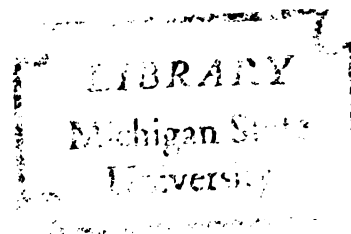


THE STRUCTURE DETERMINATION
OF 2-KETO-3-DEOXY-6-PHOSPHOGLUCONIC
ALDOLASE FROM PSEUDOMONAS PUTIDA
AT 3.56Å RESOLUTION

A Dissertation
for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
Irene Moustakali Mavridis
1975



This is to certify that the
thesis entitled
THE STRUCTURE DETERMINATION
OF 2-KETO-3-DEOXY-6-PHOSPHOGLUCONIC ALDOLASE
FROM PSEUDOMONAS PUTIDA AT 3.56 Å RESOLUTION
presented by

IRENE MOUSTAKALI MAVRAIDIS

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in PHYSICAL CHEMISTRY

A. Tulinsky

Major professor

Date July 23, 1975

0-43923

ABSTRACT

THE STRUCTURE DETERMINATION OF 2-KETO-3-DEOXY-6-PHOSPHOGLUCONIC ALDOLASE FROM PSEUDOMONAS PUTIDA AT 3.56Å RESOLUTION

by

Irene Moustakali Mavridis

The determination of the three dimensional structure and the tertiary folding of the enzyme 2-keto-3-deoxy-6-phosphogluconic aldolase (KDPG aldolase) from Pseudomonas Putida using standard protein X-ray crystallographic methods is reported.

KDPG aldolase, a trimeric enzyme, crystallizes in space group $P2_13$ with twelve protein monomers in the unit cell or one monomer in the asymmetric unit and with a unit cell dimension, $|\vec{a}| = 103.4\text{\AA}$. Three dimensional X-ray intensity data were collected from crystals of the native protein and two mercury (II)-containing derivatives at a resolution of 3.56Å and a gold-containing derivative at a resolution of 5.1Å. The positions of the heavy atom substitutions were deduced using difference Patterson, direct methods and difference Fourier techniques. The phases of the protein reflections were determined by the multiple isomorphous replacement method including anomalous scattering data of the two mercury-containing derivatives and were refined to a final mean figure of merit of 0.720.

The electron density map clearly shows trimeric arrangements of the subunits around the three-fold rotation axes of the unit cell. However, every subunit can belong to two different trimers and an ambiguity is introduced in the choice of a trimeric molecule for KDPG aldolase.

Irene Moustakali Mavridis

It is possible to follow the polypeptide chain of the protein which is composed of many helical regions distributed on the outside of the molecule and two β -sheet structures one parallel and one antiparallel in the inside. An empty channel ($9 \times 9 \times 30 \text{\AA}$) skews through each subunit passing from its center at about 45° to the three fold axis of the trimer.

THE STRUCTURE DETERMINATION
OF 2-KETO-3-DEOXY-6-PHOSPHOGLUCONIC ALDOLASE
FROM PSEUDOMONAS PUTIDA AT 3.56^oÅ RESOLUTION

by

Irene Moustakali Mavridis

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

1975

**To my parents
and Aristides**

ACKNOWLEDGMENTS

To Professor Alexander Tulinsky for his advise and support throughout this study, I would like to express my sincere thanks.

It is a pleasure to acknowledge a debt of graditude to Professor W.A. Wood for his help in the biochemical aspects of this work and to Ms. Diana Ersfeld and Mr. Paul Kuipers for their cooperation in providing the protein crystals.

I am greatly thankful to Dr. Richard Vandlen and Mr. Michael Liebman for their assistance in various aspects of my study and their continuous cooperation.

I am indebted to Dr. J.C. Ford for his help in providing the phase program and calculating the special Fourier syntheses and to Dr. Bobby Barnett for his assistance and many fruitful discussions.

Support from the Molecular Biology Section of the National Science Foundation during the course of this work is also acknowledged.

Finally I would like to thank all the members of the laboratory of Dr. A. Tulinsky for their general assistance.

TABLE OF CONTENTS

I.	INTRODUCTION	1
1.	General.	1
2.	2-Keto-3-Deoxy-6-Phosphogluconic Aldolase. Metabolic Role and Mechanism.	2
a.	The Role of Schiff-Base Formation.	5
b.	Aldolytic Cleavage	6
3.	The Isomorphous Replacement Method in Protein Crystal Structure Determination.	7
4.	Location of the Heavy Atoms.	12
5.	Protein Phase Determination.	18
6.	Refinement of the Heavy Atom Parameters and of the Protein Phase Angles	23
II.	EXPERIMENTAL	27
1.	Preparation of the Heavy Atom Derivatives.	27
2.	Data Collection.	33
3.	Data Reduction and Scaling	40
III.	SOLUTION OF THE PHASE PROBLEM AND REFINEMENT OF THE PROTEIN PHASES	47
1.	Initial Determination of the Main Heavy Atom Substitution.	47
2.	Determination of the Atomic Parameters by Difference Fourier Synthesis and Refinement of Protein Phases	59
3.	Determination of the Absolute Configuration.	68
4.	Summary and Error Analysis	70
IV.	ELECTRON DENSITY OF KDPG ALDOLASE.	78
1.	Calculation of the Electron Density Map.	78
2.	Molecular Packing.	82
3.	Molecular Structure.	89
4.	Concluding Remarks	98
	REFERENCES	101

LIST OF TABLES

Table	Page
1. Crystal Properties and Unit Cell Parameters of KDPG Aldolase.	28
2. Conditions of Preparation of Heavy Atom Derivatives.	32
3. Data Collection Parameters and Scaling Constants	42
4. Figures of Merit for the Eight Solutions Given by MULTAN for HgS and EHgTS	54
5. The Positions of the Three Substitutions of HgS and EHgTS Determined by the Difference Patterson Synthesis and by Direct Methods	58
6. Occupancy Changes of Substitution Sites after Three Cycles of Refinement	66
7. Change in the Residual Factors after Three Cycles of Refinement.	67
8. Heavy Atom Parameters.	69
9. Refinement Error Analysis.	72
10. Figure of Merit Analysis	76
11. Amino Acid Analysis of the Trimer of KDPG Aldolase (Reference 3).	91

TABLE OF FIGURES

Figure	Page
1. Mechanism of KDPG aldolase (reference 18)	4
2. Phase circle diagram of parent compound and two derivatives	10
3. (a) Vector diagrams of the structure factor of an isomorphous derivative for Friedel pair reflections showing anomalous dispersion; (b) superposition of mirror image of $(\bar{h}\bar{k}\bar{l})$ on (hkl) (reference 43)	16
4. Unit radius phase circle with line probability density, $P(\alpha_j)$, represented radially outward from phase circle, as base line (reference 45).	21
5. Morphology of an idealized crystal of KDPG aldolase	29
6. Distribution of diffraction intensities along the $\vec{a}^*$ axis of native and HgS and EHgTS (a), native and KAuCl_4 and Na_3IrCl_6 (b) containing crystals of KDPG aldolase.	34
7. Distribution of the square of the structure amplitudes of native and derivative crystals of KDPG aldolase versus 2θ angle.	45
8. Harker section of difference Patterson synthesis at an arbitrary scale based on HgS data; contour intervals at 100 beginning at 100. The numbers indicate self vector positions and the x's cross vector positions of the corresponding heavy atom sites with the principal sites 1, 2 and 3	48
9. Harker section of difference Patterson synthesis based on EHgTS data; scale and symbols as in Figure 8. Contour intervals at 100 beginning at 100	49
10. Harker section of difference Patterson synthesis based on Na_3IrCl_6 data; scale and symbols as in Figure 8. Contour intervals at 50 beginning at 100	50
11. Harker section of difference Patterson synthesis based on KAuCl_4 data; scale and symbols as in Figure 8. Contour intervals at 50 beginning at 100	51

Figure	Page
12. (a) Figure of merit with : • — no anomalous dispersion contribution (AD); o — with AD but incorrect hand; x — with AD correct hand. (b) Difference in number of reflections with $\langle m \rangle$ based on correct and incorrect assumption of the absolute configuration.	71
13. Distribution of r.m.s. heavy atom structure amplitude, r.m.s. differences and lack of closure as a function of $\sin\theta/\lambda$	75
14. Distribution of figure of merit, m , among reflections; fraction of total number of reflections with m greater than given m , in parentheses	77
15. "Best" electron density map of KDPG aldolase at 3.56Å resolution in projection down the a axis.	80
16. Schematic (hk0) projection of the arrangement of the subunits of KDPG aldolase in the unit cell of space group $P2_13$	83
17. Schematic (0kl) projection of the arrangement of the subunits of KDPG aldolase in the unit cell of space group $P2_13$	84
18. "Best" electron density map in projection down the three fold axis; (a) the trimer of the first kind, (b) the trimer of the second kind	86
19. The sequence of the known polypeptide containing the active lysine of KDPG aldolase.	90
20. "Best" electron density along the three fold axis. The trimer of the second kind is indicated; the individual monomers are outlined and denoted as A_1 , A_2 and A_3	94
21. Schematic representation of folding of the KDPG aldolase monomer indicated as A_1 in Figure 20. View approximately perpendicular to long direction of the central channel.	96

I. INTRODUCTION

1. General

The method of x-ray diffraction has been proven to be the most potent method to determine a detailed structure of a complicated molecule such as a protein. The first protein structures to be completed, myoglobin and hemoglobin¹ provided a tremendous amount of information concerning the tertiary structure of proteins. The subsequent determinations of the structure of numerous enzymes and other proteins confirmed and extended this information a great deal, so that today the determination of the structure of a new protein molecule is expected to reveal the same general characteristics observed previously, such as hydrophobic interior-hydrophilic exterior, hydrogen bonding into α -helices and β -sheets, ion pairs, etc.. When the crystal structures of oligomeric proteins showing allosteric behaviour are determined in the future, exceptions to the foregoing might be observed in unusual relations among subunits. However, the determination of the molecular structure of a protein does not terminate in the revelation of the arrangement of its amino acids in three dimensions; it extends to the specific problems of the biochemistry of these molecules and in this respect the determination of the structure of every protein provides unique information. Thus, in the case of the tetrameric hemoglobin, structural relations between the subunits in the oxy and deoxy forms have provided a very plausible explanation of the function of this important protein. In the case of enzymes, complexes with different substrates have been investigated

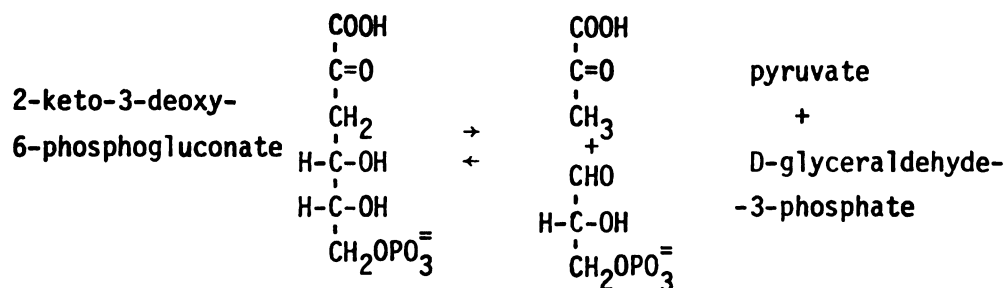
crystallographically and this has shed some insight on the very complex processes via which enzymes function. In the case of functionally related proteins, comparison of their structures has revealed common architectural features necessary for the function.

The purpose of the study to be described here was to establish the three dimensional structure of an aldolase enzyme. Aldolases are enzymes which cleave carbon-carbon bonds and the structure of an enzyme of such function has not heretofore been determined. The ultimate purpose of the determination of the structure is to study its relationship with substrates in the hope that this will elucidate more conclusively the details of its function and specificity. Moreover, physical², chemical³ and crystallographic⁴ studies have shown that the enzyme is composed of three identical subunits. This may be the first well documented case of a three subunit enzyme. Trimeric proteins are indeed rare in nature. This has led to the generally accepted view that there are fundamental constraints against the evolutionary survival of odd-numbered oligomers compared to even-numbered ones. Therefore, this unusual property of the enzyme is an additional pertinent reason for the structure determination and should lead to the exact nature of the interactions of the subunits.

2. 2-Keto-3-Deoxy-6-Phosphogluconic Aldolase. Metabolic Role and Mechanism

2-Keto-3-deoxy-6-phosphogluconic aldolase (hereafter denoted as KDPG-aldolase) from pseudomonas putida is one of the most thoroughly investigated aldolases, both from a structural²⁻⁷ and a mechanistic⁶⁻¹⁶ point of view, so that it has become a model of aldolytic catalysis second to that of fructose-1, 6 diphosphate aldolase. KDPG-aldolase catalyses

the following reaction:



This reaction is one step of a pathway which has been shown to play a major role in glucose, fructose, mannose, gluconate, glucosamine and 2-ketogluconate utilization in a variety of microorganisms. KDPG-aldolase is indeed widely distributed among eubacteria, but not found in higher forms of life. The aldolases which catalyse the cleavage of the above carbohydrates to three carbon atom units do not show consistency as to the mechanism involved. Thus, KDPG-aldolase, 2-keto-3-deoxy-6-phosphogalactonic aldolase and 4-hydroxy-2-ketoglutaric aldolase function via a Schiff-base mechanism, like the class I fructose-1,6-diphosphate aldolase found in higher animals, green plants and protozoa, whereas the rest of the bacterial aldolases require divalent metal ions such as Zn(II), Co(II) or Fe(II) as well as K^+ ions for full activity¹⁷ (class II aldolases).

KDPG aldolase is known to catalyse four reactions¹⁸: 1) cleavage of KDPG, 2) Schiff-base formation between a lysine ϵ -amino group and carbonyl compounds, 3) exchange between solvent protons and methyl hydrogens of pyruvate and 4) decarboxylation of oxalacetate. These functions are shown schematically in Figure 1.

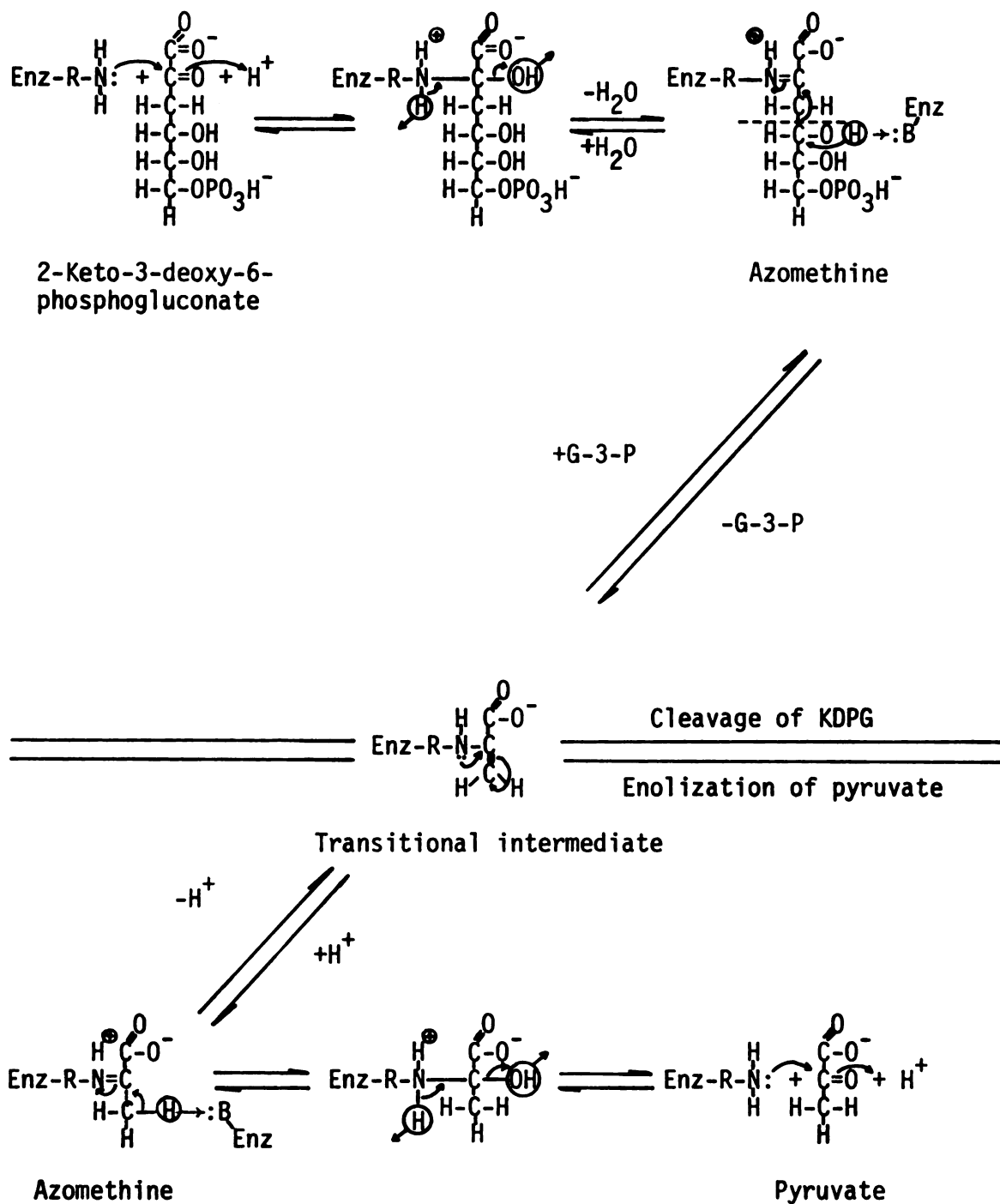


Figure 1. Mechanism of KDPG aldolase (reference 18).

S
P
n
s
tr
ar
a
ti
wi

al
cle
Do

a. The Role of Schiff-Base Formation

KDPG aldolase is inactivated by treatment with pyruvate in the presence of NaBH_4 . Subsequent treatment of the derivatized protein with LiAlH_4 and acid hydrolysis yields ϵ -N(2-n-hydroxypropyl) lysine¹¹, indicating that the pyruvate had reacted with a lysine residue of the enzyme. The experiment demonstrates that the enzyme functions via Schiff-base assisted catalysis as fructose-1,6-diphosphate aldolases.

A large number of carbonyl containing substrates inactivate KDPG aldolase¹² by this reaction. Of a series of seven analogs of pyruvate, only hydroxypyruvate and dihydroxy acetone do not inactivate the enzyme. Compounds with a longer chain like α -ketobutyrate, α -ketoisovalerate and α -ketoglutarate caused inactivation. A series of analogs of KDPG including 5-keto-4-deoxyglucarate, 2-keto-4-hydroxyglucarate, 2-keto-3-deoxygluconate, 2-keto-3-deoxygalactonate and 2-keto-3-deoxy-6-phosphogalactonate are also successful in inactivating the enzyme, but none of these was cleaved to 3-carbon atom compounds as is KDPG. The specificity of the enzyme not to form a Schiff-base with dihydroxy acetone or hydroxypyruvate, while it forms the base with hydroxyacetone and α -ketobutyrate, shows that a non-steric restraint must exist against a hydroxyl group on the deoxy position of KDPG. Apart from this restriction, KDPG aldolase is completely non-specific in forming Schiff-bases with carbonyl compounds.

The exchange and oxalacetate decarboxylation specificity of KDPG-aldolase involve the same lysine which is used in the carbon-carbon cleavage¹² and in fact, the reactions proceed via the Schiff-base compound as can be seen in Figure 1.

b. Aldolytic Cleavage

KDPG-aldolase is highly specific in the direction of cleavage. Thus, 2-keto-3-deoxy-6-phosphogalactonate⁶, 2-keto-6-phosphogluconate¹⁹, 2-keto-3-deoxygluconate, 2-ketogluconate, 5-keto-4-deoxygluconate, 2-keto-4-hydroxyglutarate, α -ketoglutarate¹², deoxyribose-5-phosphate and fructose-1,6-diphosphate²⁰ are not cleaved. The above list of compounds which are not cleaved demonstrates that:

- (i) a 3-deoxy group
- (ii) a 4-hydroxyl group in the erythro configuration
- (iii) a 6-phosphate group

have to be present simultaneously in a 6-carbon carbohydrate in order for it to be a proper substrate for aldolytic cleavage, whereas, as mentioned previously, a large number of carbonyl analogs form a Schiff-base covalent catalytic intermediate with the active lysine residue.

It is clear, from the foregoing, that the Schiff-base formation itself is not a sufficient condition for cleavage, but it has not yet been determined what additional factors are involved. There are indications²¹ that additional base assistance in Schiff-base catalysed aldolization is likely. In this context, the hydroxyl group on the fourth carbon (C_4) in the direction of cleavage or on the carbonyl group of glyceraldehyde-3-phosphate in the direction of condensation has been implicated. Thus the existence of the erythro-hydroxyl group at C_4 and the phosphate at C_6 are required for steric reasons. There are lysine residues in the active site of KDPG-aldolase, other than the one forming the Schiff-base, which have the potential for the additional nucleophile function. However, their function seems to be indirect, presumably by simply maintaining the conformation⁶. Experiments by Meloche with

bromopyruvate^{13,14,22} have resulted in the isolation of a three carbon atom adduct with both a carboxyl and an sulfhydryl group. This led him to postulate that a carboxylate or a cysteine residue plays the role of the additional base. Consequently he suggests the existence of more than one active site conformation.

