## INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

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

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

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

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

University Microfilms International
300 North Zeeb Road
Ann Arbor, Michigan 48106 USA
St. John's Road, Tyler's Green
High Wycombe, Bucks, England HP10 8HR

RICE UNIVERSITY

# 3.5 Å RESOLUTION STRUCTURE OF L-ARABINOSE BINDING PROTEIN FROM *E. coli*

by

GEORGE NEAL PHILLIPS, JR.

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

THESIS DIRECTOR'S SIGNATURE:

F. A. Quiocho

HOUSTON, TEXAS

NOVEMBER 1976

TO MY MOTHER AND FATHER,

WHOSE HELP AND ENCOURAGEMENT HAVE MADE THIS ENDEAVOR POSSIBLE

ACKNOWLEDGMENTS

I would like to acknowledge the following for their roles in this study:

-Prof. F. A. Quiocho - whose knowledge, guidance, patience, active participation, and beautiful crystals, have made my graduate experience impossible to imagine as being any better.

-Prof. R. W. Hogg - for results of the sequence analysis in advance of publication.

-Dr. V. K. Mahajan - who provided invaluable help with data collection and processing (July 1974 - Sept. 1976).

-Dr. A. K. Q. Siu - for computer programming and teaching of same to me (Oct. 1974 - Aug. 1975).

-D. Miller and T. Dupuy - for isolating the large amounts of binding protein required to do a structure analysis.

-W. Meador and J. Mays - for help in preparation of heavy atom solutions and other technical aspects.

-R. LaBrum - for help with manual contouring of electron density maps.

-B. Conner - for adaptation of ORTEP software to the visual display terminal.

-The Robert A. Welch Foundation - for a Pre-Doctoral Fellowship.

TABLE OF CONTENTS

VI. APPENDICES

# I. INTRODUCTION

## A. Active Transport and the Role of Binding Proteins

Any living cell must control the substances which enter and exit its cytoplasm. In fact, a cell is defined by its permeability barrier, the membrane. Specific mechanisms are required to regulate the transport of solutes which are essential to the life of the cell. Although transport processes have been studied extensively, relatively little progress has been made in describing the mechanisms of actual translocation of specific molecules across the membrane. The problem is complicated by the fact that transport systems are diverse, often even within the same organism.

Due primarily to their organizational simplicity and genetic accessibility, bacterial transport systems have been extensively studied. However, the evolutionary nature of life makes it exceedingly likely that the basic principles of membrane function will be universal, just as the principles of genetics and protein biosynthesis are universal.

The discussion in this thesis will be limited to bacterial active transport. Reviews on the general subject of transport have appeared (Christensen, 1975; and Oxender, 1972). In addition, a good review limited to bacterial transport has been published by Boos (1974). As reviewed by Boos, bacterial active transport can be subdivided into three biochemically distinct systems. The first type, termed group translocation, is defined as the simultaneous transport and modification of a chemical species. The best studied system of this type is the phosphotransferase system of E. coli. The system can

be described by the following reactions (Kundig and Roseman, 1966):

$$\text{Phosphoenolenolpyruvate} + \text{HPr} \xrightarrow{\text{Enzyme I}} \text{HPr-P} + \text{pyruvate}$$

$$\text{HPr-P} + \text{Sugar}_{\text{outside}} \xrightarrow{\text{Enzyme II}} \text{Sugar-6-P}_{\text{inside}} + \text{HPr}$$

Enzyme I and HPr are cytoplasmic proteins and Enzyme II is a membrane bound complex, and posseses the ligand specificity for the system.

The second type of transport involves strictly components which are embedded in the membrane. Fox and Kennedy (1965) have found a transport system for β-galactosides which requires the participation exclusively of the membrane bound 'M protein'. The M protein has been isolated only in an inactive form. Kaback et al. (1976) have extensively studied other transport systems which remain active after the formation of E. coli membrane vesicles. These include transport systems for tryptophan, proline, glycine, lysine, aspartic acid, asparagine, alanine, serine, and β-galactosides. The accumulation of solutes in these membrane vesicles is coupled primarily to a respiratory chain-dependent oxidation of D-lactate and not to the hydrolysis of ATP per se.

The third type of bacterial active transport involves the participation of high-affinity periplasmic binding proteins which are not firmly attached to the membrane. The main concern of this thesis will be with this type of system. Reviews by Rosen and Heppel (1973) and Oxender and Quay (1976) deal exclusively with these systems and are therefore particularly relevant to this study.

Since the original observation that the release of certain proteins from bacteria by mild osmotic shock was concomitant with loss of transport activity (Heppel, 1969), much research has been instigated regarding the role of the released proteins in transport processes.

The binding proteins released during shock treatment have ligand binding specificities (hence the name binding protein) which match the transport specificities affected by the shock procedure. Table 1 shows the ligands, source, $K_d$, and $K_m$ of transport for all known transport binding proteins. All of the binding proteins isolated so far consist of single polypeptide chains. Heppel (1975) and Wood (1976) have shown that the energy for the bacterial systems which involve high-affinity binding proteins is derived from ATP or other phosphate compounds.

Binding proteins with similar function as the bacterial periplasmic binding proteins have been identified in mammalian systems as well. $Ca^{++}$ binding proteins have been implicated in the transport of $Ca^{++}$ into intestinal cells of chick, rat, dog, cow, and monkey (Wasserman, 1972) and glucose binding proteins have been found in rabbit small intestinal epithelial cells (Catsimpoolas, 1975) and in rat kidney cortex brush border (Thomas, 1973).

The various arguments which implicate the binding proteins in the transport process are as follows:

1. The binding proteins are localized in the cell in the precise place expected for a transport component. The cell envelope of gram negative bacteria consists of a tri-laminar coating. The innermost layer, the cytoplasmic membrane, is a lipid bilayer which constitutes the permeability barrier of the cell. The outermost layer consists of lipopolysaccharide, lipoprotein and lipid, but is not usually considered to be a permeability barrier to small molecules.

Between the two lipid layers lies peptidoglycan. The region between the inner and outer lipid layers which contains the peptidoglycan and

Table 1: Binding Proteins and Their Properties.

| Binding Protein | References* | Source | Molecular Weight | $K_d$ of binding ($10^{-6}$) | $K_m$ of transport ($10^{-6}$) |
|---|---|---|---|---|---|
| Amino Acids- | | | | | |
| Lysine, Arginine, Ornithine | (1) | E. coli | 28,000 | 3, 1.5, 5 | - |
| Arginine | (2) | E. coli | 27,700 | 0.03 | - |
| Histidine | (3) | S. typh. | 25,500 | 0.15 | 0.08 |
| Cystine | (4) | E. coli | 27,000 | 0.01 | 0.02 |
| Glutamine | (5) | E. coli | 27,000 | 0.3 | 0.08 |
| Glutamate and Asparatate | (6) | E. coli | 31,000 | 0.7, 1.2 | - |
| Leucine, Isoleucine, Valine | (7) | E. coli | 36,000 | 1.0 | - |
| Leucine | (8) | E. coli | 36,000 | 0.7 | - |
| Phenylalanine | (9) | Comamonas | 37,000 | 0.2 | 20 |
| Tryptophan | (10) | N. crassa | - | 80 | 50 |
| Sugars- | | | | | |
| Arabinose | (11) | E. coli | 34,000 | 0.3 | 0.3 |
| Galactose | (12) | E. coli | 32,000 | 0.5 | 0.5 |
| Ribose | (13) | S. typh. | 30,400 | 0.33 | - |
| Maltose | (14) | E. coli | 40,500 | 1.5 | - |
| Glucose | (15) | rabbit intestinal brush border | - | - | - |
| Glucose | (16) | rat kidney brush border | - | - | - |
| Glucose-1-P | (17) | A. tume. | - | - | - |
| Ions- | | | | | |
| Calcium | (18) | intestinal mucosa | 24,700 | 26 | - |
| Phosphate | (19) | E. coli | 42,000 | 0.8 | - |
| Sulfate | (20) | S. typh. | 32,000 | 20 | 36 |
| Vitamins- | | | | | |
| Thiamine | (21) | E. coli | - | 0.03 | - |
| Riboflavin | (21) | E. coli | 48,000 | 30 | - |
| Cyanocobalamine | (22) | E. coli | 22,000 | 0.006 | - |

* (1) Rosen, 1971; (2) Rosen, 1973; (3) Lever, 1972; (4) Berger and Heppel, 1972; (5) Weiner and Heppel, 1971; (6) Willis and Furlong, 1975; (7) Penrose et al.,1968; (8) Furlong and Weiner, 1970; (9) Garoff and Bromwell, 1971; (10) Wiley, 1970; (11) Parsons and Hogg, 1974a;(12) Anraku, 1968; (13) Aksamit and Koshland, 1972; (14) Kellerman and Szmelcman, 1974; (15) Catsimpoolas, 1975; (16) Thomas, 1973; (17) Fukui and Miyairi, 1970; (18) Wasserman, 1972; (19) Medveczky and Rosenberg, 1969; (20) Pardee, 1966; (21) Iwashima et al., 1971; (22) Taylor et al., 1972.

perhaps some free space, has been termed the periplasmic space. The sulfate binding protein has been localized in the perimlasmic space by Pardee and Watanabe (1968). They were able to label the binding protein with a dye which can penetrate the outer portion of the cell wall, but not the cytoplasmic membrane. Antibodies were prepared against the purified sulfate binding protein which inactivated the protein *in vitro*. They were not able to inhibit the sulfate binding activity of whole cells. Since antibodies can not penetrate the cell wall, this implies that the binding proteins are not part of the outer cell membrane. Nakane *et al.* (1968) have used the peroxidase histochemical staining procedure for electron microscopy to localize the leucine-isoleucine-valine binding protein in the periplasmic space. The glucose binding protein from rat kidney cortex has been shown to be concentrated in brush border membranes (Thomas, 1973).

2. As previously mentioned, mild osmotic shock treatment causes the simultaneous loss of transport activity and release of binding proteins into the shock fluid. This finding in itself, proves nothing. However, the ligand specificities of the binding proteins exactly matches the specificities of the affected transport. This relationship has been demonstrated for all the periplasmic binding proteins in Table 1. Cells subjected to mild osmotic shock remain viable and are still able to grow in nutrient medium. Note that the substrates for the above binding protein systems are not the same as those observed in vesicle preparations.

3. Ligand binding constants ($K_d$'s) of the purified binding proteins are often similar to the transport constants ($K_m$'s of transport) *in vivo* (see Table 1). The $K_m$'s of transport are almost always slightly greater

than or equal to the $K_d$'s of binding for a particular ligand, which implies that the binding protein may be involved in transport, and furthermore, that the binding of ligand may be the rate limiting event in the transport process.

4. Often competitive inhibitors of the *in vitro* binding of ligand to binding protein are competitive inhibitors of active transport *in vivo* (Oxender, 1972). This has been shown for several cases, in particular for the tryptophan binding protein of *N. crassa*.

5. The cellular level of binding protein and transport activity are co-regulated (Oxender, 1972). In the leucine-isoleucine-valine and L-arabinose systems regulatory mutants were found with derepressed (or elevated) levels of transport and binding proteins. Furthermore, in all cases studied to date, the converse of the above has been found to hold, i.e., mutants which produce little or no binding protein show a corresponding decrease in transport of solute.

6. Mutant strains with structurally defective binding proteins are unable to carry out transport. A point mutation in the structural gene coding for the histidine binding protein has resulted in a binding protein with defective histidine binding and defective transport (Ames and Lever, 1970). Similar findings were observed for the D-galactose binding protein by Robbins and Rotman (1975) and Boos and Servas (1970). Moreover, revertants of such binding-protein-negative mutants have been found which have normal transport and binding proteins similar to the wild type.

Besides the binding proteins, there are other components associated with the shock-sensitive transport systems. Silhavy *et al*(1974) have reported that two other components of the system which requires the D-galactose binding protein have been identified by comparing

two-dimensional polyacrylamide gels of membrane fractions from cells grown under inducing versus non-inducing conditions for the transport system. These membrane bound proteins have apparent molecular weights of 80,000 and 50,000. Recent genetic studies by Ames and Spudich (1976) strongly indicate that the histidine binding protein (the J protein) from S. typhimurium contains two binding sites. One site is required for the binding of histidine, and the other for interaction with another protein component which is presumably membrane bound (P protein). A mutant with J protein defective in a region separate from the histidine binding site can function in histidine transport only if a compensating mutation is introduced in the gene which produces the P protein.

Although the necessity of the periplasmic binding proteins is now well established, little is known about the actual mechanisms by which solute is translocated across the membrane. So far, mechanisms which have been discussed for the role of binding proteins in transport invoke conformational changes in the binding protein, either to promote a vectorial movement of substrate, or to trigger some other component to respond. Much effort has been expended trying to demonstrate conformational changes for the binding proteins in solution.

Many studies, utilizing techniques of fluorescence, electrophoresis, optical rotary dispersion, circular dichroism, infrared spectroscopy, proton magnetic resonance, and chromophoric probes have attempted to demonstrate conformational changes, with conflicting results.

B. Properties of the L-Arabinose Binding Protein

The subject of this thesis is the periplasmic L-arabinose binding protein from E. coli. The protein has been well characterized both biochemically and genetically by Novotny and Englesberg (1966) and Parsons and Hogg (1974a). The synthesis of the binding protein is controlled by the araC gene. This is the same gene which contains the L-arabinose operon. Further genetic studies have shown that mutants lacking the high affinity transport system for L-arabinose have been found which also lack detectable levels of the binding protein (Hogg, 1971). The amino acid sequence has been determined recently by Hogg (personal communication) and is given in Figure 1. Based on the sequence analysis, L-arabinose binding protein has a molecular weight of 34,000 and 306 amino acid residues. The protein contains an unusual amount of methionine (10) and proline (15), and sequences of pro-pro at 277-278, and arg-arg-arg at 148-150.

As with other binding proteins, L-arabinose binding protein may undergo a conformational change upon ligand binding. This has been suggested by the fact that significant changes in the fluorescence emission spectrum occur upon ligand binding. However, this may be due to direct quenching of the fluorescence by substrate.

Miller (1976) has shown that the sugar binding site of L-arabinose binding protein probably lies near the single cysteine residue. The addition of arabinose protects the protein from reaction with mercurinitrophenol, and vice versa, presumably by direct steric interference.

The L-arabinose binding protein has been shown to be similar to the D-galactose binding protein. They have almost identical similar molecular weights and their binding specificities overlap. L-Arabinose binding protein binds L-arabinose ($K_d$ = 3 x $10^{-7}$ $M^{-1}$) and D-galactose

($K_d$ = 4 x $10^{-7}$ $M^{-1}$), whereas D-galactose binding protein binds D-galactose ($K_d$ = $10^{-7}$ $M^{-1}$) and D-glucose ($K_d$ = $10^{-7}$ $M^{-1}$). Furthermore, antibodies prepared against L-arabinose binding protein cross-react with D-galactose binding protein and vice versa (Parsons and Hogg, 1974b).

glu - asn - leu - lys - leu - gly - phe - leu - val - lys -
gln - pro - glu - glu - pro - trp - phe - gln - thr - glu - 20
trp - lys - phe - ala - lys - asp - ala - gly - lys - asp -
leu - gly - phe - glu - val - ile - lys - ile - ala - val 40
pro - asp - gly - glu - lys - thr - leu - asn - ala - ile -
asp - ser - leu - ala - ala - ser - gly - ala - lys - gly - 60
phe - val - ile - cys - thr - pro - asp - pro - lys - leu
gly - ser - ala - ile - val - ala - lys - ala - arg - gly - 80
tyr - asp - met - lys - val - ile - ala - val - asp - asp -
gln - phe - val - asn - ala - lys - gly - lys - pro - met - 100
asp - thr - val - pro - leu - val - met - met - ala - ala -
thr - lys - ile - gly - glu - arg - gln - gly - gln - glu - 120
leu - tyr - lys - glu - met - gln - lys - arg - gly - trp -
asp - val - lys - glu - ser - ala - val - met - ala - ile - 140
thr - ala - asn - glu - leu - asp - thr - ala - arg - arg -
arg - thr - thr - gly - ser - met - asp - ala - leu - lys - 160
ala - ala - gly - phe - pro - glu - lys - gln - ile - tyr -
gln - val - pro - thr - lys - ser - asn - asp - ile - pro - 180
gly - ala - phe - asp - ala - ala - asn - ser - met - leu -
val - gln - his - pro - glu - val - lys - his - trp - leu - 200
ile - val - gly - met - asn - asp - ser - thr - val - leu -
gly - gly - val - arg - ala - thr - glu - gly - gln - gly - 220
phe - lys - ala - ala - asp - ile - ile - gly - ile - gly -
ile - asn - gly - val - asp - ala - val - ser - glu - leu - 240
ser - lys - ala - gln - ala - thr - gly - phe - tyr - gly -
ser - leu - leu - pro - ser - pro - asp - val - his - gly - 260
tyr - lys - ser - ser - glu - met - leu - tyr - asn - trp -
val - ala - lys - asp - val - glu - pro - pro - lys - phe - 280
thr - glu - val - thr - asp - val - val - leu - ile - thr -
arg - asp - asn - phe - lys - glu - glu - leu - glu - lys - 300
lys - gly - leu - gly - gly - lys.

Figure 1. The amino acid sequence of L-arabinose binding protein (Hogg, unpublished, Nov. 1976, personal communication).

C. Chemotaxis and the Role of Binding Proteins

Some shockable binding proteins have also been shown to be essential in chemotaxis (Hazelbauer and Adler, 1971; Aksamit and Koshland, 1972), a process whereby organisms move toward or away from a chemical stimulus (Adler, 1974). In bacteria, chemotaxis can be conceptualized as occurring in three basic steps:

1. Recognition of the stimulus by a chemoreceptor. The same binding proteins which are essential to transport are the recognition components of this simple behavioral stimulus-response system. E. coli exhibits chemotaxis toward galactose but not arabinose. However, the similarities between L-arabinose binding protein and D-galactose binding protein in structure and transport activity warrant discussion of this phenomenon. Transport and chemotaxis toward a single chemical are not dependent on each other, i.e., some mutants have normal chemotaxis but not transport and vice versa (Adler, 1975). Therefore, the two systems are generally independent, but share a common chemoreceptor.

2. Transmission of the information from chemoeffectors to the flagella. Little is known about this aspect of chemotaxis, however other protein components have been shown to be necessary.

3. Appropriate response of the motile system (flagella).

Bacteria move into or out of chemical gradients by controlling the rate at which they stop and randomly reorient themselves. It is known that forward motion of a bacterium is created by counterclockwise rotation of its flagella, whereas the organism twiddles (reorients) if the flagella turn clockwise (Larsen et al., 1974). Apparently the binding of chemicals to the binding proteins trigger some system which controls the direction of rotation of the flagella. When moving

into a gradient of attractant they twiddle less frequently, hence moving further into the gradient. If there is no gradient, or if they are moving down a gradient, then they twiddle at a normal basal rate.

Bacteria also modulate their twiddling in temporal gradients (Koshland, 1974), implying that the organism does not receive signals from its 'head' and 'tail', but somehow remembers what the chemical level of the stimulus was a few seconds earlier, implying a 'kinetic' memory. Details of the mechanism of transfer of information about the stimulus are not at all clear. However, it is known that some protein component which lies mechanistically between the receptor (binding protein) and the flagella must be methylated to retard the twiddling rate (Armstrong, 1972).

D. Statement of Purpose

The purpose of this study is to examine the three-dimensional structure of the L-arabinose-binding protein from E. coli using the technique of X-ray crystallography. L-arabinose-binding protein is the first protein in this class to be crystallized which is suitable for atomic resolution structure analysis. The understanding of the roles of binding proteins in transport and chemotaxis at the molecular level requires the elucidation of the three-dimensional structure of at least one of these proteins. X-ray crystallography is the only technique which can reveal the nature of the polypeptide chain folding and conformation of a protein. This technique can also provide some understanding of the stereochemistry of binding and any concomitant conformational changes which may occur. This thesis will present the first structural characterization of L-arabinose-binding protein and report the results at 3.5 Å resolution. The results will be examined in the light of the biochemical and genetic studies described previously, and will form a basis for comparison of future structural studies of other binding proteins.

## II. MULTIPLE ISOMORPHOUS REPLACEMENT

A complete and detailed description of basic crystallographic methods will not be presented, since such a description would be lengthy, and excellent books are available (Stout and Jensen, 1968; Blundell and Johnson, 1976). However, the method of multiple isomorphous replacement (MIR) is of particular interest to crystallographers studying macromolecules, and since the technique was utilized exclusively in this study, a brief description of the theory will be presented. A thorough discussion of the MIR technique can be found in the book by Blundell and Johnson (1976), and in a nice review by Phillips (1966).

Briefly stated, the determination of the structure of any crystalline molecule by means of X-ray crystallographic technique entails the determination of an image of the electron density in the crystal using the well known Fourier transform:

$$\rho(x,y,z) = \frac{1}{V} \sum_{hkl}\sum\sum \hat{F}_{hkl} \exp\{-2\pi i(hx+ky+lz)\} \qquad (1)$$

where $\rho(x,y,z)$ is the electron density in the unit cell of volume, V, x,y, and z are the coordinates in the unit cell, and $\hat{F}_{hkl}$ is the structure factor. The structure factor can be further broken down as follows:

$$\hat{F}_{hkl} = |\hat{F}_{hkl}| \exp(-i\alpha_{hkl}) \qquad (2)$$

where $\hat{F}_{hkl}$ is related to the experimentally determined diffraction intensity and is complex, and $\alpha$ is the phase, which is initially unknown. Unfortunately, when one records the intensity at a given reciprocal lattice point, one loses the phase, $\alpha$, of the structure

factor because one is recording the intensity, $I_{hkl}$, which is proportional to $\hat{F}_{hkl} \times \hat{F}^{*}_{hkl}$, and is a real number. Thus the situation is that the magnitudes of the coefficients in the Fourier synthesis of the structure are known, but not the phases. Due largely to the contributions of Bokhoven et al. (1951), Bragg and Perutz (1958), Perutz (1963), and Kendrew et al. (1960), the method of multiple isomorphous replacement has been devised to provide a solution of the phase problem for macromolecular analysis.

The five steps involved in this method are as follows:

1. Crystallization of the protein and preparation of a series of at least two, preferably more, crystalline derivatives adequately isomorphous with the parent protein crystals and containing additional atoms, heavy enough to be useful in the phase determination.
2. Collection of diffraction data for the parent and derivative crystals to the desired resolution.
3. Determination of the positions of all heavy atoms within the unit cell, initially using some form of difference Patterson or Fourier synthesis. The difference Patterson synthesis is first utilized to locate heavy atoms within the unit cell. This synthesis is a modification of the original Patterson synthesis (Patterson, 1935):

$$P(u,v,w) = \sum_{hkl}\sum\sum |\hat{F}|^2 \exp\{-2\pi i(hu+kv+lw)\} \qquad (3)$$

The result of using the coefficients $\hat{F}_{hkl}^2$ instead of $\hat{F}_{hkl}$ as in the electron density synthesis is that the function has peaks at locations corresponding to the heads of all interatomic vectors with their tails at the origin. It follows that if there are N atoms in a structure, then there are $N^2$ peaks in a Patterson function

unit cell, N of which will be at the origin (i.e., the zero vectors produced by the vector between an atom and itself). If the number of atoms is small, the Patterson function can usually be interpreted, producing a trial set of atomic positions. In solving for the heavy atom positions in an isomorphous derivative protein structure, one can eliminate all the peaks in the Patterson function due to non-heavy atoms by using the coefficients;

$$\Delta F^2 = (|\hat{F}_{PH}| - |\hat{F}_P|)^2$$

where $|\hat{F}_{PH}|$ is the magnitude of the derivative structure factor, and $|\hat{F}_P|$ is the magnitude of the native structure factor. Although this is not rigorously correct, i.e., the phases of $\hat{F}_{PH}$ and $\hat{F}_P$ are not identical, the approximation is usually good enough to allow the solution to be found if the number of heavy atom sites is small. Initial heavy atom parameters can sometimes be solved from the easily obtainable centrosymmetric difference Patterson projections. After centric projection, heavy atom analysis is then confirmed in three-dimensional difference Patterson maps.

Provided that initial phases are available (from Patterson synthesis), an independent check of the validity of the solution of the heavy atom positions can be obtained by computing difference Fourier maps, using coefficients;

$$\Delta F = ||\hat{F}_{PH}| - |\hat{F}_P||$$

in equation (1). Again, this is not rigorously correct, as the phases are not, in general, identical for the $\hat{F}_{PH}$ and the $\hat{F}_P$. But the results are generally valid. Other lower substitution sites and/or sites in derivatives with large numbers of different sites which were unable to be solved by difference Patterson techniques can

usually be revealed by difference Fourier analysis.* After the heavy atoms have been located properly, refinement of the various heavy atom parameters is necessary. This includes the adjustment of parameters describing position, occupancies (to allow for incomplete substitutions), temperature factors, and scale factors for derivative data in relation to the native protein data. This step is done in conjunction with Step 4 as described below.

4. Calculation of the contribution of the heavy atom scattering for each derivative to the structure factors of each reflection, and their use with the measured intensities to determine the phase angles of the parent enzyme structure factors. Harker (1956) has presented a graphical method (phase diagram) for the determination of the phases of the native enzyme utilizing three sets of data along with the positions of the heavy atoms. It has been formulated as follows. The basic equation for calculating the intensity of a given diffracted ray, $I_{hkl}$, from a crystal lattice is:

$$I_{hkl} \propto |\hat{F}_{hkl}|^2 = |\Sigma_n f_n \exp\{2\pi i(hx_n+ky_n+lz_n)\}|^2 \qquad (6)$$

where $I_{hkl}$ and $\hat{F}_{hkl}$ are as defined previously, $f_n$ is the x-ray scattering power of atom n, and $x_n$, $y_n$, $z_n$ are the positions of the atom n, along the axes of the unit cell and are expressed in fractions of the unit cell length. If one breaks down equation (6) into 'heavy' and

---

* This difference Fourier technique can also be utilized to reveal the binding sites of ligands and to detect subtle differences (if any) in the protein structure under various conditions, for example, the conformational changes upon ligand binding. By subtracting the native structure factor amplitudes from the perturbed ones, one obtains a map where new scattering electrons show up as positive peaks, and places where electrons are no more, show up as negative peaks.

