UNIVERSITY OF CALIFORNIA

San Diego

The Crystal Structure Determination of Cytochrome c_2 from R. rubrum and Related Topics in Protein X-ray Crystallography

A dissertation submitted in partial satisfaction of the requirements for the Degree Doctor of Philosophy in Chemistry

by

Francis Raymond Salemme

Committee in charge:

Professor Joseph Kraut, Chairman
Professor Nathan O. Kaplan
Professor William S. Allison
Professor Melvin Simon
Professor Paul A. Price

1972

The dissertation of Francis Raymond Salemme is approved and is acceptable in quality and form for publication on microfilm:

N. O. Kopell

Paul Price

Martin D Kamen Melvin Calvin

William A. Allison

Joseph Kraut

Committee Chairman

University of California, San Diego

1972

This thesis is dedicated to

Alfred Moreland, who started me thinking

Patrick La Fratta, who taught me how to make things

Harold Stockman S. J., who taught me how to stand up for myself

My Father, who taught me to be patient and compassionate

and

Janet, who reawakened me to the purpose of life.

TABLE OF CONTENTS

Page

LIST OF FIGURES AND TABLE

Figure | Page

CHAPTER III

CHAPTER IV

CHAPTER II

Table | Page

ACKNOWLEDGMENT

I wish to express my gratitude to my thesis advisor, Joseph Kraut, for his constant help and encouragement during the course of the work presented in this thesis. Thanks are due to Steve Freer and Richard Alden for their assistance in the computational aspects of this investigation. I thank Martin Kamen and Robert Bartsch for their ready assistance and illuminating discussions concerning bacterial cytochromes. I thank David Fehr for his expert help in the design and construction of the coordinate hunting engine. I thank John Robertus and Charles Carter for furnishing a friendly critical audience for the past few years.

VITA

8 June 1945 - Born - Norwood, Massachusetts

1963-1964	U. S. Merchant Marine, Lamont Geological Observatory, Palisades, New Jersey
1967	B. A., Molecular Biophysics Yale University, New Haven, Conn.
1969	M. S., Chemistry University of California, San Diego, La Jolla, California

PUBLICATIONS

"Involvement in Biology Today," with Allen Vegotsky and Sam Wilson. CRM Books, Del Mar, California, 1972

FIELD OF STUDY

Molecular Biophysics: Protein X-ray Crystallographic Structure Determination, Morphological Evolution of Protein Structure and Function

Professor Joseph Kraut

ABSTRACT OF THE DISSERTATION

The Crystal Structure Determination of Cytochrome $\underline{c}_2$ from R. rubrum and Related Topics in Protein X-ray Crystallography

by

Francis Raymond Salemme

Doctor of Philosophy in Chemistry

University of California, San Diego, 1972

Professor Joseph Kraut, Chairman

This thesis consists of four chapters, each of which is concerned with a different aspect of protein structure determination by x-ray crystallography.

Chapter I contains a description of a novel general technique for the production of large single crystals for protein crystallography by free interface diffusion. Chapter II contains the description of the high resolution crystal structure determination and structural analysis of cytochrome $\underline{c}_2$ from Rhodospirillus rubrum. Chapter

III contains a description of some preliminary x-ray crystallographic studies of glyceraldehyde 3-phosphate dehydrogenase from Palinurus. Chapter IV concerns the design, construction, and application of a novel device allowing rapid and accurate measurement and construction of molecular models.

CHAPTER I

A FREE INTERFACE DIFFUSION TECHNIQUE FOR THE CRYSTALLIZATION OF PROTEINS FOR X-RAY CRYSTALLOGRAPHY

INTRODUCTION

Recent advances in protein structure determination have created great interest in improved techniques for obtaining large single crystals (average linear dimension = 0.2 mm) suitable for x-ray diffraction work. The most desirable features in such a technique are 1) that it be sufficiently simple and flexible to easily allow exploration of a wide range of crystallization parameters such as solvent pH, dielectric constant, ionic strength, or the effect of various additives and 2) that it allow experiments to be carried out with small amounts of starting material. Reports of a number of techniques fulfilling these criteria have appeared in the literature (1-4).

The method described in this communication utilizes free diffusion between a concentrated protein solution and a precipitant solution to attain the conditions of protein supersaturation requisite for the nucleation and subsequent growth of large single crystals. Protein and precipitant solutions are layered directly over each other in small crystallization cells or tubes, and allowed to diffuse to equilibrium. At the onset of interdiffusion of the protein and precipitant layers, the protein in the immediate vicinity of the interface is exposed to a transient supersaturating concentration of precipitant. This transient protein supersaturation induces the

formation of nucleii which serve as crystal growth loci when equilibrium is reached. At equilibrium the overall precipitant concentration in the total volume is substantially less than that required to spontaneously precipitate the protein, facilitating the gradual growth of a few large crystals from the nucleii created at the initial interface. This method differs in principle from previously described diffusion techniques, since in contrast to membrane or vapor diffusion techniques which slowly approach protein supersaturation by a gradual increase in precipitant concentration, this method initially exposes some of the protein to a supersaturating precipitant concentration. Subsequent interdiffusion of the protein and precipitant layers gradually reduces their relative concentrations until at equilibrium, a precipitant concentration sufficient to promote crystal growth from any nucleii present, but lower than the concentration required to induce spontaneous nucleation is attained. The free diffusion method offers manipulative advantages because of its simplicity.

MATERIALS AND METHODS

Reagents: "Special Enzyme Grade" ammonium sulfate was obtained from Mann Research Laboratory, New York, N. Y. Solutions of ammonium sulfate were prepared by dissovling a specified weight of salt in 1 liter of solution at 25° C according to the nomogram of

Dixon (5). All other chemicals were standard analytical reagent grade. Doubly glass distilled water was used for the preparation of all solutions. Parafilm is a deformable polyethylene film manufactured by the American Can Co., Neenah, Wis.

Proteins: Cytochrome c_2 of Rhodospirillus rubrum was isolated by a modification of the method of Horio and Kamen (6). Cytochrome c' of R. rubrum and cytochrome c' of Rhodopseudomonas Palustris (7) were the gift of Dr. Robert Bartsch. Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12) was isolated from commercially obtained frozen South African spiny lobster (Palinurus) tails by the method of Allison and Kaplan (8).

Concentrated stock solutions of proteins were prepared by centrifuging suspensions of microcrystals grown in ~ 90% saturated ammonium sulfate, and dissolving the resulting pellet in a minimum volume of cold water or buffer. The concentration of these solutions was assayed spectrophotometrically and adjusted to the desired "stock" concentration with the addition of more buffer. The final stock concentration of the cytochrome solutions (MW = 13-16,000) was 30 mg/ml. The concentration of Palinurus glyceraldehyde 3-phosphate dehydrogenase (MW = 144,000) was 50 mg/ml. It is advantageous to keep stock protein solutions minimally buffered so that crystallization conditions may be varied by adjustments of the pH and ion content of the precipitant. Protein solutions prepared

from centrifuged salt precipitates in this manner contain residual ammonium sulfate (5 to 10% saturated), which does not normally interfere with the course of the experiments unless crystallization from organic solvents is attempted. In this case it may be necessary to remove residual salt by dialysis. The conditions for crystallization reported below did not require desalting of the stock solutions prior to crystallization. In the case of Palinurus glyceraldehyde 3-phosphate dehydrogenase, the presence of the residual ammonium sulfate appeared to enhance the stability of the protein during storage. Concentrated protein solutions are stored at 5° C in plastic centrifuge tubes. The solutions are centrifuged for 10 minutes at 10,000 × g prior to the removal of aliquots for crystallization in order to pellet any dust or debris which may be present in the solution.

Crystallization Experiments in Microcells: An apparatus useful in searching for proper crystallization conditions in cases where only a small amount of protein (< 5 mg) is available is shown in Fig. 1. The apparatus consists of a disposable Pasteur pipette (Van Waters and Rogers, 9" × 7 mm O. D., #14672-040) which has been cut off so that the large diameter section is about 2 cm long. The cell is loaded by drawing a premeasured volume of the least dense solution into the capillary bore of the pipette by means of a screw type pipetting aid so that the lower meniscus of the solution volume remains even with the end of the capillary. A premeasured volume of the more dense solution

is then drawn into the capillary bore so that both solutions form a continuous liquid volume having a sharp interface. With the tube maintained in a vertical orientation, the protein/precipitant volume is drawn higher into the capillary bore. The end is sealed with a glob of hard wax. The pipetting aid is carefully removed and the top of the tube is sealed with Parafilm. The useful capacity of these gradient tubes ranges from 10 to 100 μl. The quantity of protein suitable to perform an individual test in a microcell ranges from 0.06 to 0.5 mg. Since the capillary bore of the pipette tapers near the top, the time required to reach equilibrium between the protein and precipitant layers may be altered by adjusting the height of the total solution volume in the capillary bore (Fig. 2). Diffusion to equilibrium in these capillaries, as observed with colored cytochrome solutions (M. W. approx. 13,000) layered over 50 to 80 per cent saturated ammonium sulfate, takes place within about an hour, depending upon conditions. The method is adaptable to smaller scale or longer diffusion times by drawing the capillary bore to a smaller diameter.

Larger Scale Crystallization: In cases where more than a few milligrams of protein are available for experimentation, gradients may be set up in miniature test tubes (Fig. 3). These tubes are easily manufactured by heat sealing a 10 to 25 mm length of 5 to 7 mm I. D. pyrex glass tubing. Protein and precipitant solutions

are carefully layered in with a micropipette. The tube is sealed with Parafilm. Diffusion to equilibrium usually takes place within one hour of setting up the gradient. The total volume used in a single tube ranges from 50 to 200 μl and contains 0.6 to 2.5 mg of protein.

Searching for Crystallization Conditions: Before a search for crystallizing conditions is undertaken for a previously uncrystallized protein, it is advantageous to make a survey of its physical and chemical properties. In this way it is possible to limit the range of crystallization parameters which must be varied and to assure that optimal conditions for the maintenance of the homogeneity and stability of the protein are met. Some important factors to be considered follow.

1) pH. Proteins exhibit widely variable abilities to maintain their "native conformation" as the pH of their solutions is varied from that at which they physiologically function. Assay of enzymatic activity, spectroscopy, or other conformational probes such as circular dichroism or optical rotary dispersion often give useful indications of the extent of deviation from the native conformation induced by a pH change.

Protein solubility may not change continuously as solution pH is varied from the protein isoionic point due to discontinuous ion binding to ionizable groups of the protein molecules. This property is especially manifest in solutions of high ionic strength. Few proteins

are crystallized from high ionic strength solutions buffered at the protein's isoionic point.

2) Specific additives for protein stabilization. Individual proteins may require additive ions or chemical modification to maintain molecular homogeneity in solution.

Proteins containing active sulfhydryl groups may react with heavy metal ions or oxidants contaminating crystallization solutions. In some cases these fortuitous reaction products are sufficiently different in conformation from the native molecule to frustrate crystallization attempts. The presence of small concentrations of reducing agents and a chelating agent such as EDTA may prevent such undesirable reactions.

Inorganic ions such as Ca^{++}, Mg^{++}, Cs^{+} or $Po_4^{=}$ in some instances bind to proteins and often stabilize them in solutions.

Proteolytic enzymes may be protected from autolysis by reaction with appropriate specific inhibiting agents. In like manner, conformational stability may be enhanced in some enzymes by the presence of competitive inhibitors or cofactors which form stable complexes with the protein.

Proteins containing metals as prosthetic groups may require special treatment to assure the integrity of the metal-protein binding. Proteins containing reversibly oxidizable and reduceable metal chelates should be stabilized in the desired oxidation state of the presence

of a low concentration of suitable redox reagent as a conformational change often accompanies oxidation and reduction of the metal.

3) Temperature. Proteins exhibit widely varying conformational sensitivity to changes in temperature. In cases where the range of stability is small, it is customary to perform all isolation and crystallization procedures at 4° C. The resultant crystals may or may not be as temperature sensitive as the protein in solution. Other proteins appear to be relatively insensitive to temperature changes in the range of 4 to 37°C. Proteins showing this type of conformational stability may be advantageously crystallized at temperatures up to 37° C. As Jacoby (9) has pointed out, proteins become less soluble in concentrated salt solutions as the temperature is increased. In practical terms, this means that at higher temperature, crystals may be grown from less concentrated salt solutions than that required at 4° C or room temperature. This is of particular importance when the protein is very soluble and requires nearly saturated salt solutions for crystallization. In such cases it is often difficult to form the protein heavy-atom derivatives required for x-ray structure analysis (10).

Once conditions requisite for the maintenance of the protein's conformational stability are established, a suitable precipitating agent must be found. The precipitant should alter the solvent properties of the aqueous protein solution sufficiently to render the

protein insoluble, and yet not disrupt its tertiary molecular conformation.

All of the proteins discussed in this communication had, as a step in their purification, been precipitated with ammonium sulfate. Consequently crystallization experiments were carried out using this salt as a precipitant, eventually resulting in the establishment of routine preparative conditions.

Zeppezauer (3) has described a variety of organic and inorganic precipitating agents which cover a wide range of pH and dielectric constants. A summary of protein crystallization conditions illustrating the usefulness of several of these precipitants is given in the recent review of Eisenberg (11).

I. Selection of a precipitating agent is performed by squirting a 1 μl aliquot of the concentrated (2-5%) protein solution into a miniature test tube containing 50 μl of concentrated precipitant solution buffered at a suitable pH together with the required reducing agents, stabilizing ions, etc. (The initial concentration of the precipitant solution used is about 90% saturated in the case of ionic salts or 50% by volume in the case of organic precipitants.)

a. If a precipitate does not form, additional tests are performed at higher precipitant concentration. Some proteins, however, remain soluble in saturated salt solutions. The ionic strength of such solutions may in some cases be further increased with the addition of a different salt.

b. If a precipitate forms, the contents of the tube are carefully diluted with buffer until the precipitate dissolves. 1) If the precipitate dissolves, subsequent tests at lower precipitant concentration will give a rough estimate of the minimum concentration required to precipitate the protein under these conditions. 2) If the protein does not redissolve, it may be surmised that it has been irreversibly denatured. Additional tests, however, may be made at lower precipitant concentration, until a minimum precipitating concentration is found. If the precipitate is still insoluble, it is best to look for another precipitating agent.

II. After a precipitating agent has been found which renders the protein reversibly soluble and insoluble, test gradients are set up in order to find the proper conditions for crystallization.

The first tests are carried out using equal volumes of protein (3 to 5%) and precipitant solutions, layered into the microcells or small test tubes as described above. The initial conditions should be adjusted so that when the cell reaches equilibrium, the precipitant concentration in the total solution volume is about 15% less than the minimum precipitating concentration found in the solubility tests. This reduction in precipitant concentration compensates for the observed decrease in protein solubility with increasing protein concentration in concentrated precipitant solutions. Since all of the precipitant is initially present in one half of the cell volume, the

precipitant solution layered into the cell is roughly twice as concentrated as the desired final precipitant concentration. When precipitants such as ammonium sulfate are used which have a non-linear % saturation vs. weight/volume curve, appropriate compensating adjustments in the precipitant concentration must be made. A useful nomogram for calculating ammonium sulfate concentrations is given in reference 5.

The appearance of a flocculent precipitate at the interface indicates that the precipitant concentration is too high. In this case, the precipitant concentration should be reduced until only slight turbidity is observed at the interface. If crystal growth or precipitation does not ensue within a week or so after the gradient has reached equilibrium, it may be surmised that the final precipitant concentration in the tube is now too low to induce crystallization.

In this case it is necessary to adjust conditions to reduce the initial precipitant concentration gradient at the interface, while simultaneously maintaining sufficient overall precipitant concentration to promote crystal growth when the cell has reached equilibrium. This is done by increasing the volume of the precipitant layer relative to the protein solution volume. The precipitant solution concentration is accordingly reduced so that the final equilibrium precipitant concentration in the total volume equals that initially determined above for the equal volume test. This procedure lessens the initial gradient

in ionic strength imposed upon the protein at the interface and reduces the final protein concentration in the cell at equilibrium.

Amorphously precipitated material protein is removed from the microcells by breaking the wax seal and pushing the solution out with a pipetting aid. Precipitated protein is easily removed from the small tubes with a Pasteur pipette. Precipitated material may be centrifuged and redissolved with a minimal volume of buffer to yield a concentrated protein solution suitable for conducting further experiments. Recycling protein in this fashion minimizes handling losses and and allows a large number of experiments to be performed with small quantities (< 10 mg) of starting material.

The presence of a shower of small crystals in the cell may indicate that the precipitant concentration is too high or that crystal nucleii were initially present in the protein solution. Once small crystals have been obtained, subsequent experiments in which the temperature or precipitant pH is altered, or in which various salts or organic ions are added will often increase the size of the crystals or reform an undesirable habit.

Although crystals may be grown in either microcells or small tubes, it is advantageous to perform experiments in tubes as soon as approximately correct crystallization conditions are established. The large number of crystals required for x-ray work can then be grown by simultaneously preparing several tubes under conditions identical to those which experimentally gave good crystals.

RESULTS

Cytochrome c_2 (R. rubrum, MW 12,800) is crystallized from ammonium sulfate at pH 5.8 as flat rhombic plates. At 23°, the crystals are grown by layering 50 µl of 30 mg/ml protein solution in water, containing approximately 5% residual saturated ammonium sulfate over 100 µl 85% saturated unbuffered ammonium sulfate (Fig. 4). At 37°, crystals are grown by layering 50 µl of protein solution over 50 µl of 65% saturated unbuffered ammonium sulfate pH 5.8 (Fig. 5a). Both crystal types have similar habits and are isomorphic. Crystals generally required two weeks to one month to grow and commonly reached 0.8 × 0.6 × 0.25 mm in size. Crystals grown at room temperature under high salt conditions formed only one useful heavy-atom isomorphic derivative in approximately 200 attempts. The crystals grown at 37°C yielded five potentially useful derivatives in the first few derivitization attempts and made possible the completion of the crystal structure analysis of this protein.

Cytochrome c' (Rps. palustris, MW approx. 15,000) crystallizes in two distinct habits under differing conditions. This protein was crystallized at 37° since its solubility characteristics are similar to those of cytochrome c_2. Crystals prepared by layering 50 µl of 30 mg/ml protein solution over 100 µl of unbuffered 65% saturated ammonium sulfate pH 5.8 and stored at 37°C grew as long, flat

rectangular prisms having dimensions 1.0 × 0.5 × 0.3 mm (Fig. 5b). Crystals prepared by layering 50 μl of 30 mg/ml protein solution over 50 μl of 65% saturated ammonium sulfate 0.1 M in $Mg(NO_3)_2$ pH 5.2 at 37°C, grew as beveled rectangular prisms (Fig. 5c) having dimensions 0.6 × 0.4 × 0.3 mm. Both crystal forms require about a month to grow to useable size.

Cytochrome c' (R. rubrum, MW approx. 16,000) also crystallizes in two distinct habits under different conditions. Crystals prepared by layering 50 μl of 30 mg/ml protein solution over 50 ml of 65% saturated unbuffered ammonium sulfate pH 5.8 at 37°C grew as long rods having a diamond-shaped cross section (Fig. 5d) having dimensions 2.0 × 0.2 × 0.2 mm. Crystals prepared by layering 50 μl of 30 mg/ml protein solution over 50 μl of 65% saturated ammonium sulfate 0.1 M in $Mg(NO_3)_2$ pH 5.2 at 37° grew as beveled rectangular prisms (Fig. 5e) about 0.4 × 0.3 × 0.2 mm in size.

All of the cytochromes discussed above are stable and exist predominantly (> 95%) in their oxidized form due to air oxidation. Preliminary attempts to grow large crystals of the reduced forms of these cytochromes indicate that reduced cytochrome crystals do not form under the same conditions producing oxidized crystals, although in one case (R. rubrum c_2) it is known that oxidized and reduced crystals are isomorphic.

Glyceraldehyde 3-phosphate dehydrogenase (Palinurus) was obtained as a 50 mg/ml solution by dissolving centrifuged microcrystals in 0.05 M sodium phosphate pH 6.5 containing 0.05 M 2-mercaptoethanol and 0.005 M EDTA. Crystals were grown at 23°C by layering 50 µl of stock solution over 100 µl of identically buffered ammonium sulfate. The protein crystallized as flat rhombic plates (Fig. 5e). The largest crystals grown were 0.6 × 0.5 × 0.2 mm.