3. The Isomorphous Replacement Method in Protein Crystal Structure

Determination

The electronic distribution of a system in the crystalline state is a periodic function of the coordinates (x,y,z) measured parallel to the generally oblique crystal axes $\vec{a}$, $\vec{b}$ and $\vec{c}$. It is therefore possible to express the electron density, ρ , (electrons per unit volume) at any point (x,y,z) in the crystal by a triple Fourier series²³⁻²⁵:

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

where $F(h,k,l)$ is the structure factor of the reflection (h,k,l) whose amplitude is proportional to the square root of the diffracted X-ray radiation of that reflection and V is the volume of the unit cell.

$F(h,k,l)$ is a complex quantity representing the scattering of X-rays by the electrons of one unit cell in a certain direction (h,k,l). It is defined as:

$$F(h,k,l) = \sum_{j=1}^N f_j(h,k,l) \exp\{2\pi i(hx_j+ky_j+lz_j)\}, \quad (2)$$

that is, it can be regarded as a sum over all atoms N in the unit cell of atomic scattering factor $f_j(h,k,l)$ (scattering power of the atom) multiplied by a phase factor which depends on the position of the atoms, j , (x_j, y_j, z_j) , and on the direction of the diffracted beam, (h,k,l).

Unfortunately, X-ray diffraction measurements do not provide the phase of the structure factor and equation (1) cannot be applied directly to determine the electron density in the unit cell, and hence the atomic arrangement; more indirect methods are therefore used to find the phases of the structure factors. The isomorphous replacement method is one such method which proved to be very successful in determining the structure factor phases of protein crystals.

The isomorphous replacement method is not a new technique of crystal structure analysis. It was used classically as long ago as 1937 by J.M. Robertson on the centric structure determination of phthalocyanine²⁶. In 1951, it was suggested by Bokhoven, Schoone and Bijvoet²⁷ that the method could be used to determine phases for non-centrosymmetric structures provided that three or more isomorphous crystals could be obtained (multiple isomorphous replacement) with different heavy atom positions in each and with the rest of the structure remaining completely unchanged by the heavy atom substitution. Finally, in 1956, Harker²⁸ outlined in detail a graphical and analytical method of the determination of the phase angles in the non-centric case using multiple isomorphous replacement. Neither, however, put the multiple isomorphous replacement method to a practical test. It was Perutz²⁹ who showed, in 1953, that two heavy atoms, such as silver or mercury, when attached to a protein of molecular weight as large as 68,000, caused measurable intensity changes in the diffraction pattern. Consequently the method was used to determine the centrosymmetric (h0l) phases of hemoglobin³⁰ and myoglobin³¹, the non-centrosymmetric (0kl) projection of hemoglobin³² and finally in the three dimensional non-centric structure determination of myoglobin³³.

Although there are detailed review articles^{34,35} discussing the method of multiple isomorphous replacement as applied to the solution of the phase problem in protein crystallography, a brief review here will best serve the purpose of clarifying the course taken in the determination of the structure of KDPG-aldolase as well as introducing the appropriate nomenclature.

We denote by $\vec{F}_p$ the structure factor of the protein crystal for a certain point (h,k,l) of reciprocal space. The structure factor for a derivative crystal composed of the structurally unchanged protein plus a small number of additional atoms of large scattering power (heavy atoms), will be denoted as $\vec{F}_{pH1}$. The corresponding amplitudes $|F_p|$ and $|F_{pH1}|$ of both vectors can be measured experimentally. If we represent the contribution of the heavy atom alone as $\vec{F}_{H1}$, the above vectors should satisfy the following equation:

$$\vec{F}_{pH1} = \vec{F}_p + \vec{F}_{H1} \quad (3)$$

It is possible to find the vector $\vec{F}_{H1}$ from the amplitudes $|F_p|$ and $|F_{pH1}|$, that is, we can determine the positions of the heavy atoms H1. The solution of equation (3), in terms of the amplitudes of the vectors $\vec{F}_p$ and $\vec{F}_{pH1}$ and the vector $\vec{F}_{H1}$, is illustrated by Harker's phase circle diagram²⁸ in Figure 2. For a certain reflection (hkl), the circle of the parent compound is drawn with its center at the origin of the complex plane and with a radius proportional to the amplitude of the protein structure factor; its phase can have any value on the circle from 0 to 2π . A second circle is drawn with center at the end of vector $-\vec{F}_{H1}$ of radius proportional to $|F_{pH1}|$; the two circles intersect at two

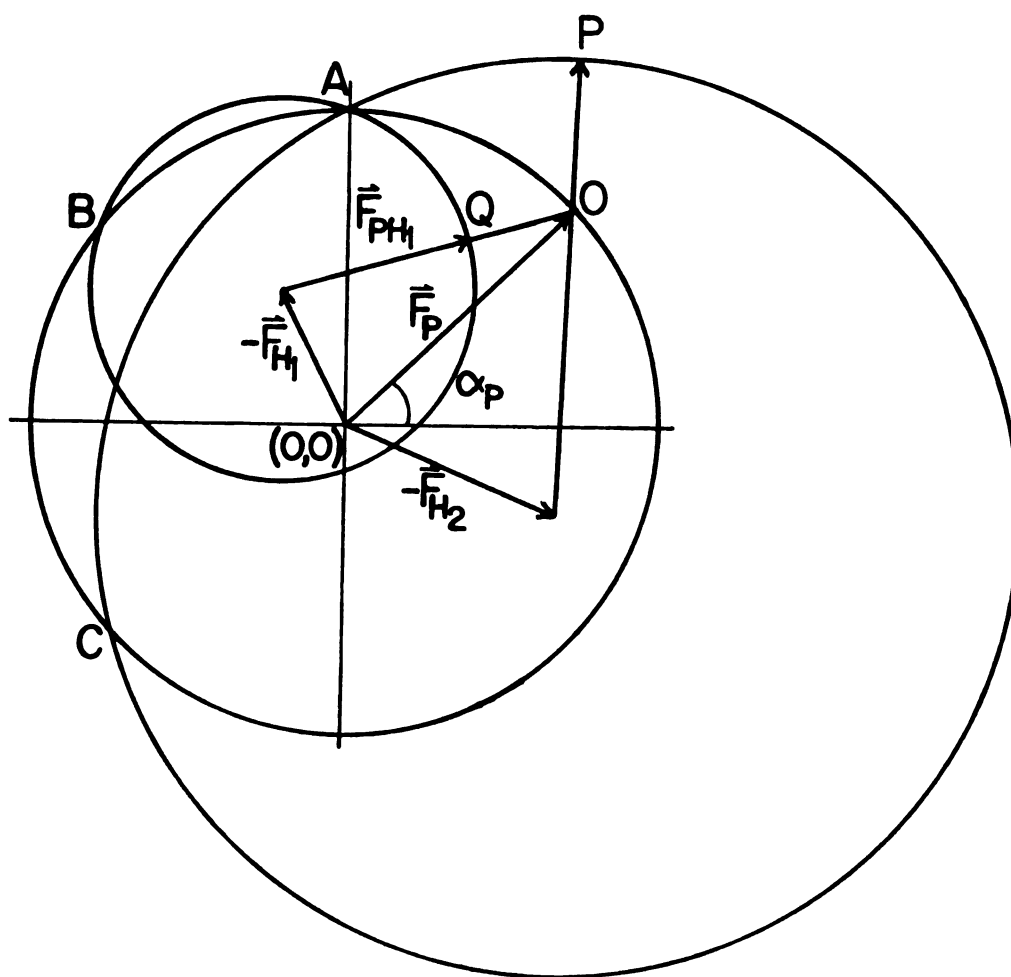


Figure 2. Phase circle diagram of parent compound and two derivatives.

points A and B. It is clear that a phase corresponding to either of these two points satisfies equation (3). Therefore the phases corresponding to these two points represent two possible phases that the protein structure factor $\vec{F}(h,k,l)$ can have, only one of which is correct. The ambiguity is resolved by introducing another isomorphous derivative with a heavy atom vector, $\vec{F}_{H2}$, which is not collinear with $\vec{F}_{H1}$ (different heavy atom position). The phase circle of $\vec{F}_{PH2}$ will also intersect the parent circle again at two points A and C, one of which coincides with one of the first pair (A).

The foregoing case is ideal and is never satisfied in practice. The two derivative circles seldom cross the parent compound circle at exactly the same point due to experimental and other errors such as imperfect isomorphism, inaccurate determination of the heavy atom vectors $\vec{F}_{H1}$ and $\vec{F}_{H2}$, etc.. Thus, in general, more than two isomorphous derivative crystals are needed to have an accurate determination of protein phases.

The crystal structure determination of a protein can be divided into the following steps:

- a. Preparation of the heavy atom isomorphous derivative crystals and collection of complete sets of diffraction intensity data of the protein and the derivative crystals.
- b. Location of the heavy atoms within the unit cell of the crystal and characterization of their parameters (positional coordinates, scaling factors, extent of substitution and temperature factors).
- c. Calculation of the protein phases.
- d. Calculation of the electron density Fourier map using the determined phases and the experimentally measured protein structure amplitudes.

o
n
(
a

T
to
la

in
si
enc

In practice, the sequence of the above steps is not always strictly followed. For instance, one can determine the positions of some of the heavy atoms in one or more isomorphous derivatives and calculate approximate protein phases which then can be used to determine the positions of additional heavy atoms for the same derivative or to locate the heavy atoms in other isomorphous derivatives, and so on, until the protein phases are determined as accurately as possible.

4. Location of the Heavy Atoms

The main problem in locating the heavy atoms in the unit cell of an isomorphous crystal is that the true magnitude of the vector $\vec{F}_H$ is unknown. For some classes of reflections which are centrosymmetric (centrosymmetric projections) the observed heavy atom amplitude, $|F_H|$, according to equation (3), is:

$$|F_H| = |F_{PH}| + |F_P| \quad (4a)$$

$$|F_H| = |F_{PH}| - |F_P| \quad (4b)$$

The first case (4a) is usually disregarded because it is rare and applies to small structure factors $\vec{F}_P$ and difference Patterson synthesis are calculated with coefficients:

$$|\Delta F|^2 = (|F_{PH}| - |F_P|)^2 \quad (5)$$

in order to find the location of heavy atoms in centrosymmetric projections. The same coefficients have been used for non-centrosymmetric projections³² and today they are used routinely for calculating the

o
o
v
T
t
by
ap

la
Th
of
co
of
the
det
con
Bra
enve
func
calc

Patterson synthesis of the heavy atoms in three dimensions. Although $|\Delta F|$ is not the true amplitude of the structure factors of the heavy atoms $|F_H|$, when $|F_H|$ is small or zero $|\Delta F|$ is small or zero; however, when $|F_H|$ is large, $|\Delta F|$ need not be large, unless the phases of $\vec{F}_{PH}$ and $\vec{F}_P$ are nearly the same. Thus, the terms that contribute most to the Patterson synthesis are those for which the approximation, that $\vec{F}_{PH}$, $\vec{F}_P$ and consequently $\vec{F}_H$ are collinear, is most nearly true.³⁶ The coefficients $|\Delta F|^2$ used for the difference Patterson synthesis can be considered as a "dampened" $|F_H|$ distribution which will show the true vector distribution, possibly obscured by a higher than normal background. This has been shown by the theoretical expansion of the coefficients of the difference Patterson synthesis^{34,37} and has been verified abundantly by experience in protein structures that have been solved using the approximation.

Protein crystals belong to non-centrosymmetric space groups. The latter do not have a well-defined origin like a center of symmetry. Therefore, it is possible to determine the coordinates of the heavy atoms of two or more derivatives with respect to a different origin. These coordinates can be correlated to the same origin by leaving the positions of the heavy atoms of one isomorphous derivative fixed and translating the coordinates of the heavy atoms of the others. The difficulty is to determine the amount and direction of the translation. Harker²⁸ first considered trial and error techniques by using selected reflections. Bragg³⁸ later suggested a method of fitting sinusoidal curves to the envelope of the differences $|\Delta F|$. Perutz³⁹ proposed two correlation functions, later modified and improved upon by Blow³², which depend on calculating certain Fourier syntheses from the structure amplitudes of

two derivatives; the positions of the heavy atoms of the derivatives can then be deduced simultaneously. These syntheses, however, have a high background which limits their usefulness severely. Rossmann³⁶ showed that a Patterson type function with coefficients $(|F_{PH1}| - |F_{PH2}|)^2$ is approximately equivalent to the self Patterson of the heavy atoms of derivative 1, plus the self Patterson of the heavy atoms in derivative 2, minus the cross Patterson between the heavy atoms in derivative 1 and 2. The synthesis has positive peaks at the end of vectors between the heavy atom of the same derivative and negative peaks at the end of vectors between heavy atoms of the different derivatives. Even this synthesis, however, can be difficult to interpret if there are more than one heavy atoms per asymmetric unit. The easiest to interpret correlation function was proposed by Kartha and Parthasarathy⁴⁰ and also suggested by Steinrauf⁴¹. Its coefficients are:

$$(\Delta F_{iso1}) (\Delta F_{iso2}) = (|F_{PH1}| - |F_P|) (|F_{PH2}| - |F_P|) \quad (6)$$

This map can be shown to have positive peaks of magnitude $f_{H1} f_{H2}$ at positions $(r_{H1} - r_{H2})$ and also positive peaks of the same height at the centrosymmetric positions $-(r_{H1} - r_{H2})$, where f_{H1} and f_{H2} are the scattering factors of the heavy atoms of derivatives 1 and 2, respectively. Complications arising from the self Patterson vectors of the heavy atoms of the individual derivative are therefore avoided and the synthesis is particularly suitable for correlation involving multiple heavy atoms substitutions. Moreover, the background can be reduced if anomalous dispersion data are combined with the isomorphous replacement data⁴⁰.

o
P
T
t
is
ve
Fr
mi
ti
gin
in
the
in
ima
fac
Y wi

and

The positions of the heavy atoms of the derivatives are determined from the difference Patterson synthesis which is centrosymmetric, so that two centrosymmetrically related sets of coordinates are consistent with the difference Patterson, only one of which is consistent with the X-ray diffraction intensity data. This ambiguity can be resolved by using the property of heavy atoms to give rise to appreciable anomalous scattering. The case of a single isomorphous substitution, where the anomalous scattering may be used to indicate which of the two enantiomorphic solutions is the correct one, was described by Bijvoet⁴². Figure 3(a) shows the vector diagrams of the structure factors of an isomorphous derivative for Friedel pair reflections showing anomalous dispersion. The diagrams are mirror images across the real axis with respect to the protein contribution $\vec{F}_p$ and the real part of the heavy atom contribution $\vec{F}_H'$. The imaginary part of the heavy atom structure factor, $\vec{F}_H''$, has the same direction in both reflections, so that the resulting vectors $\vec{F}_{pH}(+)$ and $\vec{F}_{pH}(-)$ of the derivative, in the directions (h,k,l) and $(\bar{h},\bar{k},\bar{l})$ respectively, differ in magnitude. Figure 3(b) can be constructed by superimposing the mirror image of $(\bar{h},\bar{k},\bar{l})$ on that of (h,k,l) . If $\vec{F}_{pH}$, the derivative structure factor vector in the absence of anomalous scattering, makes an angle γ with $\vec{F}_H'$ then:

$$|\vec{F}_{pH}(+)|^2 = |\vec{F}_{pH}|^2 + |\vec{F}_H''|^2 - 2|\vec{F}_{pH}||\vec{F}_H''|\sin\gamma$$

and

$$|\vec{F}_{pH}(-)|^2 = |\vec{F}_{pH}|^2 + |\vec{F}_H''|^2 + 2|\vec{F}_{pH}||\vec{F}_H''|\sin\gamma$$

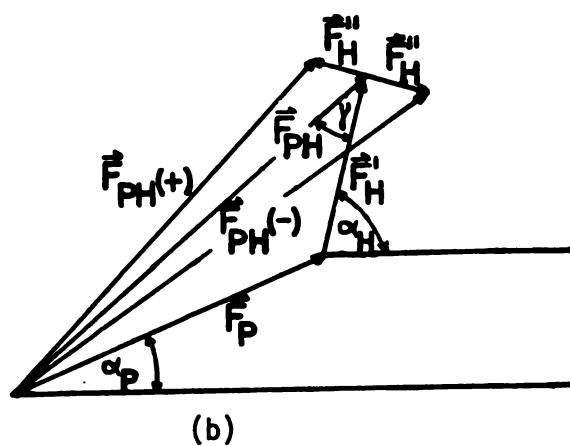
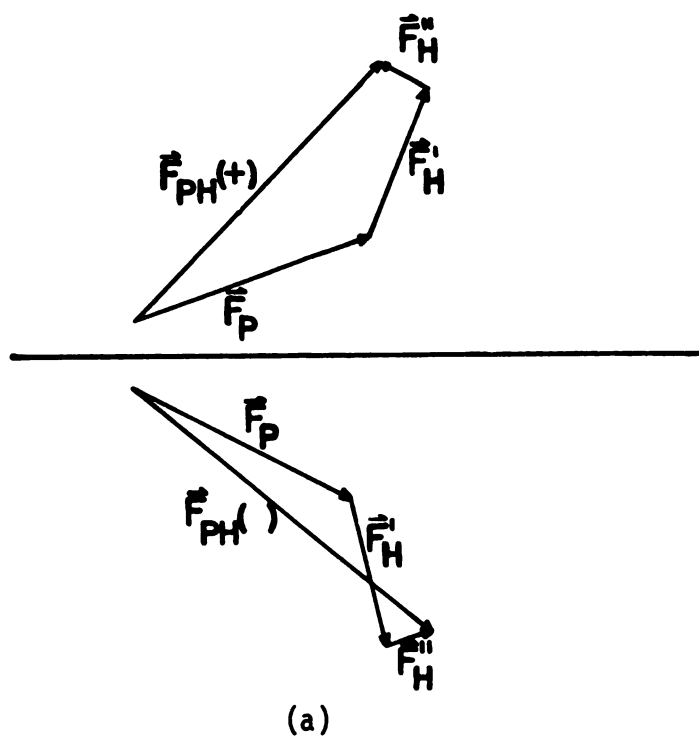


Figure 3. (a) Vector diagrams of the structure factor of an isomorphous derivative for Friedel pair reflections showing anomalous dispersion.
 (b) Superposition of mirror image of (hkl) on (hkl) (reference 43).

s

Ge

the

From

in t

ambig

phous

heavy

ulate

factor

for a g

normalo

ion eq

e

so that

$$|F_{PH}(+)|^2 - |F_{PH}(-)|^2 = -4|F_{PH}||F_H''|\sin\gamma \quad (7)$$

Generally $\vec{F}_H''$ is small, thus:

$$|F_{PH}(+)| + |F_{PH}(-)| \approx 2|F_{PH}| ;$$

therefore,

$$(\Delta\pm)_c = |F_{PH}(+)| - |F_{PH}(-)| = -2|F_H''|\sin\gamma \quad (8)$$

From (8), it can be seen that the sign of $(\Delta\pm)_c$ indicates whether γ lies in the range 0 to π or π to 2π and hence it is possible to resolve the ambiguity in phase determination which occurs with only a single isomorphous derivative⁴³. Similarly, the two sets of the coordinates of the heavy atoms of an isomorphous derivative can be used separately to calculate protein phases α_p and the phases of the heavy atom structure factor α_H . Then, because (Figure 3(b))

$$\sin\gamma = |F_p| \sin(\alpha_H - \alpha_p) / |F_{PH}| \quad (9)$$

for a given protein phase, the difference, $\epsilon'(\alpha_p)$, between the observed anomalous scattering component $(\Delta\pm)_o$ and the one calculated, $(\Delta\pm)_c$, from equation (8) is:

$$\epsilon'(\alpha_p) = (\Delta\pm)_o - 2|F_H''|\sin\gamma,$$

or

The
men

with

will
centr
factor

5. Pr

isomorp
brick 44
that er
isomorp
can be c
failure

or from equation (9),

$$\varepsilon'(\alpha_p) = (\Delta\pm)_0 - 2|F_p||F_H''|\sin(\alpha_H - \alpha_p)/|F_{PH}|. \quad (10)$$

The correct set of coordinates will be those which give the best agreement, that is, the smallest $\varepsilon'(\alpha_p)$.

