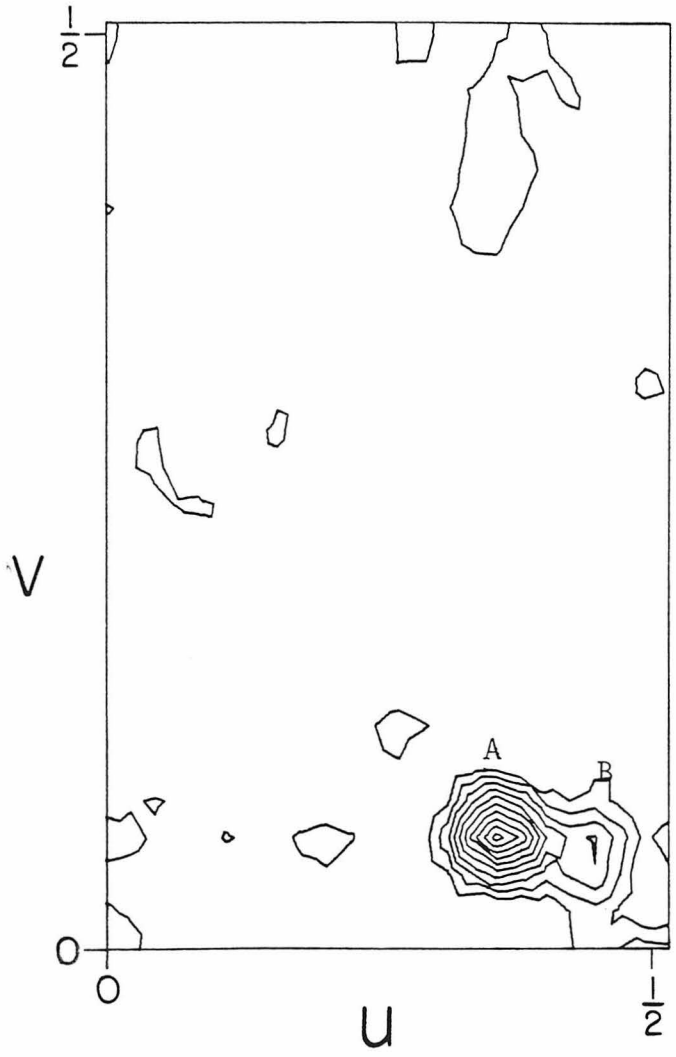


Figure 3a

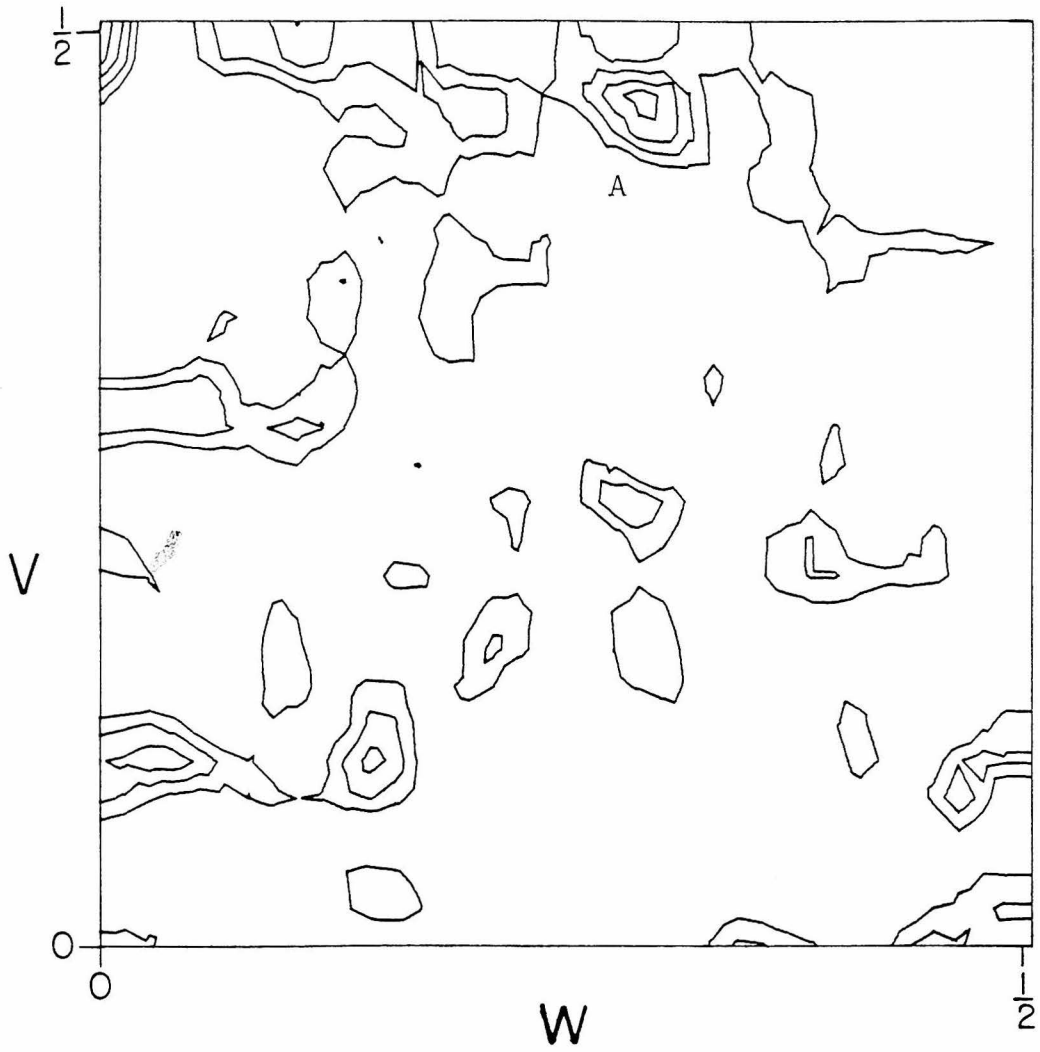


Figure 3b

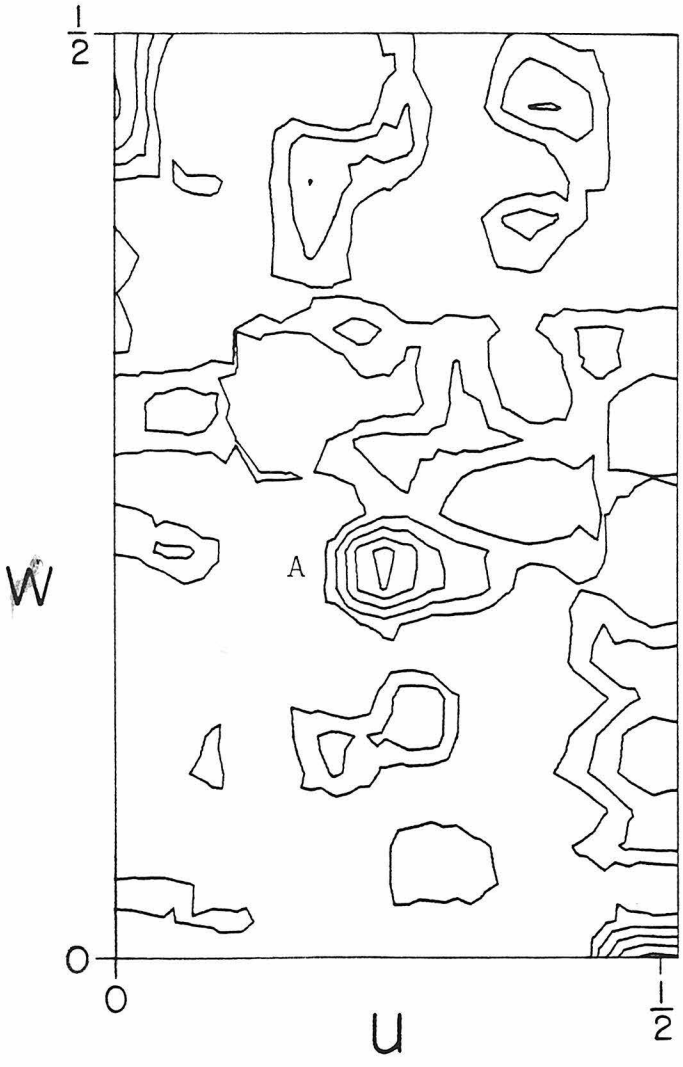


Figure 3c

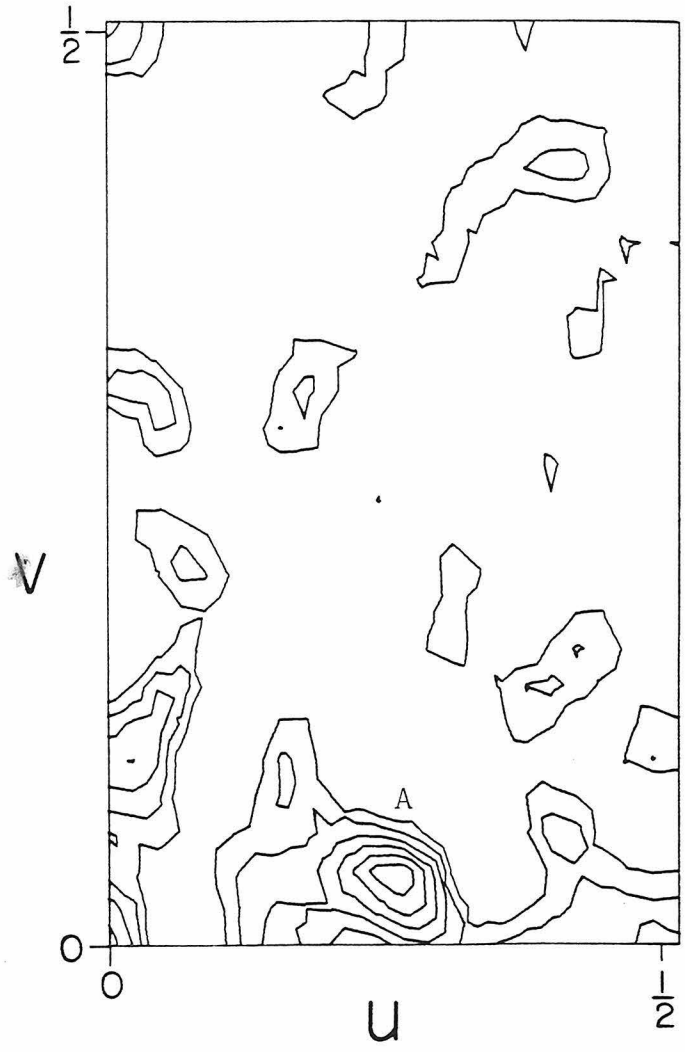


Figure 4a

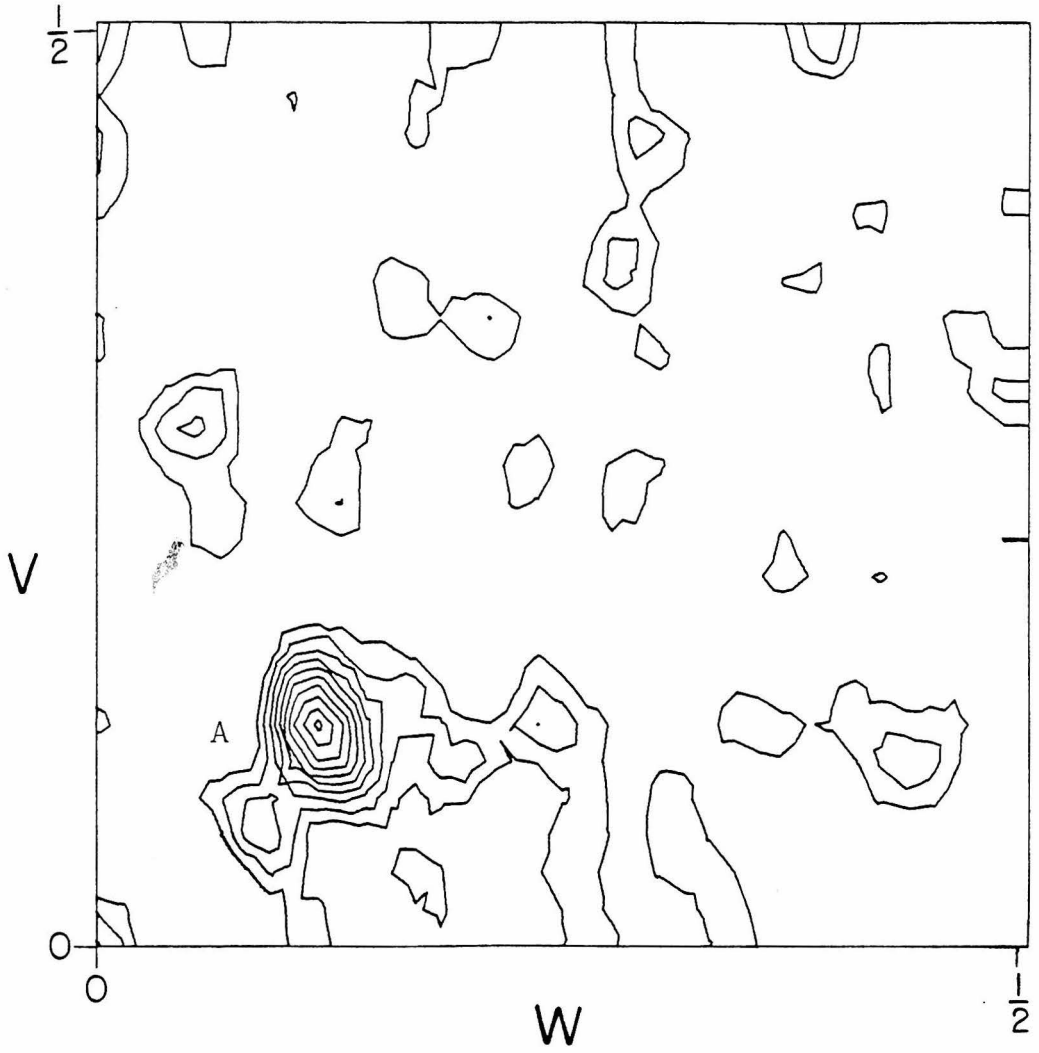


Figure 4b

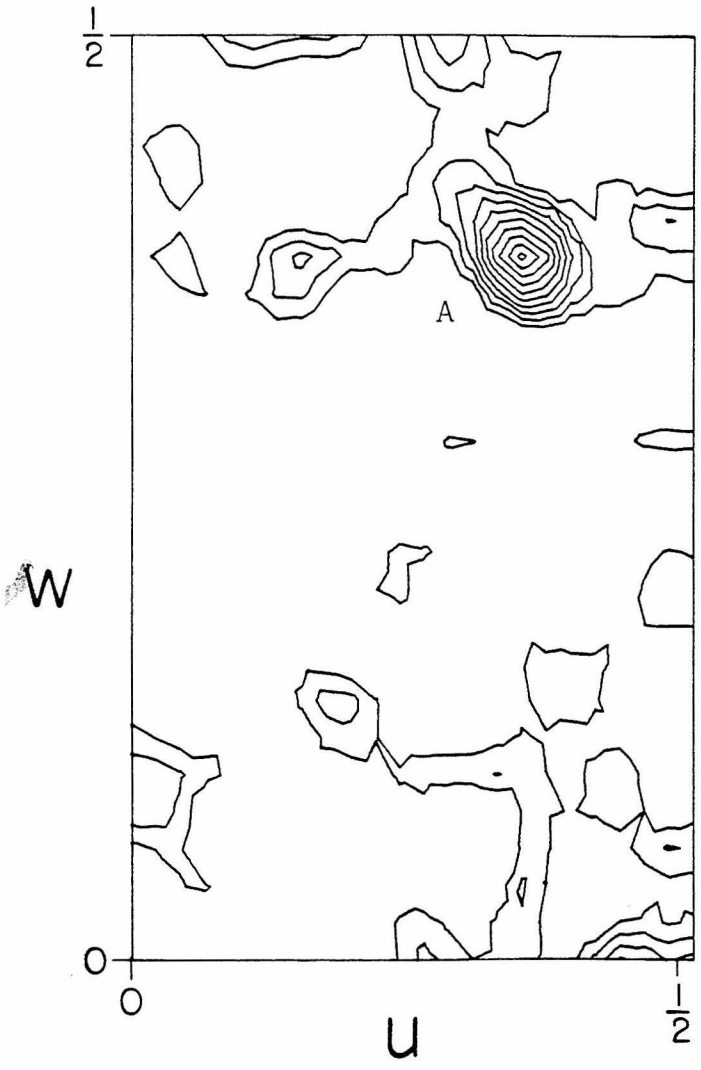


Figure 4c

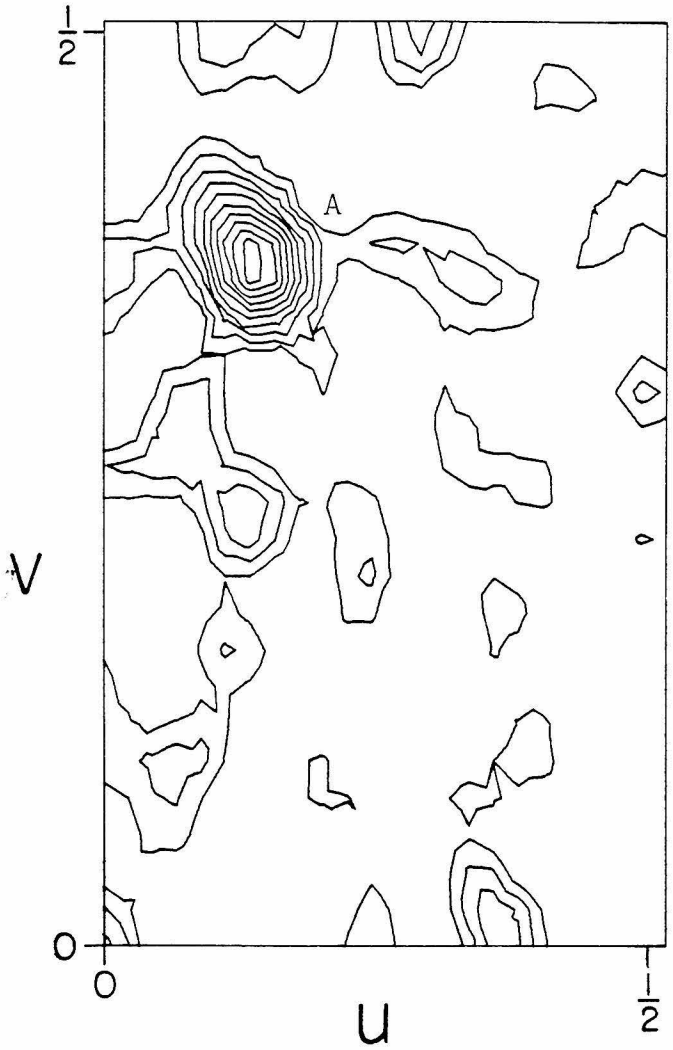


Figure 5

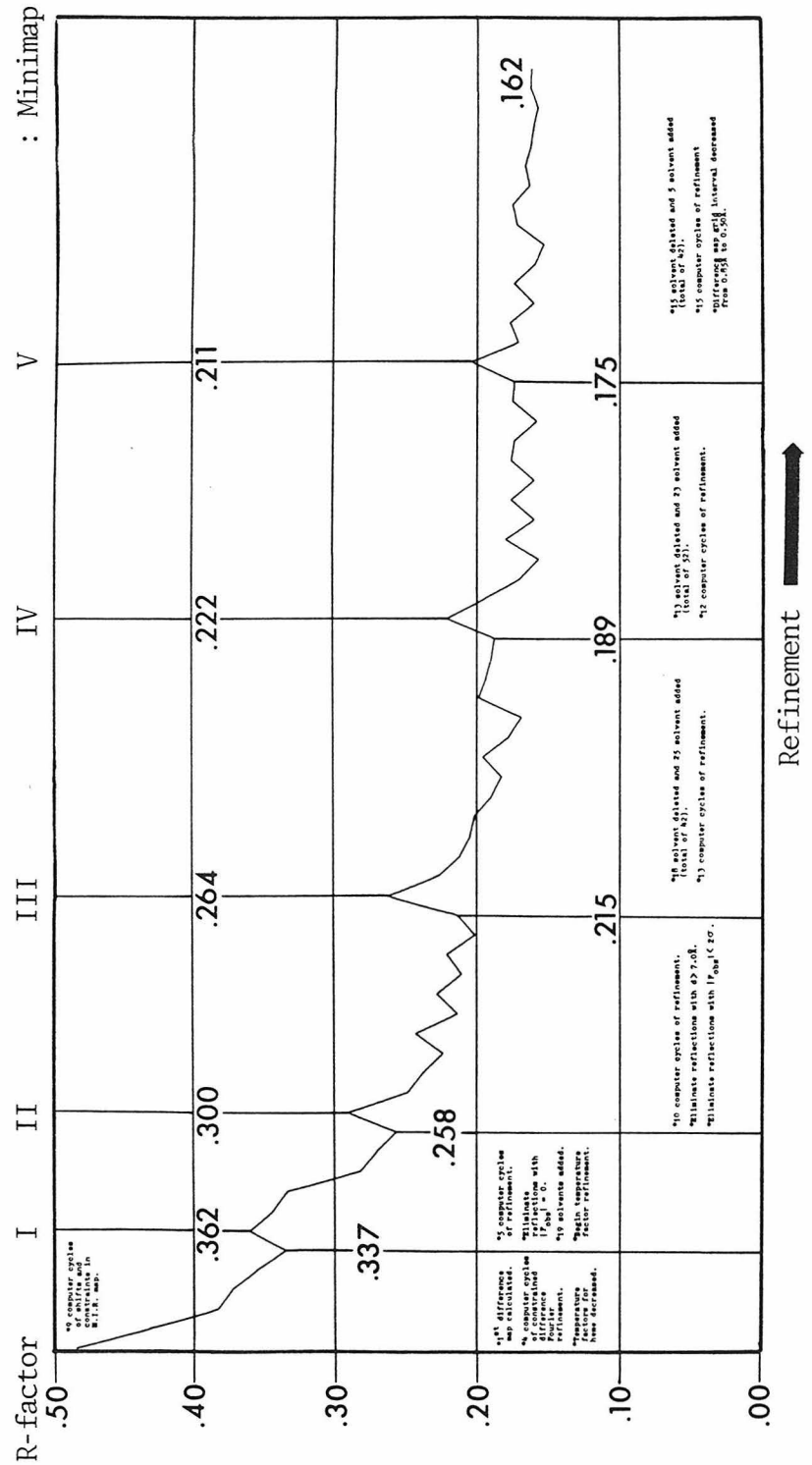

R-factor
.50
.40
.30
.20
.10
.00
I
II
III
IV
V
: Minimap
.362
.337
.300
.258
.264
.215
.222
.189
.211
.175
.162
Refinement

Figure 6

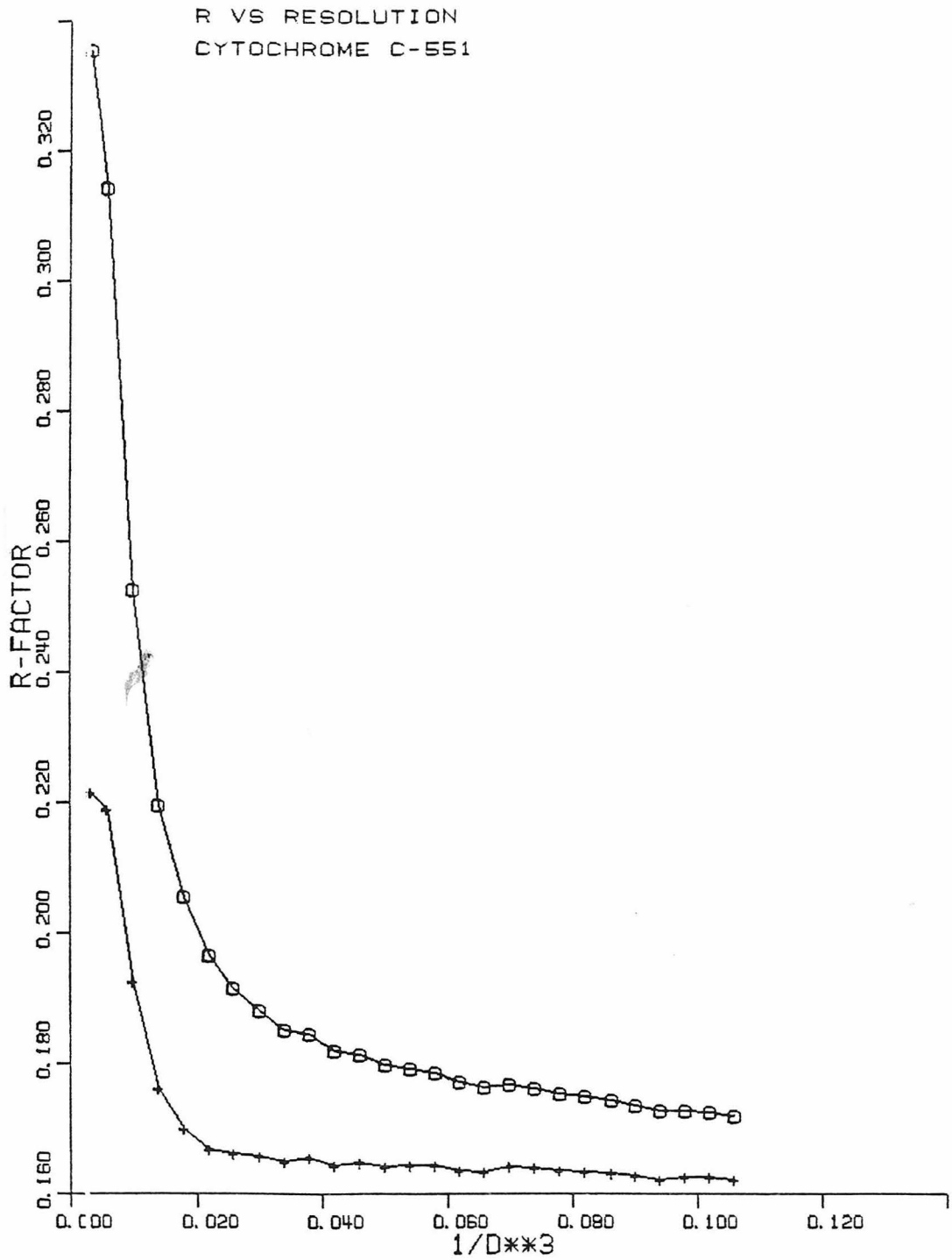

R VS RESOLUTION
CYTOCHROME C-551
R-FACTOR
0.160
0.180
0.200
0.220
0.240
0.260
0.280
0.300
0.320
0.000
0.020
0.040
0.060
0.080
0.100
0.120
1/D**3

Figure 7

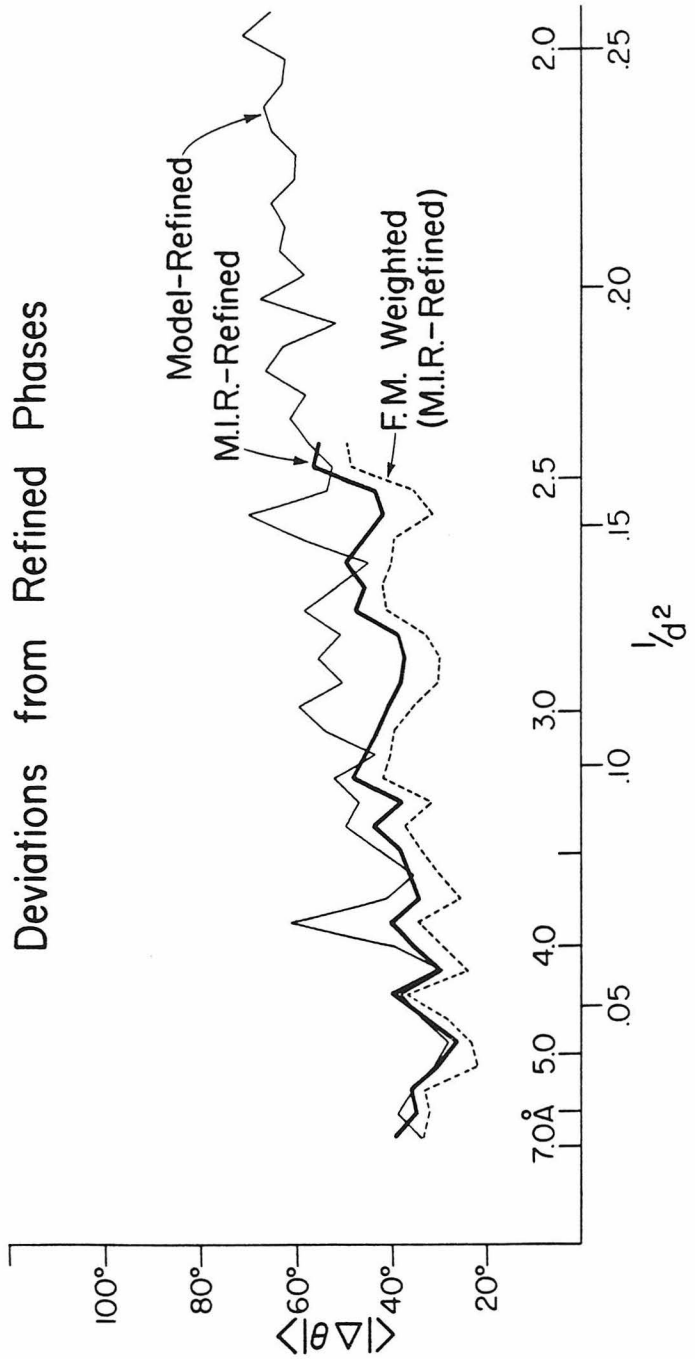

Deviations from Refined Phases
Model-Refined
M.I.R.-Refined
F.M. Weighted
(M.I.R.-Refined)
100°
80°
60°
40°
20°
⟨|Δθ|⟩
7.0Å
5.0
4.0
3.0
2.5
2.0
.05
.10
.15
.20
.25
1/d²

Figure 8

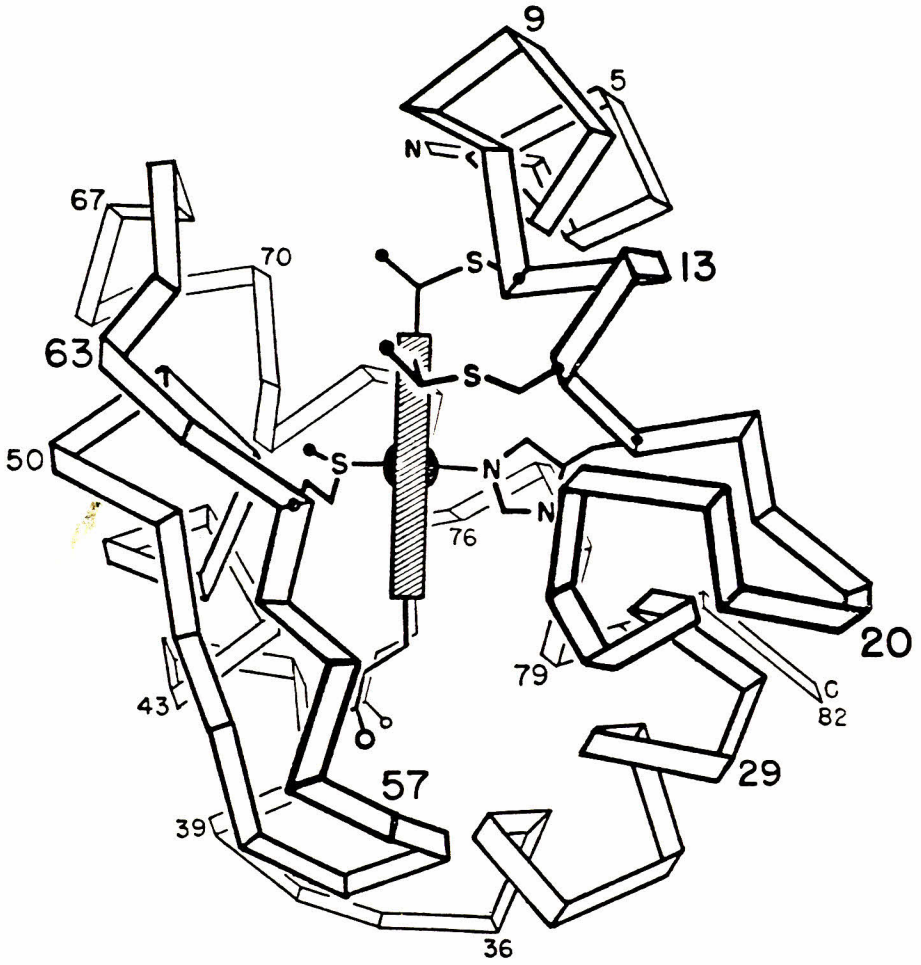


Figure 9

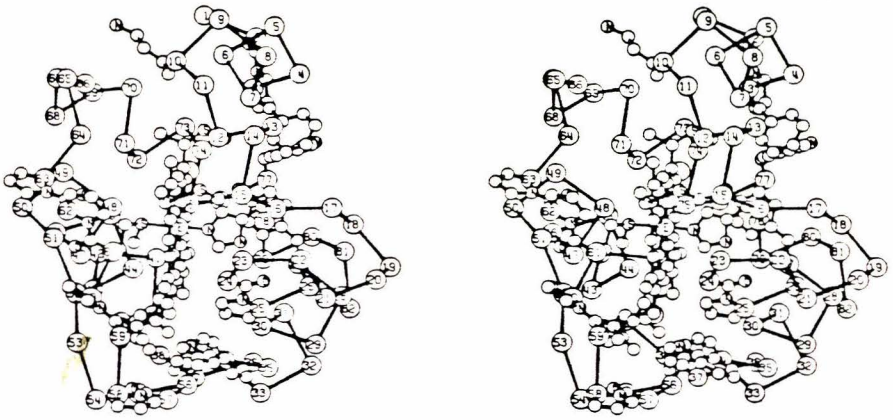


Figure 9b

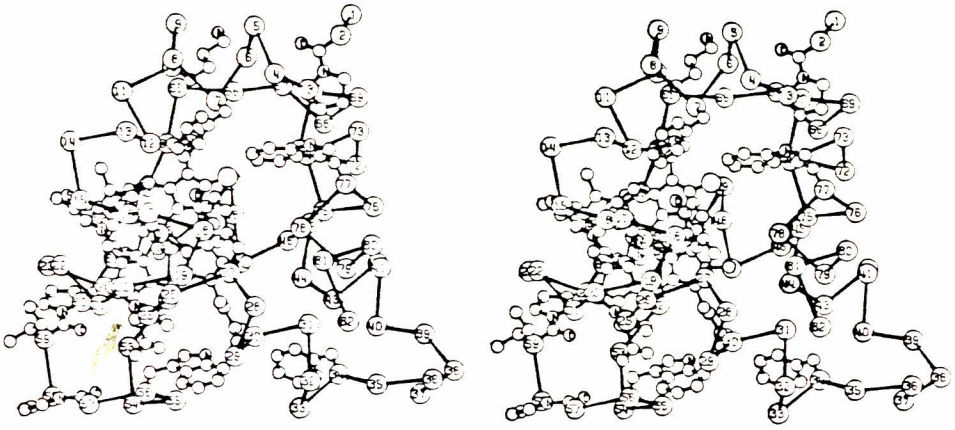


Figure 10

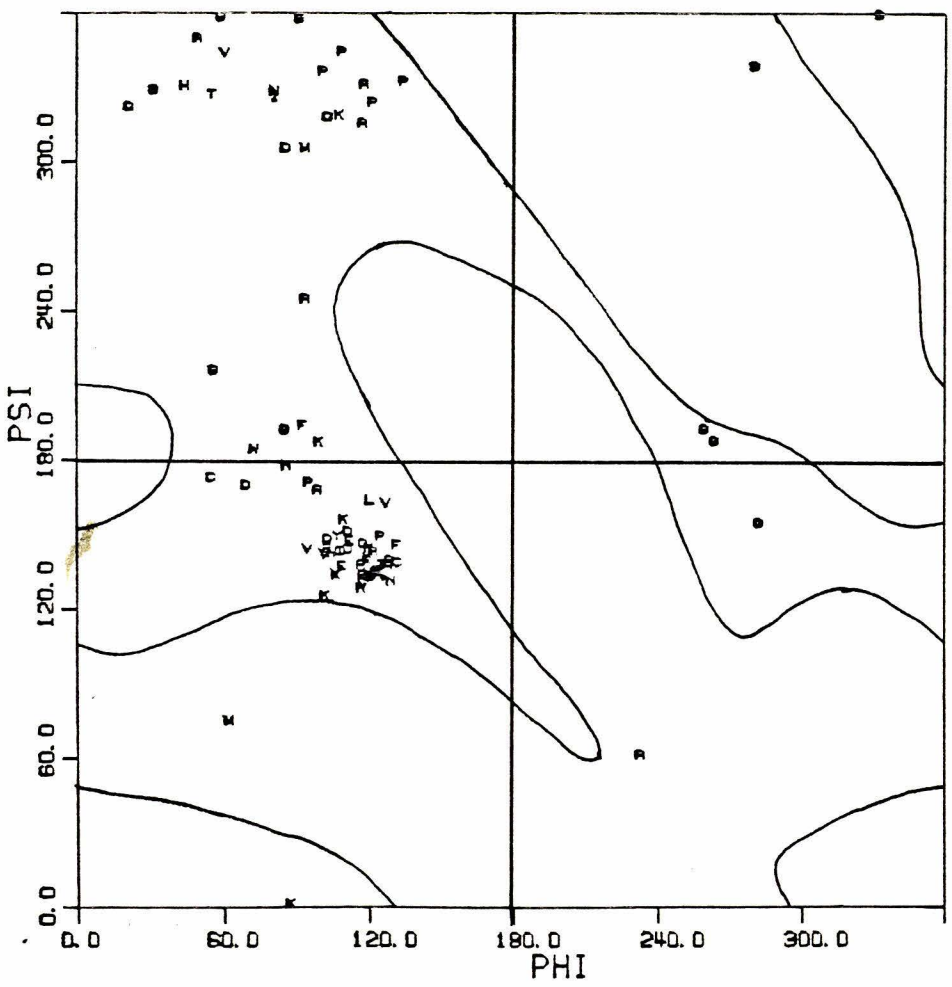


Figure 11

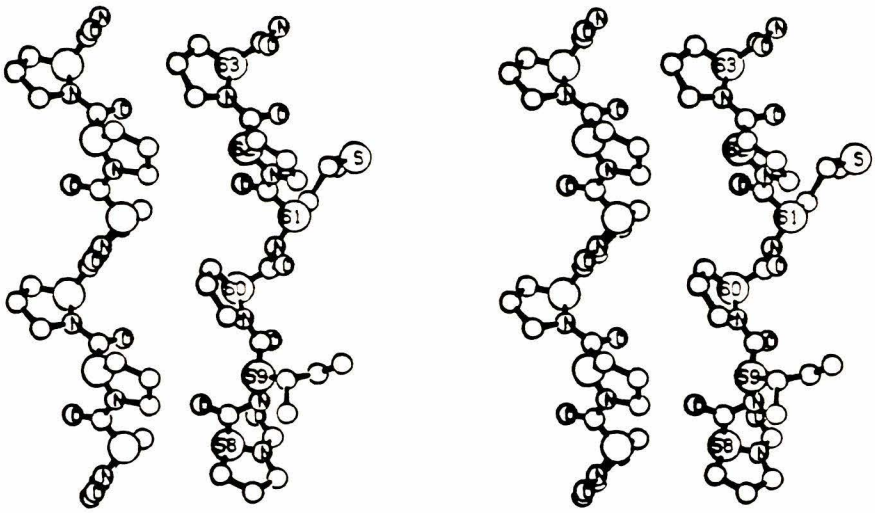


Figure 12a

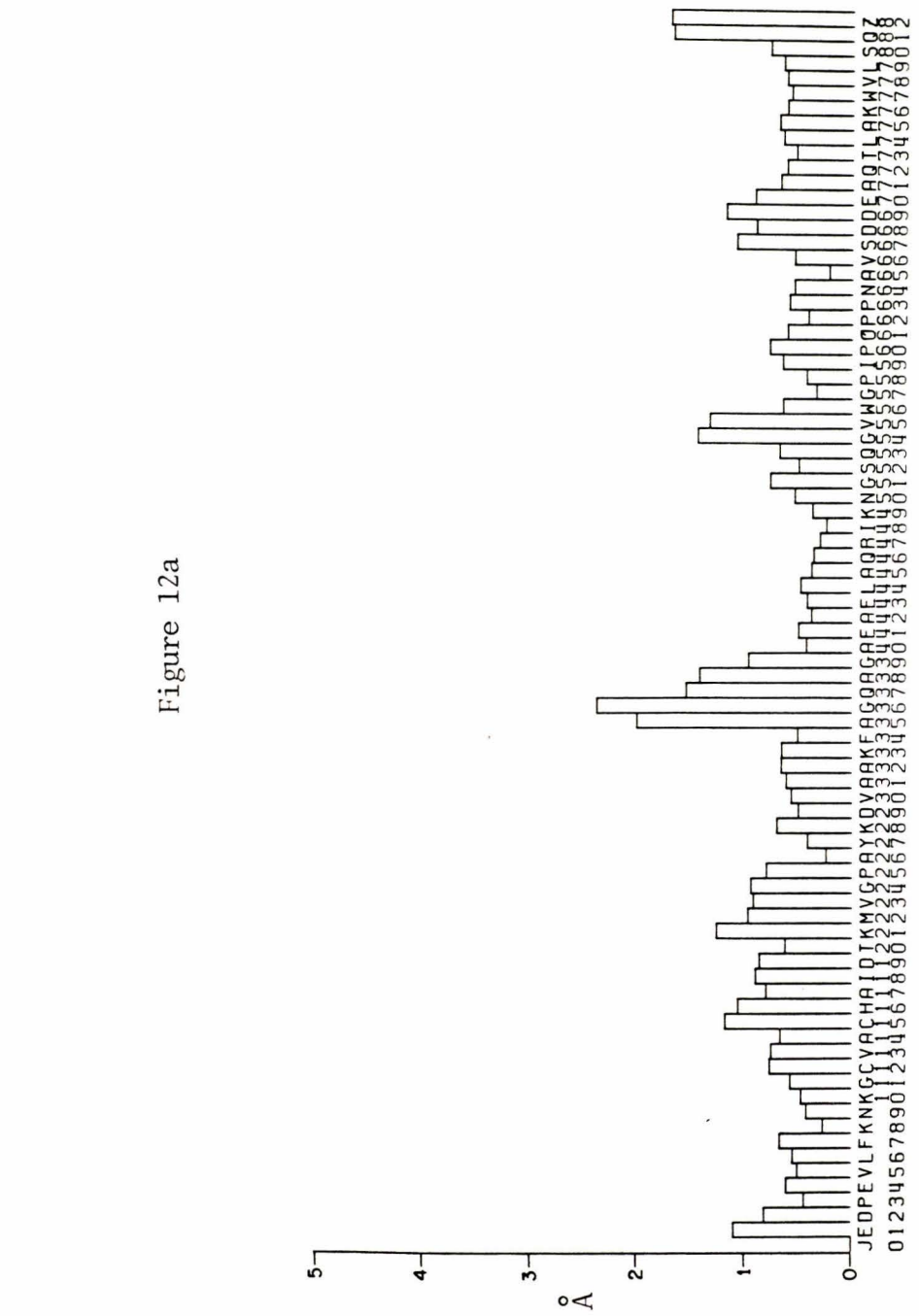


Figure 12b

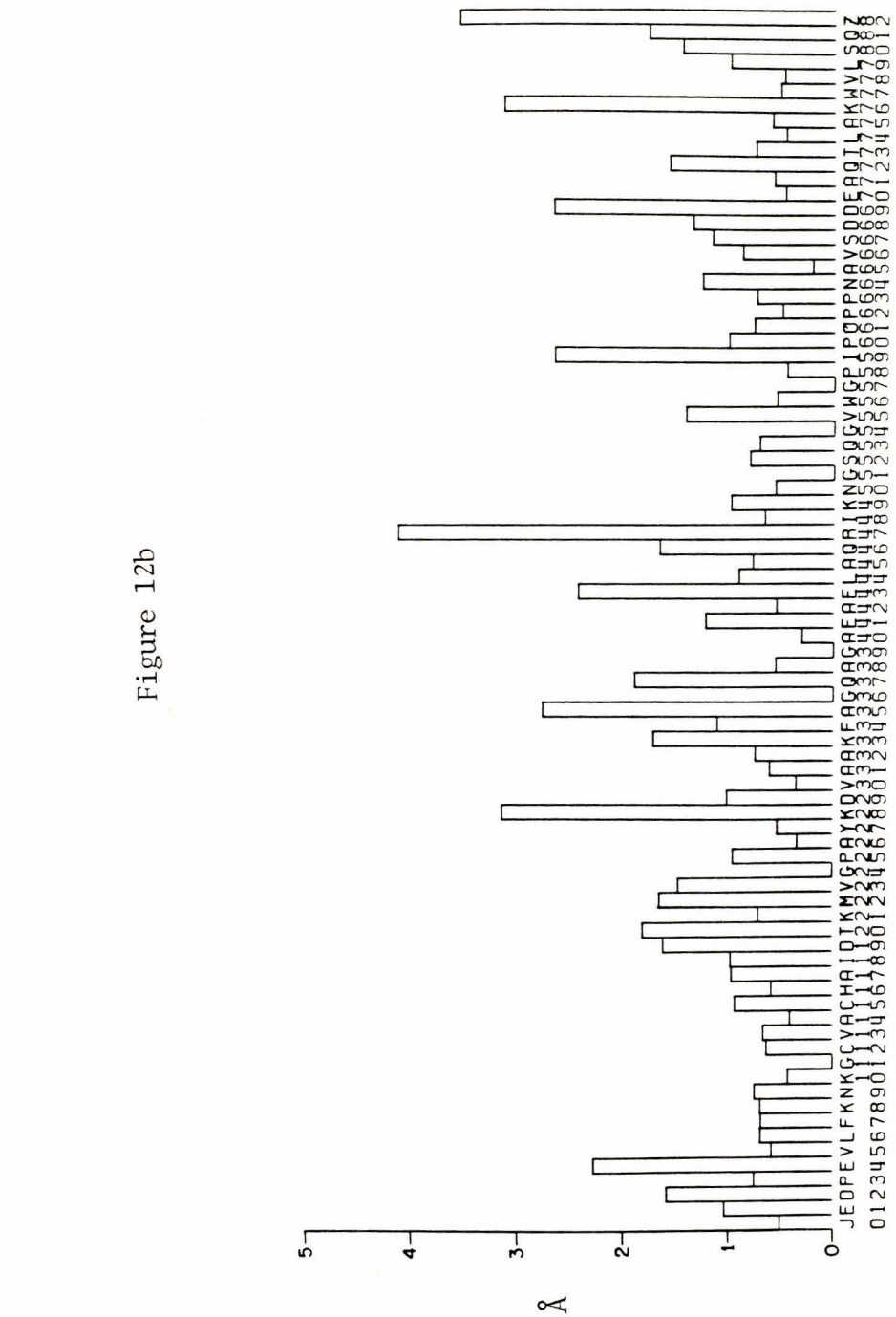
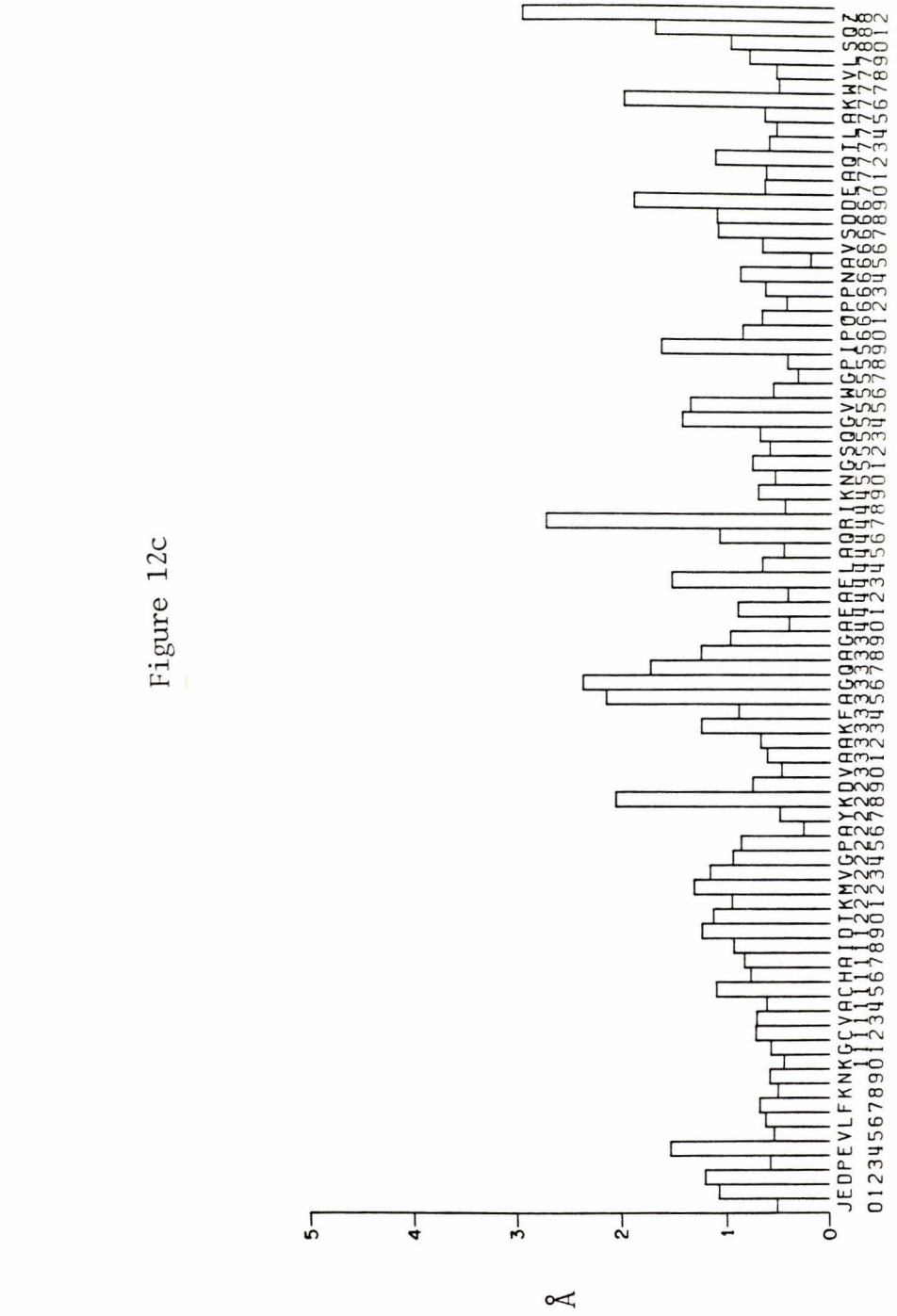