DISCUSSION

The production of large single crystals is dependent upon the introduction of a small number of nucleating sites into a solution which is saturated with protein. When growing crystals by batch methods, this condition is effected by slowly adding a precipitant to a concentrated protein solution until a slight turbidity appears, signalling the formation of microcrystalline or amorphously precipitated material. The solution is generally well supersaturated with protein at this point and crystallization or precipitation will continue to completion unless more solvent is added to the dish to just remove the turbidity, presumably leaving a few nucleii to serve as growth points for a few large crystals. The same effect may also be obtained by putting a dish showing turbidity into the cold, which usually renders the protein more soluble, or by seeding a saturated protein solution with a few

microscopic seeds obtained from crushing a previously grown crystal of the same protein (11).

The free interface diffusion technique described in this communication arrives at the required conditions in the following way: When the concentrated protein and precipitant layers begin to diffuse together at the interface, a transient condition of protein supersaturation sufficient to induce spontaneous nucleation in the protein is attained. As diffusion continues and the band of interpenetration of protein and precipitant widens, the relative concentrations decrease until finally at equilibrium the total solution volume is saturated with protein but unable of itself to induce spontaneous nucleation. A few of the nucleii formed at the initial interface may survive, however, and serve as growth sites for crystals. These nucleii usually fall to the bottom or collide with the walls of the tube, although occasionally crystals are observed either to form in a ring on the tube walls at the location of the original gradient or as an avalanche down one side of the tube.

The manner in which the conditions for nucleation and crystal growth are attained in the free diffusion technique discussed above differs from the way in which the required conditions are effected in other diffusion methods described in the literature. Both the Zeppezauer (1, 3, see also 2) microdiffusion dialysis technique and the vapor diffusion technique described by Davies and Segal (4)

gradually approach protein saturation by continuously increasing precipitant or precipitant/protein concentration. Crystallization is probably induced by the presence of some foreign particle acting as a heterogeneous nucleus for crystal growth, since sufficient protein supersaturation to induce spontaneous homogeneous nucleation usually results in a shower of microcrystals. The free diffusion technique approaches conditions of protein saturation from a transient condition of supersaturation sufficient to induce spontaneous nucleation. At equilibrium, a few of these nucleii remain to serve as growth points for crystals.

FIGURE 1

A free diffusion microcell manufactured from a cut-off Pasteur disposable pipette. The cell is stored in a vertical position to allow even diffusion between protein and precipitant layers.

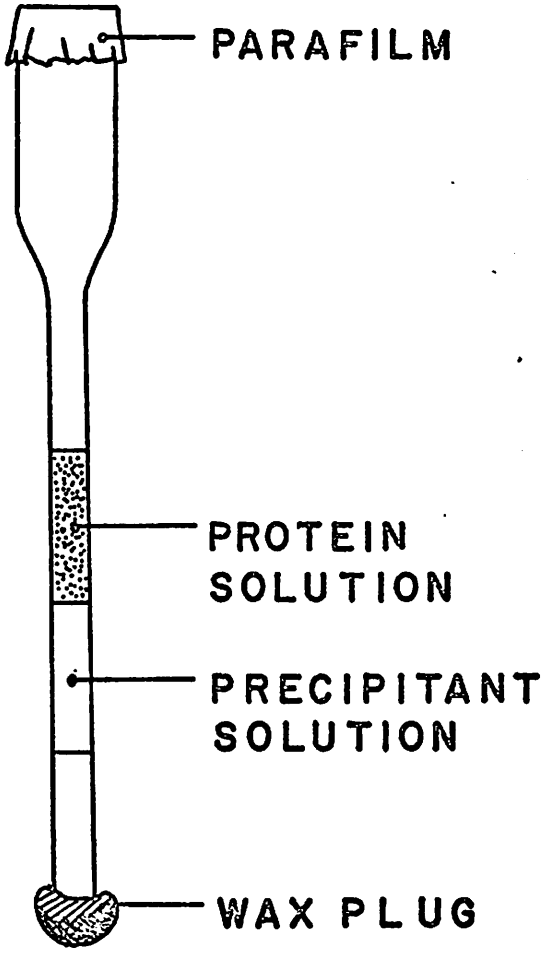
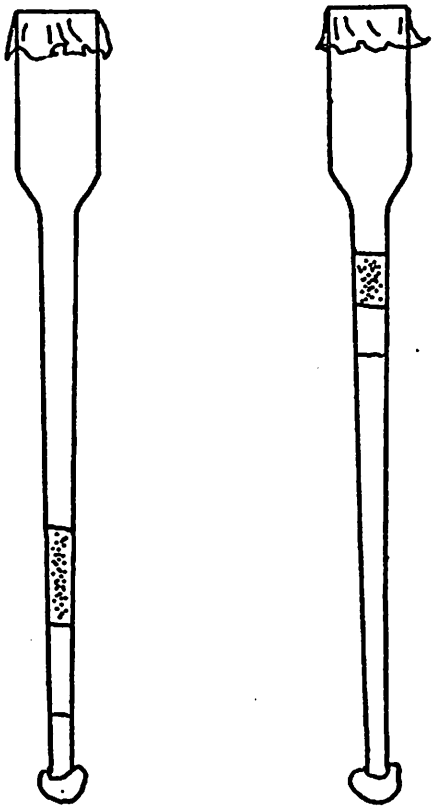

PARAFILM
PROTEIN SOLUTION
PRECIPITANT SOLUTION
WAX PLUG

FIGURE 2

The time required to reach equilibrium between protein and precipitant layers in a microcell may be decreased by drawing the protein/precipitant volume into a part of the capillary having a larger cross section. Cells A and B contain the same total solution volume. The area of the interface is greater and the diffusion path length shorter in Cell B than in Cell A. As a result, approach to equilibrium is more rapid in Cell B.



A B

FIGURE 3

Free diffusion macrocell manufactured from Pyrex glass tubing. The rate at which the cell reaches equilibrium may be varied by altering the relative dimensions of the tube, in analogy to that described in Fig. 2.

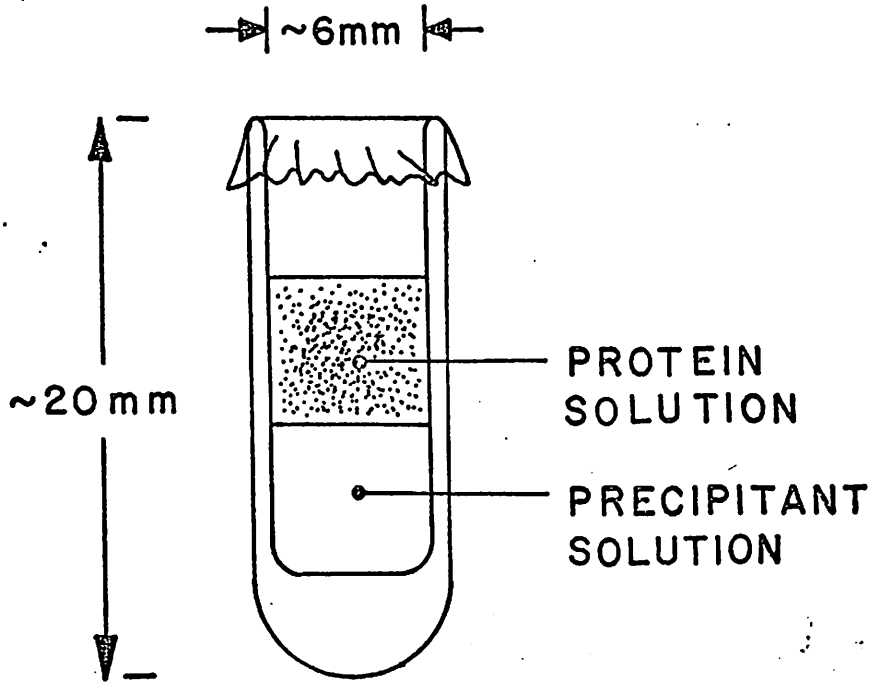

~6mm
~20mm
PROTEIN SOLUTION
PRECIPITANT SOLUTION

FIGURE 4

Cytochrome c_2 (R. rubrum) crystals grown at room temperature in a gradient tube.

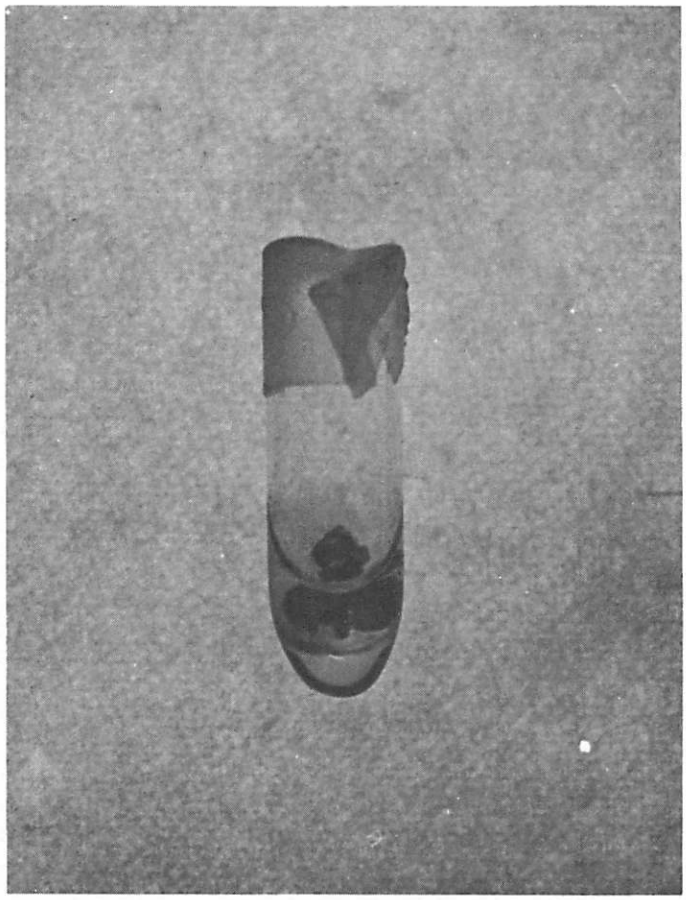


FIGURE 5

Protein crystals. The scale is given by the 0.4 mm bar at the left of each picture.

a) Cytochrome c_2 (R. rubrum)

b) Cytochrome c' (Rps. palustris), Form 1

c) Cytochrome c' (Rps. palustris), Form 2

d) Cytochrome c' (R. rubrum), Form 1

e) Cytochrome c' (R. rubrum), Form 2

f) Glyceraldehyde 3-phosphate dehydrogenase (Palinurus)

Details of crystallizing conditions are given in text.

FIGURE 5

Protein crystals. The scale is given by the 0.4 mm bar at the left of each picture.

a) Cytochrome c_2 (R. rubrum)

b) Cytochrome c' (Rps. palustris), Form 1

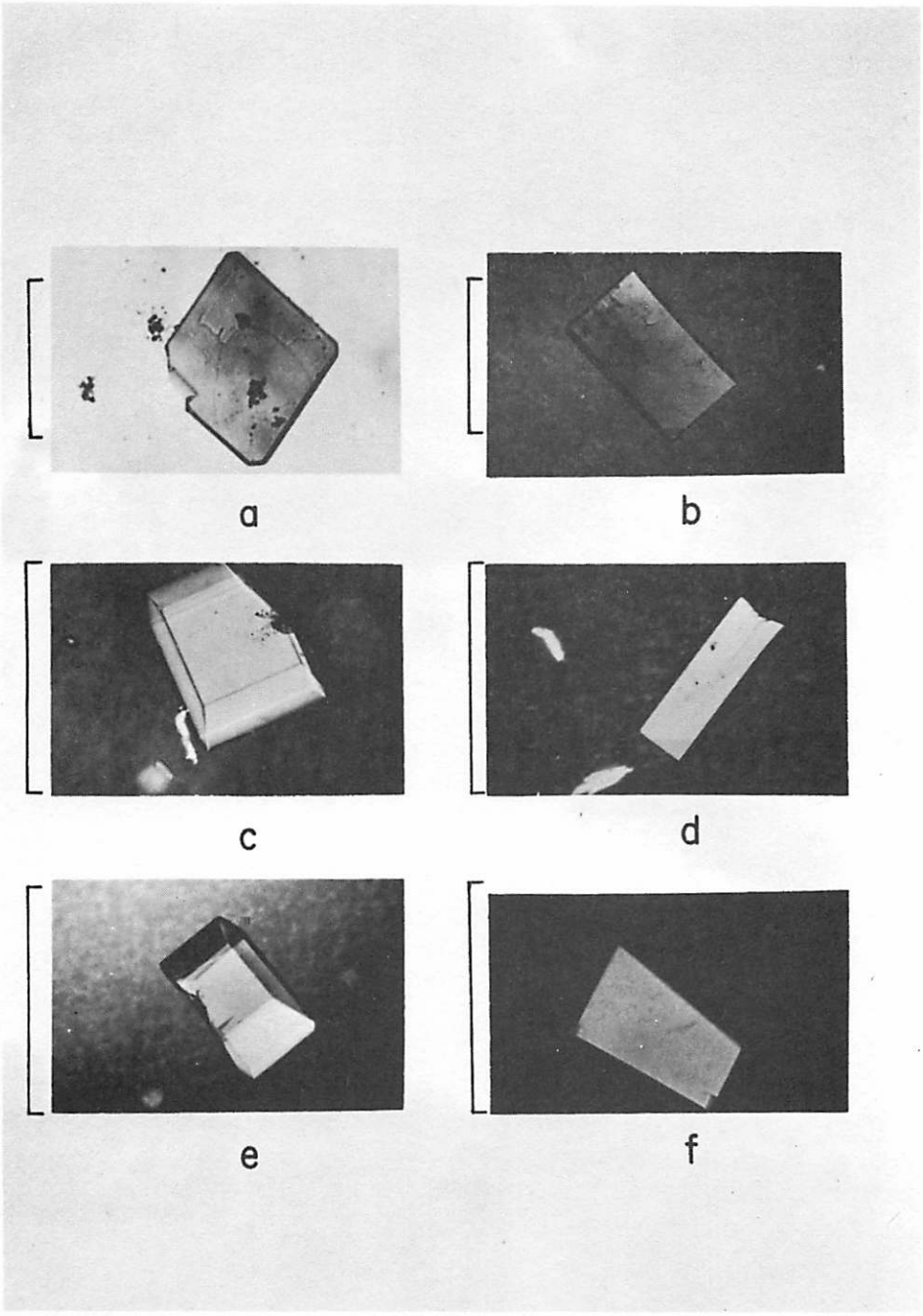
c) Cytochrome c' (Rps. palustris), Form 2

d) Cytochrome c' (R. rubrum), Form 1

e) Cytochrome c' (R. rubrum), Form 2

f) Glyceraldehyde 3-phosphate dehydrogenase (Palinurus)

Details of crystallizing conditions are given in text.


a
b
c
d
e
f

REFERENCES

1. Zeppezauer, M., Eklun, H., and Zeppezauer, E. S., Arch. Biochem. Biophys., 126, 565 (1968).

2. Weber, B. H., and Goodkin, P. E., Arch. Biochem. Biophys., 141, 489 (1970).

3. Zeppezauer, M., in "Methods in Enzymology" (W. B. Jacoby, ed.), Vol. 22, p. 253. Academic Press, New York (1971).

4. Davies, D. R., and Segal, D. M., in "Methods in Enzymology" (W. B. Jacoby, ed.), Vol. 22, p. 266. Academic Press, New York (1971).

5. Dixon, M., Biochem. J., 54, 457 (1953).

6. Horio, T., and Kamen, M. D., Biochim. Biophys. Acta., 48, 266 (1961).

7. Bartsch, R. G., in "Methods in Enzymology" (A. San Pietro, ed.), Vol. 23, p. 344. Academic Press, New York (1971).

8. Allison, W. S., and Kaplan, N. O., J. Biol. Chem., 239, 2140 (1964).

9. Jacoby, William B., in "Methods in Enzymology" (W. B. Jacoby, ed.), Vol. 22, p. 248. Academic Press, New York (1971).

10. Dickerson, R. E., Takano, T., Eisenberg, D., Kallai, O. B., Samsun, L., Cooper, A., and Margoliash, E., J. Biol. Chem., 246, 1151 (1971).

11. Eisenberg, D., in "The Enzymes" (Paul D. Boyer, ed.), 3rd edition, Vol. I, p. 1. Academic Press, New York (1970).

CHAPTER II

THE HIGH RESOLUTION CRYSTAL STRUCTURE DETERMINATION OF CYTOCHROME c_2 FROM RHODOSPIRILLUM RUBRUM

INTRODUCTION

Cytochrome c_2 of Rhodospirillum rubrum was one of the first c-type cytochromes characterized occurring in a procaryotic organism (1). The structural determination of this protein is of considerable interest in the investigation of morphological and functional evolution of c-type cytochromes (2), since substantial sequence homology with cytochrome c of eucaryotic organisms is known to exist (3).

R. rubrum is a purple non-sulfur photosynthetic bacterium commonly found in the soil. This organism can live anaerobically as a photoheterotroph dependent upon a cyclic photophosphorylating electron transport chain for the production of ATP. Alternatively, this organism may live in the dark in the presence of oxygen and reduced organic substrates, utilizing an oxidative electron transport chain for the production of ATP (Fig. 1). Considerable controversy exists concerning the details of the interactions of these electron transport systems, particularly with respect to the mechanism of production of reduced pyridine nucleotides (4). Cells of R. rubrum grown aerobically in the dark or anaerobically in the light show considerable differences in their membrane structure (which is known to be the location of the molecular arrays requisite for the generation of ATP in both electron transport systems). The organism in either case produces large amounts of cytochrome c_2, approximately

90% of which is soluble in the cytoplasm. The remainder is firmly bound in the membrane complex where it may play a role in either oxidative or photosynthetic electron transport. Although all of the specific functions of this protein have yet to be unequivocally established, it is believed to function as the primary donor to bacteriochlorophyll in the photosynthetic electron transport chain.

The protein consists of a single polypeptide chain of 112 residues and contains a single covalently bonded protoheme IX. The calculated molecular weight of cytochrome $\underline{c}_2$ is 12,480 daltons. The $E_{m,7}$ of the protein is +320 millivolts. Its isoionic point (pI) is 6.2. The amino acid sequence of the cytochrome $\underline{c}_2$ as determined by Dus, Sletten, and Kamen (5) is shown in Fig. 2. Also shown is the proposed sequence homology with horse heart mitochondrial cytochrome $\underline{c}$ whose detailed x-ray structure determination has recently been completed by R. Dickerson and coworkers (6). The detailed functional and structural comparison of R. rubrum cytochrome $\underline{c}_2$ and horse heart cytochrome $\underline{c}$ are the subject of further papers of this series. It is the purpose of this communication to describe the crystallographic structure determination of oxidized R. rubrum cytochrome $\underline{c}_2$ determined at a nominal resolution of 2 Å.

PROTEIN CRYSTALLIZATION AND ISOMORPHIC DERIVATIVE PREPARATION

Cytochrome $\underline{c}_2$ was isolated from anaerobically grown photosynthetic cultures of R. rubrum by a modification of the method of Horio and Kamen (7). Although cytochrome $\underline{c}_2$ is isolated from the bacterium in the reduced form (Fe^{II}), upon prolonged contact with air in the absence of reducing agents, it is approximately 95% converted to the oxidized form (Fe^{III}). Suspensions of oxidized cytochrome $\underline{c}_2$ microcrystals precipitated from 90% saturated ammonium sulfate (pH 5.8) were centrifuged, and the pellet dissolved with cold distilled water to obtain a concentrated (30 mg/ml) protein solution for crystallization, containing about 5 to 10% residual ammonium sulfate. Crystals were grown in small test tubes (7 × 20 mm) by a free interface diffusion technique (8). Crystals were initially grown by layering 50 µl of concentrated protein solution over 100 µl of 80% saturated (25°C) ammonium sulfate at 23°C. Crystals suitable for x-ray work (average dimension ≃ 0.3 mm) grew in these tubes within one to two weeks. The 23°C form of oxidized cytochrome $\underline{c}_2$ crystallizes as flat rhombohedral plates.