Alternatively, it has been proposed⁴⁰ that a Fourier synthesis with coefficients:

$$\begin{aligned} (\Delta F_{iso1} + i\Delta F_{ano1})(\Delta F_{iso2} + i\Delta F_{ano2}) = \\ = [|F_{PH1}| - |F_p| + i(|F_{PH1}(+)| - |F_{PH1}(-)|)/(f_{H1}'/2f_{H1}'')] \times \\ \times [|F_{PH2}| - |F_p| + i(|F_{PH2}(+)| - |F_{PH2}(-)|)/(f_{H2}'/2f_{H2}'')] \end{aligned} \quad (11)$$

will give peaks at the end of the vectors $(r_{H2} - r_{H1})$, but not at their centrosymmetric. In equation (11), f_H' is the real part of the scattering factor of the heavy atom and f_H'' is the imaginary part according to:

$$f_H = f_H^0 + \Delta f_H' + i\Delta f_H'' = f_H' + if_H''$$

5. Protein Phase Determination

The principles involved in calculating protein phases from multiple isomorphous replacement data have been described previously. Blow and Crick⁴⁴ have treated the errors involved in the method and have shown that errors from all sources (experimental measures, lack of isomorphism, incomplete or imperfect refinement of the heavy atom parameters, etc.) can be considered as residing on the magnitude of $\vec{F}_{PH}$, thus representing a failure of the phase triangle to close exactly on the $\vec{F}_{PH}$ side for a

9

$\vec{F}_j$

for

Th

te

In

l at

the

a ph

phas

were

vativ

which

given protein angle α_p (Figure 2). Since the magnitude and phase of $\vec{F}_H$ can be calculated, the calculated structure factor for the derivative for a given arbitrary protein phase α_p can be found:

$$D^2(\alpha_p) = |F_p|^2 + |F_H|^2 + 2|F_p||F_H|\cos(\alpha_p - \alpha_H). \quad (12)$$

The lack of closure error, $\epsilon_H(\alpha_p)$, can then be defined for a given protein phase angle α_p as:

$$\epsilon_H(\alpha_p) = ||F_{pH}| - |D(\alpha_p)|| \quad (13)$$

In Figure 2, these quantities are the lengths OQ and OP for derivatives 1 and 2 respectively. The lack of closure can be used properly to find the best point of intersection of the phase circles.

Assuming a Gaussian distribution of errors⁴⁴, the probability that a phase angle α_p is correct is related to the lack of closure of the phase triangle for the above angle and for the derivative i by:

$$P_i(\alpha_p) = \exp(-\epsilon^2(\alpha_p)/2E_i^2), \quad (14)$$

where E_i is the root mean square lack of closure error for the ith derivative; E_i can be determined from the centrosymmetric reflections for which $\vec{F}_{pH}$, $\vec{F}_p$ and $\vec{F}_H$ are collinear from:

$$E_i^2 = \langle (|F_{pHi}| - |F_p| - |F_{Hi}|)^2 \rangle. \quad (15)$$

T

c

m

a

in

It

for

$\vec{r}_i^2$

elec

mini

squa

coef

tribu

phase

as sh

factor

radial

gravit

$\vec{F}_D$

action

if the

be near

if the

The implication here is, of course, that the errors are the same for centrosymmetric and non-centrosymmetric reflections. When several isomorphous derivatives are used simultaneously, the total probability for a given protein phase angle, α_j , is proportional to the product of the individual probabilities:

$$P(\alpha_j) = \prod_i P_i(\alpha_j) = \exp(-\sum_i \epsilon_i^2(\alpha_j)/2E_i^2). \quad (16)$$

It would be reasonable to take as the correct phase angle of the protein for a given reflection the most probable angle, α_M , the angle for which $\sum_i \epsilon_i^2(\alpha_M)/2E_i^2$ is a minimum. However, Blow and Crick⁴⁴ have shown that the electron density calculated with amplitude $|F_p|$ and phases α_M is not minimized with respect to error. The synthesis with the smallest mean square error in electron density over the entire unit cell has as Fourier coefficients the vector of the center of gravity of the probability distribution.

If the probability $P(\alpha_j)$ is plotted for the phases α_j around the phase circle, for most of the reflections the distribution is bimodal as shown in Figure 4⁴⁵ (for clarity the diagram has been scaled by a factor of $1/|F_p|$ and the probability density has been represented as a radial distance from the center of the phase circle). The center of gravity of the resulting probability density is at the end of vector $|F_p|\vec{m}$ with polar coordinates $|F_p||m|$ and α_B . If the probability distribution is sharp, the vector $\vec{m}$ will be near the unity phase circle, but if the probability distribution is uniform around the circle, $|m|$ will be nearly zero. The magnitude of the vector $\vec{m}$, therefore, is a measure of the reliability of the phase determination for a given reflection and

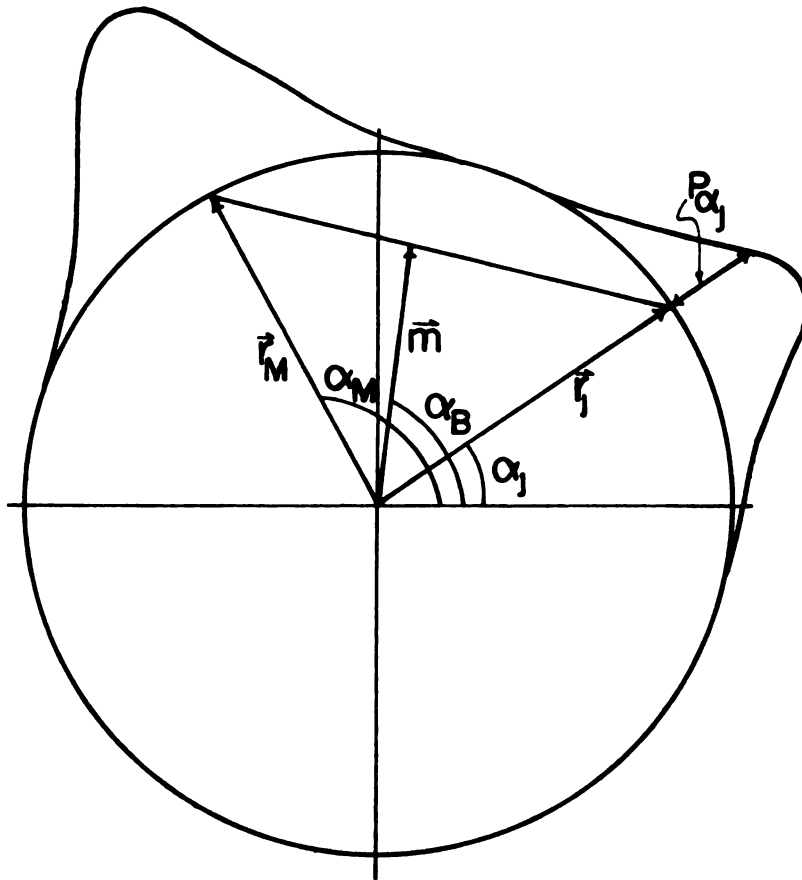


Figure 4. Unit radius phase circle with line probability density $P(\alpha_j)$ represented radially outward from phase circle as base line (reference 45).

i

th

a

ph

How

Since

by

A

to be t

mit ce

re pha

re mos

it is called the figure of merit for that reflection. It has been shown⁴⁵ that $|m|$ is the mean value of the cosine of the error in phase angle for a given reflection. If the probability, P_j , is calculated around the phase circle in constant intervals, α_j , then:

$$\vec{m} = \sum_j P_j(\alpha_j) r_j / \sum_j P_j(\alpha_j) = \sum_j P_j(\alpha_j) \exp(i\alpha_j) / \sum_j P_j(\alpha_j) . \quad (17)$$

However, r_j is unity and:

$$m \cos \alpha_B = \sum_j P_j(\alpha_j) \cos \alpha_j / \sum_j P_j(\alpha_j) \quad (18a)$$

$$m \sin \alpha_B = \sum_j P_j(\alpha_j) \sin \alpha_j / \sum_j P_j(\alpha_j) . \quad (18b)$$

Since the error in phase angle at a given α_j is defined as:

$$\Delta \alpha_j = \alpha_B - \alpha_j \quad (\text{see Figure 4}) ,$$

by changing the origin so that $\alpha_B = 0$ gives $\alpha_j = \Delta \alpha_j$ and

$$|m| = \sum_j P_j(\Delta \alpha_j) \cos \Delta \alpha_j / \sum_j P_j(\Delta \alpha_j) = \langle \cos \Delta \alpha_j \rangle \quad (19)$$

A Fourier synthesis with coefficients $|m| |F_p| \exp(2\pi i \alpha_B)$ can be shown^{44,45} to be the electron density with the least mean square error over the whole unit cell and it is called "best Fourier" or the "best electron density." The phase α_B of the vector $\vec{m}$ is called the "best phase" in contrast to the most probable phase, α_M .

de

6.

case

ampl

can

Ther

prio

cent

defin

where

scali

and $\vec{F}_1$

but it

stitut

that t

rimi

Blow and Crick⁴⁴ have shown that the mean square error in electron density of the "best Fourier" synthesis is:

$$\langle \Delta \rho^2 \rangle = \frac{2}{V^2} \sum_h^{\infty} \sum_k^{\infty} \sum_l^{\infty} |F_p(h,k,l)|^2 (1 - |m^2(h,k,l)|). \quad (20)$$

6. Refinement of the Heavy Atom Parameters and of the Protein Phase

Angles

The refinement of the heavy atom parameters in the non-centrosymmetric case presents many difficulties because there are not observed structure amplitudes to which the calculated structure factors of the heavy atoms can be compared, unless, of course the protein phase angles are known. There exist, however, two methods of refining the heavy atom parameters prior to the determination of protein phases. Hart's method⁴⁶ relies on centrosymmetric reflections, where equations (4a) and (4b) apply. He defines the best heavy atom parameters as those which minimize:

$$E_H = \sum (K_H |F_{PH}| \pm |F_P| - |F_H|)^2, \quad (21)$$

where the summation is over centrosymmetric reflections, K_H is a refining scaling factor and the change in sign allows for the possibility of $\vec{F}_{PH}$ and $\vec{F}_P$ differing in sign. The method has been found to work well^{45,47}, but it has the disadvantage of not using the bulk of intensity data constituting the non-centrosymmetric reflections. Rossmann³⁶ has suggested that the heavy atom parameters can be refined in three dimensions by minimizing the quantity:

$$E_R = \sum w [(K_R |F_{PH}| - |F_P|)^2 - |F_H|^2], \quad (22)$$

even though, in general, equation (4b) does not apply to non-centrosymmetric reflections. The assumptions made by using this method are the same as those made for the difference Patterson synthesis and the method gives satisfactory results, if the weight, w , is selected to attach more importance to reflections for which equation (4b) is nearly true.

Clearly, the most appropriate way is to refine by successive least squares iterations of the heavy atom parameters followed by protein phase calculation. Once all substitutions for all derivatives have been found, they can be used to calculate a set of protein phases which in turn can be used to refine the parameters of the heavy atoms, and so on until a desired degree of convergence has been reached.

Two quantities have been proposed for minimization by the method of least squares. Dickerson et al.^{35,45}, proposed the quantity:

$$E_{Hj} = \sum_n w_n (\vec{F}_{PHjn} - \vec{D}_{Hjn})^2, \quad (23)$$

where w_n is a weighting factor and D as defined in equation (12); the summation is over all reflections. All heavy atom parameters of a derivative j can then be refined by solving the set of normal equations:

$$\sum_i (\sum_n w_n (\partial D_{Hj} / \partial \psi_i) (\partial D_{Hj} / \partial \psi_q)) \Delta \psi_i = \sum_n w_n (\partial D_{Hj} / \partial \psi_q) (|F_{Hj}| - |D_{Hj}|), \quad (24)$$

where the subscript i and q denote the individual parameters ψ_i and ψ_q of the derivative j . Kraut et al.⁴⁸ proposed the minimization of:

$$E_k = \sum_n (K_k |F_{PH}| - ||F_P| \exp(i\alpha_M) + \vec{F}_H|)^2, \quad (25)$$

where α_M is the most probable protein phase. Both approaches seem to give reliable results.

An alternative method of refinement has been through the use of difference Fourier methods. These methods can also be used to find minor substitution sites, especially at the early stages of refinement. Two difference Fourier syntheses have been widely used. The first has been proposed by Steinrauf⁴¹; its coefficients are:

$$\Delta = (|F_{PH}| - |F_P|) \exp(i\alpha_P), \quad (26)$$

where α_P is an estimate of the protein phase angle (best phase angles or most probable phase angles). This synthesis shows the heavy atom substitutions which were included in the determination of the protein phase angles and any other substitution site that has not been included. The second difference Fourier synthesis suggested by Blake *et al.*⁴⁷ is calculated with coefficients:

$$\Delta\Delta = (|F_{PH}| - |\vec{F}_P + \vec{F}_H|) \exp(i\alpha_{PH}), \quad (27)$$

where α_{PH} is the current heavy atom derivative phase. The map reveals heavy atom sites not included in the determination of the protein phases. Fairly precise changes in the parameters of the heavy atoms used in the protein phase calculation can be estimated from this map (coordinates and occupancy). Experience has shown that such difference Fourier syntheses are generally correct and provide the best indications for minor

sites of derivatives which have not yet been included into calculations. However, refinement of the heavy atom parameters by such methods is very laborious and generally they are not used for such purposes.

II. EXPERIMENTAL

1. Preparation of the Heavy Atom Derivatives

Isolation of KDPG aldolase from the bacteria pseudomonas putida, purification, measurement of its activity and subsequent crystallization was performed in the laboratory of Dr. W.A. Wood, Department of Biochemistry, Michigan State University, according to already published methods^{2,4}. Crystals of the enzyme suitable for collection of X-ray intensity data were grown by the method of Zeppezauer⁴⁹ as described by Vandlen et al.⁴. Because the solution in which the crystals were stored initially was considered to be of very low ionic strength ($0.5M (NH_4)_2SO_4 - 0.1M KH_2PO_4$)⁴, the crystals that were used for the final sets of intensity data in this work were stored in $2.5M (NH_4)_2SO_4 - 0.1M KH_2PO_4$, at pH 3.5 and at 10-15°C. Crystal properties and unit cell parameters are summarized in Table 1.

The search for suitable isomorphous derivatives containing heavy atoms followed a general approach. The appropriate chemical reagent was dissolved in a solution of the same composition as the one in which the protein crystals were stored to form a concentrated solution. A small amount of this was added to a solution containing 5-10 protein crystals so that the mole ratio of heavy atom reagent to protein would be approximately 10/1; the crystals were then allowed to soak from several days to several weeks, depending on the reagent.

In order to calculate the number of moles of protein per crystal, the volume of the crystal was estimated visually. The crystals have a rhombohedral morphology as shown in Figure 5. The surface opposite to

TABLE 1. Crystal Properties and Unit Cell Parameters of KDPG-Aldolase

Crystal System	Cubic
Space Group	P2 ₁ 3
$ \vec{a} $	103.40(4) Å
Number of Molecules per Unit Cell	12 monomers
Number of Molecules per Asymmetric Unit	1 monomer
Mass (%) of Protein per Unit Cell	37%
Crystal Density (salt free)	1.126 g/cm ³

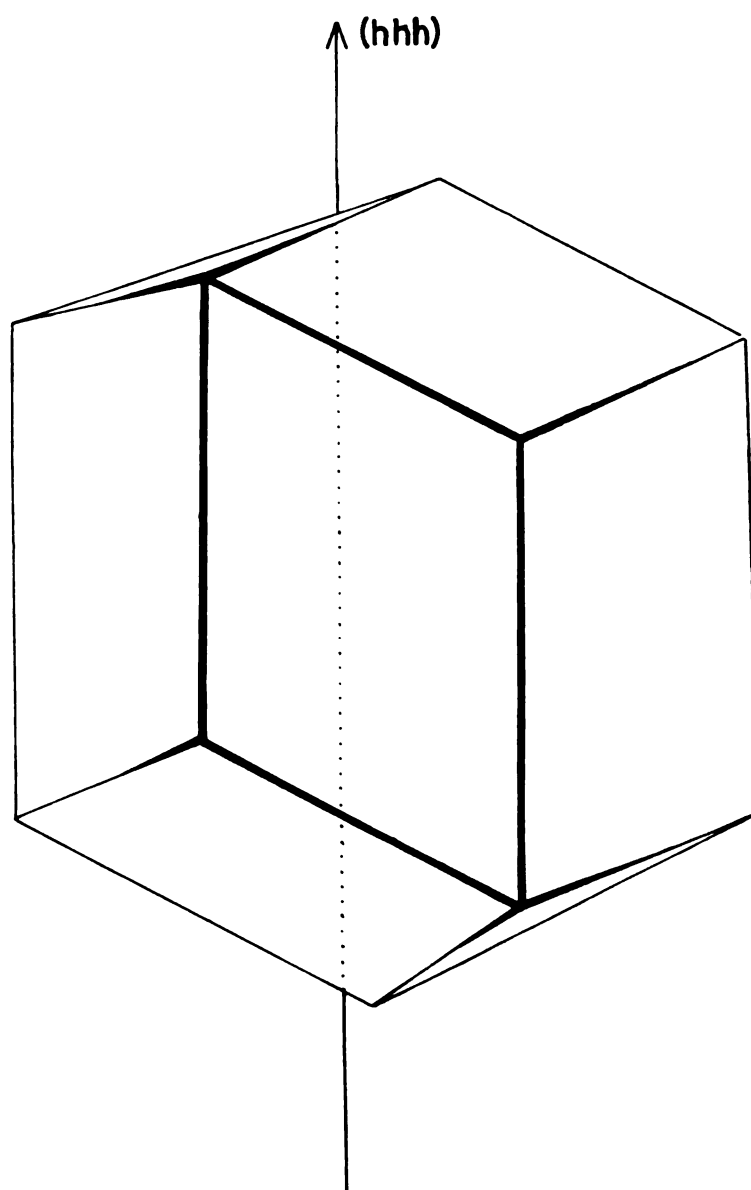


Figure 5. Morphology of an idealized crystal of KDPG aldolase.

that shown constitutes only one more face and because, in general, the overall depth is short, the crystals can be approximated as parallelepiped platelets. The amount of protein in the crystal (37%) was also taken into consideration as was the partial specific volume of KDPG-aldolase ($\bar{v}_p = 0.745 \text{ cm}^3/\text{g}$) and the molecular weight of the monomeric subunit (25,000 Daltons)⁴. For example, to calculate the number of moles of protein in a crystal of dimensions $0.5 \times 1.0 \times 1.0 \text{ mm}$, the volume of the crystal, $v = 0.5 \times 10^{-3} \text{ cm}^3$, must be multiplied by the partial specific density of the protein ($1/\bar{v}_p$) and the fraction of the crystal mass associated with the protein (0.37) in order to obtain the mass of protein in the crystal:

$$\text{mass of protein, } m_p = \frac{V \times 0.37}{\bar{v}_p} = 0.248 \times 10^{-3} \text{ g}.$$

Taking the molecular weight of the protein to be $\sim 25,000$ Daltons, the number of moles of protein in the crystal is:

$$\eta = 0.248 \times 10^{-3} \text{ g} / 25000 \text{ g moles}^{-1} = 9.93 \times 10^{-9} \text{ moles}.$$

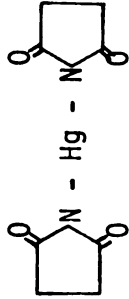
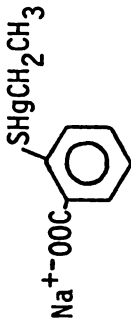
At more or less regular time intervals a crystal was removed from the solution where it was soaking in the presence of a heavy atom containing compound and the diffraction pattern along the principal axes $\vec{a}^*$, $\vec{b}^*$, $\vec{c}^*$, the body diagonal (hhh) and the face diagonal (0hh) of the reciprocal lattice was recorded. The diffraction pattern was then compared to the corresponding native enzyme pattern. If considerable changes in intensities had occurred, three dimensional intensity data were collected. If no or only small changes in the diffraction pattern occurred after a

considerable amount of soaking time, the initial concentration of the heavy atom reagent was doubled and the above process was repeated. Most of the heavy atom compounds tried either produced no change in the diffraction of the native protein crystals, even in high mole ratio of heavy atom to protein, or destroyed the crystals (e.g. phenylmercuric acetate, sodium para-chloromercuric benzenesulfonate, and related compounds).

The compounds that gave good isomorphous substitutions are listed in Table 2 along with the conditions of their preparation and some other data. Two of those, containing mercury (II), were stable for a period of 2-3 months so that it was fairly easy to collect 3.56Å resolution sets of intensity data. In fact, two sets of data were collected for each derivative. The first sets were of somewhat inferior quality because the crystals were small and the radiation damage to them severe. A second preparation of isomorphous crystals containing mercury (II) succinimide, hereafter denoted as HgS, and sodium ethylmercuric-thiosalicylate, hereafter denoted as EHgTS, was carried out on larger crystals under the same conditions as the first. Although generably more reliable, the results were comparable with those of the first sets. However, the intensities were not averaged between the two sets and only the intensities from the second preparation were employed in the structure analysis and determination of KDPG-Aldolase.

The isomorphous derivative containing KAuCl_4 was not very stable. Once intensity changes occurred in the diffraction pattern, intensity data were collected immediately, because the crystals apparently deteriorated as they soaked in the heavy atom containing solution. Attempts to reproduce the experiment of preparing the derivative in order to collect a 3.56Å resolution set of data from several crystals failed; changes in

TABLE 2. Conditions of Preparation of Heavy Atom Derivatives

Derivative	Mole Ratio Heavy Atom/Protein	Soaking Time	Resolution* (Å)	Cell Dimension (Å)	Number** of Crystals
Mercury (II) succinimide (HgS) 	10/1	20 days	3.56	103.34(5)	3
Sodium Ethylmercuric-thiosalicylate (EHgTS) 	20/1	1 month	3.56	103.53(4)	2
Na ₃ IrCl ₆	55/1	6 months	4.44	103.21(4)	1
KAuCl ₄	10/1	18 days	5.1	103.44(7)	1

* Resolution to which intensity data were measured.