Figure 12c



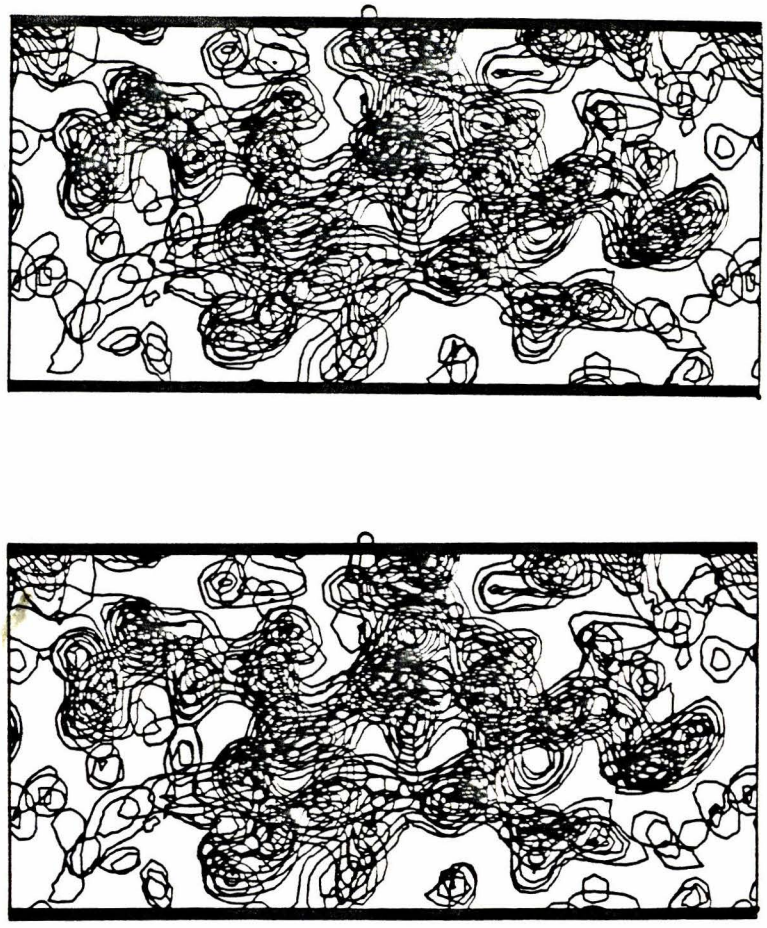


Figure 13a

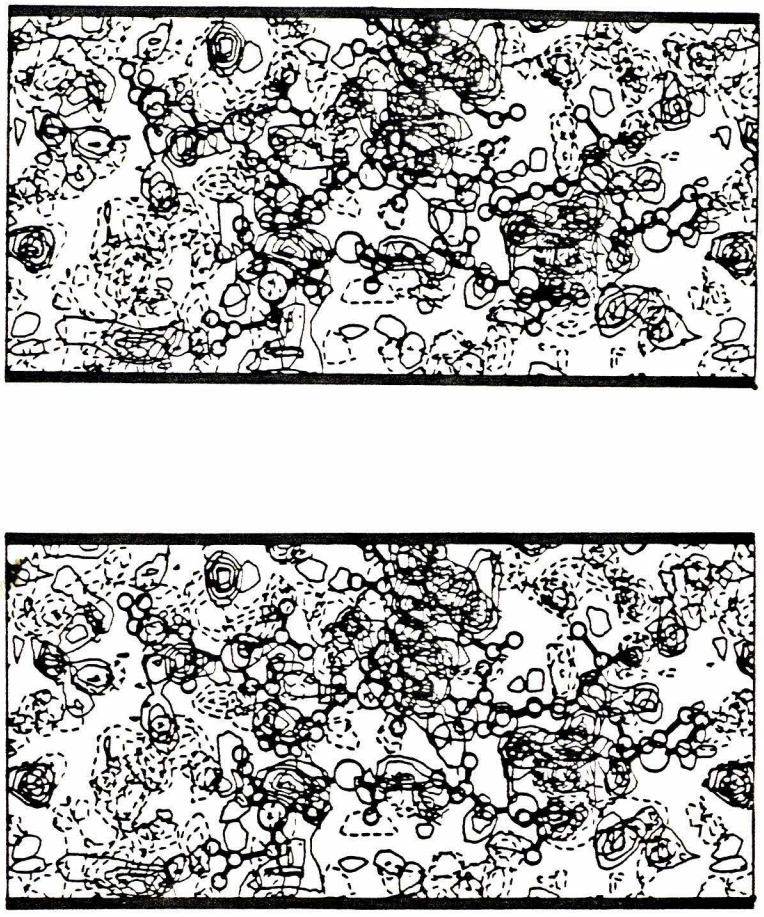


Figure 13b

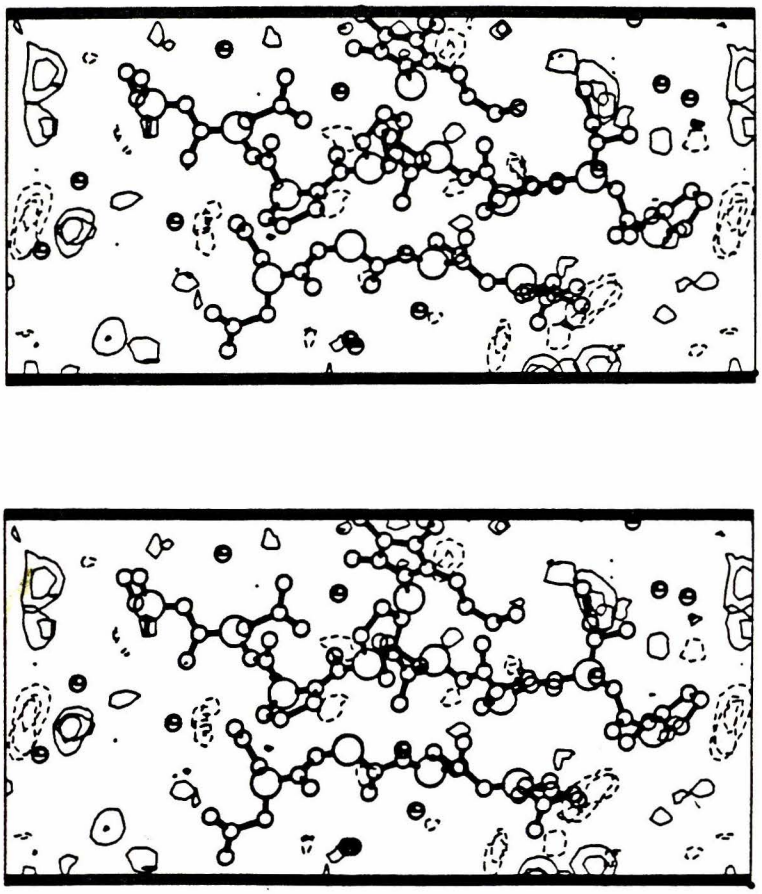


Figure 13c

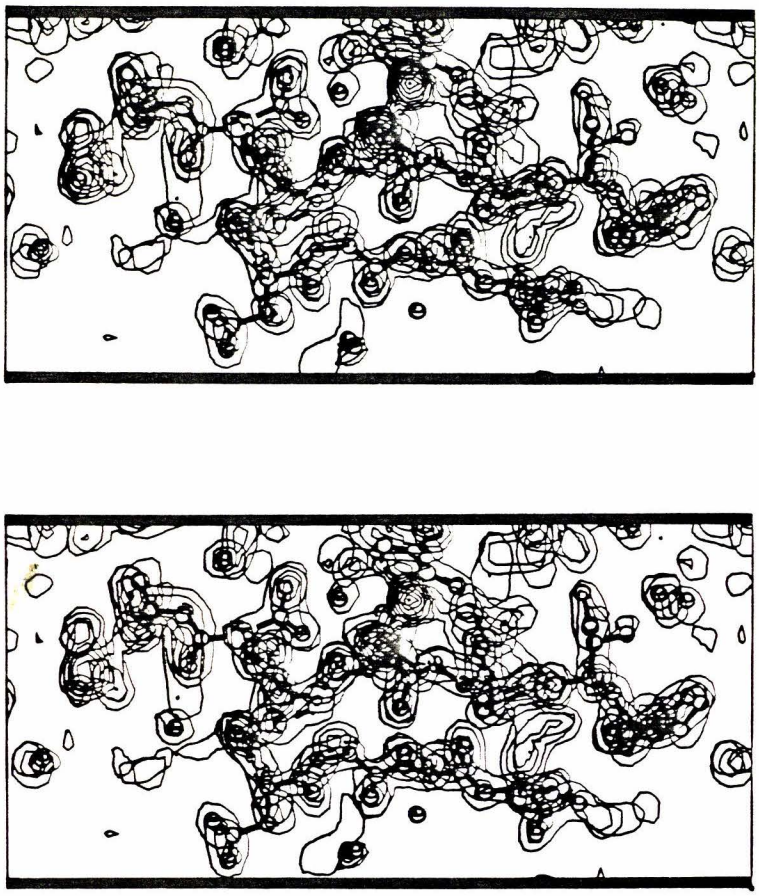


Figure 13d

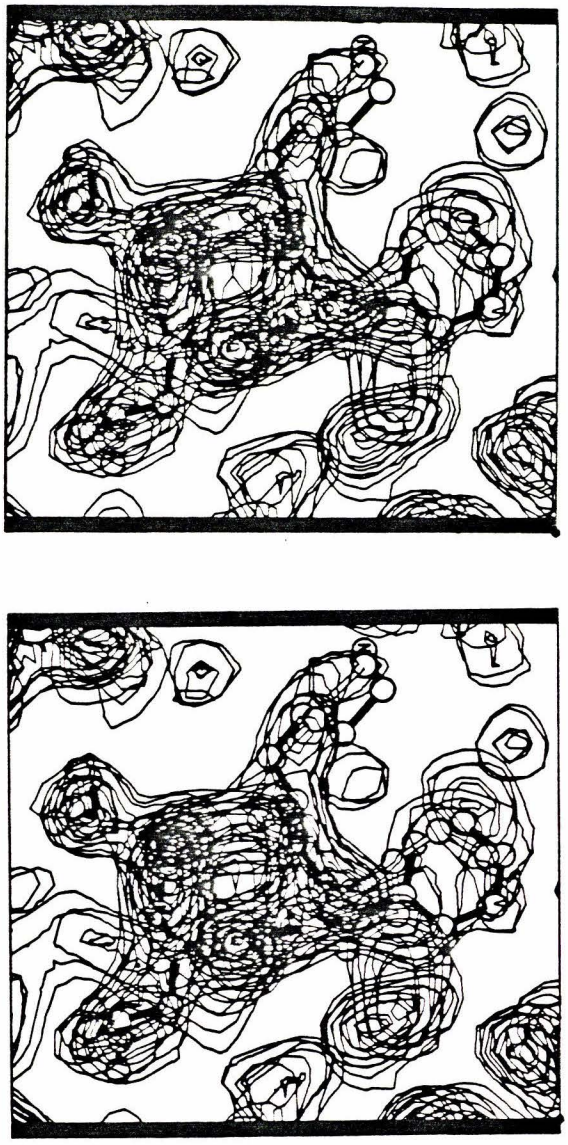


Figure 14a

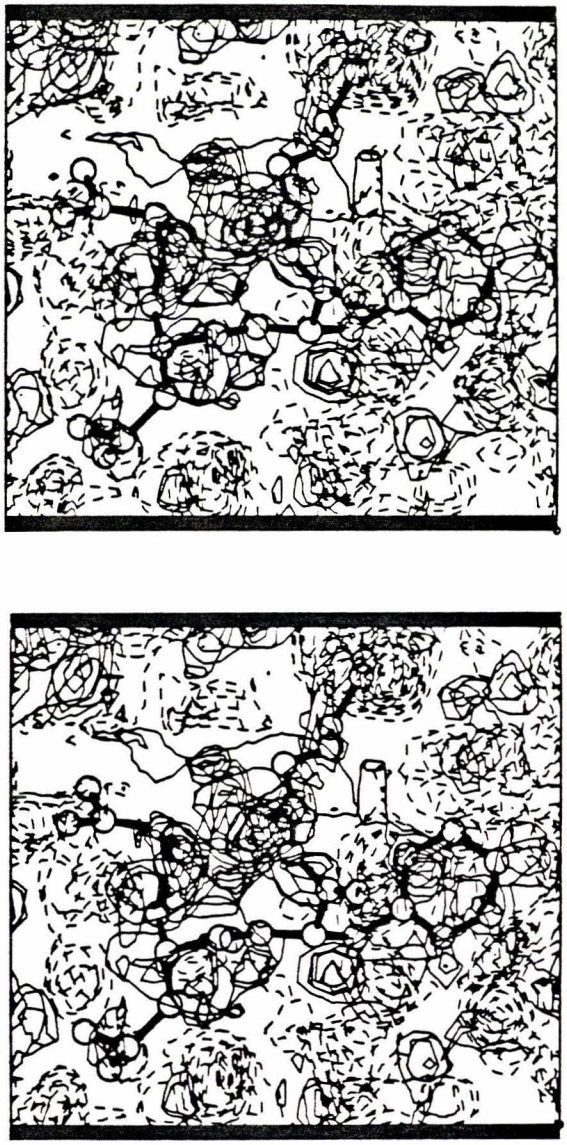


Figure 14b

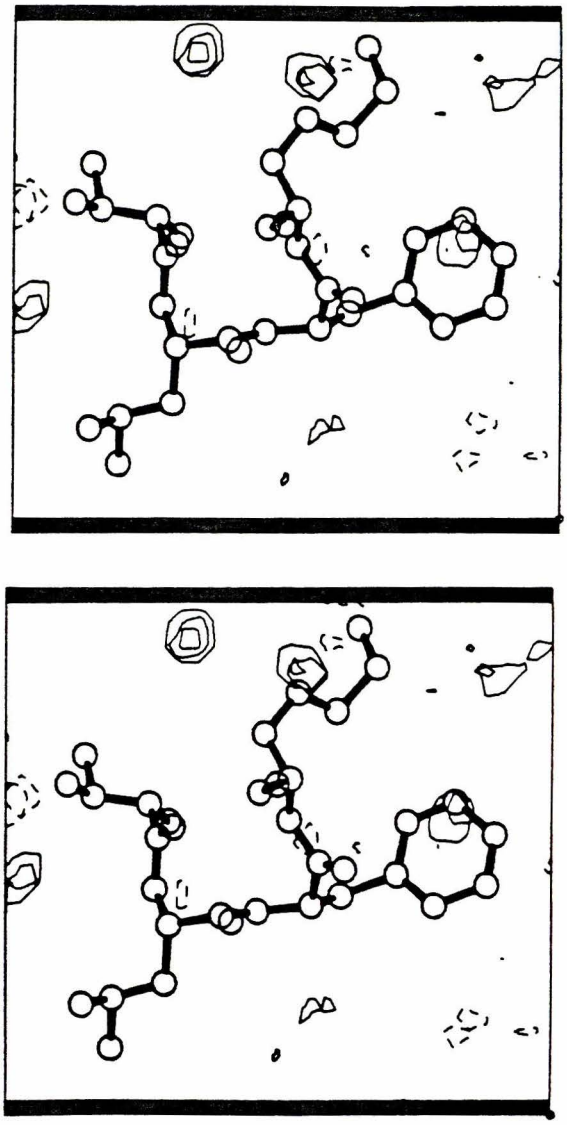


Figure 14c

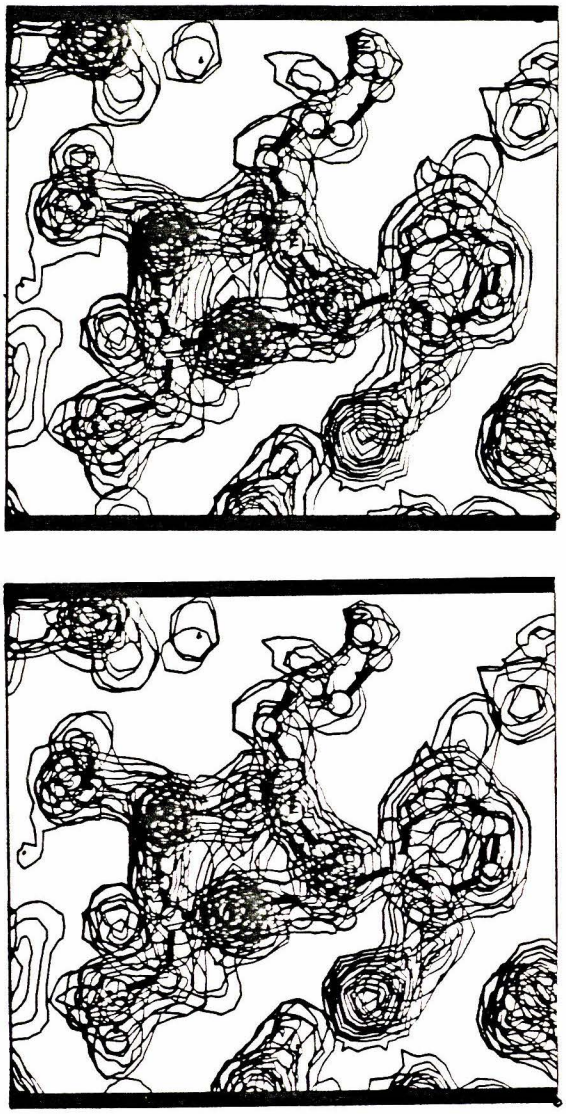


Figure 14d

Figure 15a

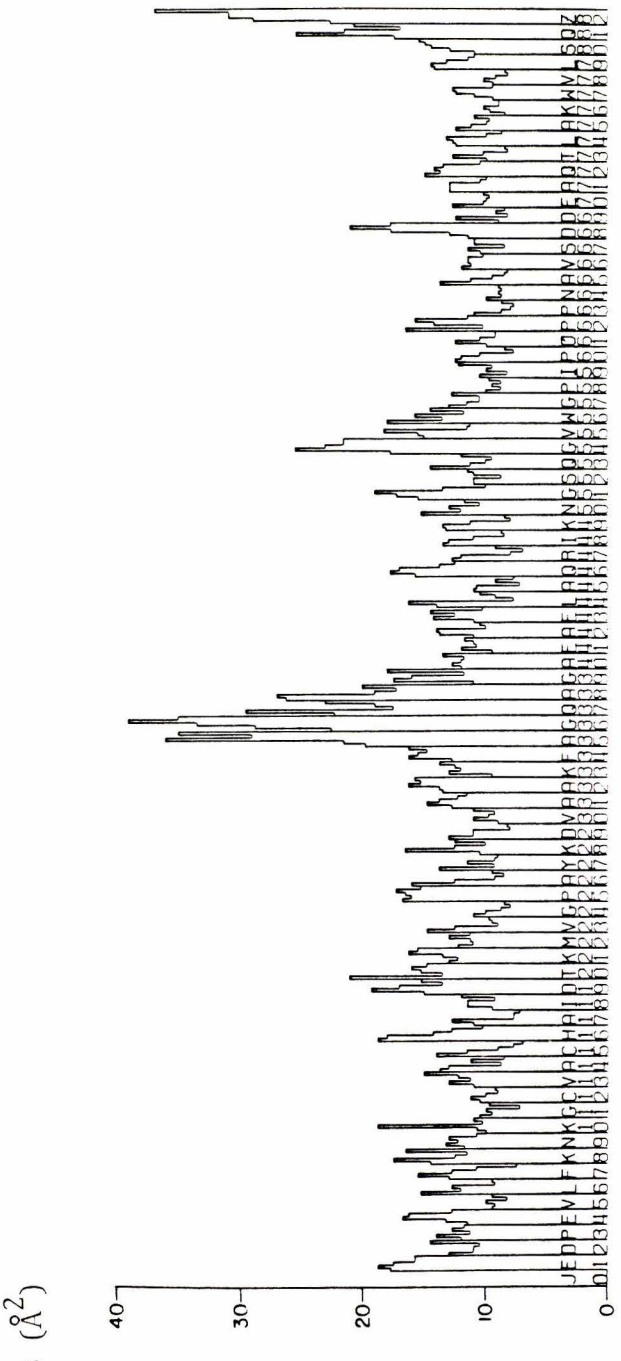


Figure 15b

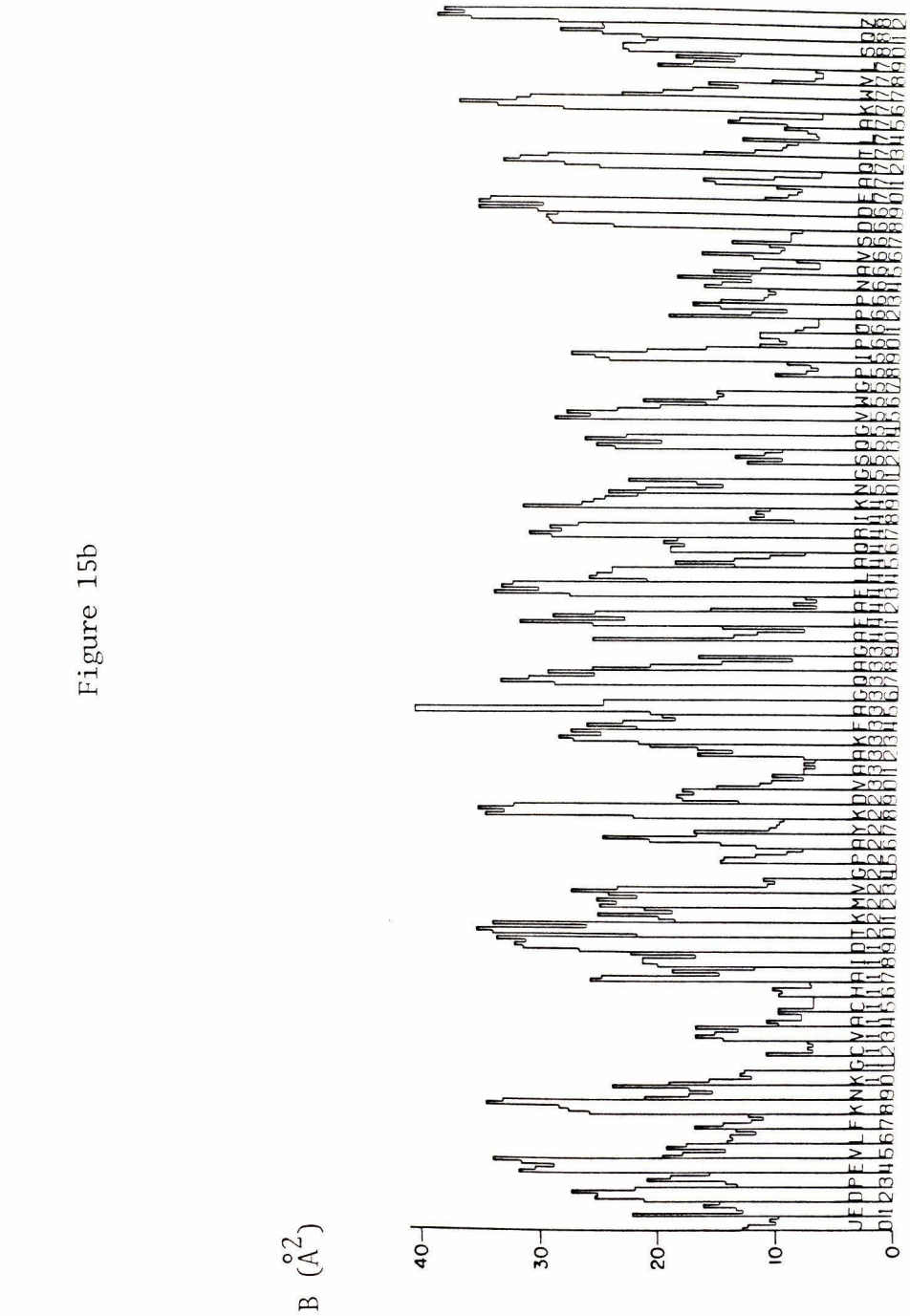
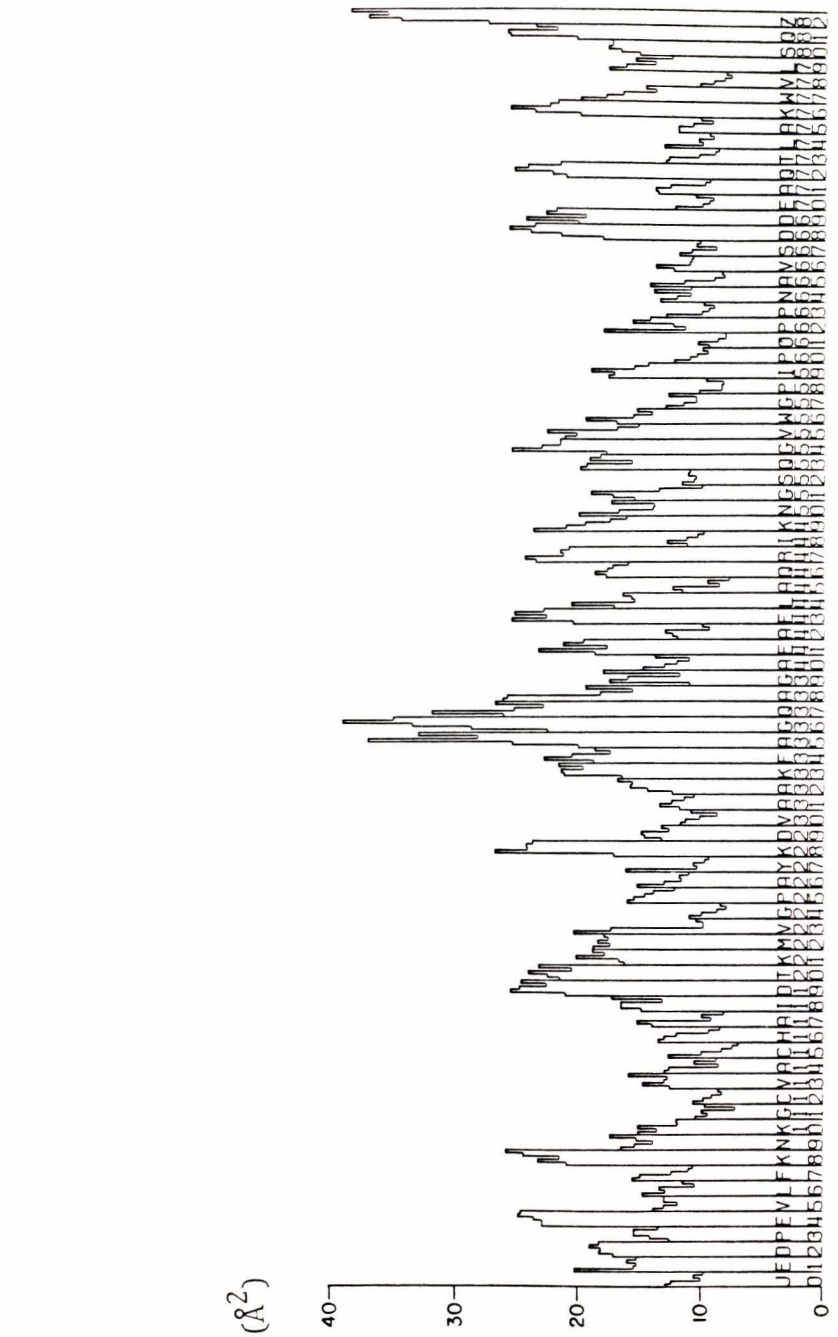


Figure 15c



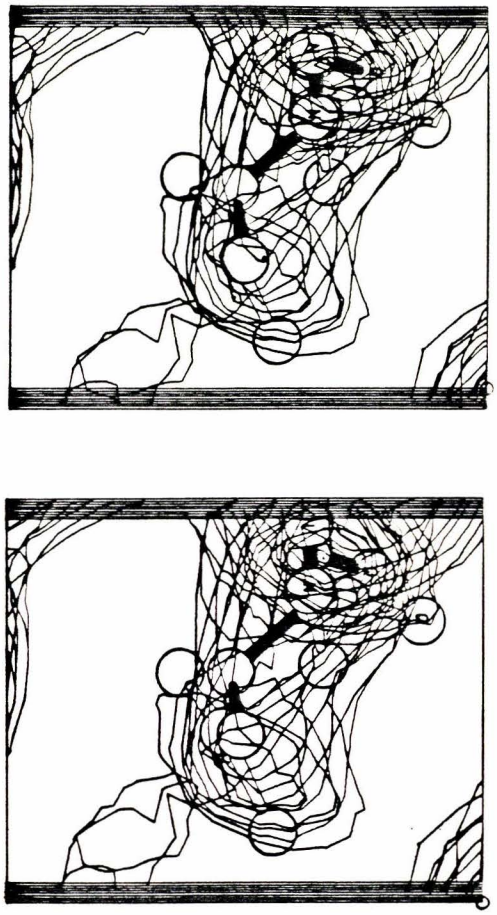


Figure 16a

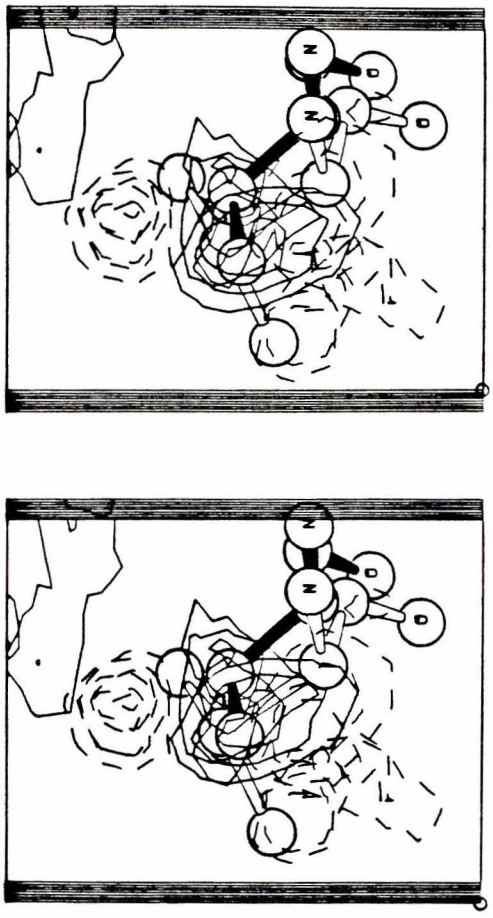


Figure 16b

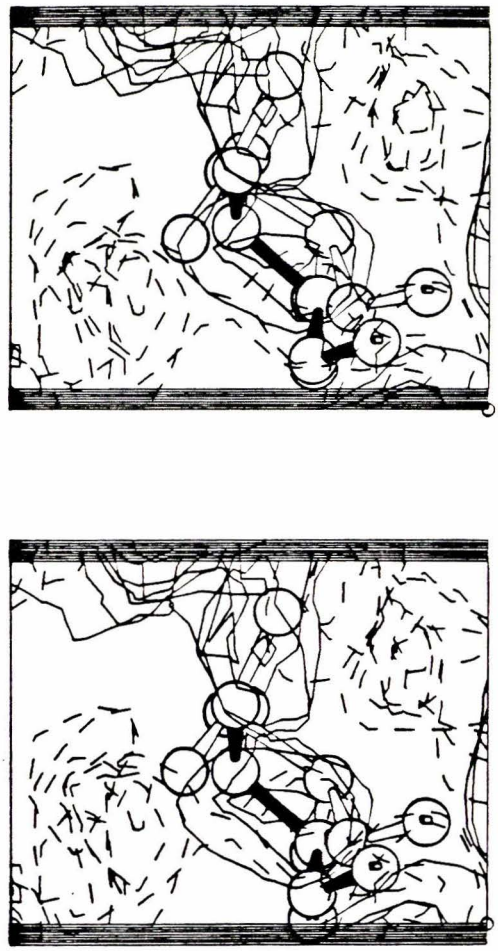


Figure 16c

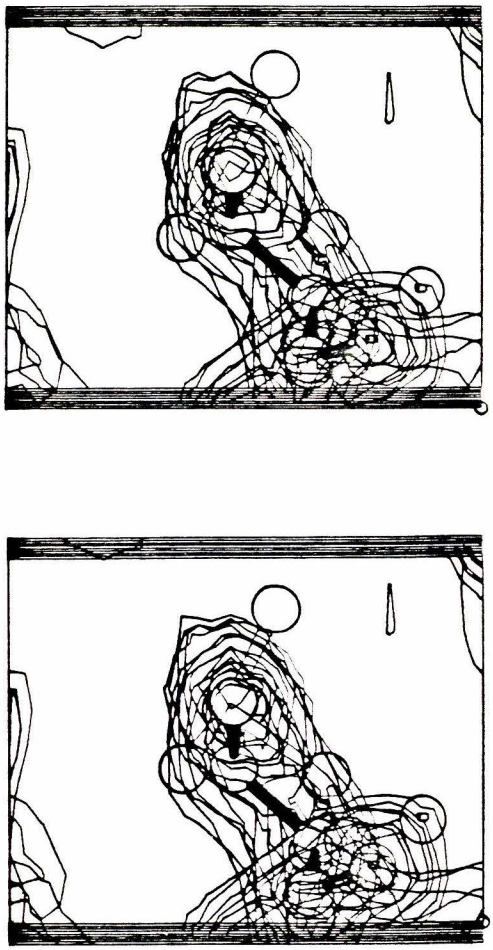


Figure 16d

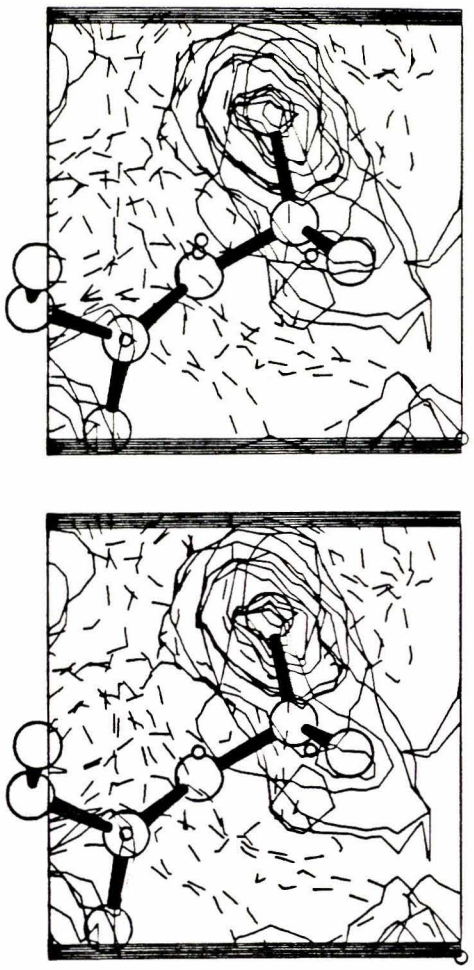


Figure 17a

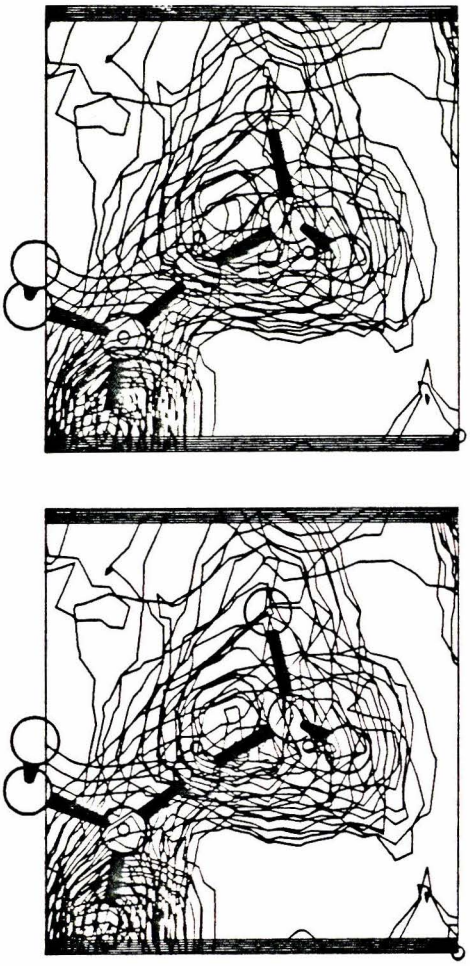


Figure 17b

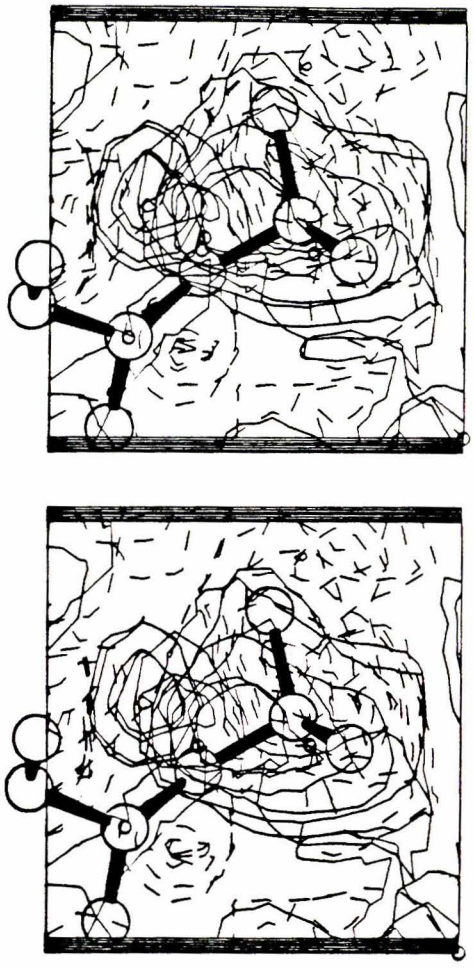


Figure 17c

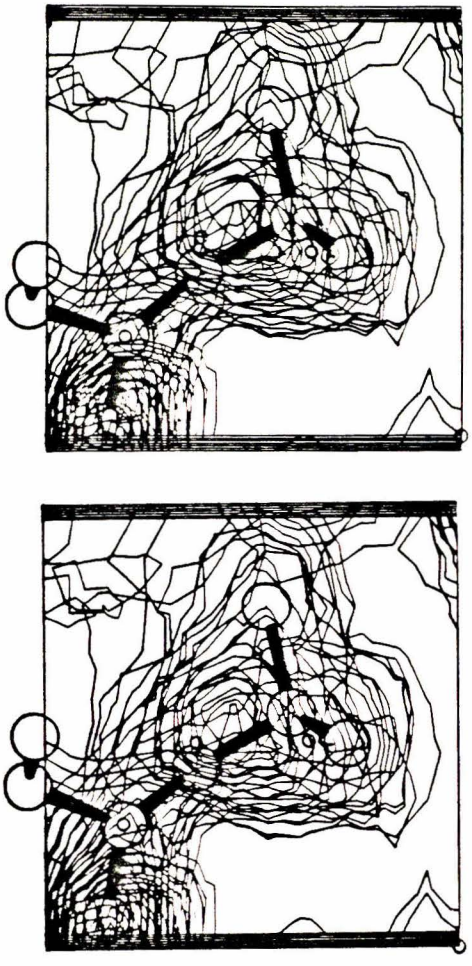


Figure 17d

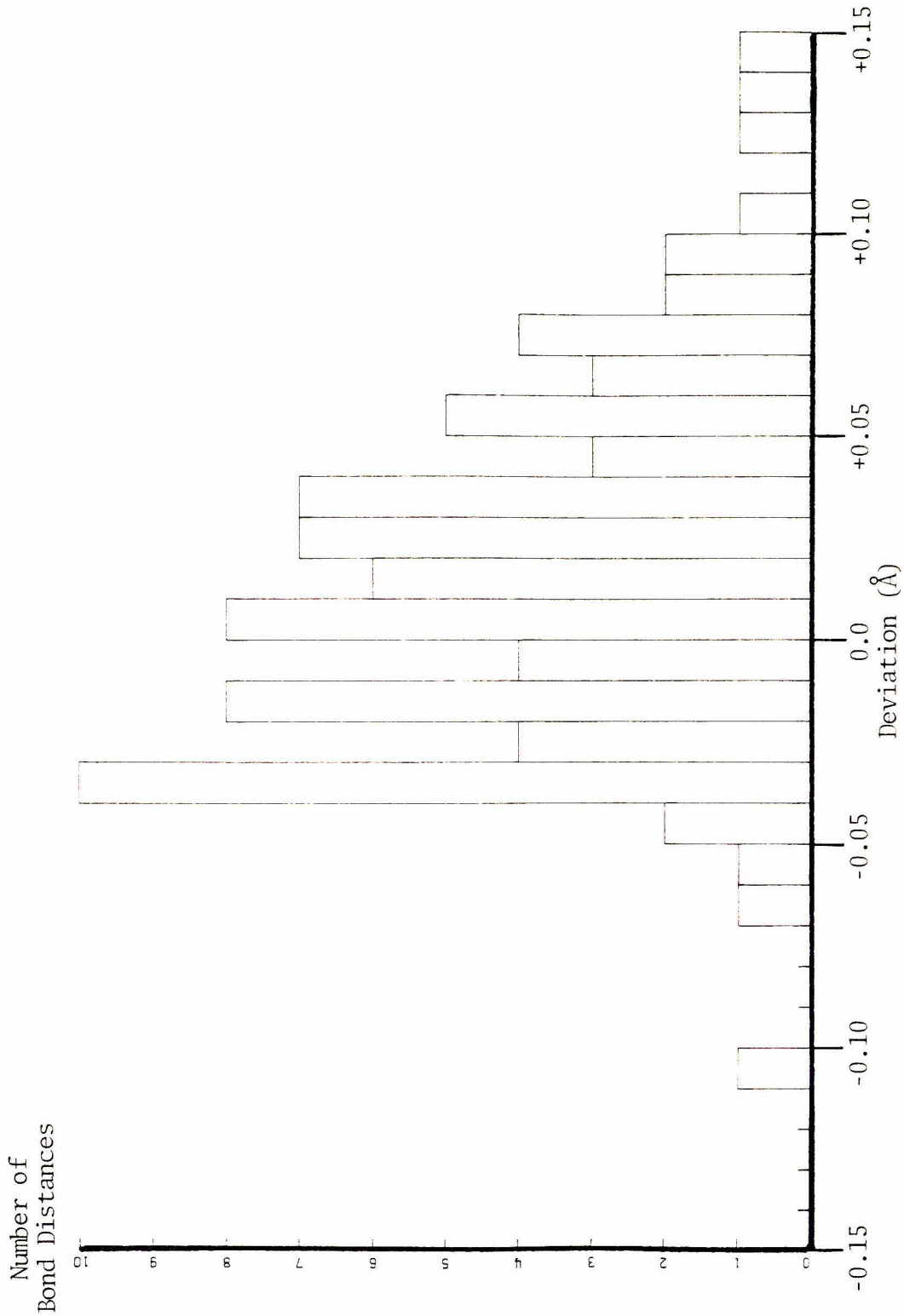


Figure 18a Distribution of N-C_α Bond Distance Deviations from Standard Values

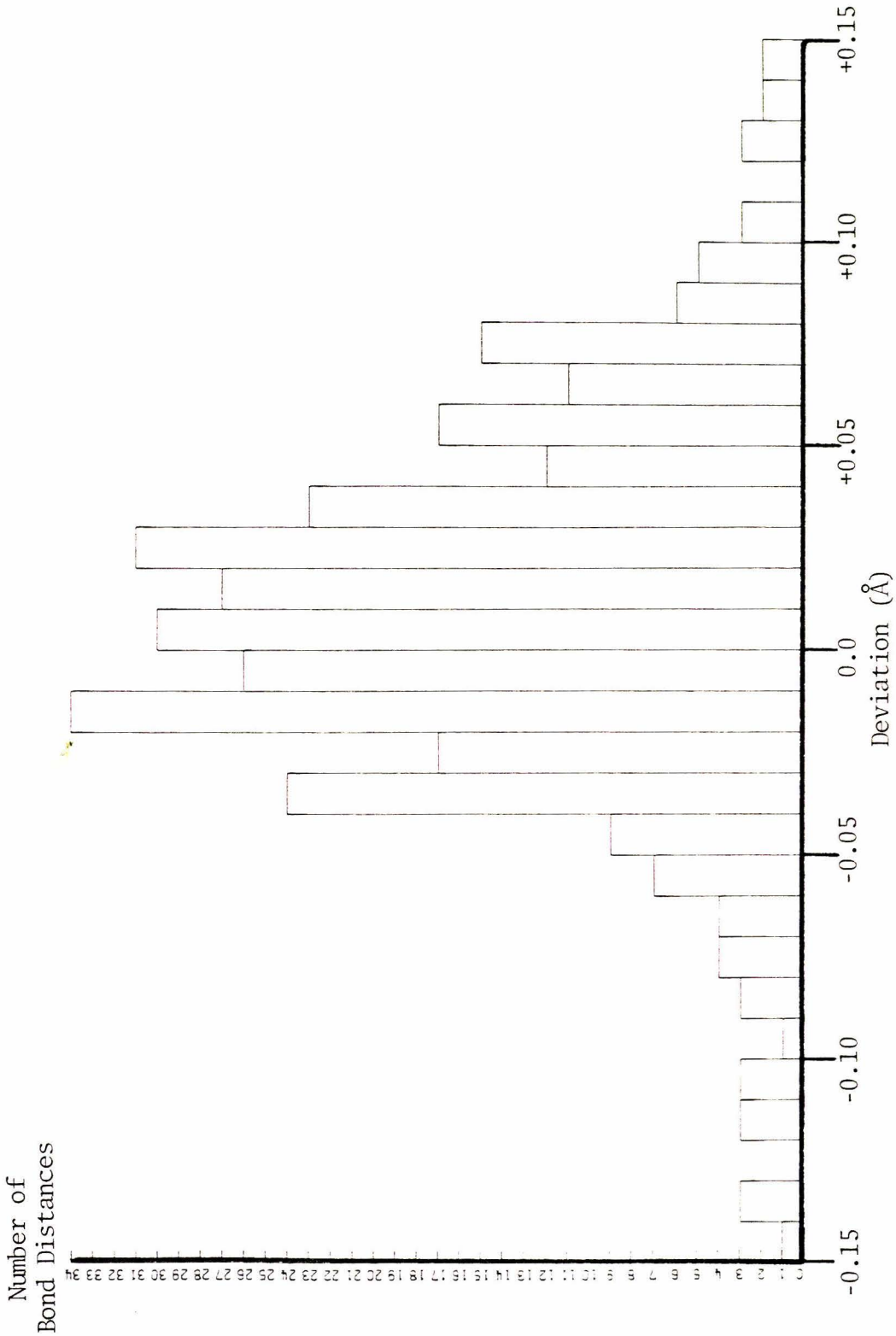


Figure 18b Distribution of All Main Chain Bond Distance Deviations from Standard Values

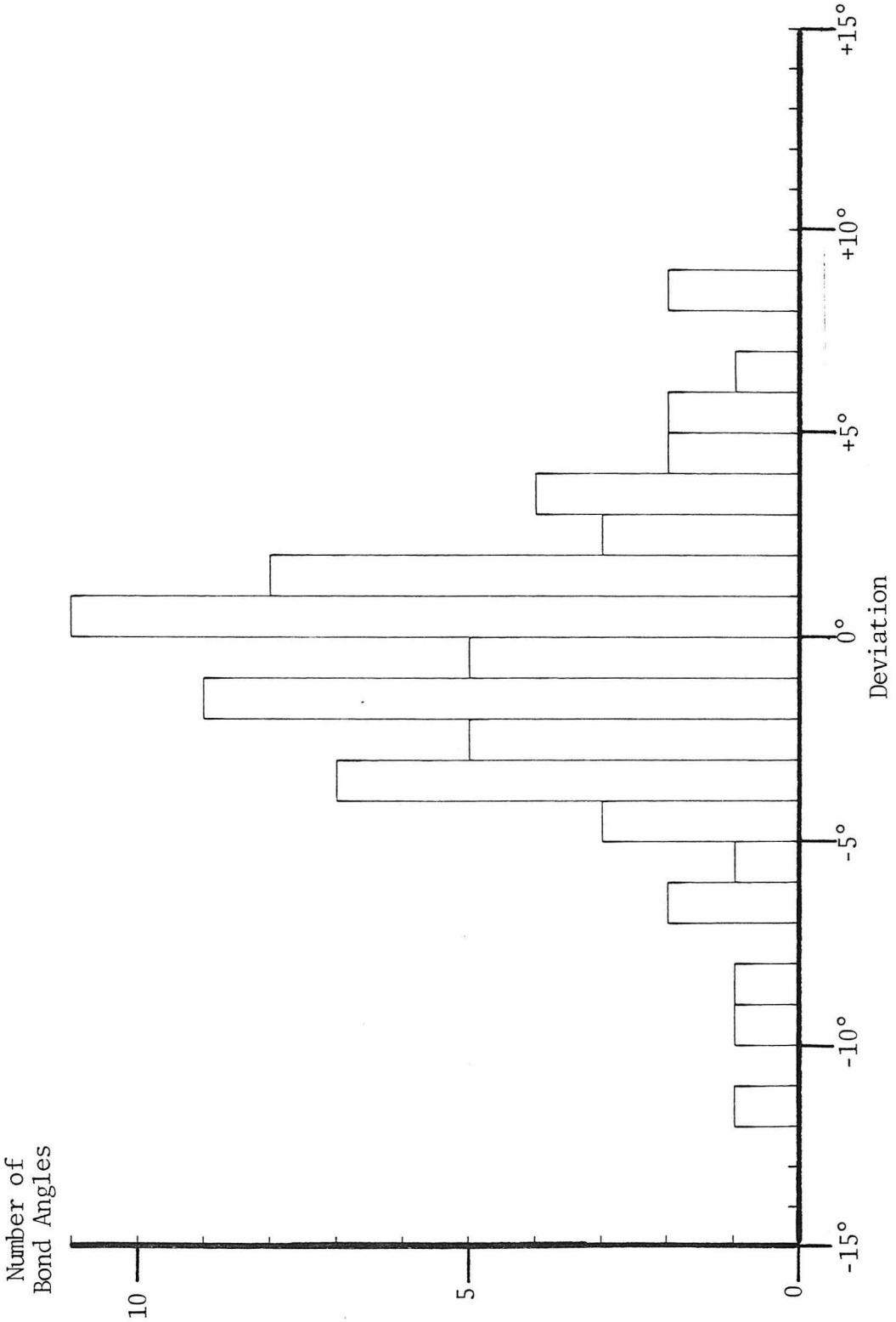


Figure 18c Distribution of C_α-C_β-C_γ Bond Angle Deviations from Standard Values