Attempts to form isomorphic derivatives of these protein crystals were carried out by soaking a few crystals in a 10- to 100-fold molar excess of heavy atom reagent in ammonium sulfate, buffered at various pH's. After one to two weeks of soaking, a representative crystal was mounted in a Lindemann glass capillary in the

conventional manner and 21° precession photographs of the okl or hol reciprocal lattice nets were obtained. The base plane projection photographs of potential isomorphic derivatives were compared with analogous projections obtained from the native parent protein crystals in order to ascertain whether or not a derivative had been formed. In approximately 200 attempts, only one heavy atom isomorphic derivative of the 23° form of these crystals, $AuCl_4$, giving an interpretable 5 Å resolution difference Patterson was found. Transfer of these crystals to mixed potassium phosphate/ammonium sulfate solutions yielded two additional derivatives, $Pt(CNS)_4$ and $OsNH_4I$. Derivatives formed in this way were not suitable for high resolution structural work, however, since crystals stored in the potassium phosphate/ ammonium sulfate solution appeared to be disordered as evidenced by a rapid fall-off of diffracted intensity on the precession films beyond a reciprocal lattice spacing corresponding to 3.5 Å resolution.

Further crystallization experiments with the native protein yielded an additional crystal form grown by layering 50 µl of oxidized cytochrome c_2 solution over 50 µl of 62% saturated unbuffered ammonium sulfate (pH 5.8) at 37°C. These crystals appear as slightly truncated flat rhombohedral plates and were isomorphic with the 23°C crystal form. Attempts to prepare heavy-atom isomorphic derivatives with these crystals gave three derivatives which yielded

clearly interpretable 5 Å difference Pattersons. These derivatives were $UO_2(NO_3)_2$, $IrNH_4I$ (in 0.1 m KSCN) and $OsNH_4I$ (in 0.1 m KSCN).

X-RAY CRYSTALLOGRAPHIC DATA

Cytochrome c_2 crystallizes in the orthorhombic space group $P2_12_12_1$ and has unit cell dimensions a = 32.33, b = 37.36, and c = 84.62 Å. The unit cell contains four molecules, with one molecule per crystallographic asymmetric unit. These crystals diffract quite well, showing fairly strong reflections out to a Bragg spacing of at least 2 Å.

Parent x-ray intensity data was collected both by screenless precession photography (9) and by the use of a Hilger and Watts four-circle automatic diffractometer.

Parent film data collected by the screenless precession technique required mounting two crystals for each data set taken. Sequential 2.75° precession photographs were taken at consecutive precession camera spindle settings of 0, 2.5, 7.5, 12.5, 17.5, and 22.5° for each of four possible crystal orientations beginning with the incident x-ray beam normal to the a*c* or b*c* plane, with c* either parallel or perpendicular to the spindle axis of the camera. The films were scanned and the integrated intensities corrected for Lorentz and polarization effects, and scaled as described previously (10). Each crystal received about 25 hours of x-irradiation in the course of

collection of a single replicated set of data. Comparison of intensities of common reflections scanned at the beginning and end of data collection by this technique indicated that essentially no degradation of crystal diffracting power took place over the time course of data collection.

Diffractometer data was collected by the peak scanning technique described by Wyckoff (11). Typically, counts were accumulated for thirteen 2.5 sec intervals on 0.01° increments of ω and θ. The maximum sum of four contiguous counts was considered proportional to the reflection intensity. Background corrections were applied from an empirically determined curve of diffuse background scatter versus 2θ for each crystal. An absorption correction curve (12) was also empirically determined for each crystal. Full sets of replicated data containing about 12,000 or 22,000 observations to 2.5 Å or 2 Å resolution respectively were taken on individual crystals. The maximum duration of x-ray exposure during data collection was about 200h for a replicated 2 Å data set. A small set of reflections was continuously monitored at intervals throughout data collection. A 15% decrease in intensity of these reflections due to crystal damage was considered the maximum allowable.

Four sets of parent intensities obtained from precession films (29,345 observations of 4,676 reflections) scaled together with an overall residual R-factor of 8.8%. Two sets of 2 Å diffractometer

data (32, 679 observations of 13, 269 observed reflections) scaled together with an R-factor of 5. 1%. The overall scaling R-factor for combined parent data (67, 024 observations of 13, 269 reflections) was 8. 6%. Observations of all centric (h, k, or l = 0), space group (hkl) and Bivjoet related reflections ($\bar{h}\bar{k}\bar{l}$) to 2. 5 Å resolution were replicated at least four times. Data from 2. 0 to 2. 5 Å resolution were replicated at least twice.

The four heavy atom derivatives used in the structure refinement were isomorphic with the parent crystals to within 0. 8 percent on all unit cell axes. Prior to the collection of high resolution data, heavy atom derivative crystals were checked by collecting a preliminary 5 Å data set and calculating a difference Patterson. All isomorphic derivative data used in the structural refinement was collected on the automatic diffractometer. Data on the gold chloride, and uranyl nitrate derivates was collected to 2 Å resolution and replicated twice. Data on the Osmium and Iridium Ammonium Iodides was collected to 2. 5 Å resolution and also replicated twice. Overall residual scaling R-factors for the derivatives ranged from 6 to 9%.

PHASE REFINEMENT

In the initial stages of phase refinement, intensities for hkl and $\bar{h}\bar{k}\bar{l}$ Bivjoet related reflections were averaged for phase calculations. Initial refinement cycles were carried out using only data to

2.5 Å resolution by the usual method of minimizing $\Sigma [k' F_H(OBS) - F_H(CALC)]^2$ in alternating cycles of phase calculation and refinement of parameters. The final cycles of phase refinement included all data to 2 Å resolution, since 2.0-2.5Å shell difference pattersons of the gold chloride and uranyl nitrate derivatives showed peaks at all heavy atom sites 2 to 16 times above background. The iridium derivative was eliminated from the refinement process, since its sites appeared nearly identical with those occupied by the osmium derivative. The final phases were refined by a program (13) which calculated centroid phases for both space group and Bivjoet related reflections. This was effected by treating each F_{hkl} and the complex conjugate of its Bivjoet related reflection $[F_{\bar{h}\bar{k}\bar{l}}]^*$ as independent reflections in the calculation of the centroid phases. This program also calculates a least-squares value of the anomalous scale constant, which is the percentage of total heavy atom electrons which anomalously scatter x-rays. The behavior of the mean figure of merit as a function of $\sin\theta/\lambda$ is shown in Fig. 3. Isomorphic derivative statistics from the last cycle of refinement are shown in Table 1.

INTERPRETATION OF THE ELECTRON DENSITY MAP

The correct structural enantiomorph was determined by calculation of a parent anomalous dispersion Bivjoet difference fourier. This map showed a single positive peak six times greater than background at the position of the heme iron determined from previous 5 Å

resolution studies (14) indicating that the proper enantiomorph had been chosen. A small preliminary map on a scale of 0.163 inches/Å was contoured on 8.5 × 11" plexiglass sheets to allow the limits of the model to be determined prior to production of a large scale map.

A 2 cm/Å map (1 Å = 0.7874") was calculated on a 4687 × 32 × 72 grid from centroid phases, but without centroid weighting factors. Forty-eight sections from x = 2300 to 6900 normal to the x axis were plotted directly on 31" square 0.005" transparent acetate plastic sheets on a CalComp flatbed plotter. The spacing between sections corresponds to 0.87 Å on the 2 cm/Å scale of the map. The map was contoured at a minimum level of 0.27 e/Å^3 and every 0.18 e/Å^3 thereafter. The acetate map sections were fixed to aluminum screen frames and inserted into a horizontal Richards optical comparitor (15).

Model construction was begun at the heme, since the positions of the heme iron atom and cysteinyl sulfur atoms of residues 14 and 17 appeared as prominent features in the map. The chain was subsequently built according to the published sequence from residue 14 to the amino terminus and from residue 17 to the carboxyl terminus of the chain. There appeared to be no breaks in the backbone chain density. The electron density was well defined for the majority of backbone carbonyl groups and most of the internal amino acid side chains of the structure. Planar aromatic residues were generally well defined in the map allowing unambiguous orientation of the ring

planes in the map density. External residues were generally less well defined in the map than those of the interior of the molecule. Density corresponding to lysine residues in most cases faded appreciably beyond the β or γ side chain carbon atom, although in some cases these residues appeared to have branched density indicating two or more alternative conformations in the crystal lattice. The overall quality of the map appeared to be poorer in the section of chain from residue 71 through 95. This effect may be due to greater flexibility of this part of the chain in the crystal lattice, or to static variations in the molecular conformation of this region.

The following structural analysis is subject to the usual caveats with respect to the interpretation of electron density, especially in regions which are somewhat ill defined in the map. It is not likely, however, that the refinement procedures which are currently in progress to improve the quality of this map will significantly alter the description of the molecule presented in this paper.

GENERAL DESCRIPTION OF THE MOLECULAR CONFIGURATION

The overall shape of the molecule is a roughly prolate ellipsoid of dimensions 25 × 33 × 40 Å. The heme group lies perpendicular to the surface of the molecule with one edge exposed to the external solvent environment. The ϵ^1 imidazole nitrogen of histidine 18 and the sulfur of methionine 91 fill the fifth and sixth coordination positions of the heme iron atom.

As expected from previous x-ray crystallographic studies of proteins, the interior of the molecule is occupied principally by uncharged hydrophobic amino acid residues or polar residues which form hydrogen bonds. The exterior of the molecule in contact with the solvent is populated by the side chains of residues which are charged at physiological pH. Some notable exceptions to these general characteristics of protein structure are to be found in the structure, however. The possible importance of these exceptional residues to the mechanism of action of this protein are discussed below.

Figure 4 shows the backbone α-carbon atom chain configuration of the oxidized cytochrome c_2 molecule. The molecule contains three segments of α helix: from residues 2 through 10 at the amino terminus, from residues 64 through 70 in the middle of the chain, and from residues 96 through 108 near the carboxyl terminus of the chain. The remainder of the chain consists of meandering extended polypeptide chain forming several hairpin loops but without obvious pleated sheet β structure. Figure 5 is a schematic drawing of the polypeptide chain showing the relative location of helical regions in the sequence and the principal interactions stabilizing the heme.

THE HEME ENVIRONMENT

The heme is covalently bonded to the backbone polypeptide chain by thioether linkages with cysteine residues 14 and 17. The ε imidazole nitrogen of histidine 18 forms a coordinate bond with the

heme iron in the fifth coordination position. The imidazole δ nitrogen of histidine 18 forms a hydrogen bond with the carbonyl oxygen atom of proline 30. These interactions, in addition to the $C_\alpha - C_\beta$ bond of the residue, fix the orientation of the imidazole ring plane normal to the heme plane and intersecting it on a line defined by the iron atom and the methylene carbon atom bridging the pyrole rings connected to the polypeptide backbone at residues 14 and 17. The heme propionic acid side chains form hydrogen bonds with tryrosines 46 and 48, serine 49, and tryptophane 62. The tyrosine hydroxyl of residue 46 appears to form an additional hydrogen bond with the carbonyl oxygen of valine 28. Likewise, the possibility exists of further hydrogen bonding of serine 49 to the carbonyl of lysine 88.

The sixth iron coordination position is filled by the sulfur atom of methionine 91. The bond between the sulfur atom of methionine 91 and the heme iron does not appear from the electron density to be strictly coaxial with the iron-ε nitrogen bond to histidine 18. The sulfur atom appears to be displaced about 0.5 Å off axis. This feature can be clearly seen in Fig. 6 which is a section of the electron density normal to the heme plane, in the plane of the nitrogen and sulfur ligands. The possible relevance of this feature to the mechanism of oxidation and reduction of this protein is discussed below.

The residues bounding the interior of the heme crevice, in addition to those already described which form the lattice of bonds

rigidly supporting the heme, are valine 10, phenylalanine 20, glycine 29, proline 30, leucine 32, valine 35, alanines 40 and 41, histidine 42, tyrosine 52, leucine 67, tyrosine 70, valine 71, serine 89, phenylalanine 93, leucine 95, asparagine 103, valine 104, and tyrosine 107 (Figs. 5 and 7). While it is unlikely that any of these residues are charged in the apolar environment of the heme crevice, it is clear that a large number of them are polar in nature. Of particular note is the fact that all five of the tyrosine residues of the molecule appear as nearest neighbors to the heme group. Tyrosines 46 and 48 are hydrogen bonded to the heme propionic acid side chains as described previously. Tyrosine 70 is oriented such that its hydroxyl oxygen is pointed roughly at the sulfur atom of methionine 91. Tyrosine 52 lies approximately 5.5 Å from the heme plane on the methionine ligand side and is roughly coplanar with the heme ring. Tyrosine 107 lies roughly normal to the heme plane at a distance of 6 Å at the upper rear of the heme crevice but does not appear to interact with any proximal chemical groups. (Distances are measured from the center of the aromatic ring to the closest point on the heme plane.)

The molecule contains five phenylalanine residues located in the sequence in positions 20, 33, 36, 77, and 93.. Phenylalanine 20 lies approximately ($\phi \simeq 20°$), coplanar with the ring of tyrosine 107, at a distance of 6 Å. This residue lies at the right rear of the heme crevice, 9 Å from the heme plane, and is accessible from the right

side of the exterior of the molecule. Phenylalanine residues 33 and 36 are both more than 12 Å distant from the heme and lie partially buried on the exterior of the molecule, forming no obvious interactions with the heme or residues associated with it. Phenylalanine 77 lies 10 Å from the heme plane and is coplanar with the ring of tryptophane 62. The rings of these residues appear to overlap edge to edge at an interplanar distance of 3.5 Å. The benzene ring of phenylalanine 93 lies in Van Der Walls contact with the heme at the upper front of the heme crevice. The plane of the benzene ring makes about a 45° angle with the heme plane and blocks access to the iron by any external solvent ions.

THE RIGHT CHAIN

The first 62 residues of the polypeptide chain form the right side, bottom, and rear of the heme crevice. This section of the sequence forms essentially a single layer of polypeptide chain consisting of the N-terminal α-helix and succeeding extended chain residues which form the lattice of bonds supporting the heme. The rigidity of this section of the chain is implicated by the relative definition of the electron density map in this region as compared to the less well defined electron density characteristic of the extended polypeptide chain in the remainder of the molecule.

Several external hairpin loops are stabilized by 3_{10} hydrogen bonds. These occur from the carbonyl oxygen of aspartic acid 21 to

the amide nitrogen of glycine 24, from the carbonyl of leucine 32 to the amide of valine 35, from the carbonyl of valine 35 to the amide of asparagine 38, from the carbonyl of glutamic acid 54 to the amide nitrogen of alanine 57, and from the carbonyl oxygen of glutamic acid 50 to the amide nitrogen of threonine 53. Additional hydrogen bonds occur between the side chain amides of residues 26 and 45, and from the hydroxyl oxygen of threonine 19 to the carbonyl oxygen of alanine 23. The ring of proline 30, whose carbonyl oxygen forms a hydrogen bond to the imidazole of histidine 18, lies directly on top of the ring of tyrosine 46, which is in turn hydrogen bonded to a heme propionic acid side chain and the carbonyl oxygen of valine 28 as described above.

THE LEFT CHAIN

Residues 64 through 70 form a segment of regular α-helix which forms the upper left side of the heme crevice. The chain subsequently winds back and forth on the left side of the molecule forming the lower left side of the heme crevice and finally ascends to the carboxyl terminus α-helix whose internal residues line the top of the heme crevice. The extended chain configuration of residues 71 through 96 forms several hairpin loops. One such loop is stabilized by a 3_{10} hydrogen bond from the carbonyl oxygen of lysine 75 to the amide nitrogen of valine 78. Two additional tight exterior chain reversals

occur at proline residues 74 and 85, both of which are in the trans-configuration. Further hydrogen bonding interactions appear probable between the hydroxyl group of threonine 63 and the side chain amide of asparagine 66, and from the hydroxyl group of serine 82 to the amide of aspartic acid 84.

The structural rationale incorporated into the extended chain conformation on the left side of the molecule is not as readily apparent as that encountered on the right side. In addition to the several strictly hydrophobic residues which occupy the interior of the left side of the molecule, there appear several polar residues. Most notable are tyrosine residues 52 and 70, and serine residues 51, 82 and 89. Tyrosine 70, as described previously, is a nearest neighbor to the heme plane and lies with its phenoxyl group pointed at the sulfur atom of methionine 91. Tyrosine 52 is also a nearest neighbor of the heme on the left side of the heme crevice. Its hydroxyl oxygen appears to form a bifurcated hydrogen bond with tyrosine 70 and serine 89. Serine 89 is half buried at the surface of the molecule on the left exterior of the heme crevice. Serine 51 lies in a shielded loop of polypeptide chain on the left side of the molecule to the left and below tyrosine 52. The hydroxyl oxygen of this residue is within hydrogen bonding distance to the hydroxyls of both serine 49, which is in turn hydrogen bonded to a heme proprionic acid side chain, and serine 83, which is situated at the left exterior of the molecule. Whether these hydrogen bond

networks incorporating buried polar residues merely fulfill a structural role or are somehow involved in the oxidation-reduction mechanism of this protein awaits comparison of this structure with that of the reduced form of cytochrome $\underline{c}_2$. The structure determination of reduced cytochrome $\underline{c}_2$ molecule by difference Fourier techniques is presently in progress. However, some speculations on the possible functional role of these interactions are considered below in the context of the discussion of the interactions of cytochrome $\underline{c}_2$ with other molecules participating in the photosynthetic electron transport chain.

HEAVY ATOM BINDING SITES

All of the principal heavy-atom binding sites to the cytochrome $\underline{c}_2$ molecule in the crystal lattice occur on the back side (relative to the heme crevice) of the molecule. Inspection of the electron density distribution in the unit cell shows as expected, that this part of the molecule forms the border of the largest solvent channel in the crystal lattice.

The principal osmium binding site occurs in the vicinity of the carboxylate group of glutamic acid 37 and the ammonium group of lysine 112. This position is also a secondary site for the uranyl nitrate derivative. Secondary osmium binding sites occur in the vicinity of the ammonium side chain groups of lysines 56 and 109. Lysine 56 is adjacent to methionine 55 which also could be involved

in the osmium binding site, although failure to form a mersalyl derivative at this site suggests that the methionine sulfur is relatively inaccessible to the external solvent environment, despite its only partially buried configuration in the structure.

The principal uranium binding sites occur near the carboxylate group of glutamic acid 64 and the hydroxyl group of threonine 63. A secondary site in addition to the one occupied in common with the osmium derivative is located near the partially exposed isoleucine 101 side chain in the vicinity of the ammonium group of lysine 97.

The most heavily occupied gold chloride binding site is adjacent to the imidazole ring of histidine 42 and the ammonium group of lysine 41. The small secondary gold binding site lies in the vicinity of the carboxylate group of aspartic acid 3.

SPECULATIONS ON THE MECHANISM OF CHAIN FOLDING OF CYTOCHROME $\underline{c}_2$

The backbone chain configuration of the molecule can be roughly considered to consist of three sections.

The first section consists of the amino terminus α-helix and succeeding extensively hydrogen bonded and tightly looped extended chain, through residue 62. This segment of the chain forms the right side of the molecule and furnishes all of the bonded interactions to the heme with the exception of the iron-methionine 91 sulfur bond.

The second section of the chain commences with the short

α-helix at residue 64 which bounds the upper left surface of the heme crevice. Succeeding residues form the looped extended chain configuration comprising the left side of the molecule. The structure in this region is also extensively hydrogen bonded, although the principal interactions with the heme appear to be hydrophobic in nature. The chain configuration of this region shields the internal polar residues tyrosines 52 and 70, both of which are nearest neighbors to the heme on the left side. The proximity of these residues to the heme, their participation in a hydrogen bond network to surface polar residues, and the apparent interaction between the hydroxyl of tyrosine 70 and the sulfur of methionine 91, implicate the importance of these residues for the mechanism of electron transport in this cytochrome.

The final section of the chain consists of the carboxyl terminus α-helix whose internal residues bound the top of the heme crevice, and whose interactions with the α-terminal helix and the remainder of the molecule appear to be essentially hydrophobic in nature, with the possible exception of a hydrogen bond between the serine hydroxyl of residue 11 and the side chain amide of asparagine 103. The polar residue tyrosine 107 is buried at the upper rear of the heme crevice by the "x" configuration formed by the descending amino terminus helix and the ascending carboxy terminus helix.