** Number of crystals used for the data collection.

the diffraction pattern of the examined lines were reproducible, but the crystals were much more sensitive to both radiation damage during data collection and to the heavy atom reagent upon prolonged soaking.

A full $3.56\text{\AA}$ resolution set of data was also not obtained for the isomorphous crystals containing Na_3IrCl_6 (Table 2) because there were not enough crystals available in the first preparation. Since the difference Patterson synthesis of the derivative was very complicated, it did not appear to be particularly interesting, at least in the early stages of this work.

The changes in the dimensions of the unit cell of the derivatives, which is an indication of the degree of isomorphism, are very small compared to native ($|\vec{a}| = 103.4\text{\AA}$; Table 2). The changes in the intensities of the diffraction pattern along the $\vec{a}^*$ axis are not and they are shown in Figure 6.

2. Data Collection

The same approach was taken to measure the diffraction intensities of both native and isomorphous derivative crystals; therefore unless otherwise stated the following comments apply to all data sets collected.

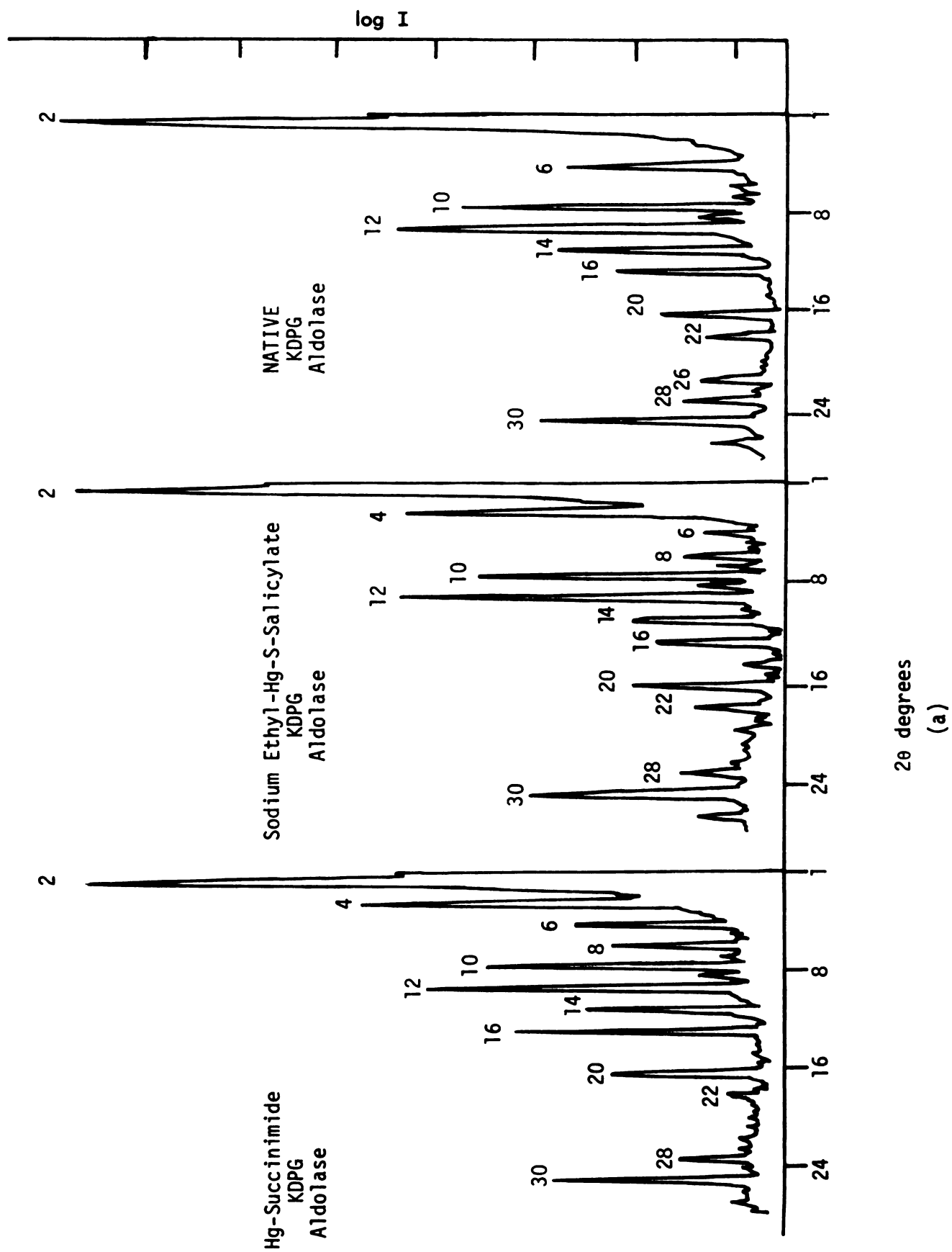
The space group $P2_13$ to which the crystals belong has 12 equivalent positions and the following relations hold among the structure amplitudes:

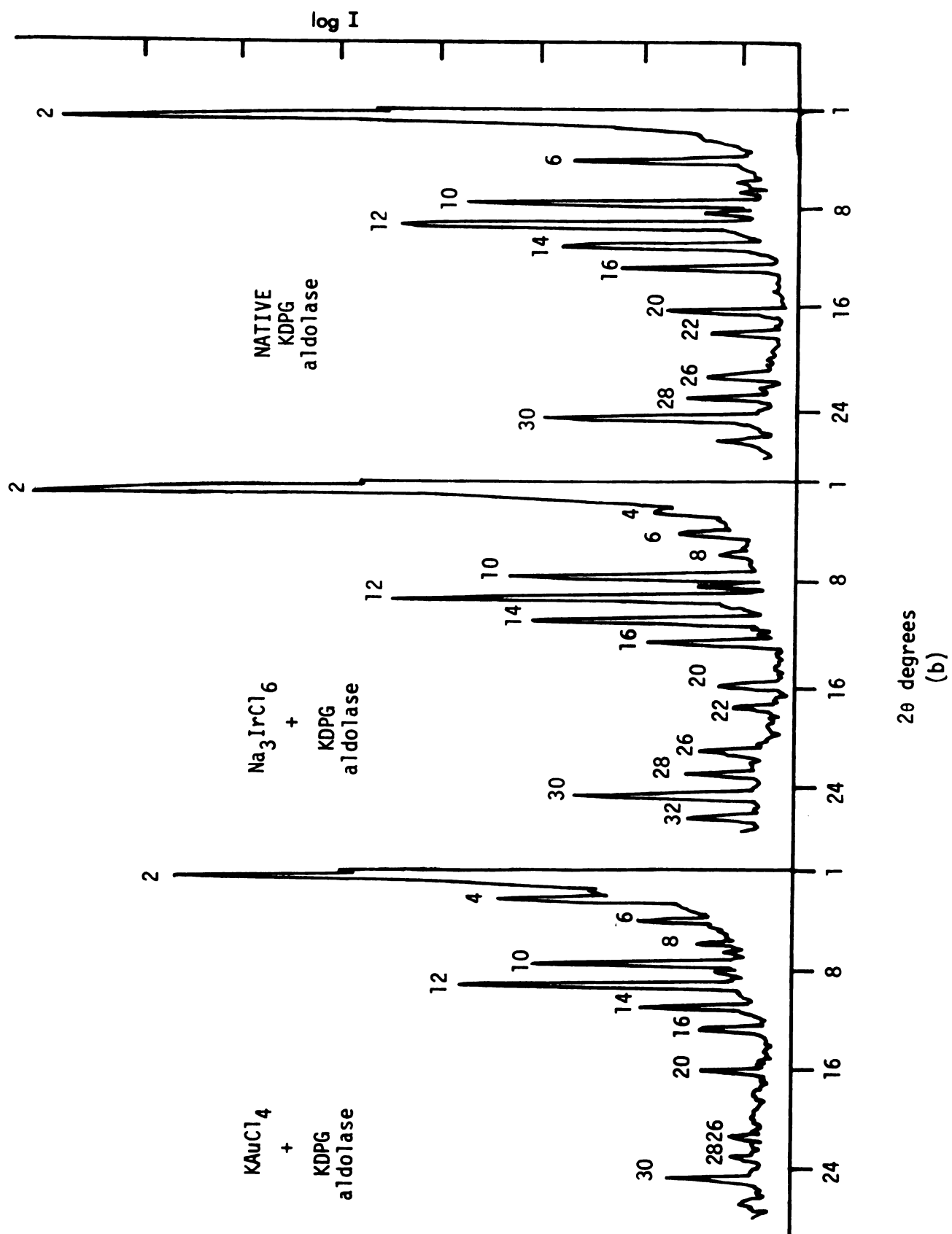
$$|F(hk1)| = |F(\bar{h}k1)| = |F(h\bar{k}1)| = |F(hk\bar{1})| = |F(\bar{h}\bar{k}\bar{1})|;$$

moreover, due to the three-fold rotation axis:

$$|F(hk1)| = |F(k1h)| = |F(1hk)|.$$

Figure 6. Distribution of diffraction intensities along the $\vec{a}^*$ axis of native and HgS and EHgTS (a), native and KAuCl_4 and Na_3IrCl_6 (b) containing crystals of KDPG aldolase.





Thus, only 1/24 of the diffraction intensities are unique. Although it was realized that measurement of redundant reflections would improve the quality of the intensity data by averaging, in the present case, such an advantage would be lost in the process of scaling the data among a much larger number of crystals. On the other hand, a unique set of intensities to 3.56 $\text{\AA}$ resolution (4577 reflections) could be obtained from only two crystal specimens.

The indices of the unique reflections that were to be measured were sorted in order of increasing 2θ angle and stored on the disc of the PDP-8 computer which controlled the data collection. The fact that the data were collected in increasing 2θ angle facilitated the combination of data from different crystals; the beginning of the data collection from a new crystal overlapped 100 or more reflections that were measured at the end of the data collection of the previous crystal. In this way the reliability of the measurements and corrections at the end of the first data collection, where the crystal had suffered the largest radiation damage, were ascertained and confirmed.

Crystals that were to be used for intensity measurements were chosen to be without cracks or spurs, single and of the largest available size (0.5x1.0x1.0mm). They were mounted with the body diagonal direction (hhh) parallel to the azimuthal axis (spindle direction) of the diffractometer. Thus, in that orientation the (hhh) line was at $\chi=90^\circ$ and the axes $\vec{a}^*$, $\vec{b}^*$ and $\vec{c}^*$ were at $\chi=35.2^\circ$ with their ϕ values differing by 120° . The crystals were mounted in capillary tubes of common glass and diameter of 1.5mm; the capillaries were not coated by any water repellant agent. After the crystals had been oriented properly with drops of mother liquor below and above it, the tube was sealed in a low flame⁵⁰.

Reflection intensities were measured by using Cu K α radiation (1.5418Å) on a Picker four-circle diffractometer controlled by a Digital Equipment Corporation (DEC) 4K PDP-8 computer (FACS-I system). The computer is coupled to a DEC 32K Disc File and an AMPEX TMZ 7-track tape transport. An ω -step scan procedure was performed utilizing balanced Ni/Co filters placed between the X-ray beam and the crystal to measure background. Each step was performed in 0.03° increments of the ω -angle and the scan extended $\pm$ 0.075° on either side of the calculated ω -value of the peak position. Each step was measured for a duration of four seconds and the four largest measurements were summed to give the intensity of the reflection⁵¹. In order to allow for small variations from the calculated ω -angle of the reflection, as when the crystal was slightly misaligned, the program carried out one or two additional steps in the correct direction⁵².

The background was measured with the Co filter at the ω -value of the maximum intensity for a time interval of four seconds and this measured intensity was multiplied by four to give the total background.

The alignment of the crystal was monitored by measuring three standard reflections periodically every 100 reflections (about 1 hour)⁵².

These reflections were particularly sensitive to misalignment: (16,16,16) at $2\theta = 23.85^\circ$, $\chi=90^\circ$ and at two different ϕ values approximately 90° apart and the general reflection (2,20,7) at $2\theta = 18.28^\circ$ and $\chi=52^\circ$.

When the intensity of a monitored reflection decreased by 10%, the crystal was considered misaligned and a sub-routine automatically found the new orientation of the crystal. The intensities of the monitored reflections were also used to estimate the decrease in the intensity of the reflections due to the radiation damage of the crystal by the X-ray beam.

The following preliminary measurements were always performed before the measurement of the intensities of a crystal.

- a. Lack of balance measurement. The measured background had to be corrected because the balance of the Ni/Co filters was not exact. The correction was obtained for each crystal empirically by measuring background with the Ni and the Co filters separately as a function of 2θ . The difference in the measurements was taken to be the lack of balance and it was added to the background.
- b. Absorption measurement. The intensities of the (4,4,4), (6,6,6), (11,11,11) and (16,16,16) reflections at $\chi=90^\circ$ were measured as a function of the azimuthal angle for use in the correction of the intensities of the reflections for absorption according to an empirical method⁵³. The maxima of the absorption ratios $I_{\max}/I$, where $I_{\max}$ is the maximum intensity of the measured reflection and I the intensity of the same reflection at any other ϕ angle, varied from 1.4 to 1.9. The majority of the ratios were at or below 1.6 and only one crystal had a maximum ratio of 1.9.

No survey of the diffracted peak intensities as a function of the ω -angle was carried out to estimate the mosaicity of the derivatives. However, the diffraction pattern along the recorded lines showed a broadening of the peak profiles of the derivatives with respect to those of the native enzyme crystals.

The unit cell parameter of the crystals was determined from the least square fit of the calculated and observed Bragg angles of twelve reflections of high 2θ angle (18° – 25°) distributed uniformly in the part of the reciprocal space to be used for intensity measurements. In addition, 2-3 reflections with a χ angle greater than 70° were always included to insure reliability of the calculated orientation angles⁵². The least

squares procedure was constrained to fit one lattice parameter and three orientation angles.

The three dimensional intensity data collection was preceded by a separate $6.0\text{\AA}$ resolution intensity data of the centrosymmetric (0kl) zone to be used to obtain scaling factors between native and derivative crystals. The three dimensional data were collected in order of increasing 2θ . The number of reflections collected from each crystal depended on the rate of radiation damage of the crystal. Upon completion of the three dimensional intensity data collection from a given crystal, another set of intensities of the centrosymmetric (0kl) zone at $6.0\text{\AA}$ resolution was measured in order to confirm the damage of the crystal deduced from the decrease of the intensities of the monitored reflections and to confirm that the (0kl) Patterson projection was the same before and after the data collection.

Finally, the absorption was remeasured for most of the crystals to assure that it did not change during the data collection. The intensity distributions along the (h00), (0h0), (00h), (hhh), and (0hh) lines were also recorded and were compared qualitatively to those corresponding to before the data collection. The relative intensity distribution was the same before and after the data collection, but in general, the intensities of the reflections at higher 2θ angles had decreased much more than those at lower 2θ angles.

3. Data Reduction and Scaling

The following corrections were made to the intensities before they were reduced to structure amplitudes.

- a. The background was corrected for lack of balance; the correction was

subtracted from the intensities measured with a Ni filter to give the net measured intensities.

b. The net measured intensities were corrected for absorption with an empirical multiplicative absorption factor which depends on the 2θ , χ and ϕ of each reflection⁵³.

c. The net measured intensities were also corrected for radiation damage (decay). The decay factors were estimated from the decrease of the intensities of the monitored reflections with X-ray exposure which was generally linear. However, the slope for the reflections with 2θ angles smaller than 15° was about half that of reflections of 2θ greater than 15° .

Each reflection was multiplied by a correction factor for decay, D , $D = N \times \text{slope} + \text{intercept}$, where N is the serial number of the reflection during the data collection proportional to the amount of time the crystal was exposed to X-ray radiation before the reflection was measured, slope and intercept are the slope and intercept of the line obtained by plotting the intensity of the monitored reflections with respect to N .

d. Finally, the intensities were corrected for Lorentz and polarization factors to give relative structure amplitudes.

Table 3 lists the data sets which were used in the solution of the structure, the number of intensity data measured from each crystal and the percent decrease of the intensity of the monitored reflections at the end of the data collection, $\Delta I/I_0$, where I_0 is the initial intensity. Since the monitors had a 2θ angle in the range $18 < 2\theta < 24$, most reflections did not suffer this amount of radiation damage.

The combined number of unique intensity observations from the two native protein crystals at $3.56\text{\AA}$ resolution is 4577. Of those, 3853 were considered observed. The observation limit was defined by the intensity

TABLE 3. Data Collection Parameters and Scaling Constants

		2θ Range (degrees)	Number of Reflections	Decay (%) Decrease in Monitors	Scale Applied to F 's	Overall Scale (S)	Overall B
Native							
crystal 1	2.5-20.3	2528	30	1.0			
crystal 2	20.0-25.0	2150	25	1.1			
HgS						0.99	2.0
crystal 1	2.5-19.0	2016	27	1.2			
crystal 2	18.8-23.0	1458	32	1.42			
crystal 3	22.0-25.0	1004	20	1.43			
EHgTS						1.0	7.0
crystal 1	2.5-22.3	2960	23	1.01			
crystal 2	20.0-25.0	1612	9	1.57			
Na ₃ IrCl ₆		2.5-20.0	2396	34	-	1.15	14.5
KAuCl ₄		2.5-17.4	1652	38	-	1.54	45.0

of the systematically unobserved reflections. Because the derivatives were more sensitive to radiation damage, only the intensities of the 3853 observed reflections were measured in the derivative data collections.

The structure amplitudes of the first native crystal were used as a reference to which the structure amplitudes of the reflections were scaled for the remainder of the crystals. The scaling for the HgS and EHgTS was carried out in two steps. First, the Patterson synthesis of the (0kl) projection was calculated from the measured intensities of the zone at $6.0\text{\AA}$ resolution before the measurement of the three dimensional data. The ratio of the peaks of the native crystal 1 which were not affected by substitution, to the corresponding peaks of the derivatives were estimated and the average was applied to the square of the structure amplitudes as a scale. This process was also used to place the two native protein crystals on the same scale. A second scaling scheme was used to refine the initial scale and to verify the validity of the corrections applied to the intensities. Scaled averaged $|F_{PH}|$ values of the derivative were plotted versus their average 2θ for eight 2θ ranges and these were compared to the corresponding distribution of the native structure amplitudes. If deviations occurred, the $|F_{PH}|$'s were multiplied by a factor Q ,

$$Q = S \times \exp(B \sin^2 \theta / \lambda^2) ,$$

where S is an overall scale factor and the exponential is an overall temperature-like factor. The above constants are shown for the derivatives in Table 3 and are indicated as overall scale and overall B, while in the same table the scaling factors obtained from the Patterson

projections for each crystal are shown under scale.

In the case of the KAuCl_4 and Na_3IrCl_6 derivatives, no preliminary scaling by a Patterson projection was attempted. The structure amplitudes were scaled to the native structure amplitudes by only considering the distribution plot and it is for this reason that the overall scales differ so much from unity. This second approach to the scaling seemed to be quite satisfactory. The first set of protein intensity data, not used for the solution of the structure, was scaled by the distribution plot to the second set. Its overall scale was 1.11 and the overall B $4.0 \text{ \AA}^2$. The residual factor between the two data sets is $R = 0.028$, where

$$R = \frac{\sum_{hkl} ||F_{\text{NAT1}}| - |F_{\text{NAT2}}||}{\sum_{hkl} |F_{\text{NAT2}}|}.$$

The relative $|F|^2$ distribution curves for the native and derivative structure amplitudes are shown in Figure 7.

The anomalous differences in the intensities of Bijvoet pair reflections²⁷ were measured for HgS and EHgTS to approximately $6 \text{ \AA}$ resolution from the same crystals that the corresponding isomorphous data were collected. The intensity of an (hkl) reflection was measured followed by its $(\bar{h}\bar{k}\bar{l})$ counterpart. The intensities of the $(\bar{h}\bar{k}\bar{l})$ reflections were processed in exactly the same way as those of the (hkl) reflections. The lack of balance and absorption corrections were measured separately at -2θ angles and separate distribution curves refined the scaling of the structure amplitudes to those of the native crystals.