Figure 18d Distribution of Main Chain C_α-C-N-C_α and O-C-N-C_α Torsion Angle Deviations from Standard Values

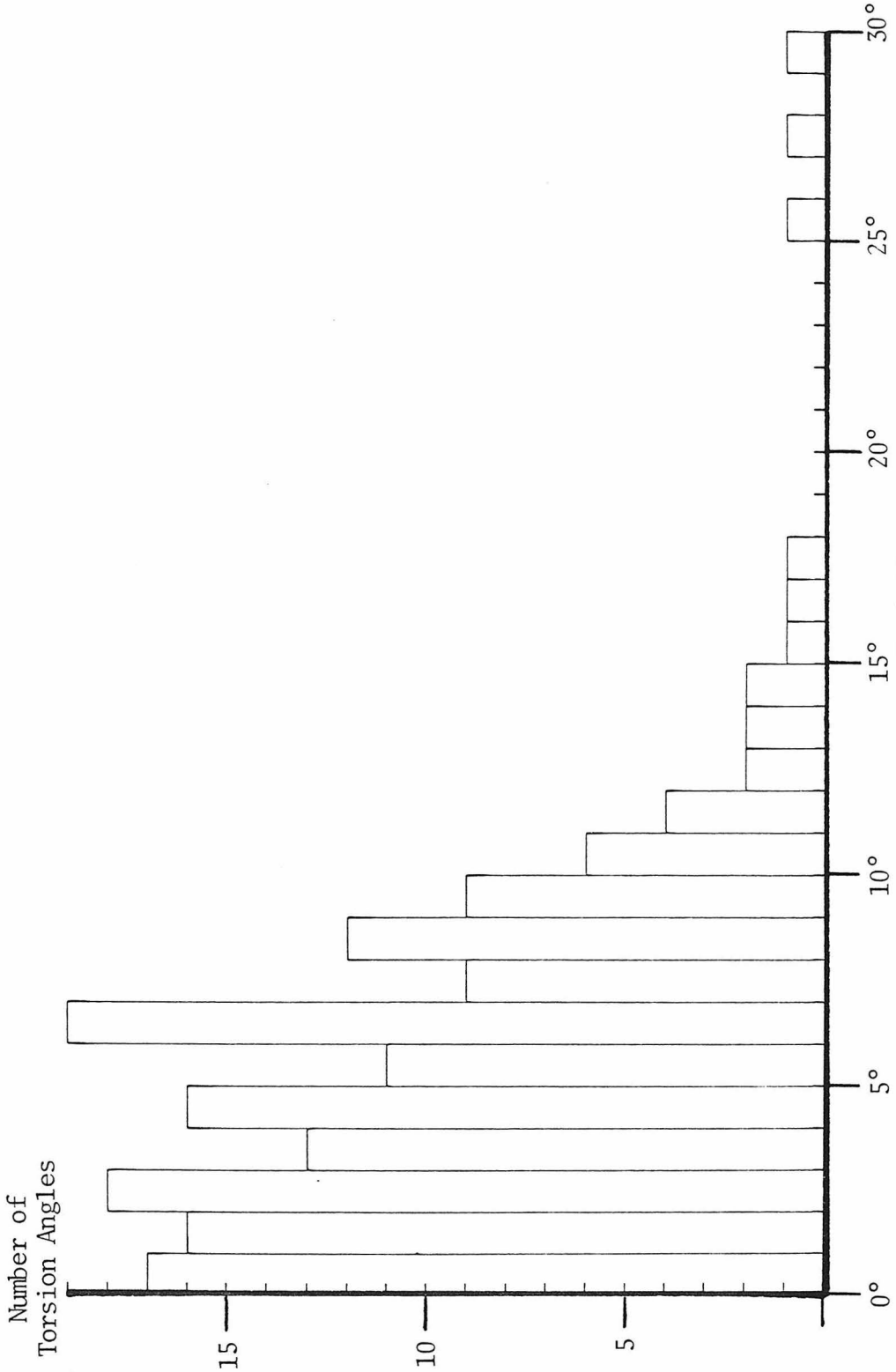


Figure 19a

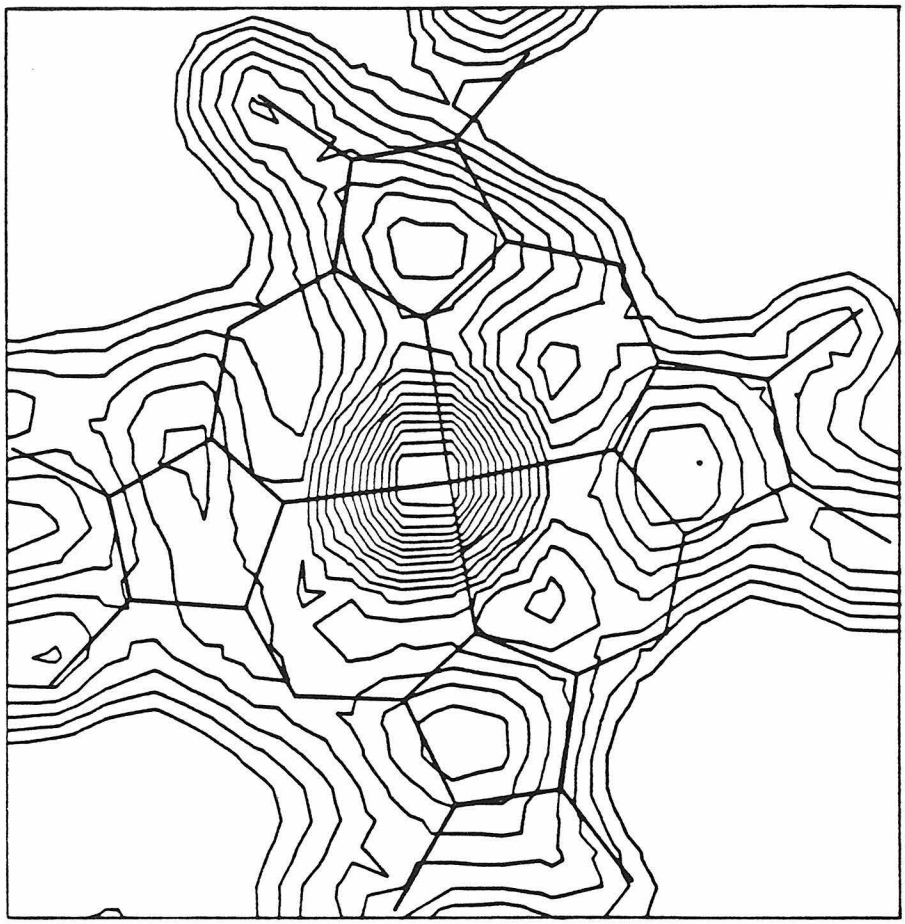


Figure 19b

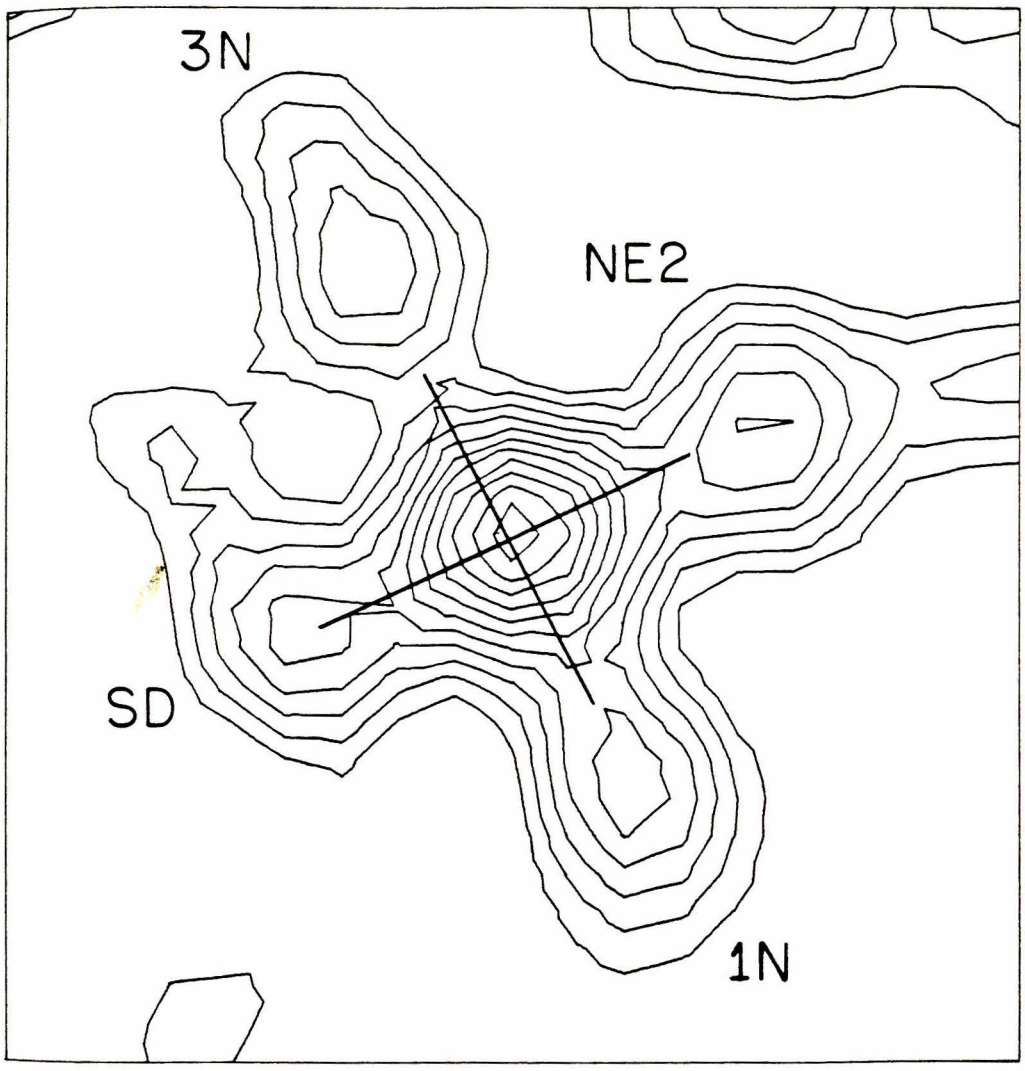


Figure 19c

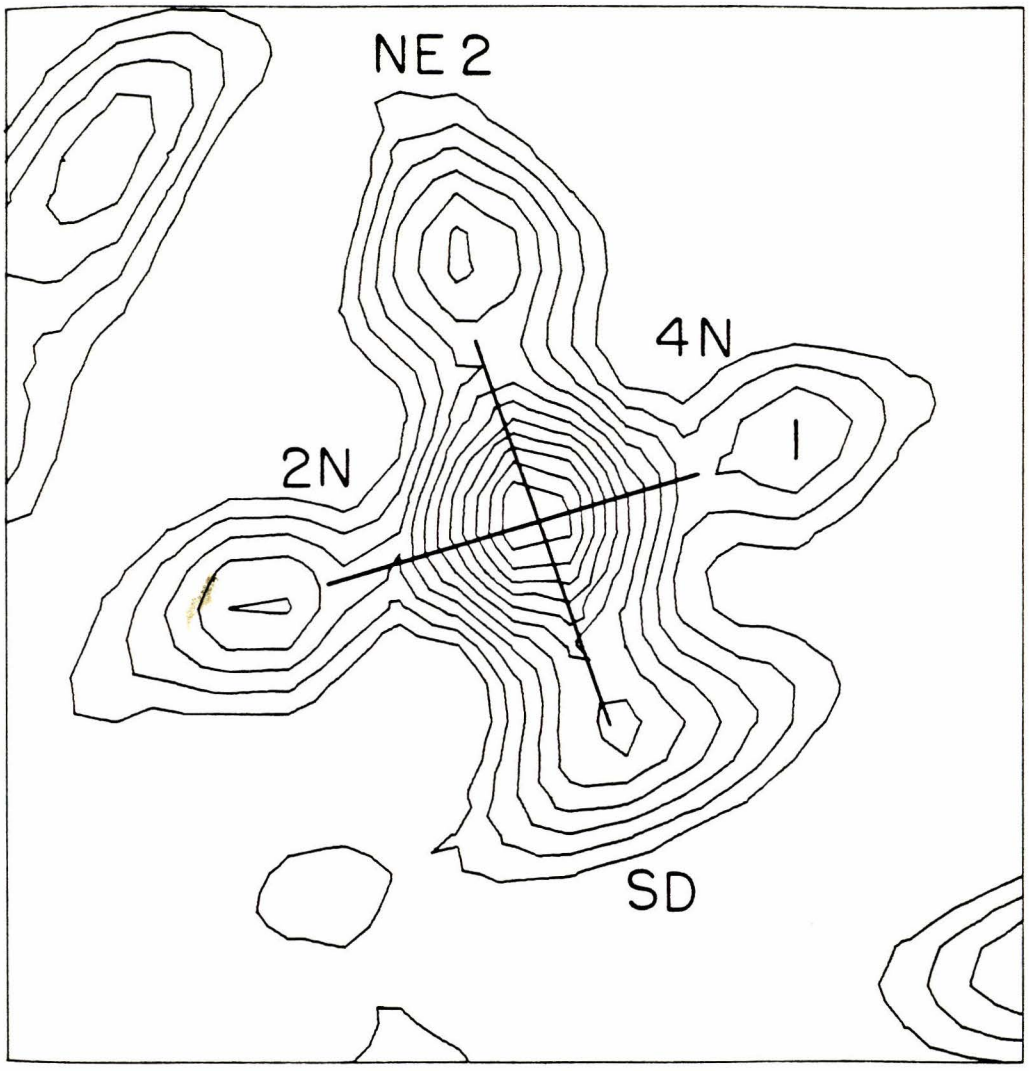


Figure 20a

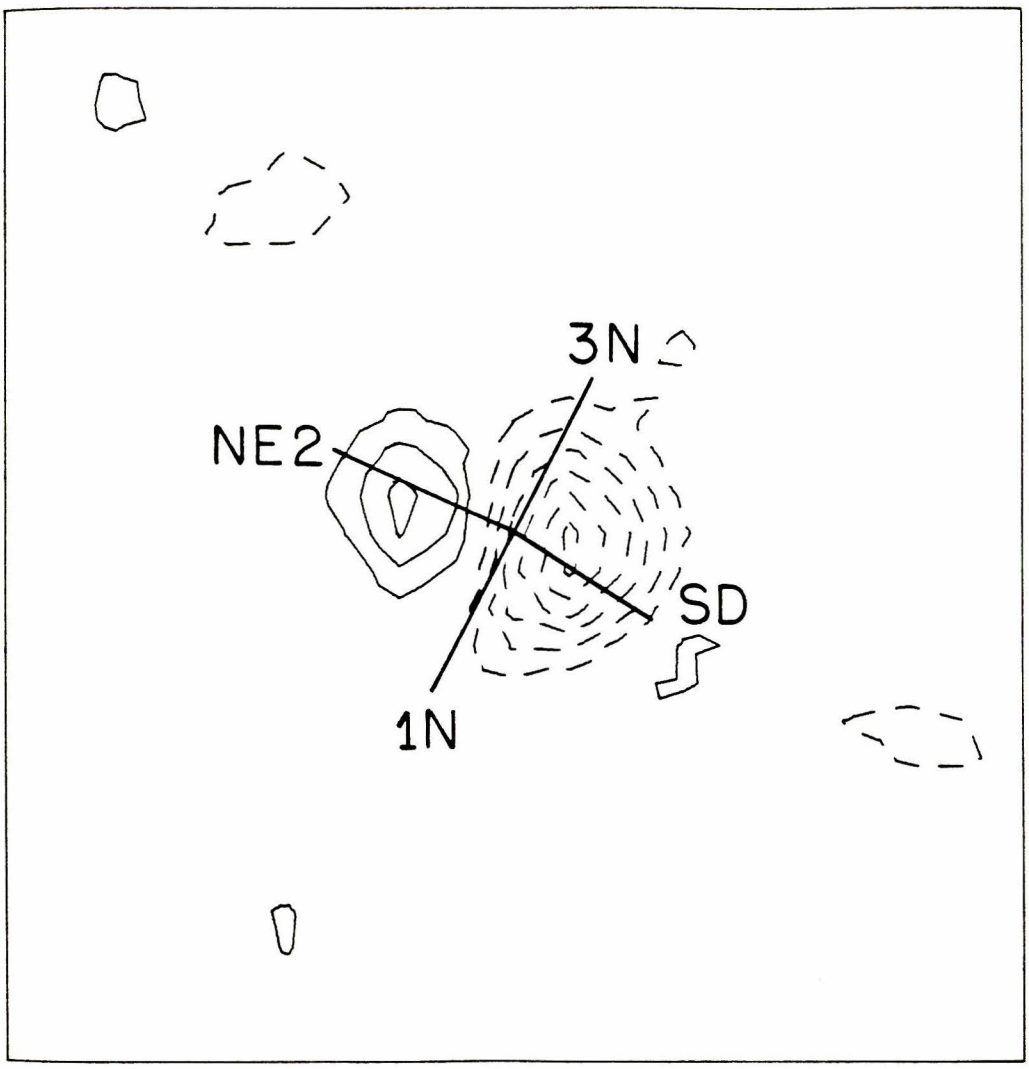


Figure 20b

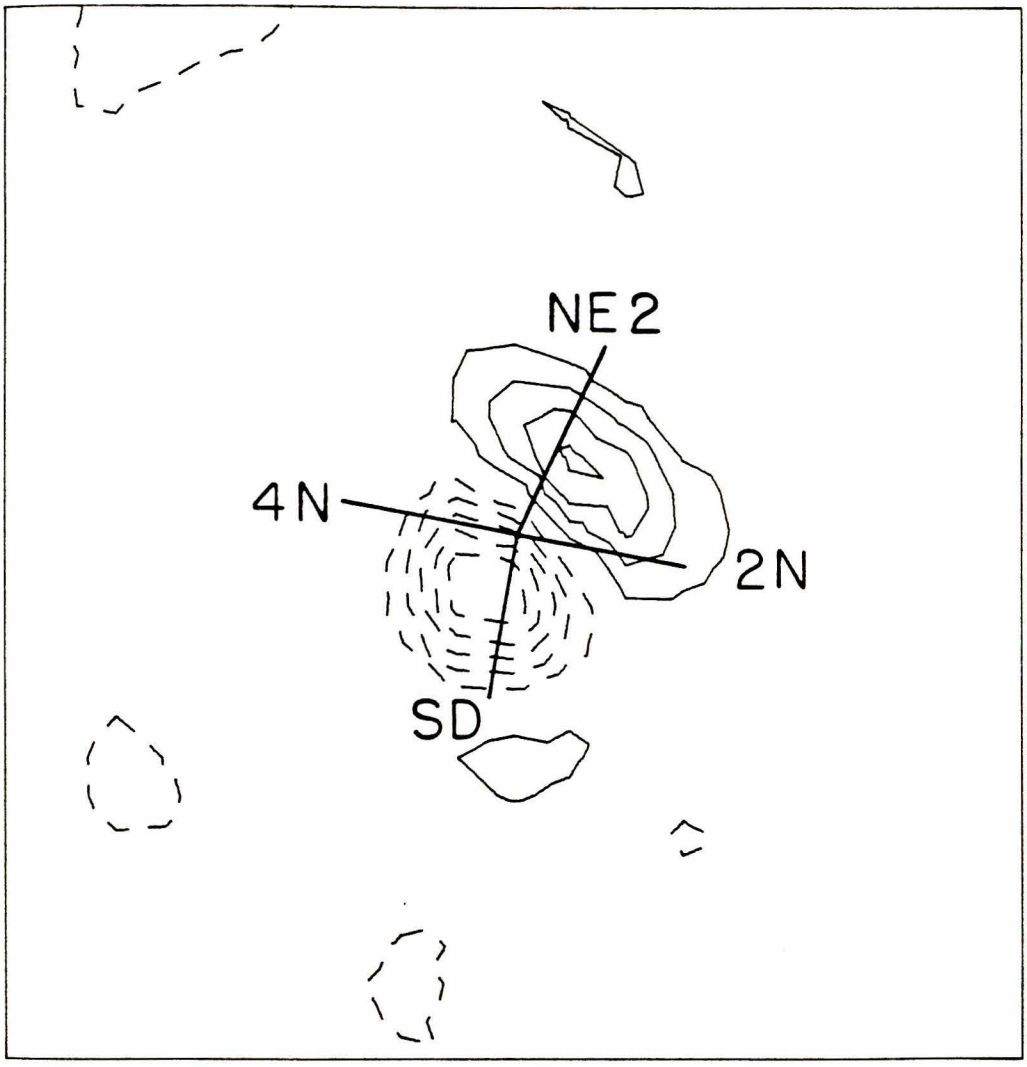


Figure 20c

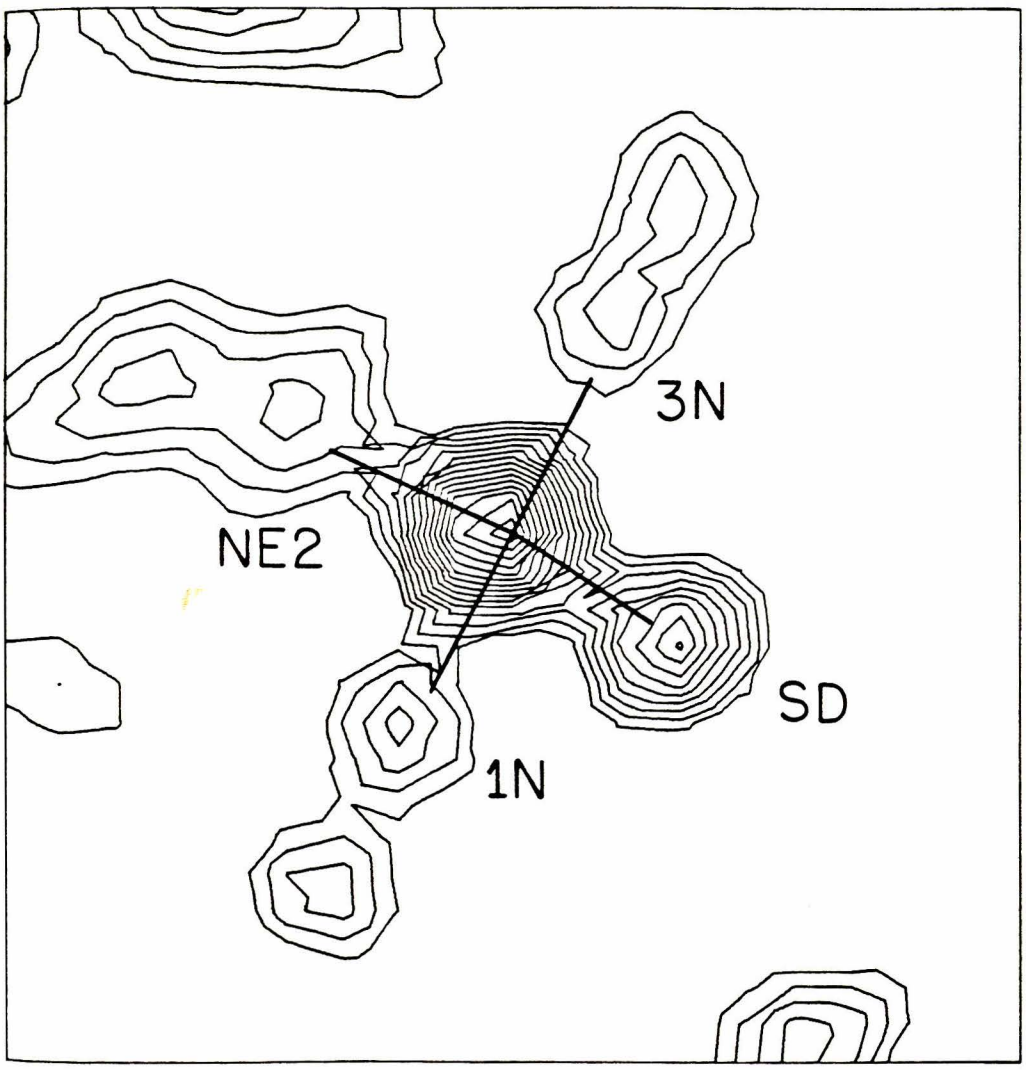


Figure 20d

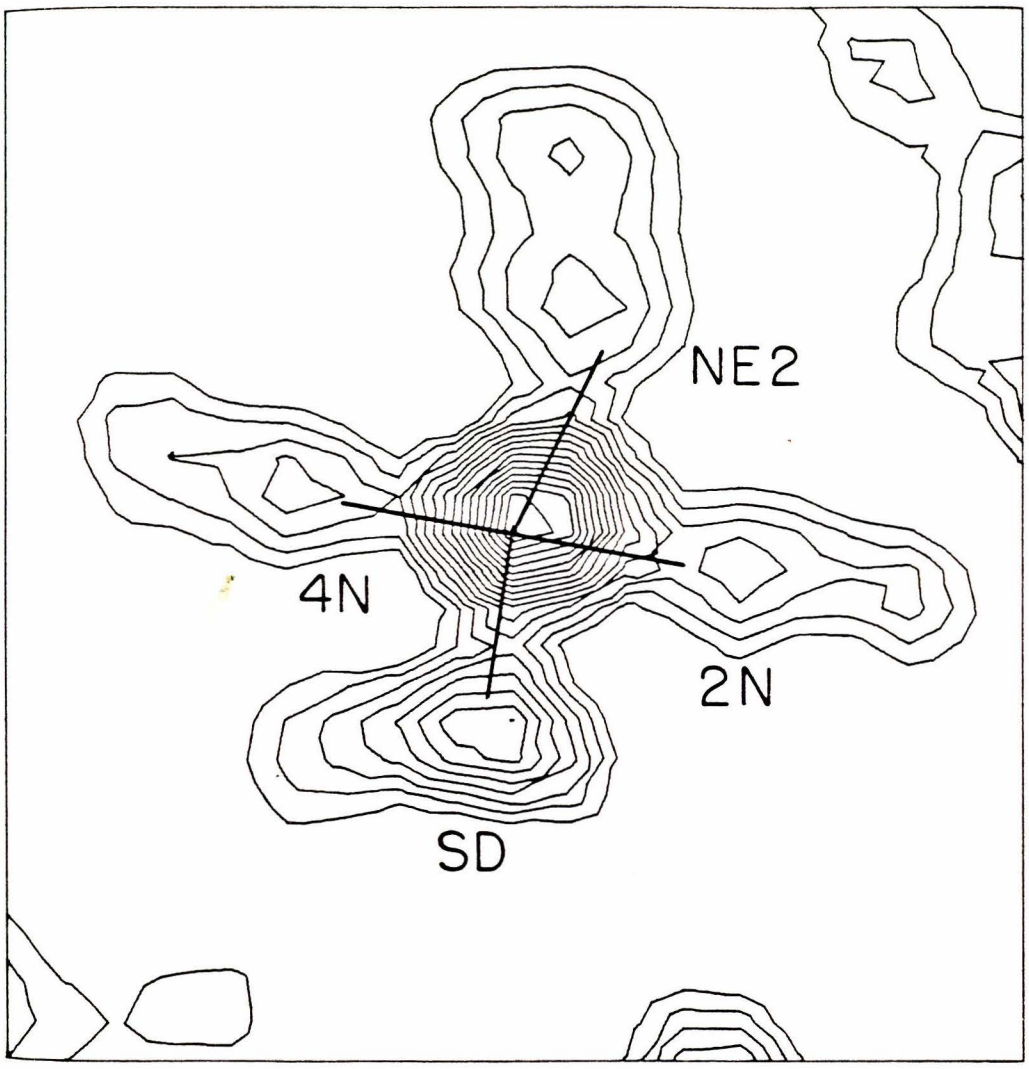


Figure 21a

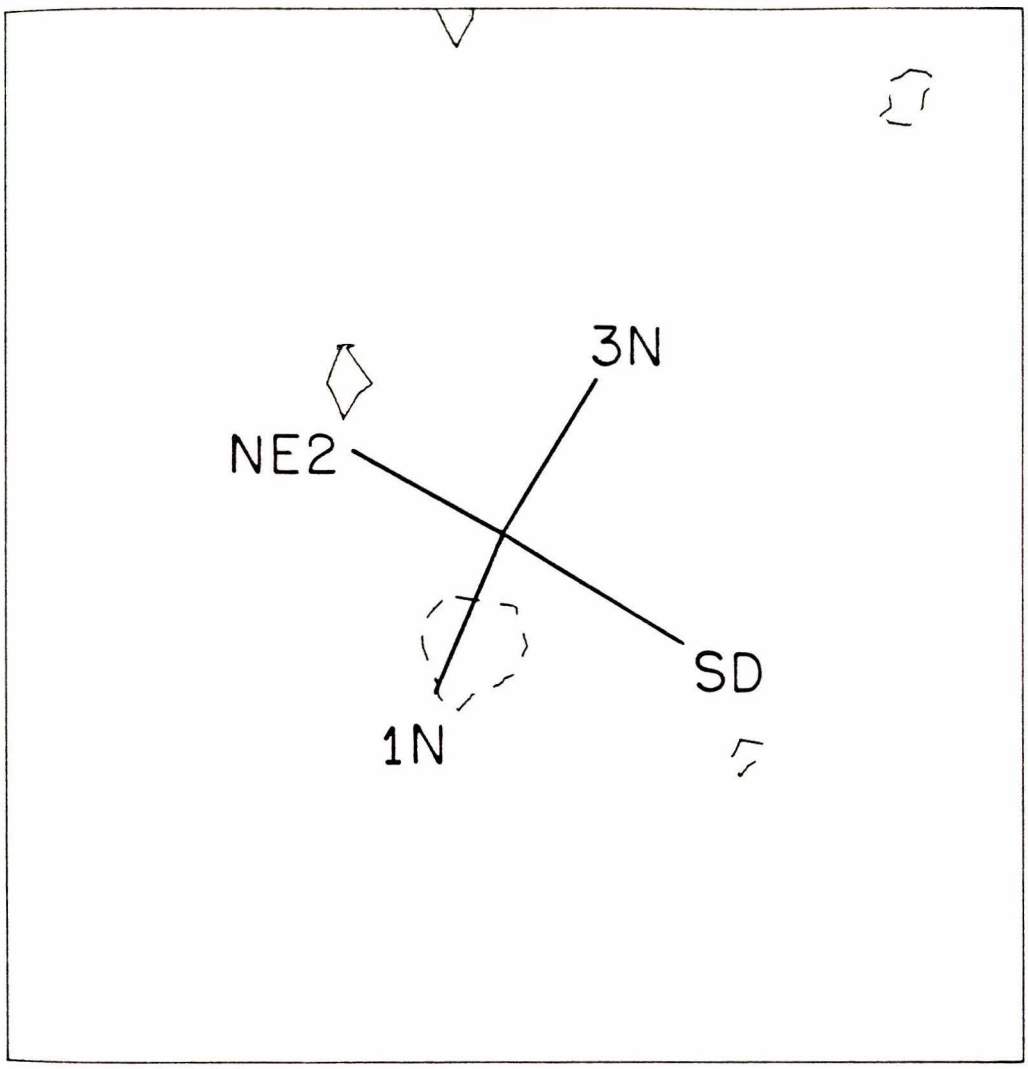


Figure 21b

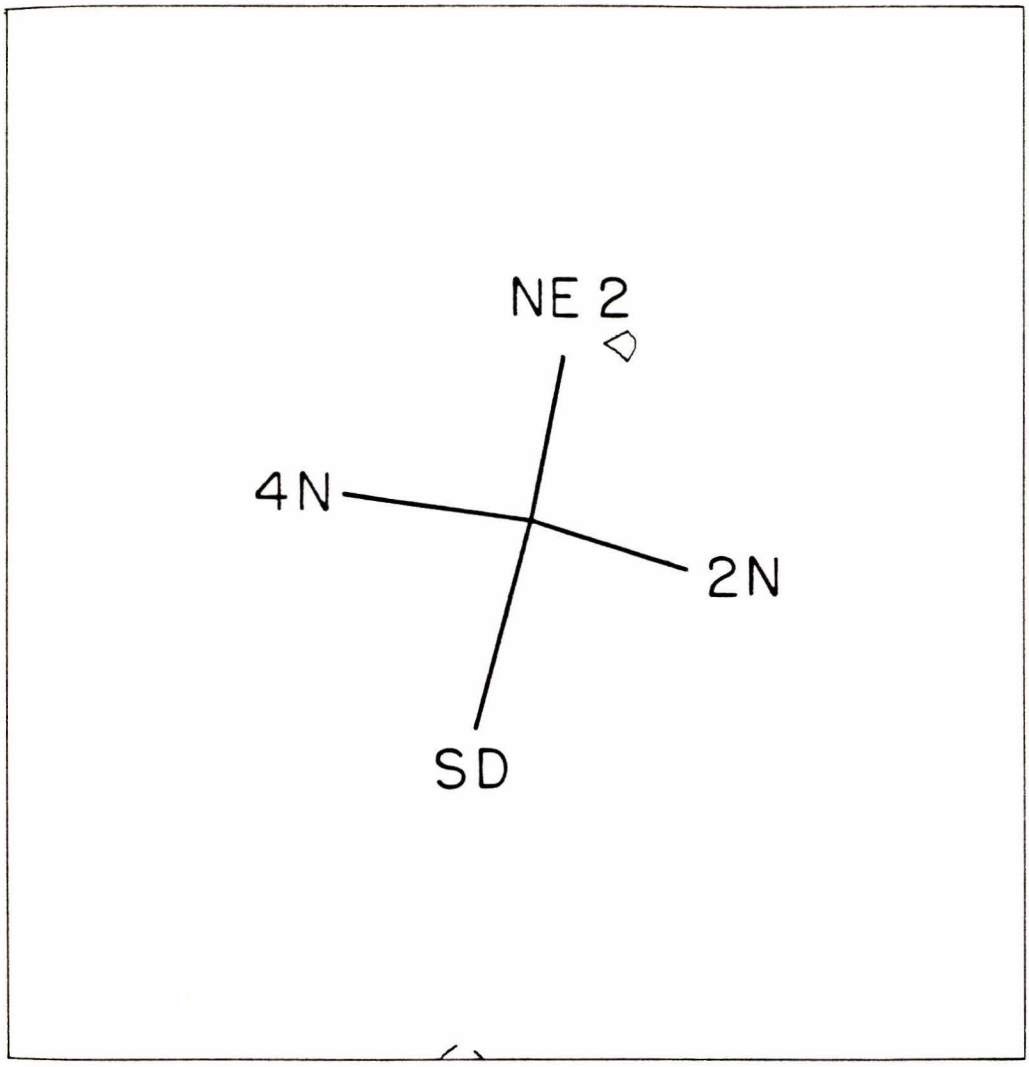


Figure 21c

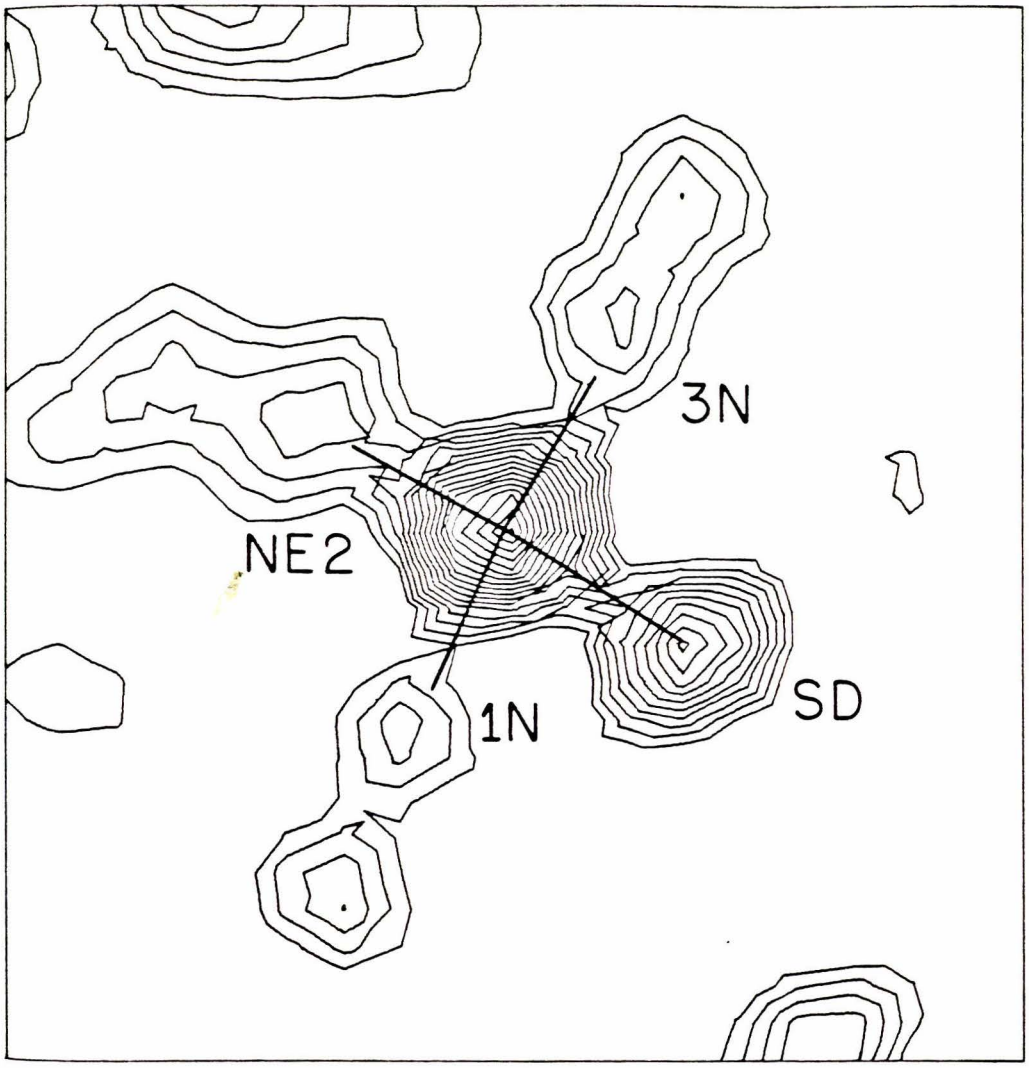


Figure 21d

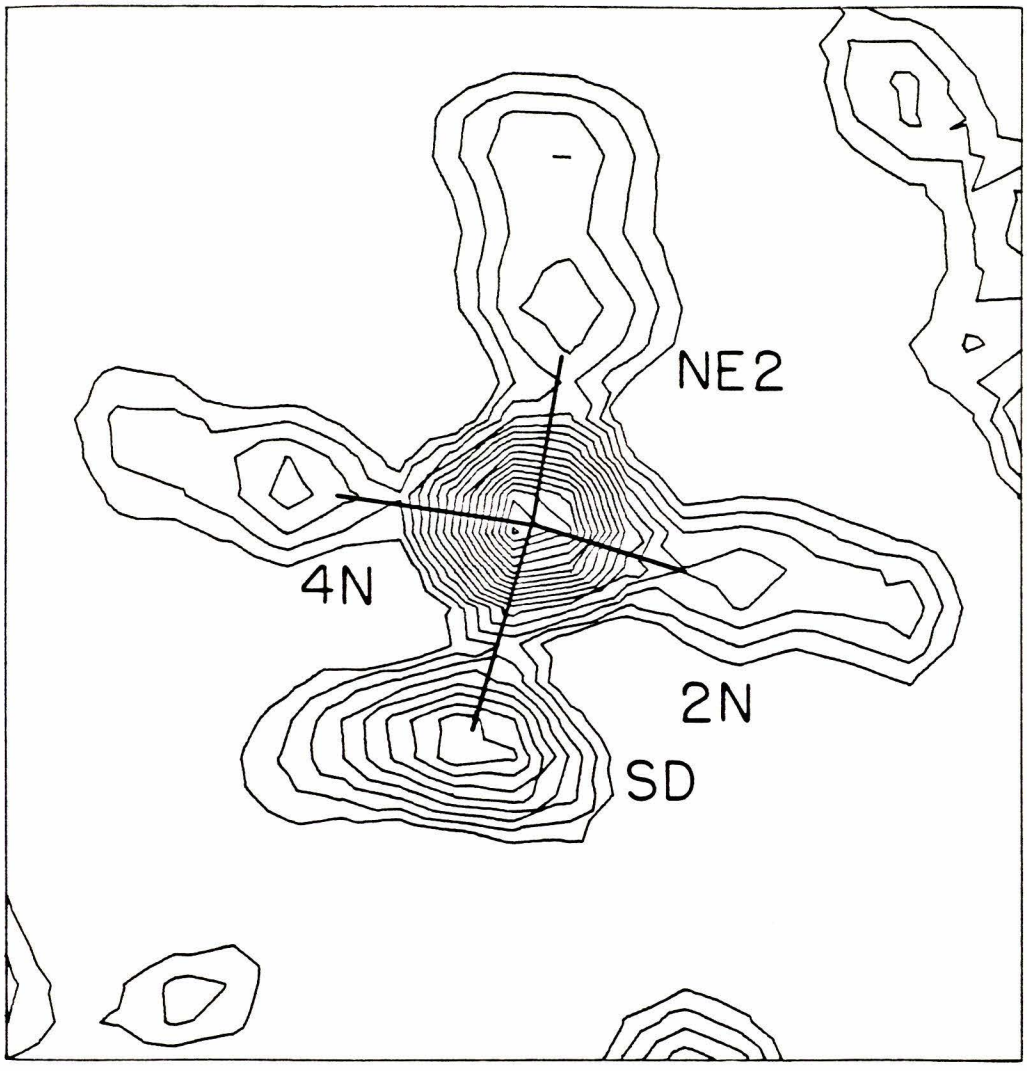


Figure 21e

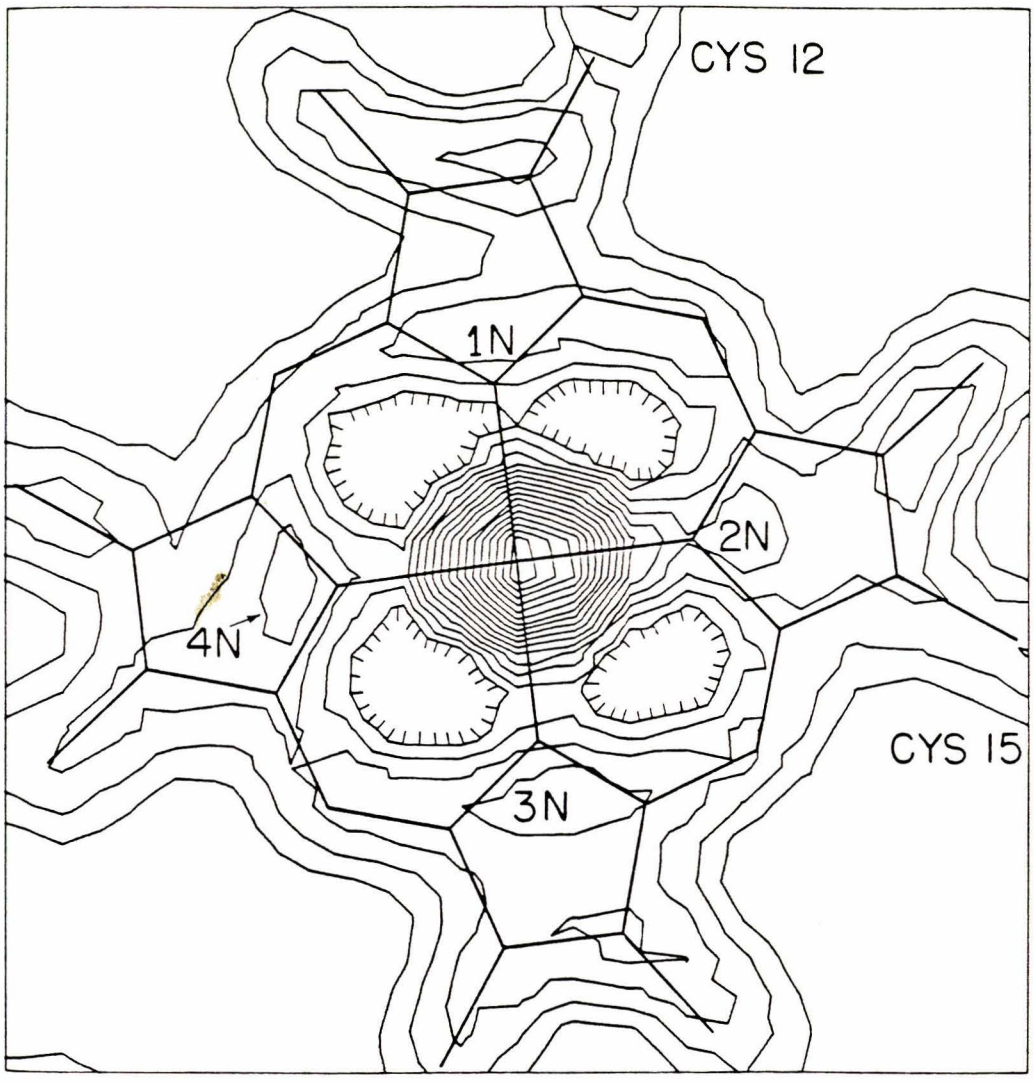


Figure 22a

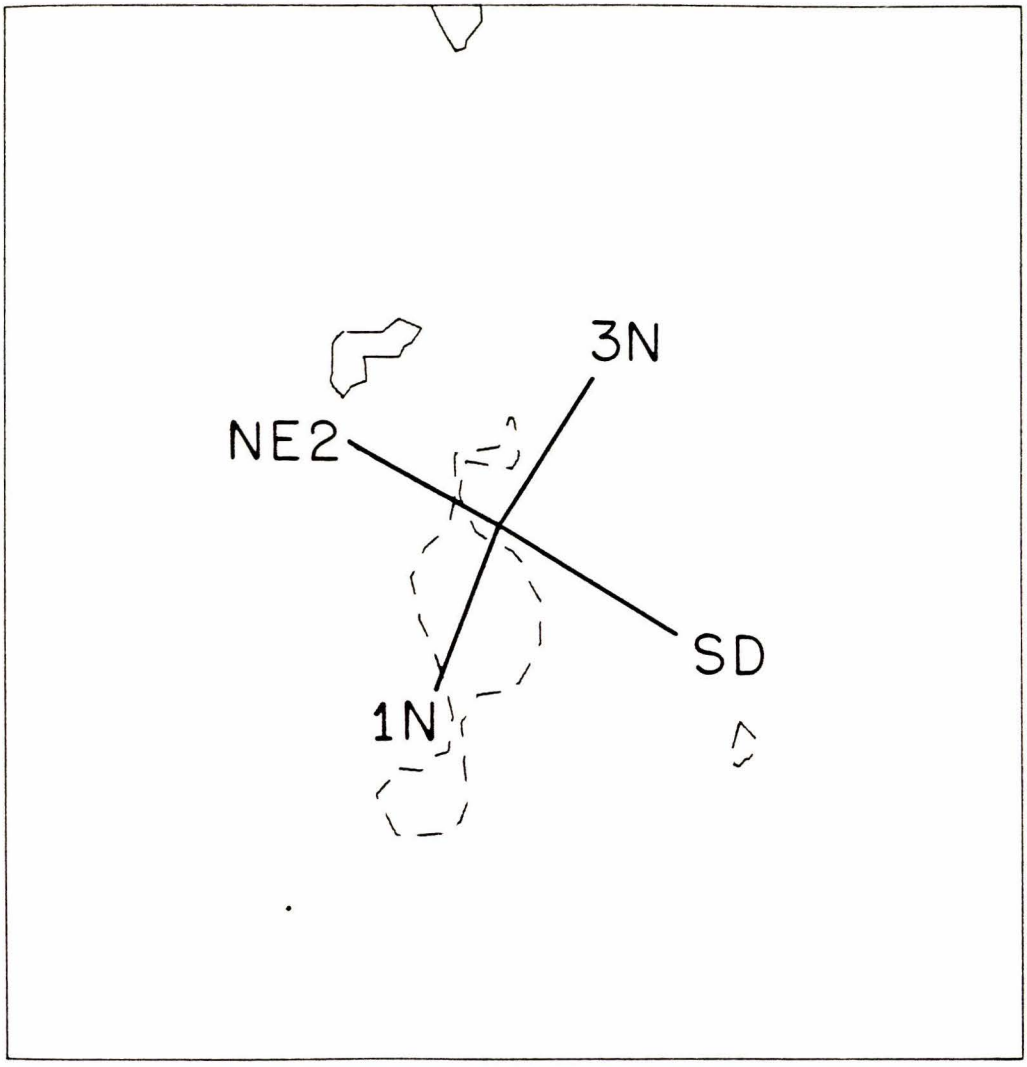


Figure 22b

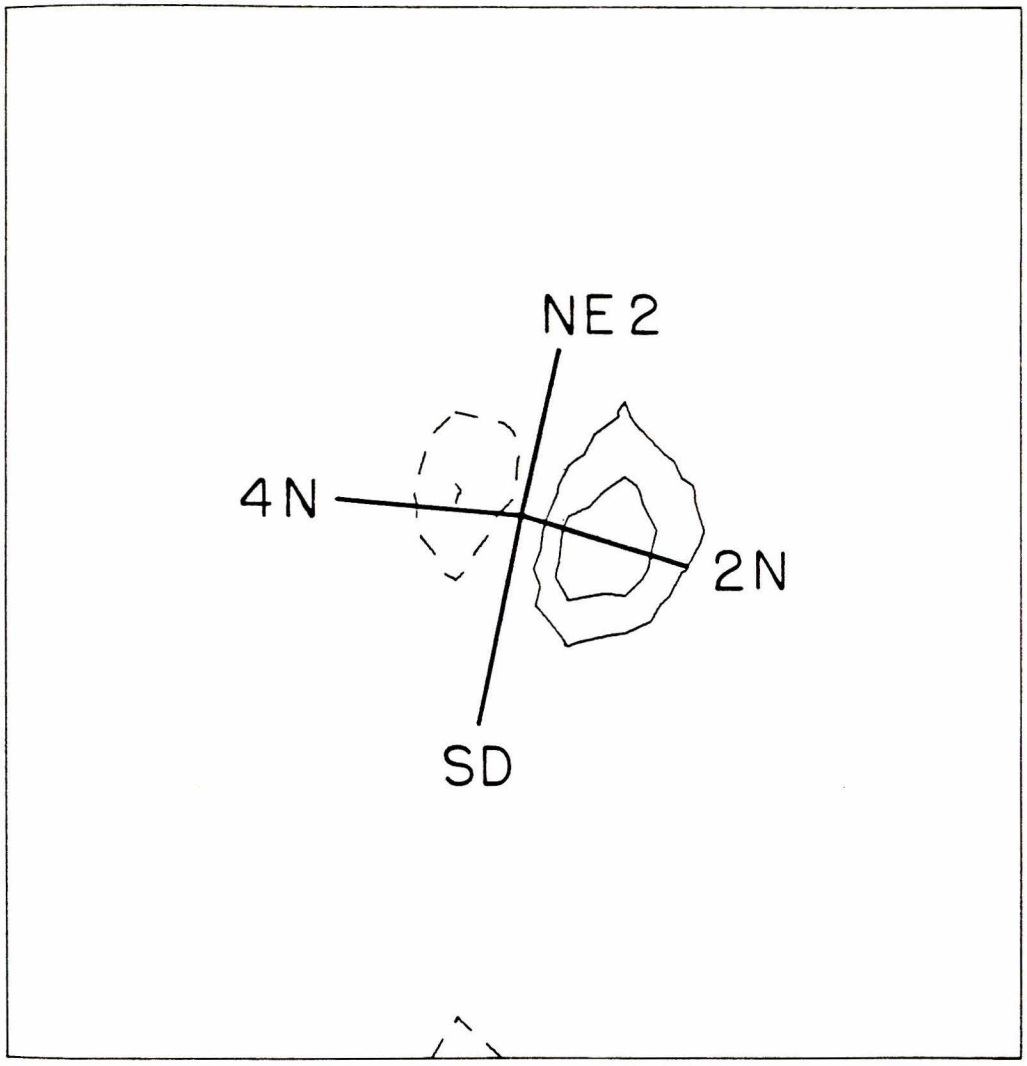


Figure 22c

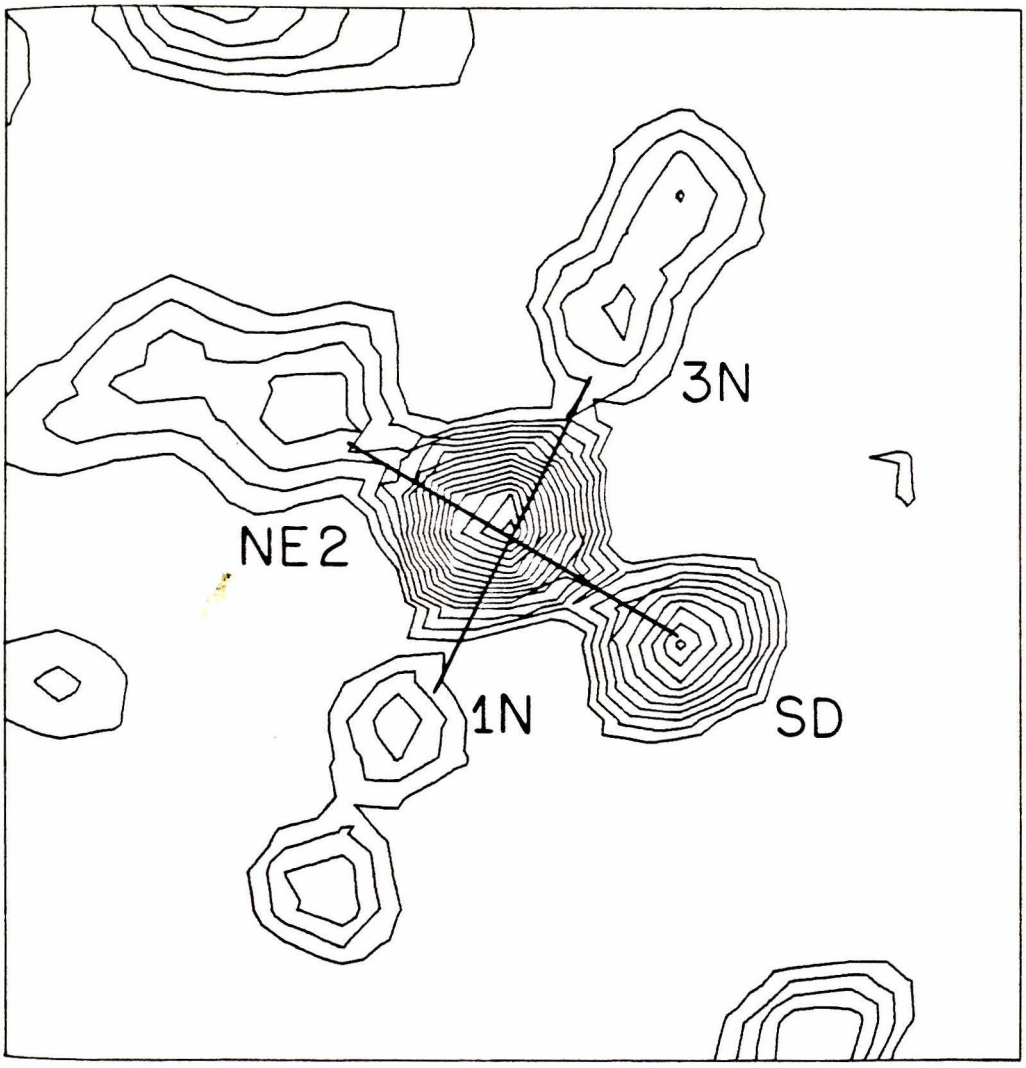


Figure 22d

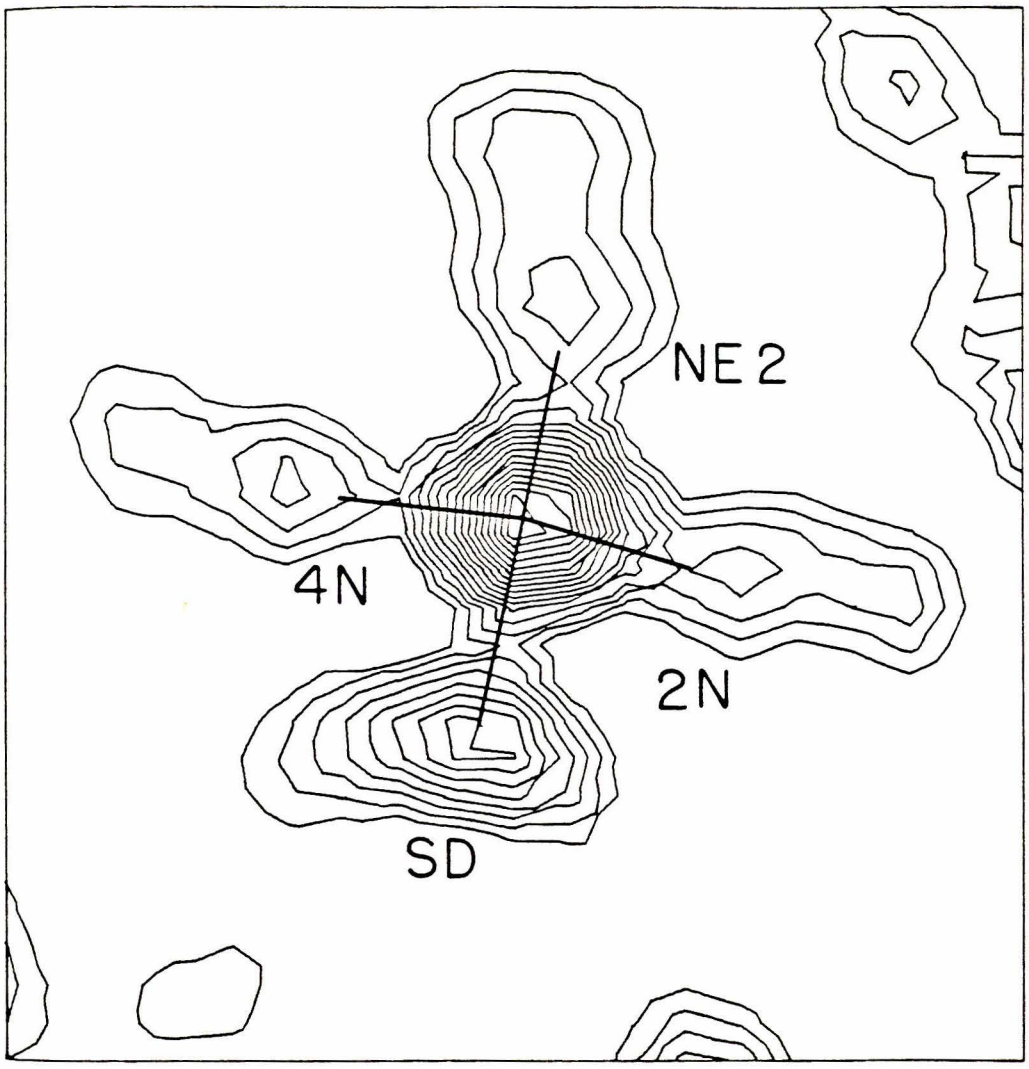


Figure 23a

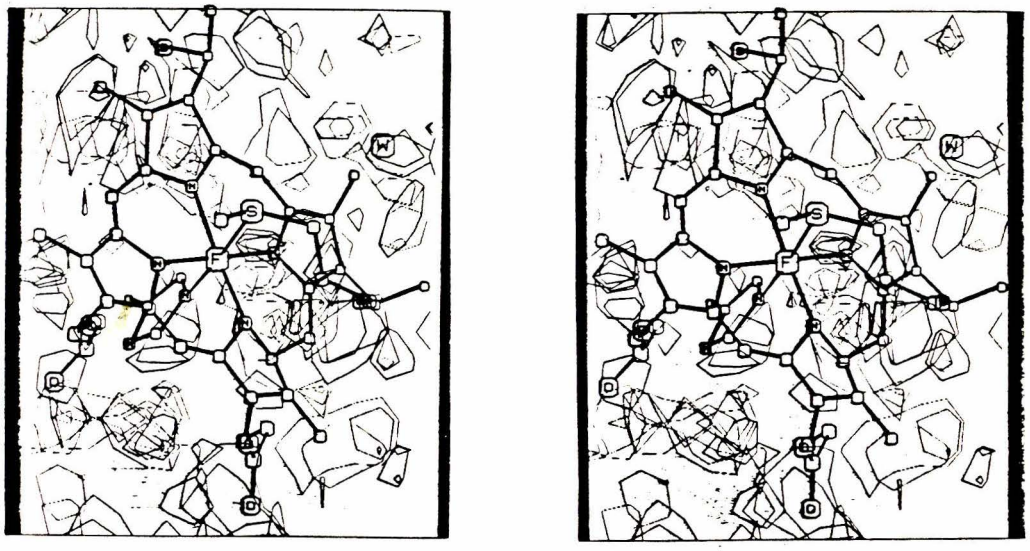


Figure 23b

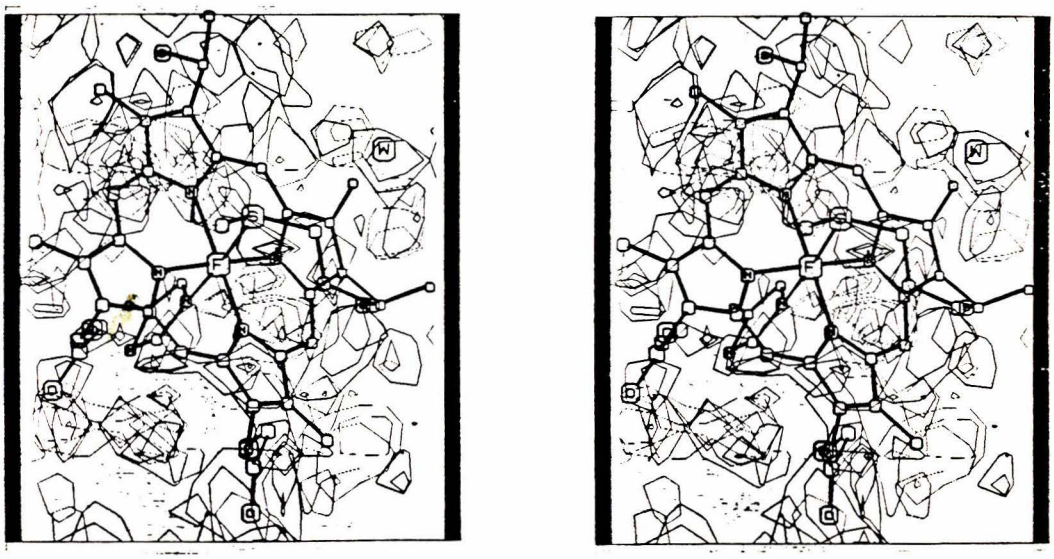


Figure 24

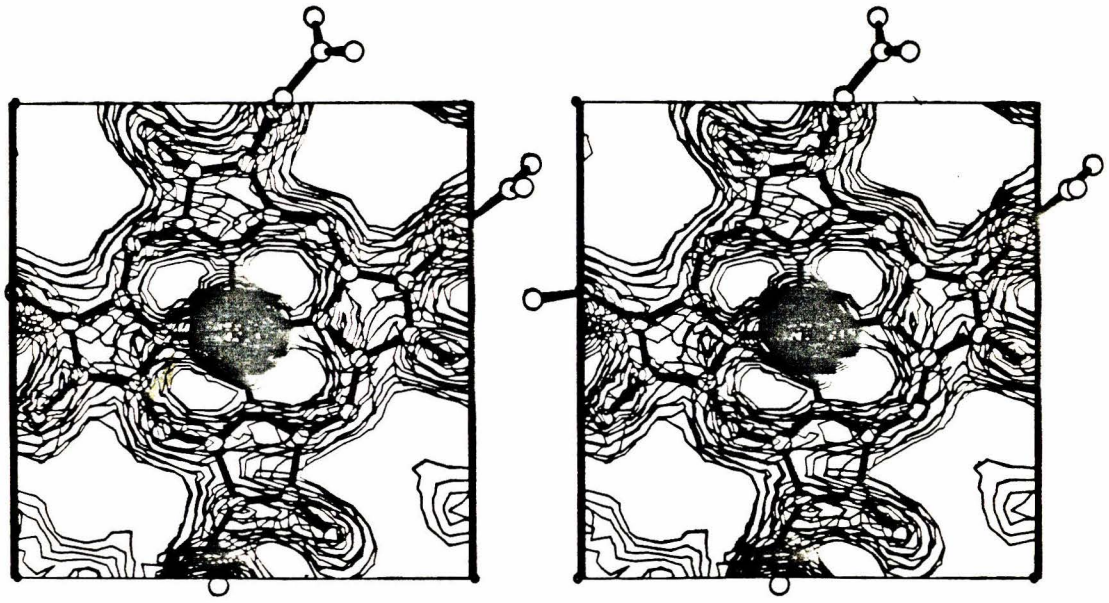


Figure 25a

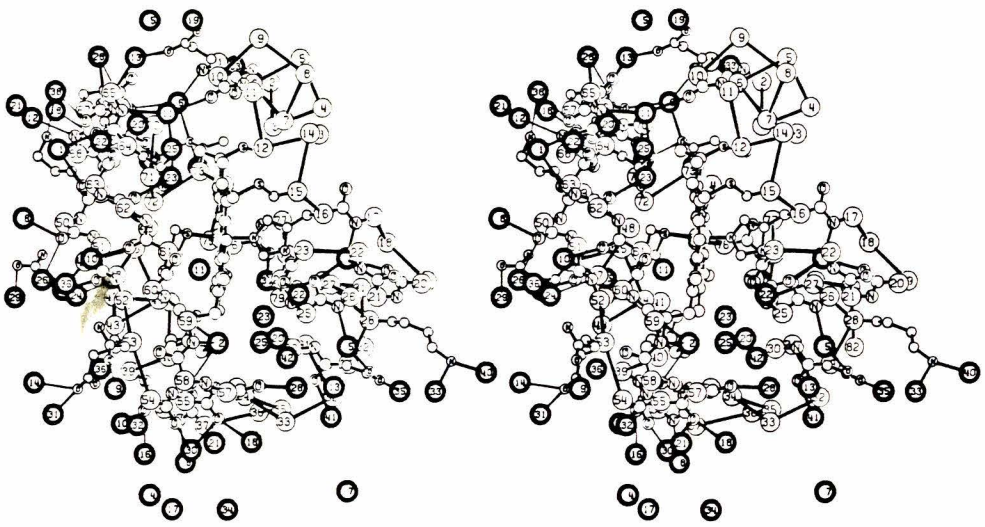


Figure 25b

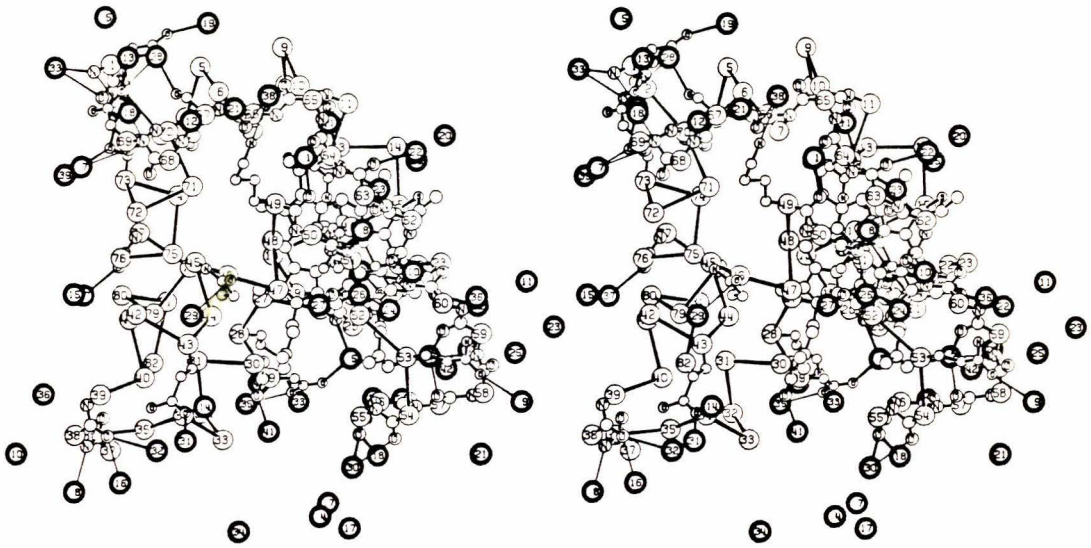


Figure 26

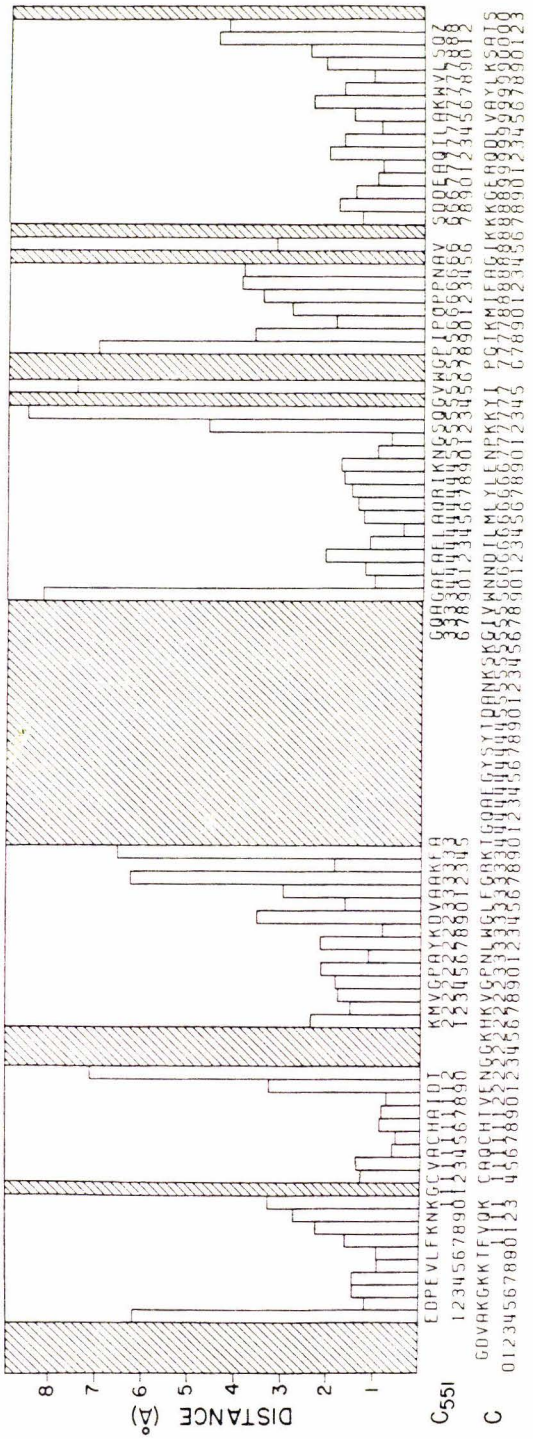
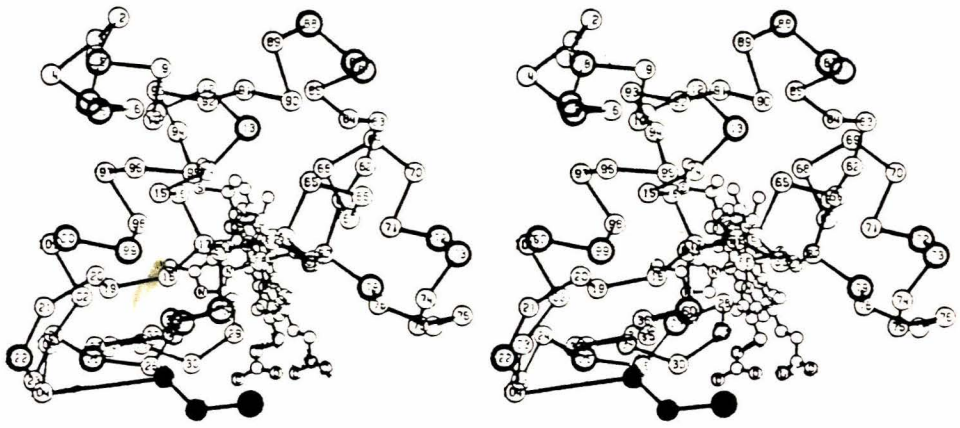


Figure 27



APPENDIX I

The Cytochrome Fold and the Evolution of Bacterial Energy Metabolism

J. Mol. Biol. (1976) **100**, 473–491

The Cytochrome Fold and the Evolution of Bacterial Energy Metabolism

RICHARD E. DICKERSON, RUSSELL TIMKOVICH AND ROBERT J. ALMASSY

Norman W. Church Laboratory of Chemical Biology
California Institute of Technology
Pasadena, Calif. 91125, U.S.A.

(Received 22 August 1975, and in revised form 21 October 1975)

Respiratory cytochrome c from eukaryotes, photosynthetic c_2 from purple non-sulfur bacteria, and respiratory c_{550} from *Paracoccus* (formerly *Micrococcus*) *denitrificans*, have been shown by X-ray structure analysis and amino acid sequence comparisons to form a unified subfamily of evolutionarily homologous proteins (cytochromes c_2), in which the differences between respiratory and photosynthetic cytochromes are no greater than the variation among photosynthetic cytochromes c_2. From sequence and structure considerations alone, these cytochromes are indistinguishable as to origin or function. Low resolution (4 Å) X-ray analysis of *Pseudomonas* cytochrome c_{551} shows it to have the same overall folding pattern as cytochrome c, with a massive deletion at the bottom of the molecule. This information enables amino acid sequence homologies to be extended with reasonable probability of correctness to all cytochromes c_{551}, to *Chlorobium* c_{555}, *Pseudomonas* c_5, algal cytochromes *f*, and even c_{553} from *Desulfovibrio*. All of these proteins appear to comprise a broader family of evolutionarily related electron carriers.

It is proposed that bacterial and eukaryotic oxygen respiration arose from the dual-function photosynthetic and respiratory electron chain in purple non-sulfur bacteria, by the loss of photosynthetic capabilities. A detailed outline for the evolution of photosynthesis and respiration in bacteria is presented, involving first a symbiotic sulfur cycle, then the development of the Krebs cycle machinery and finally the evolution of the present-day symbiotic oxygen cycle between plants and animals.

1. Introduction

One of the more intriguing dividends from protein crystal structure analysis has been the ability to perceive evolutionary relationships in proteins in a stronger way than is provided by amino acid sequence comparisons alone. Myoglobin and hemoglobin were the first such examples (Perutz *et al.*, 1960). With study, evolutionary homology in the globins can be discerned from their amino acid sequences, but the homology is instantly apparent from their three-dimensional structures. The pancreatic serine proteases are so similar that homologies are obvious from sequence as well as structure (Walsh & Neurath, 1964; Stroud, 1974). A more interesting case is provided by lactate, malate, alcohol, and glyceraldehyde-3-phosphate dehydrogenases. In these proteins the NAD-binding domains share a common folding pattern, although the remainder of each molecule is different (Buehner *et al.*, 1973; Rossmann

et al., 1975). The structural similarity of the entire binding domain, and not merely the NAD-binding site, can be used to support the concept of divergence from a common ancestral NAD-binding protein, rather than convergence to the same dinucleotide-binding active site structure.

The *c*-type cytochromes of bacteria and eukaryotes are particularly well suited for this type of homology study. They all are heme proteins of moderate molecular weight, with the same type of heme and covalent attachment of heme to the protein chain. The homologies between some of these cytochromes are obvious from amino acid sequence comparisons alone, as with cytochromes *c* of eukaryotic respiration and c_2 from photosynthetic bacteria. With others, a relationship is suggested by sequences, but is harder to see. For example, it has not been clear how the sequences of eukaryotic *c* and bacterial c_{551} should be aligned in order to demonstrate homology (Needleman & Blair, 1969; Dickerson, 1971). Still other types of cytochromes exhibiting a *c*-type spectrum may not be related at all. These proteins may represent nothing more than convergence on a particularly favorable way of binding heme to a polypeptide chain.

Can three-dimensional protein structures reveal evolutionary relationships that are so ancient as to be obscured in the amino acid sequences? Is it possible to use protein structure comparisons among distantly related molecules to decipher the evolutionary history of families of proteins, and of the metabolic chains that include them? These are the questions that this paper attempts to answer, for cytochromes *c* of eukaryotic and prokaryotic photosynthesis and respiration.

2. The Varieties of Cytochrome *c*

It should be emphasized that the term "cytochrome *c*" is a spectral classification, based on the wavelengths of the principle visible absorption bands, and that these ultimately arise from the particular substituents on the heme and the mode of attachment of heme to polypeptide chain (Dickerson & Timkovich, 1975). The standard pattern is a covalent attachment through cysteines in the sequence: Cys-*X*-*Y*-Cys-His, where *X* and *Y* are other amino acids, and the histidine side chain is the fifth ligand to the heme iron. These molecules functon as electron carriers in photosynthesis and respiration. A summary of the various kinds of photosynthesis and respiration encountered in prokaryotes and eukaryotes is given in Table 1. All of these pathways use electron transport chains that possess low molecular weight cytochromes *c*. Some, but not necessarily all, of these cytochromes *c* may be evolutionarily related; and this in turn may provide information about the evolutionary relationships between photosynthesis and respiration.

The patterns of amino acid sequences that have been observed in these small cytochromes *c* are shown diagrammatically in Figure 1. The most common pattern, in the first line, is found in respiratory *c* of eukaryotes and c_{550} of *Paracoccus denitrificans* (formerly *Micrococcus*), and in photosynthetic c_2 of purple non-sulfur bacteria. These proteins have the standard heme attachment near the amino terminus, and a methionine residue serving as the sixth iron ligand farther down the chain toward the carboxyl terminus. Other bacterial cytochromes have a chain-bending proline residue following this methionine, and cytochromes c_{551} from *Pseudomonas* bracket the methionine with no less than three prolines. As will be seen, all of these cytochromes have the same overall folding pattern in three dimensions, and almost

TABLE 1

Photosynthetic and respiratory electron transport systems where cytochromes c *are found*

Metabolic pathway	Typical small cytochromes
1. *Sulfide-using photosynthesis*: $H_2S \rightarrow S \rightarrow SO_4^{2-}$	
Green sulfur bacteria: *Chlorobium*	c_{555}
Purple sulfur bacteria: *Chromatium*	c_{553}, c_{550}
2. *Sulfate respiration*: $SO_4^{2-} \rightarrow H_2S$	
Desulfovibrio, *Desulfotomaculum*	c_3, c_{553}
3. *Cyclic photosynthesis*: No external reductant necessary	
Purple non-sulfur bacteria: *Rhodospirillum*, *Rhodopseudomonas*	c_2
4. *Water-using photosynthesis*: $H_2O \rightarrow O_2$	
Prokaryotes: Blue-green algae	*f*
Eukaryotes: Other algae, green plants	*f*
5. *Oxygen respiration*: $O_2 \rightarrow H_2O$	
Prokaryotes	
Purple non-sulfur bacteria (secondary importance)	c_2
Blue-green algae (secondary importance)	
Other non-photosynthetic, respiring bacteria (nitrate respiration sometimes an alternative)	c_{550}, c_{551}, c_4, c_5, others
Eukaryotes	
Other algae (secondary importance)	*c*
Green plants (major importance)	*c*
Animals (essential)	*c*

surely are descendants of a common protein ancestor. For cytochromes c_4, c', and the subsequent examples of Figure 1, the situation is less certain. Resolution of their evolutionary relationships, if any, must depend on future research. Similarly, there is no reason at present to suspect that the larger cytochrome c_1 of mitochondrial respiration is necessarily related to these proteins by ancestry.

3. The Cytochrome Fold

The structures of three cytochromes *c* have been solved by X-ray diffraction methods at high resolution: respiratory cytochrome *c* from tuna (Dickerson *et al.*, 1971; Swanson *et al.*, 1976; Takano *et al.*, 1973, 1976), photosynthetic c_2 from the purple non-sulfur bacterium *Rhodospirillum rubrum* (Salemme *et al.*, 1973), and c_{550} from the facultative nitrate-respiring bacterium *Paracoccus denitrificans* (Timkovich & Dickerson, 1973, 1976). These molecules are similar but not identical in size: 103 amino acids from tuna *c*, 112 amino acids from c_2, and 134 for c_{550}. Nevertheless, the overall folding pattern of these proteins is the same, as can be seen from the stereo drawings of Figures 2 to 4. In each case, the heme is wrapped in a similarly folded cocoon of polypeptide chain, leaving only one edge of the heme exposed. The heme environments and packing of hydrophobic side chains in the molecular interior are virtually identical, as are the hydrogen bonds to the heme propionic acids. With two exceptions, which will be discussed below, aromatic side chains occur in equivalent locations; and positively charged lysines are found at the upper and lower ends of the heme crevice.