'light' atoms, the following results are obtained:

$$\hat{F}_{PH} = \hat{F}_{H(\text{heavy atoms})} + \hat{F}_{P(\text{light atoms})} \tag{7}$$

Hence, when the heavy atom positions are known within the unit cell, $\hat{F}_H$ can be calculated. However, only the magnitudes are known for the complex numbers, $\hat{F}_{PH}$ and $\hat{F}_P$. An Argand diagram (Figure 2 ) best describes the above situation. Since $\hat{F}_{PH}$ must equal the sum of $\hat{F}_P + \hat{F}_H$ (equation (7)) only two solutions for the phases (real and complex components) of the $\hat{F}_{PH}$ and $\hat{F}_P$ rays are possible, namely, the intersections of the two 'magnitude' circles. Note that if the reflection is centric, i.e., symmetry dictates that $\alpha = 0, \pi/2, \pi$, or $3\pi/2$, then the phase is determined from this single derivative. If a second heavy atom derivative is used, and its calculated $\hat{F}_H$ is different from the first derivative's, a unique solution exists for the phases of the structure factors $\hat{F}_{PH1}$, $\hat{F}_{PH2}$, and most importantly $\hat{F}_P$, (Figure 3.). Hence equation (1) can be applied to calculate an electron density map.

In practice, all three 'magnitude' circles rarely intersect at a unique point, due to inaccuracies in measuring the diffraction intensities and in determining the heavy atom parameters, and possibly lack of strict isomorphism. Hence, the graphical method is not always clear cut and is inadequate especially when many reflections have to be considered. The method of Harker has been superceded by the method of Blow and Crick (1959) which makes possible a more rigorous treatment of errors. Blow and Crick have derived the following expression for the probability that a particular native phase $\alpha$,

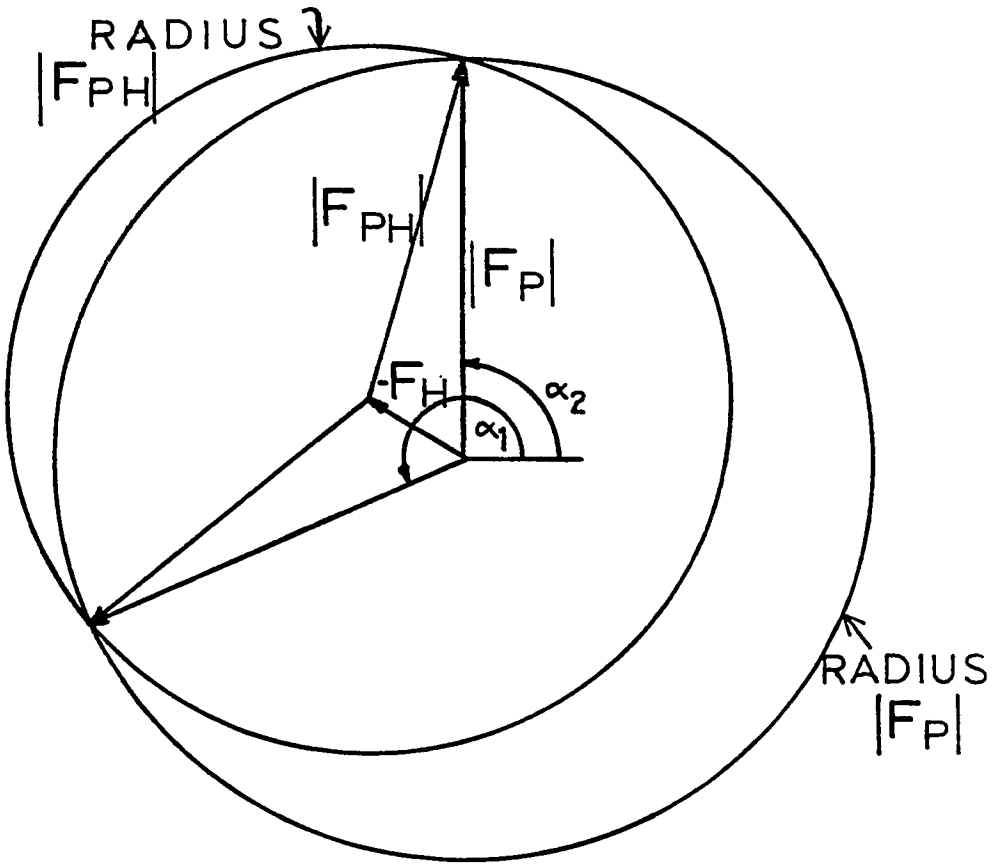



Figure 2: Phasing by an isomorphous derivative. (Harker, 1956).

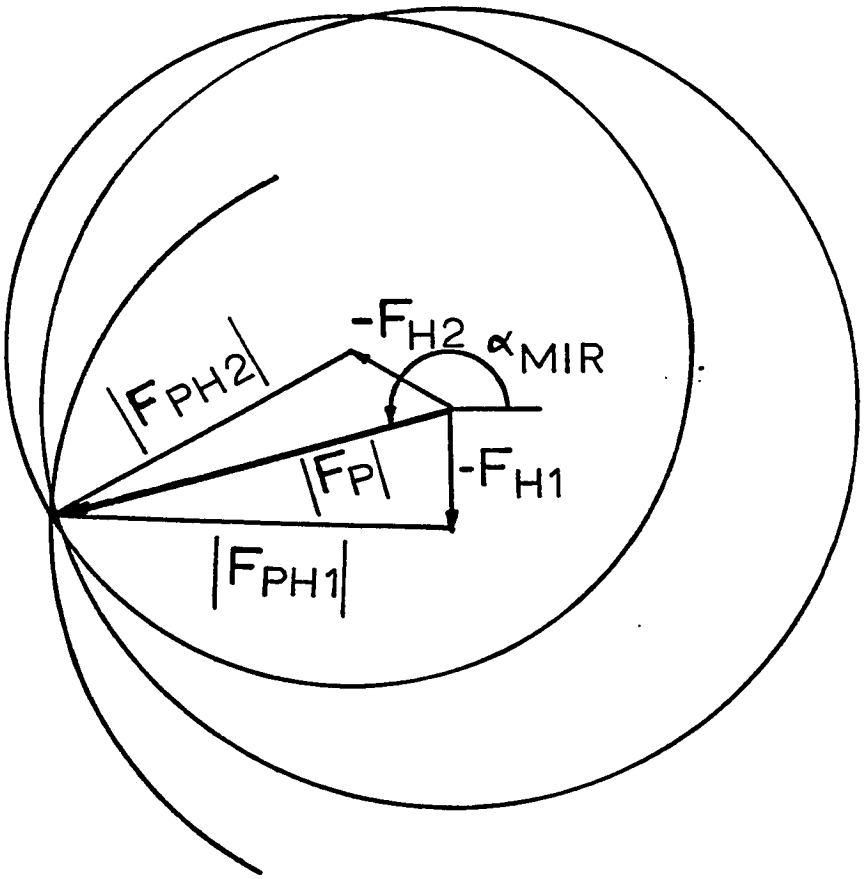



Figure 3: Multiple Isomorphous Replacement.

is correct.

$$P(\alpha) = \exp(-\sum_i w_i d_i^2 / E_i^2) \tag{7}$$

where $w_i$ the weight for derivative i, $d_i$= the lack of closure of the phase triangle, and $E_i$= the standard deviation of $F_{PHi}$. The d's are illustrated graphically in Figure 4. The E's are estimated for the first calculation, thereafter they are taken as the RMS lack of closure from given range of $\sin\theta/\lambda$. The entire range of $\sin\theta/\lambda$ is divided into annuli for this purpose.

The phase chosen is usually the phase at the centroid of the probability distribution, however, one may also specify that the phase with the highest $P(\alpha)$ be chosen. The former method results in 'best' Fourier coefficients, whereas the latter are termed 'maximum probability' phases. The centroid is defined as:

$$\hat{F}_P^{cen} = \int_0^{2\pi} P(\alpha) \, |\hat{F}_P| \exp(i\alpha) d\alpha \; / \; \int_0^{2\pi} P(\alpha) d\alpha. \tag{8}$$

The quantity,

$$m = \hat{F}_P^{cen} \; / \; \hat{F}_P \tag{9}$$

has been termed the figure of merit, and is the weighted mean of the cosine of the error in the phase angle.

As mentioned previously, it is necessary to refine the heavy atom parameters in order to obtain accurate phasing. One can refine on these values by minimizing the residual,

$$R = \sum_{hkl}\sum\sum w_{hkl}\{|\hat{F}_{PH}|^2 - (|\hat{F}_P|\exp(i\alpha) + \hat{F}_H)^2\} \tag{10}$$

where hkl means sum over all reflections, $w_{hkl}$ is the weight of

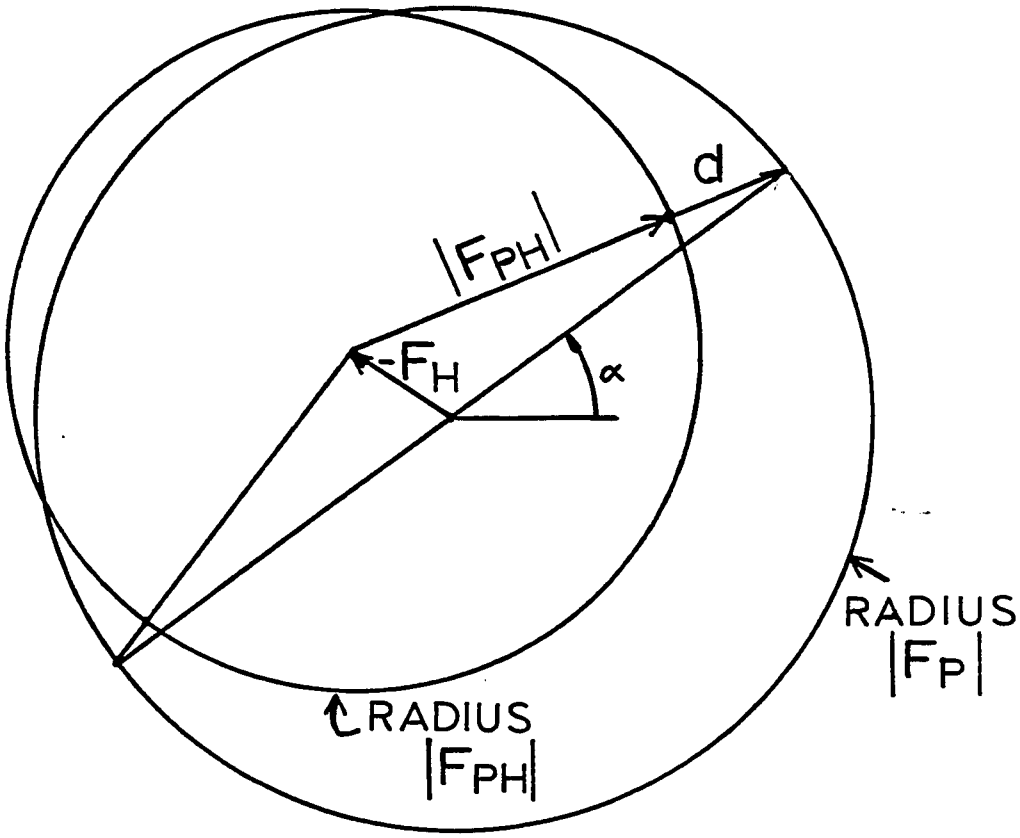



Figure 4 : Lack of closure, d, in the phase triangle.

reflection hkl, and the other variables are as described earlier.

The minimization is achieved by differentiating equation (10) with respect to the parameters of $\hat{F}_H$ and setting the results to 0. Usually, one builds a square matrix with rank equal to the number of parameters to be refined, and the matrix is inverted to find the required parameter shifts.

Alternately, one may consider only the diagonal elements, i.e., all parameters are assumed to be independent of the others. This approximation allows faster and cheaper calculation because the number of terms to be calculated is small and the matrix inversion is trivial.

Thus step 4 can be alternated with step 3, i.e. more accurate phases can be calculated followed by more accurate determination of the heavy atom parameters. The alternation is continued until no significant changes occur in the phases relative to the last cycle.

The least squares procedure will then reduce the 'lack of closure' of the phase triangles by minimizing the sum of the squares of the differences between the calculated and observed heavy atom derivative intensities. During the course of the refinement, the 'root mean square lack of closure error' and the 'root mean square heavy atom contribution' are monitored as functions of Bragg angle. From the results of a number of derivatives, the best value of the phase for a particular structure factor is determined and the 'figure of merit' given to represent the reliability of the phase determination.

5. Calculation of an electron density map using the 'best' phases with amplitudes weighted by the figure of merit, or using the 'maximum probability' phases.

# III. METHODOLOGY

## A. Purification and Crystallization of L-Arabinose Binding Protein

The L-arabinose binding protein was isolated and purified from E. coli B/R U1041 (araA39) by the method of Parsons and Hogg (1974). Crystals were obtained from protein solutions in dialysis bags immersed in 50% 2-methyl-2,4-pentanediol, 10mM potassium phosphate, pH 6.4 (Quiocho et al., 1974). A photomicrograph of the typically good crystals is shown in Figure 5. Crystals as large as 5mm in length have been obtained. It is interesting to note that crystals formed during the first crystallization are well-formed, but do not diffract to high resolution. Crystals produced by recrystallization are not so well formed, but diffract beyond 2 Å resolution.

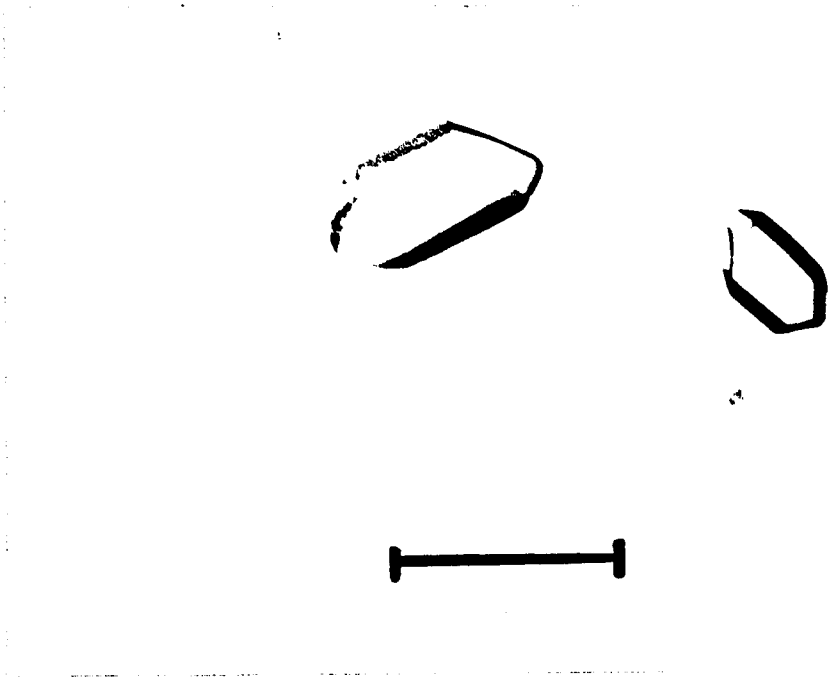


Figure 5: Crystals of L-arabinose binding protein. The bar represents 1mm.

B. Preparation of Heavy Atom Derivatives

Derivatives were prepared in the usual manner of soaking crystals in standard solutions containing heavy atom compounds. The standard solution consisted of 60% 2-methyl-2,4-pentanediol buffered with either 10mM potassium phosphate or sodium maleate, pH 6.4. Some heavy atom compounds formed cloudy precipitates in potassium phosphate whereas in maleate the solutions were clear. Because of this problem, sodium maleate was adopted as the standard buffer. In order to check the effect of different buffers, a comparison was made in both reciprocal space and real space between native crystals in the phosphate and maleate buffers. No significant differences were found.

C. Data Collection

Since diffraction data was obtained by both film and diffractometer methods, both techniques will be described. For either method crystals were mounted in thin-walled glass capillaries, using the surface tension of the mother liquor to prevent slippage of the crystal. A piece of filter paper moistened with soaking solution was placed in the end of the capillary before sealing it with bee's wax.

1. Film Methods

The x-ray source employed was an Elliot GX-6 rotating anode generator with a copper target. The only monochromating device used was a nickel filter. The wavelength of radiation (Cu $K_\alpha$) was taken to be 1.5418 Å. The exit collimator installed for any given experiment was the one just large enough to fully bathe the crystal in the x-ray beam.

The cameras used were two identical Nonius-Enraf model Y 925 precession cameras. The only modification of the cameras consisted of the addition of an x-y translation adjustment device on the layer line screen holder, to allow for accurate centering of the layer line screen. Layer line screen settings were calculated in the usual fashion except for a slight error correction in the crystal to layer line screen distance peculiar to each camera. Film packs consisted of 1, 2, or 3, 5" x 5" pieces of Ilford Industrial x-ray film. A stiff cardboard square placed in front of the film was the best way found to hold the film flat in the cassette. The films were developed using standard procedures, with the addition of a constant temperature device and nitrogen bubble bursts to ensure even development.

Films were scanned on a Syntex AD-1 densitometer. Standard

procedures were used except for a modification in the way background intensities were calculated. The total background was taken to be the average of all background samplings for a given spot, instead of including only the least intense samplings. A 0.02 × 0.01 cm spot size was used in all cases. The area scanned was adjusted to match the spot size of a particular pack of films. The data thus obtained was stored on magnetic tape and further processed using the large university computer (IBM 370/155) as described in the section on data processing.

2. Diffractometer Methods

All data except that for some from initial heavy atom searches was collected with an automated diffractometer. A Syntex Model $P2_1$ was used, equipped with a Syntex extended front collimator whose exit slit is 0.635 cm from the crystal, instead of the standard 5.0 cm.* This decreases the background scattering, especially at low values of 2θ.

In order to further increase the peak to background ratio of counts at the detector, a beam tunnel 62 cm long was placed 2 cm from the crystal. The pressure in the tube was reduced to about 60 torr. Diffraction data was collected on the Syntex supplied test crystal with and without the evacuated tube in place. Of 193 reflections monitored, the sum of all background counts with the tube in place was diminished by 43% relative to the counts with no tube. The sum of all peak heights was diminished by only 14.7%. This is equivalent to an increase of 51% in the peak height to background

---

* This was added in November of 1975. Some data was collected without the benefit of this device.

ratio. The radiation was derived from either a 1500 watt Syntex or a 1200 watt Philips sealed copper tube. The tubes were operated at the maximum power at all times. Nickel filtering was used as the only monochromating device. Again the wavelength of radiation (Cu $K_\alpha$) was taken to be 1.5418 Å. Tube alignment was checked periodically and adjusted as required.

Because the crystals have a regular external morphology, they could be mounted consistently with a* parallel to the diffractometer's Ø axis. The crystal, once optically aligned, could usually be brought to the condition for diffraction for an axial reflection with less than 1° or 2° of movement of the goniometer head relative to the original optical setting. The automatic centering routine was then used to maximize the diffraction. By further adjusting the goniometer head settings, one could align a* parallel to the Ø-axis to within 0.03°. Once this has been done, another reflection of known 2θ, ω, and χ can be found by carrying out a Ø-scan. The known location of this second reflection, taken with approximate unit cell parameters, is enough to fully determine the crystal's orientation relative to the machine's coordinate system, i.e., a trial orientation matrix can be obtained. Precise unit cell parameters of each crystal were obtained by centering on 10 to 15 intense reflections of varying 2θ and applying the least squares program supplied by Syntex.

With the crystal alligned as described above, measurements necessary for a semi-empirical absorption correction (North et al., 1968) become easy to obtain. The absorption correction consists of a Ø-independent function derived from the measurements of the net

intensity of an axial reflection (6,0,0) at 10° intervals in $\phi$, from 10° to 350°. Theoretically those measurements 180° apart should be in close agreement. Crystals with disagreements larger than 10% were generally discarded, although sometimes the completion of a data set required the use of a marginal crystal. For these cases, the quality of data (see Section III.D. Data Processing) was monitored carefully to ensure that no significant errors were introduced into the data set. The absorption correction itself was found to affect the intensities from 5% to 100%. The diffractometer software was modified to measure the necessary absorption correction parameters automatically. See Appendix I for details of this modification.

Diffraction intensities were measured using the standard $\omega$ step-scan (Wyckoff et al.,1967) program as supplied by Syntex. For each reflection seven steps were taken over a total range of 0.2°, and the intensity was taken as the sum of the highest 3 consecutive steps. The scan rate is based on a preliminary fast scan through the reflection and was allowed to vary from 0.3°/minute to 0.6°/minute. Background counts were taken for time equal to the scan time 1° away from the reflecting $\omega$ on each side of the peak.

Five reflections of different magnitudes and $2\theta$ values were measured every 150 reflections to monitor crystal decay. Data collection was terminated for a given crystal when the intensity of any check reflection dropped below 80% of its initial value. These monitors were further used to correct for the time-dependent, x-ray induced decay of intensities.

D. Data Processing

Digitized film data from the densitometer was processed on the university's computer, using a version of the LPETC program written by D. C. Wiley at Harvard University and modified to be compatible with the Syntex tape format.

The output magnetic tape from the diffractometer contains reflection intensities corrected only for measured background and the variable scan rate. Other correction factors for Lorentz, polarization, absorption, and deterioration effects were applied using the university's computer. Also, redundant data collected from one crystal were averaged at this time. The output from this program is written on tape with a format compatible with the Crym System (Lipscomb et al., 1966), a complete package of programs for protein and small molecule crystallography. The Crym System was not modified except to fix known programming bugs.

Data from different crystals were correlated using the Difcor routine of Crym. The R-factors which are indicators of the quality of data were calculated by this routine and are given in Tables 2 and 5. The Crym System was also used for all least squares refinement and Fourier calculations.

E. Refinement

In this analysis, the fastest convergence was achieved usually by refining parameters in the following way. First, only the derivative scale factors were refined, with all other approximate parameters held fixed. Usually two or three cycles using full shifts were sufficient to arrive at converged scale factors. These scale factors were not found to deviate substantially during further refinement.

Next, population and positional parameters, P, x, y, and z, were allowed to vary also, with shift factors of 0.5. The population and coordinate parameters for a given derivative were put in a separate block matrix. Once these parameters converged (usually in two or three cycles), the temperature factors could also be refined by using a diagonal matrix. Full or block matrix refinement at this stage tended to make the temperature factors and population parameters diverge. Some of the temperature factors had to be constrained because they refined to negative values. When this occurred, the temperature factor was set to 10.0 and was not refined in subsequent refinement cycles. Heavy atom sites whose population dropped below about 10% were deleted.

Convergence of all the above parameters could usually be achieved in about three to four more cycles. Next, the temperature factors could be put in a block matrix and refined with the other parameters again. After all the parameters had stabilized (three to four more cycles), true full matrix refinement could be carried out successfully. However, the parameters were usually sufficiently converged such that full matrix refinement resulted in no significant parameter shifts.

F. Electron Density Maps

Actual contour maps of the Fourier calculations were produced either by tracing onto clear plastic directly from Crym contour output maps or else were computer drawn using a Calcomp 1136 drum plotter and reproduced onto Xerox transparencies. The software for generating the plot tapes from Crym format Fourier grids was written following the algorithm of Dayhoff (1964).

In order to directly compare model structures to the electron density map, an optical device of the type described by Richards (1968) was constructed. This consists of a large half-transparent mirror with the electron density map illuminated from behind the mirror and the model in front of the mirror. The result is a superposition of the images of the model and the electron density map.

The stereographic drawings of the molecule were plotted on the Calcomp plotter using ORTEP software (Johnson, 1965). To expedite finding a 'good view' of a particular feature to be illustrated, the ORTEP software was adapted to a Tektronix graphics display terminal. Thus many views of the molecule could be seen in a few minutes.

# IV. CRYSTALLOGRAPHIC ANALYSIS OF L-ARABINOSE BINDING PROTEIN

## A. Initial Characterization

Precession photographs of L-arabinose binding protein crystals (Figure 6) show that the space group is $P2_12_12_1$ by displaying orthogonal axes and systematic extinctions of the type h=odd along h00, k=odd along 0k0, and l=odd along 00l. There is one protein molecule per asymmetric unit. The unit cell dimensions of the native protein in maleate buffer are a=55.46(09) Å, b=71.82(08) Å, and c=77.76(13) Å. The unit cell dimensions of all the heavy atom derivative crystals are given in Appendix B. All derivative crystals were found to be isomorphous on the basis of unit cell parameters.

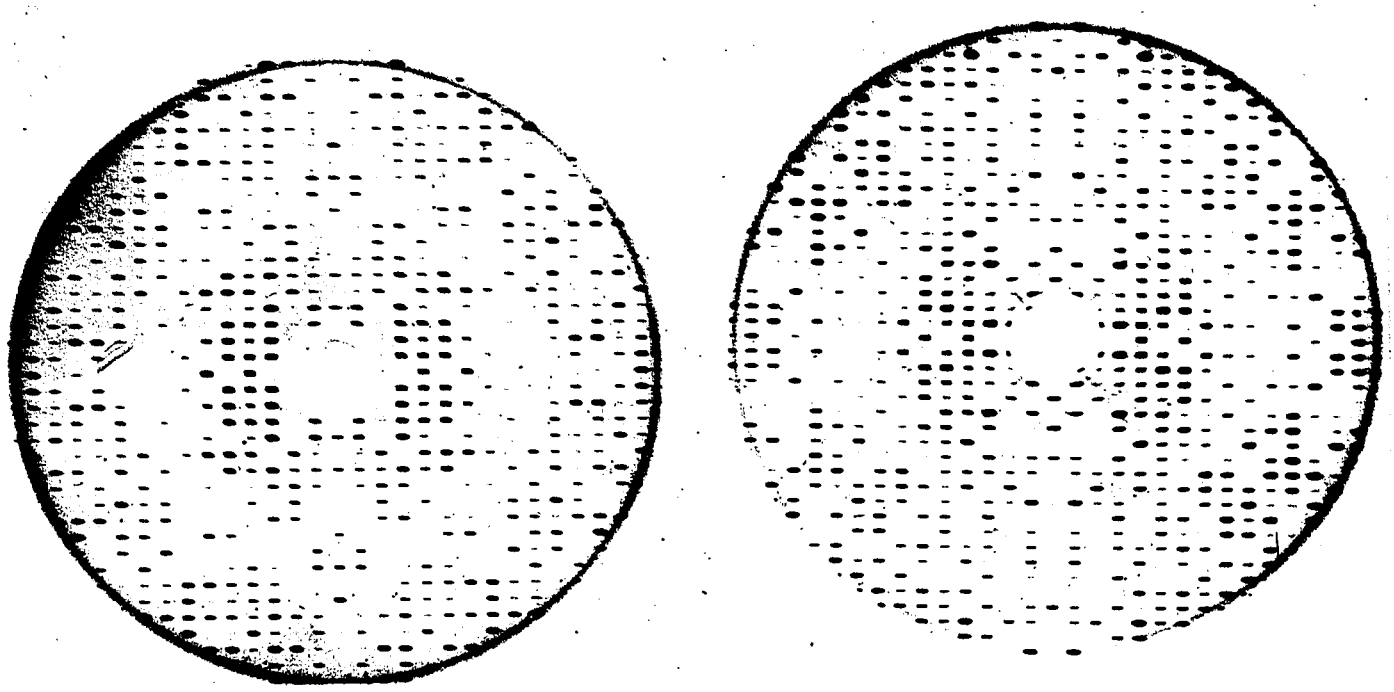


Figure 6: Precession photos (μ=12°) for hko (left) and h0l zones (right).