The Wilson plot⁵⁴ corresponding to the logarithm of the ratio of the average $|F|^2$ to the average sum of the squares of the scattering factors of the atoms of the native enzyme molecule versus $\sin^2\theta$ in seven

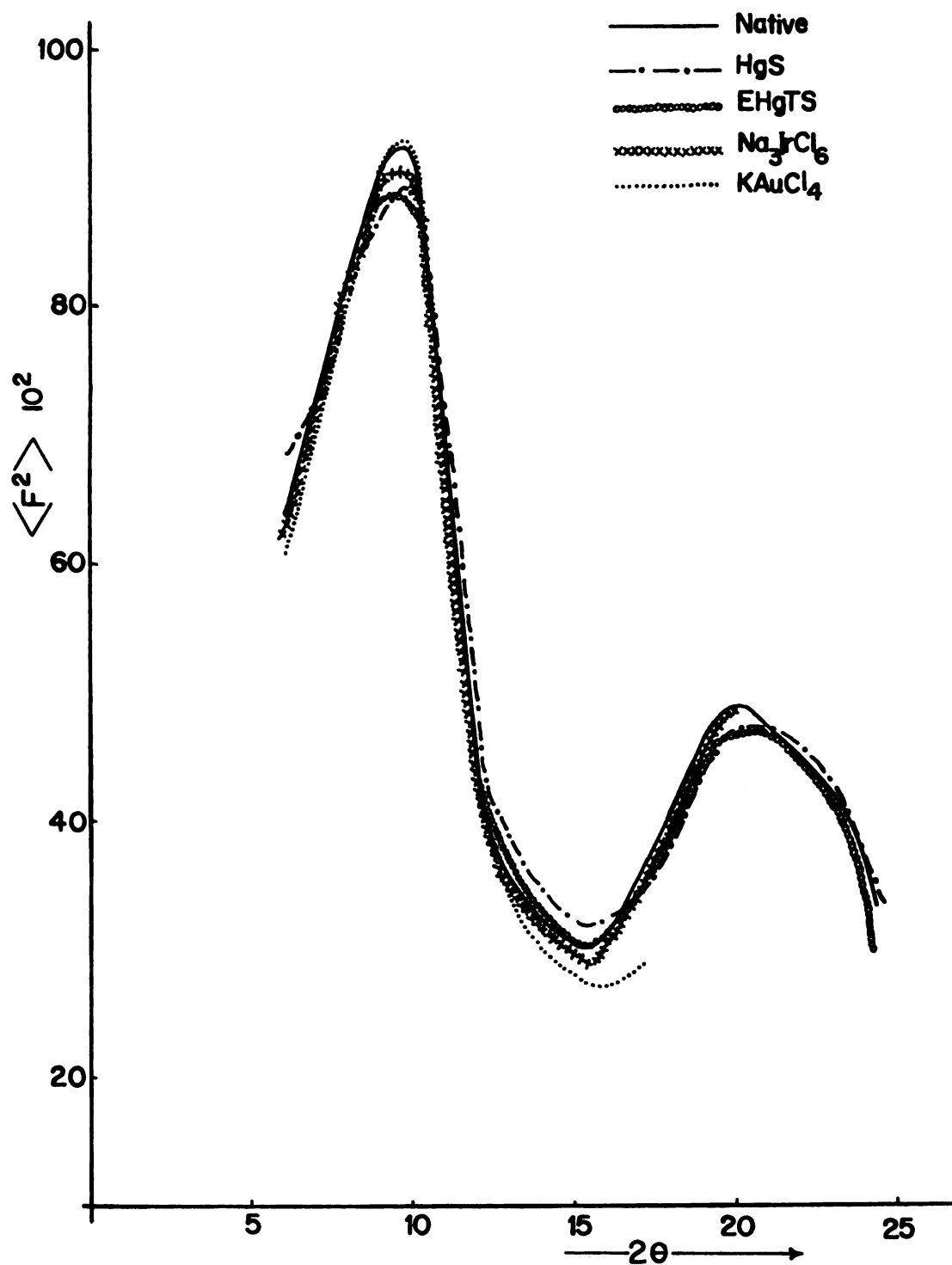


Figure 7. Distribution of the square of the structure amplitude of native and derivative crystals of KDPG aldolase versus 2θ angle.

ranges gave an absolute scale factor and an overall temperature factor for the KDPG-Aldolase crystals of:

absolute scale factor $K_W = 0.117$

temperature factor = $32.3 \text{ \AA}^2$

The calculation was based on the three ranges before the last corresponding to $19^\circ < 2\theta < 23.5^\circ$.

III. SOLUTION OF THE PHASE PROBLEM AND REFINEMENT OF THE PROTEIN PHASES

1. Initial Determination of the Main Heavy Atom Substitution

Three dimensional difference Patterson syntheses as well as centrosymmetric (0kl) projections with coefficients $|\Delta F|^2 = ||F_{PH}| - |F_P||^2$ were calculated for all derivatives. The three dimensional maps were, in general, easier to interpret than the projections because there was less overlapping of vectors and the peak height to noise ratio was higher; the projections were only used to confirm the coordinates deduced from the three-dimensional maps.

The Harker sections for the derivatives are shown in Figures 8-11. For the space group $P2_13$, there are three identical mutually perpendicular Harker sections. For a single substitution site with general coordinates (x,y,z), vector peaks are expected at:

- (a) $1/2, 1/2 \pm 2y, \pm 2z$
- (b) $\pm 2x, 1/2, 1/2 \pm 2z$
- (c) $1/2 \pm 2x, \pm 2y, 1/2$,

in the Harker sections $(1/2, v, w)$, $(u, 1/2, w)$ and $(u, v, 1/2)$, respectively. However, due to the three fold rotation axis, the positions generated by the cyclic permutation of the above peaks also exist. Therefore, three times as many vectors appear in every Harker section, so that only one needs to be examined e.g. for the $(1/2, v, w)$ Harker section there will

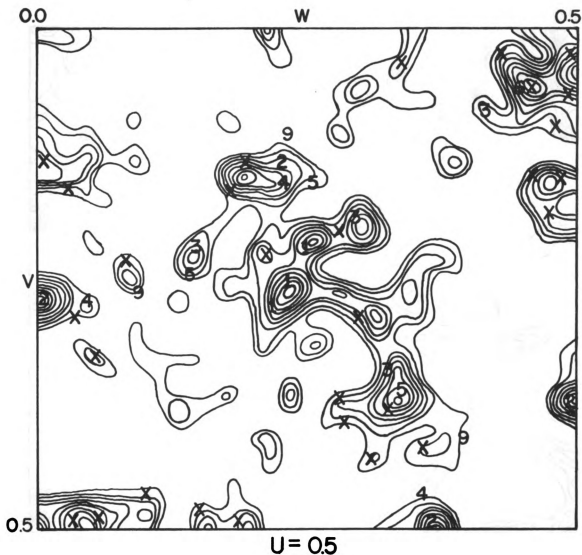


Figure 8. Harker section of difference Patterson synthesis at an arbitrary scale based on HgS data; contour intervals at 100 beginning at 100. The numbers indicate self vector positions and the X's cross vector positions of the corresponding heavy atom sites with the principal sites 1, 2 and 3.

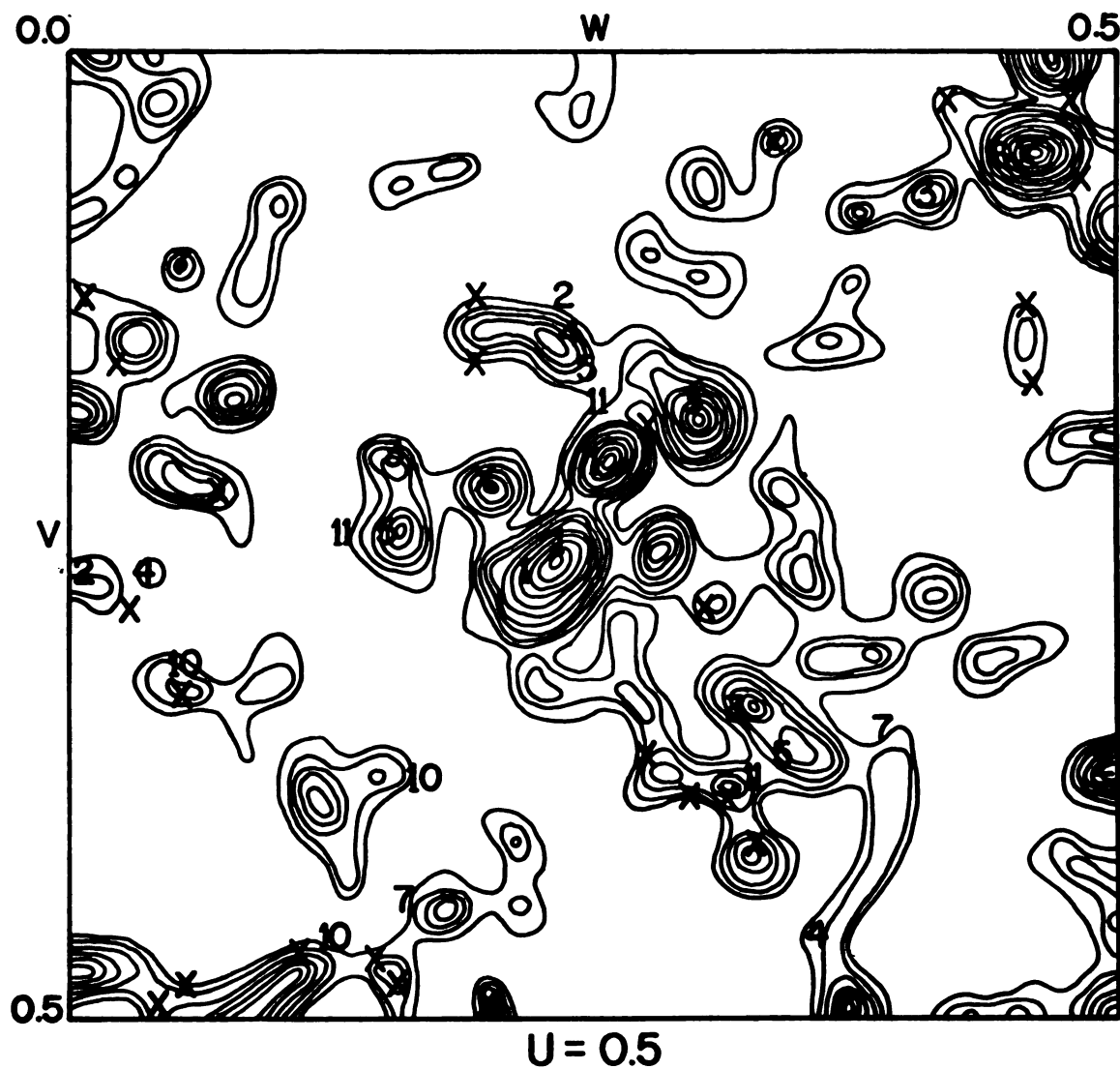


Figure 9. Harker section of difference Patterson synthesis based on EHgTS data; scale and symbols as in Figure 8. Contour intervals at 100 beginning at 100.

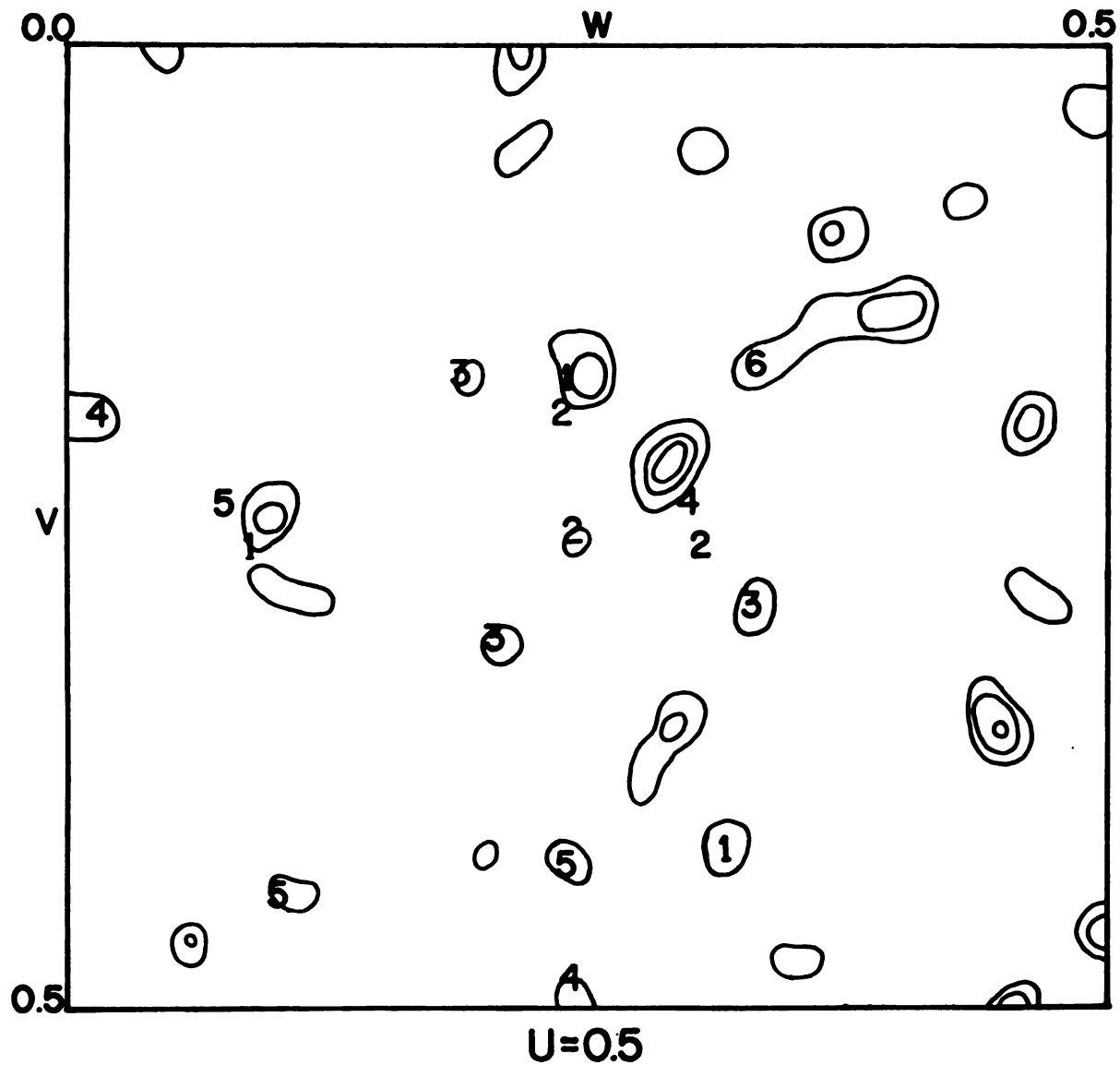


Figure 10. Harker section of difference Patterson synthesis based on Na_3IrCl_6 data; scale and symbols as in Figure 8. Contour intervals at 50 beginning at 100.

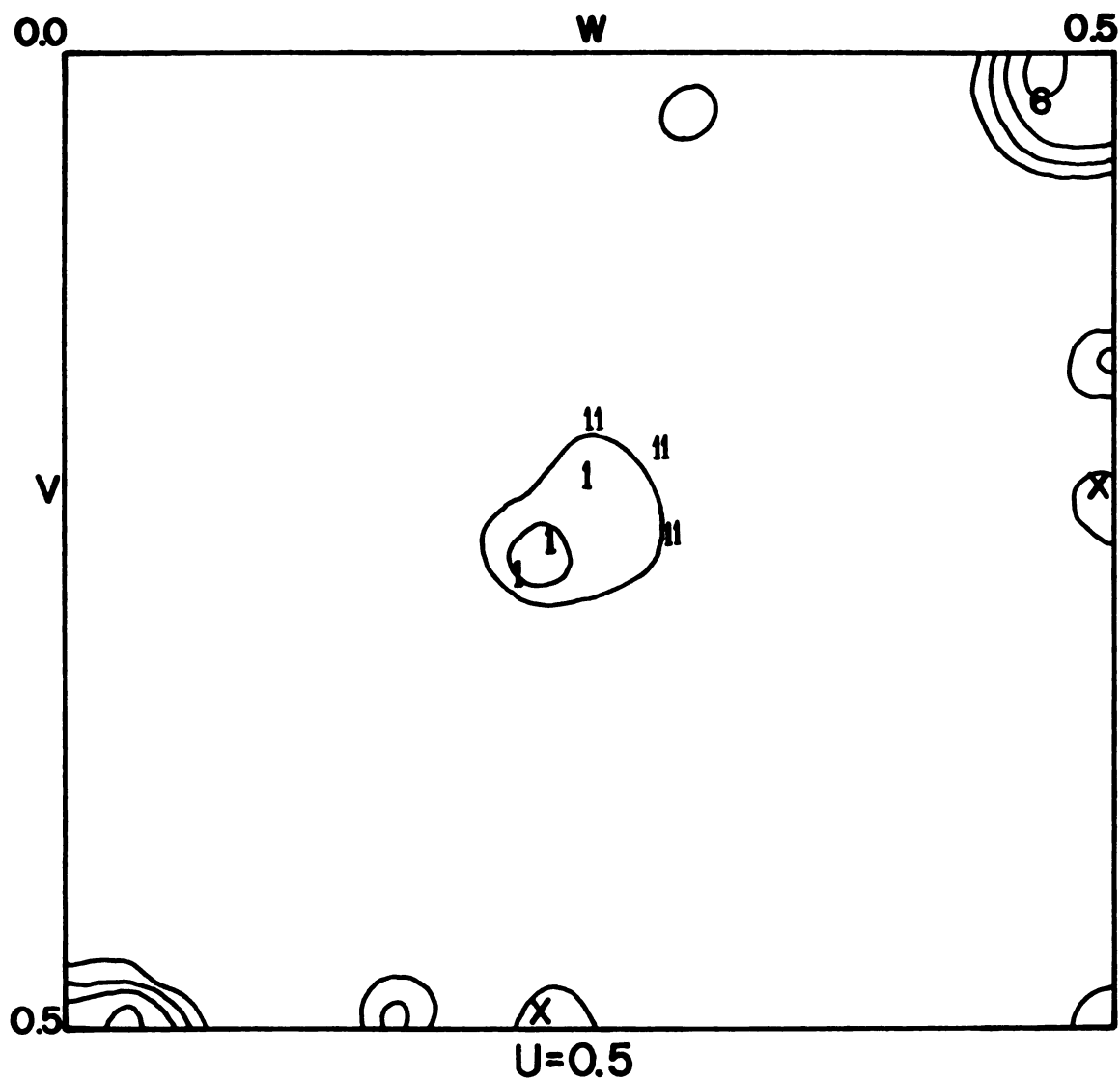


Figure 11. Harker section of difference Patterson synthesis based on KAuCl_4 data; scale and symbols as in Figure 8. Contour intervals at 50 beginning at 100.

be peaks at:

(i)	$1/2$	$1/2 \pm 2y$	$\pm 2z$
(ii)	$1/2$	$1/2 \pm 2z$	$\pm 2x$
(iii)	$1/2$	$1/2 \pm 2x$	$\pm 2y$

Multiple solutions of atomic coordinates are possible from the Harker vectors. Selection of one such solution for a substitution site of one derivative establishes the crystallographic origin to which the coordinates of the other substitution sites of this and of the other derivatives will be referred and fixes an enantiomorphic configuration. From the Harker section of HgS, two prominent sites were found, HgS1 and HgS2. The cross vectors between these sites were examined in order to refer the second position to the same origin as the first. There are 16 possible combinations of the coordinates of HgS2 whose vectors with the given coordinates of HgS1 had to be checked in order to find the correct one and yet the result was not without ambiguity. Moreover, it was also clear that at least one more site of high occupancy existed but its position could not be extracted from the Harker section. The Harker section of EHgTS is very similar to that of HgS as can be seen from Figures 8 and 9; this indicates that substitution occurs at similar positions. However, the EHgTS vectors which correspond to HgS2 are very weak and this shows that this position has lower occupancy in EHgTS. The additional peaks in the Patterson synthesis of EHgTS which do not correspond to positions 1 or 2 also indicate that, in this derivative, there is an additional site of high occupancy.

In order to gain more confidence in the determined positions of HgS1 and HgS2 and in an effort to find additional sites, an alternative approach to the difference Patterson method was desirable. The alternative investigated was the application of direct methods to locate the heavy atoms. The approximate quantities $||F_{PH}| - |F_P||$ were used as an approximation to the observed structure amplitudes for the heavy atom contribution. Normalized structure amplitudes ($|E|$'s) were computed in the usual manner²⁴ from the approximate difference structure amplitudes. The twelve equivalent positions of the space group coupled with, at least, three substitution sites makes the underlying assumption of direct methods, that there is a uniform distribution of scattering matter in the unit cell, reasonable. If one makes the additional assumption that the approximate difference structure amplitudes of the heavy atom are not very different from the exact but unknown amplitudes one should expect reasonable results from the application of the methods.

The largest 205 E's up to 1.75 were used in the direct method analysis. The program employed for the purpose was MULTAN⁵⁵, a general program for all space groups. It uses the Σ_2 formula⁵⁶ to set up the phase relationships between reflections which are later used by the tangent formula⁵⁶ to assign phases. Eight sets of phases were determined, each characterized by different reliability factors. The latter are listed in Table 4 for HgS and EHgTS. The absolute figure of merit, ABS FOM, is a measure of the consistency of the set of phases and if the space group is non-symmorphic, as $P2_13$, the correct phases should either give the highest possible value of ABS FOM or slightly below the maximum. ABS FOM is zero for random phases and it should be unity for correct phases, but because of the way the tangent formula operates it is usually

TABLE 4. Figures of Merit for the Eight Solutions Given by MULTAN
for HgS and EHgTS.