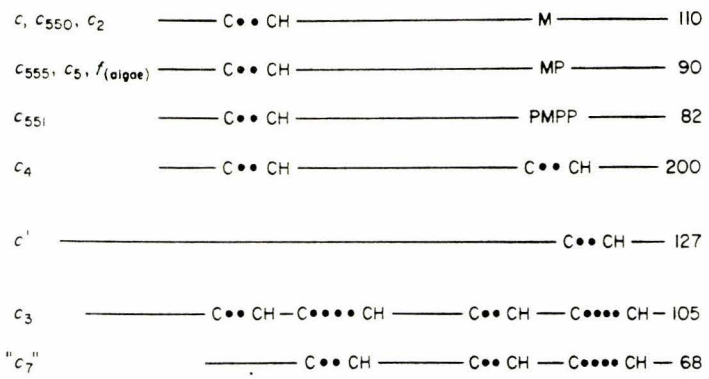


FIG. 1. Schematic representation of amino acid chains in small cytochromes *c*. Cytochrome types are listed at the left, by the amino terminus of the chain, and typical chain lengths are given at the right, beside the carboxyl terminus. Chains in first row (110) vary from 103 to 134 residues. C, H, M and P are cysteine, histidine, methionine and proline; and dots represent unspecified amino acids. From Dickerson & Timkovich (1975).

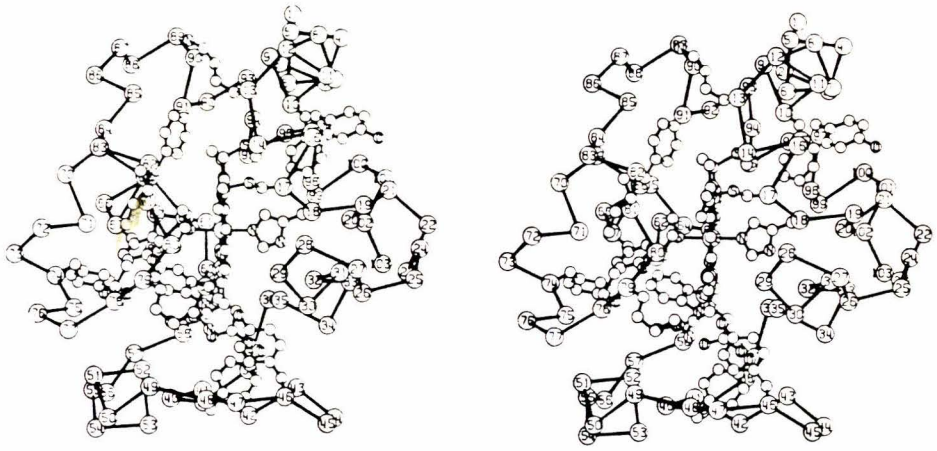


FIG. 2. Stereo pair alpha-carbon drawing of reduced tuna cytochrome *c*, viewed directly into the heme crevice. The heme group is seen on edge, with ligands histidine 18 (right) and methionine 80 (left). The 8 evolutionarily conserved aromatic side chains are shown, as are the lysines at top and bottom of the heme crevice (positions 13 and 79). From Takano *et al.* (1973).

Structural differences among these three molecules occur in the form of additional loops of chain on the molecular surface. The two bacterial cytochromes, c_2 and c_{550}, both have extra amino acids in the 50 and 70 regions. (For uniformity, residue numbering in this paper will always be that of eukaryotic cytochrome *c*.) Respiratory c_{550} has even more amino acids in an enlarged loop in the 20 region and in a 15-residue carboxyl-terminal "tail". These variations in chain length, from 103 to 134 amino acids, are all accommodated as excisions or additions at the surface of the molecules. The structural integrity of the interior of the cytochrome *c* molecule is carefully preserved, presumably by natural selection.

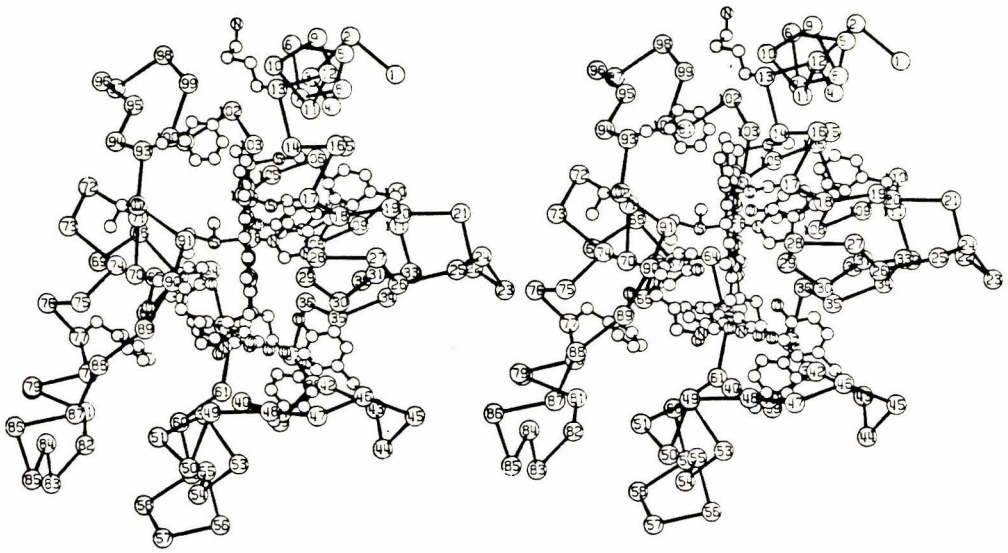


FIG. 3. Equivalent stereo drawing of oxidized cytochrome c_2 from *R. rubrum*. The same 8 aromatic side chains and 2 lysines are drawn as in Fig. 2. Notice the overall similarity of folding of the main chain, with additions in the regions around residues 50 and 80 regions (70 region in *c* numbering). Observe also the shift in positions of the 2 aromatic rings on the right side (positions 20 and 107, or 20 and 97 in *c* numbering). Co-ordinates by courtesy of F. R. Salemme and J. Kraut.

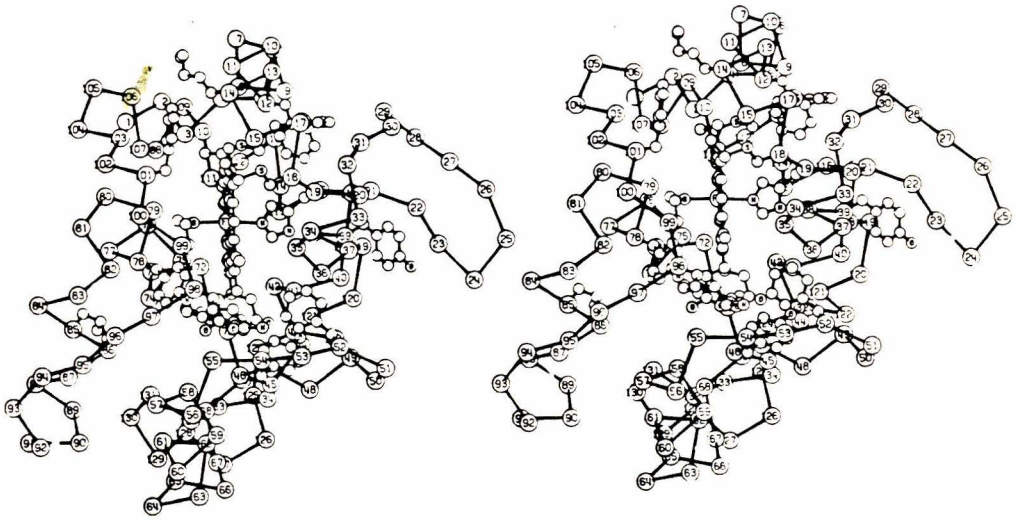


FIG. 4. Stereo drawing of oxidized cytochrome c_{550} from *Paracoccus* (formerly *Micrococcus*) *denitrificans*. Notice the similarity in location of aromatic rings, but the replacement of an aromatic ring by leucine at position 85 (74 in *c*). Note also the high positioning of rings on the right side, as in cytochrome *c*, at residues 12 and 113 (10 and 97 in *c*), and the additional tyrosine at position 39. The loop in the region of residue 20 at the right is more extensive than in Figs 2 and 3.

One interesting difference in aromatic side chain positions is observed between photosynthetic c_2 and the two respiratory cytochromes. In the latter, phenylalanine 10 and tyrosine/phenylalanine 97 form a pair of aromatic rings on the upper right of

the molecule as viewed in Figures 2 to 4. In contrast, the position structurally analogous to residue 10 is serine in c_2 rather than phenylalanine. A compensating phenylalanine is found farther down the right side of c_2 at position 20, which in the two respiratory proteins is valine or isoleucine. To accommodate this shift in phenylalanine position, tyrosine 97 is rotated by one place along the carboxyl-terminal alpha helix, as can be seen in the stereo drawings. However, sequence comparisons show that this does not represent a change in tyrosine position in the amino acid chain. Instead, it is a structural alteration, with the entire alpha helix in c_2 being rotated by 100° along its axis to bring the *same* tyrosine 97 next to position 20 rather than 10.

The other major change in aromatic side chains is the replacement of tyrosine or phenylalanine at the position corresponding to residue 74, by *leucine* in c_{550}. To a certain extent, c, c_2 and c_{550} are cross-reactive. Cytochrome c can donate electrons to the *R. spheroides* bacteriochlorophyll photocenter (Prince *et al.*, 1974), and all three proteins can interact with the reductases and oxidases of c and c_{550} with varying efficiences (Kamen & Vernon, 1955; Smith *et al.*, 1966; Davis *et al.*, 1972), suggesting a shared electron-transfer mechanism. The complete absence of an aromatic ring at position 74 in c_{550} therefore effectively eliminates the so-called "Winfield mechanism" of electron transfer through aromatic side chains on the left side of the molecule (Dickerson *et al.*, 1972).

4. The c_2 Class of Cytochromes

The X-ray structure comparisons described above suggested initially that eukaryotic respiratory c, bacterial respiratory c_{550}, and bacterial photosynthetic c_2 each had distinctive structure features:

(a) extra loops of amino acids in the 50 and 70 regions of bacterial cytochromes.

(b) An extended loop in the 20 region and a carboxyl-terminal tail in c_{550}.

(c) A lower position for the paired aromatic rings on the right side of c_2.

In reality, none of these generalizations is valid. Ambler has sequenced four other cytochromes c_2 from purple non-sulfur bacteria, and these are compared with sequences of the previously discussed cytochromes in Table 2. (Ignore c_{555} and beyond for the moment.) These comparisons reveal that:

(a) the extra chains in the 50 and 70 regions are not characteristic of all cytochromes c_2. *R. vannielii* c_2, although a photosynthetic electron carrier, has a sequence exactly paralleling that of eukaryotic c.

(b) The extra residues in the loop in the 20 region are not characteristic of respiratory c_{550}. The same features are observed in c_2 of *R. capsulata* and *R. spheroides*.

(c) The lowered aromatic rings are not general for all cytochromes c_2, but are a peculiarity of *R. rubrum*. In all other purple non-sulfur bacteria, these aromatic residues are positioned as they are in c and c_{550}.

In summary, there appear to be no clear-cut structural features that distinguish cytochromes c, c_{550} and c_2, in spite of their diversity of origin and function.

Similar conclusions can be drawn from the amino acid comparison matrix of Table 3. Two different types of comparisons are made, which reveal different aspects of structure. At the upper right of the Table, the lack of an amino acid at a point on one sequence but not the other, is counted formally as a 21st kind of amino acid. Loops of chain possessed by one protein but not the other contribute to the tally of

TABLE 2

Amino acid sequence alignments in c-type cytochromes

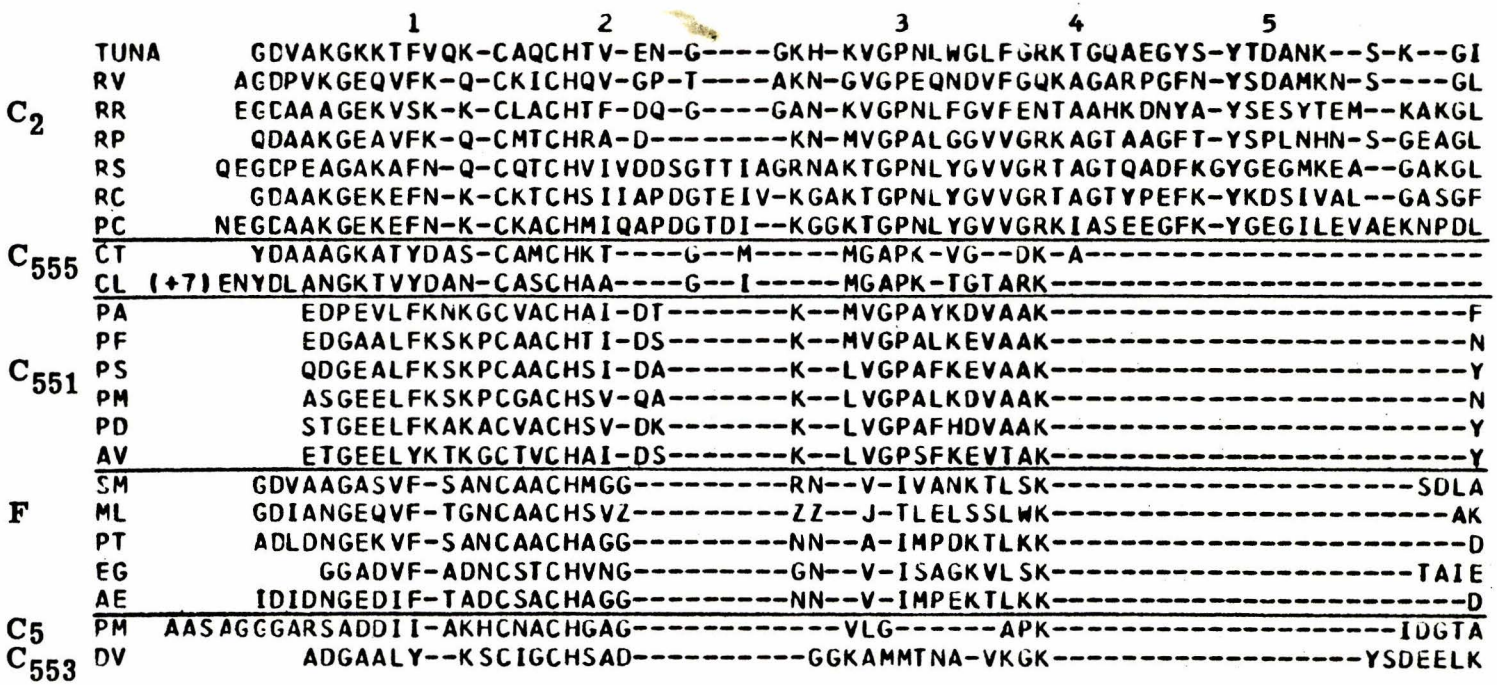


TABLE 2—*continued*

6	7	8	9	10		
VWNNDTLMEYLENPKKYI----------		PGTKMIFAGIK-K--KGERQDLVAYLKSATS--			TU	
TWDEATLDKYLENPKAVV----------		PGTKMVFVGLK-N--PQDRADVIAYLKQLSGK-			RV	
TWTEANLAAYVKDPKAFVLEKSGDPK------		AKSKMTFK-LT-K--DDEIENVIAYLKTLK---			RR	C_2
VWTQEDIIAYLPDPNAYLKKFLTDKGQADKATGSTKMTFK-LA-N--DQQRKDVAAYLATLK---					RP	
AWDEEHFVQYVQDPTKFLKEYTGDAK------		AKGKMIFK-LK-K--EADAHNIWAYLQQVAVRP			RS	
AWTEEDIATYVKDPGAFLKEKLDDKK------		AKTGMAFK-LA-K----GGEDVAAYLASVVK--			RC	
TWTEANLIEYVTDPKPLVKKMTDDKG------		AKTKMTFK-MG-K----NQADVVAFLAQBBPBAGZGZAAGAGSBSZ			PC	
AWAP-HIAKGMNVM----VANSIKGYKG----		TKGMMPAKGGNPKLTDAQVGNAVAYMVGQSK--			CT	C_{555}
-WNS-RLPQGLATMIEKSVAGYEGEYRG----		SKTFMPAKGGNPDLTDKQVGDAVAYMVNEVL--			CL	
AGQAGAEAELAQRIKNGSQGV----------		WGPIPMPPNAV----SDDEAQTLAKWVLSQK---			PA	
AGVKDADKTLAGHIKNGTQGV----------		WGPIPMPPNQV----TDAEALTLAQWVLSLK---			PF	
AGQDGAADLLAGHIKNGSQGV----------		WGPIPMPPNPV----TEEEAKILAEWILSQK---			PS	C_{551}
AGVDGAADVLAGHIKNGSTGV----------		WGAMPMPPNPV----TEEEAKTLAEWVLTLK---			PM	
GAQGDGVAHITNSIKTGSKGN----------		WGPIPMPPNAV----SPEEAKTLAEWIVTLK---			PD	
AGQAGIADTLAAKIKAGGSGN----------		WGQIPMPPNPV----SEAEAKTLAEWVLTHK---			AV	
KYLKGFDDDAVAAV-AYQVTN----------		GKNAMPGFNG--RLSPLQIEDVAAYVVDQAEKGW---			SM	
SYLANFNGDESAIV--YQVTN----------		GKNAMPAFGG--RLEDDEIABVASYVLSKAG------			ML	F
VLEANSMNTIDAIT--YQVQN----------		GKNAMPAFGG--RLVDEDIEDAANYVLSQSEKGW---			PT	
EYLDGGY-TKEAIE--YQVRN----------		GKGPMPAWEG--VLSEDEIVAVTDYVYTQAGGAWANV			EG	
ALADAKMVSVNAIT--YQVTN----------		GKNAMPAFGS--RLAETDIEDVANFVLTBZBKGWE--			AE	
AWKERADHGGLDG-ILAKAIS----------		GINAMPPKGT--CADCSDDELREAIQKMSGL			PM	C_5
ALADYMKAAMGSAK-PVKGQGAE---------		ELYKMKGYAD--GSYGGERKAMSKL			DV	C_{553}

RV, *Rhodomicrobium vannielii* (ATCC 17100)
RR, *Rhodospirillum rubrum* (ATCC 11170)
RP, *Rhodopseudomonas palustris* (ATCC 17001)
RS, *Rhodopseudomonas spheroides* (ATCC 17023)
RC, *Rhodopseudomonas capsulata* (ATCC 23782)
PC, *Paracoccus denitrificans* (ATCC 13543)
CT, *Chlorobium thiosulphatophilum* (NCIB 8346)
CL, *Chlorobium limicola* (in "Chloropseudomonas ethylica")
PA, *Pseudomonas aeruginosa* (P6009)
PF, *Pseudomonas fluorescens* (ATCC 17588)
PS, *Pseudomonas stutzeri* (ATCC 17400)
PM, *Pseudomonas mendocina* (CH-110)
PD, *Pseudomonas denitrificans* (ATCC 13867)
AV, *Azotobacter vinelandii* (ATCC 12837)
SM, *Spirulina maxima* (prokaryotic blue-green alga)
ML, *Monochrysis lutheri* (eukaryotic alga)
PT, *Porphyra tenera* (eukaryotic red alga)
EG, *Euglena gracilis* (eukaryote)
AE, *Alaria esculenta* (eukaryotic brown alga)
DV, *Desulfovibrio vulgaris* (NCIB 8303)

Paracoccus c_{550} sequence from Timkovich *et al.* (1975). All other prokaryotic sequences from Ambler (1975); algal cytochromes *f* from Ambler & Bartsch (1975).
Numbers 1, 2, 3, 4 . . . across the top of the Table indicate residues 10, 20, 30, 40 . . . in cytochrome *c* numbering.
The first seven residues in *Chlorobium limicola*, indicated by (+7) in the Table, are: A V T K A D V.
J, monomethyllysine. Other amino acid symbols are standard.