B. Solution of Heavy Atom Derivatives

The experimental conditions for attempts to obtain heavy atom derivatives are given in Appendix C. Many attempts were made, but discussion will be confined to those which resulted in intensity changes relative to native, and whose difference Patterson and difference Fourier maps showed heavy atom substitution. Prior to the availability of the diffractometer, searches for heavy atom derivatives were carried out using film technique as described in the section on film methods. Difference Fourier maps and difference Patterson maps for all successful derivatives are given in Appendix D along with their interpretations.

Photographs of hk0 and h0l projections from crystals soaked in solutions of mercurinitrophenol showed slight intensity changes when compared to native photographs. From difference Patterson maps calculated in these projections, the position of one Hg atom per protein molecule was solved. This led to the investigation of the 0kl projection. The values obtained for the position of the heavy atom were mutually consistent for all three projections.

Large intensity differences were found in crystals soaked in $K_2PtCl_6$. However, the difference Patterson map showed many peaks which were not interpretable. At this point, refinement was carried out using the single derivative, mercurinitrophenol. As described previously in the section on theory, one derivative is sufficient to determine the phase of a centric reflection. Using the phased hk0, h0l, and 0kl data, a difference Fourier map was calculated in projection, using coefficients $|\hat{F}_{Pt}|-|\hat{F}_{NATIVE}|$ and native phases. However, the resulting peaks did not yield atomic positions which

were consistent in the different projections. An iridium derivative, $K_2IrCl_6$, also showed slight intensity differences using the film technique, however it, too, was not solvable using difference Patterson or projection Fourier methods.

Following these initial attempts to obtain heavy atom derivatives, and for the next several months, many heavy atom compounds were tried, but none showed changes in intensities relative to native. At this point, the automatic diffractometer was delivered to the laboratory and the three-dimensional data were recollected from the native and mercurinitrophenol derivative crystals*. Subsequently, data was collected from crystals soaked in $NaAuCl_4$. There existed a single site in difference Fourier maps calculated for this derivative using the mercurinitrophenol derived centrosymmetric phases. The x,y, and z, were consistent in all three projections. Centrosymmetric phases based on this single Au site could indeed reestablish the location of the original Hg site in the mercurinitrophenol derivative, providing evidence that the sites were real. Three-dimensional phases derived from the Au and Hg derivatives allowed the solution by difference Fourier of two other derivatives, $(NH_4)_2PtCl_4$ and the previously unsolvable $K_2IrCl_6$. At this point, initial MIR phases were calculated using the mercurinitrophenol, Au, Pt, and Ir derivatives and a few cycles of refinement of all parameters were undertaken. These phases were then used to recalculate difference Fourier maps for each of the

---

* Three-dimensional data had been collected for native and mercurinitrophenol crystals using the film technique, however the quality of the data was poor, as evidenced by high values of the residual, R, which were 11% and 16%, respectively.

derivatives, which revealed the existence of additional, less substituted heavy atom sites within each derivative. The phase problem having been solved, attention was turned to further refinement and improvement of heavy atom parameters for the purpose of obtaining the best phases.

C. 5Å Resolution Structure Analysis

Initial refinement was carried out at 3.5 Å resolution employing the $(NH_4)_2PtCl_4$, $K_2IrCl_6$, $NaAuCl_4$, and mercurinitrophenol derivatives using the methods described in the section on refinement. The soaking conditions used and high quality of heavy atom derivative data is given in Table 2. The overall scale factor was initially obtained from a Wilson plot of the native data, but was subsequently reduced to give an occupancy of about 100% for the most substituted platinum site. Initial scale factors for the derivatives were obtained by scaling the average $|\hat{F}_{PH}|$ to the average $|\hat{F}_P|$ times 1.05, to allow for heavy atoms in the unit cell. The figure of merit reached a plateau of 0.66, and refinement was stopped. Difference Fourier maps were again calculated to verify the locations of all heavy atom positions. Although a 3.5 Å resolution electron density map was calculated at this stage, it was deemed to be of marginal quality. The map showed some errors in the form of large 'pile-ups' of density around Hg site 1 and Pt site 2 which were 2 to 3 times the predominant maximum density elsewhere in the map. Negative peaks of equal magnitude were also present near these large positive regions. The problem is believed to have been caused by the mercurinitrophenol derivative data. Since only the Pt and Hg derivatives contributed significantly to the phasing in the annulus from 5 Å to 3.5 Å, one of them being below par could conceivably cause such errors. It has been suggested that 'maximum probability' phases, instead of the usual 'centroid' phases can correct this type of error, although it did not in this case. Anisotropic temperature factors for the Hg

Table 2: Quality of data used in the 5 Å resolution analysis.*

| Reagents (Conc.) | Soak Time | No. of crystals | No. of overlaps | R† |
|---|---|---|---|---|
| Native (Phosphate) | - | 4 | 3951 | 0.031 |
| Mercurinitrophenol (2 mM) | 2 weeks | 4 | 1274 | 0.054 |
| $Na_2AuCl_4$ (0.5 mM) | 2 weeks | 2 | 1855 | 0.035 |
| $(NH_4)_2PtCl_4$ (1.2 mM) | 1 week | 5 | 2299 | 0.087 |
| $K_2IrCl_6$ (1 mM) | 11 days | 1 | - | - |
| $K_2IrCl_6$ (1mM) | 6 weeks | 3 | 537 | 0.032 |

*Although the phasing was limited to 5 Å, the overlaps include all data to 3.5 Å resolution.

†$R = \Sigma|\bar{I}(H) - I(H)|/\Sigma|\bar{I}(H)|$, where $\bar{I}(H)$ is the average intensity of measurement H and I(H) is any measurement of reflection H.

sites were also tried, but they did not help, either.

At this point, an iridium derivative soaked for six weeks was solved and found to be similar, but not identical to the previous iridium derivative. Unfortunately, it, too, was good only to 5 Å. In order to avoid the problem described above, it became necessary to restrict refinement to 5 Å resolution. This refinement resulted in a figure of merit of 0.75 and a 5 Å electron density map was calculated. This map showed no signs of phasing errors (pile-ups). The heavy atom parameters from the refinement are given in Table 3. The plot of figure of merit vs. $\sin\theta^2/\lambda^2$ is presented in Figure 7. Figure 8 shows a breakdown of phasing contribution by each derivative. Both the errors (E's) and heavy atom scattering contributions (rms $\hat{F}_H$) are given. Generally, a ratio of 1.5 or greater for the rms $\hat{F}_H$ to the rms lack of closure is considered to be an indicator of a good derivative. These ratios are given in Table 4 for the 5 Å studies.

From the 5 Å resolution electron density map, two more or less distinct domains could be seen, and the molecular envelope could be defined. The overall shape of the molecule was determined to be ellipsoidal, with dimensions 68 Å by 38 Å by 30 Å.

Several long rodlike structures could be discerned in the electron density map (e.g. Figure 9) and were assumed (correctly, as it turned out) to be helices. The handedness of the helices in the map was unclear, and were later unambiguously resolved on the basis of the 3.5 Å analysis. This implied that wrong assignments of signs were made for heavy atom parameters throughout the 5 Å structure analysis. The signs were changed for the final 3.5 Å studies.

Table 3: Heavy atom parameters based on 5 Å resolution refinement

| Site | x | y | z | $B(Å)^2$ | Fractional Occupancy |
|---|---|---|---|---|---|
| | | | MERCURINITROPHENOL DERIVATIVE | | |
| 1 | 0.1052 | 0.1259 | 0.2282 | 62.0 | 0.85 |
| 2 | 0.2072 | 0.1448 | 0.2543 | 39.4 | 0.22 |
| 3 | 0.3012 | 0.7417 | 0.2544 | 10.0* | 0.14 |
| | | | Pt DERIVATIVE | | |
| 1 | 0.4536 | 0.8230 | 0.3092 | 92.5 | 1.05 |
| 2 | 0.3954 | 013103 | 0.3327 | 99.9 | 0.68 |
| 3 | 0.0350 | 0.1741 | 0.4717 | 72.1 | 0.53 |
| 4 | 0.0327 | 0.1000 | 0.2061 | 27.1 | 0.16 |
| 5 | 0.4156 | 0.3315 | 0.1739 | 12.2 | 0.15 |
| 6 | 0.4505 | 0.1388 | 0.3133 | 10.0* | 0.15 |
| 7 | 0.4543 | 0.2307 | 0.3739 | 126.4 | 0.15 |
| 8 | 0.4769 | 0.1521 | 0.4512 | 31.7 | 0.21 |
| | | | Au DERIVATIVE | | |
| 1 | 0.3769 | 0.3135 | 0.3449 | 73.4 | 0.34 |
| 2 | 0.0048 | 0.0610 | 0.2047 | 68.2 | 0.15 |
| 3 | 0.2996 | 0.4904 | 0.2263 | 59.2 | 0.12 |
| 4 | 0.0655 | 0.0993 | 0.2353 | 10.0* | 0.10 |
| | | | Ir DERIVATIVE (11 DAY SOAK) | | |
| 1 | 0.4704 | 0.2762 | 0.4328 | 103.0 | 0.49 |
| 2 | 0.4709 | 0.2221 | 0.3900 | 101.3 | 0.35 |
| 3 | 0.0708 | 0.6545 | 0.0047 | 90.7 | 0.18 |
| | | | Ir DERIVATIVE (6 WEEKS SOAK) | | |
| 1 | 0.4728 | 0.2780 | 0.4353 | 98.0 | 0.40 |
| 2 | 0.0090 | 0.1200 | 0.8361 | 62.2 | 0.22 |
| 3 | 0.4724 | 0.2290 | 0.3985 | 115.3 | 0.23 |
| 4 | 0.4248 | 0.3455 | 0.5109 | 107.7 | 0.23 |

* Parameters which were not refined.

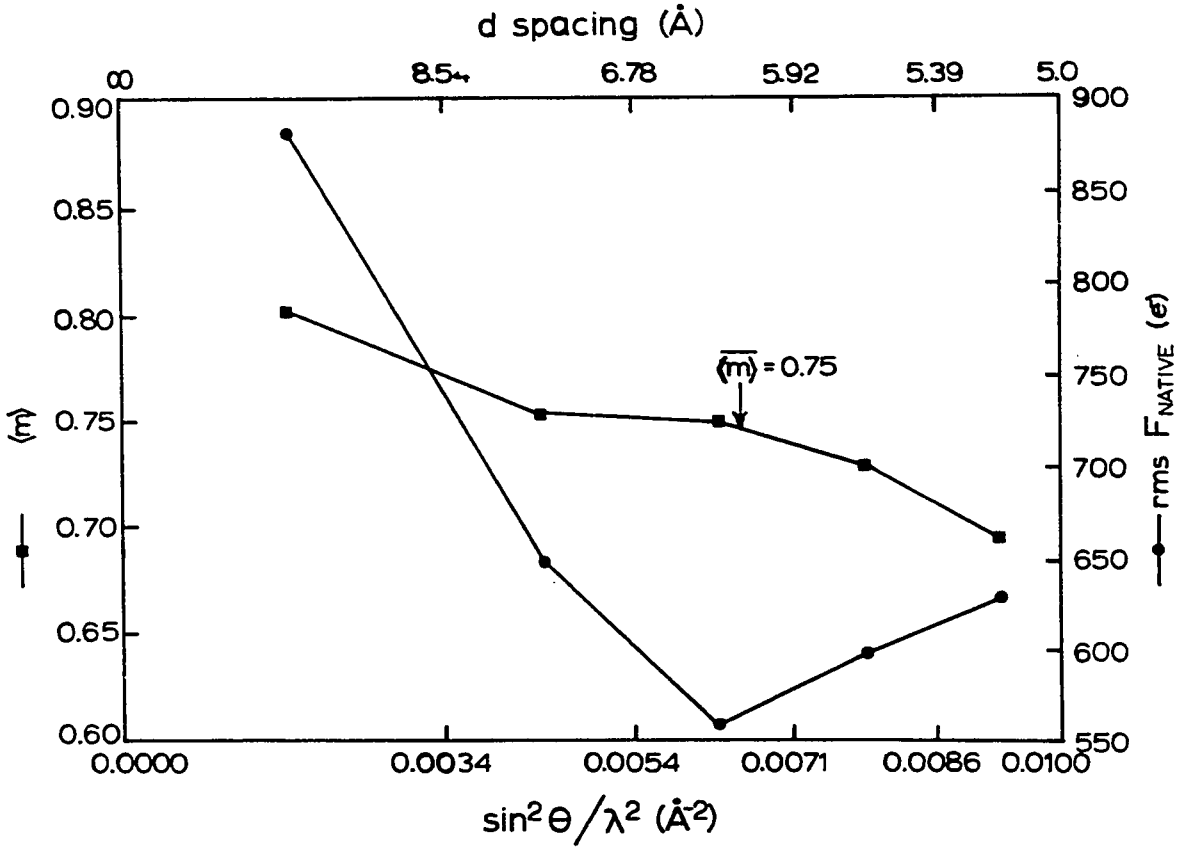



Figure 7: Distribution of figure of merit and rms $F_{NATIVE}$ for 5 Å phasing.

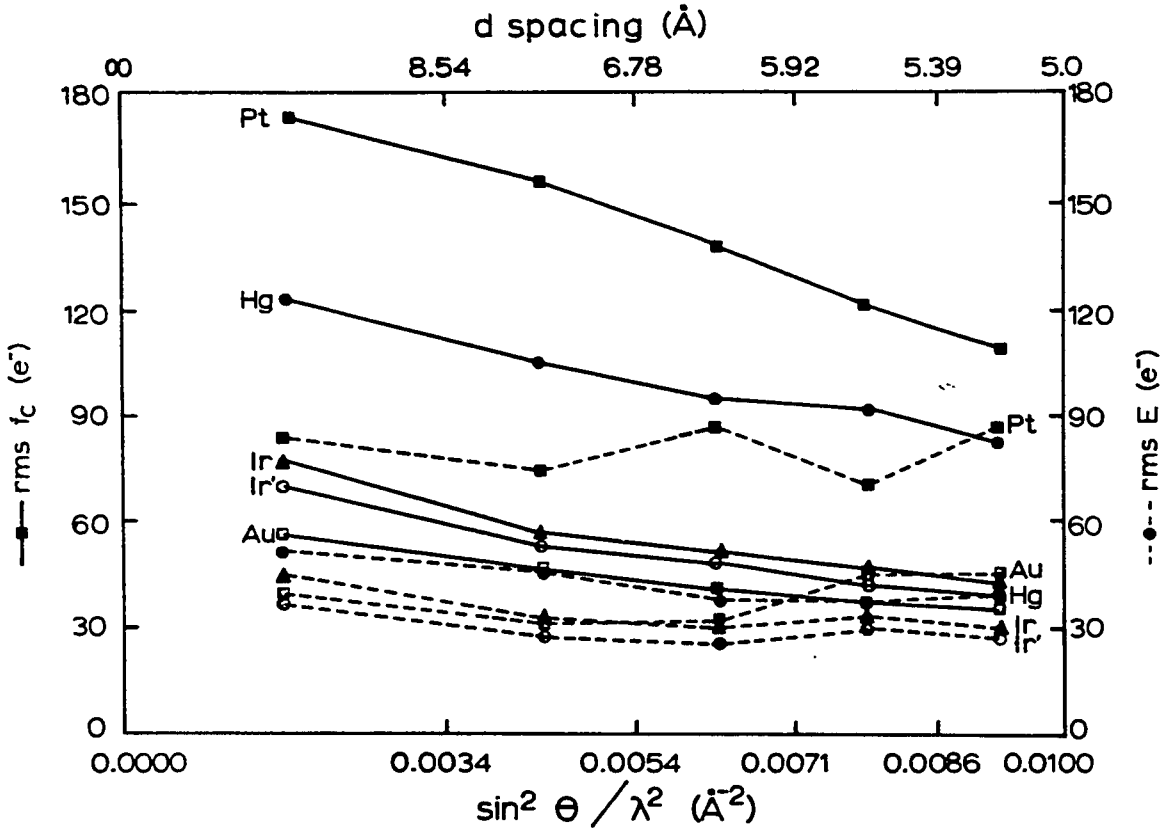



Figure 8: 5 Å phasing contributions breakdown by derivative.

Table 4: 5 Å Resolution Phasing Statistics

| Heavy Atom | RMS-E* | RMS-$F_H$† | $F_H$/ E | R-factor§ |
|---|---|---|---|---|
| Hg | 43.2 | 99.8 | 2.31 | 0.056 |
| Pt | 79.0 | 135.6 | 1.72 | 0.097 |
| Au | 37.6 | 43.9 | 1.17 | 0.049 |
| Ir, 11 day soak | 33.3 | 53.1 | 1.59 | 0.044 |
| Ir, 6 weeks soak | 29.8 | 49.6 | 1.66 | 0.039 |

*Root mean square lack-of-closure error $E = (\Sigma d_{hj}^2/n)^{\frac{1}{2}}$, $d_{hj}$ = lack-of-closure for reflection h of derivative j and n = number of reflections.

†Root mean square heavy atom contribution $F_H = (\Sigma_h f_{hj}^2/n)$, $f_{hj}$ = heavy atom scattering factor

§Kraut R factor = $(\Sigma||\hat{F}_{PH}|-|\hat{F}_P + \hat{F}_H||)/\Sigma|\hat{F}_{PH}|$

A styrofoam model was constructed from the 5 Å map to better visualize the structure. Two views of the model are given in Figures 10(a) and 10(b). Note the apparent similarity in shape between the two domains as viewed down the Z axis. Although the model was constructed as the wrong enantiomorph, the photographic negatives of the model have been inverted to produce figures with the correct handedness. Little else could be deduced from the 5 Å resolution analysis. However, the delineation of the molecular boundary, and the existence of rodlike structures were reassurance that phasing was being carried out reasonably well.

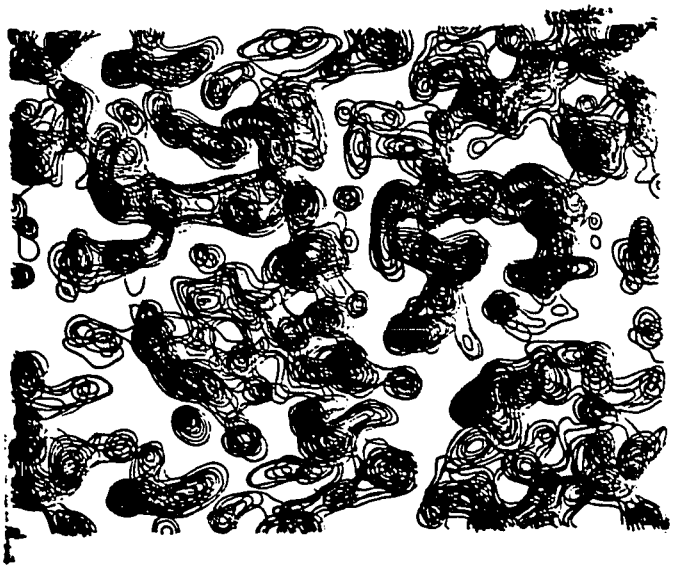
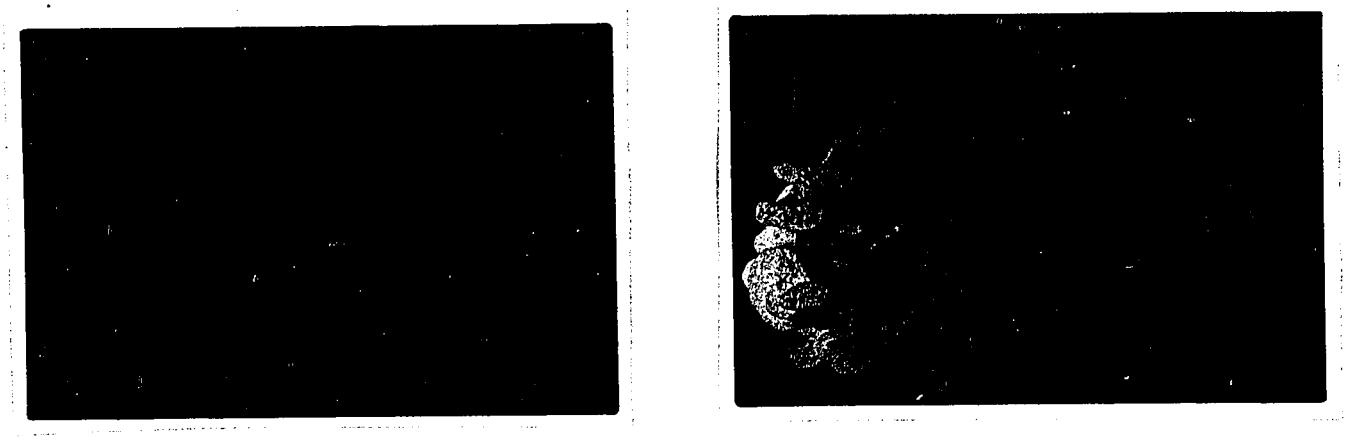


Figure 9: A portion of the 5 Å electron density map bounded by x = 0.40 to 0.75, y = 0.20 to 1.30, and z = - 0.20 to 0.28. y is horizontal, x is vertical, and z is the axis which has been sectioned.



(a) (b)

Figure 10: The styrofoam model viewed down a line bisecting x and z (a) and down z (b). y is horizontal.

D. 3.5 Å Resolution Structure Analysis

The structure analysis at 3.5 Å resolution was furthur studied after the addition of two good derivatives, $CdI_2$ and PCMBS (p-chloromercuribenzenesulfonic acid). Both show significant heavy atom scattering to at least 3.5 Å resolution. It was decided to include only the better Ir, mercurinitrophenol and Pt derivatives from the low resolution studies. The decision was based on the magnitudes of the heavy atom contribution, $\hat{F}_H$. Therefore, five derivatives were used in the 3.5 Å resolution analysis and are given in Table 5. The quality of the data is good, as indicated by the low values of the R-factors. Refinement was carried out until convergence was reached. The heavy atom parameters resulting from the refinement are given in Table 6. The positions have been shifted relative to the 5 Å results to place all of the heavy atom sites on one molecule and define the correct enantiomorph. The mean figure of merit for the 4000 reflections which were phased was 0.76, much improved relative to the initial 3.5 Å resolution mean merit of 0.66. The distribution of figure of merit and $F_{NATIVE}$ versus $\sin^2\theta/\lambda^2$ are given in Figure 11. Again the breakdown of phasing contribution by derivative is presented in Figure 12, with a summary in Table 7. The outstanding quality of the PCMBS derivative indicated that it should be very useful in future high resolution studies. The figure of merit weighted electron density map was calculated and no obvious signs of phasing errors were visible, as opposed to the earlier 3.5 Å resolution maps.

Table 5: Quality of the 3.5 A resolution data.*

| Derivative (Conc.) | Soak Time | No. of crystals | No. of overlaps | R |
|---|---|---|---|---|
| Native (Maleate) | - | 7 | 6022 | 0.047 |
| PCMBS (2mM) | 2 weeks | 5 | 1969 | 0.047 |
| $CdI_2$ (3mM) | 17 days | 6 | 1581 | 0.035 |
| Mercurinitrophenol (2mM) | 2 weeks | 4 | 1274 | 0.054 |
| $(NH_4)_2PtCl_4$ (1.2mM) | 1 week | 5 | 1849 | 0.076 |
| $K_2IrCl_6$ (1mM) | 11 days | 1 | - | - |

* As defined in Table 2.

Table 6: Heavy Atom Parameters Based on 3.5 Å Resolution Refinement

| Atom | x | y | z | B(Å)$^2$ | Fractional Occupancy |
|---|---|---|---|---|---|
| | | | Pt DERIVATIVE | | |
| 1 | 0.4535 | 0.8183 | -0.3061 | 27.3 | 1.00 |
| 2 | 0.3994 | 0.3113 | -0.3340 | 188.0 | 1.05 |
| 3 | 0.0302 | 1.1699 | -0.4702 | 31.5 | 0.64 |
| 4 | 0.0670 | 1.1058 | -0.2299 | 10.0* | 0.13 |
| 5 | 0.5287 | 0.8266 | -0.3315 | 10.0* | 0.16 |
| 6 | 0.5012 | 0.6760 | -0.2284 | 10.0* | 0.20 |
| 7 | 0.1943 | 1.1262 | -0.3139 | 12.7 | 0.18 |
| 8 | 0.3855 | 0.8422 | -0.3165 | 113.2 | 0.15 |
| 9 | 0.5021 | 0.7341 | -0.1198 | 10.0* | 0.32 |
| 10 | 0.5044 | 1.1484 | -0.4551 | 10.0* | 0.12 |
| | | | $CdI_2$ DERIVATIVE | | |
| 1 | 0.4657 | 0.9744 | -0.4013 | 43.9 | 1.08 |
| 2 | -0.0755 | 1.1325 | -0.4508 | 43.8 | 1.09 |
| 3 | 0.0102 | 1.1935 | -0.4540 | 78.9 | 1.15 |
| 4 | 0.0618 | 1.1968 | -0.2606 | 57.5 | 0.72 |
| 5 | 0.6315 | 0.6378 | -0.2311 | 50.4 | 0.35 |
| 6 | 0.1297 | 0.7055 | -0.6433 | 40.0 | 0.26 |
| | | | PCMBS DERIVATIVE | | |
| 1 | 0.1072 | 1.1317 | -0.2291 | 31.6 | 0.74 |
| 2 | 0.3067 | 0.7363 | -0.2592 | 9.4 | 0.86 |
| 3 | 0.0037 | 1.1605 | -0.4718 | 3.7 | 0.18 |
| | | | MERCURINITROPHENOL DERIVATIVE | | |
| 1 | 0.1044 | 1.1275 | -0.2293 | 52.5 | 0.83 |
| 2 | 0.3053 | 0.7374 | -0.2591 | 6.9 | 0.28 |
| | | | Ir DERIVATIVE | | |
| 1 | 0.4734 | 0.7748 | -0.0655 | 233.6 | 0.74 |
| 2 | 0.5326 | 0.7218 | -0.1137 | 145.5 | 0.39 |
| 3 | 0.4824 | 0.8741 | -0.3339 | 16.6 | 0.14 |
| 4 | -0.0569 | 1.1498 | -0.4964 | 210.4 | 0.26 |
| 5 | 0.1068 | 1.1412 | -0.2453 | 155.7 | 0.12 |

* Parameters which were not refined.