Compound	Solution Number	ABS FOM	RESID
HgS	1	.5616	46.65
	2	.5617	45.06
	3	.5642	45.52
	4	.5639	45.14
	5	.5615	46.72
	6	.5614	46.07
	7	.4722	46.56
	8	.2708	51.91
EHgTS	1	.4675	42.33
	2	.2567	36.77
	3	.2535	34.73
	4	.2365	32.85
	5	.2441	36.48
	6	.3150	34.53
	7	.2824	36.13
	8	.2737	36.20

closer to 1.2 or even higher. The second reliability criterion, RESID, is merely an ordinary crystallographic residual:

$$\text{RESID} = \sum_h ||E_h|_{\text{obs}} - |E_h|_{\text{calc}}| / \sum_h |E_h|_{\text{obs}}.$$

$|E_h|_{\text{obs}}$ is the observed amplitude of E_h and $|E_h|_{\text{cal}}$ is the calculated from the equation:

$$E_h = S \sum_{h'} E_{h'} E_{h-h'},$$

where S is a scaling factor:

$$S = \sum_h |E_h| / \sum_h \left| \sum_{h'} E_{h'} E_{h-h'} \right|.$$

Clearly RESID should be a minimum for the correct phases, but experience shows that it is not as powerful a criterion as ABS FOM.

As Table 4 indicates, all ABS FOM, especially for EHgTS, are much lower than the ABS FOM expected for correct phases in the case of small structures where exact structure amplitudes are employed; this, however, is more or less expected. Three dimensional E maps for all eight sets of phases were calculated for both derivatives. For HgS, the sets 1-6 gave approximately the coordinates of site HgS1 with respect to different origins, while the set 7 gave the coordinates of two heavy atoms corresponding to one of the possible combinations of the coordinates of HgS1 and HgS2 as determined from the difference Patterson synthesis. These are listed in Table 5. The phases of set 8 generated an E-map resembling that of set 7 but the peak heights were very low.

As the ABS FOM's in Table 4 suggest, the phase determination of EHgTS was not as good as that of HgS; there was no consistency in the coordinates of the peaks appearing in the E-maps of the various solution sets. One set showed a substitution with the same coordinates as HgS1 and two others suggested new substitutions which were later found by difference electron density methods. In general, however, the quality of the phase determination was fairly poor so that the results of the direct method were not considered in the case of EHgTS.

Having gained more confidence in the positions of two of the heavy atoms, which were the same for HgS and EHgTS, and knowing that the second is of low occupancy in EHgTS, the difference Patterson map of the latter was studied more carefully. Particular attention was paid to larger peaks in general positions that might hopefully correspond to cross vectors between position 1 and some new position. This intensive search resulted in the successful location of a new substitution, EHgTS3, whose self vectors and cross vectors with the position EHgTS1 explained almost all the strong peaks of the EHgTS difference Patterson map. The same substitution was also found in HgS, and most of the strong peaks of the difference Patterson of HgS were also explained. These results were then confirmed by the calculation of Kartha's correlation function⁴⁰, with coefficients as in equation (6), where iso1 is HgS and iso2 EHgTS. All the cross vectors between HgS1, HgS2, HgS3 and EHgTS1 and EHgTS3 were located.

The coordinates of these three substitutions of HgS and EHgTS, as initially determined, are shown in Table 5 along with other related quantities. At the time, it was not possible to detect any difference in their coordinates. Their location was in effect the most important

breakthrough in the problem since all the other substitutions of the above derivatives and other derivatives were determined very easily from them.

Considerable time was also spent in the interpretation of the three dimensional difference Patterson map of Na_3IrCl_6 ; this map shows many peaks of low peak height compared to those of HgS and EHgTS (Figure 10), and no substitution of prevailing occupancy seemed to exist. Even though three substitutions were found from the Patterson that were also consistent with the peaks of the Kartha correlation function⁴⁰ between Na_3IrCl_6 and HgS or EHgTS, this derivative was not included in the initial calculation of protein phases.

The derivative containing KAuCl_4 has a rather simple Harker section shown in Figure 11. The vectors indicated by 1 can be explained well by a substitution with the same or nearly the same coordinates as HgS1. This is also consistent with the Kartha correlation map between KAuCl_4 and HgS. Another substitution with all self vectors at the region of the Harker section indicated by 6 is possible. These vectors are also present in the difference Patterson maps of HgS and EHgTS. The above conclusions were confirmed by calculating the centrosymmetric (0kl) projection using as coefficients the quantity $||F_{p+Au}| - |F_p||\exp(i\alpha_p)$, where α_p are the protein phases determined by using the HgS and EHgTS derivatives. However, the KAuCl_4 derivative was also not used in the initial determination of the protein phases; a more reliable independent confirmation of the atomic coordinates still seemed desirable.

TABLE 5. The Positions of the Three Substitutions of HgS and EHgTS Determined by the Difference Patterson Synthesis and by Direct Methods.

Compound	Site	x	y	z	Occupancy Estimated from Difference Patterson	Occupancy from Least Squares (in electrons)
HgS	1	0.119	0.144	0.873	high	38
	2	0.020	0.182	0.885	high	28
	3	0.647	0.922	0.661	high	36
EHgTS	1	0.119	0.144	0.873	high	40
	2	0.020	0.182	0.885	low	13
	3	0.647	0.919	0.658	high	44

2. Determination of the Atomic Parameters by Difference Fourier Synthesis and Refinement of Protein Phases

The three main substitutions shown in Table 5 account for most of the largest vector peaks in the difference Patterson of HgS and EHgTS. Although the coordinates of the three are the same in both derivatives, HgS2 and EHgTS2 have different occupancies; therefore the heavy atom contribution, $\vec{F}_H$, to the structure amplitudes of the heavy atom derivatives is different and, more important, not collinear. Thus, the three dimensional protein phases calculated by the multiple isomorphous replacement method using these derivatives should be good enough to give information as to additional substitutions.

In order to establish an initial estimate of the occupancies of the above substitutions, the calculated heavy atom contribution was refined by the method of least squares varying only the occupancies. The quantity minimized was:

$$\sum_{hkl} (||F_{PH}| - |F_P|| - K|F_{Hcalc}|)^2, \quad (28)$$

where $K=1/K_w$ is the inverse of the absolute scale as determined by the Wilson plot⁵⁴. The residual factors R,

$$R = \frac{\sum_{hkl} ||F_{PH}| - |F_P|| - |F_{Hcalc}|}{\sum_{hkl} ||F_{PH}| - |F_P||}, \quad (29)$$

after the refinement were 50.2%, and 51.1% for HgS and EHgTS, respectively. The occupancies in electron counts of the sites determined by the procedure are listed in Table 5.

After these preliminary occupancies had been established the protein phase calculation gave an average figure of merit⁴⁵, $\langle m \rangle$, of 0.531. Of 3853 observed native enzyme reflections, the number of reflections with figure of merit greater than 0.5 (referred to hereafter as $N(0.5)$) was 2117.

Difference Fourier syntheses were calculated for both HgS and EHgTS using reflections with a figure of merit greater than 0.5 (as in all subsequent difference Fourier maps). The coefficients of the Fourier syntheses were those suggested by Steinrauf⁴¹:

$$\Delta = (|F_{PH}| - |F_P|) \exp(i\alpha_P) ,$$

where α_P is the best protein phase. This type of Fourier synthesis will be referred to hereafter as difference Fourier. The substitutions used for the phase determination appeared very large in these maps. In addition, two more positions, lying close together, appeared in both maps with peak heights approximately 25% of the height of the major substitutions, while the background was only about 12%. The larger of the two, HgS4 and EHgTS4, was included in the next protein phase calculation which had a mean figure of merit $\langle m \rangle = 0.534$ and $N(0.5) = 2090$.

The protein phases obtained from the above calculation were used to calculate a difference Fourier for the Na_3IrCl_4 derivative. The major features of this map suggested two substitutions, Ir1 and Ir2, which generated vectors consistent with the difference Patterson of Na_3IrCl_6 . From the peak heights, the occupancies of these substitutions were much lower than those of the mercury containing derivatives. Moreover, the mercury substitutions which had been used for the protein phase

determination appeared as "ghost" peaks with peak height higher than those of the iridium substitutions, but about 10 times smaller than in the mercury difference Fourier maps. This was due to the fact that the differences in amplitude between the Na_3IrCl_6 derivative and the protein are small and that the protein phases are dominated by the mercury contribution.

Nevertheless, the Na_3IrCl_6 derivative was used in the next protein phase determination as a third derivative. As an initial estimate of the occupancies of the two iridium substitutions, the result of the least square refinement of the iridium occupancies was used (as in the case of the mercury derivatives). The protein phase determination showed better overall statistics when the third derivative was included ($\langle m \rangle = 0.578$, $N(0.5) = 2347$). These phases were then used to calculate difference Fourier maps for all the derivatives. In addition, another type of difference Fourier map, which will subsequently be called double-difference Fourier was calculated; its coefficients are those suggested by Blake *et al*⁴⁶:

$$\Delta\Delta = (|F_{\text{PH}}| - |\vec{F}_{\text{P}} + \vec{F}_{\text{H}}|) \exp(i\alpha_{\text{PH}}) ,$$

where α_{PH} are the phases of the heavy atom derivative (equation (27)). The difference Fourier map possesses all the substitutions of a derivative, while the double-difference map additionally subtracts the substitutions included in the protein phase determination; thus, no difference density should appear at the positions of the latter unless the positions and/or the occupancies of these atoms are not exactly like those observed. The purpose of calculating both types of difference maps at this point was twofold: firstly to confirm the occupancies of the known substitutions

and secondly to ascertain if any additional minor substitutions are present in both maps. The double-difference Fourier maps showed that the occupancies of the included substitutions of HgS and EHgTS should be increased and that of the Na_3IrCl_6 should be decreased. Both types of maps showed four additional substitutions HgS5 (also apparent in the previously calculated difference Fourier) and HgS6, HgS7, HgS8 for the mercury (II) succinimide derivative and EHgTS5, EHgTS6, EHgTS7, EHgTS8' for the ethylmercuric-thiosalicylate derivative; HgS8 and EHgTS8' do not have the same coordinates. The peak heights of the newly found sites were about 15% of that of the major substitutions, while the background decreased to about 8%. The self vectors (except for HgS6 and EHgTS6) did not show in the difference Patterson map, but the cross vectors involving the major substitutions were present. In the case of Na_3IrCl_6 , two more sites were found, Ir3 and Ir4, of occupancy comparable to those of Ir1 and Ir2 and the "ghost" peaks of the mercury atoms were still present in both maps.

At this stage, a difference Fourier was calculated for the KAuCl_4 derivative. Two very prominent peaks appeared at similar positions as HgS1 and HgS6. However, since there was also positive density at the positions of the other mercury atoms used in the protein phase determination (as in the case of the Na_3IrCl_6 difference Fourier), some uncertainty remained concerning their true nature, even though there were consistencies with the difference Patterson map. Thus, the KAuCl_4 derivative was still not included in the phase determination as fourth derivative.

Including the four additional substitutions sites in the HgS and EHgTs derivatives, and the two sites in Na_3IrCl_6 , improved the mean figure of merit to $\langle m \rangle = 0.584$ but $N(0.5)$ increased only slightly

(by 43) to 2390 reflections.

Further improvement of the protein phases was carried out by refining the occupancies and coordinates of all heavy atoms by the method of least squares. The program used for the refinement as well as for the determination of protein phases has been written originally by Rossmann⁵⁷. It minimizes the quantity defined by equation (23), where w_n , the weighting factor for each reflection, is $1/E^2$ and E , the lack of closure error, was taken to be the average isomorphous error for the region of $\sin\theta$ to which the particular reflection belongs:

$$\bar{E}^2 = \bar{E}_{iso}^2 = (\overline{|F_{PHobs}| - |F_{PHcalc}|})^2,$$

where $|F_{PHobs}|$ is the observed structure amplitude of an isomorphous derivative and $|F_{PHcalc}|$ is the calculated amplitude using the protein structure amplitude for a given set of protein phases and the calculated heavy atom contribution. The coordinates of the substitution sites did not change much ($0.1\text{\AA}$ on the average) with refinement, but the occupancies of the main sites changed considerably. In agreement with the double-difference Fourier the occupancies of the mercury atoms increased, while those of the iridium atoms decreased. After the refinement, the protein phase determination had improved statistics of $\langle m \rangle = 0.602$ and $N(0.5) = 2491$. The agreement indices (defined in Table 7) improved immensely for HgS and EhgTS, but they became slightly worse for the Na_3IrCl_6 derivative. Another cycle of refinement of the same parameters marked by increased $\langle m \rangle$ to 0.670 and $N(0.5)$ to 2887.

A difference Fourier of the KAuCl_4 derivative was calculated with the phases of the second cycle of refinement. The sites with coordinates

similar to those of HgS1 and HgS6 again appeared very large while all the other mercury peaks, which were previously suspected as being "ghosts", had disappeared except for HgS3; the latter, however, also decreased in peak height. In addition, another weak substitution appeared that gave vectors consistent with the difference Patterson map of $\text{KAuCl}_4(\text{Au11})$.

The derivative containing KAuCl_4 with three substitution sites (their number matches the number of the sites of HgS and EHgTS to which their coordinates are similar) was included in the subsequent protein phase determination. The coordinates and occupancies of the atoms of the other derivatives were those determined by the first cycle of refinement. The addition of this fourth derivative increased $\langle m \rangle$ from 0.602 to 0.651 and $N(0.5)$ from 2491 to 2765.

To this stage the temperature factors for the heavy atoms had been set arbitrarily at $20\text{\AA}^2$ for the high occupancy sites and $30\text{\AA}^2$ for the low occupancy sites (based on the Wilson temperature factor of the protein). However, in the regions of small $\sin\theta$, the calculated average heavy atom contribution was comparable to the average value of $||F_{\text{PH}}| - |F_{\text{P}}||$, while in regions of high $\sin\theta$, the calculated heavy atom contribution was higher; the difference was especially large for the HgS derivative. A determination of protein phases, where the temperature factor of the high occupancy sites of HgS was increased to $32\text{\AA}^2$ and that of the low occupancy sites increased to $37\text{\AA}^2$, while the temperature factor of the high occupancy sites of EHgTs was increased to $25\text{\AA}^2$, on the one hand increased $\langle m \rangle$ to 0.660 and $N(0.5)$ to 2820 and on the other, made the distribution of the calculated structure factor of the heavy atoms and the observed difference $||F_{\text{PH}}| - |F_{\text{P}}||$ distribution comparable to all $\sin\theta$ regions.

The previous second cycle of simultaneous refinement of occupancies and coordinates of the heavy atoms showed the same rate of increase in the occupancies of the main sites of both HgS and EHgTS as the first cycle. In order to investigate if this tendency of the least square refinement was consistent with the double-difference Fourier map, the latter was calculated for all derivatives. These maps indeed verified the result of the second cycle of refinement. Moreover, additional low occupancy sites were found: one for HgS, two for EHgTS and two for Na_3IrCl_6 . A new protein phase determination with the four derivatives followed. The coordinates of the heavy atoms were those of the previous calculation and the occupancies those suggested by the double-difference Fourier. The average figure of merit $\langle m \rangle$ was 0.654 and $N(0.5)$ was 2785.

Three cycles of refinement on occupancies and coordinates of all atoms resulted in an overall improvement of the statistics to final values of $\langle m \rangle = 0.700$ and $N(0.5) = 3016$. However, although the residual factors of HgS, EHgTS and KAuCl_4 improved with the refinement, those of Na_3IrCl_6 deteriorated and became unacceptable. Moreover, the occupancies of some of the minor substitutions tended to decrease in every cycle toward an unobservable limit. Table 6 lists the occupancy changes and Table 7 lists the residual factors for each derivative before and after these three cycles of refinement. Substitutions of six electrons or less were removed from subsequent phase calculations and refinements. These are indicated by an asterisk in Table 6. The Na_3IrCl_6 derivative was also removed from calculations since it did not show acceptable behaviour with the refinement. The statistics of the protein phase calculation without the Na_3IrCl_6 derivative showed that it did not make much of a contribution to the phase determination ($\langle m \rangle$, 0.700→0.698; $N(0.5)$, 3016→2996).

TABLE 6. Occupancy Changes of Substitution Sites After Three Cycles of Refinement.

Site	<u>HgS</u>		<u>EHgTS</u>		<u>KAuCl₄</u>		<u>Na₃IrCl₆</u>	
	Initial	Final	Initial	Final	Initial	Final	Initial	Final
1	55.8	69.2	54.6	68.5	20.7	25.4	14.6	13.0
2	46.6	58.9	16.0	21.7			11.2	6.0
3	39.6	45.2	50.3	60.9			13.7	9.1
4	28.2	35.8	18.3	21.8			15.1	9.3
5	20.5	15.8	8.4	8.4			12.8	5.4
6	13.6	11.5	21.3	24.8	25.6	23.6	10.2	9.2
7	7.0	6.1*	11.6	11.8				
8	6.9	4.2*	7.0	6.1*				
9	8.5	12.1	8.5	11.3				
10			8.5	7.2				
11					8.5	6.4		

Substitutions sites 8 and 9 are different for HgS and EHgTS. For Na₃IrCl₆ sites 1 through 6 do not have the same coordinates as sites 1 through 6 of HgS and EHgTS.

TABLE 7. Change in the Residual Factors After Three Cycles of Refinement.

<u>Derivative</u>	<u>Initial R Factors</u>			<u>Final R Factors</u>		
	<u>R_{MODULUS}</u>	<u>R_{WEIGHTED}</u>	<u>R_Δ</u>	<u>R_{MODULUS}</u>	<u>R_{WEIGHTED}</u>	<u>R_Δ</u>
HgS	29.0	12.1	34.5	22.3	6.5	30.7
EHgTS	42.1	23.0	47.3	29.7	10.8	38.0
AuCl ₄	78.8	74.9	72.9	82.2	76.3	71.6
Na ₃ IrCl ₆	86.3	93.3	68.9	121.2	187.2	79.7

Let $|F_p|$ = native protein structure amplitude
 $|F_{PH}|$ = heavy atom derivative structure amplitude
 $|F_H|$ = calculated heavy atom structure amplitude
 w_{hkl} = weighting factor used during refinement
 $= 1/E^2$, where E is the lack of closure error

$$\begin{aligned} \text{then } R_{\text{MODULUS}} &= \frac{\sum_{hkl} ||F_{PH}| - |\vec{F}_p + \vec{F}_H||}{\sum_{hkl} |F_H|} \\ R_{\text{WEIGHTED}} &= \frac{\sum_{hkl} w_{hkl} (|F_{PH}| - |\vec{F}_p + \vec{F}_H|)^2}{\sum_{hkl} w_{hkl} |F_H|^2} \\ R_{\Delta} &= \frac{\sum_{hkl} ||F_{PH}| - |\vec{F}_p + \vec{F}_H||}{\sum_{hkl} ||F_{PH}| - |F_p||} \end{aligned}$$

The temperature factors of the major substitutions of the mercury derivatives were refined during the next two cycles of refinement. The changes were small and by the second cycle, the refinement had converged. The values of $\langle m \rangle$ and $N(0.5)$ increased to .710 and 3025, respectively. Three more cycles of least square refinement on occupancies and coordinates showed that the refinement had also converged with respect to these parameters ($\langle m \rangle = 0.710$ and $N(0.5) = 3038$). Table 8 lists the final parameters of all heavy atom sites used for the protein phase determination.

3. Determination of the Absolute Configuration

A right handed system of coordinates had been used for indexing the reflections during intensity data collection. The interpretation of the difference Patterson is, in general, consistent with either of two centrosymmetrically related sets of heavy atom coordinates, but only one of which is consistent with the indexing system adopted. A number of methods are used in order to resolve the ambiguity^{40,58-61}. In the present case, the simplest method was used⁶¹: after the refinement had been completed, the anomalous dispersion contributions of HgS and EHgTS were included in the calculation. The mean figure of merit for all 3853 reflections was 0.715 and $N(0.5) = 3063$. On interchanging the roles of $F_{PH}(hkl)$ and $F_{PH}(\bar{h}\bar{k}\bar{l})$ in the calculation of the anomalous dispersion contribution, $\langle m \rangle$ increased to 0.720 and $N(0.5)$ to 3082, thus suggesting that the latter correspond to the correct absolute configuration. Since the anomalous dispersion measurements extend only to about $6\text{\AA}$ resolution, the mean figure of merit for all reflections is not a very good measure of the improvement. Upon examination of the individual regions of minimum zone spacing 20, 10 and $6.67\text{\AA}$ with 27, 95 and 476 reflections,

TABLE 8. Heavy Atom Parameters

Compound	Site	Position in Fractional Coordinates			Occupancy in Electrons	B* in Å ²
		x	y	z		
HgS	1	0.1204	0.1453	0.8743	72.1	32.5*
	2	0.0022	0.1798	0.8872	62.0	29.5*
	3	0.6502	0.9228	0.6631	46.0	35.7*
	4	0.0268	0.1800	0.1220	39.3	34.1*
	5	0.1719	0.1235	0.6570	15.4	37.0
	6	0.7760	0.7760	0.7760	10.7	37.0
	9	0.0457	0.2001	0.8807	9.8	37.0
EHgTS	1	0.1208	0.1432	0.8727	69.6	21.4*
	2	-0.0016	0.1726	0.8802	22.3	21.4*
	3	0.6472	0.9204	0.6628	60.7	20.9*
	4	0.0299	0.1797	0.1197	23.4	24.1*
	5	0.1712	0.1136	0.0727	5.0	30.0
	6	0.7778	0.7778	0.7778	24.9	39.4*
	7	0.1953	0.0255	0.5764	8.5	30.0
	10	0.0898	0.4746	0.3127	10.3	30.0
	11	0.1651	0.1244	0.9367	7.7	30.0
KAuCl ₄	1	0.1133	0.1428	0.8704	30.2	45.0
	6	0.7667	0.7667	0.7667	24.2	45.0
	11	0.3768	0.1489	0.8395	5.5	45.0

* Temperature Factor = $\exp(-B \sin^2 \theta / \lambda^2)$

respectively, the first configuration gives $\langle m \rangle$ of 0.881, 0.856, 0.801 for each of the above regions, while the second configuration gives 0.909, 0.875 and 0.832. Therefore, the most probable assumption is that the centrosymmetric coordinates of those originally used give the correct absolute configuration. Figure 12 shows: (a) the difference in the figure of merit distribution when the correct and incorrect assumptions have been made in applying the anomalous dispersion observations in the protein phase determination as an additional isomorph in the multiple isomorphous replacement method and (b) the effect of the anomalous contribution and the correct configuration upon the figure of merit.