In summary, the internal structural integrity of each section appears to be principally dependent upon the formation of hydrogen

bonded interactions. Interaction between the segments are primarily stabilized by covalent bonds or hydrophobic interactions with the heme.

These observations tempt speculation upon the possible mechanism of chain folding in this molecule. Lewis and Sheraga (16) have predicted the α-helical content of several homologous eucaryotic cytochromes c, based on the assumption that the short range interactions which elicit stabilization of an α-helical configuration of a polypeptide chain in solution predominate in establishing the polypeptide chain configuration in a globular protein. As can be seen from Fig. 1, predictions of α-helical regions in cytochrome c_2 appear to correlate well with the observed helical region of the molecule, with the exception of the predicted α-helical region at residues 76 through 81. In this context it is perhaps not too presumptious to speculate that the initial event in the folding of the cytochrome c_2 molecule is the spontaneous condensation of the helical regions of the chain. As has been noted previously, the sequence Cys-X-X-Cys-His, if placed in an α-helical configuration establishes the proper stereochemical relationships to allow both condensation of the protoheme IX vinyl groups with the cysteine side chains, and ligation at the fifth heme coordination position by the ε-imidazole nitrogen atom of histidine 18. The second step in the postulated folding of the molecule would therefore involve attachment of the heme and histidine coordination to the heme iron atom. The next step would involve formation of the hydrogen

bonds to the heme propionic acid side chains by tyrosines 46 and 48, serine 49, and perhaps a main chain amide or carbonyl at residue 40-42, or the ring nitrogen of tryptophane 62. These interactions would appear stereochemically favorable due to the relatively fixed stereochemistry of the heme propionic acid side chains relative to the proximal location in the sequence of the polar residues forming the hydrogen bonds holding the heme. Once these constraints have been established, structuring of the extended chain on the right side of the molecule could ensue, burying phenylalanine 20 and leucine 32 next to the heme, stacking proline 30 on top of the ring of tyrosine 46, and forming the hydrogen bonds stabilizing the lower right side of the molecule.

The next step would involve the association of the three α-helical sections of the chain to form the top of the molecule, primarily by hydrophobic interaction between themselves and the heme. This step necessitates folding the amino terminus α-helix back upon itself at residue 11, breaking the α-helical hydrogen bonds in that region. This feature is observed in the cytochrome c_2 molecular structure as manifest by the hydrogen bonding pattern in the sequence from valine 10 to histidine 18 and fortuitously satisfies the discrepancy between the observed and predicted helicity of this region.

The remainder of the molecular folding is most likely driven by the tendency to form favorable hydrophobic interactions among

aromatic and aliphatic side chains of residues comprising the extended chain of the left side of the molecule. Notable is the stacking interaction of phenylalanine 77 with tryptophane 62 and the tucking of phenylalanine 93 into the heme crevice. Concomitant with the attainment of this configuration is the formation of the iron-methionine 91 sulfur bond.

Examination of the predicted α-helical region at residues 76 through 81 shows this region to be a sequence of tight loops, which are not, however, obviously derivative of α-helix. As noted above, the ring of phenylalanine 77, which is stacked with tryptophane 62 and the side chain of valine 78 form the hydrophobic environment for tyorsine residues 52 and 70. It is possible that strong side-chain hydrophobic interactions at this location have distorted this potentially helical region of the chain.

INTERACTIONS OF CYTOCHROME $\underline{c}_2$ WITH CELLULAR REDOX COMPONENTS

The principle, or at least most unambiguous, function served by cytochrome $\underline{c}_2$ in the living organism is the donation of an electron from the reduced form of the molecule (FeII) to an excited membrane-bound positively charged bacteriochlorophyll molecule. This electron is subsequently boosted in potential by the primary photoact in photosynthetic cyclic electron transport. The question immediately posed is the location on the molecular surface of the site of interaction

between cytochrome $\underline{c}_2$, which may or may not be bound in the membrane complex (17), and the excited bacteriochlorophyll molecule.

One of the most striking structural features of the cytochrome $\underline{c}_2$ molecule is the preponderance of bonded interactions between the polypeptide backbone and the heme. These interactions rigidly fix the heme in what initially appears an unnatural configuration with the hydrophobic edge of the ring exposed to the solvent and the hydrophilic propionic acid chains firmly hydrogen bonded to several internal polar groups of the molecule. The situation of the heme in this molecule differs from that found in the oxygen binding heme proteins (18, 19) and calf liver cytochrome $\underline{b}_5$ (20) which contain non-covalently bonded protoheme IX and whose propionic acid groups extend into the external solvent.

The exposure of the hydrophobic edge of the heme at the molecular surface suggests that this is the site of interaction with the membrane-bound bacteriochlorophyll molecule. Presumably this interaction would take place by partial intercalation of the chlorophyll and cytochrome porphyrin rings. Support for this hypothesis is inferred by consideration of the properties of the external amino acid side chains bounding the perimeter of the heme crevice. Figure 8 shows the distribution of 11 of the 17 lysine residues of the molecule. It can be clearly seen that the external side chains of these residues completely bound the perimeter of the heme crevice, and that this

charge distribution is uninterrupted by the presence of any other external ionic side chains with the exception of aspartic acid residues 98 and 99 which are clustered at the very top exterior of the heme crevice.

If it is supposed that the bacteriochlorophyll molecule is situated in a phospholipid membrane, with its phytol side chain buried in the hydrophobic interior of the membrane leaving its Mg-porphyrin ring partially exposed, and that the membrane surface is populated by the negatively charged phosphate groups of the membrane phospholipid components, the rationale behind the clustering of the lysine side chains becomes clear, that is, ionic interactions between the positively charged lysine amino groups and the negatively charged membrane bound phosphate groups would stabilize the juxtaposition of the heme and Mg-porphyrin rings in the orientation requisite for efficient electron transport between these two molecules.

It is relevant to note here that approximately 10% of the total cytochrome $\underline{c}_2$ found in the organism is tightly bound in a membrane complex. The remainder circulates freely in the cytoplasm. It is not known whether the freely diffusing material reversibly binds to the membrane <u>in vivo</u>. It is apparent, however, that the free form of the protein and the membrane-bound form which has been solvent extracted are chemically and functionally identical (21).

Any speculation on the mechanism of the reduction of the

cytochrome $\underline{c}_2$ molecule in the organism centers upon the question of whether or not the molecule interacts with the reductant at the same or a different site with which it interacts with the oxidase (bacteriochlorophyll).

Relatively little is known about the mechanism of reduction of cytochrome $\underline{c}_2$ *in vivo*. The current data (22) indicates that either ubiquinone 50 or rhodoquinone are necessary for the enzymatic NADH-dependent reduction of cytochrome $\underline{c}_2$ in chromatotphore preparations. Specific interaction with an enzyme does not appear to be implicated since the NADH:heme protein oxidoreductase requisite for the reduction of cytochrome $\underline{c}_2$ is also capable of the reduction of several other cytochromes of the organism. This suggests that the specific chemical moiety responsible for the reduction of cytochrome $\underline{c}_2$ (if there is one) is some small freely diffusible redox component of intermediate potential such as ubiquinone.

Inspection of the remainder of molecular surface of the cytochrome $\underline{c}_2$ does not show any large scale clustering of either charged or polar groups suggestive of specific charge interactions with another macromolecule. The rings of phenylalanine residues 20, 36, and 77 are partially exposed at the surface of the molecule and are situated in hydrophobic niches. Coplanar interactions between the rings of phenylalanine 20 and tyrosine 107, and between phenylalanine 77 and tryptophane 62 could conceivably serve as electron conduction routes

from some external reductant (or oxidant) molecule. The ring of phenylalanine 36, which bounds the rear of the heme crevice, would have to move several angstroms for effective interaction with any groups (e. g., tyrosine 107) proximal to the heme. This appears unlikely since this conformational alteration would necessitate the breakage of a hairpin loop hydrogen bond. The last obvious possible site for reductant binding is the heme crevice. Electron transfer from the reductant molecule would in this case probably follow nearly the same path as that utilized in the withdrawal of the electron during oxidation by bacteriochlorophyll. Current conceptions (21) about the localization of membrane-bound redox components tend toward configurations in which the macromolecular components are rigidly held in the proper relative orientation to facilitate efficient electron transport. If the heme crevice served as the binding site for both the oxidant and reductant, it would appear that some gross molecular movement would be necessary during the process of electron transport. It is to be noted, however, that the exterior heme crevice of the cytochrome $\underline{c}_2$ molecule is nearly 20 Å in length and that alternative and perhaps even simultaneous binding of oxidizing and reducing molecular species may be possible, especially if the physiological reductant is in fact a small aromatic compound such as a quinone. Presumably such a small molecule reductant would be quite soluble and capable of free diffusion in the hydrophobic matrix of the membrane.

Verification of these speculations awaits further biochemical and crystallographic investigation of the binding of redox components to the cytochrome $\underline{c}_2$ molecule.

The specific path followed by the electron in the reversible oxidation and reduction of the heme iron is at present unknown. As inferred in the discussion of possible redox reagent binding sites, this path may differ in the mechanism of oxidation and reduction of the heme iron. The predominant structural feature of the molecule relevant to facilitation of a redox change in the iron atom is the proximity of the hydroxyl oxygen atom of tyrosine 67 to the methionine 91 sulfur atom coordinated to the iron in the sixth position. If the sulfur atom is assumed to have a tetrahedral sp^3 orbital configuration, one filled sp^3 orbital lobe interacts with the d_z^2 axial iron orbital. The remaining unbonded filled sp^3 orbital then appears to point directly coaxially at the filled sp^2 orbital of the hydroxyl oxygen of tyrosine 67. This tyrosine is potentially involved in a hydrogen bond network to the hydroxyl oxygen of tyorsine 52, which is in turn hydrogen bonded to serine 89. A small degree of torsion in the backbone chain at residue 91 would allow further or alternative hydrogen bonding of serine 89 to threonine 92. This region of the chain is apparently strained in any case, as evidenced by the non-axial bond from the sulfur of methionine 91 to the heme iron. Whether or not this configuration is static or dynamically involved in the oxidation and reduction mechanism awaits

examination of the reduced structure. At this stage it is not possible to infer whether this network serves to predispose the iron atom to a change in oxidation state by affecting its ligand field, or serves as the actual pathway for electron conduction.

CORRELATIONS WITH PHYSICAL AND CHEMICAL DATA

Bartsch and coworkers (17) have found that cytochrome c_2 isolated from photosynthetically grown cultures of R. rubrum is composed of eight distinct species having isoelectric points ranging from pI 9.3 to 4.1. These multiple isoelectric forms of the protein presumably arise from successive deamidation of glutamine or asparagine residues to form the corresponding carboxylic acids. They observed that the only form which was readily crystallized was a species having a pI of 6.2, which is the predominant species in the mixture isolated from the organism. Examination of cytochrome c_2 crystals used for crystal structure analysis by polyacrylamide gel electrophesis indicates that the crystals are composed essentially (> 95%) of the pI 6.2 form.

This was also the predominant form isolated in the extraction procedure utilized to obtain material for sequence determination (23). The question remains whether it is possible to rationalize the non-linear distribution of the various isoelectric species and their apparent inability to crystallize in terms of the structural features of the molecule.

side chain amides of residues 26, 31, and 45. Presumably hydrolysis of one of these amides would destabilize the remaining ones with respect to hydrolytic attack and hence account for the observed distribution of isoelectric species. It is not clear whether the effect of hydrolysis of the side chain amides on the ability of the protein to crystallize arises from solubility changes induced by ionization of external residues or by conformational disruption caused by the loss of structurally important hydrogen bonds.

Flatmark and Robinson (24) have studied the circular dichroic absorption spectra of cytochrome $\underline{c}_2$ in both the reduced and oxidized form. They calculated that the cytochrome $\underline{c}_2$ molecule contained approximately 36% α-helical structure based upon the ellipicity value at 222 mμ. The observed "good" α-helical content of the cytochrome $\underline{c}_2$ molecule is 26%. They did not observe any difference in α-helical content of the molecule upon oxidation and reduction of the molecule. This observation is consistent with the experimental observation that the cytochrome $\underline{c}_2$ molecule is essentially crystallographically isomorphic in the oxidized and reduced form, and can in fact be reversibly oxidized and reduced in the crystalline state.

FIGURE 1

Simplified schematic of electron transport system of the non-sulfur purple photosynthetic bacteria Rhodospirillus rubrum indicating possible roles of cytochrome c_2 in the system.

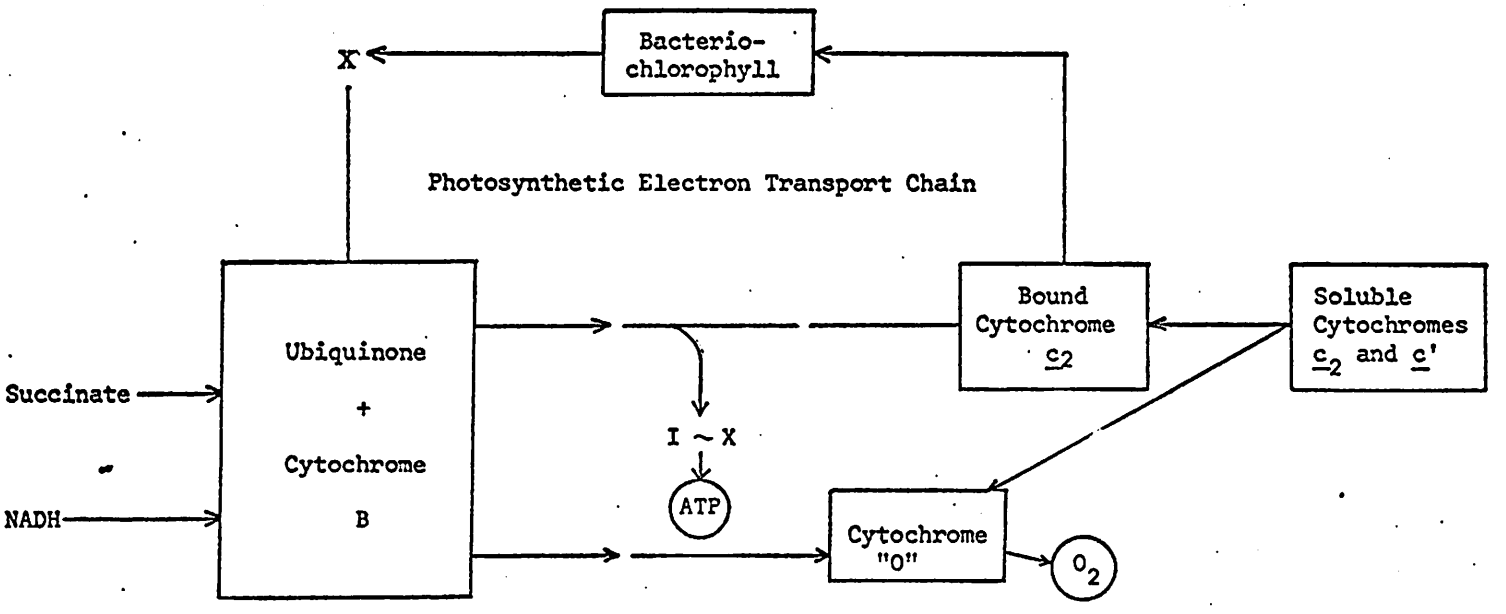

Bacterio-
chlorophyll
X
Photosynthetic Electron Transport Chain
Ubiquinone
+
Cytochrome
B
Succinate
NADH
Bound
Cytochrome
c2
Soluble
Cytochromes
c2 and c'
I ~ X
ATP
Cytochrome
"O"
O2

FIGURE 2

Sequence homology between cytochrome $\underline{c}_2$ of Rhodospirillum rubrum and mitochondrial horse heart cytochrome c (5). Solid bars above sequence indicate observed regions of α-helix in the cytochrome $\underline{c}_2$ molecule. Dashed bars indicate areas of predicted α-helix by the criteria of Lewis and Sheraga (16). Dot code: ● = helix maker, ◍ = helix indifferent, O = helix breaker.

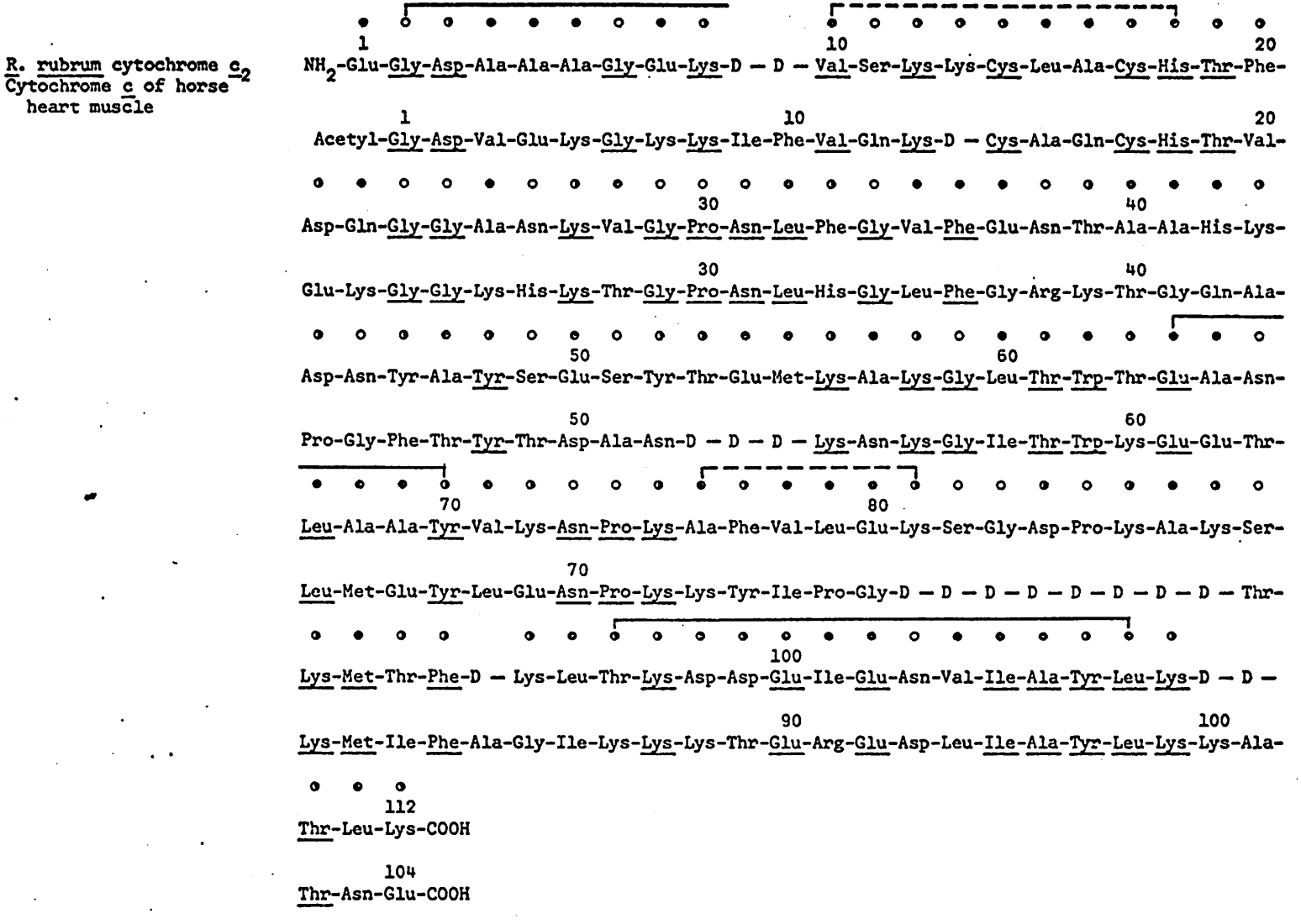

R. rubrum cytochrome c2
Cytochrome c of horse heart muscle
NH2-Glu-Gly-Asp-Ala-Ala-Ala-Gly-Glu-Lys-D — D — Val-Ser-Lys-Lys-Cys-Leu-Ala-Cys-His-Thr-Phe-
Acetyl-Gly-Asp-Val-Glu-Lys-Gly-Lys-Lys-Ile-Phe-Val-Gln-Lys-D — Cys-Ala-Gln-Cys-His-Thr-Val-
Asp-Gln-Gly-Gly-Ala-Asn-Lys-Val-Gly-Pro-Asn-Leu-Phe-Gly-Val-Phe-Glu-Asn-Thr-Ala-Ala-His-Lys-
Glu-Lys-Gly-Gly-Lys-His-Lys-Thr-Gly-Pro-Asn-Leu-His-Gly-Leu-Phe-Gly-Arg-Lys-Thr-Gly-Gln-Ala-
Asp-Asn-Tyr-Ala-Tyr-Ser-Glu-Ser-Tyr-Thr-Glu-Met-Lys-Ala-Lys-Gly-Leu-Thr-Trp-Thr-Glu-Ala-Asn-
Pro-Gly-Phe-Thr-Tyr-Thr-Asp-Ala-Asn-D — D — D — Lys-Asn-Lys-Gly-Ile-Thr-Trp-Lys-Glu-Glu-Thr-
Leu-Ala-Ala-Tyr-Val-Lys-Asn-Pro-Lys-Ala-Phe-Val-Leu-Glu-Lys-Ser-Gly-Asp-Pro-Lys-Ala-Lys-Ser-
Leu-Met-Glu-Tyr-Leu-Glu-Asn-Pro-Lys-Lys-Tyr-Ile-Pro-Gly-D — D — D — D — D — D — D — Thr-
Lys-Met-Thr-Phe-D — Lys-Leu-Thr-Lys-Asp-Asp-Glu-Ile-Glu-Asn-Val-Ile-Ala-Tyr-Leu-Lys-D — D —
Lys-Met-Ile-Phe-Ala-Gly-Ile-Lys-Lys-Lys-Thr-Glu-Arg-Glu-Asp-Leu-Ile-Ala-Tyr-Leu-Lys-Lys-Ala-
Thr-Leu-Lys-COOH
Thr-Asn-Glu-COOH

FIGURE 3

Behavior of the Figure of Merit ($\overline{m}$) as a function of $\sin\theta/\lambda$.