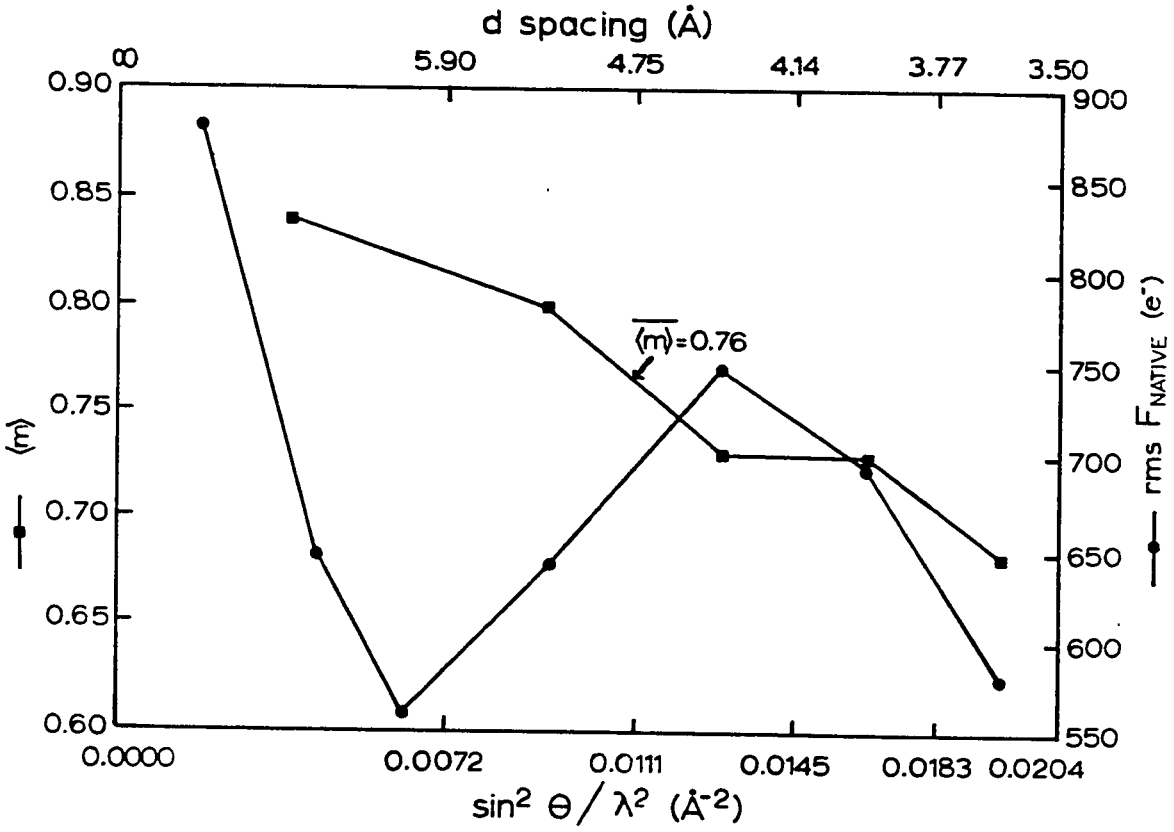



Figure 11. Distribution of figure of merit and rms $F_{NATIVE}$ vs. resolution for 3.5 Å phasing.

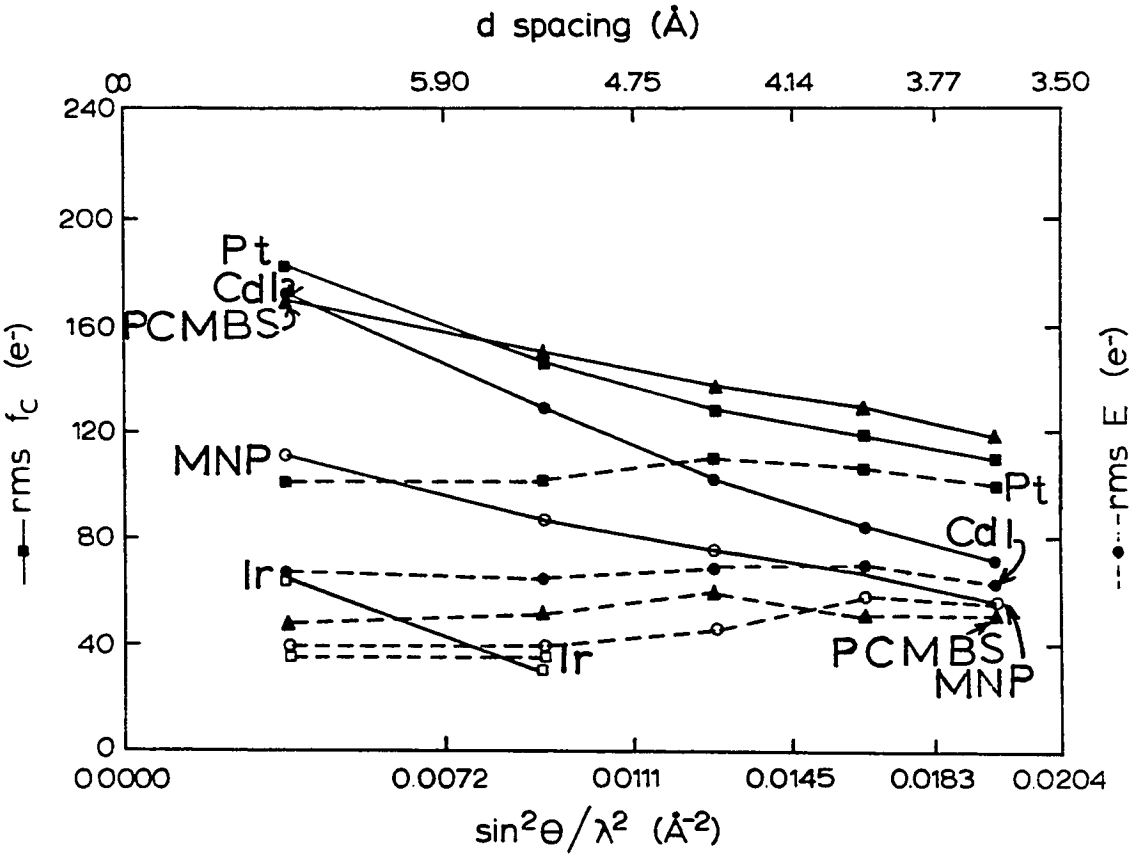



Figure 12. 3.5 Å Phasing contributions by derivative.

Table 7: 3.5 Å Phasing Statistics*

| Heavy Atom | RMS-E | RMS-$F_H$ | $F_H$/ E | R-factor |
|---|---|---|---|---|
| Hg, PCMBS | 51.6 | 143.2 | 2.78 | 0.065 |
| Hg, mercurinitrophenol | 47.8 | 82.3 | 1.72 | 0.061 |
| $CdI_2$ | 66.4 | 119.3 | 1.80 | 0.085 |
| Pt | 104.6 | 142.3 | 1.36 | 0.130 |
| Ir† | 35.0 | 53.1 | 1.52 | 0.048 |

*As defined in Table 4.

†5 Å resolution only.

First of all, the interpretations based on the 3.5 Å resolution electron density map are in basic agreement with the 5 Å results. The rodlike structures were confirmed as helices, and the correct enantiomorph was defined as described in the previous section. The molecular boundary is more clear at this resolution, as expected. The secondary and tertiary structural features of the molecule are also quite clear and will be discussed below.

The entire 3.5 Å resolution electron density map on which the structural interpretations are based is given in Figure 13. The region of the unit cell shown is that required to encompass a single molecule (more than just an asymmetric unit). The map is presented as a series of photographs of the transparencies generated as described in the methods section. Each photo is a composite of seven sections, representing a thickness of 5.9 Å. A brief description of the structural features of each composite is given along with each photo. The fraction of the unit cell desplayed is x = -0.20 to 0.70, y = 0.25 to 1.25, and z = 0.92 to 0.39. The y-axis is horizontal, x is vertical, and z is the axis which has been sectioned. Note that this map is in the same orientation as Figure 16(a), so that a qualitative comparison can be made between the polypeptide chain tracing and the electron density maps, if desired. The scale of the photographs is 4.7 Å/cm.

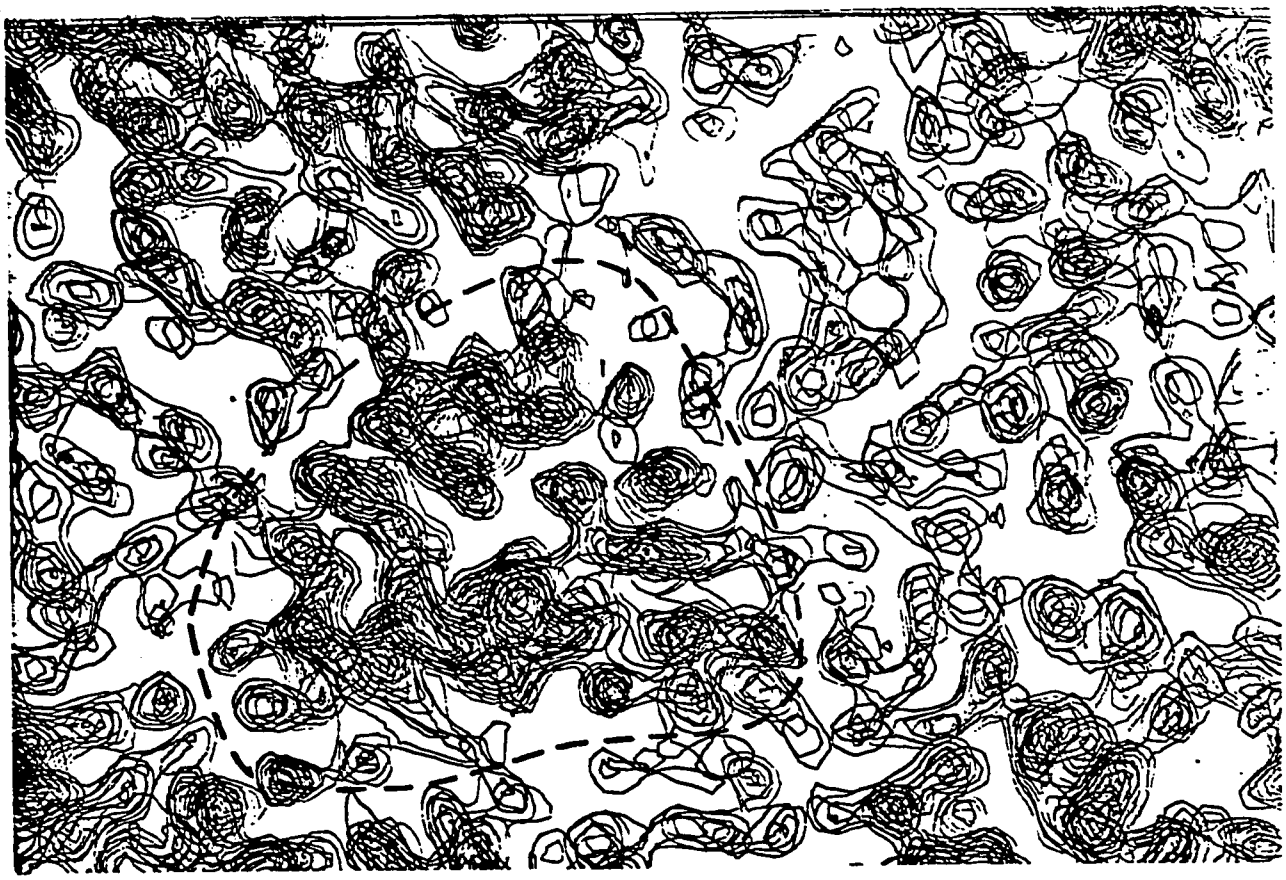


Figure 13(a): Z = 0.919 to 0.843

This photo shows the extreme top * of the molecule. Two helices† (b and c) of the P domain can be seen at center left. Parts of neighboring molecules can be seen in each of the four corners, with disordered, weakly diffracting solvent elsewhere in the map.

---

* The portion of the centered molecule with the highest Z has arbitrarily been designated as the 'top' of the molecule. Similarly, low Z portions are referred to as the 'bottom'. This is consistent with the physical procedure of stacking the transparencies.

† The helices are lettered consecutively from $NH_2$ to COOH-terminus. See also Figure 15.

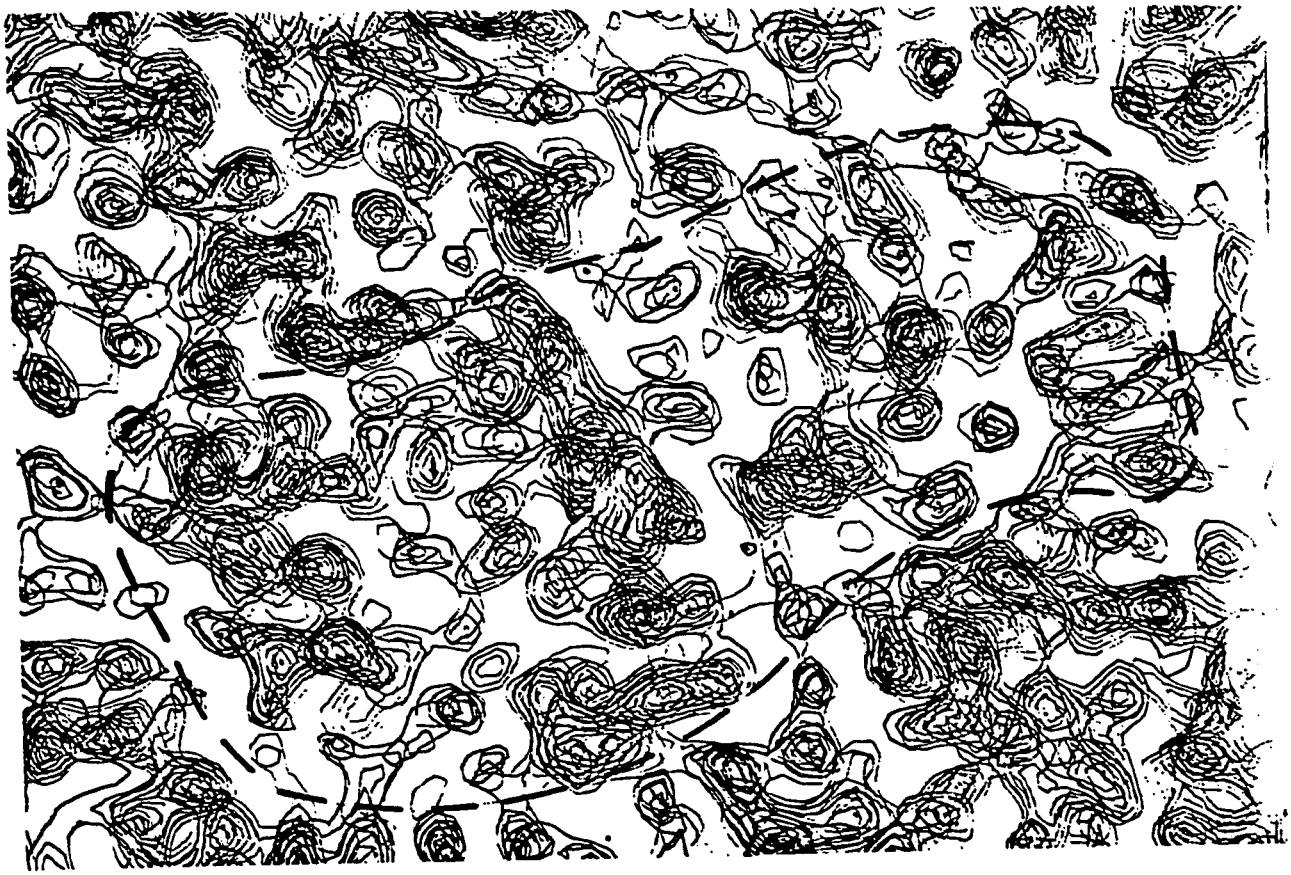


Figure 13(b): Z = 0.843 to 0.768

This section shows the extreme top of the Q domain (right). The density at left is part of the P domain. The helix at upper left is from a neighboring molecule, as are the dense regions at lower left and lower right.

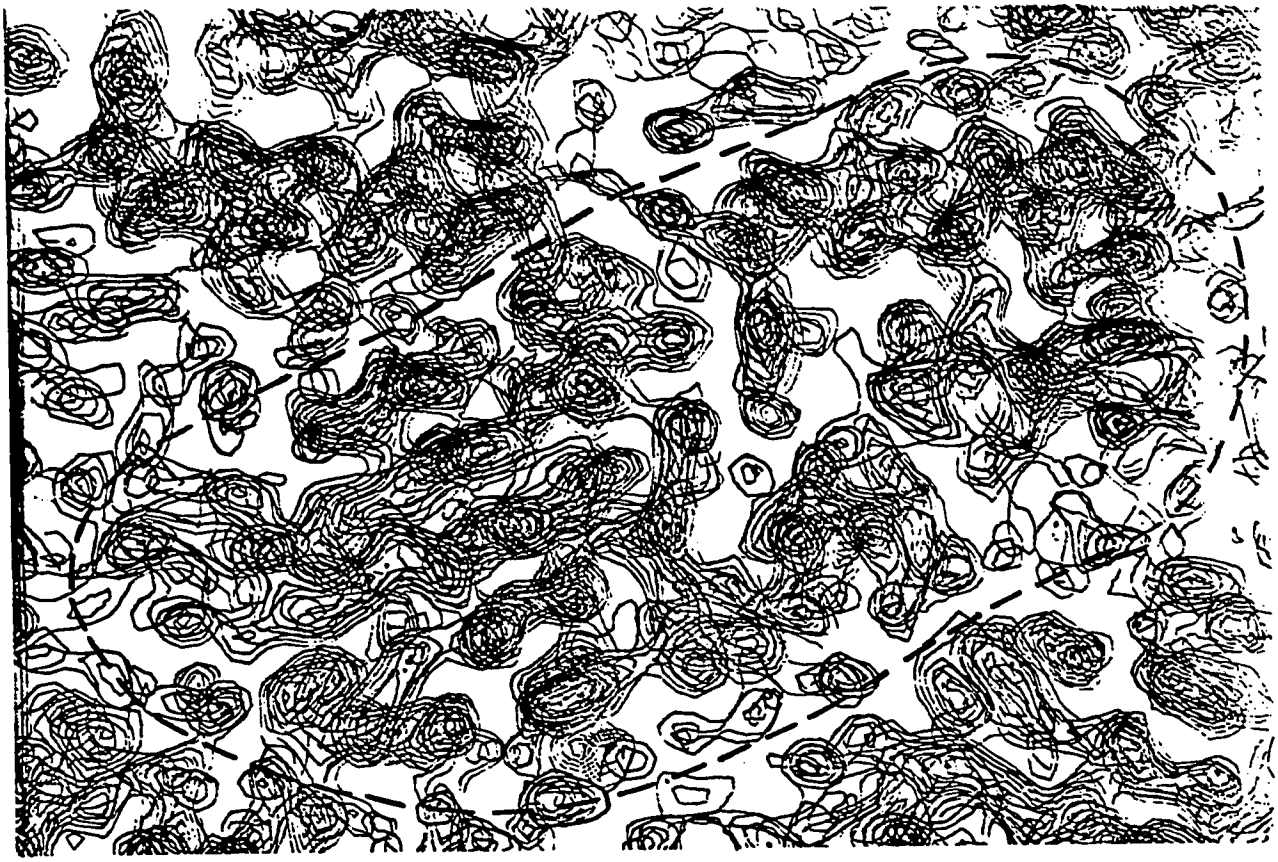


Figure 13(c): Z = 0.768 to 0.692

This section is interesting for several reasons. First of all, it contains the screw axis along y ( z = 0.75). Two screw related f helices can be seen in the upper portion of the photo, with their axes almost exactly along the screw axis itself. The molecular boundary is clear, with the molecule running diagonally from lower left to upper right. Note the general absence of electron density between the P domain (left) and the Q domain (right). The photo also displays prominently 5 of the β-sheet strands in the P domain (left center).

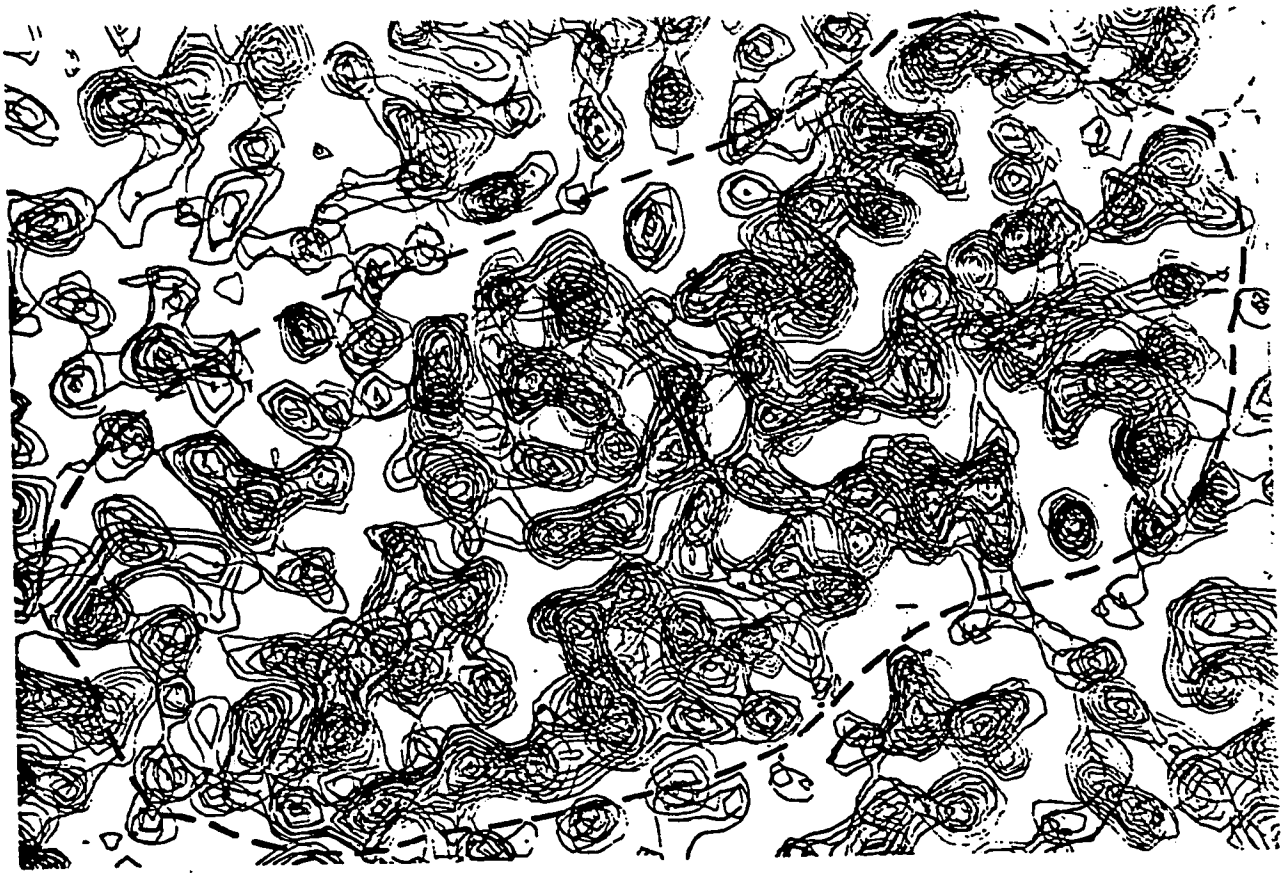


Figure 13(d): Z = 0.692 to 0.617

This section slices through both domains, showing again the overall ellipsoidal shape of the protein. At upper right, helix g can be seen forming an intermolecular contact.