4. Summary and Error Analysis

The refinement of the protein phases was terminated at that point, since better phases cannot be obtained with the present heavy atom derivatives. An analysis of the refinement after anomalous dispersion had been applied correctly is summarized in Table 9. The ratio κ , which is an experimental estimate of $\Delta f''/(f + \Delta f')$, where $\Delta f'$ and $\Delta f''$ are the real and imaginary parts of the anomalous scattering components of the scattering factor f of the heavy atoms, is close to the expected theoretical value (0.12 for the region of $\sin\theta$ $0 < \sin\theta < 0.05$). Except for the KAuCl_4 derivative, the mean scattering contribution of the heavy atoms ($\text{r.m.s.}|F_H|$) is considerably greater than the r.m.s. closure error E . This can be seen over the scattering range in Figure 13.

The figure of merit distribution as a function of $\sin\theta$ (Figure 12a) and the distribution of the reflections with respect to figure of merit are listed in Table 10. A histogram of the latter is shown in Figure 14.

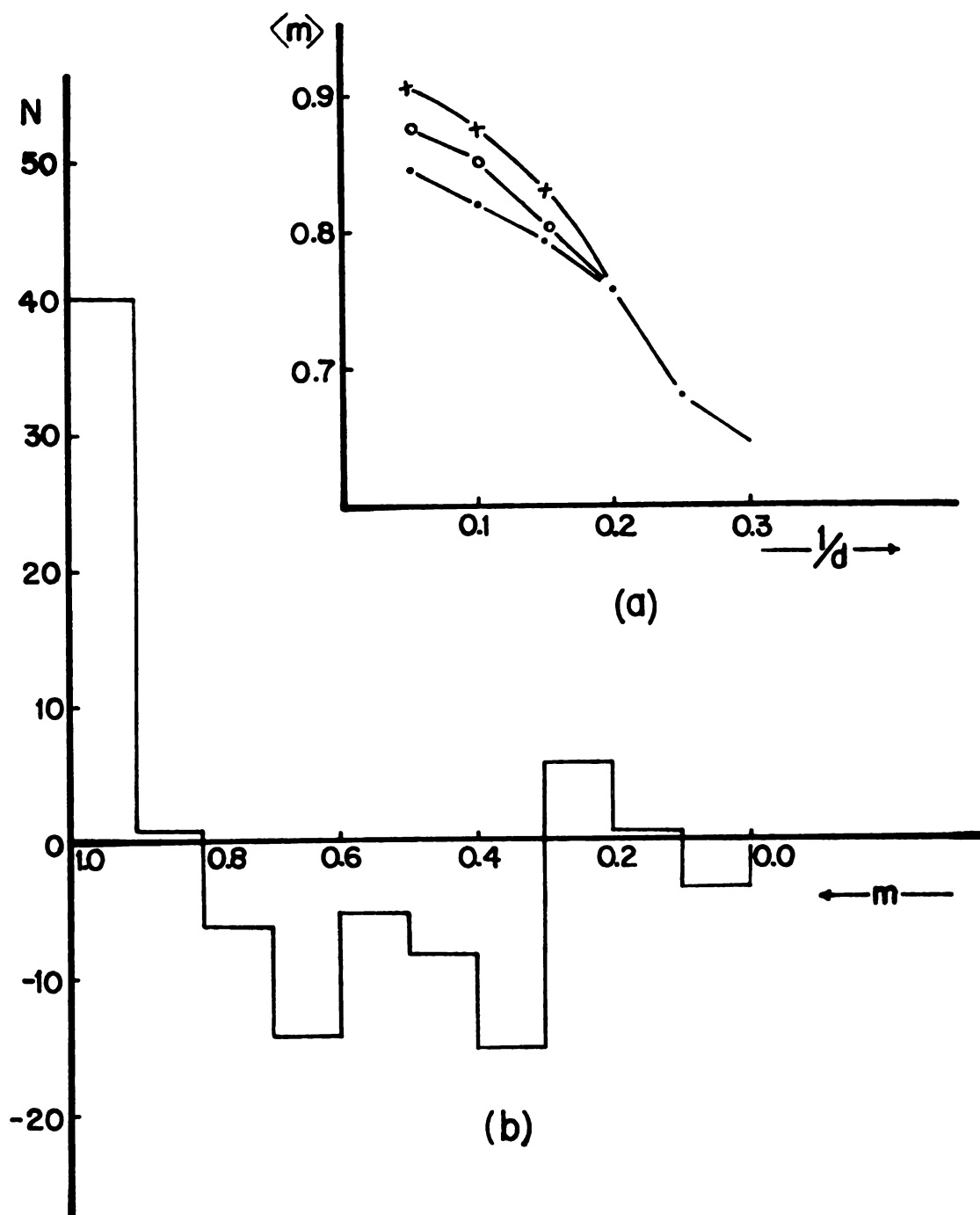


Figure 12. (a) Figure of merit: $\cdot$ — no anomalous dispersion (AD) contribution; $\circ$ — with AD but incorrect hand; $\times$ — with AD correct hand. (b) Difference in number of reflections with $\langle m \rangle$ based on correct and incorrect assumption of the absolute configuration.

TABLE 9. Refinement Error Analysis

Resolution	20Å	10Å	6.67Å	5Å	4Å	3.33Å	Overall
HgS							
κ	0.14	0.11	0.09				
r.m.s. closure E	80	66	60	50	64	64	
r.m.s. $ F_H $	506	397	313	266	223	183	
r.m.s. differences	431	296	216	160	158	131	
A.D. residuals	87	90	92				
R_{MODULUS}	.155	.146	.160	.147	.220	.287	.197
R_{WEIGHTED}	.022	.024	.031	.030	.070	.104	.049
R_{Δ}	.183	.186	.220	.245	.313	.404	.256
number per zone, n	26	193	473	887	1333	889	
EHgTS							
κ	0.11	0.08	0.06				
r.m.s. closure E	102	108	85	64	82	78	
r.m.s. $ F_H $	448	362	280	248	227	195	
r.m.s. differences	354	284	200	165	172	157	

TABLE 9. (continued)

Resolution	20Å	10Å	6.67Å	5Å	4Å	3.33Å	Overall
A.D. residuals	.92	.93	.94				
R_{MODULUS}	.219	.259	.269	.217	.304	.335	.279
R_{WEIGHTED}	.044	.077	.079	.058	.113	.139	.092
R_{Δ}	.278	.323	.368	.325	.405	.416	.374
number per zone, n	26	193	467	858	1321	828	
KAuCl₄							
r.m.s. closure E	0.0	71	65	75			
r.m.s. $ F_H $	0.0	112	93	74			
r.m.s. differences	0.0	127	92	85			
R_{MODULUS}	0.0	.617	.644	.919			.774
R_{WEIGHTED}	0.0	.388	.474	.997			.669
R_{Δ}	0.0	.547	.650	.794			.708
number per zone, n	0	193	476	822			

TABLE 9. (continued)

Let	n	=	number of reflections
	$ F_p $	=	native protein structure amplitude
	$ F_{PH} $	=	heavy atom derivative structure amplitude
	$ F_H $	=	calculated heavy atom structure amplitude
	$ \Delta^\pm _O, \Delta^\pm _C$	=	observed and calculated Bijvoet differences: $ F(+) - F(-) $
	r.m.s. closure E	=	$\sqrt{\frac{1}{n} \sum_{hk1} (F_{PH} - \vec{F}_p + \vec{F}_H)^2 / n}$
	r.m.s. $ F_H $	=	$\sqrt{\frac{1}{n} \sum_{hk1} F_H ^2 / n}$
	r.m.s. differences	=	$\sqrt{\frac{1}{n} \sum_{hk1} (F_{PH} - F_p)^2 / n}$
	κ	=	$ \Delta^\pm _O F_{PH} / 2 F_p F_H \sin(\alpha_p - \alpha_H)$
	A.D. residual	=	$\sqrt{\frac{1}{n} \sum_{hk1} (\Delta^\pm _C - \Delta^\pm _O)^2 / \sum_{hk1} \Delta^\pm _C^2}$

R_{MODULUS} , R_{WEIGHTED} and R_Δ have been defined in Table 7.

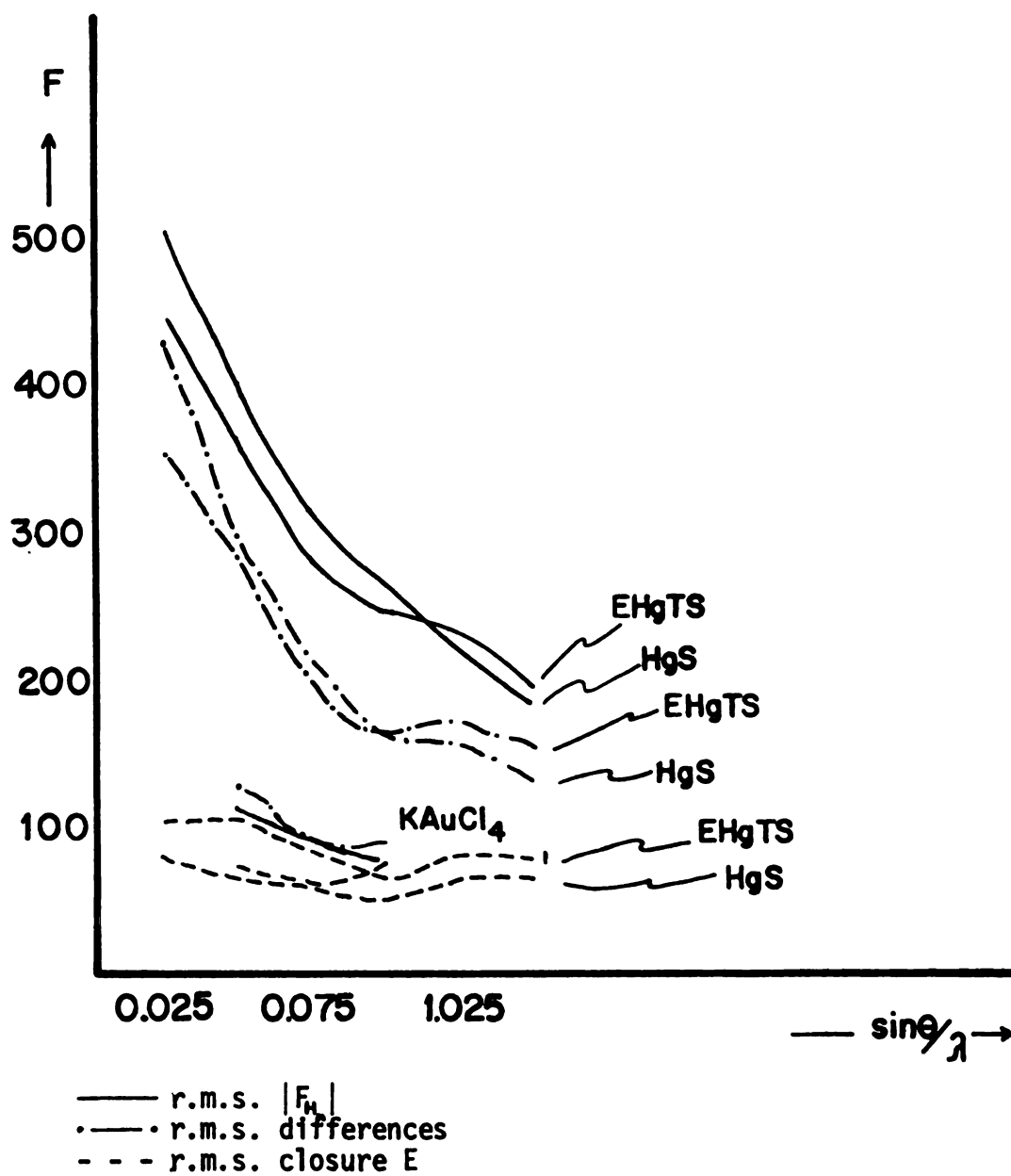


Figure 13. Distribution of r.m.s. heavy atom structure amplitude, r.m.s. differences and lack of closure as a function of $\sin\theta/\lambda$.

TABLE 10. Figure of Merit Analysis
(a) analysis of $\langle m \rangle$ with respect to resolution

Resolution	20Å	10Å	6.67Å	5Å	4Å	3.33Å	Overall
$\langle m \rangle$	.909	.875	.832	.756	.681	.644	0.720
n	27	195	476	892	1363	900	3853

(b) distribution of reflections in different figure of merit ranges

Range of average $\langle m \rangle$	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95
n	47	104	162	211	247	342	357	494	724	1165

$\langle m \rangle$ mean figure of merit
n number of reflections

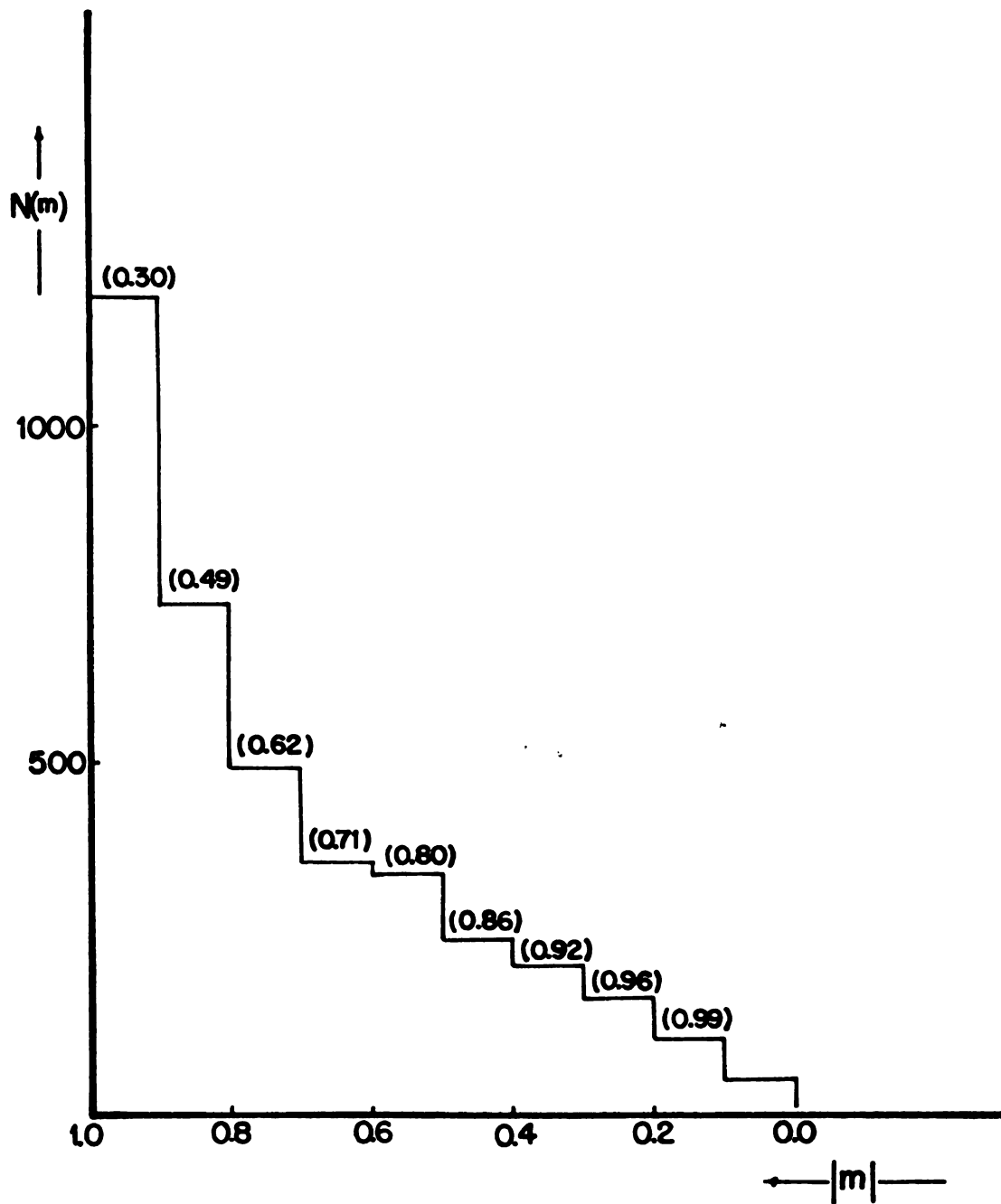


Figure 14. Distribution of figure of merit, m , among reflections; fraction of total number of reflections with m greater than given m , in parentheses.

IV. ELECTRON DENSITY OF KDPG ALDOLASE

1. Calculation of the Electron Density Map

The "best" electron density was calculated at a nominal resolution of 3.56Å. The Fourier coefficients used were the observed protein structure amplitudes converted to electrons by multiplying them by the Wilson scale factor, K_W , with the "best" phase angles, α_B , determined using the three isomorphous derivatives HgS, EHgTS and KAuCl_4 and the anomalous dispersion contribution of HgS and EHgTS. Only reflections with a figure of merit greater than 0.3 were included in the calculation and every structure factor was weighted by its figure of merit. It was mentioned in Chapter III that the coordinates of the heavy atom derivatives which give the correct enantiomorph are centrosymmetric to the ones shown in Table 8. Since the coordinates of the heavy atom derivatives were not inverted during the final calculation of phases, the image of the protein which corresponds to the L enantiomer is the mirror image of the one calculated. A theoretical value of $F(0,0,0)$ based on the amino acid composition was included in the summation as: $F(0,0,0)/V = 0.127 \text{ e}/\text{\AA}^3$.

The mean square error in the electron density of the "best" Fourier⁴⁴ of KDPG aldolase, according to equation (2), is given by the following expression in a cubic crystal system:

$$\langle \rho^2 \rangle = (24/V) \sum_{hkl}^N |F_p(h,k,l)|^2 (1 - m^2(h,k,l)),$$

where N is the number of the unique reflections used in the summation (3450 terms) and the factor 24 is required to include all the reflections,

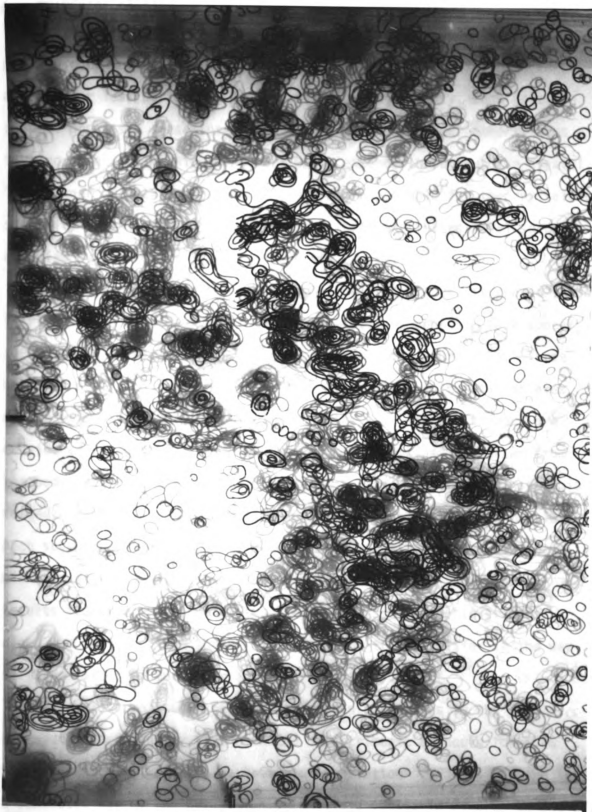
since only 1/24 of the limiting sphere is unique. The mean square error was calculated to be $\langle \rho^2 \rangle = 0.00802 \text{ e}^2/\text{\AA}^6$. Therefore, the error of the electron density map was:

$$\sigma(\rho) = (\langle \rho^2 \rangle)^{1/2} = \pm 0.088 \text{ e}/\text{\AA}^3.$$

The electron density was calculated in planes perpendicular to the $\vec{a}$ crystal axis at intervals $a/90$, $b/90$ and $c/90$, which corresponds to $1.149\text{\AA}$ between grid points. Contours of the electron density were drawn starting at $0.3 \text{ e}/\text{\AA}^3$ (e.g. greater than $3\sigma(\rho) = 0.26 \text{ e}/\text{\AA}^3$) at equal intervals of $0.1\text{e}/\text{\AA}^3$. The largest value of the electron density was $0.74 \text{ e}/\text{\AA}^3$ (about $8.5\sigma(\rho)$). The contours were traced onto plexiglass sheets which in turn were staked at proper distances to give a three dimensional presentation for inspection.