TABLE 3

Amino acid differences between cytochromes c, c_2, c_{550}

	Tuna	*RV*	*RR*	*RP*	*RS*	*RC*	*PD*	
Tuna *c*		59	78	81	92	85	98	Missing amino acids counted
R. vannielii c_2	54		75	73	89	86	98	
R. rubrum c_2	61	59		74	72	68	85	
R. palustris c_2	58	49	62		88	71	96	
R. spheroides c_2	65	65	60	64		62	86	
R. capsulata c_2	60	60	56	51	54		71	
Paracoccus c_{550}	57	60	57	60	66	51		

Missing amino acids eliminated from count

Compare these differences with those of eukaryotic cytochromes *c* given in Table XI of Dickerson & Timkovich (1975).

differences. At the lower left of the Table, positions at which amino acids occur in only one of the two chains being compared are deleted from the count, so these numbers reflect only the differences between chains in common structural regions. Most of the large difference between tuna and *Paracoccus* at the upper right corner comes from the extra chain in the latter cytochrome, but when these extra chains are eliminated from consideration at the lower left corner of the Table, the two sequences appear much more similar. This Table demonstrates that, in regions of common tertiary structure, respiratory cytochromes *c* and c_{550} are no more different from c_2 than the various cytochromes c_2 of photosynthetic bacteria are from one another.

In both sequence and structure, the cytochromes c_2 span the entire range from small cytochrome *c* to large c_{550}. We are left with the unsettling conclusion that there is nothing in the three-dimensional folding or the amino acid sequences of eukaryotic respiratory cytochrome *c* and bacterial respiratory c_{550} to indicate that they are not photosynthetic cytochromes from purple non-sulfur bacteria! Since it is not yet clear that other cytochromes with a *c*-type absorption spectrum are necessarily related to these by ancestry, it would be convenient to have a special terminology to distinguish this close-knit subfamily of proteins that are evolutionarily homologous with eukaryotic *c*. The designation by principal alpha-band absorption wavelength in nanometers is not discriminating enough: Tuna cytochrome *c* (actually a c_{550}), *Paracoccus* c_{550}, and *Euglena* c_{558} are members of this subfamily, but *Pseudomonas* c_{551} and *Desulfovibrio* c_{553} are not. One alternative would be to choose the only designation for this subfamily that does not conflict with other cytochromes, and to refer to all of these proteins—photosynthetic c_2 and respiratory c_{550}, including eukaryotic cytochrome *c*—collectively as cytochrome c_2. This would have the virtue of comprehensiveness, and also an after-the-fact logic in the mitochondrial respiratory chain of having cytochrome c_1 pass its electrons on to cytochrome c_2.

5. Evolutionary Implications

The most probable conclusions to be drawn from the foregoing comparisons are that all of these photosynthetic and respiratory cytochromes *c* (or c_2) are closely

related by evolutionary homology. In principle, it is not possible to differentiate absolutely between divergence and convergence in explaining protein structure similarities—to decide whether two proteins resemble one another because they have a common ancestor, or because they have been shaped in the same way by natural selection. One could say absolutely that corresponding structural features were homologous only if one could prove that they served no useful purpose, and hence had no functional reason for being alike. In practice, matters usually are simpler. One would find it incredible to believe that the near-identity of chain folding in trypsin and chymotrypsin, even in regions far from the active site, was entirely due to natural selection operating on unrelated polypeptides. At the other extreme, the great similarity of trypsin and subtilisin at their active sites, but total dissimilarity elsewhere in the molecules, suggests that two unrelated enzymes have been shaped alike at their active centers because they have the same catalytic mechanism. The similarities among cytochromes c, c_2 and c_{550} are like those between trypsin and chymotrypsin; greater correspondence of chain folding than could be explained by any reasonable picture of functionality. The overwhelmingly most likely explanation is that these cytochromes are folded in the same way because they have a common protein ancestor.

With this interpretation, it follows either that the electron transport chains of photosynthesis and respiration share a common evolutionary origin, or that a molecular component of one chain has been adapted and taken over by the other. The other common features of these electrons chains: ubiquinone-like molecules at the low-potential end, cytochromes b in the middle, and cytochromes c at the high-potential end, suggest that the entire chains are related, and not merely the cytochrome c elements. The traditional view of the sequence of events during the evolution of life makes it probable that respiration evolved from bacterial photosynthesis by the development of new reducing donors and oxidizing acceptors at the two ends of the chain.

The electron pathways in photosynthesis and respiration are shown schematically in Figure 5, in diagrams emphasizing their common features. The c-containing electron transport chain that is proposed as being evolutionarily homologous is represented by a heavy black arrow through the quinone (Q) and cytochromes b and c (or f). Not all of the chain components and ATP-generating sites are shown, in some cases because they would clutter the diagram, and in other cases because they are not yet known in detail.

The biochemical evidence suggests that the point of divergence between photosynthesis and respiration occurred in the ancestors of purple non-sulfur photosynthetic bacteria. Aside from blue-green algae, which have the two-photocenter, water-splitting photosynthesis of higher plants, these are the only prokaryotes capable of both photosynthesis and respiration. Photosynthetic green and purple sulfur bacteria such as *Chlorobium* and *Chromatium* do not respire. They carry out both cyclic and non-cyclic photosynthesis of the types shown in Figure 5(a) and (b), depending on their relative needs for energy in the form of ATP, or energy and reducing power in NADH. The non-cyclic photosynthesis of Figure 5(a) requires a source of reducing equivalents, which are obtained by oxidizing H_2S to free sulfur, and ultimately to sulfate. Some green and purple sulfur bacteria have a partial or complete Krebs cycle, but apparently use it as a means of synthesizing glutamate and other molecules, or run it in reverse as an auxiliary means of fixing CO_2

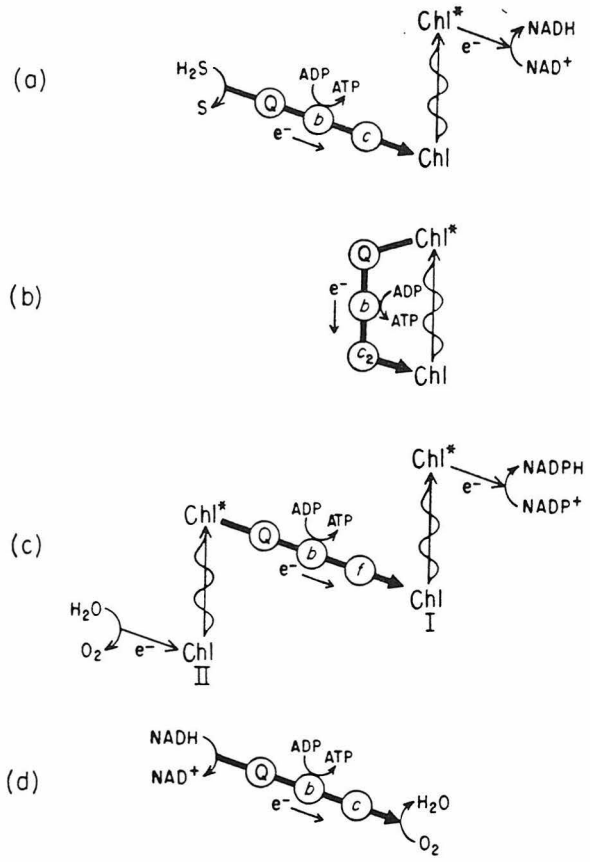


FIG. 5. Patterns of electron transport. (a) Non-cyclic bacterial photosynthesis with H_2S as an external source of reducing power. (b) Cyclic bacterial photosynthesis without an external donor. (c) Two-center, oxygen-releasing photosynthesis in algae and green plants. (d) Mitochondrial respiration. Q, Ubiquinone, menaquinone or plastoquinone. Chl, chlorophyll, and Chl*, electronically excited chlorophyll. *b*, *c* and *f* represent cytochromes. The electron transport chain that is proposed as being evolutionarily homologous in all these mechanisms is emphasized by the dark arrow.

(Evans *et al.*, 1966; Fuller, 1969; Doelle, 1969; Sokatch, 1969; Stanier *et al.*, 1970). They do not employ the Krebs cycle as a source of NADH for respiration.

In contrast, purple non-sulfur bacteria possess a complete Krebs cycle, which can supply them with NADH (Fuller, 1969). Under anaerobic conditions in the light, they carry out cyclic photophosphorylation (Fig. 5(b)) to produce ATP energy but not NADH and, therefore, do not need an external source of reducing power. They can obtain the necessary NADH for synthetic purposes from other reactions, including the Krebs cycle. Under aerobic conditions in the dark, they can shift to a purely respiratory mode of life, producing NADH by the Krebs cycle and extracting energy by reoxidizing it in a respiratory chain like that of eukaryotes (Fig. 5(d)). Metabolic control is provided in part by the fact that the production of bacteriochlorophyll is inhibited by the presence of free O_2 (Lascelles & Wertlieb, 1971; Marrs & Gest, 1973*b*).

At least in *R. capsulata* and *R. spheroides*, and perhaps in other purple non-sulfur bacteria as well, the same electron transport chain is employed in both photosynthesis and respiration (Marrs & Gest, 1973*a*,*b*; Connelly *et al.*, 1973; Dutton & Wilson, 1974). A pool of quinone molecules at the low-potential end of the chain can be reduced by NADH, succinate, or excited bacteriochlorophyll. Transfer of electrons into this pool from either photosynthetic or respiratory donors can be blocked by the same inhibitors that are effective at this point in mitochondria (Parson, 1974). At the high-potential end, cytochrome *c* can pass electrons either to cytochrome oxidase or to a bacteriochlorophyll center, depending on which is oxidized or electron-depleted. The dual-purpose electron pathway in these bacteria is outlined in Figure 6. All that would be required to turn this diagram into one of conventional mitochondrial respiration would be for a mutation to incapacitate the photosynthetic pathway; and such photosynthesis-deficient mutants of *R. spheroides* have actually been studied (Saunders & Jones, 1974; Jones & Plewis, 1974). The oxidase in *R. spheroides* is even a cytochrome a/a_3 as in eukaryotes (Saunders & Jones, 1974), although the oxidase in *R. capsulata* appears to be a cytochrome *b* or *o* (Zannoni *et al.*, 1974).

Present-day purple non-sulfur bacteria rely mainly on photosynthesis for energy, with respiration as only a "standby" alternative. This would have been a reasonable state of affairs on a primitive Earth in which free oxygen was in short supply, or was localized in the immediate vicinity of blue-green algae carrying out O_2-releasing photosynthesis. At a later stage, with more abundant atmospheric oxygen, accidental loss of photosynthesis ability could have been relatively harmless, and could have led to the ancestors of present-day respiring bacteria.

From the foregoing arguments, it is proposed that cytochromes *c*, c_2 and c_{550} comprise one virtually indistinguishable subfamily of proteins (cytochromes c_2), whose close structural and sequence similarities arise because of their close evolution-

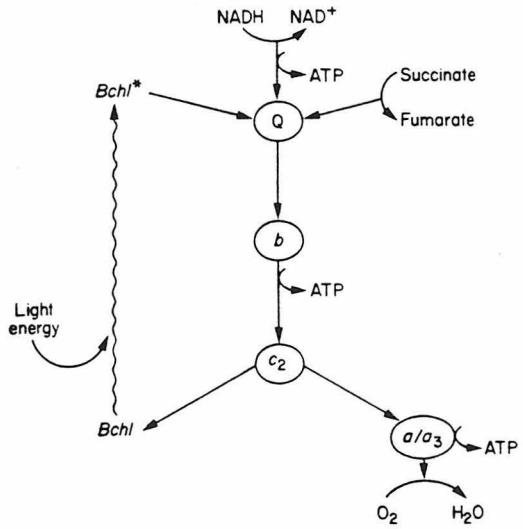


FIG. 6. Photosynthetic and respiratory electron flow in *R. spheroides*, showing the common electron chain from ubiquinone (Q) to cytochrome c_2, and the probable sites of ATP synthesis. Adapted from Marrs & Gest (1973*a*,*b*), Dutton & Wilson (1974), and Jones & Plewis (1974).

ary relationships. Eukaryotes, *Paracoccus denitrificans*, and perhaps other respiring bacteria, are all metabolic descendants of dual-function bacteria resembling purple non-sulfur bacteria, having arisen from them initially by the loss of photosynthetic ability.

The great similarity of *Paracoccus* c_{550} and eukaryotic *c* harmonizes well with the suggestion of John & Whatley (1975), that a *Paracoccus*-like bacterium was the symbiont which degenerated into the mitochondrion during the evolution of eukaryotes (Margulis, 1970). John & Whatley base their selection of *Paracoccus* as the mitochondrial precursor on membrane chemistry, respiratory chain structure, the presence of cytochrome a/a_3 as an oxidase, and the mechanisms of respiratory inhibition and control. No other bacterium, they point out, has such a large assembly of mitochondrion-like features, and no other bacterium has significant mitochondrial characteristics not possessed by *Paracoccus* as well. They also point out the similarities between mitochondria, *Paracoccus* and *R. spheroides*, which include the use only of ubiquinone-10 as the low-potential acceptor pool of reducing power (Peters & Cellarius, 1972). These findings are consistent with our structural comparisons: there is an apparent evolutionary progression from *Rhodopseudomonas*, through *Paracoccus*, to the eukaryotic mitochondrion.

6. Folding in Cytochrome c_{551}

The search for the cytochrome fold in prokaryotes has recently been extended by a preliminary, low resolution (4 Å)X-ray analysis of cytochrome c_{551} from *Pseudomonas aeruginosa* (Almassy & Dickerson, manuscript in preparation). This is a member of a class of smaller respiratory cytochromes, with 82 amino acids and, like *Paracoccus* c_{550}, it is involved in facultative nitrate respiration. A sequence homology with cytochromes *c* and c_2 has been suggested (Needleman & Blair, 1969; Dickerson, 1971), but the extensive deletions that would be necessary have made these comparisons inconclusive. In such a case, three-dimensional structure comparisons can be particularly useful.

As usual, the low-resolution electron density map from X-ray analysis would not have been interpretable by itself, and structural comments ordinarily would have to await the completion of the high-resolution analysis, for which data are now being collected. However, common structural features in related proteins can be recognized even in the blurred picture that one obtains at low resolution. The c_{550} analysis at 4 Å (Timkovich & Dickerson, 1973) showed that a correct interpretation of the map could be made using the high-resolution structure of eukaryotic *c* as a guide; and that regions of both similarity and difference could be discerned. This has given us confidence in the interpretation of the low resolution c_{551} map, and the tentative chain path deduced is shown in the stereo drawings of Figure 7. Only the overall path of polypeptide chain is represented, since individual amino acid positions cannot be distinguished at this resolution. The heme group is easy to see, as are the two long alpha-helices at the amino and carboxyl termini of the polypeptide chain. These features have the same relative positions and orientations as in *c*, c_2 and c_{550}. In addition, the path of the polypeptide chain between these helices can be traced, with the heme attachments and the loop in the region of residue 20 on the right side being especially easy to follow.

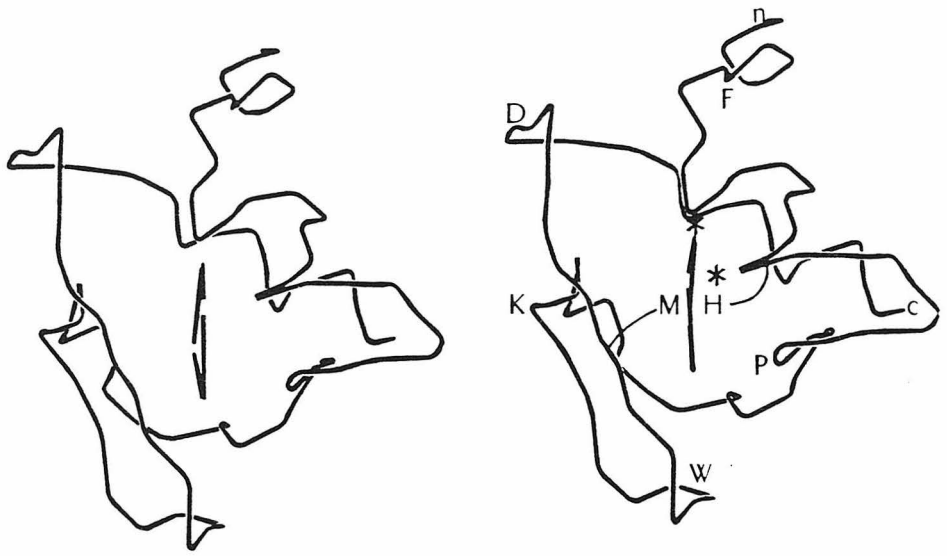


FIG. 7. Stereo pair drawing of the main chain path and heme (see on edge) in oxidized cytochrome c_{551} from *Pseudomonas aeruginosa*. Same molecular orientation as Figs 2 to 4. n and c indicate amino and carboxyl ends of chain. Asterisks locate cysteine attachments of heme to chain at positions 12 and 15 (14 and 17 in cytochrome *c*). H and M, heme ligands, histidine 16 and methionine 61 (18 and 80 in *c*). F, phenylalanine 7 (10 in *c*), P, proline 25 (30 in *c*), K, lysine 49 (72 in *c*), and D, aspartic 68 (88 in *c*). W, tryptophan 56, with side chain density in the map suggesting that it is hydrogen-bonded to the innermost heme propionic acid like tryptophan 59 in *c*, although the chain positions are *not* sequentially homologous. Positioning of amino acids is based partly on side-chain shapes (especially aromatic rings) and partly on total running chain-length through the map.

To a first approximation, if this interpretation of the c_{551} structure is correct, c_{551} can be obtained from cytochrome *c* by short-circuiting the chain from residue 38 to 57, removing the bottom of the molecule, and then pushing together the 20 and 70 regions from either side at the bottom to close the gap. This key piece of information about the region of deletion then makes a sequence comparison between c_{551} and the larger cytochromes possible. A deletion similar to this at the bottom of the molecule had been predicted in 1971 purely on the basis of sequence comparisons (Dickerson, 1971), but the suggested deletion involved residues 30 to 47, a loop running to the back of the molecule and forward again, rather than residues 38 to 57, a loop running forward and then back.

The alignment of cytochrome c_{551} with other cytochromes suggested by the 4 Å X-ray analysis is shown in Table 2. Once this critical deletion in the regions around residues 40 and 50 is recognized, it then becomes possible to extend the sequence alignment on a more tentative but suggestive basis to the photosynthetic c_{555} of green sulfur bacteria, photosynthetic *f* of algae, respiratory c_5 from *Pseudomonas*, and even the low-potential but mono-heme cytochrome c_{553} of sulfate-respiring *Desulfovibrio vulgaris*. The conclusions reached by Dus *et al.* (1968) regarding the phylogenetic relationships between the carriers of cytochromes *c*, c_2 and c_{551} seem to be essentially correct.

These proteins may all make up a broader evolutionary family than what we have

called the cytochromes c_2, with more distant relationships. There appears to be a minimal set of common structural features in these proteins, including:

(a) the Cys-*X*-*Y*-Cys-His-----Met framework, with His and Met as the heme ligands.

(b) A basic residue (lysine or arginine) at position 38 in *c* numbering, near the innermost propionic acid of the heme group.

(c) An aromatic ring at position 97, followed by a hydrophobic residue, and aligned next to another aromatic ring at position 10.

(d) Glycine at position 6, where the initial alpha-helix packs against the carboxyl-terminal helix.

(e) Acidic side chains in the bend at the beginning of the carboxyl-terminal helix.

(f) A hydrogen-bonding residue (serine, threonine, glutamine) at position 100 near the end of the carboxyl-terminal helix.

These generalizations to c_{555}, *f*, c_5 and c_{553} have been based on the assumption that their three-dimensional folding patterns resemble that of c_{551}. X-ray structure analyses of these cytochromes are needed in order to test this assumption. The cytochromes of Table 2 apparently form an extensive family of evolutionarily related molecules, of which cytochrome c_2 (including c_{550} and eukaryotic *c*) is a subfamily. Other cytochromes with several hemes, with heme attachment near the carboxyl terminus, or with bound flavins, may not be related to this family at all; but Table 2 would appear to extend the "cytochrome fold" through all types of photosynthesis and respiration, and through three billion years of evolutionary history.

7. The Evolution of Bacterial Metabolism

A pattern begins to emerge of the development of an electron transport chain operating between high and low free energy, storing some of the liberated free energy by the synthesis of ATP. In some applications, the donor of reducing power into the chain is an external reductant, and the ultimate acceptor is NAD, although the immediate electron acceptor is an electron-depleted chlorophyll center (Fig. 5(a)). In other cases, both donor and acceptor are bacteriochlorophyll and the process is a cyclic one (Fig. 5(b)). In a third type of photosynthesis donor and acceptor both are chlorophylls, but at different photocenters (Fig. 5(c)). Finally, in the respiratory adaptation, the donor of reducing power is a flavoprotein enzyme which has been reduced by NADH, and the acceptor is a cytochrome oxidase complex which ultimately will reduce oxygen to water, or nitrate to a lower oxidation state of nitrogen (Fig. 5(d)). The electron chain of sulfate respiration in *Desulfovibrio* may prove to be similar to these, but relatively little is known about its components yet.

The electron transport chains shown in Figure 5 have several common features, including quinones as early acceptors of reducing equivalents, followed by cytochromes *b* and *c*, with synthesis of ATP in at least one site along the chain. These similarities lead to a self-consistent picture of the evolution of bacterial energy-managing metabolism, shown diagrammatically in Figure 8. One can propose a series of seven evolutionary events which account satisfactorily for what we know about electron transport chains and cytochrome *c* structures, and which can serve as a framework for thinking about bacterial metabolism.

(1) Ancient bacteria on the primitive Earth under reducing atmospheric conditions were anaerobic fermenters resembling present-day *Clostridia*. They obtained free

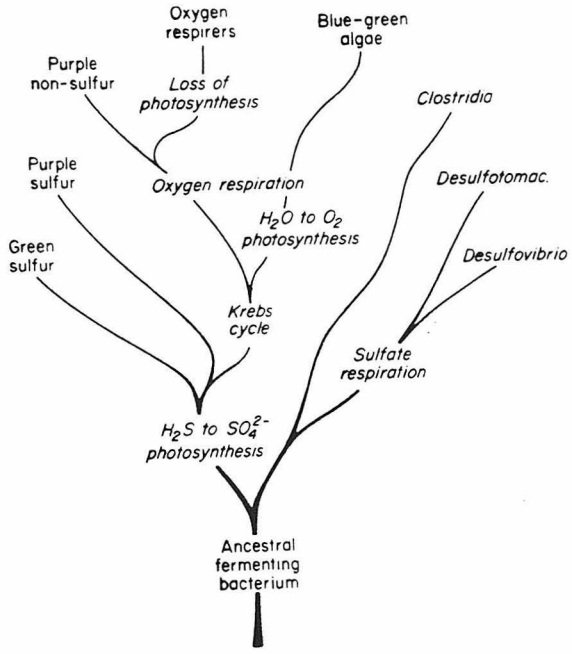


FIG. 8. Proposed evolutionary tree showing the development of cytochrome-containing electron transport chains involved in energy extraction or storage in bacteria. In this model, sulfate respiration arose in response to sulfate-releasing photosynthesis, as oxygen respiration arose later in response to oxygen-releasing photosynthesis. The Krebs cycle developed for non-respiratory purposes and was adapted to respiratory use by purple non-sulfur bacteria; and oxygen-respiring bacteria evolved from purple non-sulfur bacteria and blue-green algae by the loss of photosynthetic capabilities. See text for discussion.

energy by anaerobic glycolysis or the breakdown of other smaller organic molecules.

(2) The first scheme for tapping solar energy that has left modern descendants was sulfate-producing photosynthesis, like that carried out by green and purple sulfur bacteria such as *Chlorobium* and *Chromatium*. Electron transport chains with cytochromes *c* are first recognizable in these bacteria. Such organisms would have used the readily available H_2S as a source of reducing equivalents in photosynthesis of NADH, and would have produced local concentrations of sulfate, even under reducing atmospheric conditions.

(3) With the availability of sulfate from this or from geological sources, the ancestors of *Desulfovibrio* evolved the ability to extract more energy from their metabolites by oxidizing them with sulfate rather than simply degrading them by fermentation. A sulfur cycle could have developed between sulfide photosynthesizers and sulfate respirers, such as is encountered today between *Chlorobium* and *Desulfovibrio* (Gray *et al.*, 1972).

(4) The components of the Krebs cycle evolved piecemeal in the photosynthetic sulfur bacteria, probably for synthetic or CO_2-fixation purposes. This cycle, when complete, allowed its hosts to become the ancestors of purple non-sulfur bacteria by providing an alternative source of NADH and eliminating the need for an external reductant.

(5) A different alternative to sulfide-using photosynthesis evolved in the ancestors of blue-green algae. These found a way to use a plentiful but poor reducing agent,

H_2O, by adding photocenter II as a second energy source. "Green plant" photosynthesis evolved, splitting water and liberating O_2. A slow alteration in the atmosphere began, which may have required hundreds of millions of years to complete, but which ultimately led to the present oxidizing atmosphere.

(6) With the availability of free oxygen, even in local concentrations, blue-green algae and purple non-sulfur bacteria evolved O_2 respiration as a means of extracting more energy from metabolites. A sun-powered oxygen cycle was set up, similar in principle to the sun-powered sulfur cycle developed earlier. Oxygen respiration could have arisen independently in purple non-sulfur bacteria and blue-green algae, and molecular structure comparisons among proteins of similar function may help to decide this point.

(7) Photosynthesis-impairing mutations may have occurred repeatedly among purple non-sulfur bacteria. As long as the atmosphere contained little free oxygen, these would have been deleterious, and would have been eliminated from the population by natural selection. But after blue-green algal photosynthesis had raised the oxygen content of the atmosphere beyond a certain level, respiration became efficient enough that the loss of photosynthesis was relatively harmless. Some of the present-day respiratory bacteria, including *Paracoccus*, evolved from purple non-sulfur bacteria by loss of photosynthesis. Others such as the gliding bacteria (*Beggiatoa*, *Thiothrix*) may have evolved from blue-green algae by a similar process.

This seven-point scenario was based originally on the distribution of *c*-type cytochromes among electron transfer chains in bacteria, and on sequence and structure comparisons of cytochromes *c*. It provides a coherent explanation for the observed similarities between cytochromes of photosynthesis and respiration, and agrees with what is known so far about bacterial energy metabolism. It accounts for the occurrence in prokaryotes and eukaryotes of a closely-knit subfamily of cytochromes c_2 (c_2, c_{550}, eukaryotic *c*), with a related but less uniform family of molecules possessing the essentials of the "cytochrome fold" (c_5, c_{553}, c_{555}, algal *f*). If this evolutionary model is correct, then it should be confirmed and even extended backward in time by similar structural comparisons of other electron-transfer proteins such as ferredoxins and cytochromes *b*. While the scenario may require correction in details, it at least offers a rational picture of the evolution of bacterial metabolism in terms of successive adaptations of old mechanisms to new uses when external conditions change, and of the gradual development and diversification of an almost universal energy-extracting mechanism.

We would like to thank Drs Martin Kamen and Barry Marrs for catalytic discussions about bacterial energy metabolism. We are indebted to Drs Lucile Smith and George McLain for the gift of the original sample of cytochrome c_{550}, to Drs E. Margoliash, W. Schroeder and L. Hood for advice and assistance in determining its amino acid sequence, and to Dr Henry Harbury for guidance in the isolation and purification of cytochrome c_{551}. This work was performed with the support of National Science Foundation grant no. BMS71-00825 and National Institutes of Health grant no. GM-12121. This paper is contribution no. 5144 from the Norman W. Church Laboratory of Chemistry Biology, California Institute of Technology.

REFERENCES

Ambler, R. P. (1975). In *Handbook of Biochemistry and Molecular Biology* (Fasman, G. D., ed.), 2nd edit. Chemical Rubber Co., Cleveland.

Ambler, R. P. & Bartsch, R. E. (1975). *Nature (London)*, **253**, 285–288.

Buehner, M., Ford, G. C., Moras, D., Olsen, K. W. & Rossmann, M. G. (1973). *Proc. Nat. Acad. Sci., U.S.A.* **70**, 3052–3054.

Connelly, J. L., Jones, O. T. G., Saunders, V. A. & Yates, D. W. (1973). *Biochim. Biophys. Acta*, **292**, 644–653.
Davis, K. A., Hatefi, Y., Salemme, F. R. & Kamen, M. D. (1972). *Biochem. Biophys. Res. Commun.* **49**, 1329–1335.
Dickerson, R. E. (1971). *J. Mol. Biol.* **57**, 1–15.
Dickerson, R. E. & Timkovich, R. (1975). In *The Enzymes* (Boyer, P., ed.), 3rd edit., vol. 11, pp. 397–547, Academic Press, New York.
Dickerson, R. E., Takano, T., Eisenberg, D., Kallai, O. B., Samson, L., Cooper, A. & Margoliash, E. (1971). *J. Biol. Chem.* **246**, 1511–1535.
Dickerson, R. E., Takano, T., Kallai, O. B. & Samson, L. (1972). In *Structure and Function of Oxidation–Reduction Enzymes* (Akeson, A. & Ehrenberg, A., eds), pp. 69–83, Pergamon Press, Oxford.
Doelle, H. W. (1969). *Bacterial Metabolism*, pp. 72, 157, Academic Press, New York.
Dus, K., Sletten, K. & Kamen, M. D. (1968). *J. Biol. Chem.* **243**, 5507–5518.
Dutton, P. L. & Wilson, D. F. (1974). *Biochim. Biophys. Acta*, **346**, 165–212.
Evans, M. C. W., Buchanan, B. B. & Arnon, D. I. (1966). *Proc. Nat. Acad. Sci., U.S.A.* **55**, 928–934.
Fuller, R. C. (1969). *Prog. Photosynthesis Res.* **3**, 1579–1592.
Gray, B. H., Fowler, C. F., Nugent, N. A. & Fuller, R. C. (1972). *Biochem. Biophys. Res. Commun.* **47**, 322–327.
John, P. & Whatley, F. R. (1975). *Nature (London)*, **254**, 495–498.
Jones, O. T. G. & Plewis, K. M. (1974). *Biochim. Biophys. Acta*, **357**, 204–214.
Kamen, M. D. & Vernon, L. P. (1955). *Biochim. Biophys. Acta*, **17**, 10–22.
Lascelles, J. & Wertlieb, D. (1971). *Biochim. Biophys. Acta*, **226**, 328–340.
Margulis, L. (1970). *Origin of Eukaryotic Cells*, Yale University Press, New Haven.
Marrs, B. & Gest, H. (1973*a*). *J. Bacteriol.* **114**, 1045–1051.
Marrs, B. & Gest, H. (1973*b*). *J. Bacteriol.* **114**, 1052–1057.
Needleman, S. B. & Blair, T. T. (1969). *Proc. Nat. Acad. Sci., U.S.A.* **63**, 1227–1233.
Parson, W. W. (1974). *Annu. Rev. Microbiol.* **28**, 41–59.
Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G. & North, A. C. T. (1960). *Nature (London)*, **185**, 416–422.
Peters, G. A. & Cellarius, R. A. (1972). *Biochim. Biophys. Acta*, **256**, 544–547.
Prince, R. C., Cogdell, R. J. & Crofts, A. (1974). *Biochim. Biophys. Acta*, **347**, 1–13.
Rossmann, M. G., Liljas, A., Brändén, C.-I., Banaszak, L. J. (1975). In *The Enzymes* (Boyer, P., ed.) 3rd edit., vol. 11, pp. 61–102, Academic Press, New York.
Salemme, F. R., Freer, S. T., Xuong, Ng. H., Alden, R. A. & Kraut, J. (1973). *J. Biol. Chem.* **248**, 3910–3921.
Saunders, V. A. & Jones, O. T. G. (1974). *Biochim. Biophys. Acta*, **333**, 439–445.
Smith, L., Newton, N. & Scholes, P. (1966). In *Hemes and Hemoproteins* (Chance, B., Estabrook, R. W. & Yonetani, T., eds), pp. 395–403, Academic Press, New York.
Sokatch, J. R. (1969). *Bacterial Physiology and Metabolism*, pp. 224, 254, Academic Press, New York.
Stanier, R. Y., Douderoff, M. & Adelberg, E. A. (1970). *The Microbial World*, 3rd edit., pp. 262–263, Prentice-Hall, New Jersey.
Stroud, R. M. (1974). *Scientific American*, **231**, no. 1, pp. 74–88.
Swanson, R., Trus, B. L., Mandel, N., Kallai, O. B. & Dickerson, R. E. (1976). *J. Biol. Chem.* In the press.
Takano, T., Kallai, O. B., Swanson, R. & Dickerson, R. E. (1973). *J. Biol. Chem.* **248**, 5234–5255.
Takano, T., Trus, B. L., Mandel, N., Kallai, O. B., Swanson, R. & Dickerson, R. E. (1976). *J. Biol Chem.* In the press.
Timkovich, R. & Dickerson, R. E. (1973). *J. Mol. Biol.* **79**, 39–56.
Timkovich, R. & Dickerson, R. E. (1976). *J. Biol. Chem.* In the press.
Timkovich, R., Margoliash, E. & Dickerson, R. E. (1976). *J. Biol. Chem.* In the press.
Walsh, K. A. & Neurath, H. (1964). *Proc. Nat. Acad. Sci., U.S.A.* **52**, 884–889.
Zannoni, D., Baccarini-Melandri, A. & Melandri, B. A. (1974). *FEBS Letters*, **48**, 152–155

APPENDIX II

Pseudomonas Cytochrome c_{551} at 2.0 Å Resolution: Enlargement of the Cytochrome c Family

Pseudomonas Cytochrome c_{551} at 2.0 Å Resolution: Enlargement of the Cytochrome *c* Family *

Robert J. Almassy and Richard E. Dickerson

Norman W. Church Laboratory of Chemical Biology, California Institute of Technology, Pasadena, California 91125, U.S.A.

---

* Contribution No. 5632 of the Norman W. Church Laboratory of Chemical Biology.

SUMMARY

The structure of Pseudomonas aeruginosa respiratory cytochrome c_{551}, with 82 amino acids, has been solved by x-ray analysis and refined to a crystallographic R factor of 16.2%. It has the same basic folding pattern and hydrophobic heme environment as cytochromes c, c_2, and c_{550}, except for a large deletion at the bottom of the heme crevice. This same "cytochrome fold" appears to be present in photosynthetic cytochromes c of green and purple sulfur bacteria, and algal cytochromes f, suggesting a common evolutionary origin for electron transport chains in photosynthesis and respiration.

(End of summary)

Cytochrome c_{551} is found in various Pseudomonads and Azotobacter vinelandii, where it plays a respiratory role analogous with mitochondrial cytochrome c in eukaryotes and cytochrome c_{550} in Paracoccus denitrificans[1-4]. It is significantly smaller than these latter proteins, however, having only 82 amino acids instead of 103-134. Amino acid sequence comparisons (Table 2 of Reference 4) have suggested a similarity of folding between c_{551} and the larger cytochromes, but in the absence of x-ray data it was difficult to decide where the "deletions" in the c_{551}chain should be placed to make the proper sequence alignment[5-8]. The preliminary low resolution x-ray analysis of Pseudomonas aeruginosa cytochrome c_{551} (discussed in Reference 4)

showed that the folding patterns in $\underline{c}$, $\underline{c}_2$, $\underline{c}_{550}$ and $\underline{c}_{551}$ were indeed the same. This has now been confirmed by x-ray analysis and constrained difference map refinement at 2.0 Å resolution.

Experimental procedures

The original cytochrome $\underline{c}_{551}$ was the gift of Henry Harbury, who then trained one of us (RJA) in the techniques of growth of cultures of *P. aeruginosa* and purification of the cytochrome[2] (and H. Harbury, personal communication). Crystals were grown in 40-50% saturated ammonium sulfate solutions, with 1 M NaCl and 0.01 M ammonium phosphate buffer, pH 5.6 - 5.9. The crystals are space group $P2_12_12_1$ with unit cell dimensions: $\underline{a}$ = 29.43 Å, $\underline{b}$ = 49.00 Å, $\underline{c}$ = 49.66 Å. Three heavy atom derivatives were prepared by soaking crystals in stock solutions: K_2PtCl_4, $UO_2(NO_3)_2$, and $NaAu(CN)_2$. All high-resolution data were collected on a modified General Electric XRD-490 x-ray diffractometer, to a resolution of 2.0 Å for native protein crystals and 2.4 Å for the isomorphous derivatives.

Full details of the structure analysis will be reported elsewhere, but the strategy may be outlined here. The mean figure of merit for multiple isomorphous replacement phase refinement[9] at 2.4 Å resolution was 0.924 for centric reflections and 0.763 for all data. Atomic coordinates were measured from a Kendrew wire model built in a Richards box. These were used as the starting point for refinement on a minicomputer, alternating cycles of (a) automated and

occasionally manual shifts in atomic position based on Fourier difference maps, and (b) adjustment of bond distances, angles, and torsion angles toward standard values followed by calculation of a new difference map[10]. At present, with stereochemically acceptable bond lengths and angles, a planar heme, and the addition of 42 water molecules per cytochrome molecule, the conventional crystallographic R factor* is 16.2% at 2.0 Å resolution.

As a rough indication of the effect of refinement, two sections through the plane of the heme group are compared in Figure 1, with heme skeletons superimposed. Figure 1a shows the multiple isomorphous replacement map at 2.4 Å resolution, and Figure 1b the refined $(2F_o - F_c)\, e^{i\phi c}$ map at 2.0 Å.

Description of the structure

The α-carbon atoms, heme group, and a few key side chains of cytochrome $\underline{c}_{551}$ are shown in the stereo drawings of Figure 2. The side chains and molecular orientation have been chosen to facilitate comparison with equivalent stereo drawings of cytochrome $\underline{c}$ from tuna[11-13], $\underline{c}_{550}$ from *Paracoccus*[3], and $\underline{c}_2$ from *Rhodospirillum rubrum*[14] (see stereos in Reference 4). The similarities and differences in

---

* $$R \equiv \frac{\sum_{hk\ell} ||F_o| - |F_c||}{\sum_{hk\ell} |F_o|} \times 100\%$$

main chain folding are perhaps easier to see in the ribbon drawings of cytochromes $\underline{c}_{551}$ and $\underline{c}$ in Figure 3. Among the common structural features are the amino-terminal alpha helix, cysteine and histidine attachments to the heme, the 20's loop at the right, the 60's helix (40's helix in $\underline{c}_{551}$), methionine ligand to the heme, and the carboxy-terminal alpha helix. The principal differences are the deletion of the bottom of the $\underline{c}$ molecule in $\underline{c}_{551}$, its replacement by a 30's helix at the lower rear, and the pulling downward of the 70's loop of $\underline{c}$ to close off the bottom of the $\underline{c}_{551}$ molecule. The deletion in cytochrome $\underline{c}$ required to produce the $\underline{c}_{551}$ molecule can be described approximately as a joining of residue 40 to 56, and removal of residues 41-55.

One interesting aspect, an apparent consequence of the deletion of chain at the bottom of the molecule, is the tilting of the heme within its polypeptide cage relative to that in cytochromes $\underline{c}$, $\underline{c}_2$, and $\underline{c}_{550}$. This is illustrated in the stereo drawings of front and right sides in Figure 2. These views were obtained by rotating 44 alpha-carbon atoms onto corresponding atoms in homologous regions of tuna cytochrome $\underline{c}$ using a least-squares fitting program. The polypeptide backbones thus have been oriented to correspond to Figure 16 of Reference 11 and a view 90° to the right, in which the heme is seen directly on edge and in the plane of the paper, respectively. Relative to the $\underline{c}$ orientation, the heme in $\underline{c}_{551}$ is rotated approximately 11° to the left about a vertical axis in Figure 2a, and tilted forward 16° about a horizontal axis. (A similar tilt of the heme has subsequently been observed in cytochrome $\underline{c}_{555}$, as mentioned in the

Discussion.)

These structure comparisons now enable a precise alignment of amino acid sequences to be made, as shown in Figure 4 for tuna c and Pseudomonas c_{551}. Key considerations in this alignment are the positions of the heme ligands, the structural equivalence of the valine-glycine-proline sequences at positions 28-30 in c and 23-25 in c_{551}, the correspondence between the 60's helix ending in asparagine 70 in c with the 40's helix ending in asparagine 50 in c_{551}, and the carboxy-terminal helices beginning at residue 87 in c and 67 in c_{551}. This new alignment requires a local three-residue shift to the left to be made in the middle of the c_{551} sequences in the more comprehensive Table 2 of Reference 4--that table having been made prior to the high-resolution structure analysis of c_{551}. It is obvious from the x-ray results that leucines 64 and 68, packed against the heme in tuna c, correspond to the similarly placed leucine 44 and isoleucine 48 of c_{551}.

The hydrophobic environment around the heme group is almost identical in the two proteins. Table 1 shows the comparison of structurally and sequentially equivalent hydrophobic side chains packed around the heme in 60 eukaryotic c sequences and six prokaryotic c_{551}. At only three places along the sequences is a hydrophobic heme contact in one protein not matched in the other. At two of these places the residue even remains hydrophobic, but is only turned so as to be in less obvious contact with the heme plane. This near-identity of heme environments is striking support for the evolutionary relatedness of

the two proteins in spite of their size differences.

Aromatic side chains are less well conserved (Table 2). A cytochrome heme seems to demand the nearby presence of aromaticity, but with the placement in three dimensions being less critical. Only the two rings in what has been called the "right channel"[15], positions 10 and 97 in *c* or 7 and 77 in c_{551}, are invariant in all species. None of the aromatic rings that once were invoked in the electron transfer mechanism--59, 67, 74, 82--has an exact equivalent in c_{551}. However, the tryptophan 59 that is hydrogen bonded to the inner propionic acid group of the heme in all molecules of *c*, c_{550}, and c_2 has a functional equivalent in tryptophan 56 of c_{551}, in a way that suggests evolutionary convergence. There can be no precise homologue of tryptophan 59 in c_{551}, since that portion of the chain is involved in the large deletion in c_{551}. (Residues 40 *et seq.* in c_{551} correspond in three-dimensional structure to 60 *et seq.* in *c*, but the two chains just before these positions adopt quite different conformations.) Nevertheless, c_{551} does have a tryptophan residue that occupies a corresponding position in the folded molecule and is hydrogen bonded to the buried propionic acid. Its position along the amino acid chain of c_{551}, number 56, corresponds to the mid-70's region of eukaryotic cytochrome *c*. It is as though the molecule, lacking an essential tryptophan at one position, developed a compensating residue somewhere else, and made the requisite main chain adjustments to bring the tryptophan to its former position. These adjustments, a drawing down of the loop at the left of the heme, also help to close the "wound" at

the bottom of the molecule created by the large deletion in $\underline{c}_{551}$.

The environments of the buried and exposed propionic acid groups on the heme are similar in $\underline{c}$ and $\underline{c}_{551}$, although the results are achieved in different ways from a sequence standpoint. The aromatic rings of invariant tyrosine 48 and tryptophan 59 in $\underline{c}$ are matched by invariant tryptophan 56 and semi-invariant phenylalanine/tyrosine/asparagine 34 around the buried propionic group in $\underline{c}_{551}$. Serine/threonine 49 and invariant threonine 78 in $\underline{c}$, hydrogen bonded to the outer propionic group, are paralleled by a semi-invariant serine/threonine at position 52 or 53 in $\underline{c}_{551}$, also hydrogen bonded to the outer propionic.

These examples of structural convergence are reminiscent of the replacement of phenylalanine 11 by phenylalanine 20 in cytochrome $\underline{c}_2$ of R. rubrum, (but not in any other $\underline{c}_2$ molecule), and the compensating axial rotation of the carboxy-terminal alpha helix to adjust the position of tyrosine 107 (see Reference 4 for details). In all known cytochrome $\underline{c}$ structures--$\underline{c}$, $\underline{c}_2$, $\underline{c}_{550}$ and $\underline{c}_{551}$--there seems to be a need for (1) a tryptophan residue hydrogen bonded to the buried heme propionic, (2) another nearby aromatic ring, (3) a serine/threonine hydrogen bond to the outer propionate, and (4) a pair of aromatic side chains to the right of the heme.

One of the most remarkable features of the amino acid sequence of cytochrome $\underline{c}_{551}$ is the presence of so many prolines around the heme-liganding methionine. In Pseudomonas aeruginosa, the sequence is:

Trp56--Gly57--Pro58--Ile59--Pro60--Met61--Pro62--Pro63--Asn64

This was a cause for curiosity and discussion prior to the structure analysis. The answer, in Figure 5, is obvious by hindsight. In each stereo pair of the figure, the right polypeptide chain is a drawing of residues 58-63 of c_{551} using refined coordinates, and the left polypeptide is an idealized polyproline helix using coordinates from Reference 16. This portion of the chain in c_{551} is essentially a short stretch of threefold polyproline helix, with the heme-liganding methionine side chain branching off from its midpoint. This seems to be a structural device to ensure an extended chain conformation, necessary to bridge the long distance between tryptophan 56 at the bottom of the molecule and the carboxy-terminal alpha helix at the top. Szent-Gyorgyi and Cohen proposed the use of a high proline content to induce a polyproline conformation in proteins just twenty years ago[17]. In an introduction to a collagen symposium a few years later, G. N. Ramachandran half-seriously suggested that since the α-helix and β-sheet fibrous structures had been found in globular proteins, we might someday expect to find a globular protein with a collagen helix. It now appears that he was correct in principle!

The highly asymmetric distribution of positively and negatively charged side chains that has been remarked for other cytochromes c[3,14,15] is present in cytochrome c_{551} also. The molecule has 8 lysines, 1 arginine, 5 aspartic and 5 glutamic acids, for a net charge count among side chains of -1. However, if a plane is drawn parallel

to the page in Figure 2a, through the center of gravity of the molecule, then six positive charges (residues 8, 10, 28, 33, 47, and 49) and one negative charge (residue 29) lie on this bisecting perimeter, eight negative charges (1, 2, 4, 41, 43, 68, 69, 70) and two positive (76, 82) lie on the back half of the molecular surface, and only one side chain of each charge (aspartic 19 and lysine 21) are found on the front hemisphere. The cytochrome $\underline{c}_{551}$ molecule is essentially a sphere with the heme crevice opening to a hydrophobic front hemisphere, diametrically opposed to a negatively charged back hemisphere, with a belt of positive charges separating the two. The positive charges around the perimeter of the heme face of the molecule, that have been commented upon in other cytochromes c, are present here also. Lysines 10 at the top of the crevice, 21 at the right, and 47 at the left, correspond to lysines 13, 25/27, and 72/73 in eukaryotes. These residues are evolutionarily invariant among the six $\underline{c}_{551}$ sequences known (as are most of the other lysines). The acidic side chains are more variable, as was also the case for the eukaryotic cytochromes c.

The family of cytochromes c

The x-ray crystal structure shows clearly that cytochrome $\underline{c}_{551}$ belongs in the same evolutionary family with eukaryotic, mitochondrial c, Paracoccus $\underline{c}_{550}$, and $\underline{c}_2$ from purple nonsulfur photosynthetic/respiratory bacteria such as Rhodospirillum. Possible evolutionary relationships between these proteins, and between their host

organisms, have been discussed in Reference 4. At the time of that paper it appeared as if cytochromes c, c_2, and c_{550} were the most closely related group, and that the c_{551}'s with their chain deletion at the bottom of the molecule were more distantly related. Among the cytochromes c_2 from various purple nonsulfur photosynthetic bacteria that had been sequenced by Richard Ambler and others[18,19] were molecules that were as small as eukaryotic c or as large as c_{550}, but with nothing in the size range of c_{551}. Recent work has changed this picture (R. Ambler, private communication). Sequences have been determined for at least one c_2 from every formal species listed in Bergey's Manual[20]. In at least two of the three genera, Rhodospirillum and Rhodopseudomonas, Ambler found species containing cytochromes c_2 that in both size and amino acid sequence appear to be homologous with Pseudomonas c_{551}. It appears that all of these small cytochromes, from c_{551} with 82 amino acids to c_{550} with 134, should be considered as one evolutionary family, with a standard pattern of acceptable insertions and deletions that is followed in diverse organisms: respirers and photosynthesizers, prokaryotes and eukaryotes.

Other members of this family probably include cytochrome c_{553} of purple sulfur bacteria (Chromatiaceae) with 112 amino acids, c_{555} of green sulfur bacteria (Chlorobiaceae) with 86 amino acids, cytochromes f or c_{554} of prokaryotic and eukaryotic algae with 83-89 amino acids, and perhaps even the single-heme cytochrome c_{553} of the sulfate-respiring Desulfovibrio with 82 amino acids. Salemme has recently reported the folding of the polypeptide chain backbone in cytochrome

c_{555} of *Chlorobium thiosulfatophilum*[21] as obtained from an unrefined MIR analysis at 2.7 Å resolution, and it is apparent that this backbone is virtually identical with that reported for *Pseudomonas* c_{551} in Reference 4 and this paper. The same large deletion is observed at the bottom of the molecule, and from the ribbon drawing of the c_{555} chain path, the heme appears to be tilted forward as in c_{551}. One interesting difference observed in c_{555} is that the tryptophan hydrogen bonded to the buried propionic acid (R. Salemme, private communication) is present in the *same* position along the chain in eukaryotic cytochromes *c* and all published c_{555} sequences, even though these two proteins probably are the most distantly related in an evolutionary sense. The shifted tryptophan observed in c_{551} may represent a special adaptation in the *Pseudomonads*.