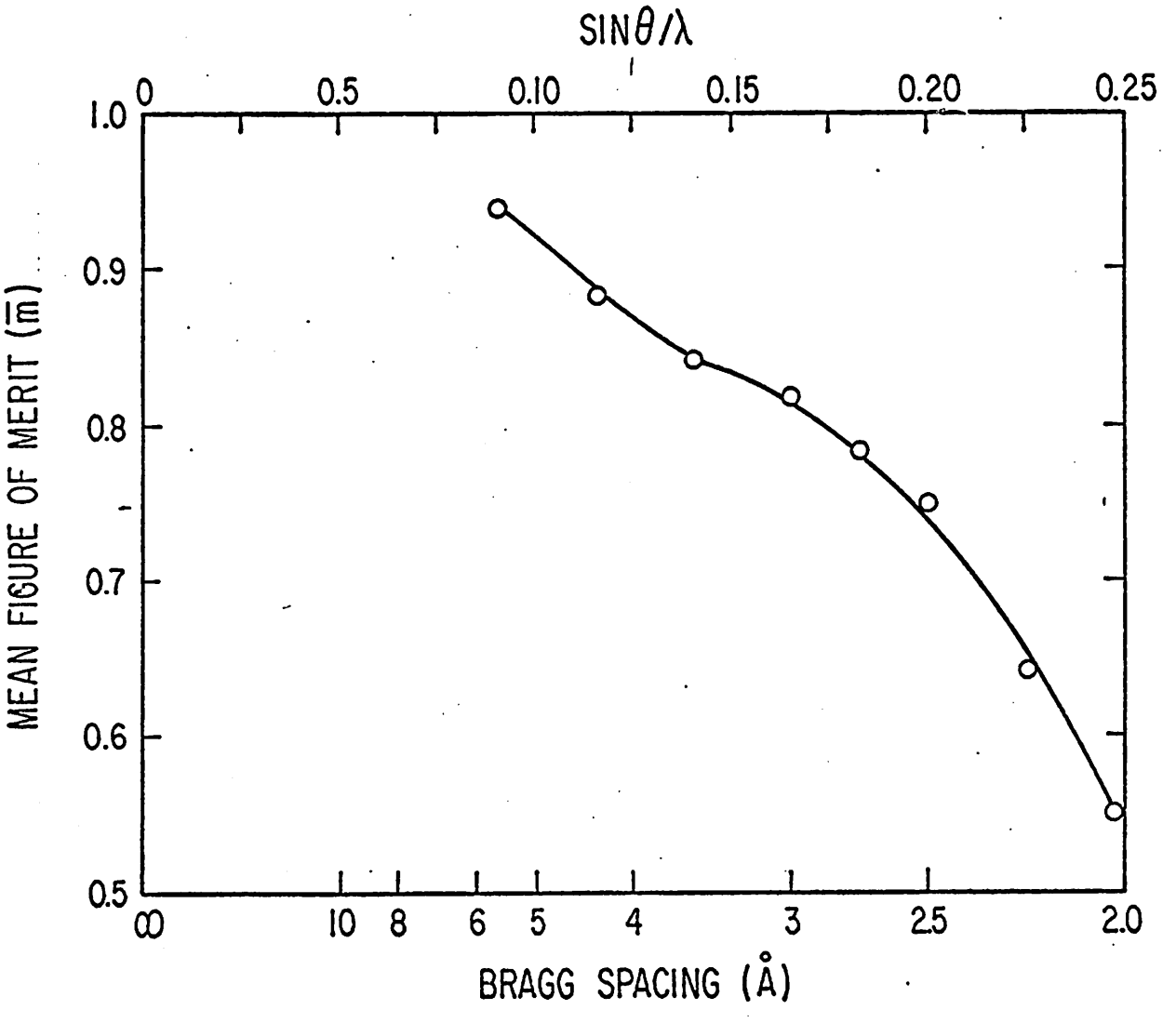

SIN θ/λ
0
0.5
0.10
0.15
0.20
0.25
1.0
0.9
0.8
0.7
0.6
0.5
MEAN FIGURE OF MERIT (m̄)
∞
10
8
6
5
4
3
2.5
2.0
BRAGG SPACING (Å)

FIGURE 4

Stereo photograph of the cytochrome c_2 molecule viewed from the 'front' (facing the heme crevice).

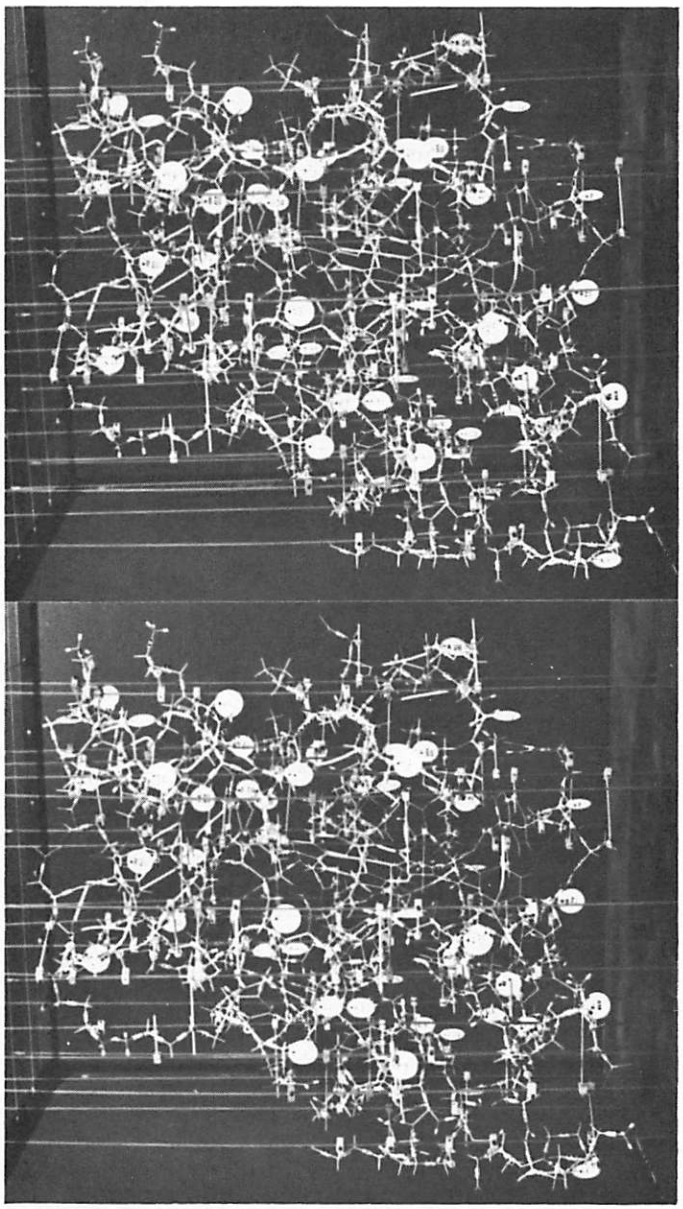


FIGURE 5

Schematic of the chain folding in the cytochrome $\underline{c}_2$ molecule. α-helical regions are indicated by folded ribbon. Blacked ribbon indicates residues which are hydrogen bonded to heme (- - -), or are near Van Der Waals contact with the heme porphyrin ring. Dashed ribbon indicates residues which bound the interior of the heme crevice. Dotted line (. . .) indicates hydrogen bonding pattern of polar residues proximal to the sulfur of methionine 91.

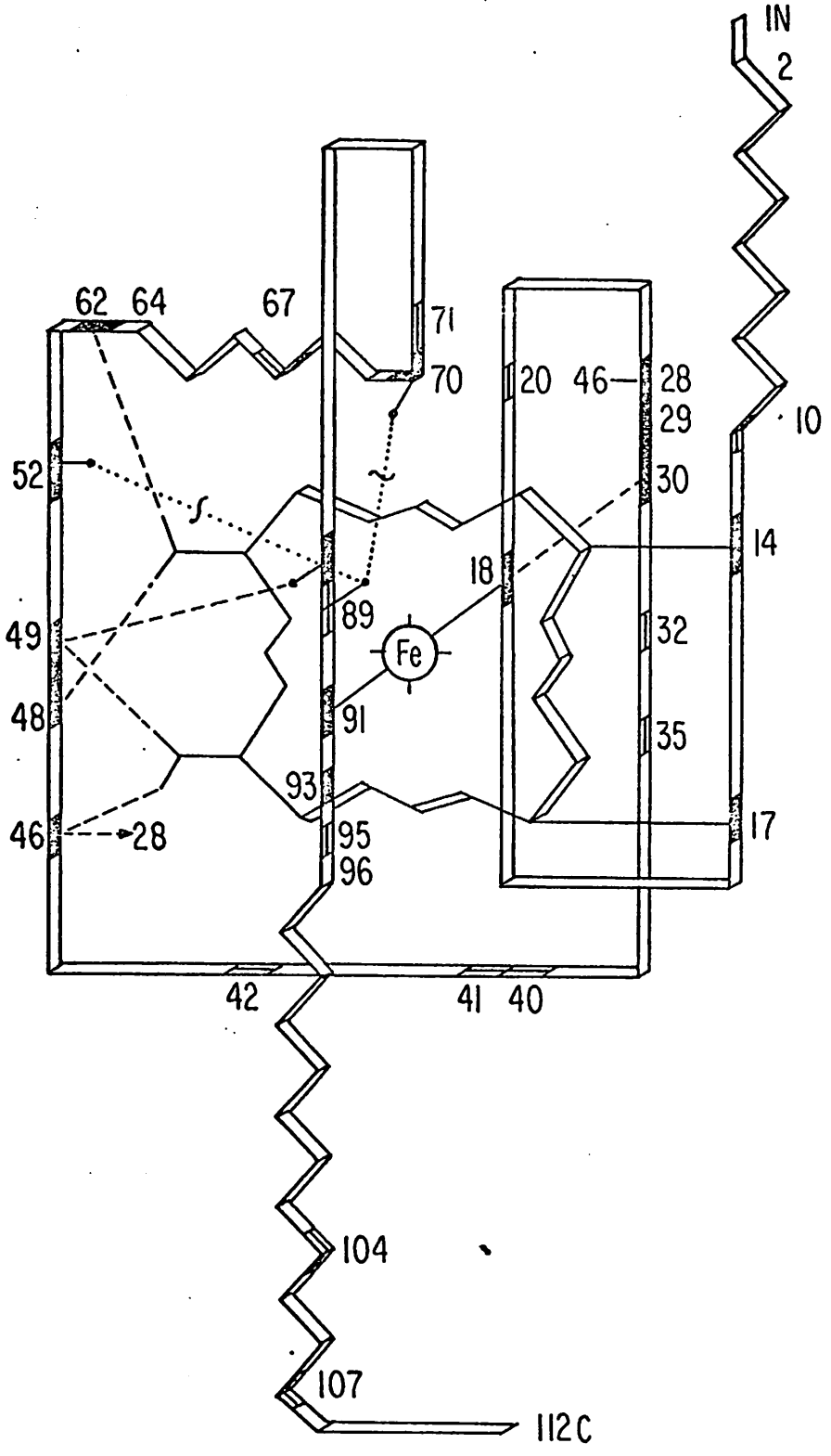

IN
2
10
62
64
67
71
70
20
46
28
29
30
14
52
18
89
49
Fe
32
48
91
35
93
46
28
95
96
17
42
41
40
104
107
112C

FIGURE 6

Fourier section through electron density in ligand plane normal to heme porphyrin ring. Electron density corresponding to the sulfur atom of methionine 91 appears displaced off axis of Fe-N bond by approximately 0.5 Å.

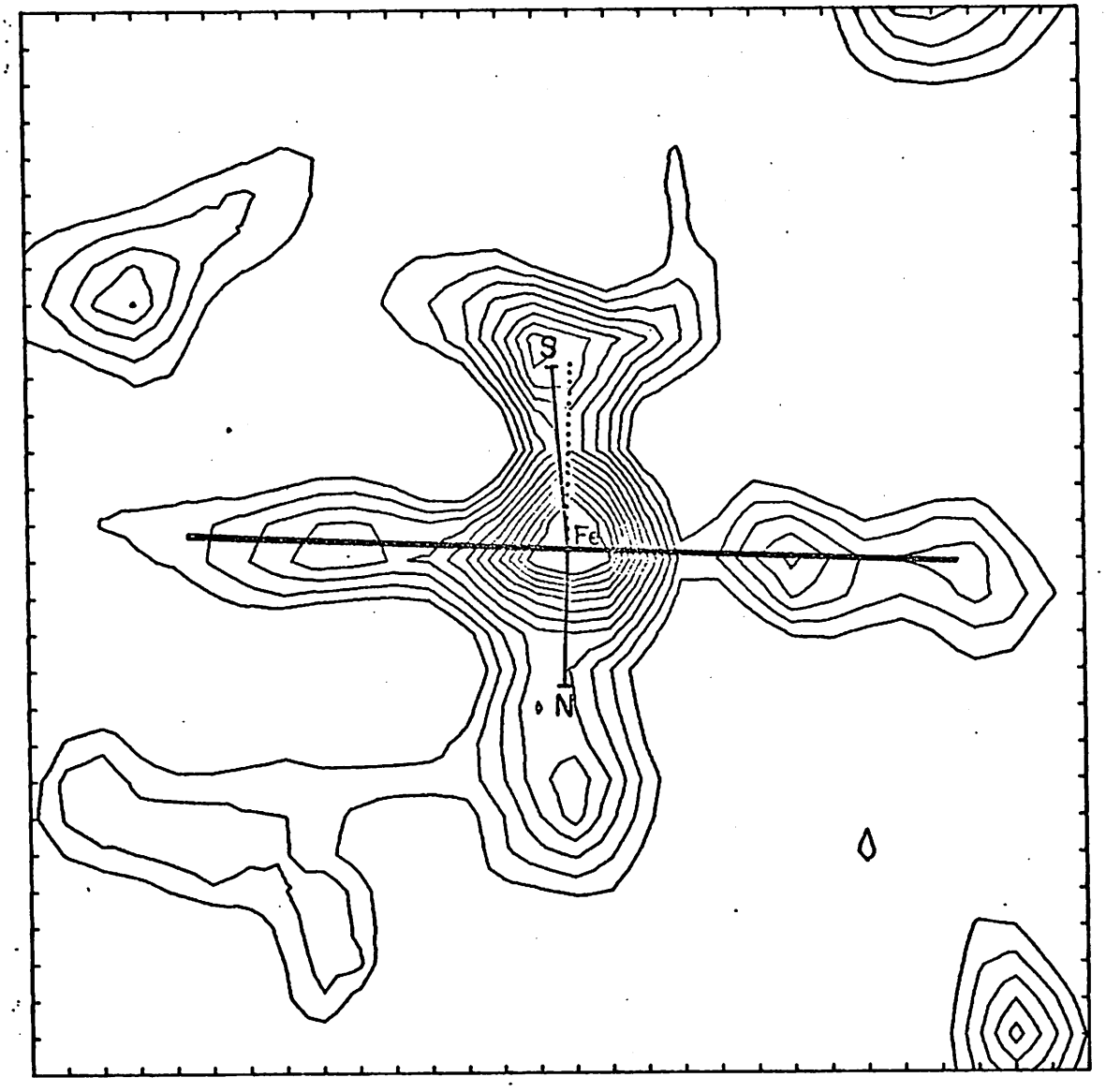

S
Fe
N

FIGURE 7

Stereo photograph of the heme crevice.

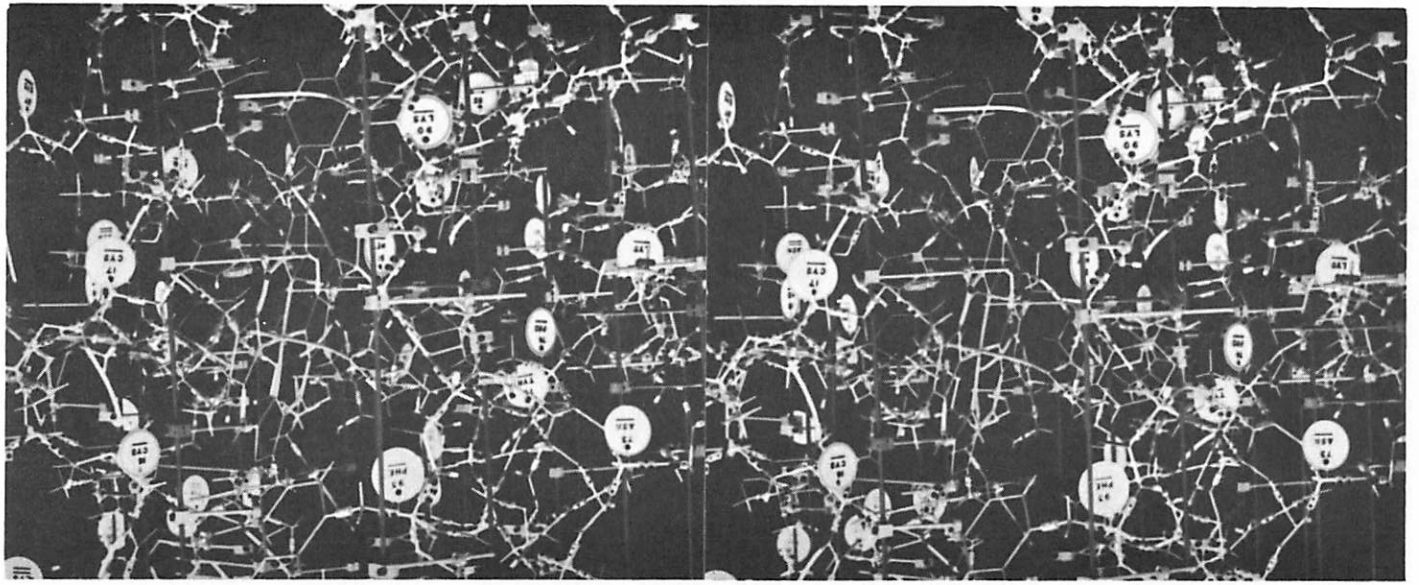


FIGURE 8

Location of 11 lysine residues bounding the perimeter of the heme crevice in cytochrome $\underline{c}_2$.