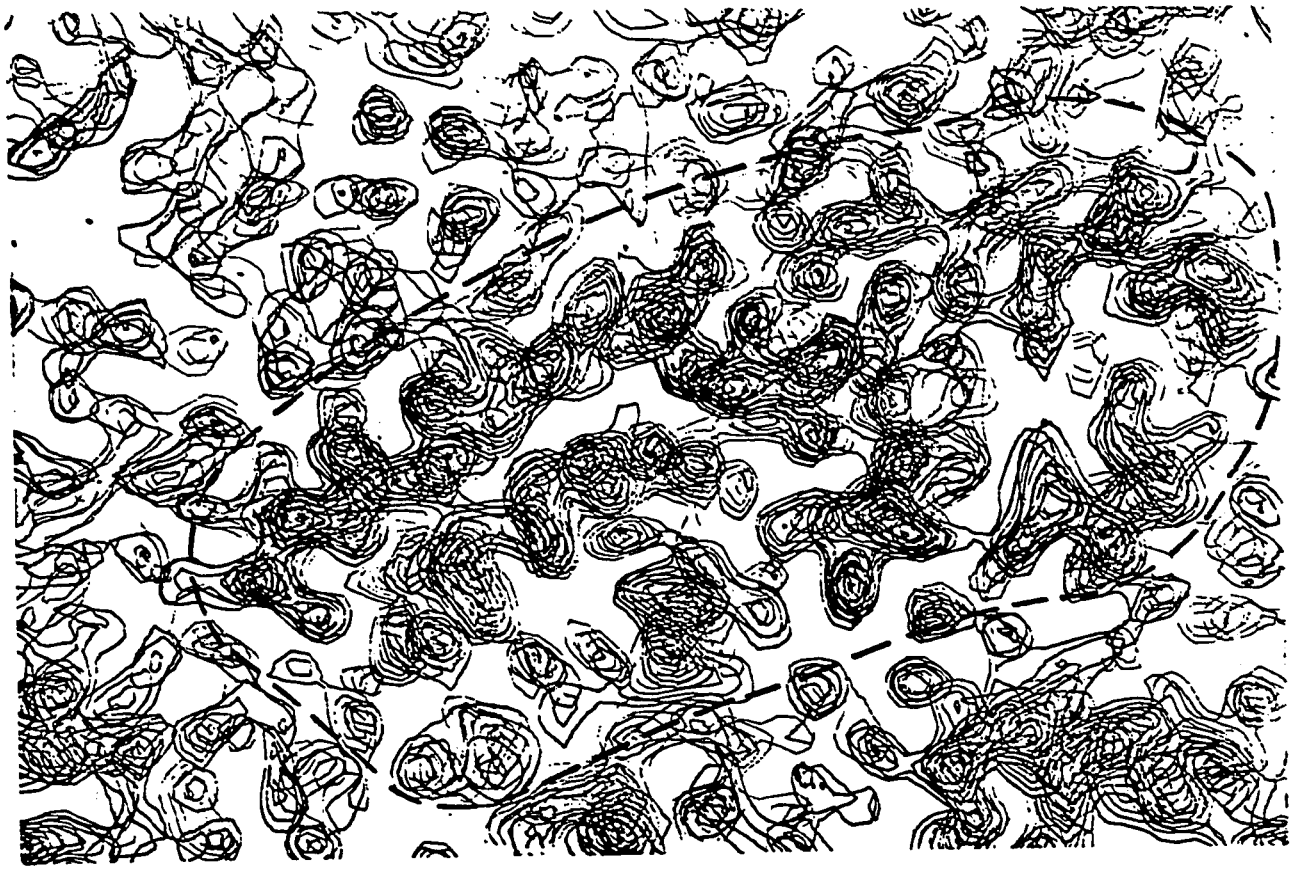


Figure 13(e): Z = 0.617 to 0.541

This portion of the molecule contains most of the a and h, P domain helices (left center, a helix left, h helix right). Parts of neighboring molecules can be seen at the lower right and lower left corners. Note the intermolecular solvent region at upper left.

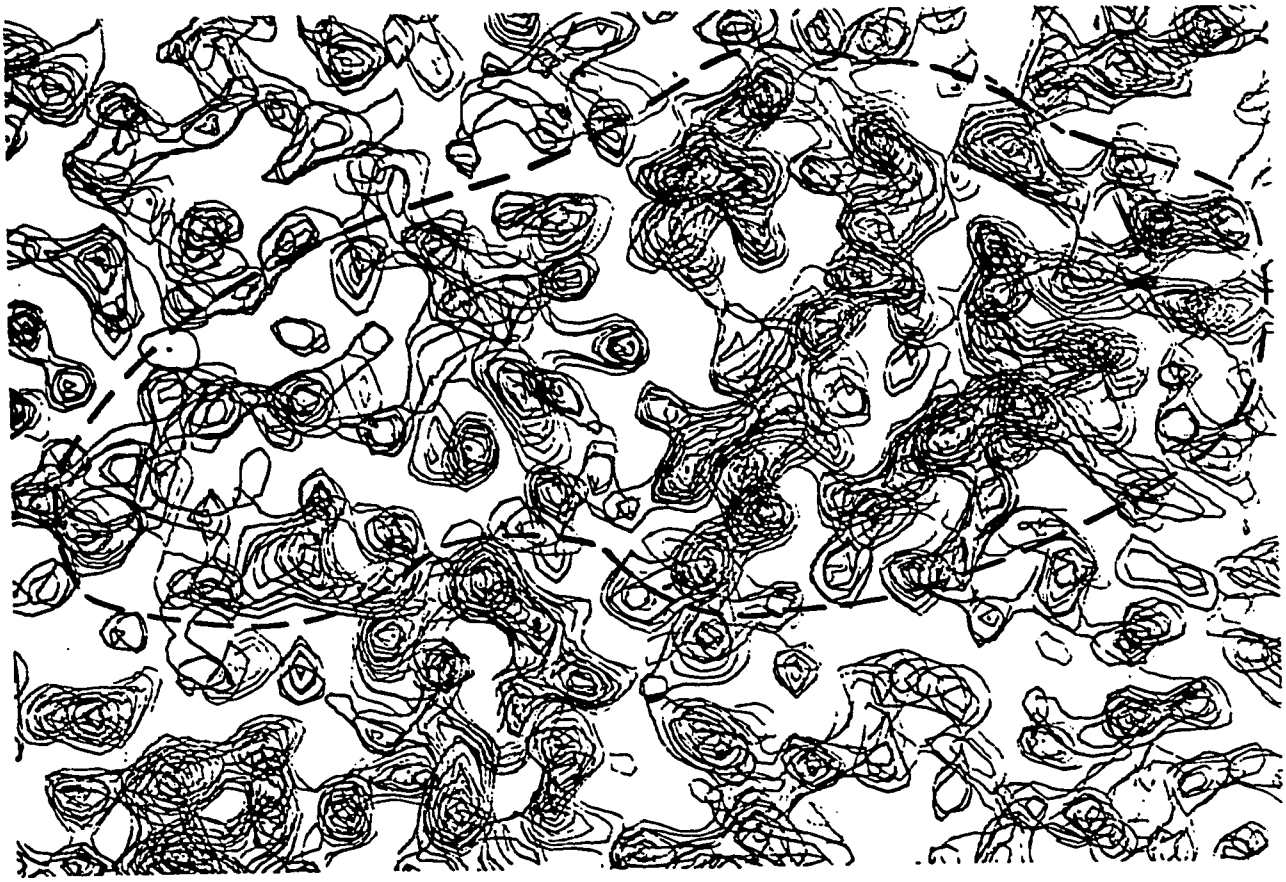


Figure 13(f): Z = 0.541 to 0.465

This figure shows the bottom of the Q domain (right). The P domain (left) has almost completely ceased to exist at this value of Z. The large loop, which is so conspicuous in Figure 16 (a) can be seen at top center in this slice of the molecule. There is also a region of intermolecular contact shown here (dashed line). The density at lower left is related to that at upper right by the screw axis along X at Y = 0.75, Z = 0.50.

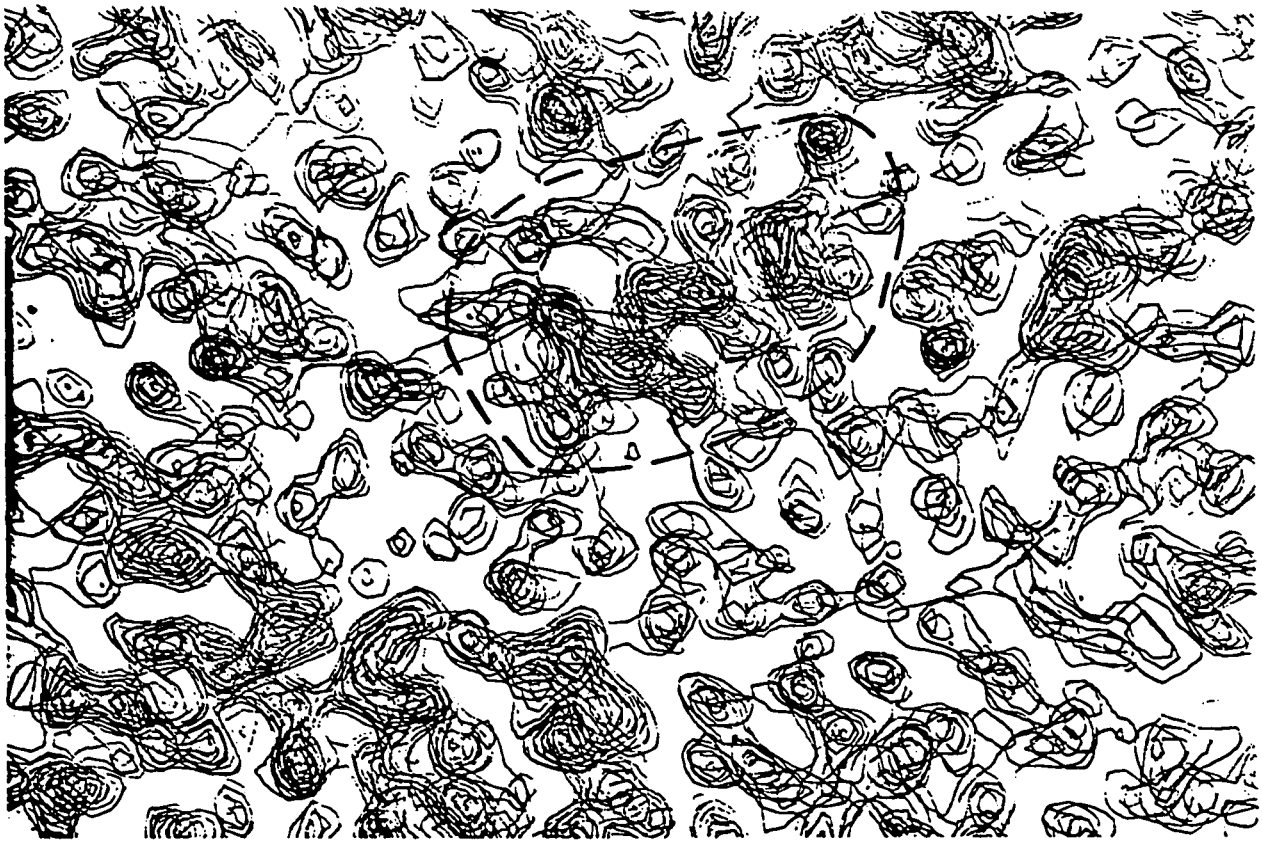


Figure 13(g): Z = 0.465 to 0.390

This photo shows the i helix of the COOH - terminal portion of the molecule. Since it lies on the extreme bottom of the entire structure, it has been caught in section, but the area surrounding it represents only solvent. Part of neighboring molecules can be seen at the lower left and top center.

Many long stretches of presumed polypeptide chain could be traced throughout the map, and although there existed several places where ambiguities in the chain direction occurred, a 'best trace' of the entire course of the polypeptide chain was made. Alternate courses of the chain in the ambiguous regions resulted in closed loops or else left out significant stretches of electron density.

At this resolution it must be stated that the polypeptide chain tracing and the sequence are not in total agreement, and rather than force the structure to contain 306 amino acids, the number determined by sequence studies, α-carbon atoms were positioned as best determined from the electron density map. This yielded 291 amino acids. Using a Richard's box (Richards, 1968) a direct study of the matching of the sequence to the electron density map was undertaken. The sequence and structure only matched well for the first 85 residues from the N-terminus and the last 30 from the COOH-terminus. The number of resolved side chains at 3.5 Å resolution is small, however, and it is not possible to assume that the sequence is incorrect. A higher resolution map will be necessary to verify more side chains, and hence clarify this discrepancy. A further discussion of the comparison between sequence and structure is presented in Section IV. F.

Stereo drawings of the entire polypeptide chain were drawn using a computerized plotter. The plots consist of points representing the α-carbon atoms of each amino acid, and lines connecting them. Precise locations of α-carbon atoms at this resolution can not be obtained and often atoms were just placed an appropriate distance from the previous α-carbon position along the stretch of electron density.

Figure 14 shows a schematic drawing of the molecule as determined from the 3.5 Å resolution studies. The cylinders represent helices and the arrows represent β-sheet. The direction of the arrows is from $NH_2$ - to COOH-terminus. For convenience, the domain at left has been labeled the P domain, and the one at right labeled the Q domain. Figure 15 shows only the α-helix and β-sheet structures of the molecule. Note that the β-sheet consists entirely of parallel strands. Also, the lettering scheme of the helices is given here.

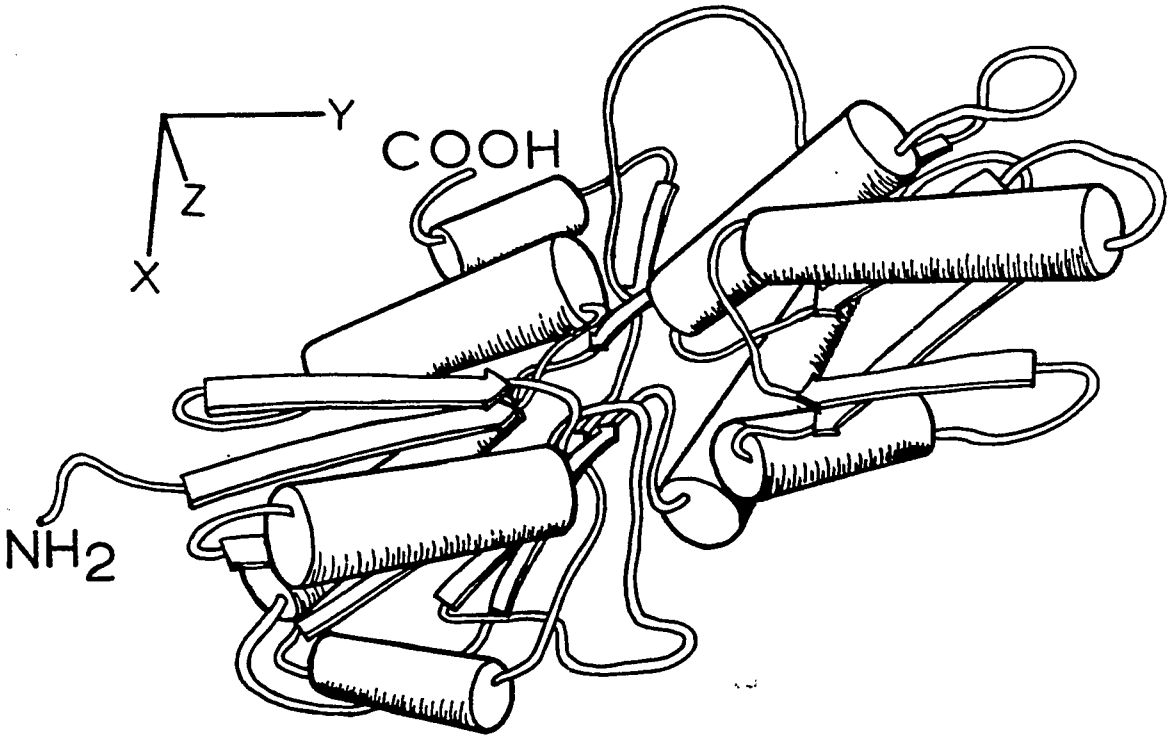



Figure 14: Schematic drawing of L-arabinose binding protein.

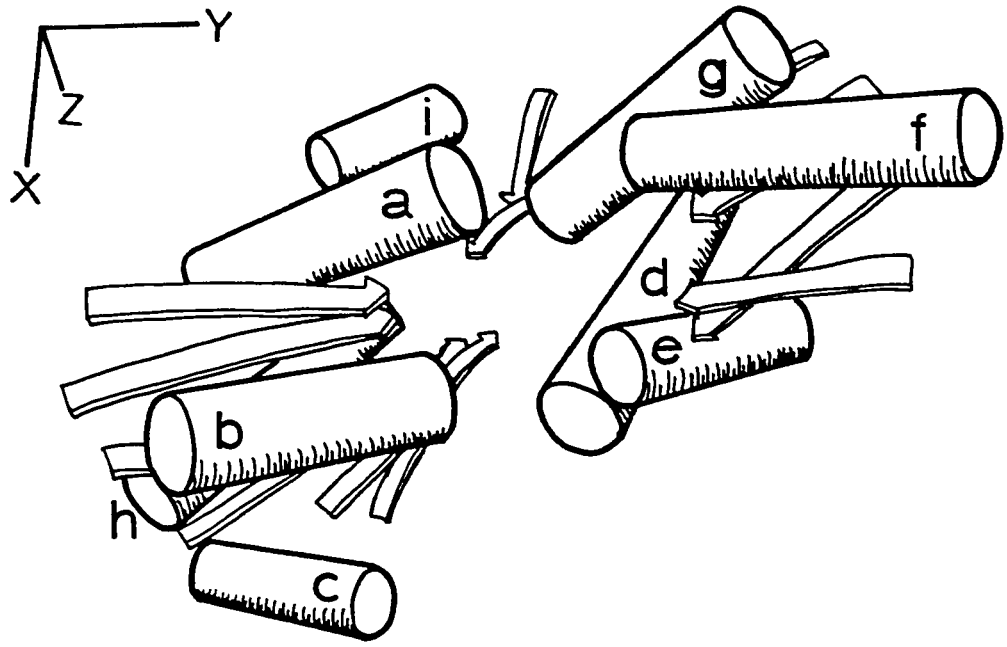



Figure 15: The arrangement of secondary structure in L-arabinose binding protein.

A more detailed description of the structure can be presented with the aid of the stereographic plots which follow.* Figure 16(a) shows the molecule viewed directly down the z axis, with y horizontal. The N-terminus is visible at the extreme left of the molecule and the COOH-terminus at the top center. Note the ellipsoidal shape of the molecule and the two domains. Figure 16(b) shows the molecule rotated 30° about the crystallographic y axis. Figure 16(c) continues the rotation to 60° relative to 16(a). In this view, the limited degree of connection between the domains is quite evident. It appears that only three chains of polypeptide chain cross between domains. Figure 16(d) shows the last of the series of rotations. This view is rotated 90° relative to 16(a), i.e. down the x-axis. The arrangement of the two domains creates a cleft, which is shown clearly in this view. An end view is shown in Figure 16 (e).

Of particular interest is a breakdown of the molecule into linear segments of polypeptide chain. Figure 17(a) shows the first third of the linear trace, starting from the N-terminus. It lies entirely in the P domain. This segment of chain shows alternate stretches of β-sheet and α-helix. The β-sheet is in the interior of the molecule, and helices on the outside. This is typical of most protein structures studied so far.

The second third of the amino acids form most of the Q domain of the molecule (Figure 17(b)). As compared to the P domain, the chain tracing of this domain was more difficult. Alternation of α - and β- structure as one traces the chain can also be seen here. This region of the molecule can not be reconciled with the sequence

---

* Stereographic glasses can be purchased from Hubbard Scientific Co. P.O. 105, Northbrook, Il. 60062.

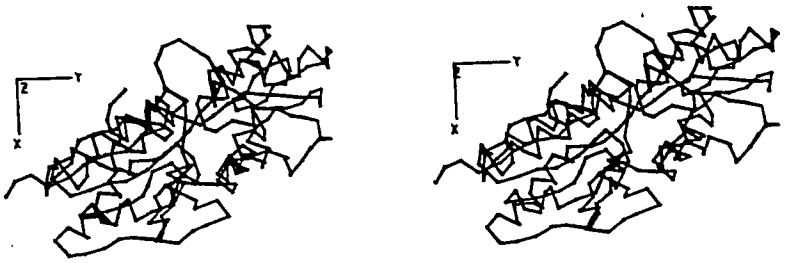


Figure 16(a): 0° rotation about y.

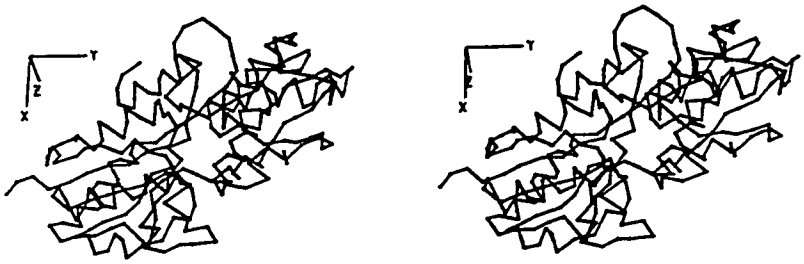


Figure 16(b): 30° rotation.

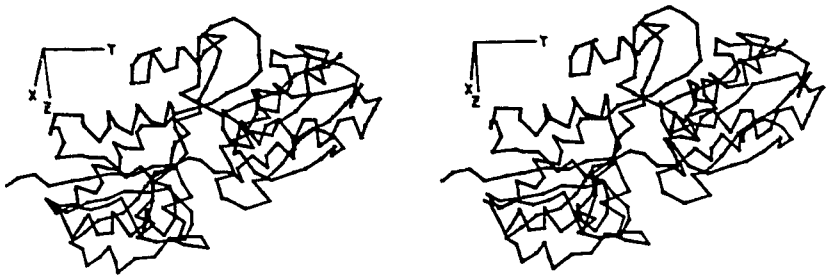


Figure 16(c): 60° rotation.

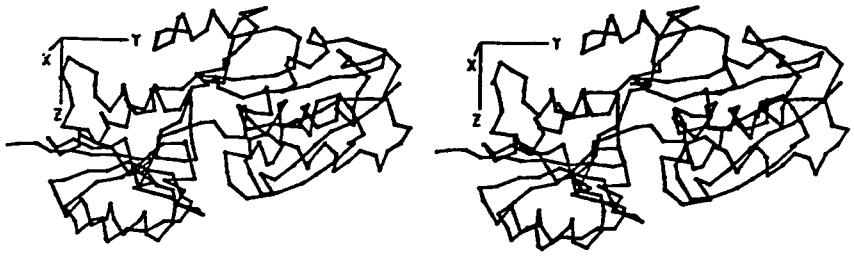


Figure 16(d): 90° rotation.

Figure 16(e): The molecule viewed end-on (approximately down y).

Figure 17(a): First third of the linear trace (Domain P).

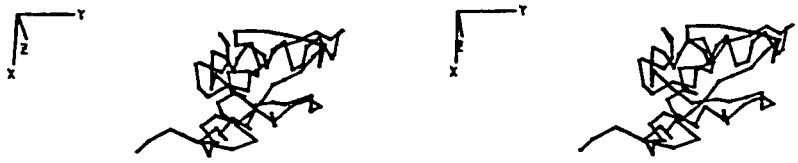


Figure 17(b): Second third (Domain Q).

Figure 17(c): Last third (linkage between the domains).

at this resolution.

The connections between the domains are produced from the last third of the amino acids (Figure 17(c)), with a large loop of about 33 amino acids extending back into the P domain before forming a final COOH-terminal helix.

With regard to the overall secondary structure, there exists extensive β-sheet in both the P and Q domains. In both domains there are six strands of entirely parallel sheet. These two β-sheets account for about 22% of the molecule, by amino acid. A plot of the α-carbon atom positions for residues involved in β-sheet is given in Figure 18. The strands in both domains are 'fanned' about 90°. The structural integrity of the two domains undoubtedly lies in the formation of these two backbone structures. Note also that the individual strands of the β-sheet tend to run in the direction of the long axis of the molecule.

The molecule contains nine helices of at least two turns. Their distribution is shown in Figure 19. Four helices lie in domain P, four in domain Q, and the ninth one is near the COOH-terminus and is more or less shared by the two domains. The helices comprise about 38% of the molecule. It is interesting that the axes of the helices are, in general, roughly parallel to the long axis of the molecule. One might speculate that the three dimensional folding depends on the formation of two helically bounded globular pieces, which then conbine end-to-end on the non-helical sides to produce a stable elongated structure.

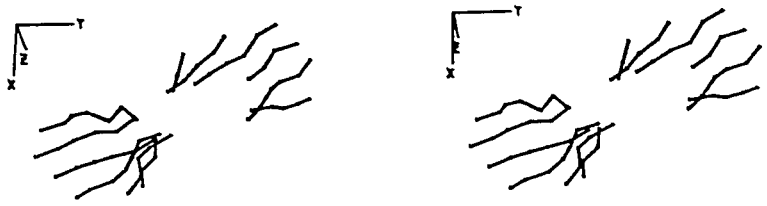


Figure 18: β-Sheet structures. Orientation is the same as in Figure 16(b).

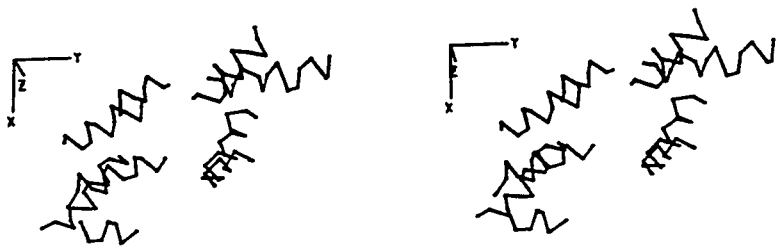


Figure 19: α-Helices. The orientation is a 20° rotation with respect to Figure 16(a).

E. Heavy Atom Binding Sites

The refined locations of all heavy atoms used in the 3.5 Å analysis are shown in Figure 20. The distribution is bizarre, to say the least. It is not known why the heavy atom compounds bind almost exclusively to the lower half of the molecule. Perhaps it can be explained when the entire sequence has been reconciled with the chain trace and the destribution of amino acids can be seen. One external side of the molecule may tend to be less hydrophilic, such that it doesn't bind the charged heavy atom compounds. If this is true, it would have functional implications, since this may be a membrane associated protein.

Platinum site 2 is bound to what is believed to be Glu-1, i.e. the N-terminal amino acid. Platinum-glutamic acid associations are common in protein structures. Platinum site 10, which appears to be between two molecules, has a rather low occupancy (0.12) and may, in fact, be artifactual.

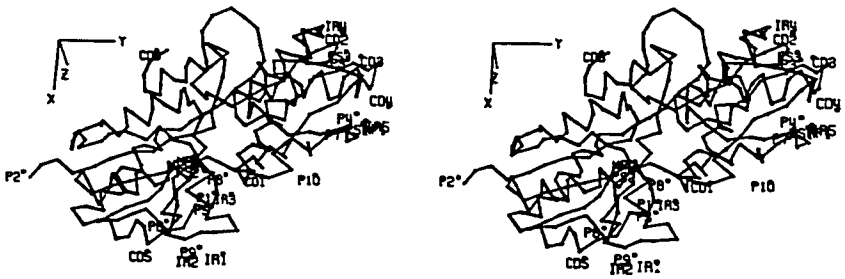


Figure 20: Distribution of heavy atom binding sites.

PS = PCMBS MP = mercurinitrophenol

P = $(NH_4)_2PtCl_4$ CD = $CdI_2$

IR = $K_2IrCl_6$

F. Crystal Packing

The strange heavy atom distribution can not be explained on the basis of packing considerations. Figure 21 shows the packing scheme of the protein. The symmetry notation is taken from the International Tables (Henry and Lonsdale, 1969). And the origin is shown as was defined for this structure analysis. The four equivalent positions are also given in the legend.