The probable evolutionary relationships between the cytochrome *c*-containing metabolic pathways in bacteria and eukaryotes are outlined in Figure 6, which is an extension of an earlier metabolic tree in Reference 4. At least the main chain pathway is now known for five of the cytochromes indicated in this figure: c_{555}, c_2, c_{551}, c_{550}, and *c*, in the probable order of evolution. The "small" form of cytochrome *c* with the bottom chain deletion is found in green sulfur and purple nonsulfur photosynthetic bacteria, respiratory bacteria (c_{551}) and cyanobacteria or blue-green algae. The "large" form of cytochrome *c*, in contrast, has been observed so far only in the branch of Figure 6 leading to the purple sulfur and nonsulfur bacteria. Hence it seems more likely that the small cytochrome is the ancestral form, and that

one branch has seen the addition of more chain at the bottom of the heme crevice.

All of the bacteria discussed in this paper and shown in Figure 6 are similar enough in general morphology to be grouped by Brock[22] into one category: the gram negative, polarly flagellated rods. It is interesting to find this morphological similarity matched by a biochemical similarity: all members perform either photosynthesis or respiration or both, all possess electron transport chains with a cytochrome c near the high-potential end, and all of these cytochromes c appear to be structurally or sequentially homologous. This class of cytochromes arose soon after bacteria developed the ability to trap light and use it for chemical purposes, and has been retained in the divergent evolution of present-day photosynthesis and respiration.

REFERENCES

1. Horio, T., Higashi, T., Sasagawa, M., Kusai, K., Nakai, M., and Okunuki, K., Biochem. J. 77, 194-201 (1960).
2. Ambler, R., Biochem. J. 89, 341-349 and 349-378 (1963).
3. Timkovich, R. and Dickerson, R.E., J. Biol. Chem. 251, 4033-4046 (1976).
4. Dickerson, R.E., Timkovich, R., and Almassy, R.J., J. Mol. Biol. 100, 473-491 (1976).
5. Needleman, S.B. and Blair, T.T., Proc. Nat. Acad. Sci. USA, 63, 1227-1233 (1969).
6. Dickerson, R.E., J. Mol. Biol. 57, 1-15 (1971).
7. McLachlan, A.D., J. Mol. Biol. 61, 409-424 (1971).
8. Dayhoff, M.O., Atlas of Protein Sequence, Vol. 5, Suppl. 2, pp. 26-27, Washington (1976).
9. Dickerson, R.E., Weinzierl, J.E., and Palmer, R.A., Acta Crystallogr. B24, 997-1003 (1968).
10. Chambers, J.L. and Stroud, R.M., Acta Crystallogr. B33, 1824- (1977).
11. Swanson, R., Trus, B.L., Mandel, N., Mandel, G., Kallai, O.B., and Dickerson, R.E., J. Biol. Chem. 252, 759-775 (1977).
12. Takano, T., Trus, B.L., Mandel, N., Mandel, G., Kallai, O.B., Swanson, R., and Dickerson, R.E., J. Biol. Chem. 252, 776-785 (1977).

13. Mandel, N., Mandel, G., Trus, B.L., Rosenberg, J., Carlson, G., and Dickerson, R.E., J. Biol. Chem. 252--in press 1977.

14. Salemme, F.R., Freer, S.T., Xuong, Ng. H., Alden, R.A., and Kraut, J., J. Biol. Chem. 248, 3910-3921 (1973).

15. Dickerson, R.E., Takano, T., Eisenberg, D., Kallai, O.B., Samson, L., Cooper, A., and Margoliash, E., J. Biol. Chem. 246, 1511-1535 (1971).

16. Sasisekharan, V., Acta Crystallogr. 12, 897-903 (1959). (Note: The x coordinate of O(1) in Table 5 should have a negative sign.)

17. Szent-Gyorgyi, A.G. and Cohen, C., Science, 126, 697-698 (1957).

18. Ambler, R.P., in CRC Handbook of Biochemistry and Molecular Biology, 3rd edition (G.D. Fasman, ed.), Proteins Section, Vol. III, pp. 294-298 (CRC Press, Cleveland, 1976).

19. Ambler, R.P., Meyer, T.E., and Kamen, M.D., Proc. Nat. Acad. Sci. USA, 73, 472-475 (1976).

20. Buchanan, R.E. and Gibbons, N.E. (eds.). Bergey's Manual of Determinative Bacteriology. 8th ed. (Williams and Wilkins, Baltimore, 1974).

21. Korszun, Z.R. and Salemme, F.R., Proc. Nat. Acad. Sci. USA (1977)--in the press.

22. Brock, T.D., Biology of Microorganisms (Prentice Hall, New Jersey, 1970), pp. 579-606.

23. Dickerson, R.E. and Timkovich, R., in The Enzymes, 3rd edition (Paul Boyer, ed.), Vol. XI, pp. 397-547 (Academic Press, New York, 1975).

TABLE 1. Hydrophobic heme contacts in cytochromes *c* and c_{551}

Cytochrome *c*		Cytochrome c_{551}	
Tuna	60 Eukaryotes	P. aeruginosa	6 Prokaryotes
Phe 10	60 Phe	Phe 7	5 Phe, 1 Tyr
Pro 30	60 Pro	Pro 25	6 Pro
Leu 32	60 Leu	Tyr 27	3 Phe, 2 Leu, 1 Tyr
Leu 35	51 Leu, 5 Ilu, 2 Val, 2 Phe	Val 30	6 Val
Leu 64	57 Leu, 2 Met, 1 Phe	Leu 44	5 Leu, 1 Ile
Tyr 67	59 Tyr, 1 Phe	No heme contact in c_{551}	
Leu 68	60 Leu	Ile 48	6 Ile
Pro 71	60 Pro	Gly 51	6 Gly
No heme contact in *c*		Pro 62	6 Pro
Phe 82	60 Phe	No heme contact in c_{551}	
Ile 85	36 Leu, 24 Ile	Val 66	6 Val
Leu 94	58 Leu, 2 Ile	Leu 74	6 Leu
Val 95	55 Ile, 4 Val, 1 Leu	Ala 75	6 Ala
Leu 98	57 Leu, 3 Met	Val 78	4 Val, 2 Ile

Eukaryotic cytochrome *c* sequences from Tables IX and X of Reference 23, prokaryotic c_{551} sequences from Table 2 of Reference 4.

TABLE 2. Aromatic side chains in cytochromes *c* and c_{551}

Cytochrome *c*		Cytochrome c_{551}	
Tuna	60 Eukaryotes	*P. aeruginosa*	6 Prokaryotes
Phe 10	60 Phe	Phe 7	5 Phe, 1 Tyr
Not aromatic		Tyr 27	3 Phe, 2 Leu, 1 Tyr
Phe 36	53 Phe, 3 Ile, 2 Val, 2 Tyr	Not aromatic	
Not aromatic		Phe 34	3 Tyr, 2 Asn, 1 Phe
Tyr 46	33 Tyr, 27 Phe	Deletion region	
Tyr 48	60 Tyr	Deletion region	
Trp 59*	60 Trp	Trp 56*	6 Trp
Tyr 67	59 Tyr, 1 Phe	Not aromatic	
Tyr 74	58 Tyr, 2 Phe	Not aromatic	
Phe 82	60 Phe	Not aromatic	
Tyr 97	59 Tyr, 1 Phe	Trp 77	6 Trp

*Molecular convergence: same position in three dimensions, although different positions along polypeptide chains.

FIGURE CAPTIONS

FIGURE 1. (a) Section through the heme group, from the multiple isomorphous replacement electron density map of Pseudomonas aeruginosa cytochrome c_{551}. Heme coordinates as obtained from the MIR map are superimposed. (b) Section through the heme group in the final $(2F_o - F_c)$ map from the refined structure, with the refined coordinates superimposed. The five-membered rings appear as flat plates, with their substituent groups clearly defined, and the six-membered rings including the heme iron have deep negative centers. The cysteine 12 attachment is at the left side of the lower pyrrole ring, and the cysteine 15 attachment is at the top of the left pyrrole ring. The two propionic acid groups extend out of the sectioning plane at the upper right corner.

FIGURE 2. Stereo drawings of the main chain backbone, heme, and key side chains in c_{551}. (a) Front view, with the main chain oriented as in the stereos of c, c_2, and c_{550} in Reference 4. (b) Right side view, at 90° from the front view. Notice the tilt of the heme compared with the stereos of the other c's. Residues 58-63 form a nearly ideal polyproline threefold helix, as shown in Figure 5.

FIGURE 3. Ribbon drawings of main chain pathways in (a) cytochrome c_{551} from P. aeruginosa and (b) cytochrome c from tuna. The basic molecular folding is the same. One can think of making c_{551}

from c by joining residues 40 and 56, deleting the intervening residues, and pulling the loop at the left downward to close off the bottom of the molecule once more. The heme group with its iron are represented by a cross-hatched slab with a central black ball. Covalent cysteine attachments, the methionine and histidine iron ligands, and the inner and outer heme propionic acid groups (at the bottom) are also shown.

FIGURE 4. Amino acid sequence alignments for tuna cytochrome c and P. aeruginosa c_{551}, made on the basis of the x-ray analyses. Brackets with α indicate regions of α helix. Ala 17-Ile 18 in c_{551} are considered α-helical because they make the proper hydrogen bonds to the beginning of helix 26-34.

FIGURE 5. Stereo pair drawings of an idealized polyproline threefold helix (left in each stereo pair) and the c_{551} chain from residues 58 to 63, showing the similarity in conformation. The Pro-Ile-Pro-Met-Pro-Pro sequence apparently is a device for obtaining an extended chain configuration past the site for the sixth (methionine) ligand to the heme iron.

FIGURE 6. Proposed family tree of the gram-negative, polarly flagellated rod bacteria and their offshoots. These include the photoautotrophs and chemoautotrophs, and the respiring organisms that are believed to have evolved from them, including Pseudomonas

aeruginosa. Included are the probable steps in metabolic evolution (rectangular boxes), and the identity and size of the cytochromes c that are related by structure and probably by descent from a common ancestral gene. For more discussion and justification, see References 4 and 23. Abbreviations: GSB = Green sulfur bacteria, PSB = Purple sulfur bacteria, PNSB = Purple non-sulfur bacteria, MITO = Mitochondria, PARA = Paracoccus, PSEU = Pseudomonas, BEGG = Beggiatoa and related gliding bacteria, BGA = Blue-green algae or Cyanobacteria, CHLO = Chloroplasts, DESUL = Desulfovibrio.

FIGURE 1a

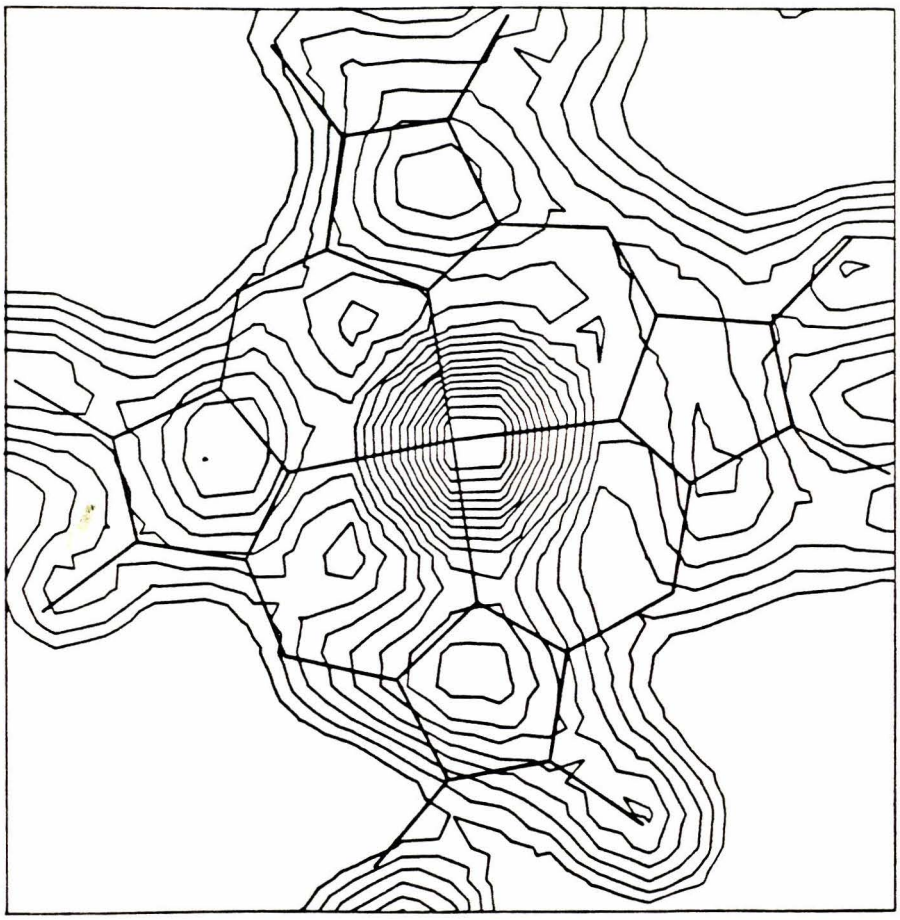


FIGURE 1b

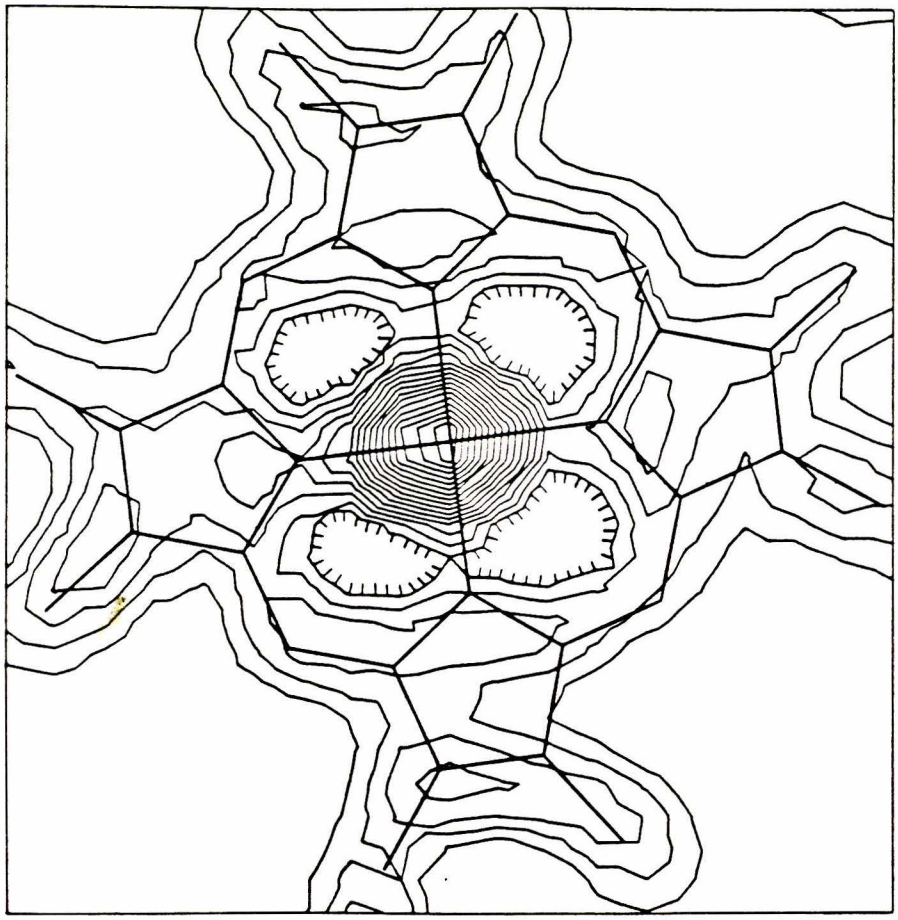


FIGURE 2a

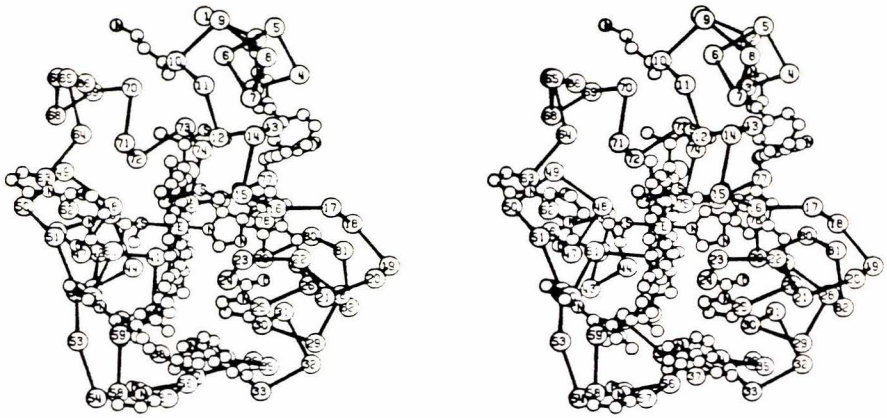


FIGURE 2b

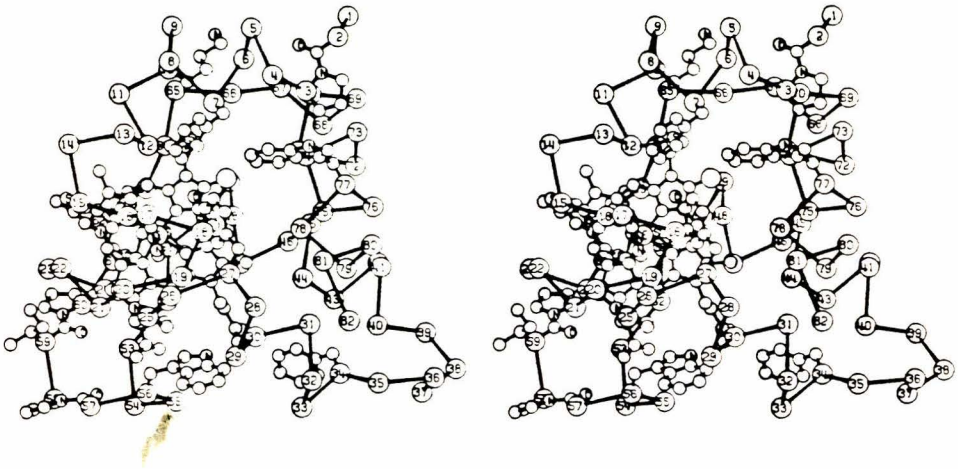


FIGURE 3a

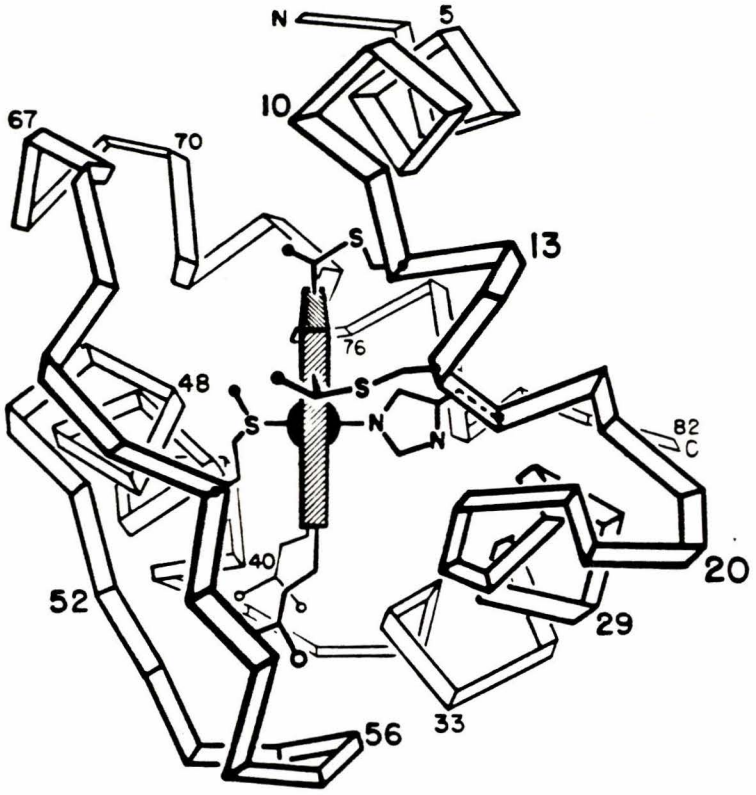


FIGURE 3b

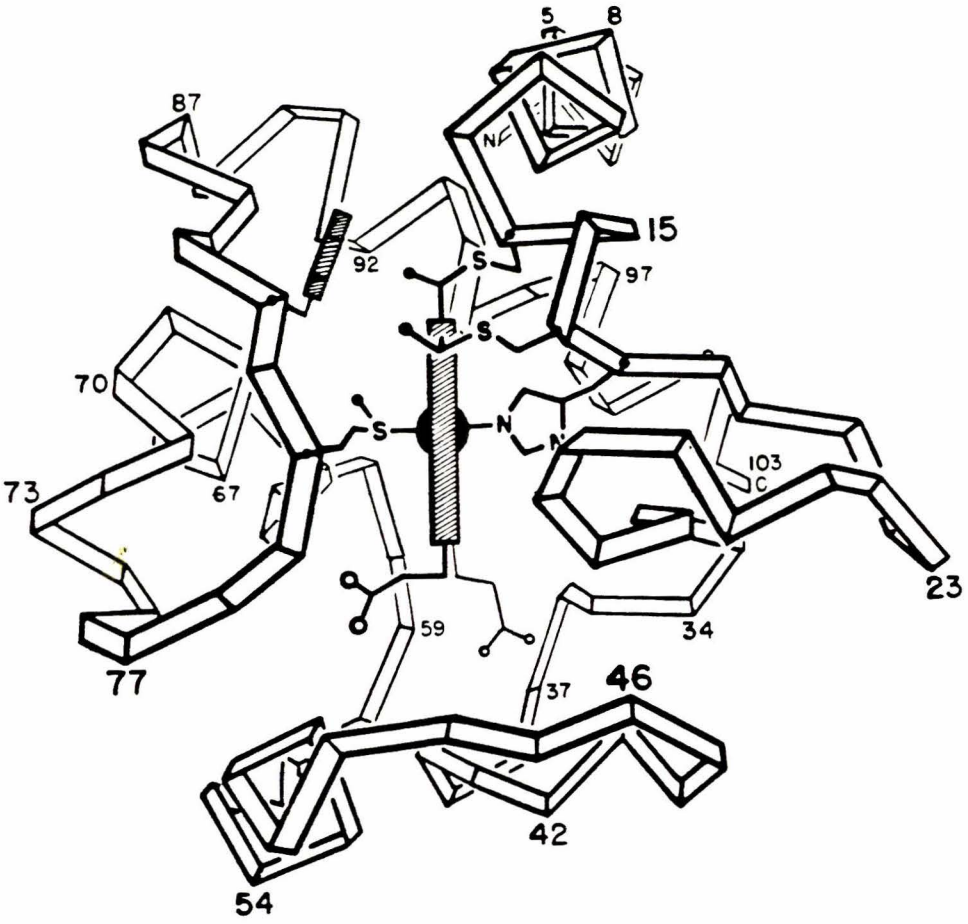


FIGURE 4

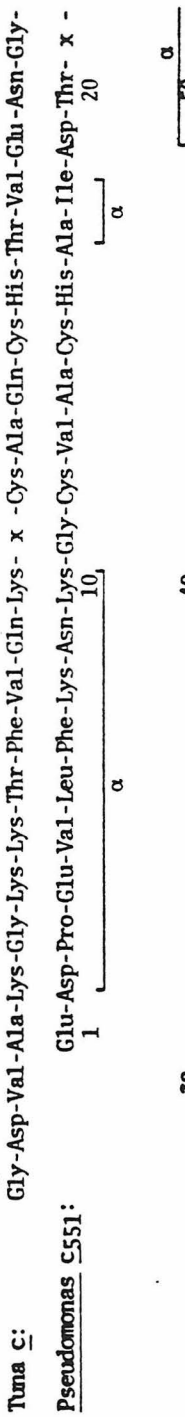

Tuna c:
Gly-Asp-Val-Ala-Lys-Gly-Lys-Lys-Thr-Phe-Val-Gln-Lys- x -Cys-Ala-Gln-Cys-His-Thr-Val-Glu-Asn-Gly-
Pseudomonas c551:
Glu-Asp-Pro-Glu-Val-Leu-Phe-Lys-Asn-Lys-Gly-Cys-Val-Ala-Cys-His-Ala-Ile-Asp-Thr- x -
-Gly-Lys-His-Lys-Val-Gly-Pro-Asn-Leu-Trp-Gly-Leu-Phe-Gly-Arg-Lys-Thr-Gly-Gln-Ala-Glu-Gly-Tyr-Ser-Tyr-Thr-Asp-Ala-
- x - x -Lys-Met-Val-Gly-Pro-Ala-Tyr-Lys-Asp-Val-Ala-Ala-Lys-Phe-Ala- x - x - x - x - x - x - x - x - x - x -
-Asn-Lys-Ser-Lys-Gly-Ile-Val-Trp-Asn-Asn-Asp-Thr-Leu-Met-Glu-Tyr-Leu-Glu-Asn-Pro-Lys-Lys-Tyr-Ile- x -Pro-Gly-Thr-
- x - x - x - x -Gly-Gln-Ala-Gly-Ala-Glu-Ala-Glu-Leu-Ala-Gln-Arg-Ile-Lys-Asn-Gly-Ser-Gln-Gly-Val-Trp-Gly-Pro-Ile-
-Lys-Met-Ile-Phe-Ala-Gly-Ile-Lys-Lys-Lys-Gly-Glu-Arg-Gln-Asp-Leu-Val-Ala-Tyr-Leu-Lys-Ser-Ala-Thr-Ser.
-Pro-Met-Pro-Pro-Asn-Ala-Val- x -Ser-Asp-Asp-Glu-Ala-Gln-Thr-Leu-Ala-Lys-Trp-Val-Leu-Ser-Gln-Lys.

FIGURE 5

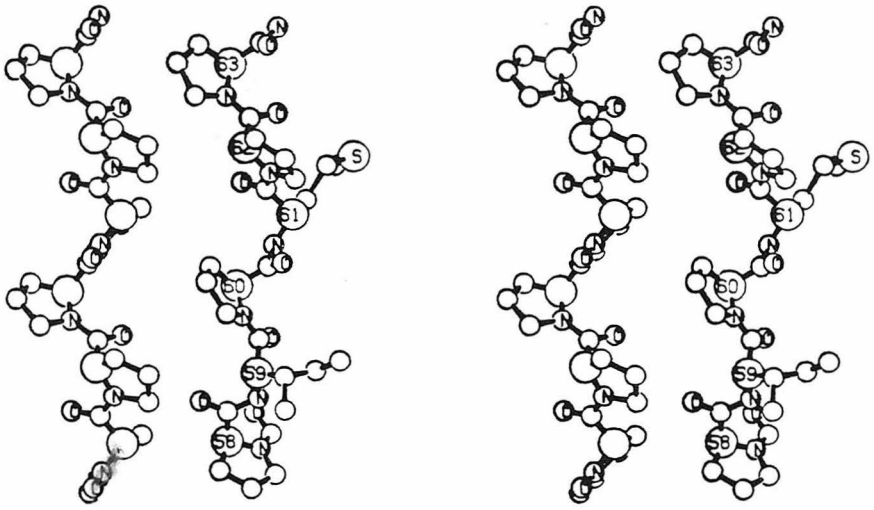
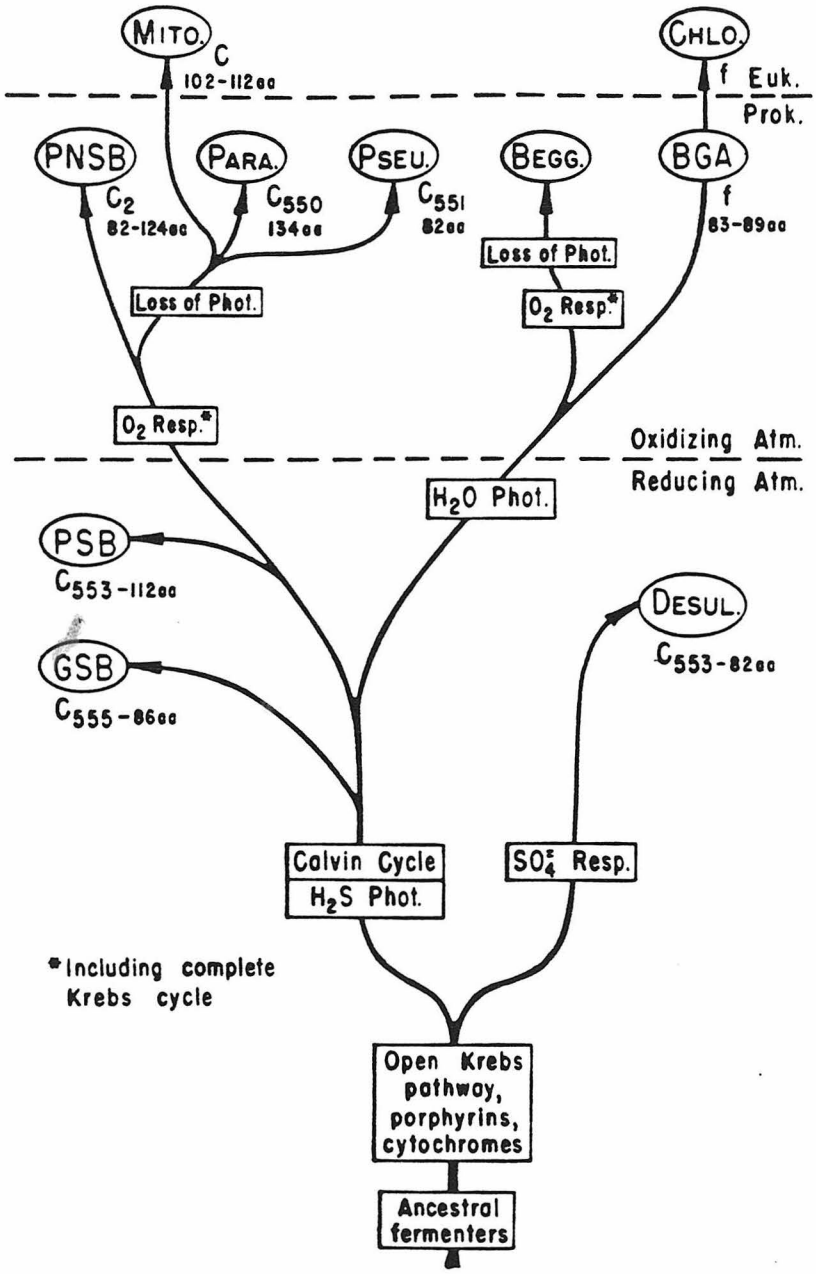


FIGURE 6


MITO.
C
102-112aa
CHLO.
f
Euk.
Prok.
PNSB
C2
82-124aa
PARA.
C550
134aa
PSEU.
C551
82aa
BEGG.
BGA
f
83-89aa
Loss of Phot.
Loss of Phot.
O2 Resp.*
O2 Resp.*
Oxidizing Atm.
Reducing Atm.
H2O Phot.
PSB
C553-112aa
DESUL.
C553-82aa
GSB
C555-86aa
Calvin Cycle
H2S Phot.
SO4= Resp.
*Including complete Krebs cycle
Open Krebs pathway, porphyrins, cytochromes
Ancestral fermenters

APPENDIX III

Refinement History of Cytochrome c_{551}

CYTOCHROME c_{551} REFINEMENT HISTORY

R Factor = 48.6 — Wire model built to 2.4Å resolution 3-derivative M.I.R. map and a complete set to 5431 F_{calc} (to 1.96Å resolution) evaluated

9 computer cycles of shifts and constraints in M.I.R. map including 8 shifts and 14 constraints

43.3 — First $(F_{obs} - F_{calc})\phi_{calc}$ map calculated

4 computer cycles of constrained ΔF refinement including 7 shifts and 7 constraints

Temperature factors for heme decreased (B_{heme}=7, $B_{protein}$=14)

33.7 — Minimap I plotted and coordinates adjusted (Paco016)

R = 36.2 after coordinate adjustments

5 computer cycles of constrained ΔF refinement including 5 shifts and 5 constraints

Begin temperature factor refinement

Eliminate reflections with F_{obs} = 0 (ΔR = -2.9, # refl. = 5050)

Add 19 solvent molecules (ΔR = +0.20)

25.8 — Minimap II plotted and coordinate and solvent adjusted (Paco031)

2 solvents deleted and 18 solvents added (total of 35)

R = 30.0 after coordinate and solvent adjustments

10 computer cycles of constrained ΔF refinement including 10 shifts and 9 constraints

Eliminated reflections with $d > 7.0$Å (ΔR = -0.70, # refl. = 4903)

Eliminated reflections with $F_{obs} < 2\sigma$(unobserved) (ΔR = -1.3, # refl. = 4625)

21.5 — Minimap-III plotted and coordinate and solvent adjusted (Paco048)

18 solvents deleted and 25 solvents added (total of 42)

R = 26.4 after coordinate and solvent adjustments

13 computer cycles of constrained ΔF refinement including 13 shifts and 9 constraints

Difference between planar Fe and non-planar Fe R Factors = -0.6%

18.9 — Minimap IV plotted and coordinate and solvent adjusted (Paco065)

13 solvents deleted and 23 solvents added (total of 52)

R = 22.2 after coordinate and solvent adjustments

12 computer cycles of constrained ΔF refinement including 12 shifts and 5 constraints

17.5 — Minimap V plotted and coordinate and solvent adjusted (Paco078)

15 solvents deleted and 5 solvents added (total of 42)

R = 21.1 after coordinate and solvent adjustments

15 computer cycles of constrained ΔF refinement including 15 shifts and 1 constraints

ΔF map grid interval decreased from 0.85Å to 0.50Å

16.2 — Final coordinates

```
********** SHFXXX *******
* FILE 1                *
* FILE 2                *
* A.AA B.BB CC.C        *
* DD EE FF G.G          *
* HH.H II.I             *
* J.JJJ                 *
*************************
```

SHFXXX - Atomic Shifting Program with output coordinate filename SHFXXX (old version with proportional curvatures, Chambers and Stroud, Acta Cryst B33, 1824, 1977)

FILE 1	- Input coordinate filename	FF	- RMS z coord shift x 10^4 (fr)
FILE 2	- Input Fourier map filename	G.G	- RMS temperature factor shift
A.AA	- Shift factor for coordinates	HH.H	- Average map density @ starting coord
B.BB	- Shift factor for temp. factors	II.I	- Average map density @ ending coord
CC.C	- Overall temperature factor	J.JJJ	- Average coordinate shift (Å)
DD	- RMS x coord shift x 10^4 (fr)		
EE	- RMS y coord shift x 10^4 (fr)		

```
********** SHFXXX *******
* FILE 1       CURVYXXX *
* FILE 2                *
* A.AA B.BB             *
* CC DD EE F.FFFF       *
* G.GGG H.HHH           *
* I.III J.JJJ K.KKKKK   *
*************************
```

SHFXXX - Atomic Shifting Program with output coordinate filename SHFXXX (new version with real curvatures)

FILE 1	- Input coordinate filename	EE	- RMS z coord shift x 10^4 (fr)
CURVYXXX	Input curvature filename	F.FFFF	- RMS temperature factor shift
FILE 2	- Input Fourier map filename	G.GGG	- Average map density @ starting coord
A.AA	- Shift factor for coordinates	H.HHH	- Average map density @ ending coord
B.BB	- Shift factor for temp. factors	I.III	- Average coordinate shift (Å)
CC	- RMS x coord shift x 10^4 (fr)	J.JJJ	- RMS coordinate shift (Å)
DD	- RMS y coord shift x 10^4 (fr)	K.KKKKK	- Average map curvature @ atomic centers

```
********** SCRXXX *******
* FILE                  *
* CONSX                 *
* A.AA B.BB C.CC D.DD   *
* EE                    *
* F.FFF G.GGG H.HHH     *
* I.III J.JJJ K.KKK     *
*************************
```

SCRXXX - Constraints Program with output coordinate filename SRCXXX

FILE	- Input coordinate filename	F.FFF	- Starting ave. bond dist. deviations
CONSX	- Constraints filename	G.GGG	- Starting ave. bond angle deviations
A.AA	- Overall wt. - Bond distances	H.HHH	- Starting ave. tor. angle deviations
B.BB	- Overall wt. - Bond angles	I.III	- Final ave. bond dist. deviations
C.CC	- Overall wt. - Torsion angles	J.JJJ	- Final ave. bond angle deviations
D.DD	- Overall wt. - Asymmetric atoms	K.KKK	- Final ave. tor. angle deviations
EE	- # iterations per molten zone		

```
********** FACOXXX ******
* FILE                  *
* AA.A                  *
* BBB                   *
* CCCC                  *
*                       *
*                       *
*************************
```

FACOXXX - Structure Factor Program followed by Fourier transformation program with output map filename FACOXXXMAP, FACOXXXPLT, or FACOXXXBIG

FILE	- Input coordinate filename
AA.A	- R Factor
BBB	- # of refl. omitted from calc. because $(F_o-F_c)/(\frac{1}{2}(F_o+F_c)) > 120\%$
CCCC	- Total number of reflections

```
********** CURVY200 *****
* FILE                  *
* AAAA                  *
*                       *
*                       *
*                       *
*                       *
*************************
```

CURVYXXX- Curvature Program for calculation of atomic curvatures

FILE	- Input coordinate filename
AAAA	- # of reflections included in calculation

```
**************************
* CONS ------------------*
* DISTANCE WT. = 2       *
* ANGLE WT. = 3          *
* TORSION ANGLE WT. = 4  *
* ALL CONSTRAINTS        *
*                        *
**************************
```

```
**************************
* CONSE -----------------*
* DISTANCE WT. = 6       *
* ANGLE WT. = 9          *
* TORSION ANGLE WT. = 12 *
* ALL CONSTRAINTS TO FE  *
* REMOVED                *
**************************
```

```
**************************
* CONSA -----------------*
* DISTANCE WT. = 5       *
* ANGLE WT. = 10         *
* TORSION ANGLE WT. = 10 *
* ALL CONSTRAINTS        *
*                        *
**************************
```

```
**************************
* CONSF -----------------*
* HEME BOND DISTANCES    *
* ONLY ARE CONSTRAINED   *
* DISTANCE WT. = 6       *
*                        *
*                        *
**************************
```

```
**************************
* CONSB -----------------*
* DISTANCE WT. = 20      *
* ANGLE WT. = 30         *
* TORSION ANGLE WT. = 40 *
* ALL CONSTRAINTS        *
*                        *
**************************
```

```
**************************
* CONSC -----------------*
* DISTANCE WT. = 2       *
* ANGLE WT. = 3          *
* TORSION ANGLE WT. = 4  *
* SOME FE CONSTRAINTS    *
* WEIGHTS SET TO ZERO    *
**************************
```

```
**************************
* CONSD -----------------*
* DISTANCE WT. = 6       *
* ANGLE WT. = 9          *
* TORSION ANGLE WT. = 12 *
* ALL CONSTRAINTS        *
*                        *
**************************
```

Map grid size to Fourier transformations

	A axis	B axis	C axis
PACOXXXMAP	0.0312(0.92Å)	0.0167(0.82Å)	0.0167(0.82Å)
PACOXXXPLT	0.0222(0.65Å)	0.0125(0.61Å)	0.0125(0.62Å)
PACOXXXBIG	0.0167(0.49Å)	0.0100(0.49Å)	0.0100(0.50Å)

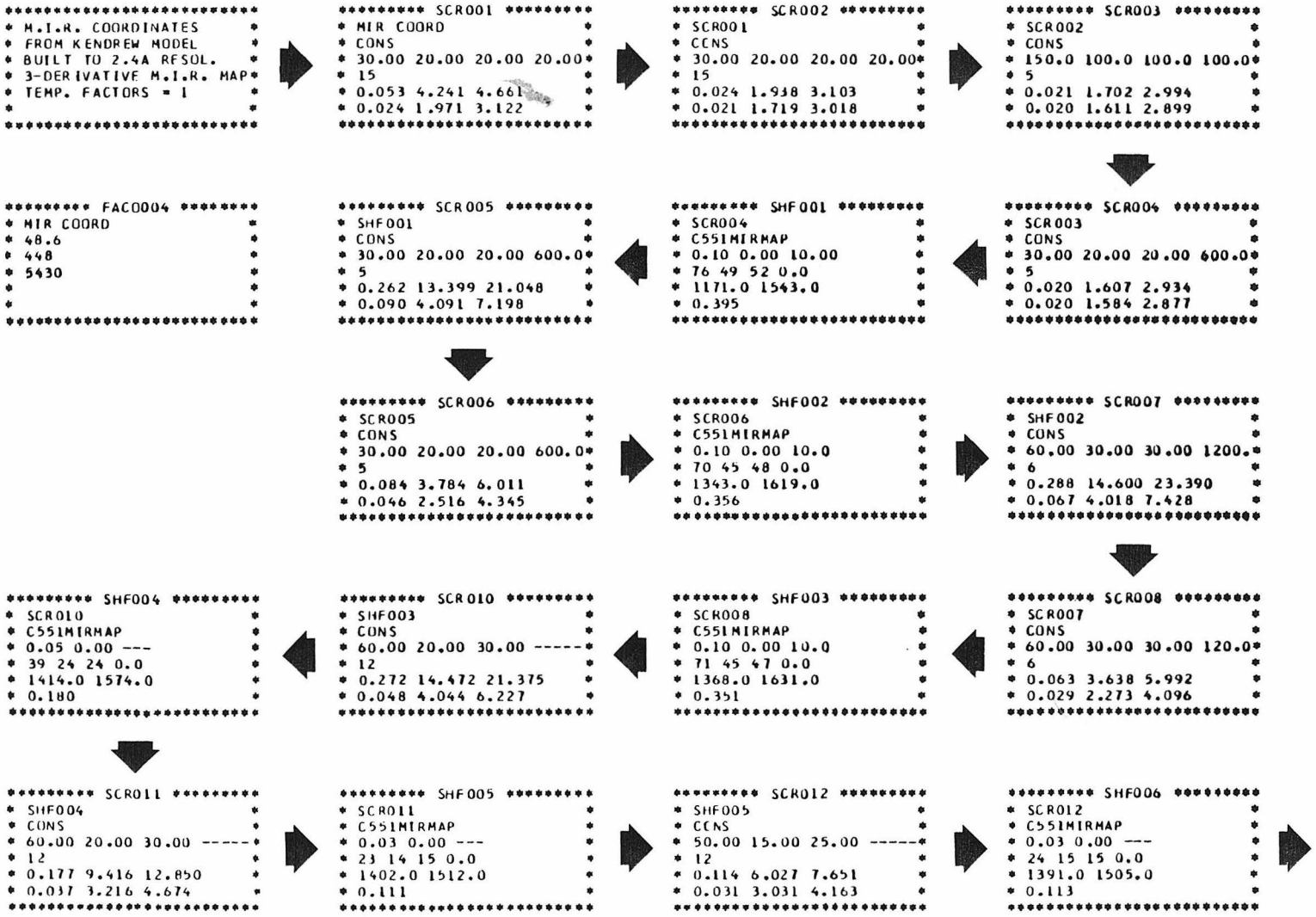
F_{obs} Reflections

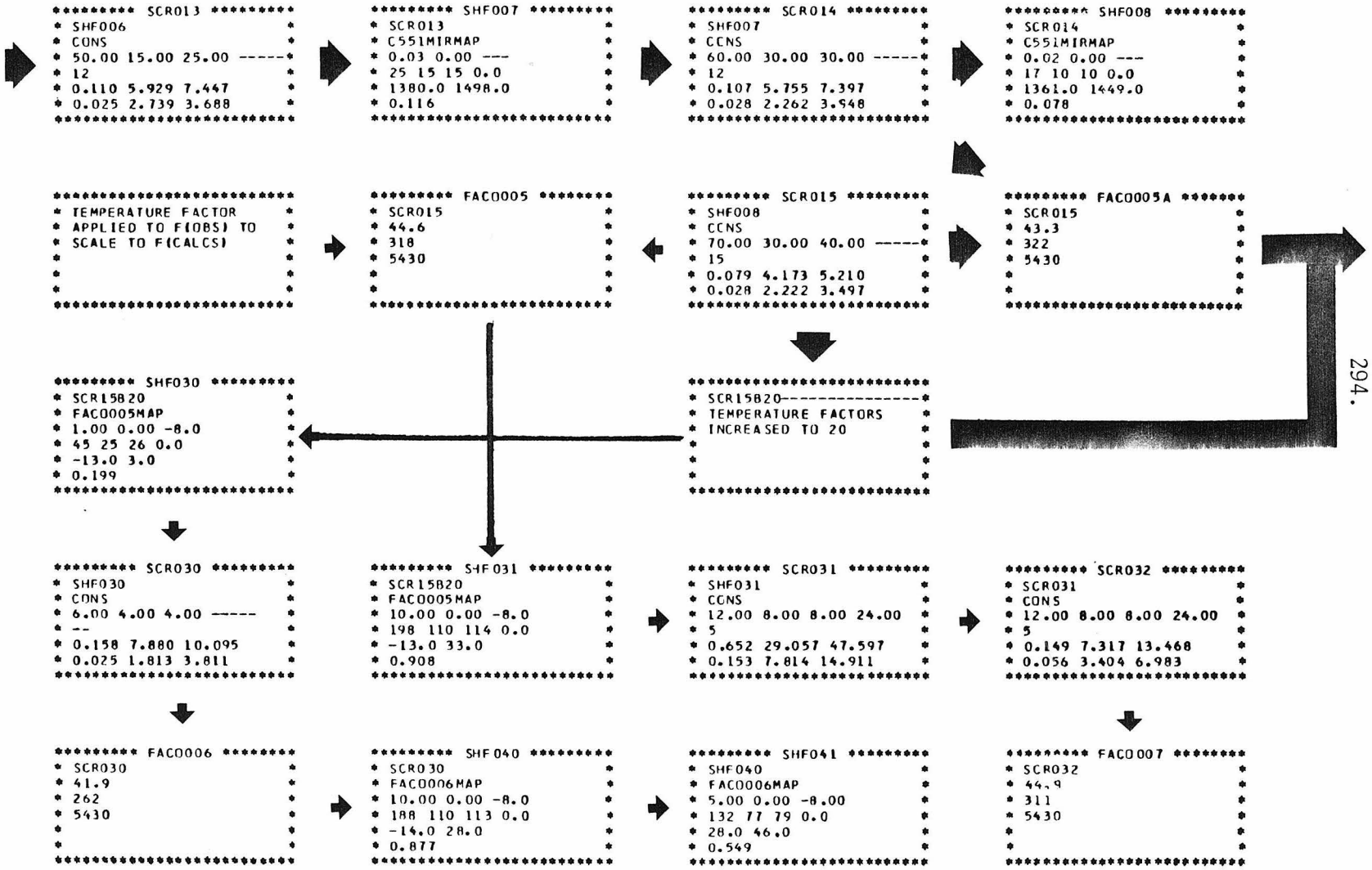
5431 - complete data set to 2.0Å

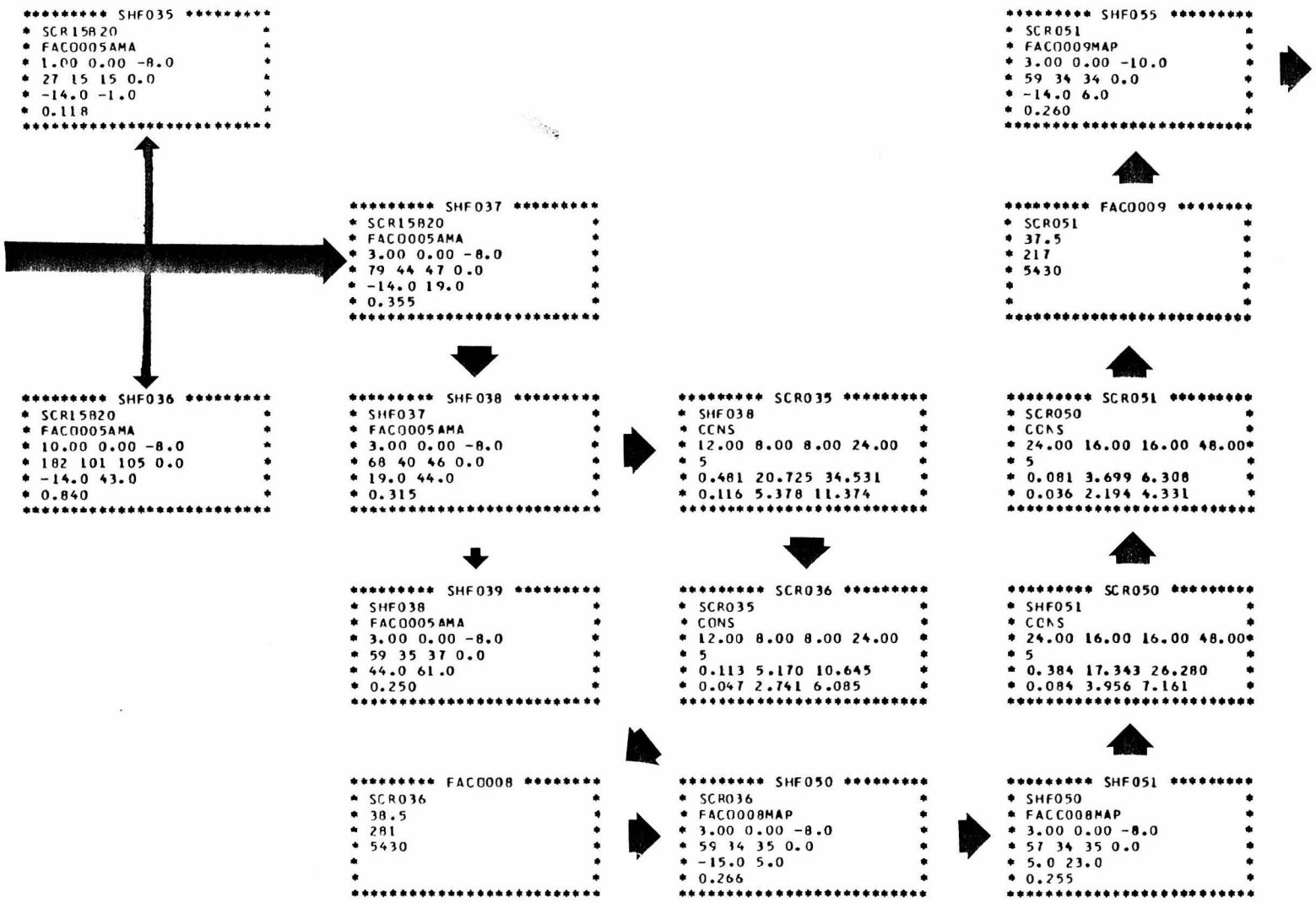
5050 - $|F_{obs}| < 0$ eliminated

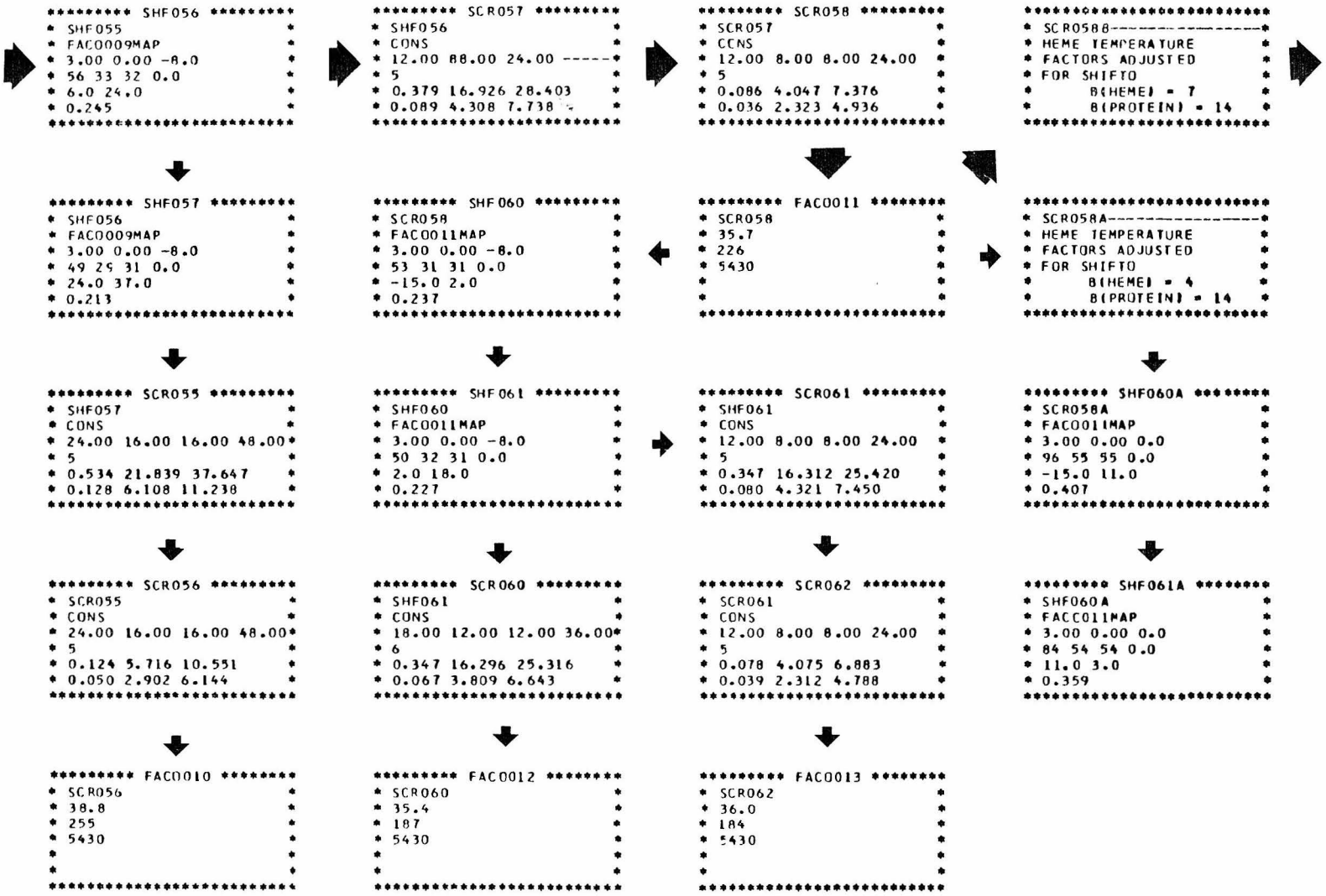
4903 - $|F_{obs}| < 0$ and d < 7Å eliminated