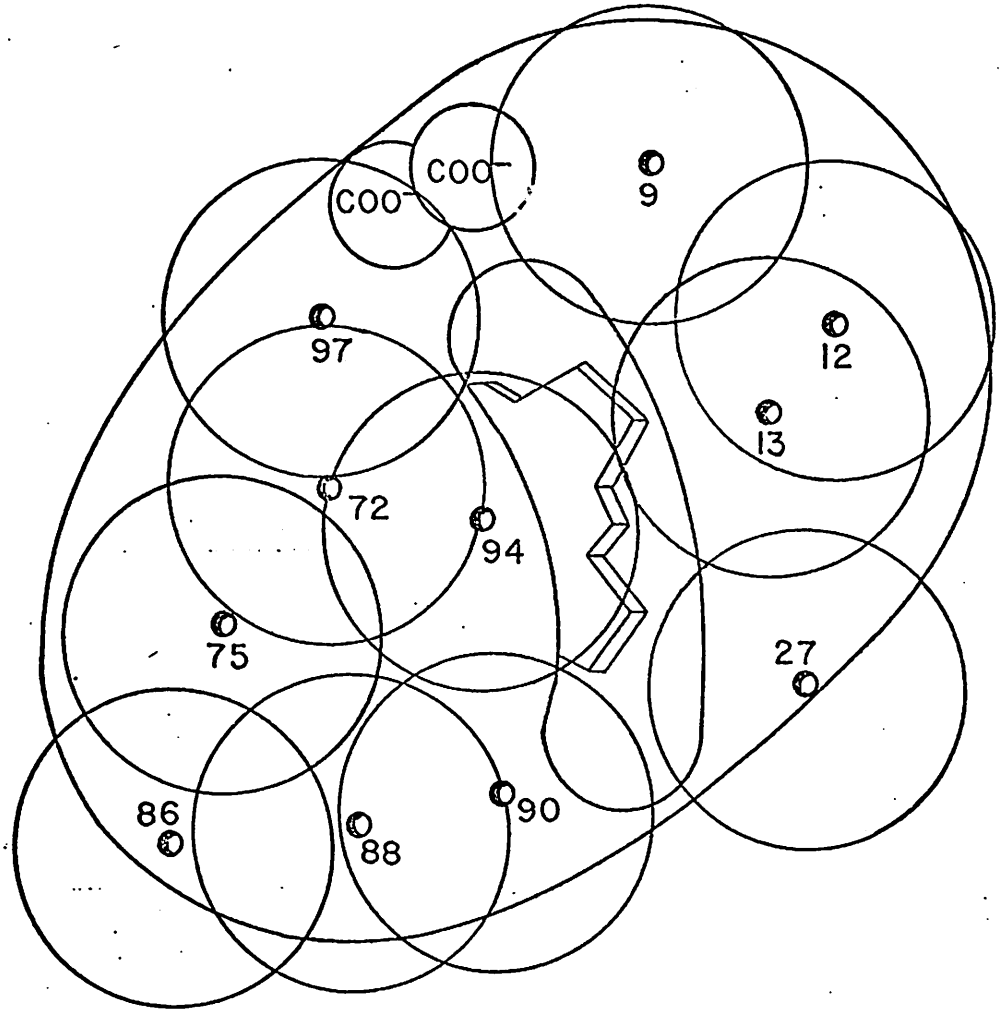

COO⁻
COO⁻
9
97
12
13
72
94
75
27
86
88
90

Table 1. Statistics for Last Cycle of Refinement

Derivative	Site No.	$\frac{x}{a}$	$\frac{y}{b}$	$\frac{z}{c}$	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}	z	k	c	rms F_H	rms E	R_1	R_2	Anom. K
$OsNH_4$	1	0.43	0.46	0.12	110.9	101.8	14.4	22.4	0.2	7.0	39.6	0.98	-1.88	42.2	30.2	12.07	55.34	0.08
	2	0.85	0.44	0.29	254.5	99.6	22.3	23.8	26.4	7.9	42.3							
	3	0.40	0.31	0.11	95.3	205.5	27.6	4.3	4.4	-9.2	21.2							
$UO_2(NO_3)_2$	1	0.11	0.19	0.14	48.8	36.3	7.7	5.1	1.3	1.6	48.6	1.35	-1.81	77.7	26.8	10.88	44.61	0.14
	2	0.25	0.41	0.26	55.0	46.0	6.4	4.5	1.3	1.4	52.4							
	3	0.90	0.03	0.17	59.6	53.2	14.4	2.2	-2.8	4.9	26.8							
	4	0.39	0.41	0.13	84.3	59.0	14.2	-2.3	-0.3	5.0	12.8							
$AuCl_4$	1	0.85	0.18	0.43	29.2	37.1	3.7	1.8	1.5	0.5	37.7	1.21	-1.38	48.6	28.2	12.28	58.26	0.10
	2	0.14	0.11	0.07	15.4	6.3	2.2	0.1	-0.2	-0.6	2.2							

Mean figure of merit = 0.74.

β as given here is defined by: $T = \exp\{-10^{-4}(\beta_{11}h^2 + \beta_{22}k^2 + \beta_{33}l^2 + 2\beta_{12}hk + 2\beta_{13}hl + 2\beta_{23}kl)\}$.

Z is the occupancy in electrons obtained from a refinement cycle of calculated parent structure factors.

k is the sacle factor to put the structure factor for a derivative on the same scale as the parent.

c is the factor that compensates for any difference in the temperature factor between the derivative and the parent: $k' = k \exp c \left(\frac{\sin\theta}{\lambda}\right)^2$.

f_H is the calculated heavy atom contribution to the structure factor.

E is the close error: $|k' F_H(obs) - F_H(calc)|$.

$$R_1 = \frac{\Sigma |k' F_H(obs) - F_H(calc)|}{\Sigma k' F_H(obs)}$$ sums over all hkl.

$$R_2 = \frac{\Sigma |k' F_H(obs) \mp F(obs)| - f_H|}{\Sigma |k' F_H(obs) \mp F(obs)|}$$

REFERENCES

1. Elsden, S. R., Kamen, M. D., and Vernon, L. P., J. Am. Chem. Soc., 75, 6347 (1953).

2. Dickerson, R. E., J. Mol. Biol., 57, 1 (1971).

3. Sletten, K., Dus, K., De Klerk, H., and Kamen, M. D., J. Biol. Chem., 243, 5492 (1968).

4. Taniguchi, S., and Kamen, M. D., Biochim. Biophys. Acta, 96, 395 (1965).

5. Dus, K., Sletten, K., and Kamen, M. D., J. Biol. Chem., 243, 5507 (1968).

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

7. Hotio, T., and Kamen, M. D., Biochim. Biophys. Acta, 48, 266 (1961).

8. Salemme, F. R. (to be published).

9. Xuong, Ng. H., Kraut, J., Seely, O., Freer, S. T., and Wright, C. S., Acta Cryst., B24, 289 (1968).

10. Freer, S. T., Kraut, J., Robertus, J. D., Wright, H. T., and Xuong, Ng. H., Biochemistry, 9, 1997 (1970).

11. Wyckoff, H. W., Doscher, M., Tsernoglou, D., Inagami, T., Johnson, L. N., Hardman, K. D., Allewell, N. M., Kelly, D. M., and Richards, F. M., J. Mol. Biol., 27, 563 (1967).

12. Furnas, T. C., Single Crystal Orienter Instruction Manual (General Electric Company, Milwaukee, 1967).

13. Freer, S. T., and Carter, C. W. (personal communication).

14. Kraut, J., Singh, S., and Alden, R. A., in Structure and Function of Cytochromes (K. Okunuki, M. D. Kamen and I. Seduzu, eds.), University of Tokyo Press and University Park Press, Tokyo, Japan and Baltimore, Maryland, 1968, p. 252.

15. Richards, F. M., J. Mol. Biol., 37, 225 (1968).

16. Lewis, P. N., and Scheraga, H. A., Arch. Biochem. Biophys., 144, 576 (1971).

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

18. Perutz, M. F., Proc. Roy. Soc. B, 173, 113 (1969).

19. Kewdrew, J. C., Sci. Amer., Offprint 121, Dec. 1961.

20. Mathews, F. S., Levine, M., and Argos, P., Nature New Biology, 233, 16 (1971).

21. Kakuno, T., Bartsch, R. G., Nishikawa, K., and Horio, T., J. Biochem., 70, 79 (1971).

22. Horio, T., Bartsch, R. G., Kakuno, T., and Kamen, M. D., J. Biol. Chem., 244, 5899 (1969).

23. Bartsch, R. G., personal communication.

24. Flatmark, T., and Robinson, A. B., in Structure and Function of Cytochromes (K. Okunuki, M. D. Kamen and I. Sekuzu, eds.), University of Tokyo Press and University Park Press, Tokyo, Japan and Baltimore, Maryland, 1968, p. 318.

CHAPTER III

PRELIMINARY CRYSTALLOGRAPHIC INVESTIGATION OF GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE FROM PALINURUS

SUMMARY

Glyceraldehyde 3-phosphate dehydrogenase isolated from South African spiny lobster (Palinurus) crystallizes from ammonium sulfate as flat rhombic plates. Preliminary investigation of these crystals by x-ray precession photography indicates that they possess an orthorhombic space group I222 or $I2_12_12_1$. The unit cell dimensions are $a = 98 \pm 1$ Å, $b = 66 \pm 1$ Å, and $c = 143 \pm 2$ Å. Assuming a molecular weight for this protein of 144,000 daltons, it is calculated that the crystallographic asymmetric unit consists of one quarter of the molecule. Based upon packing considerations assuming a tetrahedral tetramer as the molecular unit, Palinurus GPD is assigned the space group I222. A comparison of the molecular packing in human and Palinurus GPD is also discussed.

Glyceraldehyde 3-phosphate dehydrogenase (GPD, EC 1.2.1.12) is the enzyme catalyzing the phosphorylation of D-glyceraldehyde-3-phosphate to form 1, 3-diphosphoglycerate, with the concomitant reduction of NAD^+ to NADH. The molecular weight of GPD is 144,000 daltons as determined for the yeast and rabbit muscle enzymes by physical methods (1) and sequence determination of the European lobster (2) and pig muscle (3) enzymes. The native physiologically active molecule is composed of four sequentially identical subunits

about 335 amino acid residues long which have a molecular weight of 36,000 daltons.

Comparison of peptide fingerprint maps of proteolytic enzymic digests of GPD isolated from the tail muscle of the South African spiny lobster (Palinurus) with those obtained from other species having a known sequence establishes an essential similarity in the sequence of Palinurus GPD with those of ther crustaceans (Allison, unpublished data). This communication describes the crystallization and space group determination of Palinurus GPD.

GPD was isolated from commercially obtained frozen spiny lobster tails by the method described by Allison and Kaplan (4). The final step in this preparation yields a solution containing about 3.5 mg of GPD per ml. When this solution is saturated with solid ammonium sulfate at 4°C, GPD is precipitated as microcrystals. The microcrystals are collected by centrifugation and dissolved in a minimum volume of 0.05 M sodium phosphate buffer pH 6.5 containing 0.05 M 2-mercaptoethanol and 0.005 M EDTA. Using this procedure, a "stock" crystallization solution having a protein concentration of about 50 mg/ml may be obtained.

Crystals suitable for x-ray studies are grown by layering 50 ml aliquots of the "stock" concentrated GPD solution over 100 µl of identically buffered 70% saturated ammonium sulfate in test tubes measuring 3.5 mm (ID) × 20 mm. The tubes are sealed with Parafilm

and left undisturbed for two to three weeks at 25°C. Palinurus GPD crystallizes as flat rhombic plates. The largest crystals grown by the method outlined above are 0.4 × 0.4 × 0.2 mm. The crystals are sensitive to changes in temperature and will crack if cooled to 5°C and then allowed to return to room temperature. However, transferring them into buffered 90% saturated ammonium sulfate appeared to stabilize these crystals against thermal fracture.

Single crystals were mounted for x-ray analysis in 1 mm diameter Lindemann glass capillaries in the conventional manner. Reciprocal lattice photography was carried out on a Supper precession camera, using a General Electric CA8-F/CU 550 watt x-ray source. Eight degree precession photographs showing reciprocal lattice spacings to 5.6 Å were obtained for the okl and hol reciprocal lattice planes. In addition a 2° screenless precession picture was obtained showing sections of upper level lattice planes normal to a^*. About 24 hours of exposure was required to obtain a photograph of sufficient density to allow cell parameter measurements to be made. In some cases, the diffracting power of these crystals deteriorated rapidly after 48 to 72 hours of exposure. This effect is probably a manifestation of the temperature sensitivity of these crystals, since in one case, during a dry spell when the ambient temperature and barometric pressure were unusually stable, a crystal was observed to diffract strongly after over 96 hours of x-ray exposure. The large

mean deviations found for the unit cell dimensions of Palinurus GPD reflect variations found between different individual crystals. The unit cell dimensions of Palinurus GPD are:

$a = 98 \pm 1$ Å $\quad b = 66 \pm 1$ Å $\quad c = 143 \pm 2$ Å

$\alpha = \beta = \gamma = 90°$ $\quad V = 925,000 \pm 36,000$ Å^3

The systematic absences observed in the hol and okl base planes are:

$$hol\ (l + h \neq 2n)$$

$$okl\ (k + l \neq 2n)$$

Sections of upper level planes normal to a^* show systematic absences for reflections where $(h + k + l) \neq 2n$. These restrictions indicate that Palinurus GPD crystallizes in one of the orthorhombic space groups I222 or $I2_12_12_1$ which are indistinguishable by reciprocal lattice photography.

Assuming the density of Palinurus GPD crystals to be similar to that found for the European lobster enzyme (5) which has a density of 0.34 daltons/Å^3, it is calculated that the protein mass contained in the Palinurus GPD unit cell equals

925,000 Å^3 × 0.34 daltons/Å^3, or 316,000 daltons.

Since both of the orthorhombic space groups I222 and $I2_12_12_1$ contain eight crystallographically equivalent positions, the molecular weight of the crystallographic asymmetric unit equals 316,000/8, or

39,200 daltons. This value is in reasonable agreement with the figure of 36,000 daltons for the molecular weight of a single GPD subunit comprising one-quarter of the physiologically active GPD tetramer. Thus the Palinurus unit cell contains eight subunits or two tetrameric molecules of GPD.

Possible packing arrangements of the eight asymmetric subunits in both I222 and $I2_12_12_1$ unit cells are shown in Figs. 1 and 2. The packing of the subunits in the space group I222 leads to the formation of a tetrahedral tetramer whose three molecular diad axes are coincident with crystallographic diad axes. Packing subunits into the space group $I2_12_12_1$ generates a continuous helical "molecule" which lies on a crystallographic 2-fold screw axis parallel to c. Since this type of packing is not possible for real protein subunits, which are not spherically symmetric, it is possible to tentatively assign the space group of Palinurus GPD as I222.

COMPARISON OF MOLECULAR PACKING IN PALINURUS AND HUMAN GPD

Goryonov and Andreeya (6) have reported that human GPD crystallizes in the orthorhombic space group $C222_1$ with unit cell dimensions:

$$a = 97 \pm 0.5\ \text{Å},\quad b = 131 \pm 1\ \text{Å},\quad c = 147 \pm 1\ \text{Å}$$

$$V = 1,870,000 \pm 40,000\ \text{Å}^3$$

These dimensions closely resemble those found for Palinurus GPD with the exception that the b axis, and hence the unit cell volume, has been doubled in the human GPD crystals. The space group $C222_1$ contains eight crystallographic asymmetric units per unit cell, and it has been shown that the asymmetric unit of the human GPD crystals consists of one-half of the GPD tetramer. In addition to extinctions characteristic of $C222_1$, supplementary extinctions were detected:

$$okl: \; k + 2l \neq 4n$$
$$oko: \; k \neq 4n$$

which indicate the existence of pseudosymmetry with space group $I4_122$, which contains sixteen crystallographic asymmetric units per unit cell. Also supplementary extinctions hol: $l \neq 2n$ were observed. The $l \neq 2n$ extinctions on the hol plane were shown to arise from the presence of two additional symmetry elements; two perpendicular diad axes parallel to the a and c unit cell axes. This arrangment is no longer consistent with a unit cell containing eight subunits, but can be accounted for by locating a tetrahedral molecule of four subunits on the intersection of the three mutually perpendicular diad axes. Figure 3 shows these additional elements of human GPD symmetry projected into one-half (b/2) of the $C222_1$ unit cell. Note that only one of the molecular diad axes of the tetrahedral tetramer is coincident with a crystallographic diad axes. Figure 4 shows the I222 symmetry

operations characteristic of Palinurus GPD. The symmetry representations of the packing in human and Palinurus GPD are quite similar. The difference is that while the Palinurus I222 cell has a two-fold screw axis parallel to b at x = a/4, z = c/4, in the human GPD cell the corresponding two-fold screw axis is located at x = a/4, z = 0.

The molecular packing observed in the human crystal lattice necessitates that the two halves of each molecule, consisting of two MW - 36,000 subunits each, are structurally equivalent. The $I4_122$ pseudosymmetry implies that all four molecular subunits are structurally identical, and that the appearance of a two-subunit crystallographic asymmetric unit is an artifact of molecular packing in the crystal lattice. The molecular packing in the Palinurus crystal lattice, having one molecular subunit per crystallographic asymmetric unit unequivocally establishes that the Palinurus enzyme consists of a tetrahedral tetramer having four structurally identical subunits.

FIGURE 1

The molecular crystal packing in the space group I222.

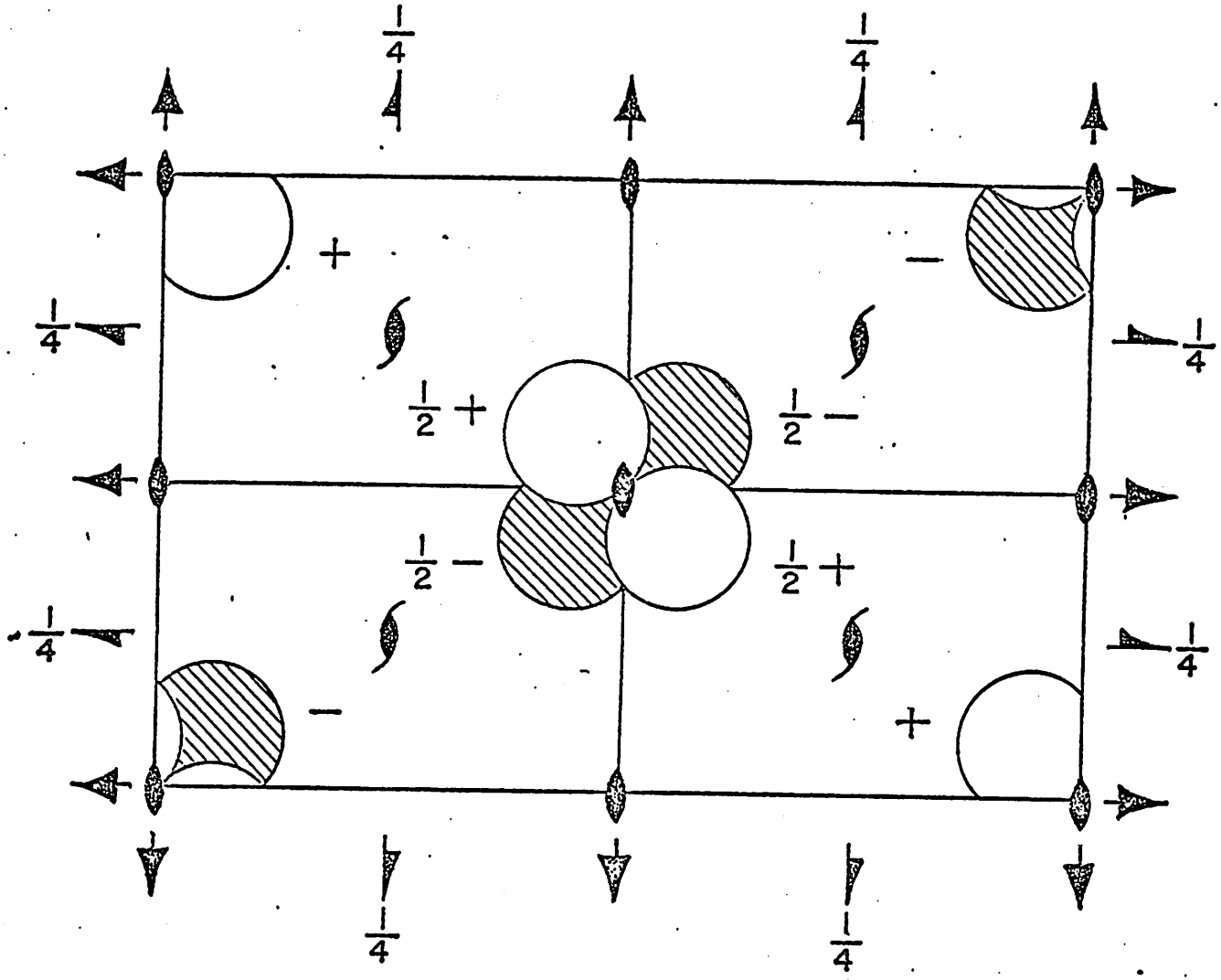

1/4
1/4
+
−
1/4
1/4
1/2 +
1/2 −
1/2 −
1/2 +
1/4
1/4
−
+
1/4
1/4

FIGURE 2

The molecular crystal packing in the space group $I2_12_12_1$.