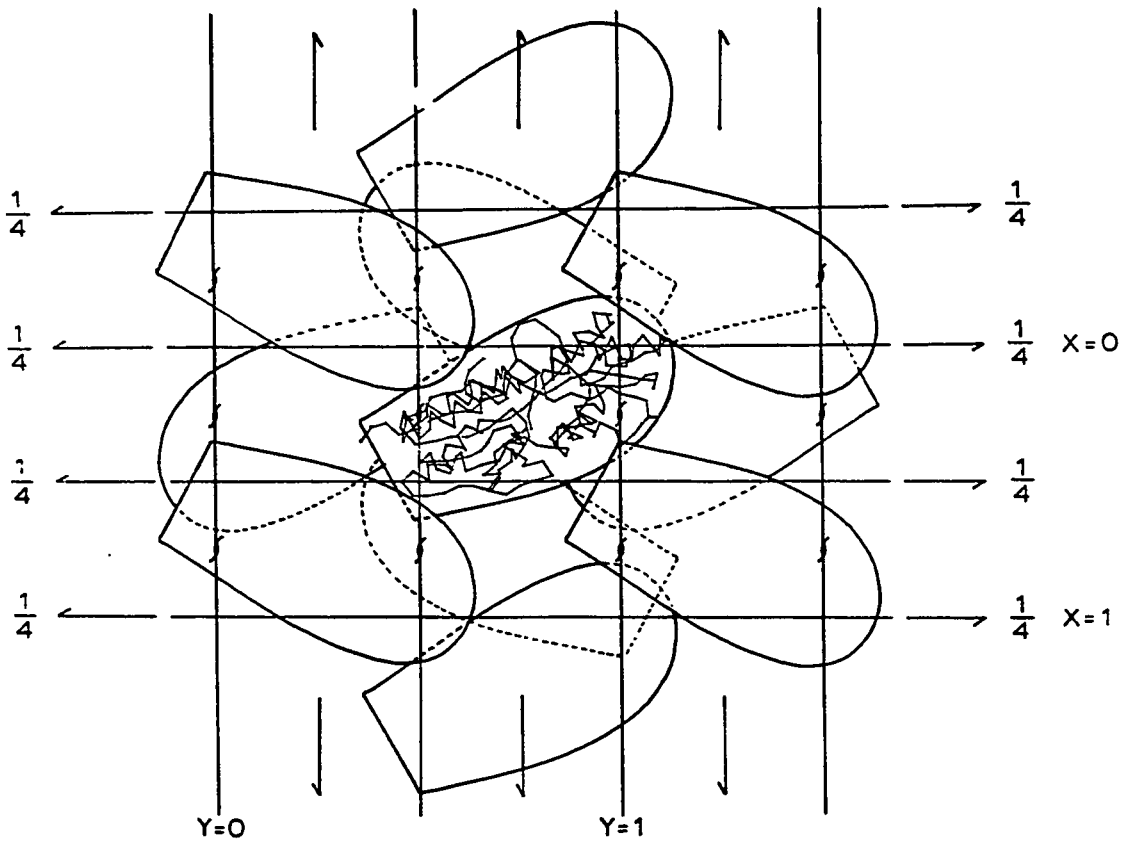



Figure 21: Packing diagram for L-arabinose binding protein. The origin is defined to be consistent with Appendix E. Equivalent positions for this space group ($P2_12_12_1$) are: x,y,z; $\frac{1}{2}$-x,-y,$\frac{1}{2}$+z; $\frac{1}{2}$+x,$\frac{1}{2}$-y,-z; and -x,$\frac{1}{2}$+y,$\frac{1}{2}$-z.

## F. Comparison Between Sequence and Structure

### 1. Matching of Side Chain Residues to Electron Density Maps

Although a complete matching of sequence to structure has proven impossible at this resolution, approximately one third of the molecule has been successfully fitted. This has enabled the two termini to be distinguished unambiguously. The large hydrophobic residues trp-15 and phe-16 at the N-terminus of the a helix, are clear, as is trp-21. The proline residues at 41, 66, and 68 are consistent with directional changes of the polypeptide chain in the electron density map. The map also shows an unambiguous continuous linear strand continuing to what would correspond to about residue number 110, however, past the region corresponding to residues 80-85, the agreement between side chain size and residue type is poor.

On the COOH-terminal end of the molecule, the agreement between side chain and electron density is good from the terminus itself back to about residue 278. At that point several stretches of polypeptide chain pass near each other, and it is not possible to continue the fitting process. Examination of a preliminary 2.8 Å resolution map by Gary Gilliland in this laboratory, has verified the qualitative aspects of the 3.5 Å resolution analysis, and has found improved agreement between sequence and structure in the regions examined so far.

### 2. Location of the Single Cysteine

Although it was originally thought that the most highly substituted Hg site of the mercurinitrophenol derivative represented binding of Hg to the single sulfhydryl, it became apparent in trying other mercurials as derivatives that this might not be true.

Hence a more systematic study was undertaken. Table 8 shows the mercurials tried and the relative occupancies at sites 1 and 2. Site 1 is on the end of the molecule, on the Q domain side. Site 2 is in the cleft between the domains, slightly towards the P side. The chemical structures of the various mercurials are shown in Figure 22.

First of all it was noticed that the color imparted to fluorescein mercuric acetate treated crystals could not be dialyzed away. Difference Fourier maps of that derivative showed 90% occupancy at site 2, the only 10% at site 1. This suggested that site 2, rather than site 1 represented the covalent Hg binding site. This was confirmed by backwashing PCMBS treated crystals in buffer and collecting diffraction data. A difference Fourier map revealed occupancy at site 2 of such backwashed crystals, and none at site 1. This places the covalent Hg at site 2, and hence, the single cysteine residue in the cleft between the domains.

In the sequence, the cysteine was originally placed at 128, however this was not consistent with the fitting of residues to the electron density map. Assignments of 64, or 97-99 were more reasonable, based on the x-ray studies. Reexamination of the sequences by Hogg (personal communication) has placed the cysteine at 64.

3. Observed versus Predicted Secondary Structure

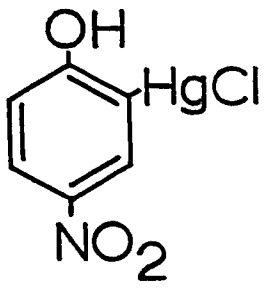
There exist various schemes for predicting secondary structure in protein molecules (Chou and Fasman, 1974; Wu and Kabat, 1973; Robson and Pain, 1971; and others). Although none can produce consistently accurate results (Matthews, 1975), these schemes have

Table 8: Mercurial Binding Results

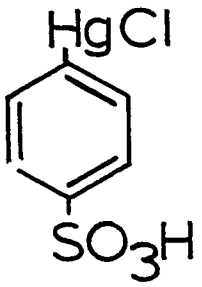
| Reagent | Fractional Occupancy | |
|---|---|---|
| | Site 1 | Site 2 |
| Mercurinitrophenol* | 80% | 26% |
| Mersalyl Na* | 60% | 20% |
| Dichloromercurinitrophenol* | 40% | 10% |
| Acetoxymercurisalicylaldehyde† | 0% | 90% |
| Fluorescein Mercuric Acetate† | 10% | 90% |
| $HgI_2$* | 5% | 37% |
| PCMBS* | 78% | 90% |
| PCMBS (backwashed)† | 0% | 90% |

* Refined values for the occupancy

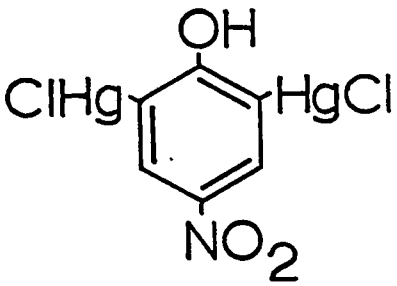
† Estimated values from difference Fourier maps



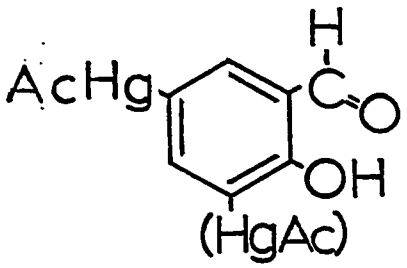

Mercurinitrophenol



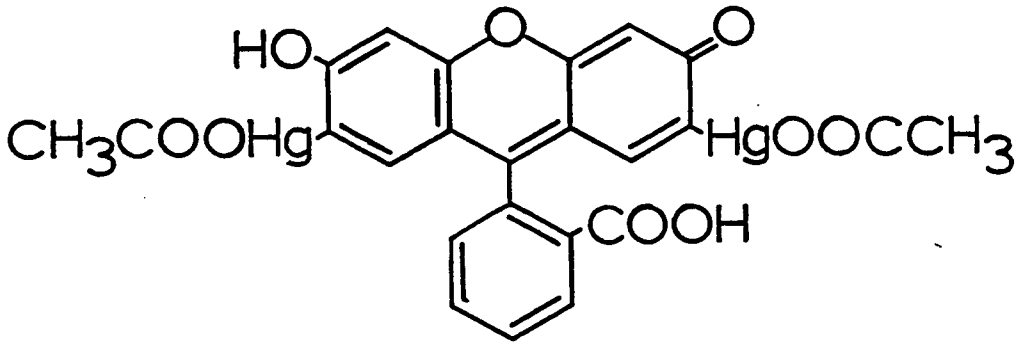

PCMBS
(p-Chloromercuri-
benzenesulfonic acid)



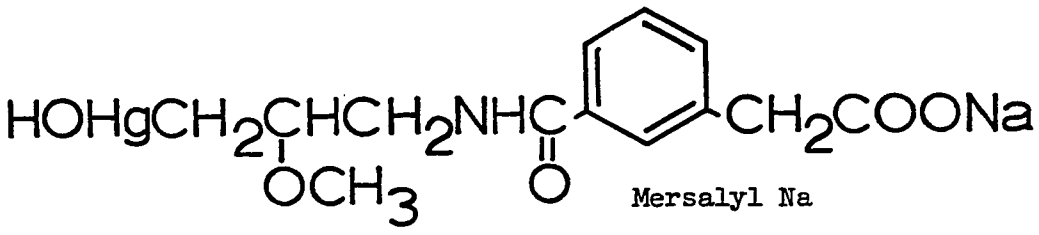

Dichloromercurinitrophenol




Acetomercurisalicylaldehyde
(stereochemistry unclear)




Fluoroscein Mercuric Acetate




Mersalyl Na

Figure 22. Organomercurials used in the study.

had limited success. One of the most successful schemes devised so far is that of Chou and Fasman. Their method is based on the relative occurance of each amino acid in $\alpha$-helix, $\beta$-sheet and random coil configurations in fifteen known protein structures. In predicting a new proteins's secondary structure, those probabilities are applied. Appendix E gives the observed and predicted regions of secondary structure. Note that since the sequence contains 306 amino acids versus 291 in the chain tracing, residues have been inserted in the listing. In this way, direct comparisons between observed and predicted structure can be made. The agreement is defined as:

$$\%_x = 100 \frac{n_x - n_{predicted}}{n_x}$$

where $n_x$ is the number of residues observed as secondary structure of type x ($\alpha$ , $\beta$, or random coil). Random predictions would result in $\%_x$ equal to 33.

If one includes the entire protein sequence in the comparison, $\%_\alpha = 54$, $\%_\beta = 57$, and $\%_{rc} = 39$. However, in many regions the agreement would be much better with a two or three residue shift in the placement of amino acids in a particular type of secondary structure, easily within the range of the experimental error of the 3.5 Å resolution trace.

If one includes only those portions of the chain tracing which have been 'fitted' to the sequence by matching of electron density to amino acid side chain size ( 1 - 85, 278 - 306), the agreement is much better, $\%_\alpha = 80$, $\%_\beta = 61$, and $\%_{rc} = 66$.

Considering the errors in deciding on which type of secondary structure a particular amino acid is involved in, and the fact that this is only a 3.5 Å resolution trace, the agreement is good for the portions of the molecule which have been fitted. As shown in Appendix E only $\beta$-sheet P-1 was totally incorrectly predicted.

G. Attempts to locate the Sugar Binding Site by Difference Fourier Methods

Originally it was hoped that the sugar binding site could be located by simply diffusing L-arabinose into the crystals and calculating difference Fourier maps, searching for an electron density peak. However, diffraction data collected from crystals soaked in solutions up to 20 mM in L-arabinose or D-galactose failed to produce any reproducible peaks in difference Fourier maps. In order to test the possibility that the binding protein was not active in the crystallization solvent, D. Miller (personal communication) carried out the standard binding assay in 60% 2-methyl-2,4-pentanediol. The protein was determined to be fully active.

The possibility also exists that sugar molecules are prevented from accessing their binding sites by direct steric interference from neighboring molecules in the crystal lattice. However, the packing diagram (Figure 21) shows that the protein is open to the solvent on all sides. Furthermore, the various organomercurials tested as derivatives have access to the presumed sugar binding site (see Introduction), and they are no smaller than the binding substrate.

A third possibility is that the structure has been solved with the sugar already bound to the protein. There are several observations that are consistent with this possibility. First of all, the $K_d$ of binding is $10^{-6}$M. And in the process of isolating the protein, the bacteria are suspended in 20 mM L-arabinose. Also, it has been reported that sugar has been found bound to the purified protein, and denaturation is required to fully remove it (Parsons and Hogg, 1974a).

Secondly, M. Newcomer (personal communication) has crystallized L-arabinose binding protein in the presence of 20 mM arabinose, and this form, although different in external morphology, shows no peaks above background in difference Fourier maps, just as in the soaking experiments.

Searches for extra 'non-protein' electron density in 3.5 Å resolution maps may not be realistic, but there appears to be a region of density too large for an amino acid side chain in the cleft region, near the position of the single thiol. However, the level of confidence for this observation is low, and more definitive work on locating the sugar binding site is underway by M. Newcomer.

## V. DISCUSSION

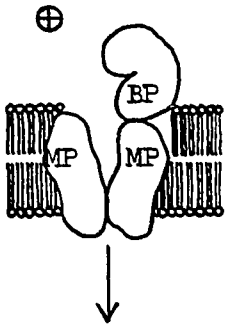
This study provides the low and medium resolution structural characterizations of L-arabinose binding protein. Obviously these steps are required for the complete elucidation of the structure and hence the function of the binding protein at the molecular level. Furthermore, the study has suggested possibilities and raised interesting questions regarding the involvement of this and other binding proteins in the processes of transport and chemotaxis. First of all, what is the function of the two separate domains? Have they evolved just to combine and produce an elongated molecule from the usually more stable and predominant globular folding pattern? This explanation would still leave in question the need for elongated molecule. The elongated nature of a binding protein has also been observed by Langridge et al. (1970). They reported that the sulfate binding protein from S. typhimurium has an axial ratio of about four to one. In addition, Moffat et al. (1975) have found that the Ca++ binding protein from bovine intestine may have psuedo-symmetry or two structural domains. Perhaps this elongation has fulfilled a requirement for the binding protein to span a large area with a minimum number of amino acids (maximum efficiency).

Another possibility is that the two domains provide for an easily induced conformational change necessary for transport and by extension, chemotaxis. The conformational change could then affect the the interaction of the binding protein with other components or affect the affinity of the binding protein for substrate, or both. There are only three polypeptide chains which connect the two domains. It is conceivable that significant conformational changes could take place in the molecule by shifting one domain relative to the other. The relationship between

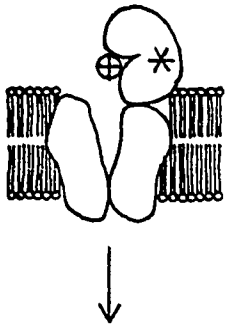
domains is especially important since the sugar binidng site is presumably between the domains. An extension of the study to higher resolution, and the examination of more controlled sugar binding conditions, should allow these and other similar questions concerning the nature of ligand binding and any resulting conformational changes to be answered.

This study has also shown that the predominantly charged heavy atoms bind almost exclusively to one side of the molecule. Since the sequence contains no regions devoid of charged residues, a totally 'hydrophobic' side of the protein is not possible. However, it is possible that the charged residues on the side of the molecule with no heavy atoms bound are paired in salt links, resulting in a generally neutral region. If this is true, then it may be that the apparent paradox of a soluble protein being involved in a membrane function has been resolved. The protein may be associated with the membrane on only one side. This weak association would be consistent with the 'shockability' of this class of proteins, as opposed to other proteins which are firmly embedded in the membrane.

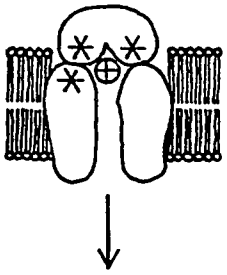
Finally, a conceptual model for binding protein mediated transport which is consistent with all the experimental facts (as discussed in the introduction) is given in Figure 23. It has been adapted and modified from schemes presented by Singer (1974) and Silhavy et al. (1974).



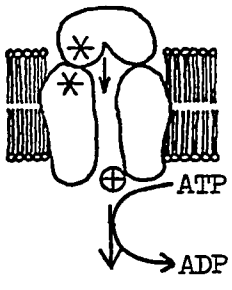

1. Initially, substrate(⊕), binding protein(BP), and membrane protein(MP) are in the ready state. The binding protein may be free in the periplasmic space or weakly associated with the membrane protein.



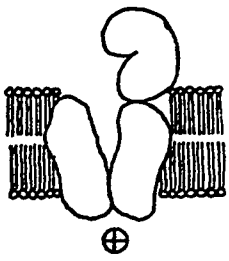

2. The binding of substrate results in a primary conformational change (*) in the binding protein which allows it to interact (further) with the membrane protein.




3. This interaction produces a secondary conformational change in the binding protein which decreases the affinity of the protein for substrate. A conformational change in the membrane protein permits the substrate access to the cytosol.




4. The substrate is released, and diffuses across the membrane. Finally, the system is restored to the higher energy ready state by the hydrolysis of ATP.




5. The cycle is completed, with the net result of the movement of substrate into the cytosol, and the hydrolysis of ATP.

Figure 23. Suggested mechanism for the role of binding proteins in transport.

## APPENDIX A. ABSORPTION CORRECTION PROGRAM

Given the omega scan program provided by Syntex ($P2_1$ recenter/omega scan version 9/11/74), the modifications below allow for the automatic collection of data for the semi-empirical absorption correction described previously.

To use the program, one must first obtain the usual orientation matrix. The indices of any Ø-independent reflection are then entered in the location shown. Other parameters such as scan rate, $\Delta\omega$, etc., can be varied as usual. The execution command (!) will initiate the data collection. The results will be in standard omega scan output format, with intensity measurements every 10° in Ø, starting with 0° and ending at 350°.

```
@216, #h,k,l (of  Ø-independent reflection)
@240, #36
@246, *15003,2274
@1606, *401
@1641, *401
@1660, *401
@2233, *2246
@14344, #0
@15003, *40475, 44476, 50477, 54500, 20503, 24504, 104477, 24502,
 44066, 20462, 24463, 30464, 34465, 2247
@15112, #-10.0, 10.0, 270.0
```

APPENDIX B. UNIT CELL PARAMETERS

| Derivative | # of Crystals | a | b | c |
|---|---|---|---|---|
| Native (Maleate buffer) | 7 | 55.46(09) | 71.82(08) | 77.76(13) |
| Native (Phosphate buffer) | 4 | 55.43(10) | 71.76(31) | 77.80(40) |
| PCMBS | 5 | 55.88(06) | 71.88(06) | 77.85(06) |
| $CdI_2$ | 4 | 55.74(04) | 71.83(08) | 78.02(06) |
| Mercurinitrophenol | 4 | 55.52(10) | 72.01(10) | 77.91(14) |
| $(NH_4)_2PtCl_4$ | 5 | 55.61(16) | 71.75(31) | 77.44(33) |
| $K_2IrCl_6$ | 3 | 55.32(04) | 71.81(03) | 77.83(11) |
| $K_2IrCl_6$' | 1 | 55.66 | 72.05 | 77.93 |
| $NaAuCl_4$ | 2 | 55.62(02) | 71.82(06) | 77.64(04) |

APPENDIX C. HEAVY ATOM DERIVATIVE ATTEMPTS

| Compound | (source) | Conc. (mM) | Soak time | Result |
|---|---|---|---|---|
| Bi subsalicylate | (1) | | | Not soluble |
| $CdI_2$ | (2) | 1 | 2 weeks | Low substitution |
| $CdI_2$ | | 1 | 3 weeks | Low substitution |
| $CdI_2$ | | 3 | 17 days* | Good to at least 3.5 Å |
| $CH_3CH_2HgCl$ | (3) | 2 | 8 days | Low substitution |
| $CH_3HgI$ | (3) | 2 | 9 days | Low substitution |
| Chloromercuri-salicylaldehyde | (4) | 2 | 2 weeks | Low substitution |
| Dichloromercuri-nitrophenol | (4) | (sat.) | 2 weeks | Low substitution |
| $EuCl_3$ | (3) | 1 | 13 days | No substitution |
| Fluorescein-mercuric acetate | (5) | (sat.) | 2 weeks | Low substitution |
| $Hg_3$ aniline | (4) | | | Not soluble |
| $HgI_2$ | (6) | 1 | 1 week | Low substitution |
| Hg salicylate | (4) | (sat.) | 3 weeks | Low substitution |
| Hg (Baker's)† | (4) | | | Not soluble |
| $IrCl_3$ | (3) | .75 | 1 week | No substitution |
| $IrCl_3$ | | 1 | 2 months | No substitution |
| $KAuCN_2$ | (3) | 20 | 8 days | No substitution |
| $K(CH_2CH_2PtCl_4)$ | (3) | 5 | | Crystals cracked |
| KI | (7) | 1 | 18 days | No substitution |
| $K_2IrCl_6$ | (1) | 1 | 11 days | Good to 5 Å |
| $K_2IrCl_6$ | | 1 | 6 weeks | Good to 5 Å |
| $K_2PtCl_4$ | (3) | 1 | (various) | Disordered crystals |
| $K_2PtCl_4$ | | 0.25 | 3 days | Disordered crystals |
| $K_3UO_2F_5$ | (4) | 0.25 | 3 days | No substitution |
| $K_3UO_2F_5$ | | 0.25 | 1 week | Low substitution |
| $LiI_2$ salicylate | (8) | 2.5 | 6 days | No substitution |
| Mersalyl Na | (4) | 2 | 2 weeks | Fair substitution |
| Mercurinitro-phenol | (4) | 2 | 2 weeks | Good to at least 3.5 Å |
| $NaAuCl_4$ | (1) | 4 | | Crystals cracked |

APPENDIX C. (continued)

| Compound | (source) | Conc. (mM) | Soak time | Result |
|---|---|---|---|---|
| $NaAuCl_4$ | | 0.5 | 2 weeks | Fair to 5 Å |
| $(NH_4)_2IrCl_6$ | (3) | 0.5 | 1 week | No substitution |
| $(NH_4)_2PtCl_6$ | | 0.1 | 1 week | Good to 3.5 Å |
| $Pb(ClO_4)_2$ | (1) | | | Not soluble |
| $Pb(SCN)_2$ | (3) | | | Not soluble |
| $Pb(NO_3)$ | (7) | 0.5 | | Heavy precipitate |
| PCMBS§ | (4) | 2 | 12 days | Good to at least 3.5 Å |
| PCMBS/(backwash) | | 2 | 7/3 days | One site only |
| PCMBS/$CdI_2$ | | 2/3 | 2 weeks | $CdI_2$ sites only |
| Pt ethylamine-diamine | (3) | 5 | 2 weeks | No substitution |
| $SmCl_3$ | (3) | 0.3 | 2 weeks | No substitution |
| TlF | (1) | 5 | 2 weeks | No substitution |
| $UO_2$ acetate | (10) | 1 | | Crystals cracked |
| $UO_2$ acetate | | 0.2 | | Crystals cracked |
| $UO_2$ formate | (8) | 0.3 | 1 week | No substitution |
| $UO_2(NO_3)_2$ | (1) | 1 | | Crystals cracked |
| $UO_2(NO_3)_2$ | | 0.2 | | Crystals cracked |

*Since slight precipitation occured, after 10 days the crystals were transferred to a clean vial containing some of the supernatant of the solution.

†Baker's mercurial: $CH_3\overset{\overset{O}{\|}}{C}HgCH_2\overset{\overset{CH_3O}{|}}{C}H\overset{\overset{OCH_3}{|}}{C}HCH_2HgO\overset{\overset{O}{\|}}{C}CH_3$

§p-Chloromercuribenzenesulfonic acid.

(1) Research Organic/Inorganic Chemical; (2) Baker Chemical; (3) Alfa Inorganic; (4) F.A. Quiocho, personal communication; (5) United States Biochemical; (6) ICN·K&K Laboratories; (7) Malinckrodt Chemical; (8) Eastman Kodak; (9) Sigma Chemical; (10) Fisher Scientific.

## APPENDIX D. DIFFERENCE PATTERSON AND FOURIER MAPS

Following are selected features of the difference Patterson maps and difference Fourier maps for the derivatives used in the 5 Å and 3.5 Å resolution analyses. Since Patterson maps are not biased in any way by phases, their interpretability and clarity are vital in guaranteeing the correct solution of the heavy atom positions, and hence correct phasing of the structure factor amplitudes. The unambiguous interpretations of the mercurinitrophenol and PCMBS difference Patterson maps prove that the phasing based on them is valid. In addition, the other more complicated derivatives, Pt and $CdI_2$, show peaks in most places where they are expected based on difference Fourier solutions. The Au and Ir derivatives have very low substitutuion, and the failure of the Patterson maps to show the sites is understandable.