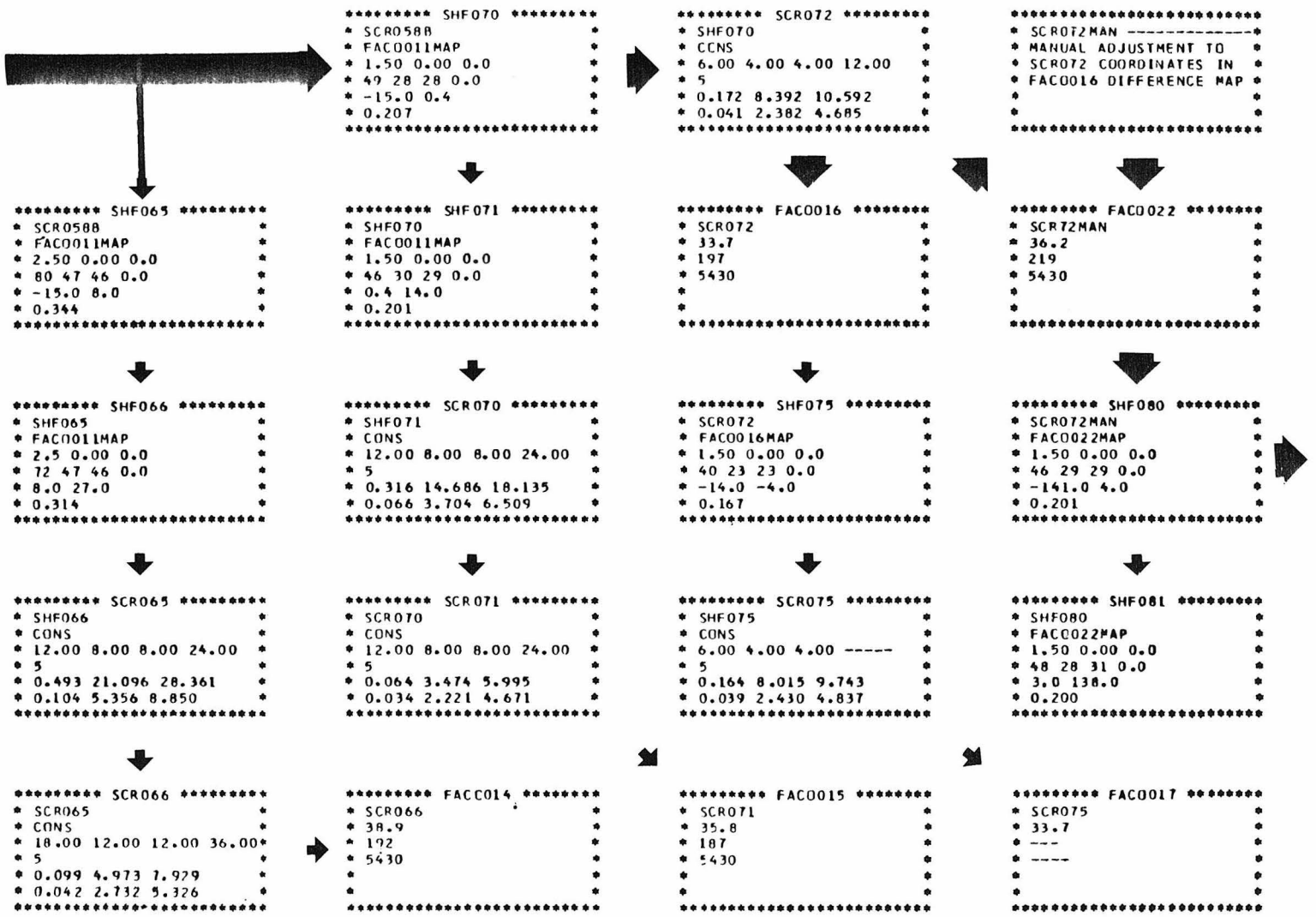
4625 - $|F_{obs}| < 0$ and d < 7Å and $|F_{obs}| < 2\sigma$ eliminated

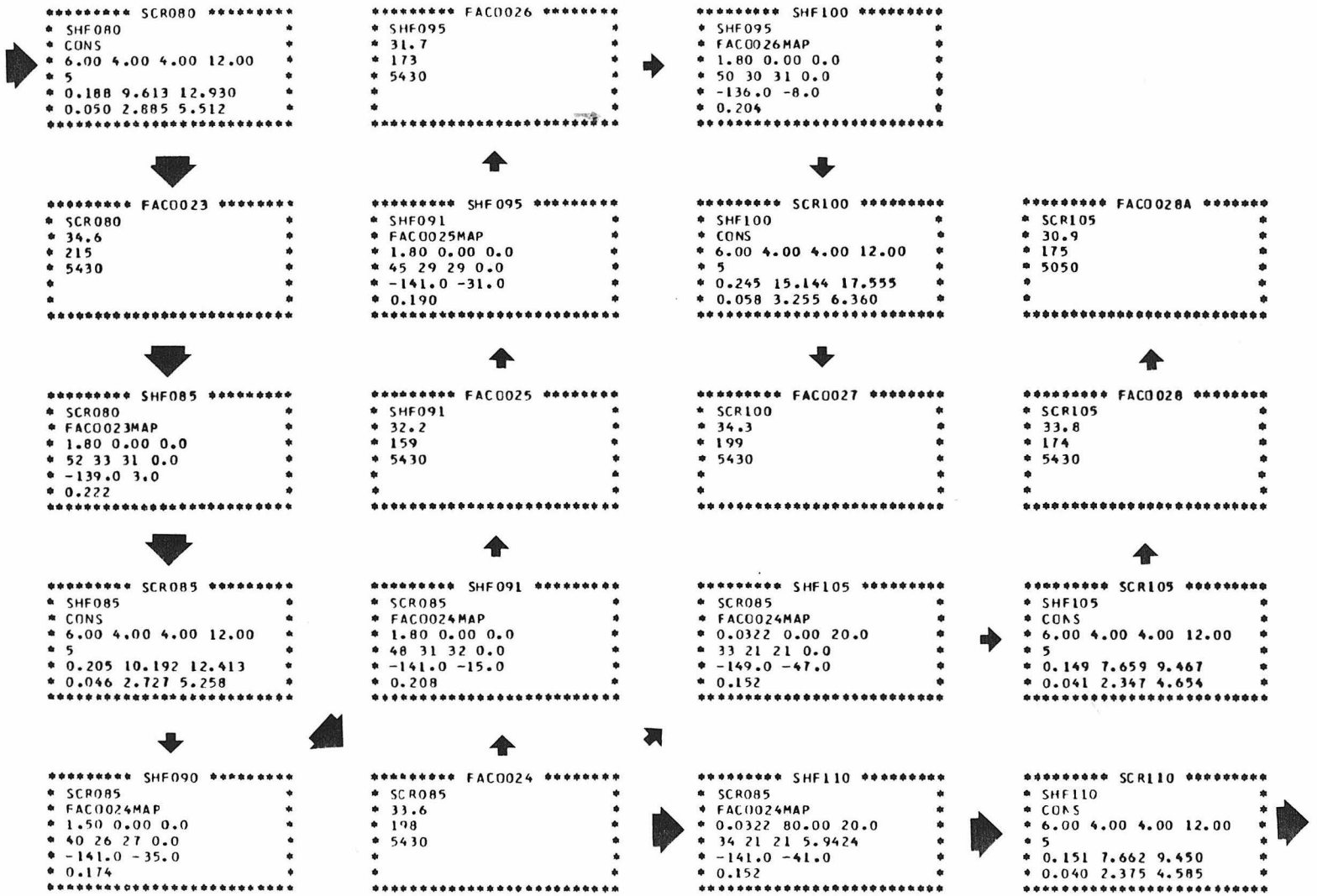

M.I.R. COORDINATES
FROM KENDREW MODEL
BUILT TO 2.4A RESOL.
3-DERIVATIVE M.I.R. MAP
TEMP. FACTORS = 1
SCR001
MIR COORD
CONS
30.00 20.00 20.00 20.00
15
0.053 4.241 4.661
0.024 1.971 3.122
SCR002
SCR001
CONS
30.00 20.00 20.00 20.00
15
0.024 1.938 3.103
0.021 1.719 3.018
SCR003
SCR002
CONS
150.0 100.0 100.0 100.0
5
0.021 1.702 2.994
0.020 1.611 2.899
SCR004
SCR003
CONS
30.00 20.00 20.00 600.0
5
0.020 1.607 2.934
0.020 1.584 2.877
SHF001
SCR004
C551MIRMAP
0.10 0.00 10.00
76 49 52 0.0
1171.0 1543.0
0.395
SCR005
SHF001
CONS
30.00 20.00 20.00 600.0
5
0.262 13.399 21.048
0.090 4.091 7.198
FAC0004
MIR COORD
48.6
448
5430
SCR006
SCR005
CONS
30.00 20.00 20.00 600.0
5
0.084 3.784 6.011
0.046 2.516 4.345
SHF002
SCR006
C551MIRMAP
0.10 0.00 10.0
70 45 48 0.0
1343.0 1619.0
0.356
SCR007
SHF002
CONS
60.00 30.00 30.00 1200.
6
0.288 14.600 23.390
0.067 4.018 7.428
SCR008
SCR007
CONS
60.00 30.00 30.00 120.0
6
0.063 3.638 5.992
0.029 2.273 4.096
SHF003
SCR008
C551MIRMAP
0.10 0.00 10.0
71 45 47 0.0
1368.0 1631.0
0.351
SCR010
SHF003
CONS
60.00 20.00 30.00 -----
12
0.272 14.472 21.375
0.048 4.044 6.227
SHF004
SCR010
C551MIRMAP
0.05 0.00 ---
39 24 24 0.0
1414.0 1574.0
0.180
SCR011
SHF004
CONS
60.00 20.00 30.00 -----
12
0.177 9.416 12.850
0.037 3.216 4.674
SHF005
SCR011
C551MIRMAP
0.03 0.00 ---
23 14 15 0.0
1402.0 1512.0
0.111
SCR012
SHF005
CONS
50.00 15.00 25.00 -----
12
0.114 6.027 7.651
0.031 3.031 4.163
SHF006
SCR012
C551MIRMAP
0.03 0.00 ---
24 15 15 0.0
1391.0 1505.0
0.113

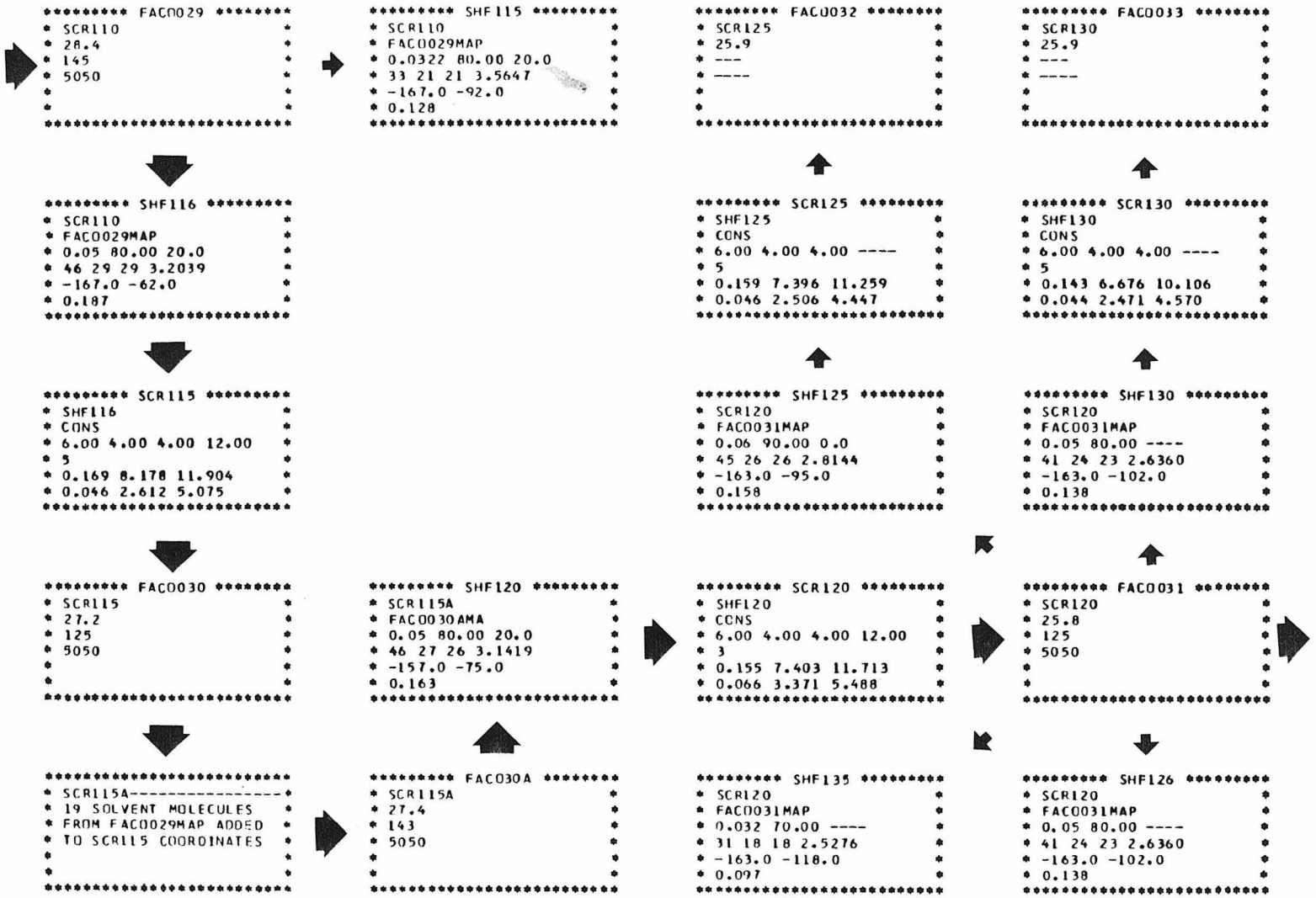

SCR013
SHF006
CONS
50.00 15.00 25.00 -----
12
0.110 5.929 7.447
0.025 2.739 3.688
SHF007
SCR013
C551MIRMAP
0.03 0.00 ---
25 15 15 0.0
1380.0 1498.0
0.116
SCR014
SHF007
CONS
60.00 30.00 30.00 -----
12
0.107 5.755 7.397
0.028 2.262 3.948
SHF008
SCR014
C551MIRMAP
0.02 0.00 ---
17 10 10 0.0
1361.0 1449.0
0.078
TEMPERATURE FACTOR
APPLIED TO F(OBS) TO
SCALE TO F(CALCS)
FAC0005
SCR015
44.6
318
5430
SCR015
SHF008
CONS
70.00 30.00 40.00 -----
15
0.079 4.173 5.210
0.028 2.222 3.497
FAC0005A
SCR015
43.3
322
5430
SHF030
SCR15B20
FAC0005MAP
1.00 0.00 -8.0
45 25 26 0.0
-13.0 3.0
0.199
SCR15B20----------
TEMPERATURE FACTORS
INCREASED TO 20
SCR030
SHF030
CONS
6.00 4.00 4.00 -----
--
0.158 7.880 10.095
0.025 1.813 3.811
SHF031
SCR15B20
FAC0005MAP
10.00 0.00 -8.0
198 110 114 0.0
-13.0 33.0
0.908
SCR031
SHF031
CONS
12.00 8.00 8.00 24.00
5
0.652 29.057 47.597
0.153 7.814 14.911
SCR032
SCR031
CONS
12.00 8.00 8.00 24.00
5
0.149 7.317 13.468
0.056 3.404 6.983
FAC0006
SCR030
41.9
262
5430
SHF040
SCR030
FAC0006MAP
10.00 0.00 -8.0
188 110 113 0.0
-14.0 28.0
0.877
SHF041
SHF040
FAC0006MAP
5.00 0.00 -8.00
132 77 79 0.0
28.0 46.0
0.549
FAC0007
SCR032
44.9
311
5430


SHF035
SCR15B20
FAC0005AMA
1.00 0.00 -8.0
27 15 15 0.0
-14.0 -1.0
0.118
SHF037
SCR15B20
FAC0005AMA
3.00 0.00 -8.0
79 44 47 0.0
-14.0 19.0
0.355
SHF036
SCR15B20
FAC0005AMA
10.00 0.00 -8.0
182 101 105 0.0
-14.0 43.0
0.840
SHF038
SHF037
FAC0005AMA
3.00 0.00 -8.0
68 40 46 0.0
19.0 44.0
0.315
SCR035
SHF038
CONS
12.00 8.00 8.00 24.00
5
0.481 20.725 34.531
0.116 5.378 11.374
SHF055
SCR051
FAC0009MAP
3.00 0.00 -10.0
59 34 34 0.0
-14.0 6.0
0.260
FAC0009
SCR051
37.5
217
5430
SCR051
SCR050
CONS
24.00 16.00 16.00 48.00
5
0.081 3.699 6.308
0.036 2.194 4.331
SHF039
SHF038
FAC0005AMA
3.00 0.00 -8.0
59 35 37 0.0
44.0 61.0
0.250
SCR036
SCR035
CONS
12.00 8.00 8.00 24.00
5
0.113 5.170 10.645
0.047 2.741 6.085
SCR050
SHF051
CONS
24.00 16.00 16.00 48.00
5
0.384 17.343 26.280
0.084 3.956 7.161
FAC0008
SCR036
38.5
281
5430
SHF050
SCR036
FAC0008MAP
3.00 0.00 -8.0
59 34 35 0.0
-15.0 5.0
0.266
SHF051
SHF050
FAC0008MAP
3.00 0.00 -8.0
57 34 35 0.0
5.0 23.0
0.255


SHF056
SHF055
FAC0009MAP
3.00 0.00 -8.0
56 33 32 0.0
6.0 24.0
0.245
SCR057
SHF056
CONS
12.00 88.00 24.00 -----
5
0.379 16.926 28.403
0.089 4.308 7.738
SCR058
SCR057
CONS
12.00 8.00 8.00 24.00
5
0.086 4.047 7.376
0.036 2.323 4.936
SCR058B----------
HEME TEMPERATURE
FACTORS ADJUSTED
FOR SHIFTO
B(HEME) = 7
B(PROTEIN) = 14
SHF057
SHF056
FAC0009MAP
3.00 0.00 -8.0
49 29 31 0.0
24.0 37.0
0.213
SHF060
SCR058
FAC0011MAP
3.00 0.00 -8.0
53 31 31 0.0
-15.0 2.0
0.237
FAC0011
SCR058
35.7
226
5430
SCR058A----------
HEME TEMPERATURE
FACTORS ADJUSTED
FOR SHIFTO
B(HEME) = 4
B(PROTEIN) = 14
SCR055
SHF057
CONS
24.00 16.00 16.00 48.00
5
0.534 21.839 37.647
0.128 6.108 11.238
SHF061
SHF060
FAC0011MAP
3.00 0.00 -8.0
50 32 31 0.0
2.0 18.0
0.227
SCR061
SHF061
CONS
12.00 8.00 8.00 24.00
5
0.347 16.312 25.420
0.080 4.321 7.450
SHF060A
SCR058A
FAC0011MAP
3.00 0.00 0.0
96 55 55 0.0
-15.0 11.0
0.407
SCR056
SCR055
CONS
24.00 16.00 16.00 48.00
5
0.124 5.716 10.551
0.050 2.902 6.144
SCR060
SHF061
CONS
18.00 12.00 12.00 36.00
6
0.347 16.296 25.316
0.067 3.809 6.643
SCR062
SCR061
CONS
12.00 8.00 8.00 24.00
5
0.078 4.075 6.883
0.039 2.312 4.788
SHF061A
SHF060A
FAC0011MAP
3.00 0.00 0.0
84 54 54 0.0
11.0 3.0
0.359
FAC0010
SCR056
38.8
255
5430
FAC0012
SCR060
35.4
187
5430
FAC0013
SCR062
36.0
184
5430

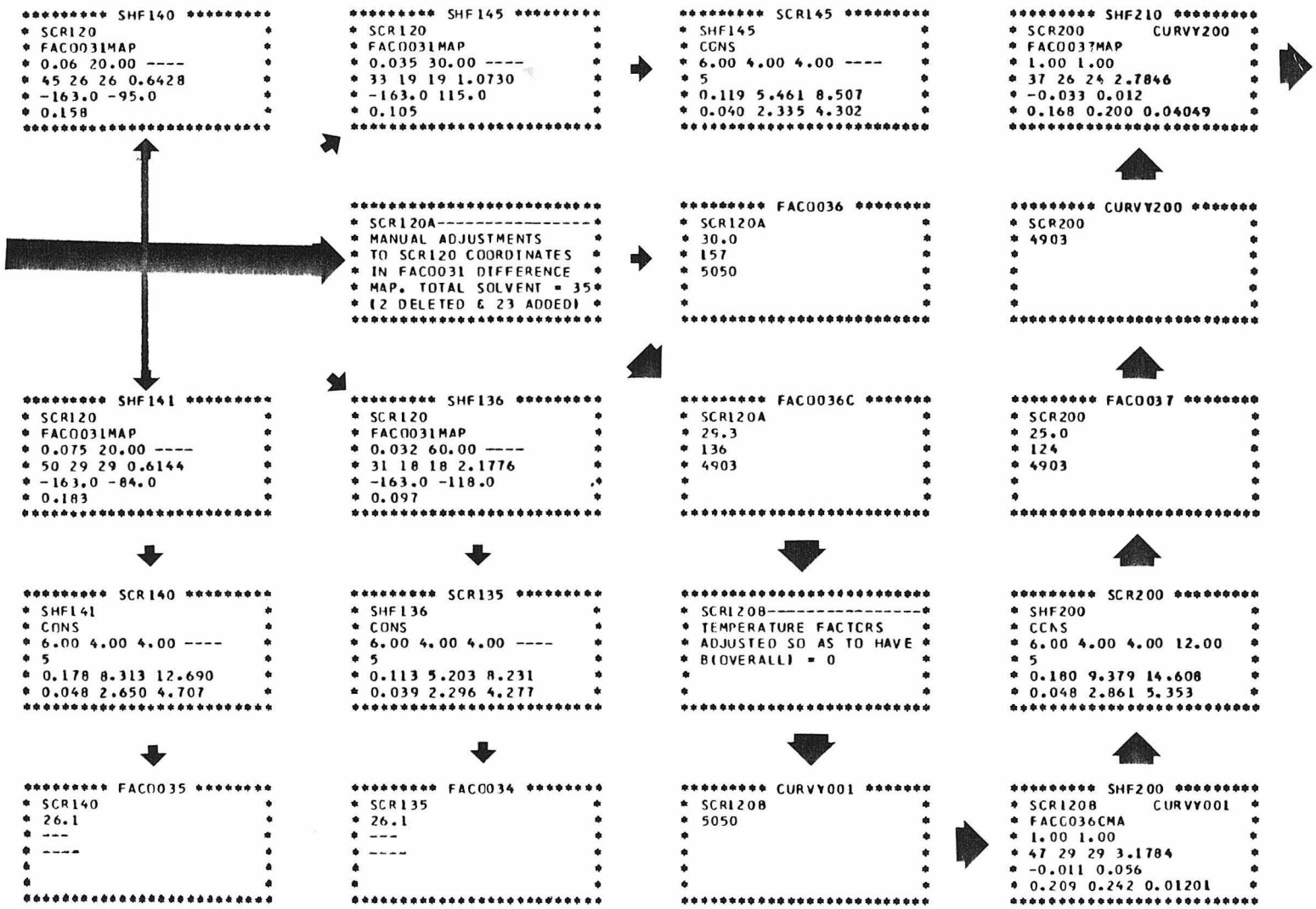

SHF065
SCR058B
FAC0011MAP
2.50 0.00 0.0
80 47 46 0.0
-15.0 8.0
0.344
SHF066
SHF065
FAC0011MAP
2.5 0.00 0.0
72 47 46 0.0
8.0 27.0
0.314
SCR065
SHF066
CONS
12.00 8.00 8.00 24.00
5
0.493 21.096 28.361
0.104 5.356 8.850
SCR066
SCR065
CONS
18.00 12.00 12.00 36.00
5
0.099 4.973 7.929
0.042 2.732 5.326
SHF070
SCR058B
FAC0011MAP
1.50 0.00 0.0
49 28 28 0.0
-15.0 0.4
0.207
SHF071
SHF070
FAC0011MAP
1.50 0.00 0.0
46 30 29 0.0
0.4 14.0
0.201
SCR070
SHF071
CONS
12.00 8.00 8.00 24.00
5
0.316 14.686 18.135
0.066 3.704 6.509
SCR071
SCR070
CONS
12.00 8.00 8.00 24.00
5
0.064 3.474 5.995
0.034 2.221 4.671
FACC014
SCR066
38.9
192
5430
SCR072
SHF070
CONS
6.00 4.00 4.00 12.00
5
0.172 8.392 10.592
0.041 2.382 4.685
FAC0016
SCR072
33.7
197
5430
SHF075
SCR072
FAC0016MAP
1.50 0.00 0.0
40 23 23 0.0
-14.0 -4.0
0.167
SCR075
SHF075
CONS
6.00 4.00 4.00 -----
5
0.164 8.015 9.743
0.039 2.430 4.837
FAC0015
SCR071
35.8
187
5430
SCR72MAN ----------
MANUAL ADJUSTMENT TO
SCR072 COORDINATES IN
FAC0016 DIFFERENCE MAP
FAC0022
SCR72MAN
36.2
219
5430
SHF080
SCR72MAN
FAC0022MAP
1.50 0.00 0.0
46 29 29 0.0
-141.0 4.0
0.201
SHF081
SHF080
FAC0022MAP
1.50 0.00 0.0
48 28 31 0.0
3.0 138.0
0.200
FAC0017
SCR075
33.7


SCR080
SHF080
CONS
6.00 4.00 4.00 12.00
5
0.188 9.613 12.930
0.050 2.885 5.512
FAC0026
SHF095
31.7
173
5430
SHF100
SHF095
FAC0026MAP
1.80 0.00 0.0
50 30 31 0.0
-136.0 -8.0
0.204
FAC0023
SCR080
34.6
215
5430
SHF095
SHF091
FAC0025MAP
1.80 0.00 0.0
45 29 29 0.0
-141.0 -31.0
0.190
SCR100
SHF100
CONS
6.00 4.00 4.00 12.00
5
0.245 15.144 17.555
0.058 3.255 6.360
FAC028A
SCR105
30.9
175
5050
SHF085
SCR080
FAC0023MAP
1.80 0.00 0.0
52 33 31 0.0
-139.0 3.0
0.222
FAC0025
SHF091
32.2
159
5430
FAC0027
SCR100
34.3
199
5430
FAC028
SCR105
33.8
174
5430
SCR085
SHF085
CONS
6.00 4.00 4.00 12.00
5
0.205 10.192 12.413
0.046 2.727 5.258
SHF091
SCR085
FAC0024MAP
1.80 0.00 0.0
48 31 32 0.0
-141.0 -15.0
0.208
SHF105
SCR085
FAC0024MAP
0.0322 0.00 20.0
33 21 21 0.0
-149.0 -47.0
0.152
SCR105
SHF105
CONS
6.00 4.00 4.00 12.00
5
0.149 7.659 9.467
0.041 2.347 4.654
SHF090
SCR085
FAC0024MAP
1.50 0.00 0.0
40 26 27 0.0
-141.0 -35.0
0.174
FAC0024
SCR085
33.6
198
5430
SHF110
SCR085
FAC0024MAP
0.0322 80.00 20.0
34 21 21 5.9424
-141.0 -41.0
0.152
SCR110
SHF110
CONS
6.00 4.00 4.00 12.00
5
0.151 7.662 9.450
0.040 2.375 4.585


FAC0029
SCR110
28.4
145
5050
SHF115
SCR110
FAC0029MAP
0.0322 80.00 20.0
33 21 21 3.5647
-167.0 -92.0
0.128
FAC0032
SCR125
25.9

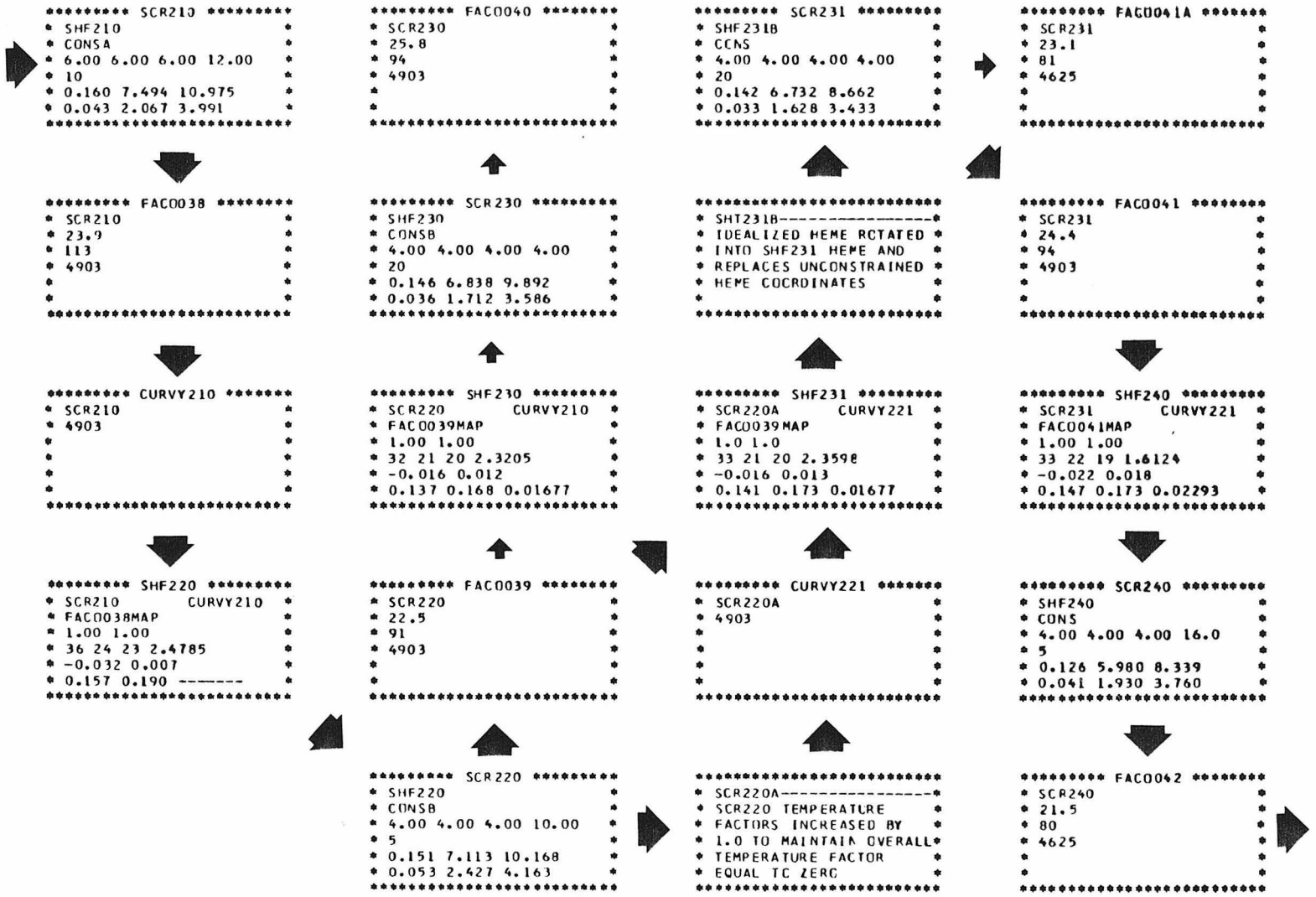
FAC0033
SCR130
25.9

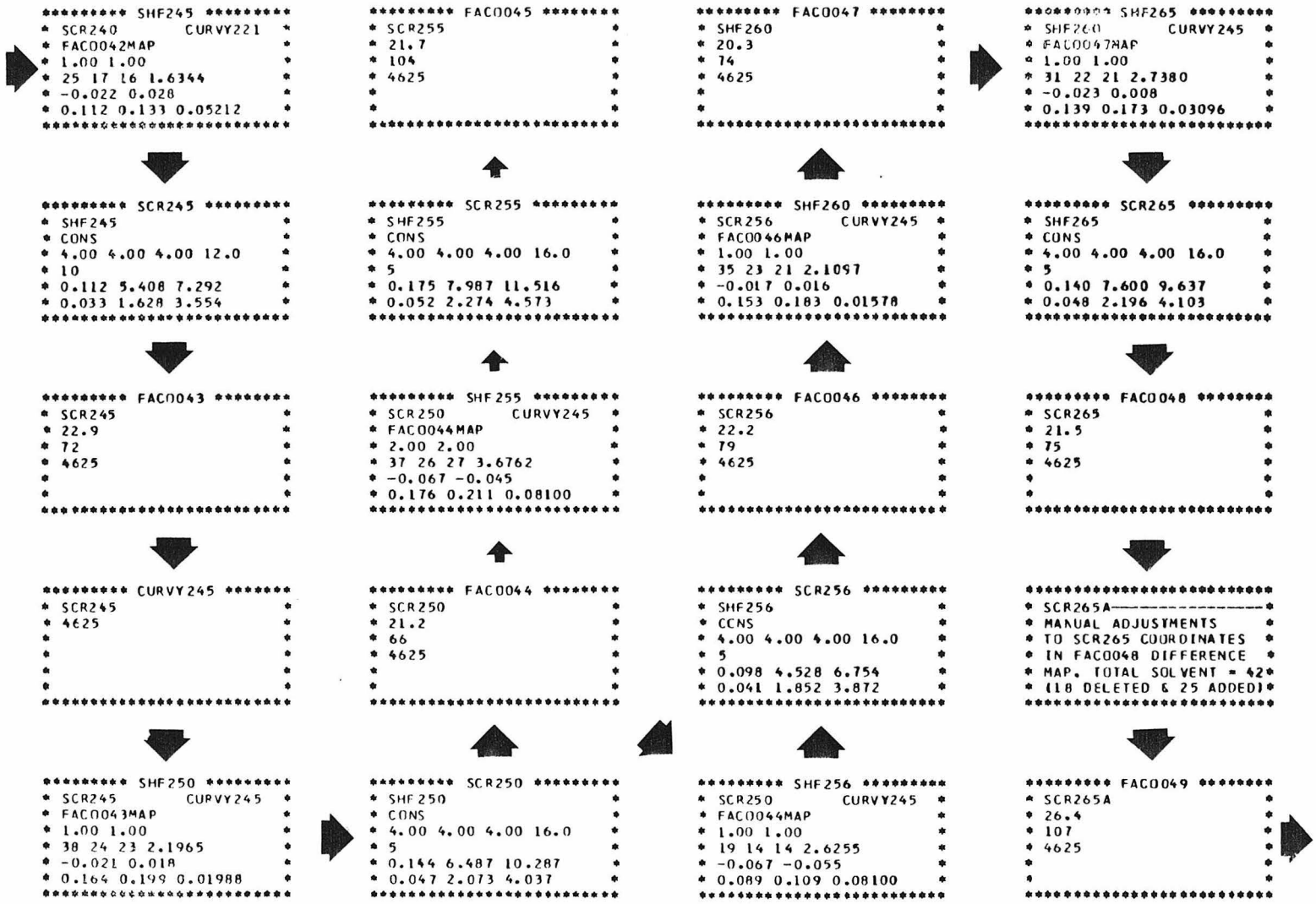
SHF116
SCR110
FAC0029MAP
0.05 80.00 20.0
46 29 29 3.2039
-167.0 -62.0
0.187
SCR125
SHF125
CONS
6.00 4.00 4.00 ----
5
0.159 7.396 11.259
0.046 2.506 4.447
SCR130
SHF130
CONS
6.00 4.00 4.00 ----
5
0.143 6.676 10.106
0.044 2.471 4.570
SCR115
SHF116
CONS
6.00 4.00 4.00 12.00
5
0.169 8.178 11.904
0.046 2.612 5.075
SHF125
SCR120
FAC0031MAP
0.06 90.00 0.0
45 26 26 2.8144
-163.0 -95.0
0.158
SHF130
SCR120
FAC0031MAP
0.05 80.00 ----
41 24 23 2.6360
-163.0 -102.0
0.138
FAC0030
SCR115
27.2
125
5050
SHF120
SCR115A
FAC0030AMA
0.05 80.00 20.0
46 27 26 3.1419
-157.0 -75.0
0.163
SCR120
SHF120
CONS
6.00 4.00 4.00 12.00
3
0.155 7.403 11.713
0.066 3.371 5.488
FAC0031
SCR120
25.8
125
5050
SCR115A
19 SOLVENT MOLECULES
FROM FAC0029MAP ADDED
TO SCR115 COORDINATES
FAC030A
SCR115A
27.4
143
5050
SHF135
SCR120
FAC0031MAP
0.032 70.00 ----
31 18 18 2.5276
-163.0 -118.0
0.097
SHF126
SCR120
FAC0031MAP
0.05 80.00 ----
41 24 23 2.6360
-163.0 -102.0
0.138

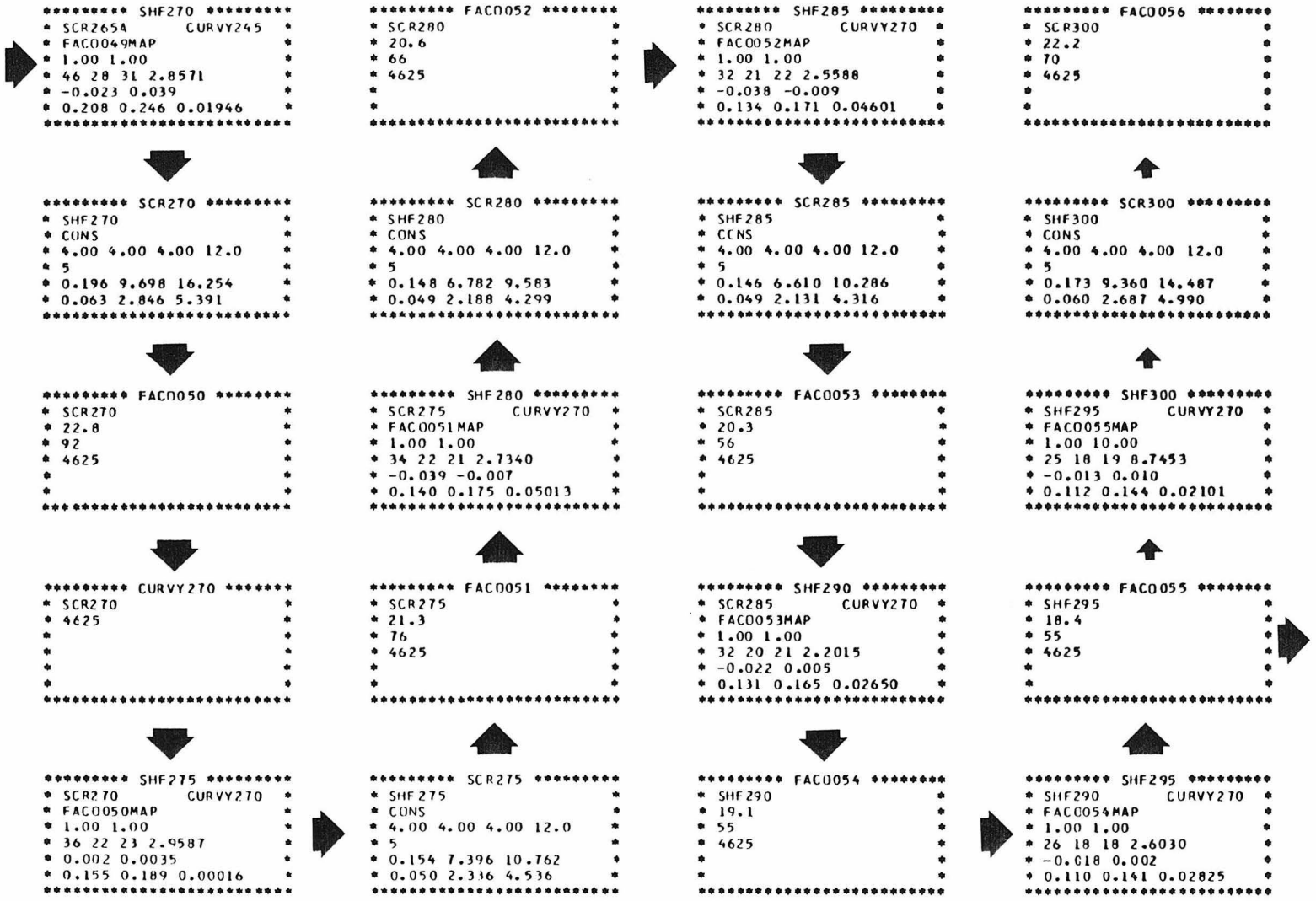

SHF140
SCR120
FAC0031MAP
0.06 20.00 ----
45 26 26 0.6428
-163.0 -95.0
0.158
SHF145
SCR120
FAC0031MAP
0.035 30.00 ----
33 19 19 1.0730
-163.0 115.0
0.105
SCR145
SHF145
CONS
6.00 4.00 4.00 ----
5
0.119 5.461 8.507
0.040 2.335 4.302
SHF210
SCR200 CURVY200
FAC0037MAP
1.00 1.00
37 26 24 2.7846
-0.033 0.012
0.168 0.200 0.04049
SCR120A
MANUAL ADJUSTMENTS
TO SCR120 COORDINATES
IN FAC0031 DIFFERENCE
MAP. TOTAL SOLVENT = 35
(2 DELETED & 23 ADDED)
FAC0036
SCR120A
30.0
157
5050
CURVY200
SCR200
4903
SHF141
SCR120
FAC0031MAP
0.075 20.00 ----
50 29 29 0.6144
-163.0 -84.0
0.183
SHF136
SCR120
FAC0031MAP
0.032 60.00 ----
31 18 18 2.1776
-163.0 -118.0
0.097
FAC0036C
SCR120A
29.3
136
4903
FAC0037
SCR200
25.0
124
4903
SCR140
SHF141
CONS
6.00 4.00 4.00 ----
5
0.178 8.313 12.690
0.048 2.650 4.707
SCR135
SHF136
CONS
6.00 4.00 4.00 ----
5
0.113 5.203 8.231
0.039 2.296 4.277
SCR120B
TEMPERATURE FACTORS
ADJUSTED SO AS TO HAVE
B(OVERALL) = 0
SCR200
SHF200
CONS
6.00 4.00 4.00 12.00
5
0.180 9.379 14.608
0.048 2.861 5.353
FAC0035
SCR140
26.1