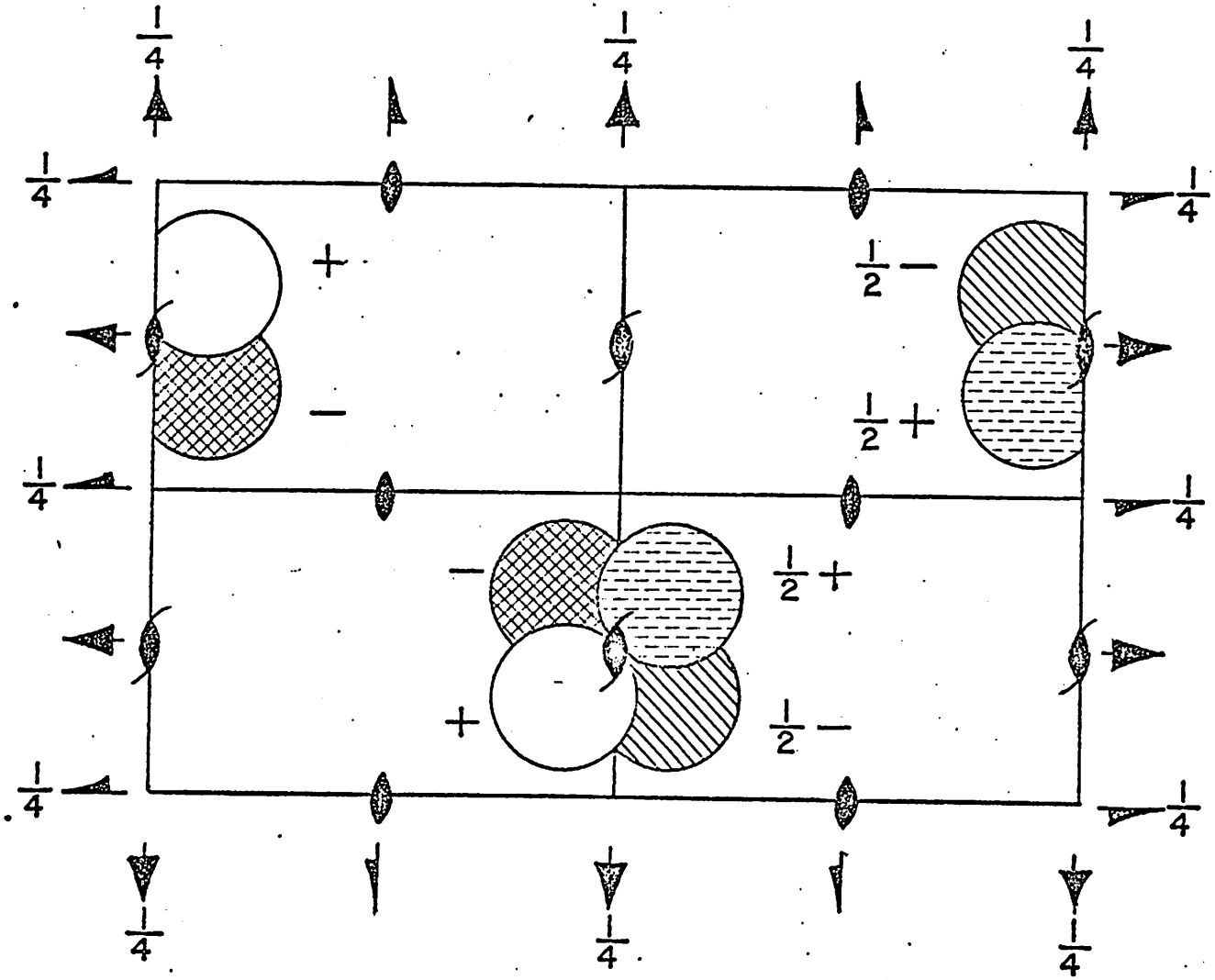

1/4
1/4
1/4
1/4
1/4
1/4
+
−
1/2 −
1/2 +
1/4
1/4
−
1/2 +
+
1/2 −
1/4
1/4
1/4
1/4
1/4

FIGURE 3

Pseudo-symmetry elements of human GPD crystals projected into one half (b/2) of the $C222_1$ unit cell.

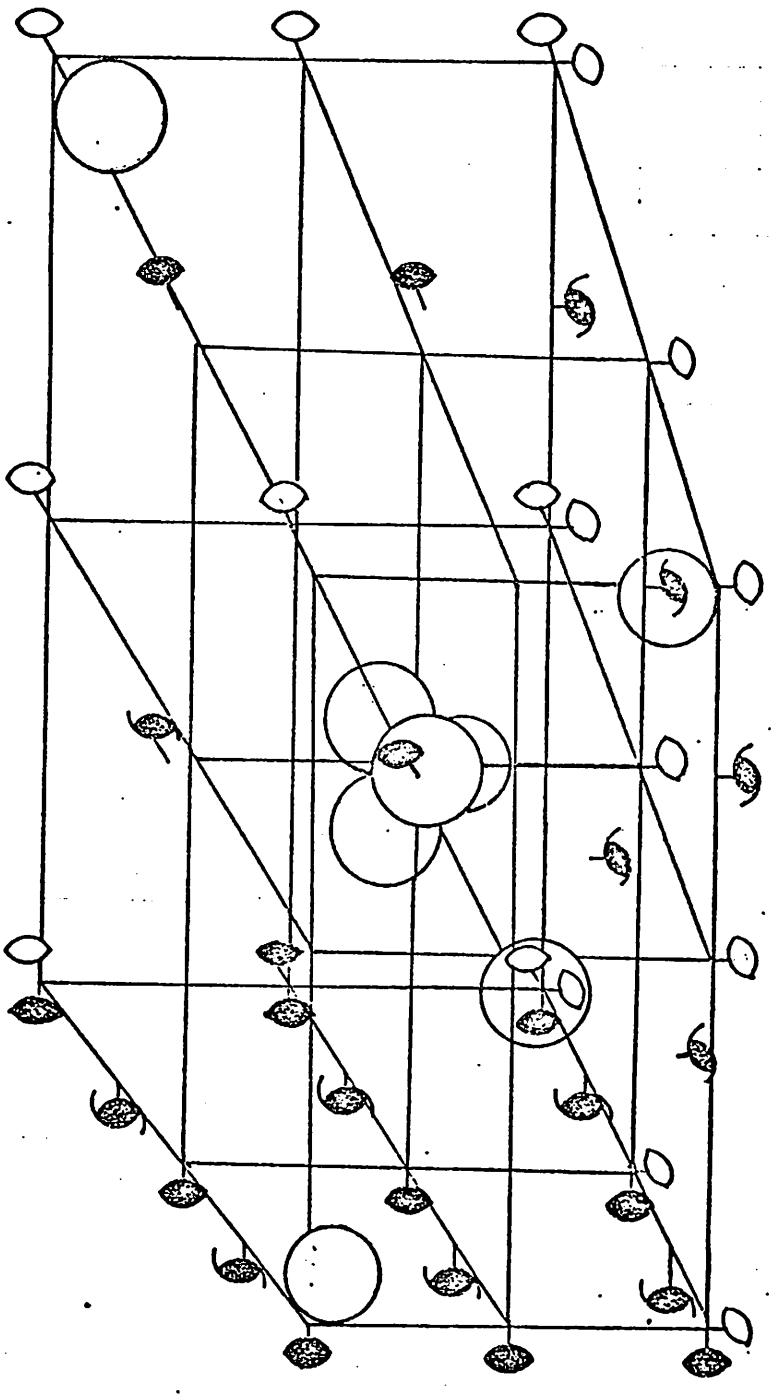
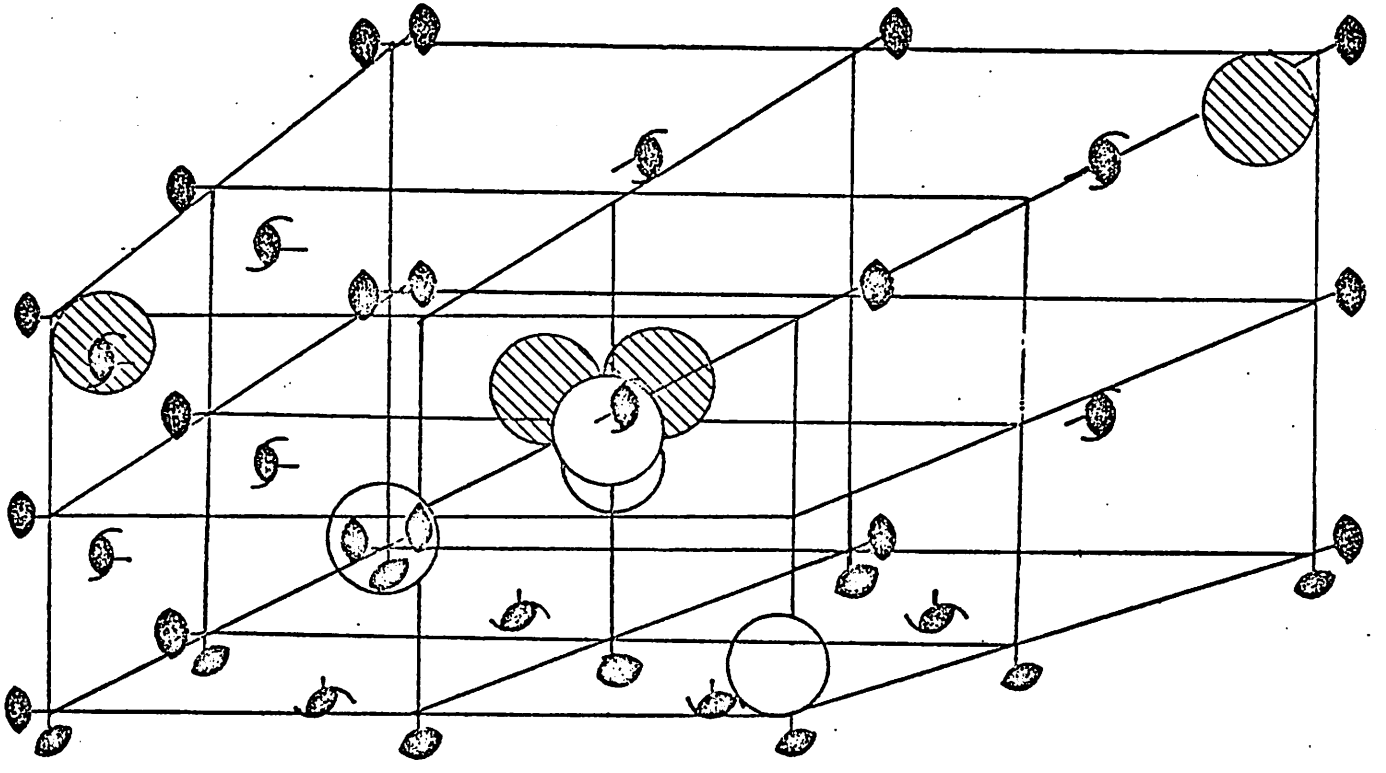


FIGURE 4

Crystal packing and symmetry elements characteristic of the Palinurus space group, I222.



REFERENCES

1. Jacnicke, R., Schmid, D., and Knof, S., Biochemistry, 7, 919 (1968).

2. Davidson, B. E., Sajgo, M., Noller, H. F., and Harris, J. I., Nature, 216, 1181 (1967).

3. Harris, J. I., and Perham, R. N., Nature, 219, 1025 (1968).

4. Allison, W. C., and Kaplan, N. O., J. Biol. Chem., 239, 2140 (1964).

5. Watson, H. C., and Banaszak, L. J., Nature, 204, 918 (1964).

6. Goryonov, A. I., and Andreeva, N. S., Molecular Biology (transl. from Russian), 1, No. 2, 261 (1967).

CHAPTER IV

A DEVICE FOR THE RAPID MEASUREMENT OF MOLECULAR MODEL COORDINATES FOR X-RAY CRYSTALLOGRAPHY

The recent proliferation of x-ray protein structure determinations has dictated the requirements for rapid and accurate model building and measuring techniques. Most of the methods currently in use have mechanical limitations which render them inaccurate, or are at least sufficiently tedious so that the prospect of repeated measurement and adjustment of atomic positions required during refinement procedures is contemplated with apprehension. The device described in this communication was constructed to improve the speed and precision with which coordinates are measured, and to facilitate construction and comparison of structures whose atomic coordinates are already known.

GENERAL OPERATING DESCRIPTION

The automatic coordinate hunting engine is used in conjunction with a Richards optical comparitor. Measurement of coordinates is based upon the same principle of optical superposition by which a molecular model is initially constructed from Fourier electron density maps. Figure 1 shows a schematic diagram of the coordinate engine mounted on a horizontal Richards optical comparitor. In order to measure the coordinates of an atom A in the model, a small spherical light source is mechanically positioned at B so that its transmitted image superimposes with the reflected image of the atom at A when viewed normal to the x'z' plane at C. Model

coordinates are measured in a right-handed cartesian coordinate system (x, y, z). Since the reflected image of the model is inverted by the mirror, the machine axes in the Fourier reference frame (x', y', z') define a left-handed coordinate system.

In the manual mode, the operator positions the marker light source by means of a "joy stick" controlling digital pulse motors which drive the light in the Fourier reference frame. An electronic counting circuit records the number of pulses required to superposition the images of the atom and the marker light and continuously displays coordinates relative to a preselected origin. The position of the light is indicated in 0.1 Å increments on the 2 cm/1 Å scale commonly used for the construction of molecular models. The origin is set by "loading" the pulse counting circuits with some predetermined values by means of thumblever switches on the electronics console. Travel on the axes is restricted by limit switches which shut off the drive motors when contacted. The limit switches may be used to mechanically define the origin of the cartesian system by driving the motors to their minimum extremum and then "loading" the counter values to x = 00.0, y = 00.0, z = 00.0 by means of the thumblever switches.

Once the origin has been loaded, it is possible to automatically position the light at any desired location relative to the origin by (1) switching to the automatic mode, which disables the joy stick

(2) setting the desired coordinates on the thumblever switches, (3) hitting a "seek" button activating comparator and pulse generating circuits which automatically generate the proper number of pulses in the required direction to drive the motors to the selected position.

Technical description. The automatic coordinate hunting engine consists of two major components; the mechanical support for positioning the marker light source, and the electronics module containing the control logic necessary to keep track of its position (Fig. 2).

The mechanical component (Fig. 3) consists of a rigid base upon which are mounted the y' axis drive motor and two one-inch diameter ball bearing slides which support the vertically mounted z' carriage. The y' axis motor drives a one-half inch pitch ball bearing screw which is meshed to a recirculating ball-nut fixed to the lower end of the z' carriage. An incoming pulse from the electronics module causes the drive motor to incrementally step through a fixed angle of 1.8°. The resulting rotary motion of the ball screw imparts a horizontal translation in the y' direction of 0.0025 inches to the z' carriage.

The z' carriage mounts the z' drive motor and two one-half inch diameter ball bearing slides supporting the x' axis beam. The z'-drive motor imparts vertical motion to the x' beam by means of an identical ball screw and nut arrangement as used to position by z' carriage.

The marker light is fixed to a small sled riding on 4 wires strung along the length of the x' beam, which is constructed of 3-inch diameter thin wall alloy tubing. The marker light is translated along x' by a cable and pully arrangement driven by the x' axis digital pulse motor.

A suitably small marker light is manufactured from a 1/8" diameter length of transparent acrylic plastic which has been rough ground to a hemisphere on one end and fixed to a lens type 1.5 volt penlight bulb. All but 1/8" from the tip of the rod is painted black, which gives a small well-defined source for superpositioning with model images. The largest volume in which coordinates may be measured is a cube 94 cm or 47 Å on a side at the 2 cm/1 Å scale.

The electronics for the coordinate engine are mounted in a console connected to the mechanical chassis by an umbilical cord (Fig. 2). The basic function of the electronics module is to generate and count pulses sent to the digital pulse drive motors used to position the marker light. A simplified schematic for one axis of the system is shown in Fig. 4. A 200 Hz oscillator generates "count up" and "count down" pulses which simultaneously incrementally step the digital pulse drive motors and are counted by the scaler. The scaler consists of an alternating modulus (31-32) counter which generates the 2 cm/1 Å scale factor. The average count of 31.5 pulses represents a displacement along the screw of 31.5 × 0.0025"

= 0.07875" which differs from 0.07874" = 0.1 Å at 2 cm/Å scale by about 0.01%. The scaled pulse train then passes into a cascade of three bidirectional binary coded decimal counters. The numerical states of the counters are continuously displayed on a three digit decimal readout indicating position in 0.1 Å increments. The origin is established by "loading" the counters with any desired values centered from the digital thumblever switches.

In the manual mode, the gating of the pulse train from the oscillator is controlled by the joy stick. Since the bidirectional pulse train passes simultaneously to both the counting chain and the motor driver, the readout always displays the position of the marker light relative to the preset origin.

Switching to automatic mode of operation disables the joy stick. When the "seek" button is depressed, a chain of four magnitude comparators compares the numerical state of the BCD counters with whatever preselected position has been entered on the thumblever switches. The output from the comparator passes to the control logic which gates the oscillator and automatically supplies the proper number of pulses to bring the marker to the position entered on the thumblever switches.

Mechanical limit switches prevent overdriving in both automatic and manual operational modes. The position of the limit switches is adjustable to allow the absolute mechanical placement of

an origin coincident with any desired position in a model being measured.

The electronics module of the hunting engine utilizes integrated circuitry throughout for high reliability and compactness. The scaling and counting circuitry is modularized with one circuit board per axis. The entire electronics unit including power supplies and motor drivers occupies a chassis 15" × 22" × 19" in size.

APPLICATIONS AND RESULTS

Coordinate measurement: The use of the coordinate hunting engine allows an operator to measure molecular model coordinates continuously from one end of the chain to the other at a rate of approximately ten amino acid residues (about 90 atoms) per hour. Use of the automatic mode allows rapid checking in the event that an error in transcription of an atom's position or sequence designation is made.

The distribution of the N-C_α, C_α-C(carbonyl) and C-N bond lengths calculated to the nearest 0.1 Å from a single measurement of the coordinates of a model of cytochrome c_2 (R. rubrum, 112 residues) (2) is shown in Fig. 5. The average bond lengths and mean deviations calculated from these figures are N-C_α = 1.47 (0.08), C_α-C = 1.58 (0.06), and C-N = 1.34 (0.07) Å. The comparable bond lengths fixed by the 2 cm/Å Kendrew skeletal models are 1.44, 1.53, and 1.32 Å respectively. Remeasurement of atoms having a

calculated bond length more than 0.1 Å from the expected value can be rapidly accomplished by utilizing the automatic mode to preposition the marker light at the initially determined coordinates. In this way it is possible to rapidly obtain a precise set of coordinates for a protein structural model.

Computer drawings produced from machine measured coordinates show symmetrical and flat rings. Consequently, the machine has been useful in preparing computer drawings of model substrate and inhibitor complexes of the enzyme subtilisin (3).

Model building: When the coordinate engine is operated in the automatic mode, it is possible to automatically position the marker light at any predetermined coordinates entered from the digital thumblever switches. This feature has been used to rapidly build analogue substrates of α-chymotrypsin into the active site of the enzyme subtilisin (3).