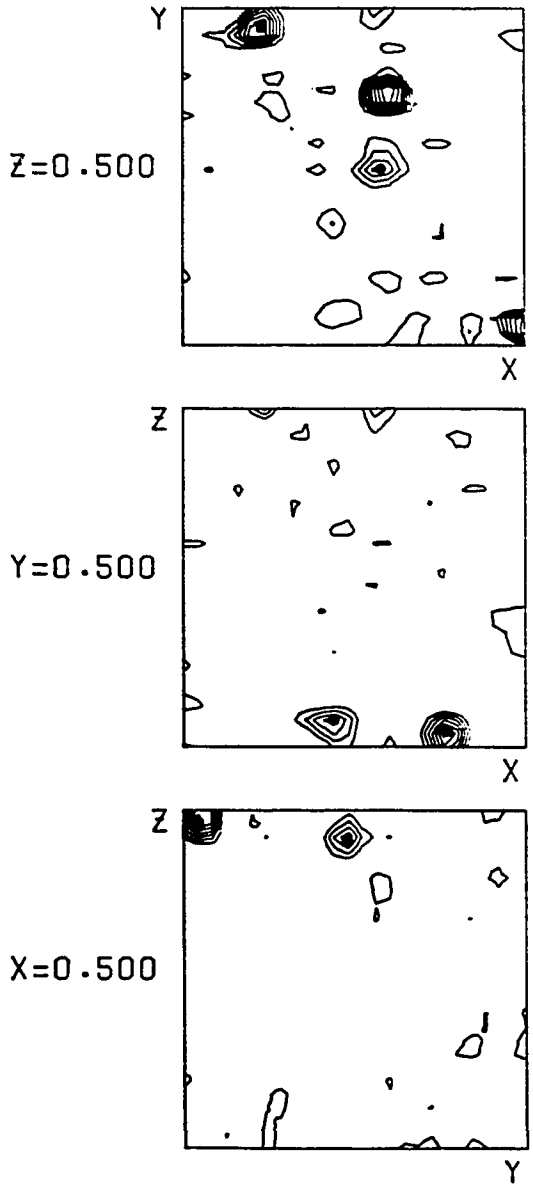
All Patterson maps were calculated by setting the average $\hat{F}_{PH}$ equal to the average $\hat{F}_P$ times 1.05, to allow for the increased number of electrons in the unit cell. All 3.5 Å Patterson maps were contoured at values of 0.04, 0.08, 0.16, 0.24, 0.32, 0.64, and 0.80 $e^2/Å^{3*}$. The 5 Å maps (iridium only) were contoured at 0.02, 0.03, 0.04, 0.08, 0.16, 0.32, and 0.48 $e^2/Å^{3*}$.

For the difference Fourier maps, sections of electron density have been calculated through the center of the major heavy atom sites of each derivative, using the final 3.5 Å phases. Derivative data were scaled to the native data by using the least squares refined scale factors. Contour levels were chosen in arbitrary by equal increments for a given derviative.

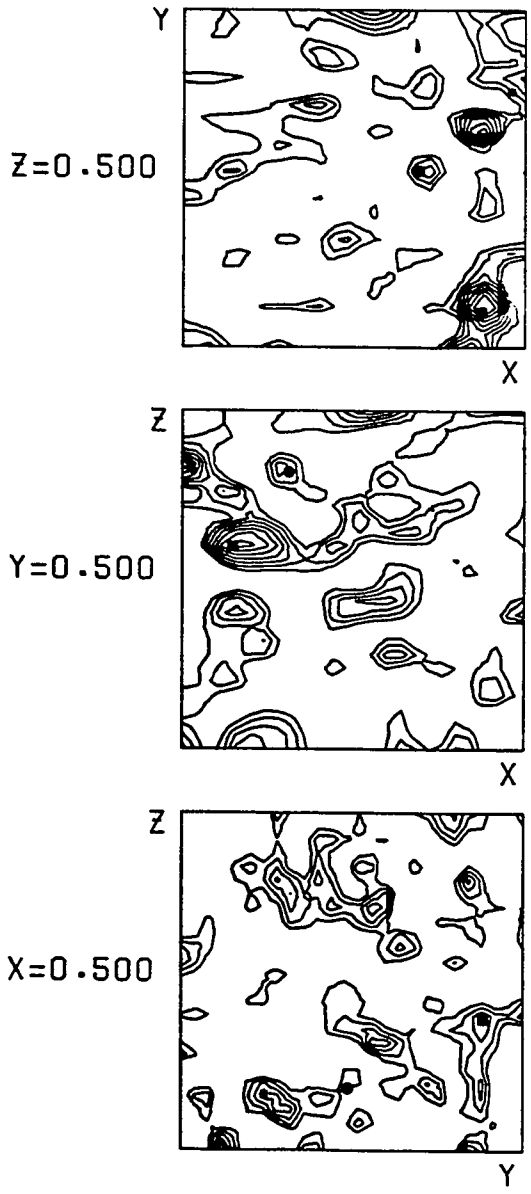
The bounds of the maps in the directions not sectioned are always 0.00 to 0.50.

---

*$F_{000}$ uncorrected.

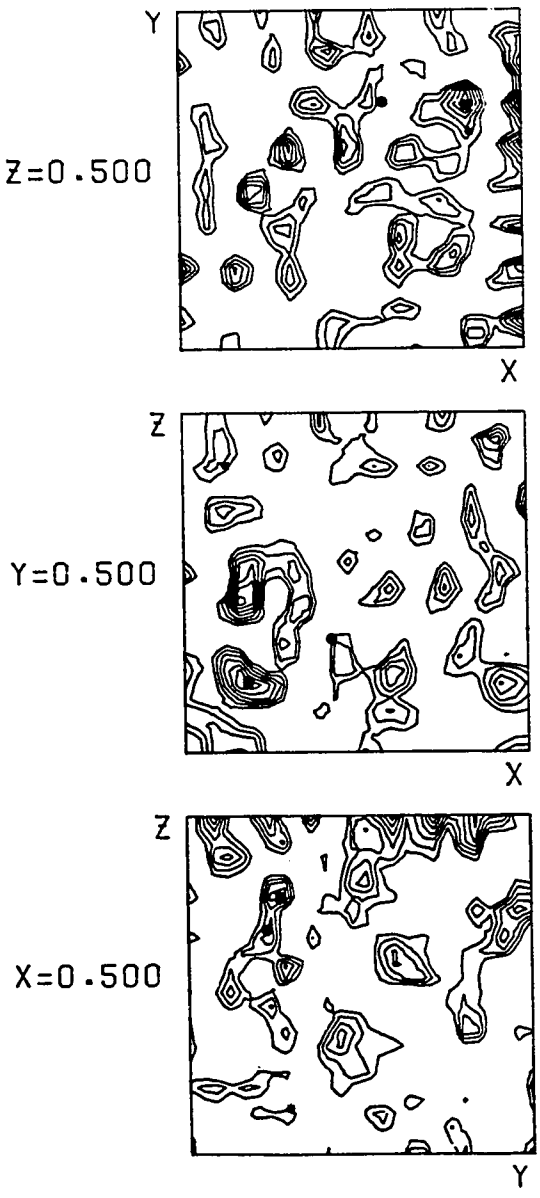



PCMBS difference Patterson map
■ = site 1, ● = site 2



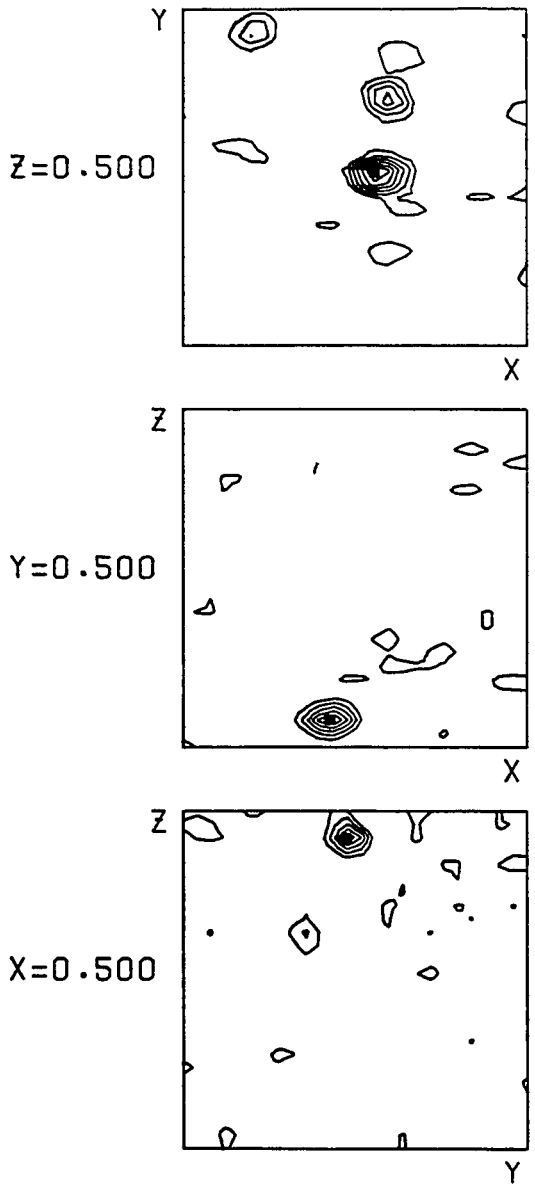

$CdI_2$ difference Patterson map

■= site 1, ●= site 2, ✱= site 3



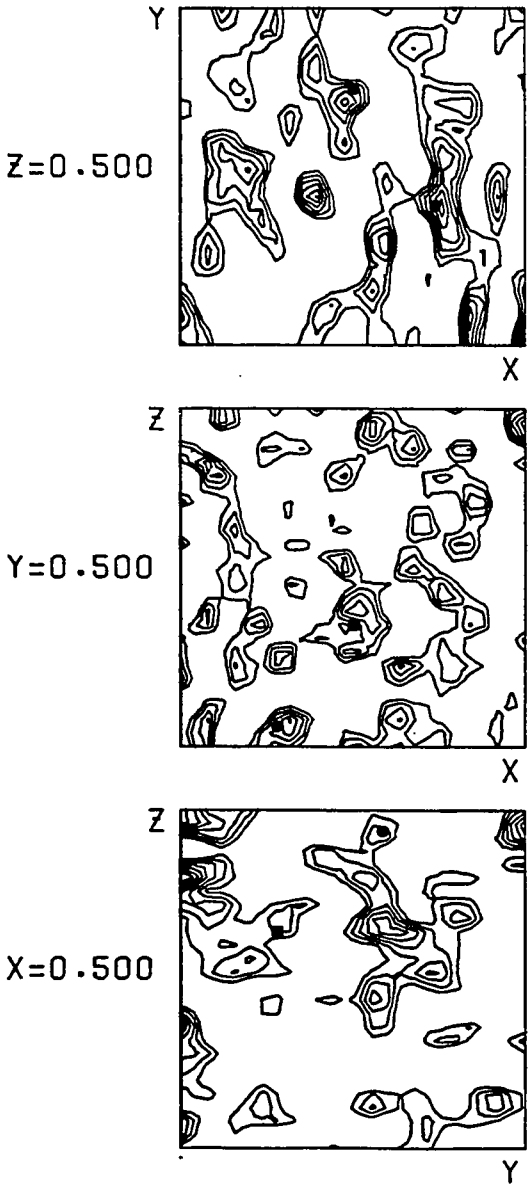

$(NH_4)_2PtCl_4$ difference Patterson map

■ = site 1, ● = site 2



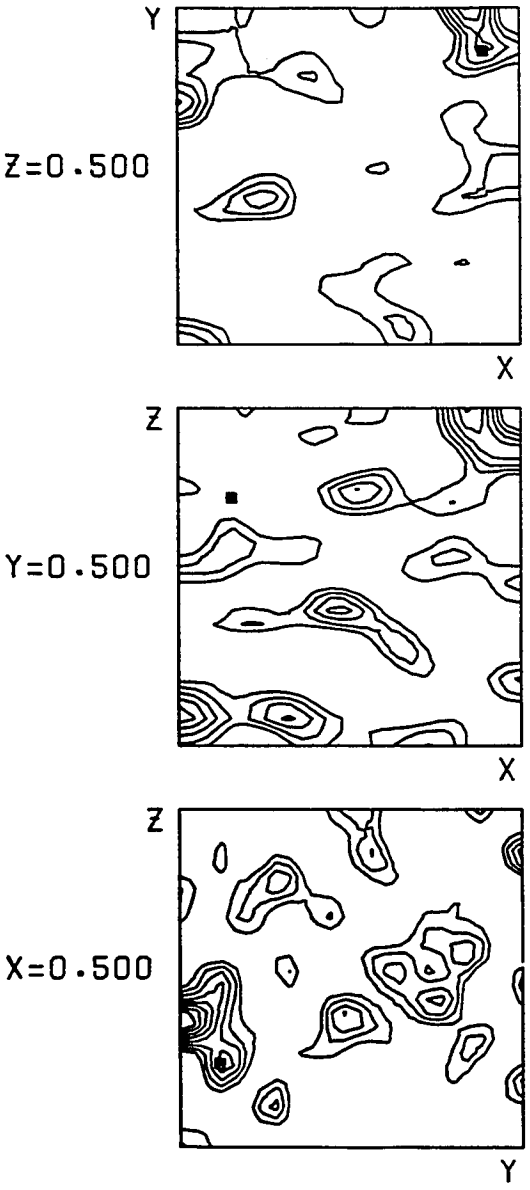

Mercurinitrophenol difference Patterson map

■ = site 1



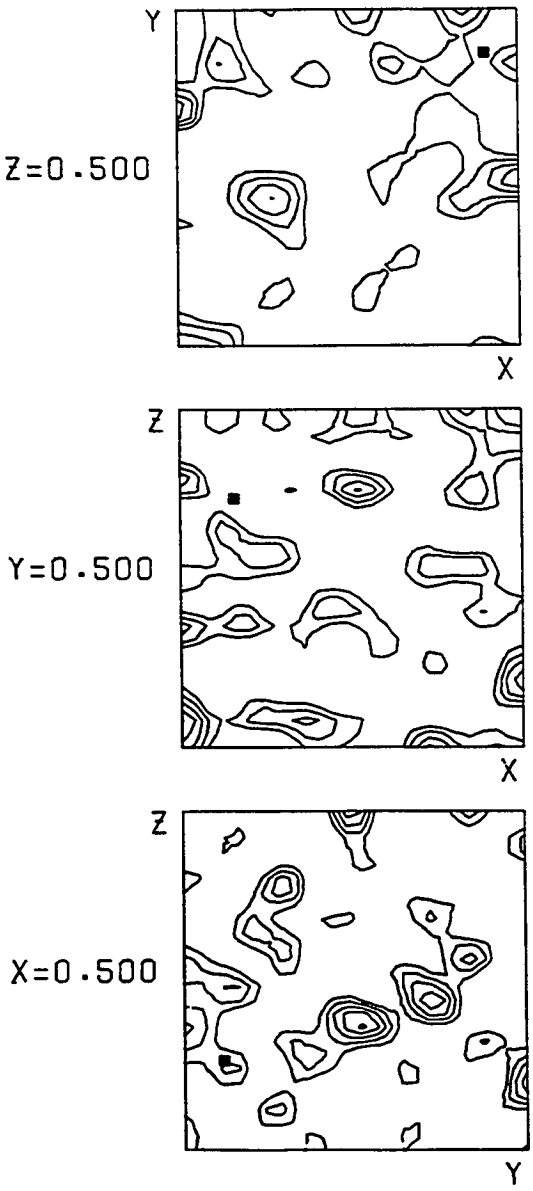

$NaAuCl_4$ difference Patterson map

■ = site 1, ● = site 2



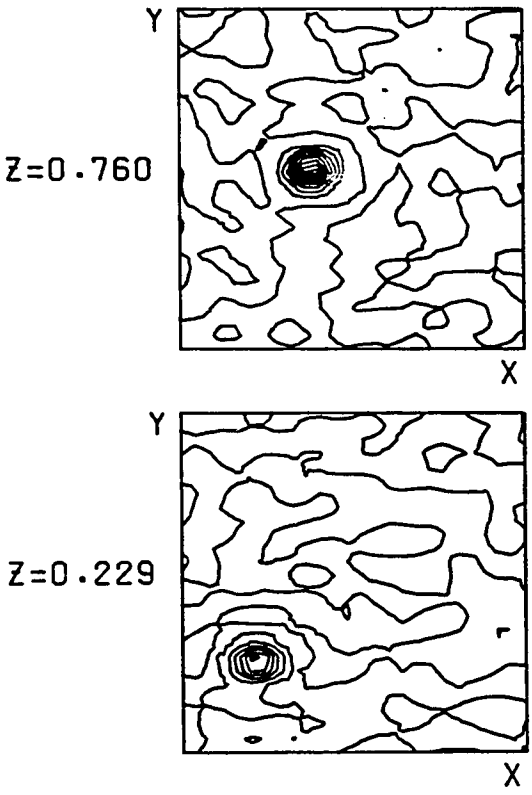

$K_2IrCl_6$ (11-day soak) difference Patterson map

■ = site 1



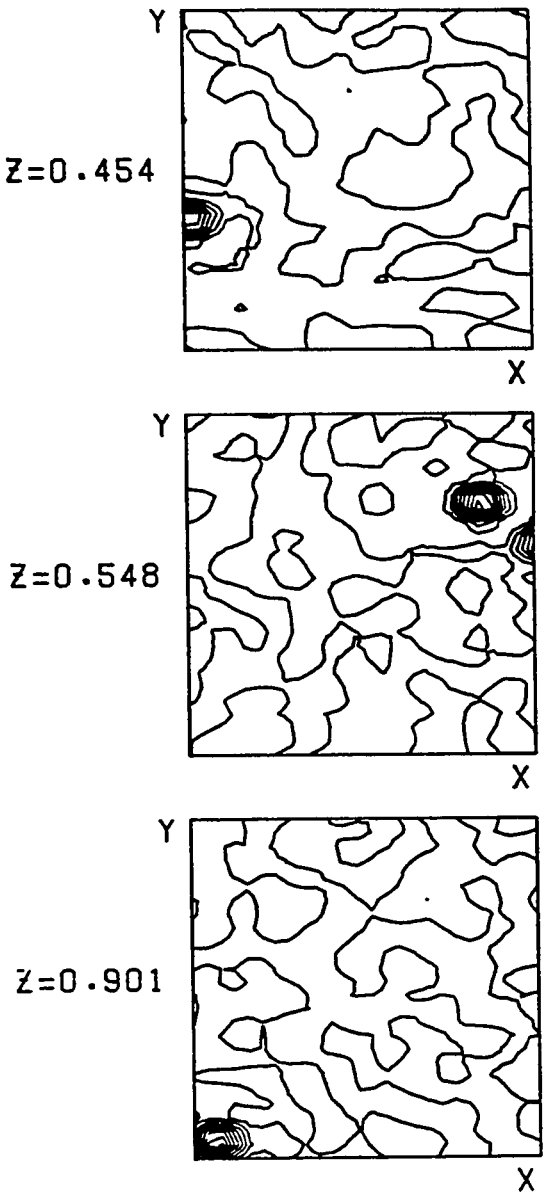

$K_2IrCl_6$ (6-week soak) difference map

■ = site 1



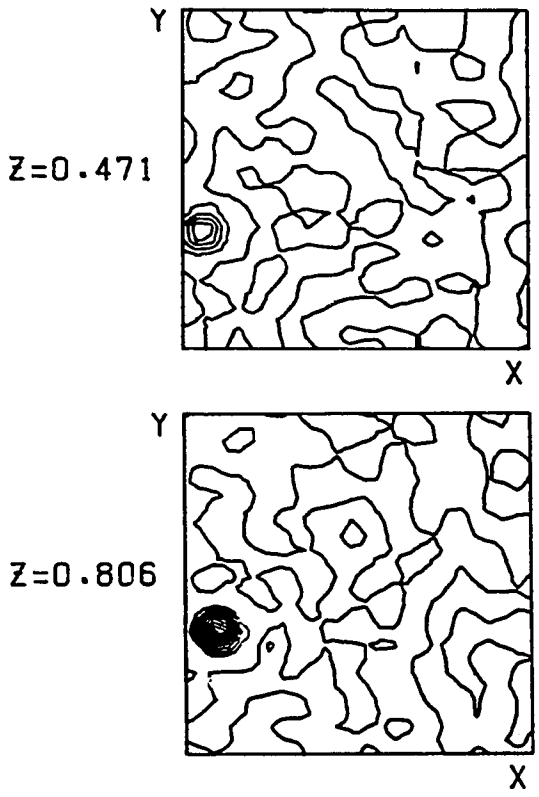

PCMBS difference Fourier map

site 1 (bottom) and site 2 (top)



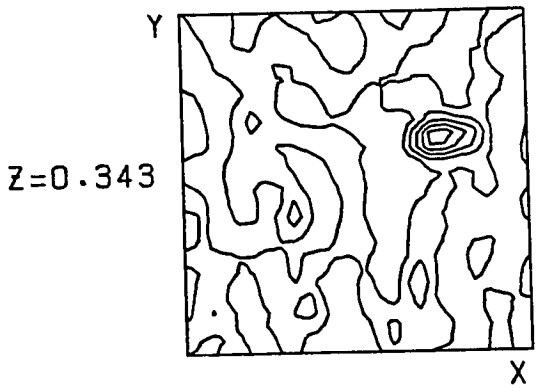

$CdI_2$ difference Fourier map

sites 1, 2, and 3 (bottom to top)



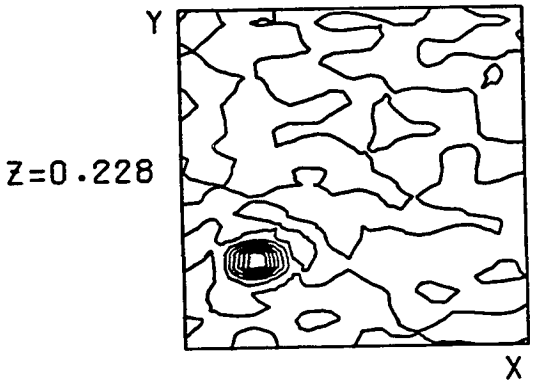

$(NH_4)_2PtCl_4$ difference Fourier map

site 1 (bottom) and site 2 (top)



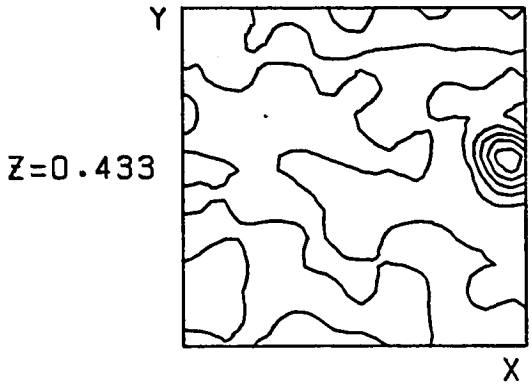

$NaAuCl_4$ difference Fourier map

site 1



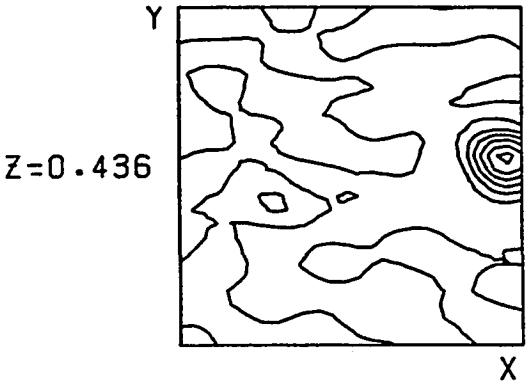

Mercurinitrophenol difference Fourier map

site 1




$K_2IrCl_6$ (11-day soak) difference Fourier map

site 1




$K_2IrCl_4$ (6-week soak) difference Fourier map

site 1

APPENDIX E. SECONDARY STRUCTURE AND COORDINATES OF α-CARBON ATOMS

| Amino Acid # | Secondary Structure Observed | Secondary Structure Predicted | X | Y | Z |
|---|---|---|---|---|---|
| 1 | | | 0.390 | 0.325 | 0.698 |
| 2 | | | 0.340 | 0.350 | 0.698 |
| 3 | | | 0.320 | 0.390 | 0.698 |
| 4 | | | 0.360 | 0.430 | 0.713 |
| 5 | | | 0.330 | 0.485 | 0.713 |
| 6 | β-sheet P-2 | | 0.290 | 0.520 | 0.724 |
| 7 | | | 0.260 | 0.575 | 0.724 |
| 8 | | β -sheet | 0.250 | 0.625 | 0.734 |
| 9 | | | 0.200 | 0.665 | 0.734 |
| 10 | | | 0.175 | 0.720 | 0.741 |
| 11 | | | 0.105 | 0.732 | 0.741 |
| 12 | | | 0.075 | 0.715 | 0.708 |
| 13 | | | 0.120 | 0.720 | 0.665 |
| 14 | | | 0.090 | 0.715 | 0.622 |
| 15 | | | 0.125 | 0.720 | 0.590 |
| 16 | | | 0.180 | 0.709 | 0.626 |
| 17 | | | 0.146 | 0.667 | 0.633 |
| 18 | | | 0.144 | 0.642 | 0.587 |
| – | | | – | – | – |
| 19 | | α -helix | 0.210 | 0.655 | 0.572 |
| 20 | α-helix a | | 0.235 | 0.635 | 0.623 |
| 21 | | | 0.192 | 0.590 | 0.620 |
| 22 | | | 0.210 | 0.580 | 0.572 |
| 23 | | | 0.275 | 0.590 | 0.585 |
| 24 | | | 0.280 | 0.560 | 0.623 |
| 25 | | | 0.245 | 0.520 | 0.595 |
| 26 | | | 0.285 | 0.525 | 0.566 |
| 27 | | | 0.300 | 0.530 | 0.536 |
| 28 | | | 0.340 | 0.500 | 0.514 |
| 29 | | | 0.380 | 0.475 | 0.492 |
| 30 | | | 0.410 | 0.460 | 0.536 |
| 31 | | | 0.375 | 0.460 | 0.536 |
| 32 | | | 0.310 | 0.465 | 0.468 |
| 33 | | | 0.310 | 0.475 | 0.611 |
| 34 | | | 0.300 | 0.450 | 0.660 |
| 35 | | | 0.265 | 0.505 | 0.670 |
| 36 | | α -helix | 0.210 | 0.515 | 0.702 |
| – | | | – | – | – |
| 37 | β-sheet P-1 | | 0.190 | 0.550 | 0.713 |
| 38 | | | 0.173 | 0.590 | 0.774 |
| 39 | | | 0.130 | 0.620 | 0.766 |
| 40 | | | 0.150 | 0.650 | 0.798 |
| 41 | | | 0.160 | 0.690 | 0.806 |
| 42 | | | 0.175 | 0.700 | 0.838 |
| 43 | | | 0.182 | 0.695 | 0.880 |
| 44 | | | 0.202 | 0.665 | 0.918 |
| 45 | | | 0.165 | 0.630 | 0.905 |
| 46 | | | 0.190 | 0.630 | 0.856 |