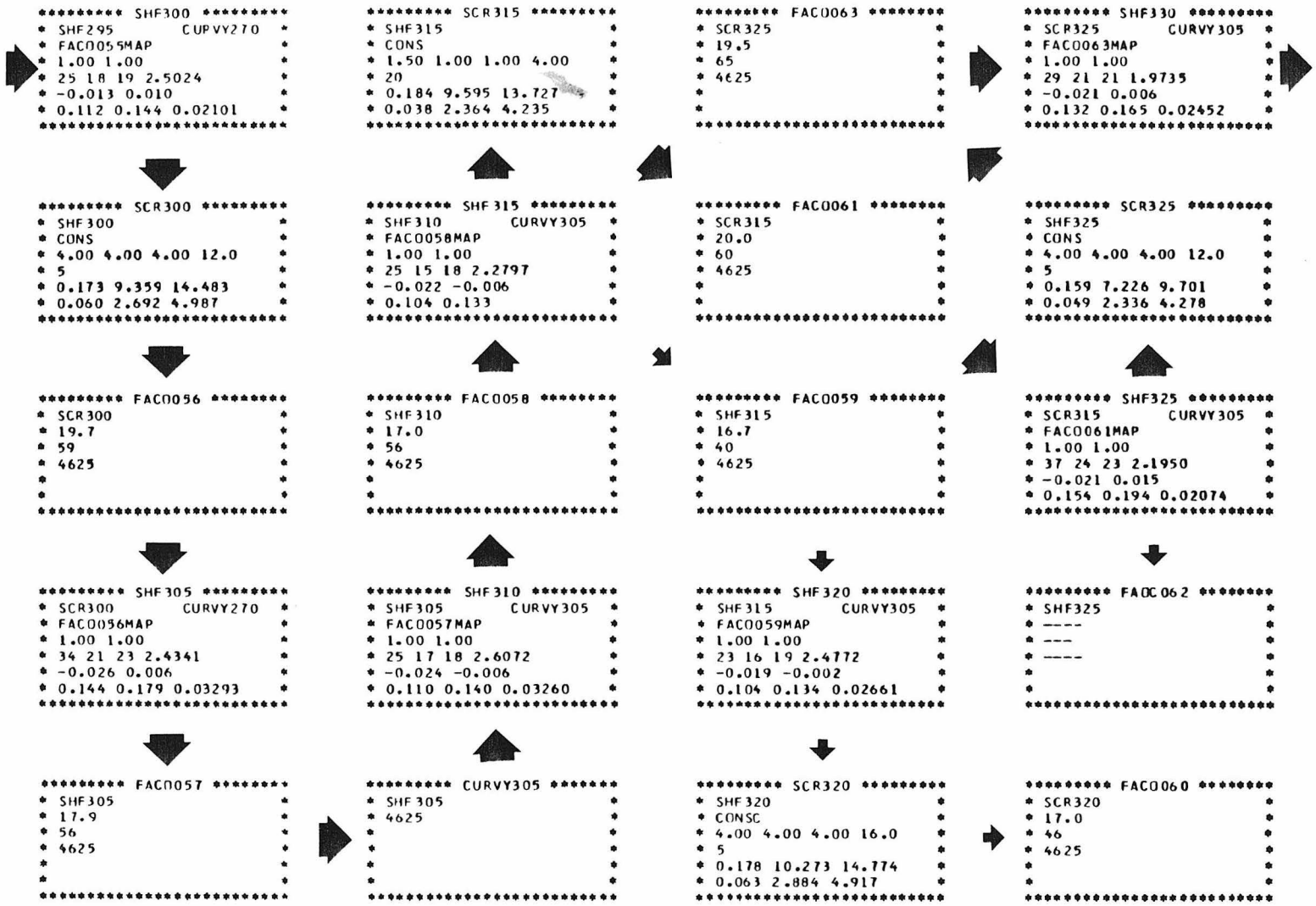
FAC0034
SCR135
26.1

CURVY001
SCR120B
5050
SHF200
SCR120B CURVY001
FAC0036CMA
1.00 1.00
47 29 29 3.1784
-0.011 0.056
0.209 0.242 0.01201

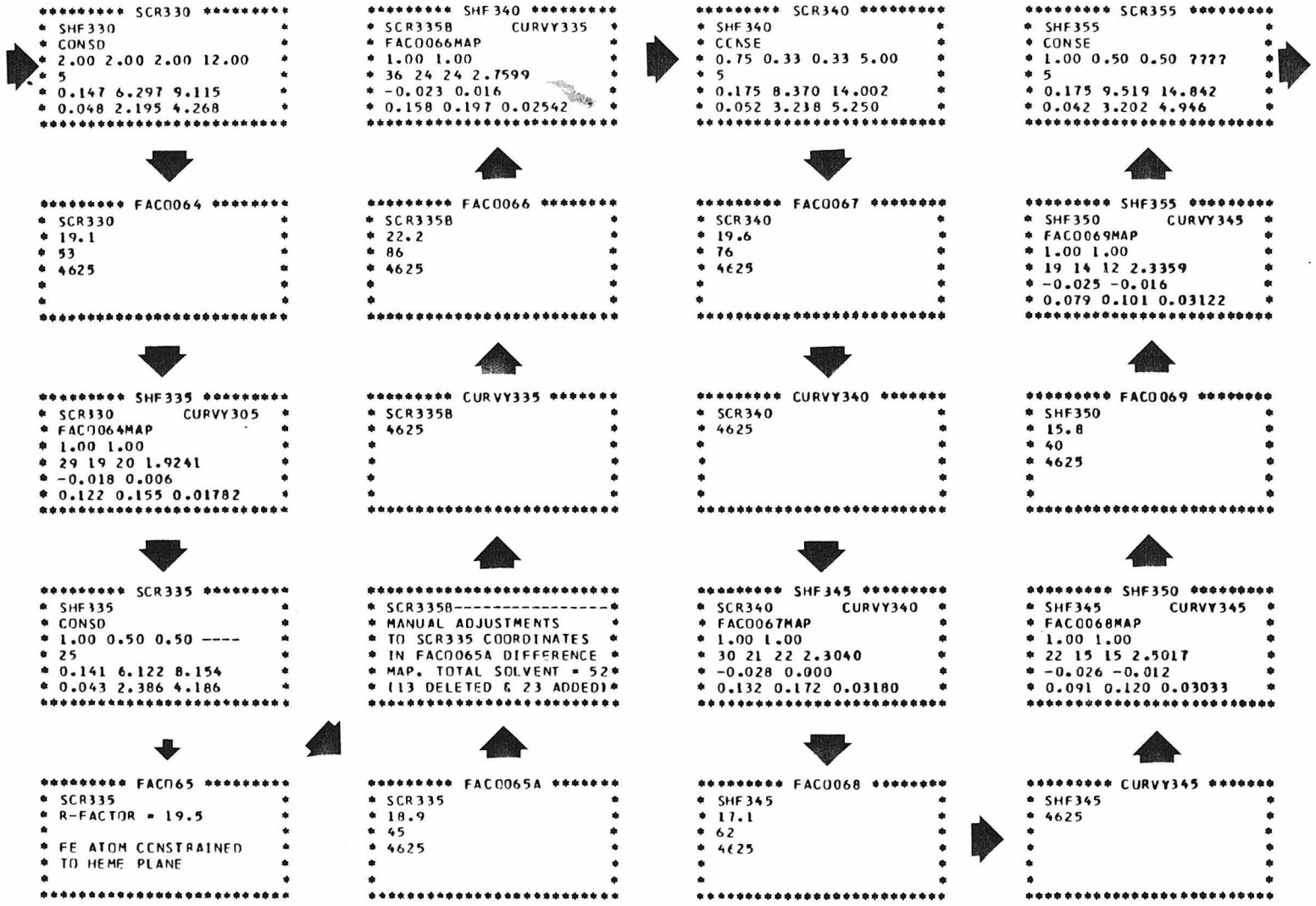

********** SCR210 **********
* SHF210 *
* CONSA *
* 6.00 6.00 6.00 12.00 *
* 10 *
* 0.160 7.494 10.975 *
* 0.043 2.067 3.991 *
****
********** FAC0038 **********
* SCR210 *
* 23.9 *
* 113 *
* 4903 *
****
********** CURVY210 **********
* SCR210 *
* 4903 *
****
********** SHF220 **********
* SCR210 CURVY210 *
* FAC0038MAP *
* 1.00 1.00 *
* 36 24 23 2.4785 *
* -0.032 0.007 *
* 0.157 0.190 ------- *
****
********** SCR220 **********
* SHF220 *
* CONSB *
* 4.00 4.00 4.00 10.00 *
* 5 *
* 0.151 7.113 10.168 *
* 0.053 2.427 4.163 *
****
********** FAC0039 **********
* SCR220 *
* 22.5 *
* 91 *
* 4903 *
****
********** SHF230 **********
* SCR220 CURVY210 *
* FAC0039MAP *
* 1.00 1.00 *
* 32 21 20 2.3205 *
* -0.016 0.012 *
* 0.137 0.168 0.01677 *
****
********** SCR230 **********
* SHF230 *
* CONSB *
* 4.00 4.00 4.00 4.00 *
* 20 *
* 0.146 6.838 9.892 *
* 0.036 1.712 3.586 *
****
********** FAC0040 **********
* SCR230 *
* 25.8 *
* 94 *
* 4903 *
****
********** SCR220A **********
* SCR220 TEMPERATURE *
* FACTORS INCREASED BY *
* 1.0 TO MAINTAIN OVERALL *
* TEMPERATURE FACTOR *
* EQUAL TO ZERO *
****
********** CURVY221 **********
* SCR220A *
* 4903 *
****
********** SHF231 **********
* SCR220A CURVY221 *
* FAC0039MAP *
* 1.0 1.0 *
* 33 21 20 2.3598 *
* -0.016 0.013 *
* 0.141 0.173 0.01677 *
****
********** SCR231 **********
* SHF231B *
* CONS *
* 4.00 4.00 4.00 4.00 *
* 20 *
* 0.142 6.732 8.662 *
* 0.033 1.628 3.433 *
****
********** SHT231B----------*
* IDEALIZED HEME ROTATED *
* INTO SHF231 HEME AND *
* REPLACES UNCONSTRAINED *
* HEME COORDINATES *
****
********** FAC0041 **********
* SCR231 *
* 24.4 *
* 94 *
* 4903 *
****
********** SHF240 **********
* SCR231 CURVY221 *
* FAC0041MAP *
* 1.00 1.00 *
* 33 22 19 1.6124 *
* -0.022 0.018 *
* 0.147 0.173 0.02293 *
****
********** SCR240 **********
* SHF240 *
* CONS *
* 4.00 4.00 4.00 16.0 *
* 5 *
* 0.126 5.980 8.339 *
* 0.041 1.930 3.760 *
****
********** FAC0042 **********
* SCR240 *
* 21.5 *
* 80 *
* 4625 *
****
********** FAC0041A **********
* SCR231 *
* 23.1 *
* 81 *
* 4625 *
****

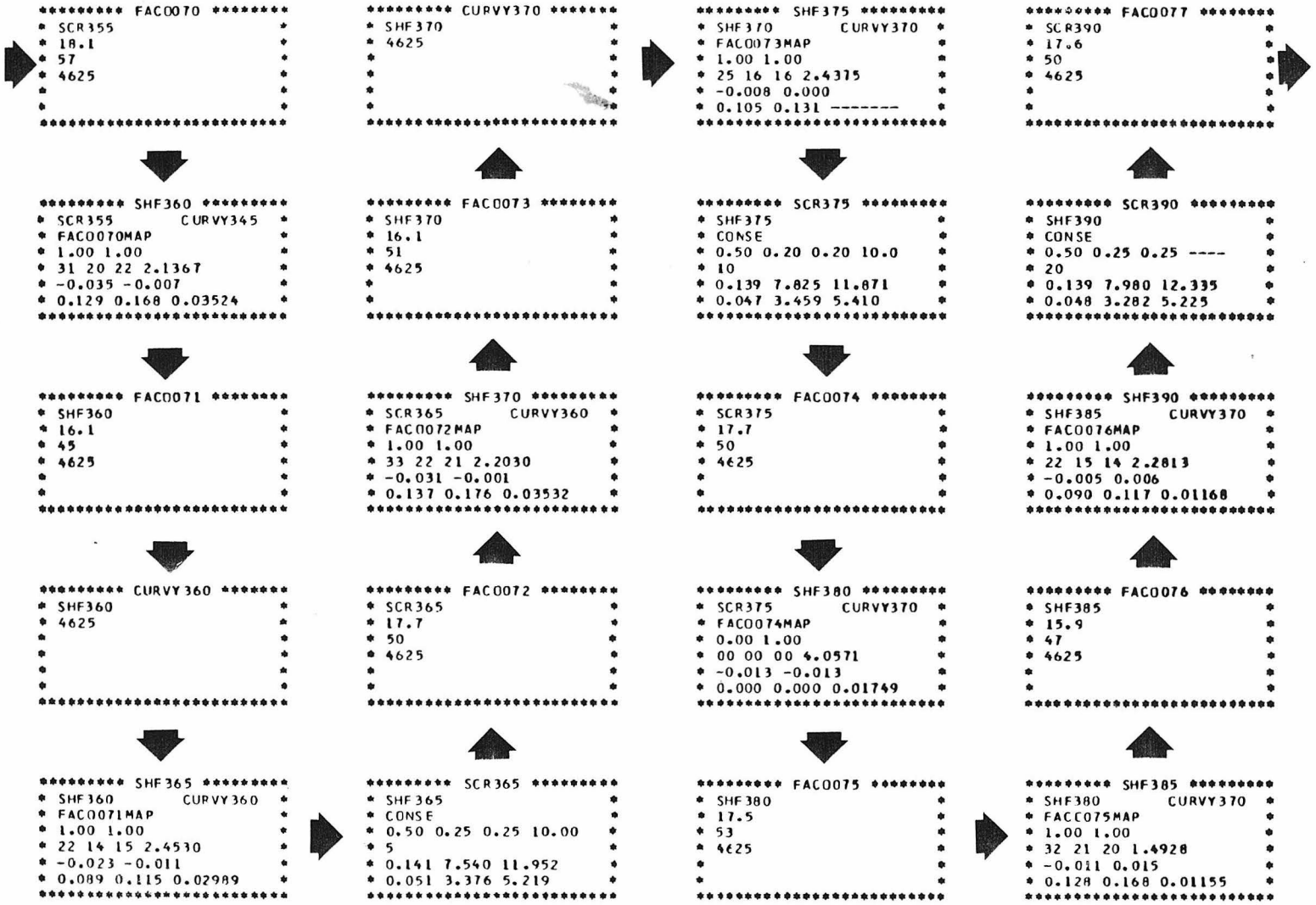

SHF245
SCR240 CURVY221
FAC0042MAP
1.00 1.00
25 17 16 1.6344
-0.022 0.028
0.112 0.133 0.05212
SCR245
SHF245
CONS
4.00 4.00 4.00 12.0
10
0.112 5.408 7.292
0.033 1.628 3.554
FAC0043
SCR245
22.9
72
4625
CURVY245
SCR245
4625
SHF250
SCR245 CURVY245
FAC0043MAP
1.00 1.00
38 24 23 2.1965
-0.021 0.018
0.164 0.199 0.01988
FAC0045
SCR255
21.7
104
4625
SCR255
SHF255
CONS
4.00 4.00 4.00 16.0
5
0.175 7.987 11.516
0.052 2.274 4.573
SHF255
SCR250 CURVY245
FAC0044MAP
2.00 2.00
37 26 27 3.6762
-0.067 -0.045
0.176 0.211 0.08100
FAC0044
SCR250
21.2
66
4625
SCR250
SHF250
CONS
4.00 4.00 4.00 16.0
5
0.144 6.487 10.287
0.047 2.073 4.037
FAC0047
SHF260
20.3
74
4625
SHF260
SCR256 CURVY245
FAC0046MAP
1.00 1.00
35 23 21 2.1097
-0.017 0.016
0.153 0.183 0.01578
FAC0046
SCR256
22.2
79
4625
SCR256
SHF256
CONS
4.00 4.00 4.00 16.0
5
0.098 4.528 6.754
0.041 1.852 3.872
SHF256
SCR250 CURVY245
FAC0044MAP
1.00 1.00
19 14 14 2.6255
-0.067 -0.055
0.089 0.109 0.08100
SHF265
SHF260 CURVY245
FAC0047MAP
1.00 1.00
31 22 21 2.7380
-0.023 0.008
0.139 0.173 0.03096
SCR265
SHF265
CONS
4.00 4.00 4.00 16.0
5
0.140 7.600 9.637
0.048 2.196 4.103
FAC0048
SCR265
21.5
75
4625
SCR265A
MANUAL ADJUSTMENTS
TO SCR265 COORDINATES
IN FAC0048 DIFFERENCE
MAP. TOTAL SOLVENT = 42
(18 DELETED & 25 ADDED)
FAC0049
SCR265A
26.4
107
4625

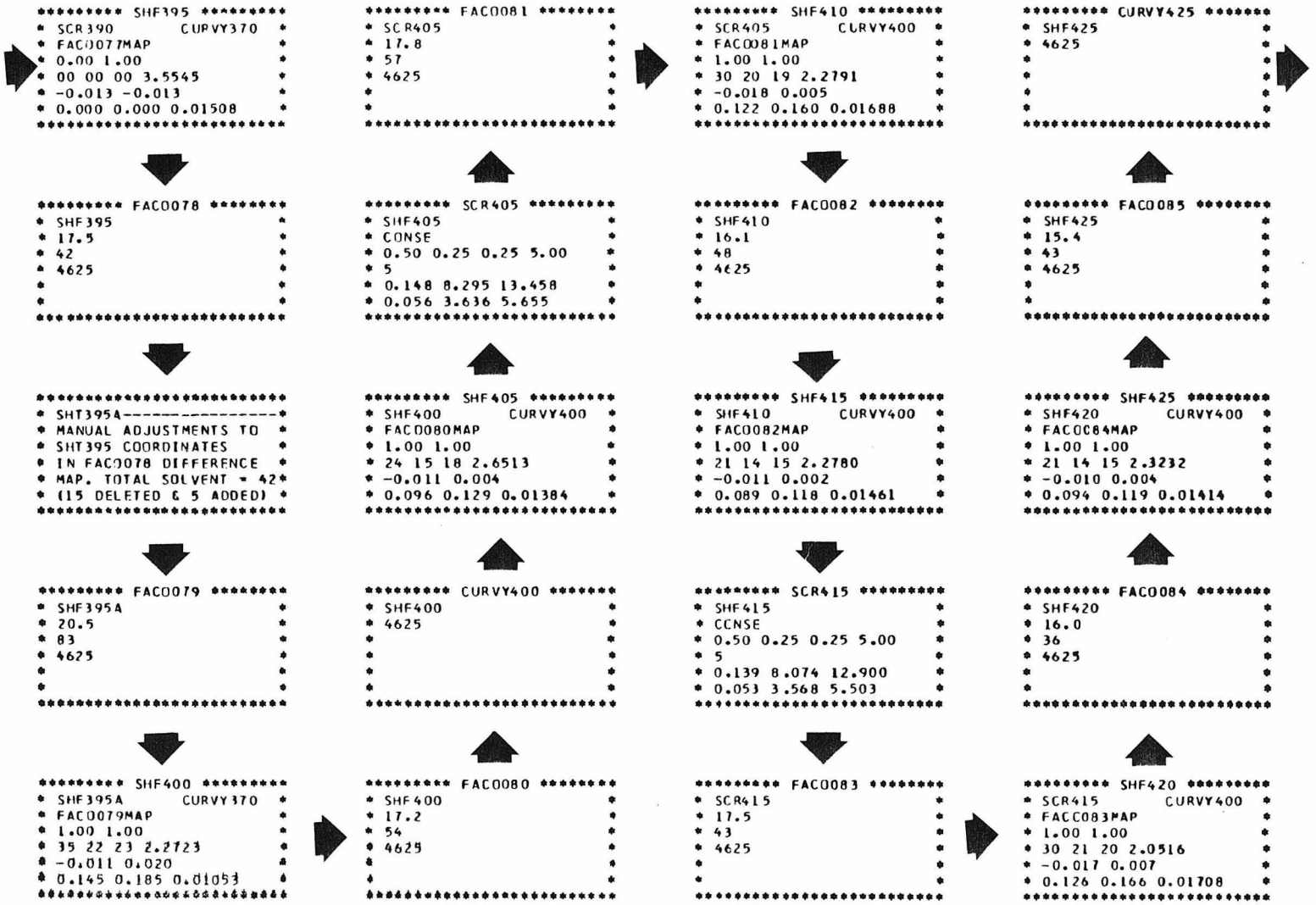

SHF270
SCR265A CURVY245
FAC0049MAP
1.00 1.00
46 28 31 2.8571
-0.023 0.039
0.208 0.246 0.01946
SCR270
SHF270
CONS
4.00 4.00 4.00 12.0
5
0.196 9.698 16.254
0.063 2.846 5.391
FAC0050
SCR270
22.8
92
4625
CURVY270
SCR270
4625
SHF275
SCR270 CURVY270
FAC0050MAP
1.00 1.00
36 22 23 2.9587
0.002 0.0035
0.155 0.189 0.00016
SCR275
SHF275
CONS
4.00 4.00 4.00 12.0
5
0.154 7.396 10.762
0.050 2.336 4.536
FAC0051
SCR275
21.3
76
4625
SHF280
SCR275 CURVY270
FAC0051MAP
1.00 1.00
34 22 21 2.7340
-0.039 -0.007
0.140 0.175 0.05013
SCR280
SHF280
CONS
4.00 4.00 4.00 12.0
5
0.148 6.782 9.583
0.049 2.188 4.299
FAC0052
SCR280
20.6
66
4625
SHF285
SCR280 CURVY270
FAC0052MAP
1.00 1.00
32 21 22 2.5588
-0.038 -0.009
0.134 0.171 0.04601
SCR285
SHF285
CONS
4.00 4.00 4.00 12.0
5
0.146 6.610 10.286
0.049 2.131 4.316
FAC0053
SCR285
20.3
56
4625
SHF290
SCR285 CURVY270
FAC0053MAP
1.00 1.00
32 20 21 2.2015
-0.022 0.005
0.131 0.165 0.02650
FAC0054
SHF290
19.1
55
4625
SHF295
SHF290 CURVY270
FAC0054MAP
1.00 1.00
26 18 18 2.6030
-0.018 0.002
0.110 0.141 0.02825
FAC0055
SHF295
18.4
55
4625
SHF300
SHF295 CURVY270
FAC0055MAP
1.00 10.00
25 18 19 8.7453
-0.013 0.010
0.112 0.144 0.02101
SCR300
SHF300
CONS
4.00 4.00 4.00 12.0
5
0.173 9.360 14.487
0.060 2.687 4.990
FAC0056
SCR300
22.2
70
4625

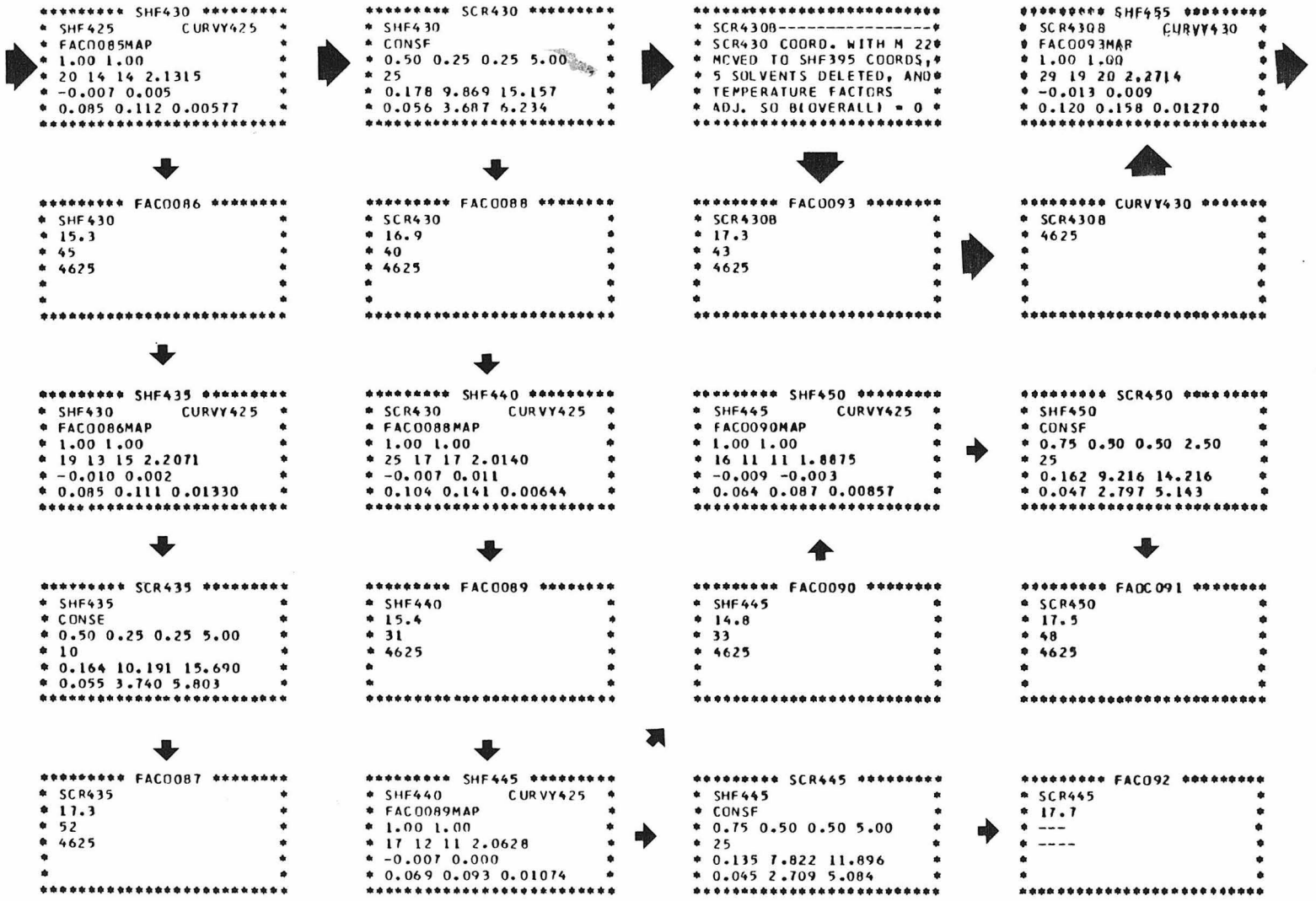

SHF300
SHF295 CURVY270
FAC0055MAP
1.00 1.00
25 18 19 2.5024
-0.013 0.010
0.112 0.144 0.02101
SCR315
SHF315
CONS
1.50 1.00 1.00 4.00
20
0.184 9.595 13.727
0.038 2.364 4.235
FAC0063
SCR325
19.5
65
4625
SHF330
SCR325 CURVY305
FAC0063MAP
1.00 1.00
29 21 21 1.9735
-0.021 0.006
0.132 0.165 0.02452
SCR300
SHF300
CONS
4.00 4.00 4.00 12.0
5
0.173 9.359 14.483
0.060 2.692 4.987
SHF315
SHF310 CURVY305
FAC0058MAP
1.00 1.00
25 15 18 2.2797
-0.022 -0.006
0.104 0.133
FAC0061
SCR315
20.0
60
4625
SCR325
SHF325
CONS
4.00 4.00 4.00 12.0
5
0.159 7.226 9.701
0.049 2.336 4.278
FAC0056
SCR300
19.7
59
4625
FAC0058
SHF310
17.0
56
4625
FAC0059
SHF315
16.7
40
4625
SHF325
SCR315 CURVY305
FAC0061MAP
1.00 1.00
37 24 23 2.1950
-0.021 0.015
0.154 0.194 0.02074
SHF305
SCR300 CURVY270
FAC0056MAP
1.00 1.00
34 21 23 2.4341
-0.026 0.006
0.144 0.179 0.03293
SHF310
SHF305 CURVY305
FAC0057MAP
1.00 1.00
25 17 18 2.6072
-0.024 -0.006
0.110 0.140 0.03260
SHF320
SHF315 CURVY305
FAC0059MAP
1.00 1.00
23 16 19 2.4772
-0.019 -0.002
0.104 0.134 0.02661
FAC0062
SHF325

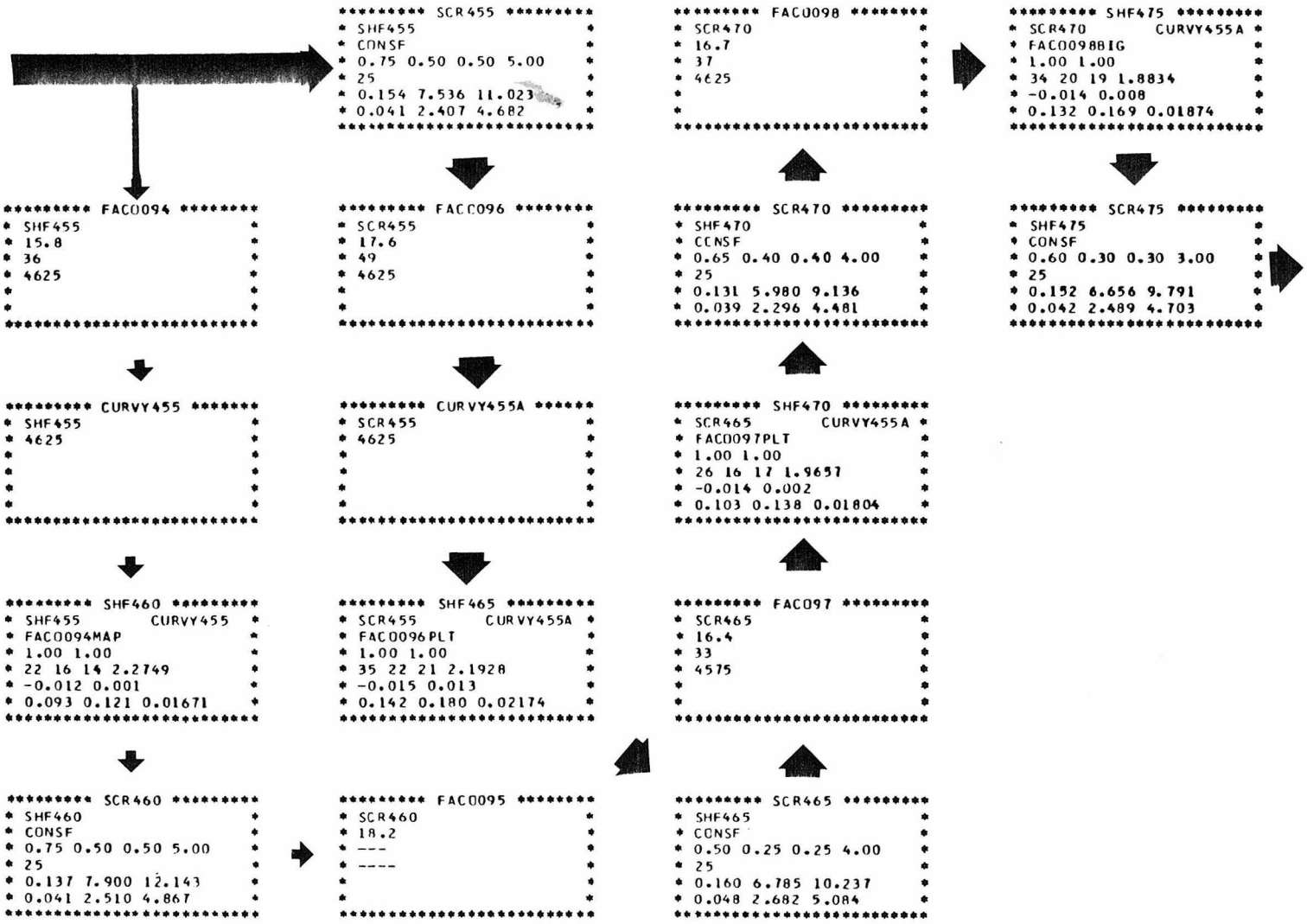
FAC0057
SHF305
17.9
56
4625
CURVY305
SHF305
4625
SCR320
SHF320
CONSC
4.00 4.00 4.00 16.0
5
0.178 10.273 14.774
0.063 2.884 4.917
FAC0060
SCR320
17.0
46
4625


SCR330
SHF330
CONSD
2.00 2.00 2.00 12.00
5
0.147 6.297 9.115
0.048 2.195 4.268
SHF340
SCR335B CURVY335
FAC0066MAP
1.00 1.00
36 24 24 2.7599
-0.023 0.016
0.158 0.197 0.02542
SCR340
SHF340
CONSE
0.75 0.33 0.33 5.00
5
0.175 8.370 14.002
0.052 3.238 5.250
SCR355
SHF355
CONSE
1.00 0.50 0.50 ????
5
0.175 9.519 14.842
0.042 3.202 4.946
FAC0064
SCR330
19.1
53
4625
FAC0066
SCR335B
22.2
86
4625
FAC0067
SCR340
19.6
76
4625
SHF355
SHF350 CURVY345
FAC0069MAP
1.00 1.00
19 14 12 2.3359
-0.025 -0.016
0.079 0.101 0.03122
SHF335
SCR330 CURVY305
FAC0064MAP
1.00 1.00
29 19 20 1.9241
-0.018 0.006
0.122 0.155 0.01782
CURVY335
SCR335B
4625
CURVY340
SCR340
4625
FAC0069
SHF350
15.8
40
4625
SCR335
SHF335
CONSD
1.00 0.50 0.50 ----
25
0.141 6.122 8.154
0.043 2.386 4.186
SCR335B----------
MANUAL ADJUSTMENTS
TO SCR335 COORDINATES
IN FAC0065A DIFFERENCE
MAP. TOTAL SOLVENT = 52
(13 DELETED & 23 ADDED)
SHF345
SCR340 CURVY340
FAC0067MAP
1.00 1.00
30 21 22 2.3040
-0.028 0.000
0.132 0.172 0.03180
SHF350
SHF345 CURVY345
FAC0068MAP
1.00 1.00
22 15 15 2.5017
-0.026 -0.012
0.091 0.120 0.03033
FAC065
SCR335
R-FACTOR = 19.5
FE ATOM CONSTRAINED
TO HEME PLANE
FAC0065A
SCR335
18.9
45
4625
FAC0068
SHF345
17.1
62
4625
CURVY345
SHF345
4625

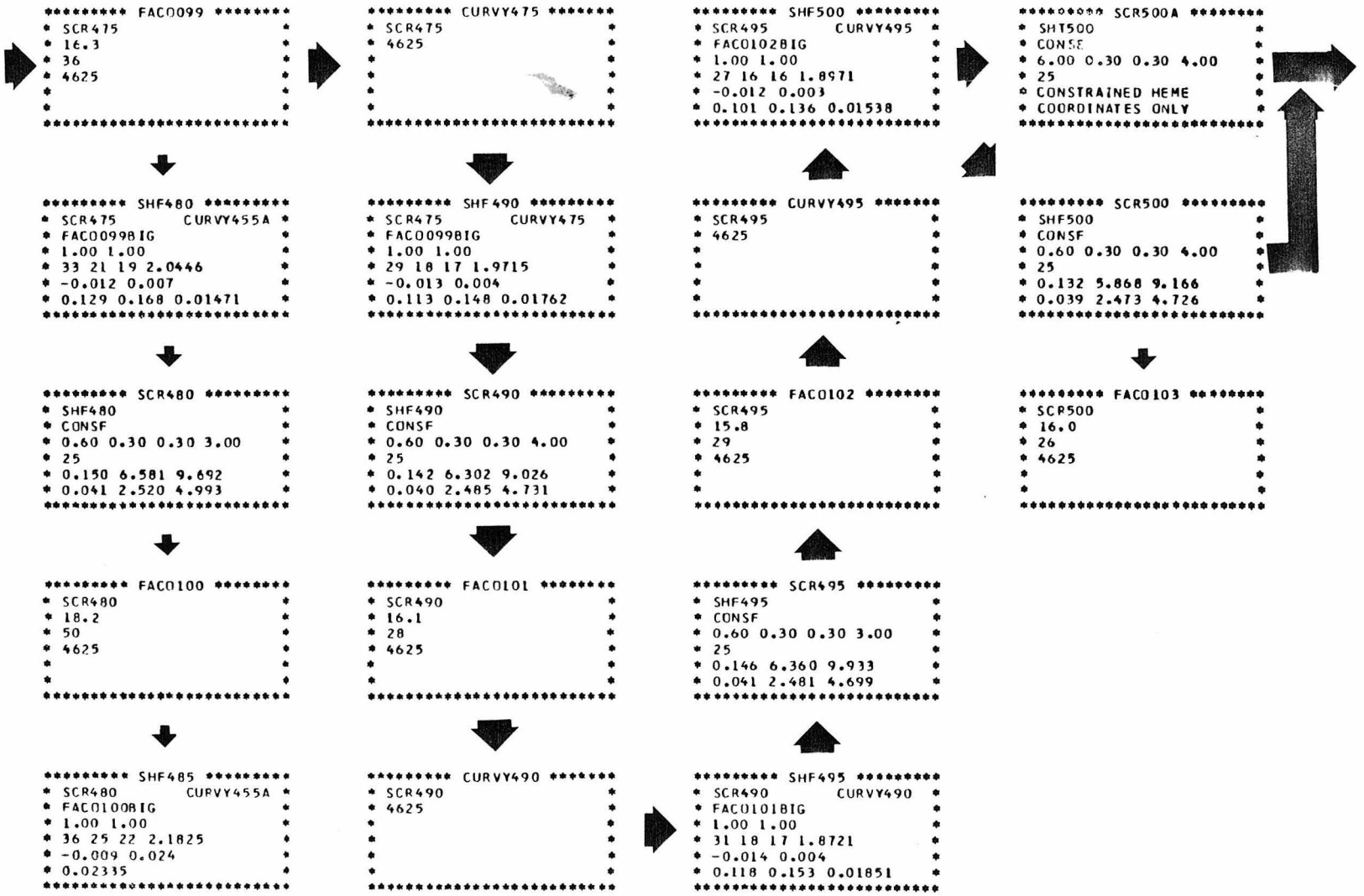

FAC0070
SCR355
18.1
57
4625
SHF360
SCR355 CURVY345
FAC0070MAP
1.00 1.00
31 20 22 2.1367
-0.035 -0.007
0.129 0.168 0.03524
FAC0071
SHF360
16.1
45
4625
CURVY360
SHF360
4625
SHF365
SHF360 CURVY360
FAC0071MAP
1.00 1.00
22 14 15 2.4530
-0.023 -0.011
0.089 0.115 0.02989
SCR365
SHF365
CONSE
0.50 0.25 0.25 10.00
5
0.141 7.540 11.952
0.051 3.376 5.219
FAC0072
SCR365
17.7
50
4625
SHF370
SCR365 CURVY360
FAC0072MAP
1.00 1.00
33 22 21 2.2030
-0.031 -0.001
0.137 0.176 0.03532
FAC0073
SHF370
16.1
51
4625
CURVY370
SHF370
4625
SHF375
SHF370 CURVY370
FAC0073MAP
1.00 1.00
25 16 16 2.4375
-0.008 0.000
0.105 0.131 -------
SCR375
SHF375
CONSE
0.50 0.20 0.20 10.0
10
0.139 7.825 11.871
0.047 3.459 5.410
FAC0074
SCR375
17.7
50
4625
SHF380
SCR375 CURVY370
FAC0074MAP
0.00 1.00
00 00 00 4.0571
-0.013 -0.013
0.000 0.000 0.01749
FAC0075
SHF380
17.5
53
4625
SHF385
SHF380 CURVY370
FAC0075MAP
1.00 1.00
32 21 20 1.4928
-0.011 0.015
0.128 0.168 0.01155
FAC0076
SHF385
15.9
47
4625
SHF390
SHF385 CURVY370
FAC0076MAP
1.00 1.00
22 15 14 2.2813
-0.005 0.006
0.090 0.117 0.01168
FAC0077
SCR390
17.6
50
4625
SCR390
SHF390
CONSE
0.50 0.25 0.25 ----
20
0.139 7.980 12.335
0.048 3.282 5.225

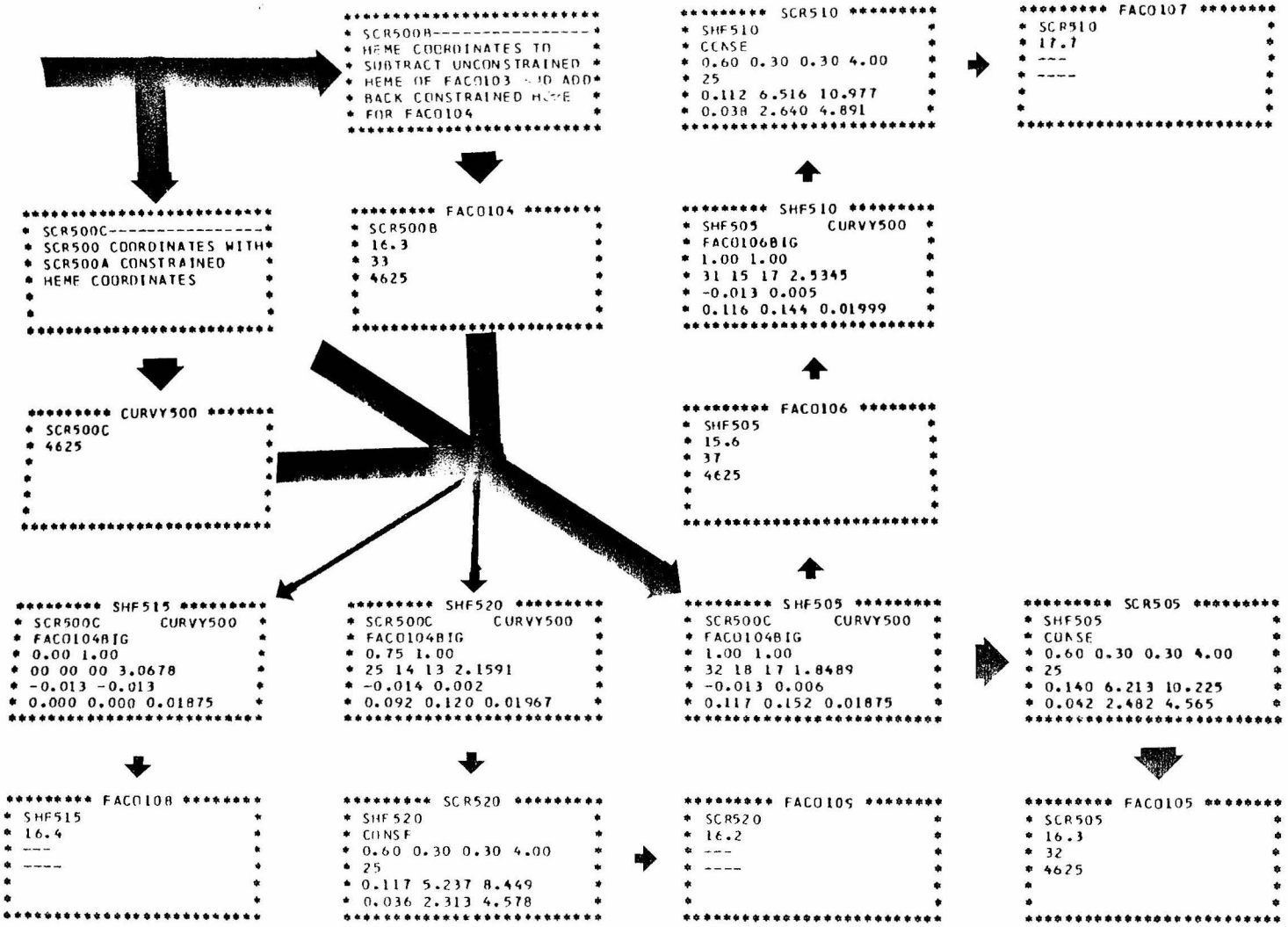

SHF395
SCR390 CURVY370
FAC0077MAP
0.00 1.00
00 00 00 3.5545
-0.013 -0.013
0.000 0.000 0.01508
FAC0078
SHF395
17.5
42
4625
SHT395A
MANUAL ADJUSTMENTS TO
SHT395 COORDINATES
IN FAC0078 DIFFERENCE
MAP. TOTAL SOLVENT = 42
(15 DELETED & 5 ADDED)
FAC0079
SHF395A
20.5
83
4625
SHF400
SHF395A CURVY370
FAC0079MAP
1.00 1.00
35 22 23 2.2723
-0.011 0.020
0.145 0.185 0.01053
FAC0080
SHF400
17.2
54
4625
CURVY400
SHF400
4625
SCR405
SHF405
CONSE
0.50 0.25 0.25 5.00
5
0.148 8.295 13.458
0.056 3.636 5.655
SHF405
SHF400 CURVY400
FAC0080MAP
1.00 1.00
24 15 18 2.6513
-0.011 0.004
0.096 0.129 0.01384
FAC0081
SCR405
17.8
57
4625
SHF410
SCR405 CURVY400
FAC0081MAP
1.00 1.00
30 20 19 2.2791
-0.018 0.005
0.122 0.160 0.01688
FAC0082
SHF410
16.1
48
4625
SHF415
SHF410 CURVY400
FAC0082MAP
1.00 1.00
21 14 15 2.2780
-0.011 0.002
0.089 0.118 0.01461
SCR415
SHF415
CONSE
0.50 0.25 0.25 5.00
5
0.139 8.074 12.900
0.053 3.568 5.503
FAC0083
SCR415
17.5
43
4625
SHF420
SCR415 CURVY400
FAC0083MAP
1.00 1.00
30 21 20 2.0516
-0.017 0.007
0.126 0.166 0.01708
FAC0084
SHF420
16.0
36
4625
SHF425
SHF420 CURVY400
FAC0084MAP
1.00 1.00
21 14 15 2.3232
-0.010 0.004
0.094 0.119 0.01414
FAC0085
SHF425
15.4
43
4625
CURVY425
SHF425
4625


SHF430
SHF425 CURVY425
FAC0085MAP
1.00 1.00
20 14 14 2.1315
-0.007 0.005
0.085 0.112 0.00577
SCR430
SHF430
CONSF
0.50 0.25 0.25 5.00
25
0.178 9.869 15.157
0.056 3.687 6.234
SCR430 COORD. WITH M 22
MOVED TO SHF395 COORDS,
5 SOLVENTS DELETED, AND
TEMPERATURE FACTORS
ADJ. SO B(OVERALL) = 0
SHF455
SCR430 CURVY430
FAC0093MAP
1.00 1.00
29 19 20 2.2714
-0.013 0.009
0.120 0.158 0.01270
FAC0086
SHF430
15.3
45
4625
FAC0088
SCR430
16.9
40
4625
FAC0093
SCR430
17.3
43
4625
CURVY430
SCR430
4625
SHF435
SHF430 CURVY425
FAC0086MAP
1.00 1.00
19 13 15 2.2071
-0.010 0.002
0.085 0.111 0.01330
SHF440
SCR430 CURVY425
FAC0088MAP
1.00 1.00
25 17 17 2.0140
-0.007 0.011
0.104 0.141 0.00644
SHF450
SHF445 CURVY425
FAC0090MAP
1.00 1.00
16 11 11 1.8875
-0.009 -0.003
0.064 0.087 0.00857
SCR450
SHF450
CONSF
0.75 0.50 0.50 2.50
25
0.162 9.216 14.216
0.047 2.797 5.143
SCR435
SHF435
CONSE
0.50 0.25 0.25 5.00
10
0.164 10.191 15.690
0.055 3.740 5.803
FAC0089
SHF440
15.4
31
4625
FAC0090
SHF445
14.8
33
4625
FAC0091
SCR450
17.5
48
4625
FAC0087
SCR435
17.3
52
4625
SHF445
SHF440 CURVY425
FAC0089MAP
1.00 1.00
17 12 11 2.0628
-0.007 0.000
0.069 0.093 0.01074
SCR445
SHF445
CONSF
0.75 0.50 0.50 5.00
25
0.135 7.822 11.896
0.045 2.709 5.084
FAC092
SCR445
17.7


SCR455
SHF455
CONSF
0.75 0.50 0.50 5.00
25
0.154 7.536 11.023
0.041 2.407 4.682
FAC0098
SCR470
16.7
37
4625
SHF475
SCR470 CURVY455A
FAC0098BIG
1.00 1.00
34 20 19 1.8834
-0.014 0.008
0.132 0.169 0.01874
FAC0094
SHF455
15.8
36
4625
FAC0096
SCR455
17.6
49
4625
SCR470
SHF470
CONSF
0.65 0.40 0.40 4.00
25
0.131 5.980 9.136
0.039 2.296 4.481
SCR475
SHF475
CONSF
0.60 0.30 0.30 3.00
25
0.152 6.656 9.791
0.042 2.489 4.703
CURVY455
SHF455
4625
CURVY455A
SCR455
4625
SHF470
SCR465 CURVY455A
FAC0097PLT
1.00 1.00
26 16 17 1.9657
-0.014 0.002
0.103 0.138 0.01804
SHF460
SHF455 CURVY455
FAC0094MAP
1.00 1.00
22 16 14 2.2749
-0.012 0.001
0.093 0.121 0.01671
SHF465
SCR455 CURVY455A
FAC0096PLT
1.00 1.00
35 22 21 2.1928
-0.015 0.013
0.142 0.180 0.02174
FAC097
SCR465
16.4
33
4575
SCR460
SHF460
CONSF
0.75 0.50 0.50 5.00
25
0.137 7.900 12.143
0.041 2.510 4.867
FAC0095
SCR460
18.2

SCR465
SHF465
CONSF
0.50 0.25 0.25 4.00
25
0.160 6.785 10.237
0.048 2.682 5.084


FAC0099
SCR475
16.3
36
4625
CURVY475
SCR475
4625
SHF500
SCR495 CURVY495
FAC0102BIG
1.00 1.00
27 16 16 1.8971
-0.012 0.003
0.101 0.136 0.01538
SCR500A
SHT500
CONSF
6.00 0.30 0.30 4.00
25
CONSTRAINED HEME
COORDINATES ONLY
SHF480
SCR475 CURVY455A
FAC0099BIG
1.00 1.00
33 21 19 2.0446
-0.012 0.007
0.129 0.168 0.01471
SHF490
SCR475 CURVY475
FAC0099BIG
1.00 1.00
29 18 17 1.9715
-0.013 0.004
0.113 0.148 0.01762
CURVY495
SCR495
4625
SCR500
SHF500
CONSF
0.60 0.30 0.30 4.00
25
0.132 5.868 9.166
0.039 2.473 4.726
SCR480
SHF480
CONSF
0.60 0.30 0.30 3.00
25
0.150 6.581 9.692
0.041 2.520 4.993
SCR490
SHF490
CONSF
0.60 0.30 0.30 4.00
25
0.142 6.302 9.026
0.040 2.485 4.731
FAC0102
SCR495
15.8
29
4625
FAC0103
SCR500
16.0
26
4625
FAC0100
SCR480
18.2
50
4625
FAC0101
SCR490
16.1
28
4625
SCR495
SHF495
CONSF
0.60 0.30 0.30 3.00
25
0.146 6.360 9.933
0.041 2.481 4.699
SHF485
SCR480 CURVY455A
FAC0100BIG
1.00 1.00
36 25 22 2.1825
-0.009 0.024
0.02335
CURVY490
SCR490
4625
SHF495
SCR490 CURVY490
FAC0101BIG
1.00 1.00
31 18 17 1.8721
-0.014 0.004
0.118 0.153 0.01851


SCR500B
HEME COORDINATES TO
SUBTRACT UNCONSTRAINED
HEME OF FAC0103 AND ADD
BACK CONSTRAINED HEME
FOR FAC0104
SCR510
SHF510
CONSE
0.60 0.30 0.30 4.00
25
0.112 6.516 10.977
0.038 2.640 4.891
FAC0107
SCR510
17.1
SCR500C
SCR500 COORDINATES WITH
SCR500A CONSTRAINED
HEME COORDINATES
FAC0104
SCR500B
16.3
33
4625
SHF510
SHF505 CURVY500
FAC0106BIG
1.00 1.00
31 15 17 2.5345
-0.013 0.005
0.116 0.144 0.01999
CURVY500
SCR500C
4625
FAC0106
SHF505
15.6
37
4625
SHF515
SCR500C CURVY500
FAC0104BIG
0.00 1.00
00 00 00 3.0678
-0.013 -0.013
0.000 0.000 0.01875
SHF520
SCR500C CURVY500
FAC0104BIG
0.75 1.00
25 14 13 2.1591
-0.014 0.002
0.092 0.120 0.01967
SHF505
SCR500C CURVY500
FAC0104BIG
1.00 1.00
32 18 17 1.8489
-0.013 0.006
0.117 0.152 0.01875
SCR505
SHF505
CONSE
0.60 0.30 0.30 4.00
25
0.140 6.213 10.225
0.042 2.482 4.565
FAC0108
SHF515
16.4
SCR520
SHF520
CONSE
0.60 0.30 0.30 4.00
25
0.117 5.237 8.449
0.036 2.313 4.578
FAC0109
SCR520
16.2
FAC0105
SCR505
16.3
32
4625