In order to compare the detailed tertiary conformation of mitochondrial cytochrome c (4) and the photosynthetic cytochrome c_2 of R. rubrum, the coordinates of cytochrome c were transformed and rotated into the proper orientation in the engine coordinate system used to measure cytochrome c_2. Small sections of the chain were built using calculated interatomic dihedral angles. These sections were subsequently positioned relative to one another in the Richards box using the automatic mode of the hunting engine. Model

building by this technique is quite fast because of the rapidity and ease with which the reference marker light can be positioned.

Refinements: The engine as designed measured atomic positions with a precision of 0.1 Å, which would certainly seem sufficient given the uncertainty of properly orienting the skeletal models in the electron density map when the structure is initially constructed. However, experience in this laboratory with various structure refinement procedures has indicated that a measurement precision of 0.05 Å would be desirable. Reduction of the size of the marker light source and addition of an illuminated reticle should allow measurements of this precision to be made.

FIGURE 1

Schematic drawing of the automatic coordinate hunting engine mounted on a horizontal Richards optical comparitor. The moving axes of the machine define a left-handed (x', y', z') cartesian coordinate system which is coincident with the mirror image of the right-handed coordinate system in the model (x, y, z) space. Coordinate measurement is made by mechanically superpositioning the transmitted image of the marker light B with the reflected image of the atom at A.

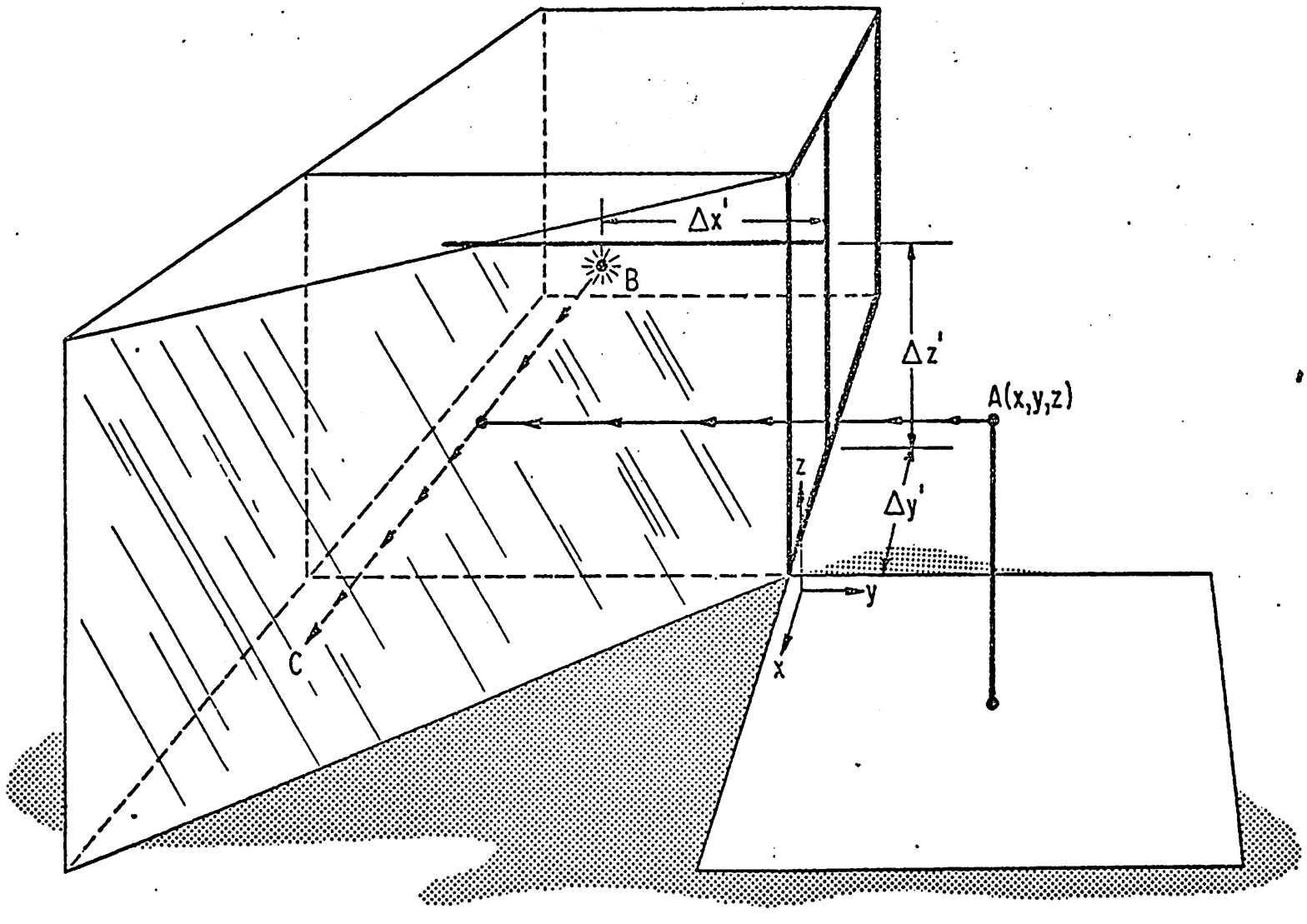

Δx'
B
Δz'
A(x,y,z)
z
Δy'
y
x
C

FIGURE 2

A photograph showing the coordinate engine electronics console, the horizontal Richards optical comparator, and a molecular model. The transmitted image of the x' axis support beam holding the marker light can be seen behind the reflected image of the molecule in the half-silvered mirror of the Richards box.

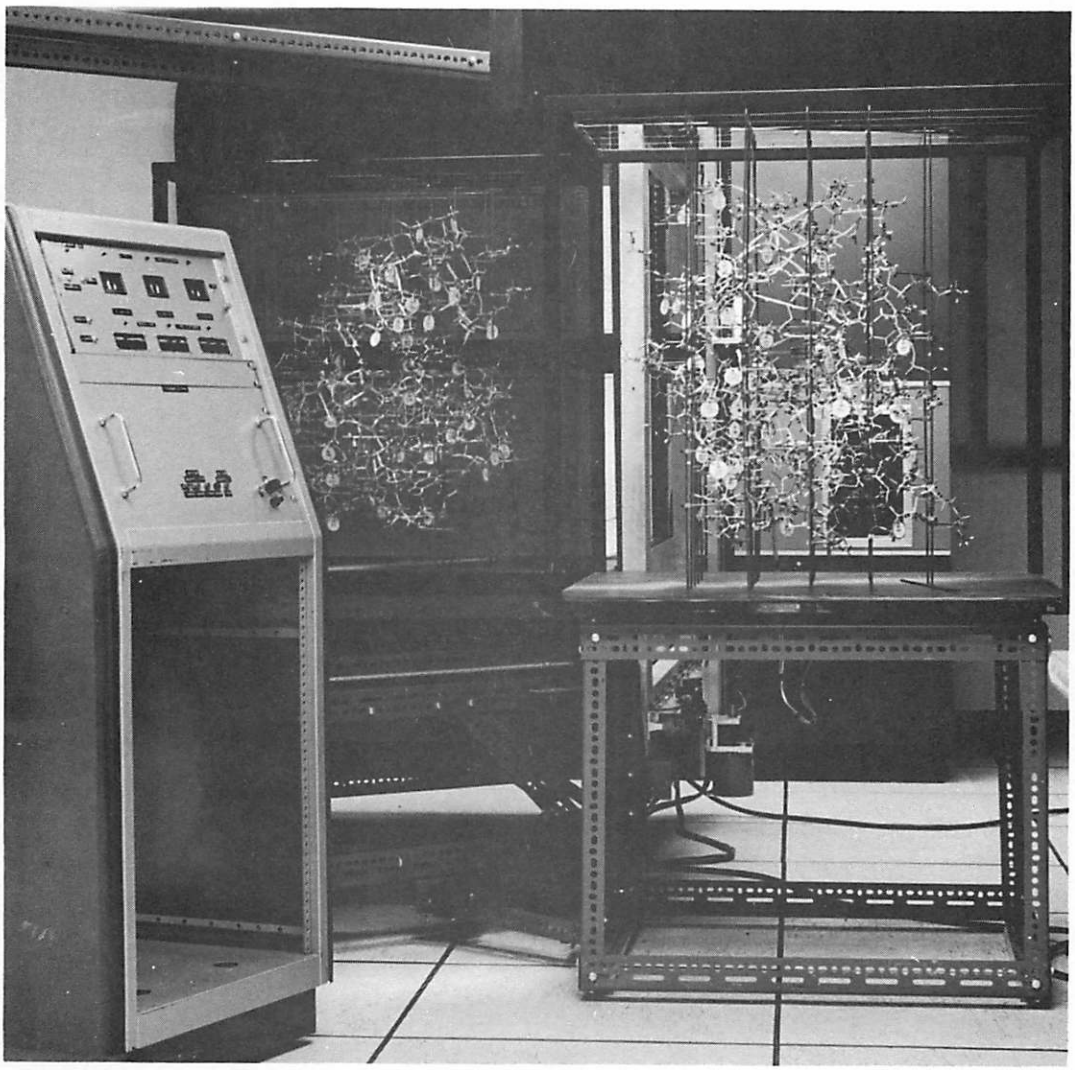


FIGURE 3

A side view of the mechanical unit of the coordinate hunting engine. Reference to Fig. 1 will allow the reader to identify individual mechanical components described in the text.

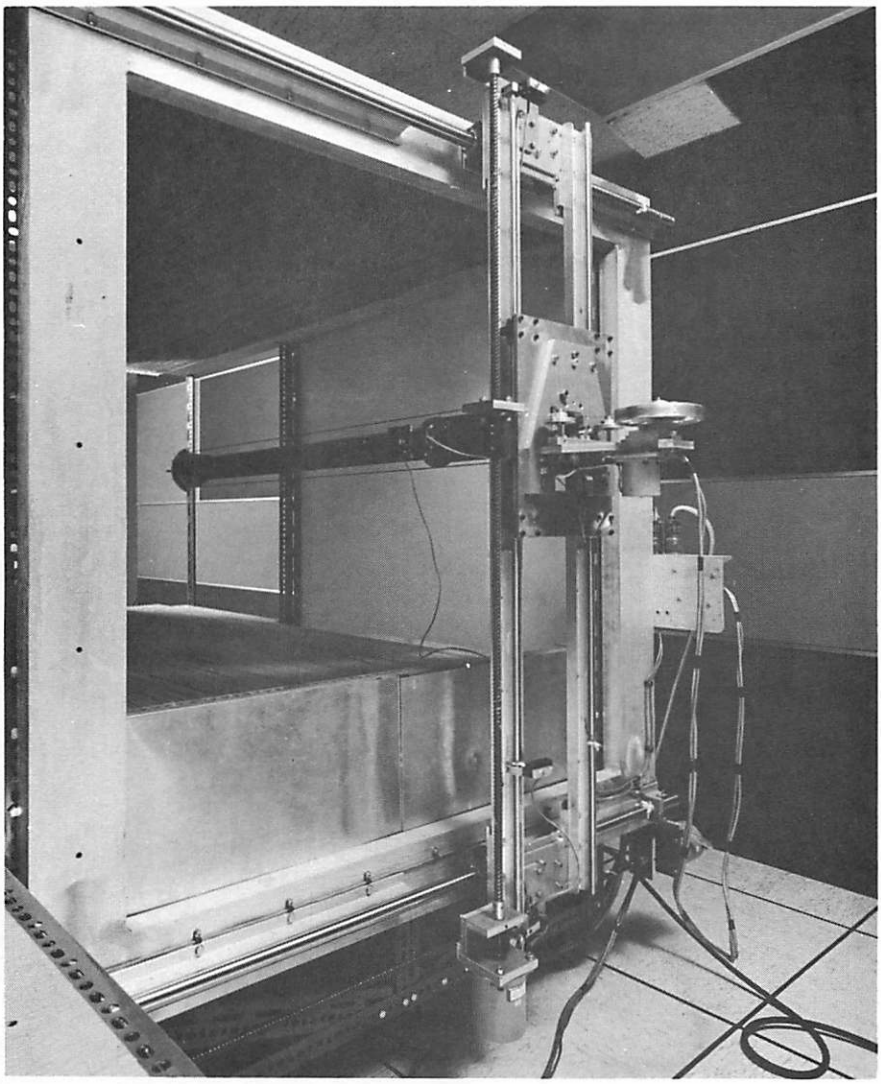


FIGURE 4

A schematic diagram of the digital pulse generating and counting circuitry of one axis of the coordinate engine.

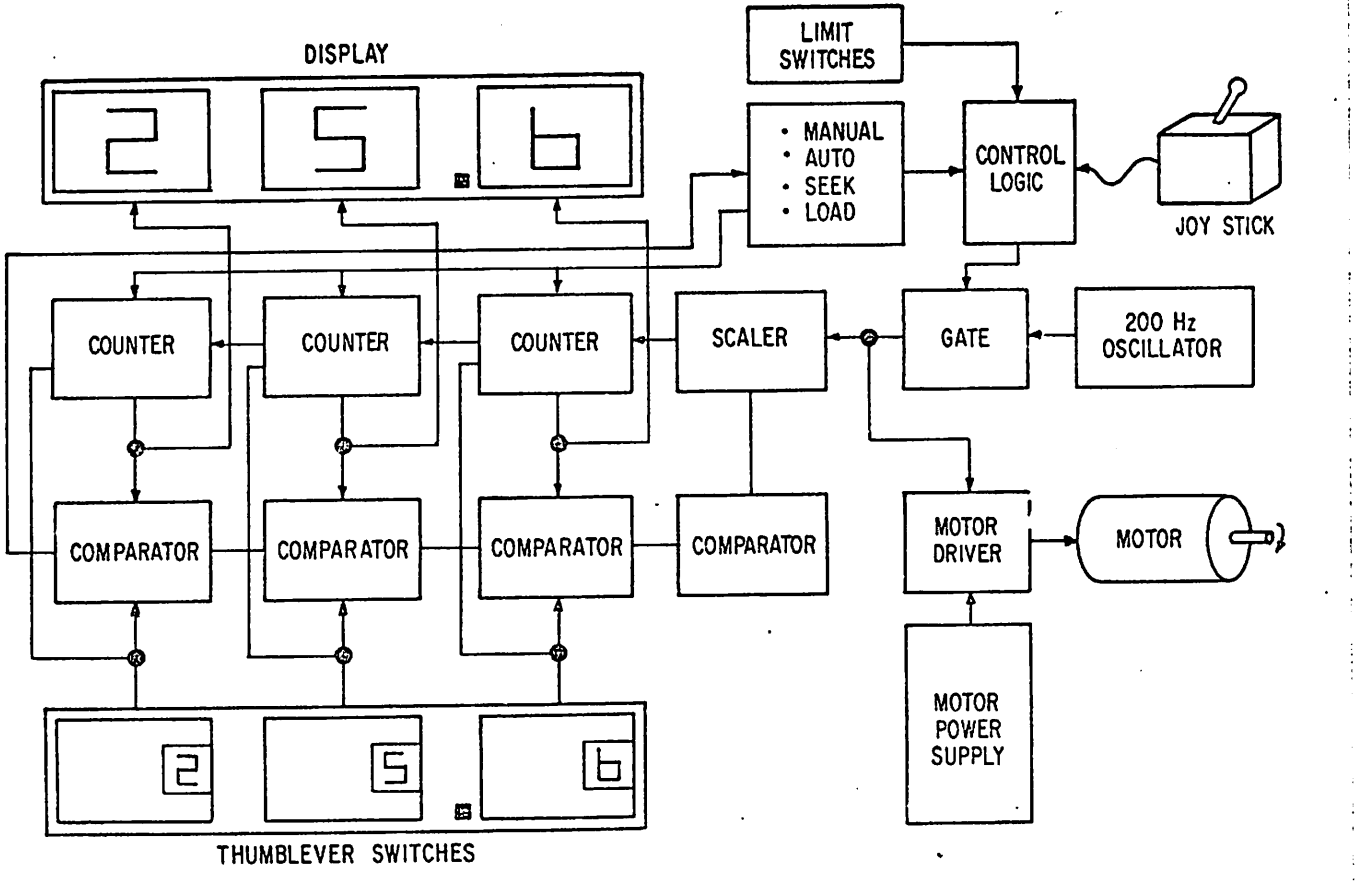
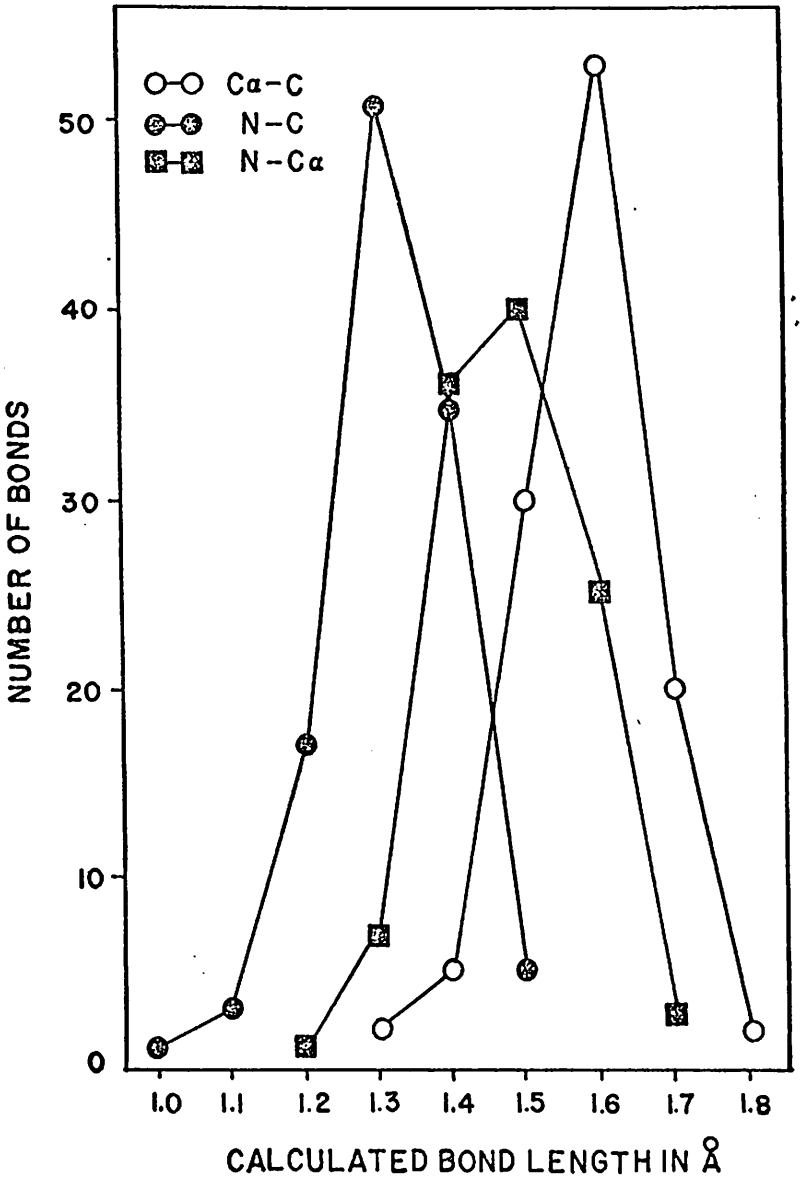

DISPLAY
2
5
6
LIMIT SWITCHES
• MANUAL
• AUTO
• SEEK
• LOAD
CONTROL LOGIC
JOY STICK
COUNTER
COUNTER
COUNTER
SCALER
GATE
200 Hz OSCILLATOR
COMPARATOR
COMPARATOR
COMPARATOR
COMPARATOR
MOTOR DRIVER
MOTOR
MOTOR POWER SUPPLY
2
5
6
THUMBLEVER SWITCHES

FIGURE 5

A graph of the distribution of bond lengths calculated to nearest 0.1 Å from a single measurement of the atomic coordinates of a model of cytochrome c_2 (R. rubrum, 112 residues).


Cα-C
N-C
N-Cα
NUMBER OF BONDS
0
10
20
30
40
50
1.0
1.1
1.2
1.3
1.4
1.5
1.6
1.7
1.8
CALCULATED BOND LENGTH IN Å

REFERENCES

1. Richards, F. M., J. Mol. Biol., 37, 225 (1968).

2. Salemme, F. R., Freer, S. T., and Kraut, J. (in preparation).

3. Robertus, J. D. (personal communication).

4. Dickerson, R. (personal communication).