| Amino Acid # | Secondary Structure Observed | Secondary Structure Predicted | X | Y | Z |
|---|---|---|---|---|---|
| 47 | | | 0.242 | 0.630 | 0.872 |
| 48 | α -helix b | | 0.235 | 0.590 | 0.905 |
| 49 | | | 0.205 | 0.560 | 0.872 |
| 50 | | α-helix | 0.251 | 0.560 | 0.836 |
| 51 | | | 0.300 | 0.560 | 0.869 |
| 52 | | | 0.275 | 0.520 | 0.884 |
| 53 | | | 0.251 | 0.500 | 0.840 |
| 54 | | | 0.317 | 0.502 | 0.824 |
| - | | | - | - | - |
| 55 | | | 0.350 | 0.470 | 0.863 |
| 56 | | | 0.260 | 0.430 | 0.816 |
| 57 | | | 0.325 | 0.425 | 0.784 |
| 58 | | | 0.385 | 0.470 | 0.777 |
| 59 | | β -sheet | 0.370 | 0.525 | 0.756 |
| 60 | β -sheet P-3 | | 0.350 | 0.570 | 0.756 |
| 61 | | | 0.350 | 0.620 | 0.734 |
| 62 | | | 0.310 | 0.670 | 0.756 |
| 63 | | | 0.260 | 0.710 | 0.766 |
| 64 | | | 0.225 | 0.725 | 0.816 |
| 65 | | | 0.270 | 0.760 | 0.827 |
| 66 | | | 0.325 | 0.770 | 0.838 |
| 67 | | | 0.350 | 0.750 | 0.870 |
| 78 | | | 0.340 | 0.710 | 0.870 |
| 79 | | | 0.410 | 0.680 | 0.892 |
| 70 | | | 0.430 | 0.635 | 0.888 |
| 71 | | | 0.450 | 0.590 | 0.917 |
| 72 | | α-helix | 0.385 | 0.575 | 0.913 |
| - | | | - | - | - |
| 73 | α-helix c | | 0.408 | 0.570 | 0.864 |
| 74 | | | 0.470 | 0.555 | 0.867 |
| 75 | | | 0.452 | 0.515 | 0.897 |
| 76 | | | 0.400 | 0.500 | 0.872 |
| 77 | | | 0.440 | 0.480 | 0.838 |
| 78 | | | 0.440 | 0.520 | 0.809 |
| 79 | | | 0.440 | 0.560 | 0.777 |
| 80 | | | 0.420 | 0.620 | 0.766 |
| 81 | β-sheet P-4 | | 0.400 | 0.660 | 0.734 |
| 82 | | | 0.370 | 0.680 | 0.713 |
| 83 | | | 0.325 | 0.700 | 0.692 |
| 84 | | | 0.300 | 0.750 | 0.692 |
| 85 | | β-sheet | 0.300 | 0.750 | 0.698 |
| 86 | | | 0.340 | 0.770 | 0.719 |
| 87 | | | 0.300 | 0.775 | 0.773 |
| 88 | | | 0.370 | 0.715 | 0.773 |
| 89 | | | 0.415 | 0.775 | 0.795 |
| 90 | | | 0.435 | 0.760 | 0.827 |
| - | | | - | - | - |
| 91 | | | 0.410 | 0.810 | 0.838 |
| 92 | | | 0.460 | 0.835 | 0.849 |
| 93 | | | 0.490 | 0.785 | 0.827 |
| 94 | | | 0.510 | 0.760 | 0.806 |
| 95 | | | 0.550 | 0.725 | 0.795 |

| Amino Acid # | Secondary Structure Observed | Secondary Structure Predicted | X | Y | Z |
|---|---|---|---|---|---|
| 96 | | | 0.540 | 0.675 | 0.784 |
| 97 | | | 0.475 | 0.710 | 0.773 |
| 98 | | | 0.490 | 0.670 | 0.745 |
| 99 | | β-sheet | 0.450 | 0.710 | 0.713 |
| 100 | β-sheet P-5 | | 0.415 | 0.710 | 0.670 |
| 101 | | | 0.380 | 0.740 | 0.660 |
| 102 | | | 0.350 | 0.800 | 0.639 |
| 103 | | | 0.380 | 0.840 | 0.633 |
| 104 | | β-sheet | 0.380 | 0.870 | 0.668 |
| 105 | | | 0.382 | 0.922 | 0.668 |
| 106 | | ( α-helix ) | 0.415 | 0.927 | 0.626 |
| 107 | | | 0.375 | 0.900 | 0.595 |
| 108 | | | 0.315 | 0.925 | 0.616 |
| - | | | - | - | - |
| 109 | | | 0.340 | 0.974 | 0.615 |
| 110 | α-helix d | | 0.360 | 0.967 | 0.551 |
| 111 | | | 0.302 | 0.945 | 0.566 |
| 112 | | | 0.259 | 0.980 | 0.571 |
| 113 | | (α-helix) | 0.290 | 1.020 | 0.550 |
| 114 | | | 0.300 | 1.000 | 0.505 |
| 115 | | | 0.230 | 0.985 | 0.497 |
| 116 | | | 0.210 | 1.030 | 0.519 |
| 117 | | | 0.250 | 1.070 | 0.491 |
| 118 | | | 0.215 | 1.075 | 0.482 |
| 119 | | | 0.125 | 1.040 | 0.482 |
| 120 | | α-helix | 0.140 | 1.005 | 0.471 |
| 121 | | | 0.100 | 1.005 | 0.449 |
| 122 | | | 0.075 | 1.025 | 0.482 |
| 123 | | | 0.085 | 1.080 | 0.471 |
| 124 | | | 0.090 | 1.125 | 0.482 |
| 125 | | | 0.100 | 1.175 | 0.492 |
| 126 | | | 0.100 | 1.210 | 0.525 |
| - | | | | - | - |
| 127 | | α-helix | 0.130 | 1.180 | 0.600 |
| 128 | | | 0.070 | 1.200 | 0.568 |
| 129 | | | 0.080 | 1.150 | 0.617 |
| 130 | | | 0.115 | 1.110 | 0.649 |
| 131 | β-sheet Q-5 | | 0.130 | 1.060 | 0.670 |
| 132 | | | 0.125 | 1.020 | 0.702 |
| 133 | | β-sheet | 0.160 | 0.980 | 0.745 |
| 134 | | | 0.180 | 0.940 | 0.777 |
| 135 | | | 0.230 | 0.920 | 0.762 |
| 136 | | | 0.255 | 0.958 | 0.806 |
| 137 | | | 0.280 | 0.900 | 0.816 |
| 138 | | | 0.290 | 0.860 | 0.784 |
| 139 | | | 0.360 | 0.870 | 0.752 |
| 140 | | | 0.340 | 0.860 | 0.708 |
| 141 | | α-helix | 0.280 | 0.890 | 0.719 |
| 142 | α-helix e | | 0.325 | 0.920 | 0.719 |
| 143 | | | 0.340 | 0.920 | 0.676 |
| 144 | | | 0.285 | 0.950 | 0.665 |
| - | | | - | - | - |

| Amino Acid # | Secondary Structure Observed | Secondary Structure Predicted | X | Y | Z |
|---|---|---|---|---|---|
| 145 | | | 0.310 | 0.990 | 0.687 |
| 146 | | | 0.330 | 1.000 | 0.645 |
| 147 | | | 0.275 | 1.020 | 0.633 |
| 148 | | | 0.300 | 1.060 | 0.665 |
| 149 | | | 0.310 | 1.075 | 0.611 |
| 150 | | | 0.360 | 1.090 | 0.611 |
| 151 | | | 0.370 | 1.100 | 0.579 |
| 152 | | | 0.320 | 1.140 | 0.568 |
| 153 | | α-helix | 0.275 | 1.150 | 0.600 |
| 154 | | | 0.240 | 1.160 | 0.633 |
| 155 | | | 0.240 | 1.175 | 0.665 |
| 156 | | | 0.240 | 1.140 | 0.687 |
| 157 | | | 0.175 | 1.130 | 0.692 |
| 158 | | | 0.160 | 1.080 | 0.734 |
| 159 | β-sheet Q-6 | | 0.150 | 1.030 | 0.766 |
| 160 | | | 0.125 | 0.990 | 0.787 |
| 161 | | | 0.120 | 0.930 | 0.809 |
| 162 | | | 0.060 | 0.875 | 0.816 |
| - | | β-sheet | - | - | - |
| 163 | | | 0.030 | 0.840 | 0.773 |
| 164 | | | 0.110 | 0.850 | 0.741 |
| 165 | | | 0.030 | 0.845 | 0.730 |
| 166 | | | 0.005 | 0.850 | 0.709 |
| 167 | | | 0.020 | 0.865 | 0.757 |
| 168 | | | 0.040 | 0.895 | 0.748 |
| 169 | | | 0.020 | 0.915 | 0.708 |
| 170 | | | 0.040 | 0.928 | 0.721 |
| 171 | | | 0.030 | 0.955 | 0.763 |
| 172 | | | 0.025 | 0.975 | 0.747 |
| 173 | α-helix f | | 0.010 | 1.000 | 0.708 |
| 174 | | | 0.062 | 1.010 | 0.735 |
| 175 | | | 0.020 | 1.032 | 0.765 |
| 176 | | | 0.020 | 1.065 | 0.744 |
| 177 | | (α-helix) | 0.036 | 1.075 | 0.724 |
| 178 | | | 0.076 | 1.095 | 0.736 |
| 179 | | | 0.020 | 1.125 | 0.763 |
| 180 | | | 0.004 | 1.150 | 0.721 |
| - | | | - | - | - |
| 181 | | | 0.070 | 1.155 | 0.709 |
| 182 | | α-helix | 0.060 | 1.180 | 0.676 |
| 183 | | | 0.020 | 1.160 | 0.665 |
| 184 | | | 0.010 | 1.145 | 0.611 |
| 185 | | | 0.025 | 1.110 | 0.607 |
| 186 | β-sheet Q-4 | | 0.050 | 1.055 | 0.607 |
| 187 | | | 0.100 | 1.025 | 0.628 |
| 188 | | | 0.125 | 0.980 | 0.649 |
| 189 | | | 0.185 | 0.980 | 0.654 |
| 190 | | | 0.165 | 0.925 | 0.654 |
| 191 | | | 0.200 | 0.875 | 0.665 |
| 192 | | β-sheet | 0.190 | 0.860 | 0.633 |

| Amino Acid # | Secondary Structure Observed | Secondary Structure Predicted | X | Y | Z |
|---|---|---|---|---|---|
| 193 | | | 0.105 | 0.845 | 0.623 |
| 194 | | | 0.080 | 0.845 | 0.649 |
| 195 | | | 0.090 | 0.890 | 0.685 |
| 196 | | | 0.085 | 0.920 | 0.642 |
| 197 | | | 0.030 | 0.920 | 0.616 |
| 198 | | | 0.500 | 0.500 | 0.500 |
| - | | | - | - | - |
| 199 | | | 0.030 | 0.920 | 0.616 |
| 200 | α-helix g | | 0.015 | 0.892 | 0.629 |
| 201 | | β-sheet | 0.002 | 0.900 | 0.672 |
| 202 | | | 0.015 | 0.950 | 0.672 |
| 203 | | | 0.010 | 0.960 | 0.628 |
| 204 | | | 0.070 | 0.930 | 0.640 |
| 205 | | | 0.080 | 0.960 | 0.681 |
| 206 | | | 0.060 | 1.010 | 0.663 |
| 207 | | | 0.102 | 1.000 | 0.623 |
| 208 | | | 0.154 | 0.980 | 0.649 |
| 209 | | | 0.150 | 1.020 | 0.680 |
| 210 | | | 0.150 | 1.020 | 0.644 |
| 211 | | | 0.120 | 1.040 | 0.611 |
| 212 | | | 0.125 | 1.060 | 0.579 |
| 213 | | | 0.085 | 1.040 | 0.546 |
| 214 | | | 0.040 | 1.020 | 0.514 |
| 215 | | | 0.010 | 1.055 | 0.525 |
| 216 | | | 0.050 | 1.060 | 0.536 |
| - | | β-sheet | - | - | - |
| 217 | | | 0.050 | 1.120 | 0.525 |
| 218 | | | 0.000 | 1.140 | 0.514 |
| 219 | | | 0.040 | 1.120 | 0.546 |
| 220 | β-sheet Q-3 | | 0.020 | 1.065 | 0.564 |
| 221 | | | 0.020 | 1.020 | 0.564 |
| 222 | | | 0.090 | 0.990 | 0.575 |
| 223 | | | 0.110 | 0.940 | 0.585 |
| 224 | | | 0.150 | 0.900 | 0.585 |
| 225 | | | 0.180 | 0.860 | 0.596 |
| 226 | | | 0.140 | 0.790 | 0.579 |
| 227 | | | 0.180 | 0.800 | 0.557 |
| 228 | | α-helix | 0.115 | 0.750 | 0.568 |
| 229 | | | 0.050 | 0.780 | 0.579 |
| 230 | | | 0.050 | 0.790 | 0.536 |
| 231 | | | 0.010 | 0.775 | 0.536 |
| 232 | | | 0.040 | 0.780 | 0.492 |
| 233 | | | 0.100 | 0.800 | 0.492 |
| 234 | | | 0.125 | 0.850 | 0.482 |
| - | | | - | - | - |
| 235 | | | 0.090 | 0.890 | 0.482 |
| 236 | | β-sheet | 0.055 | 0.925 | 0.492 |
| 237 | | | 0.010 | 0.930 | 0.525 |
| 238 | β-sheet Q-2 | | 0.035 | 0.940 | 0.543 |
| 239 | | | 0.090 | 0.915 | 0.554 |
| 240 | | | 0.140 | 0.875 | 0.554 |
| 241 | | | 0.200 | 0.845 | 0.554 |

| Amino Acid # | Secondary Structure Observed | Secondary Structure Predicted | X | Y | Z |
|---|---|---|---|---|---|
| 242 | | | 0.250 | 0.810 | 0.543 |
| 243 | | | 0.290 | 0.775 | 0.536 |
| 244 | | | 0.300 | 0.740 | 0.568 |
| 245 | | | 0.300 | 0.680 | 0.568 |
| 246 | | | 0.345 | 0.675 | 0.566 |
| 247 | | | 0.345 | 0.682 | 0.614 |
| 248 | | | 0.320 | 0.632 | 0.624 |
| 249 | | | 0.365 | 0.615 | 0.587 |
| 250 | | | 0.420 | 0.638 | 0.603 |
| 251 | α-helix h | | 0.405 | 0.625 | 0.649 |
| 252 | | | 0.398 | 0.575 | 0.640 |
| - | | | - | - | - |
| 253 | | | 0.450 | 0.570 | 0.614 |
| 254 | | α-helix | 0.495 | 0.595 | 0.644 |
| 255 | | | 0.470 | 0.568 | 0.686 |
| 256 | | | 0.478 | 0.522 | 0.663 |
| 257 | | | 0.540 | 0.530 | 0.645 |
| 258 | | | 0.540 | 0.530 | 0.686 |
| 259 | | | 0.500 | 0.475 | 0.710 |
| 260 | | | 0.550 | 0.450 | 0.668 |
| 261 | | | 0.540 | 0.685 | 0.784 |
| 262 | | | 0.550 | 0.650 | 0.665 |
| 263 | | | 0.550 | 0.610 | 0.687 |
| 264 | | | 0.580 | 0.570 | 0.687 |
| 265 | | | 0.600 | 0.525 | 0.665 |
| 266 | β-sheet P-6 | | 0.600 | 0.480 | 0.687 |
| 267 | | | 0.410 | 0.780 | 0.585 |
| 268 | | | 0.450 | 0.770 | 0.628 |
| 269 | | | 0.485 | 0.730 | 0.649 |
| 270 | | | 0.510 | 0.725 | 0.692 |
| - | | | - | - | - |
| 271 | | β-sheet | 0.450 | 0.800 | 0.546 |
| 272 | | | 0.390 | 0.825 | 0.546 |
| 273 | | | 0.355 | 0.790 | 0.525 |
| 274 | β-sheet Q-1 | | 0.330 | 0.850 | 0.514 |
| 275 | | | 0.280 | 0.840 | 0.500 |
| 276 | | | 0.225 | 0.850 | 0.490 |
| 277 | | | 0.155 | 0.865 | 0.468 |
| 278 | | | 0.100 | 0.870 | 0.449 |
| 279 | | | 0.050 | 0.820 | 0.428 |
| 280 | | | 0.100 | 0.890 | 0.406 |
| 281 | | | 0.145 | 0.841 | 0.427 |
| 282 | | | 0.135 | 0.815 | 0.467 |
| 283 | α-helix i | α-helix | 0.107 | 0.778 | 0.435 |
| 284 | | | 0.162 | 0.780 | 0.407 |
| 285 | | | 0.195 | 0.774 | 0.445 |
| 286 | | | 0.170 | 0.734 | 0.471 |
| 287 | | | 0.167 | 0.710 | 0.416 |
| 288 | | | 0.232 | 0.714 | 0.415 |
| - | | | - | - | - |
| 289 | | | 0.235 | 0.700 | 0.468 |
| 290 | | | 0.105 | 0.660 | 0.460 |
| 291 | | | 0.160 | 0.650 | 0.460 |

## APPENDIX F. PUBLICATIONS, INCLUDING OTHER RESEARCH

The preceding work has been published at various stages in its development. These include:

Quiocho, F.A., Phillips, G.N., Jr., Parsons, R.G., and Hogg, R.W. (1974), Crystallographic data of an L-arabinose binding protein from E. coli, J. Mol. Biol., 86, 491.

Phillips, G.N., Jr., Mahajan, V.K., Siu, A.K.Q., and Quiocho, F.A. (1975) Crystallographic analysis of an L-arabinose binding protein, Acta. Cryst., A31, 527.

Phillips, G.N., Jr., Mahajan, V.K., Siu, A.K.Q., and Quiocho, F.A. (1976) The Structure of L-arabinose binding protein at 5 Å resolution and preliminary results at 3.5 Å, Proc. Nat. Acad. Sci. USA 73, 2186.

Phillips, G.N., Jr., Mahajan, V.K., and Quiocho, F.A. (1976) The Structure of L-arabinose binding protein at 3.5 Å resolution, Fed. Proc. 35, 1367.

(3.5 Å resolution and chain trace, manuscript in preparation)

In addition, other research has been carried out as follows:

Phillips, G.N., Jr., Quiocho, F.A., Emery, H., Knapp, F.F., Jr., and Schroepfer, G.J., Jr. (1975) Sterol Synthesis. Structure of 3-β-p-bromobenzoyloxy-cholest-8(14)-en-15β-ol, Acta Cryst., A31, S113.

Phillips, G.N. Jr., Quiocho, F.A., Sass, R.L., Werness, P., Emery, H., Knapp, F.F., Jr., and Schroepfer, G.J., Jr. (1976) Sterol Biosynthesis. Establishment of the structure of 3β-p-bromobenzoyloxy-5α-cholest-8(14)-en-15β-ol, Bioorganic Chem. 5, 1.

Phillips, G.N. Jr., and Quiocho, F.A. (in progress) The crystal structure of 1-carboxymethyl-2-imino-3-phosphonoimidazolidine.

# VII. BIBLIOGRAPHY


Adler, J. (1974) in Biochemistry of Sensory Functions, ed. Jaernicke, L., Springer-Verlag, Berlin, pp. 107-131.

Adler, J. (1975) Ann. Rev. Biochem. 44, 341.

Aksamit, R., Koshland, D.E., Jr. (1972) Biochem. Biophys. Res. Commun. 48, 1348.

Ames, G.F.-L., and Lever, A.L. (1970) Proc. Nat. Acad. Sci. U.S.A. 66, 1096.

Ames, G.F.-L., and Spudich, E.N. (1976) Proc. Natl. Acad. Sci. U.S.A. 73, 1877.

Anraku, Y. (1968) J. Biol. Chem. 243, 3123.

Armstrong, J.B. (1972) Can. J. Microbio., 18, 1695.

Berger, E.A., and Heppel, L.A. (1972) J. Biol. Chem. 247, 7684.

Blow, D.M., and Crick, F.H.C. (1959) Acta Cryst. 12, 794.

Blundell, T.L., and Johnson, L.N. (1976) Protein Crystallography, Academic Press, New York.

Bokhoven, C., Schoone, J.C., and Bijvoet, J.M. (1951) Acta Cryst. 4, 275.

Boos, W., and Sarvas, M. (1970) Eur. J. Biochem. 13, 526.

Boos, W., Gordon, A.J., Hall, R.E., and Price, H.D. (1972) J. Biol. Chem. 247, 917.

Boos, W. (1974) Ann. Rev. Biochem. 43, 123.

Bragg, Sir L., and Perutz, M.F. (1954) Proc. Roy. Soc. A225, 315.

Branden, C.I., Holmes, K.C., and Kendrew, J.C. (1963) Acta Cryst. 16, A184.

Catsimpoolas, N. (1975) in Methods of Separation, Vol. I, Plenum Press, New York.

Chou, P.Y., and Fasman, G.D. (1974) Biochem. 13, 222.

Christensen, H.N. (1975) Biological Transport, Benjamin, Inc., Reading Mass.

Dayhoff, M.O. (1963) Commun. ACM 6, 620.

Fox, C.F., and Kennedy, E.P. (1965) Proc. Nat. Acad. Sci. USA 54, 891.

Fukui, S., and Miyairi, S. (1970) J. Bacteriol. 101, 685.

Furlong, C.E., and Weiner, J.H. (1970) Biochem. Biophys. Res. Commun. 38, 1076.

Garoff, G., and Bromwell, K.E. (1971) Arch. Biochem. Biophys. 137, 379.

Harker, D. (1956) Acta Cryst. 9, 1.

Hazelbauer, G.L., and Adler, J. (1971) Nature (Lond.) New Biol. 230, 101.

Henry, N.F.M., and Lonsdale, K., eds. (1969) International Tables for X-ray Crystallography, Kynoch Press, London.

Heppel, L.A. (1969) J. Gen. Physiol. 54, 955.

Hogg, R. W. (1971) J. Bacteriol. 105, 604.

Iwashima, A., Matsumura, A., Nose, Y. (1971) J. Bacteriol. 108, 1419.

Johnson, C.K. (1965) Oak Ridge Nat. Lab. #ORNL-3794, Revised.

Kaback, H.R., Rudnick, G., Schuldiner, S., Short, S.A. (1976) Annals N.Y. Acad. Sci. 264, 350.

Kellerman, O. and Szmelcman, S. (1974) Eur. J. Biochem. 47, 139.

Kendrow, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D.R., Phillips, D.C., and Shore, V.C. (1960) Nature, (Lond.) 185, 422.

Koshland, D.E., Jr. (1974) in Biochemistry of Sensory Functions, ed. Jaernicke, L., Springer-Verlag, Berlin, pp. 133-163.

Kundig, W., and Roseman, S (1966) Methods Enzymol. 9, 396.

Langridge, R., Shinagawa, H., and Pardee, A.B. (1970) Science, 169, 59.

Larsen, S.H., Reader, R.W., Kort, E.N., Tso, W.-W., and Adler, J. (1974) Nature (Lond.) New Biol. 249, 74.

Lever, J.E. (1972) J. Biol. Chem. 247, 4317.

Lipscomb, W.N., Coppola, J.C., Hartsuck, J.A., Ludwig, M.L., Muirhead, H., Searl, J., and Steitz, T.A. (1966) J. Mol. Biol. 19, 423.

Matthews, B.W. (1975) Biochem. Biophy. Acta, 405, 442.

Medveczky, N., and Rosenberg, H. (1970) Biochem. Biophys. Acta, 211, 158.

Miller, D. (1976) Fed. Proc. 35, 1586.

Moffat, K., Fullmer, C.S., and Wasserman, R.H. (1975) J. Mod Biol. 97, 661.

Nakane, P.K., Nichoalds, G.E., and Oxenden, D.L. (1968) Science 161, 182.

North, A.C.T., Phillips, D.C., and Matthews, F.S. (1968) Acta Cryst. Sec. A24, 351.

Novotny, C., and Englesberg, E. (1966) Biochem. Biophys. Acta, 17, 217.

Oxender, D.L. (1972) Ann. Rev. Biochem. 41, 777.

Oxender, D.L., and Quay, S. (1976) Annals N.Y. Acad. Sci., 264, 358.

Pardee, A.B., and Watanabe, K. (1968) J. Bacteriol. 96, 1049.

Parsons, R.G., and Hogg, R. W. (1974) J. Biol. Chem. 249, 3602.

Parsons, R.G. and Hogg, R. W. (1974) J. Biol. Chem. 249, 3608.

Penrose, W.R., Nichoalds, G.E., Piperno, J.R., and Oxender, D.L. (1968) J. Biol. Chem. 243, 5921.

Perutz, M.F. (1963) Nobel Lectures Yearbook, 1962, Elsevier, Amsterdam.

Phillips, D.C. (1966) Adv. Res. Diff. Meth. 2, 75.

Quiocho, F.A., Phillips, G.N., Jr., Parsons, R.G. and Hogg, R.W. (1974) J. Mol. Biol. 86, 491.

Robbins, A.R., and Rotman, B. (1975) Proc. Natl. Acad. Sci. USA 72, 423.

Robson, B., and Pain, R.H. (1971) J. Mod. Biol. 58, 237.

Rosen, B.P. (1971) J. Biol. Chem. 246, 3653.

Rosen, B.P. (1973) J. Biol. Chem. 248, 1211.

Rosen, B.P., and Heppel, L.A. (1973) in Bacterial Membranes and Walls, ed. L. Leive, Marcel Dekker, Inc. New York, pp. 209.

Silhavy, T.J., Boos, W., Kalckar, H.M (1974) 25th Mosbach Colloquium, 165.

Singer, S.J. (1974) Ann. Rev. Biochem, 43, 805.

Stout, G.H. and Jensen, L.H. (1968) X-Ray Structure Determination a Practical Guide, Macmillan, New York.

Taylor, R.T., Norrell, S.A., and Hanna, M.L. (1972) Arch. Biochem. Biophys, 148, 366.

Thomas, L.(1968) Biochem. Biophys. Acta. 291, 454. England.

Wasserman, R.H. (1972) in Metabolic Transport, L.E. Hokin, Ed., Vol. VI, Academic Press, New York.

Wasserman, R.H., Corradino, R.A., Fullmer, C.S., and Taylor, A.N. (1974) Vitamins and Hormones, 32, 299.

Weiner, J.H., and Heppel, L.A. (1971) J. Biol. Chem. 247, 7684.

Wiley, W.R. (1970) J. Bacteriol. 103, 656.

Willis, R.C. and Furlong, C.E. (1975) J. Biol. Chem. 250, 2574.

Wu, T.T., and Kabat, E.A. (1973) J. Mol. Biol. 75, 13.

Wyckoff, H. W., Poscher, M. Tsernoglou, D., Allewell, N.M., Kelly, D.M., and Richards, F.M. (1967) J. Mol. Biol. 27, 563.