The electron density of KDPG aldolase in projection along the $\vec{a}$ axis is shown in Figure 15. The $\vec{b}$ and $\vec{c}$ axes are indicated from 0 to 1, while the depth of the density is more than one third of the length of the $\vec{a}$ axis ($42\text{\AA}$). The general view of the electron density is consistent with certain facts previously known about the crystals. The large amount of space with a mean electron density less than $0.3 \text{ e}/\text{\AA}^3$ has been interpreted as representing regions of mother liquid which had been independently determined to be 63% of the mass of the crystal from variable equilibrium density measurements⁴. Regions of electron density of more than $0.4\text{e}/\text{\AA}^3$ form well defined and clearly connected peaks which have been interpreted as representing the image of the protein molecules in the crystal.

Figure 15. "Best" electron density map of KDPG aldolase at 3.56Å[•] resolution in projection down the $\hat{a}$ axis.



2. Molecular Packing

It can be seen from Figure 15 that the electron density of the protein does not possess isolated regions of one or more subunits which are more or less surrounded by solvent. To the contrary, each protein subunit possesses close intermolecular contacts with four others, thus forming a three dimensional network which extends throughout the crystal. This manner of packing presumably might be responsible for the crystalline stability of the system despite the fact that huge interstitial spaces occupied by the mother liquor pervade the crystal.

Even though there are many intermolecular contacts, it proved to be fairly easy to distinguish the monomeric protein subunits. They appear to be irregular ellipsoids of approximate dimensions $40 \times 40 \times 25 \text{ \AA}$. Three subunits are situated around each of the four three-fold rotation axes of the unit cell forming four trimeric molecules. If the trimers are considered as points at their centers of mass and lines are drawn from the body center of the unit cell to these points, a tetrahedral array is formed. However, these are not the only trimeric arrangements possible. Each subunit of the foregoing trimers makes additional close contacts with two other subunits, each from a different trimer. The latter results from the existence of two kinds of crystallographic trimers along the length of the three-fold axis. The trimers differ in the way different sides of their surface interact.

The close molecular packing in certain directions in the crystal introduces an ambiguity in the choice of a trimeric molecule for KDPG aldolase which is known to be trimeric in solution²⁻⁴. This is shown in a schematic way by two projections, $(hk0)$ and $(0kl)$, in Figures 16 and 17, respectively. The subunits are represented as spheres scaled to an approximate diameter of $35 \text{ \AA}$. The coordinates of the center of every

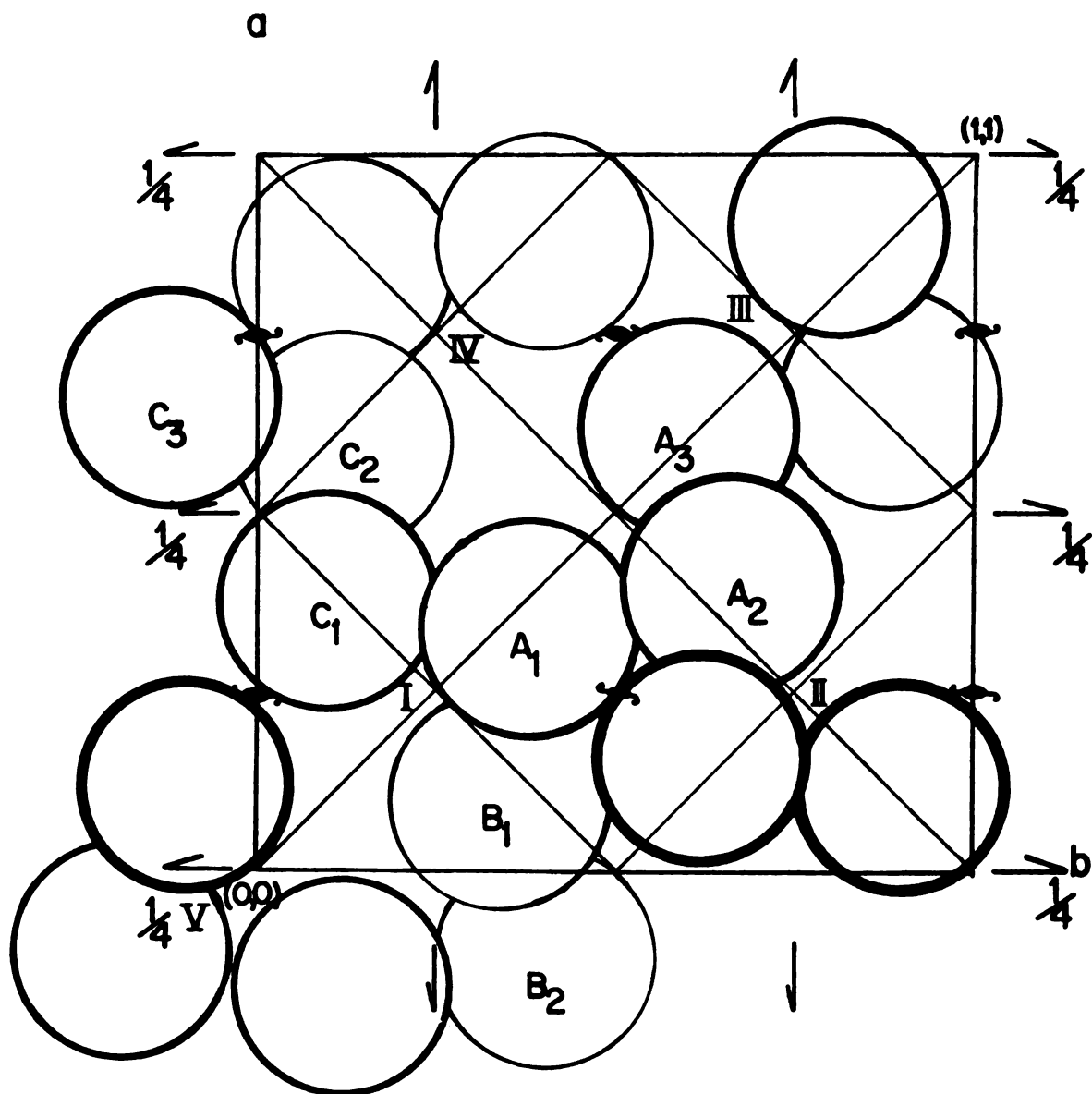


Figure 16. Schematic $(hk0)$ projection of the arrangement of the molecules of KDPG aldolase in the unit cell of space group $P2_13$.

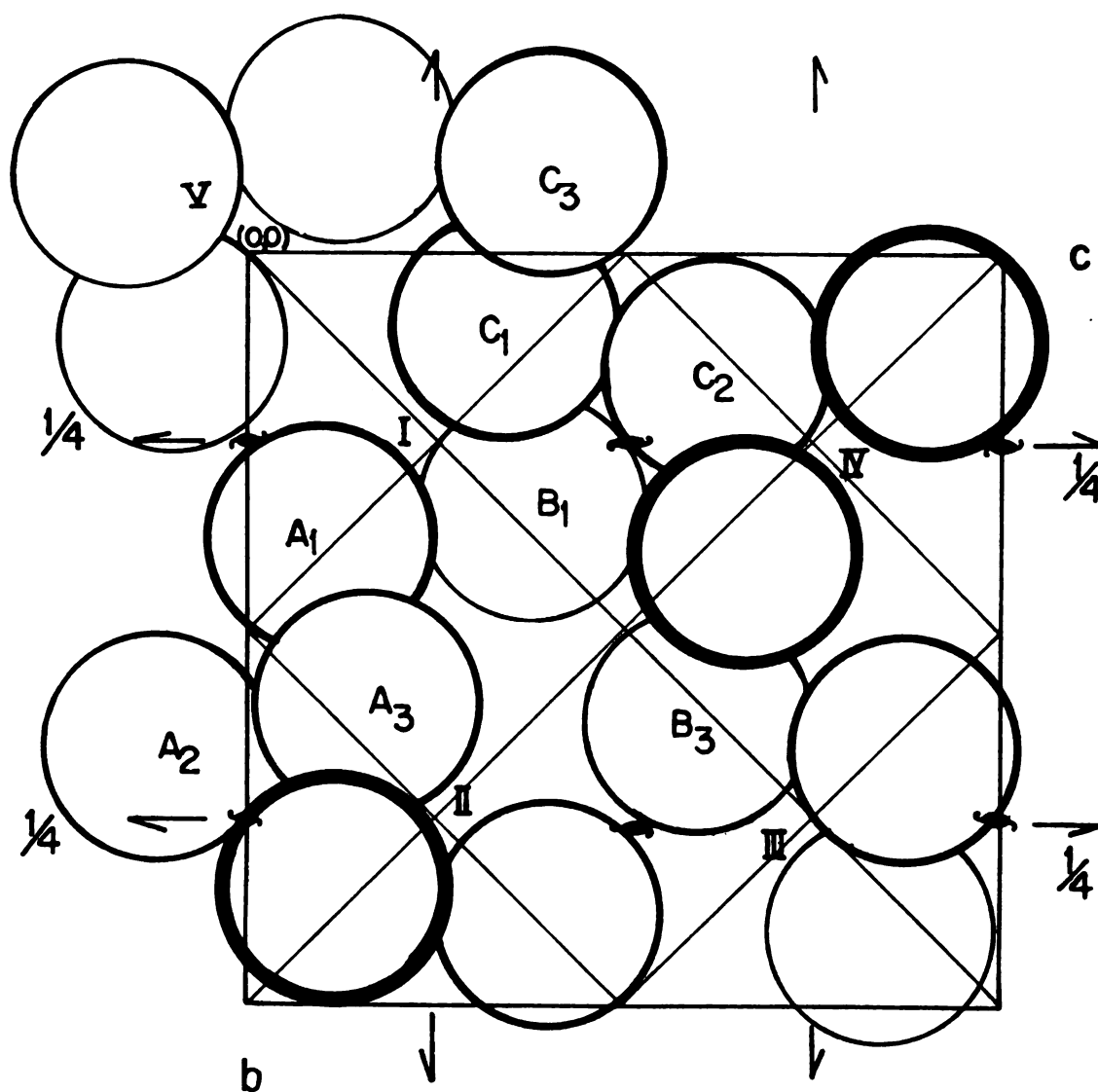
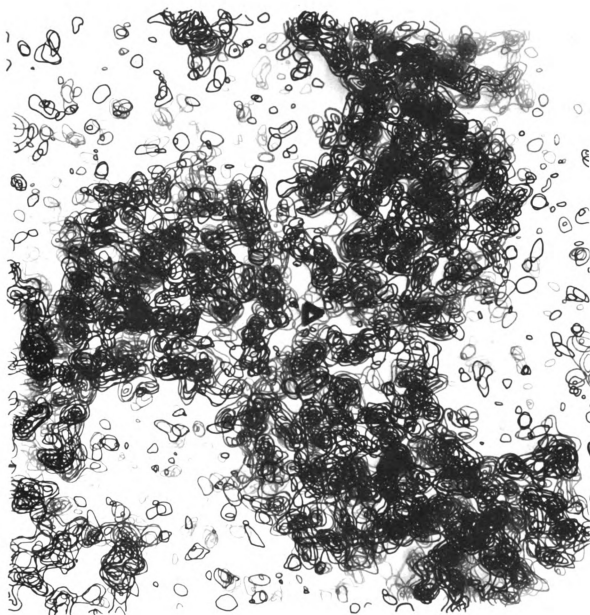


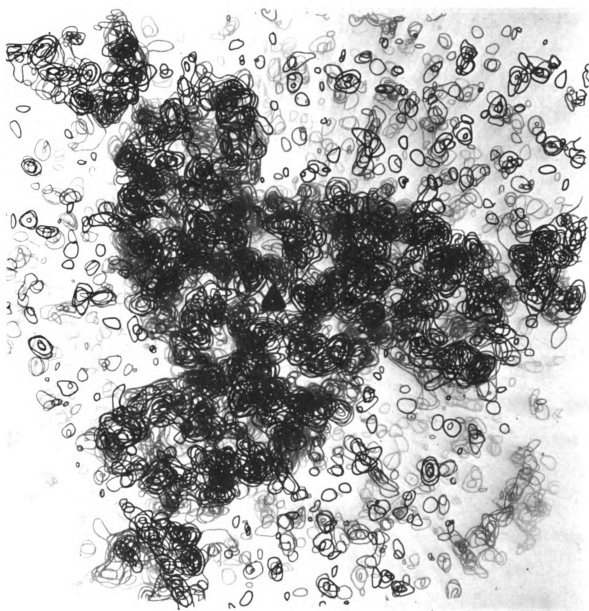
Figure 17. Schematic (0kl) projection of the arrangement of the molecules of KDPG aldolase in the unit cell of space group $P2_13$.

sphere are those of the approximate center of mass of the protein subunit. The thickness of the circles representing the spheres indicates the height of the molecule perpendicular to the projection. The trimers indicated by Roman numerals (denoted hereafter as of the first kind of trimers) are those oriented tetrahedrally near the corners of the unit cell. The diagonal lines represent the projections of the three-fold axes on the *ab* and *bc* planes of the crystal. The other kind of trimers (denoted as of the second kind) is indicated by the letters A_j , B_j , and C_j , where $j = 1, 2, 3$. It is clear from Figures 16 and 17 that subunit A_1 can be a part of trimer I or of trimer $A_1A_2A_3$ formed by the close contact of three subunits of the trimers I, II and III. The same considerations hold for all other subunits. This arises because of the two possible ways of arranging the subunits along each of the three-fold axes e.g. trimers I and V around the three-fold axis passing through (0,0,0) and (1,1,1). The crystallographic relation between these different trimers is such that the subunits of the first kind of trimer are related to the subunits of the second kind by the three different non intersecting two-fold screw axes which are projected in Figures 16 and 17. In contrast, trimers of both the first or second kind are related to themselves by only one of the three two-fold axes.

It is impossible to resolve the foregoing ambiguity from symmetry considerations alone. However, a study of the interactions of the subunits in each of the two kinds of trimer tends to break the ambiguity in a definite way. The general indication of the electron density map is that the trimer of the second kind is probably the one that exists in solution: it shows more and closer interactions between the subunits of trimers of the second kind than trimers of the first kind as it is shown in Figure 18.

Figure 18. "Best" electron density map in projection down the three fold axis; (a) the trimer of the first kind, (b) the trimer of the second kind.





3. Molecular Structure

It is difficult to give any detailed account of the molecular structure of a protein based on an electron density map with a resolution much less than atomic. In the present case, this is compounded by the fact that the electron density was calculated only one month ago. There are proteins, whose structures are determined at higher resolution, that have been studied for many years but there are still ambiguities associated with some details of their electron density. The account of the structure to be described here is very general and will be confirmed or otherwise established many times in the future when improved, higher resolution electron density maps based on more isomorphous derivatives will be available and when the amino acid sequence of the protein will be known.

The sequence of KDPG aldolase has not yet been completely determined. Of about 227 amino acid residues per subunit, the sequence of 50 consecutive residues, shown in Figure 19, containing the active lysine has been established from the sequence and the correlation of smaller polypeptides⁶². The amino acid composition of the trimer of KDPG aldolase is listed in Table 11³. The composition can be helpful in the initial interpretation of the electron density. The fact that the protein has many amino acid residues with long side chains like glutamic acid, leucine, isoleucine, lysine and arginine can be of some assistance when one establishes its folding. Another striking feature of the protein composition is that it contains many proline residues (about sixteen per subunit). Proline, the only amino acid in which the side chain condenses with the main chain, has the property of forcing a bend in a main chain and of disrupting an α -helix⁶³. Not all bends in proteins contain proline but every proline bends the chain. A final point

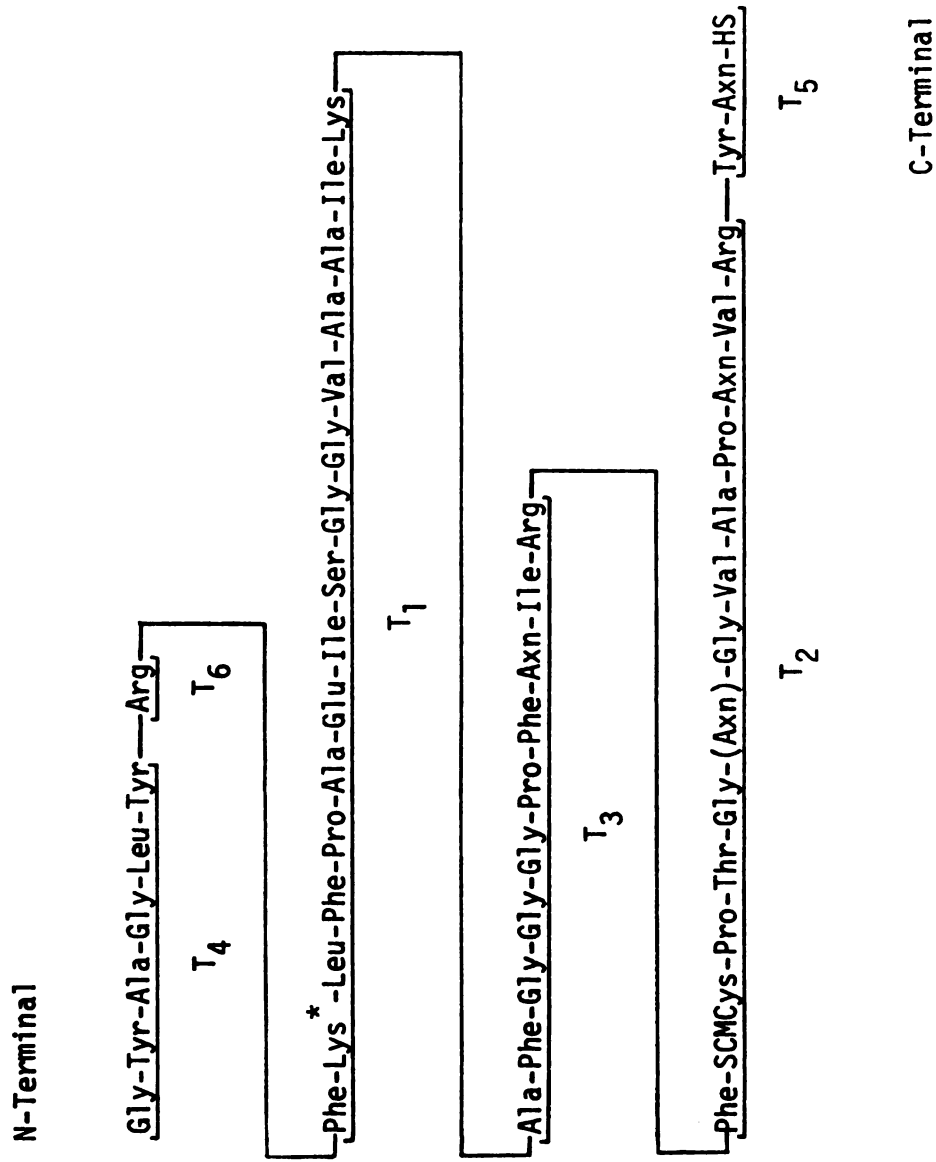


Figure 19. The sequence of the known polypeptide containing the active lysine of KDPG aldolase.

TABLE 11. Amino Acid Analysis^a of the Trimer of KDPG Aldolase
(Reference 3).

Amino Acid	Residues per mole
Alanine	92
Arginine	45
Aspartic Acid	50
Cysteine ^c (Cm ^b)	12
Glutamic Acid	60
Glycine	66
Histidine	3
Isoleucine	55
Leucine	65
Lysine	21
Methionine	20
Phenylalanine	22
Proline	47
Serine ^d	25
Threonine ^d	31
Tryptophane ^e	11
Tyrosine	9
Valine	44
NH ₃ ^f	10

- a - obtained from the average of duplicate analyses of 0.5mg each of KDPG aldolase hydrolyzed for 24 and 70 hours
b - Cm, carboxymethyl. Determined from a carboxymethylated preparation of KDPG aldolase.
c - calculated from 24-hour analysis
d - extrapolated to zero hour hydrolysis.
e - determined by spectral methods
f - ten amide residues were assumed.

worthy of note is the large alanine (30) and glycine (20) content per subunit.

Many methods, of varying degrees of accuracy, have been proposed in the past to determine the conformation of proteins from their amino acid composition and/or sequence. In the beginning, there was the simple classification that some amino acids were helix breakers⁶⁴. Later, methods were developed which rapidly became complicated and which attempted to classify all amino acids in terms of their ability to form or break helices^{65,66}. Their accuracy, investigated with a number of proteins of known three dimensional structure, has been tested with a predictability success of 60-70%. Recently, a more complete model has been proposed^{67,68} which classifies the amino acids not only in their ability to form and break helices but also, β -sheet structures. It has been successful in locating 88% of the helical and 95% of β -sheet from known sequence, as well as correctly predicting 80% of the helical and 86% of the β -sheet forming residues, in 19 proteins of known three dimensional structure. The clearest result of the method is that the amino acid residues of glutamic acid, alanine and leucine are classified as strong helix formers. This means that if a part of a protein sequence is mainly composed of these residues, it should have a helical conformation. In the same context, the amino acid residues of methionine, valine and isoleucine are strong β -sheet formers and residues which terminate α -helix most strongly are proline and glycine; the one that terminates β -sheet is glutamic acid.

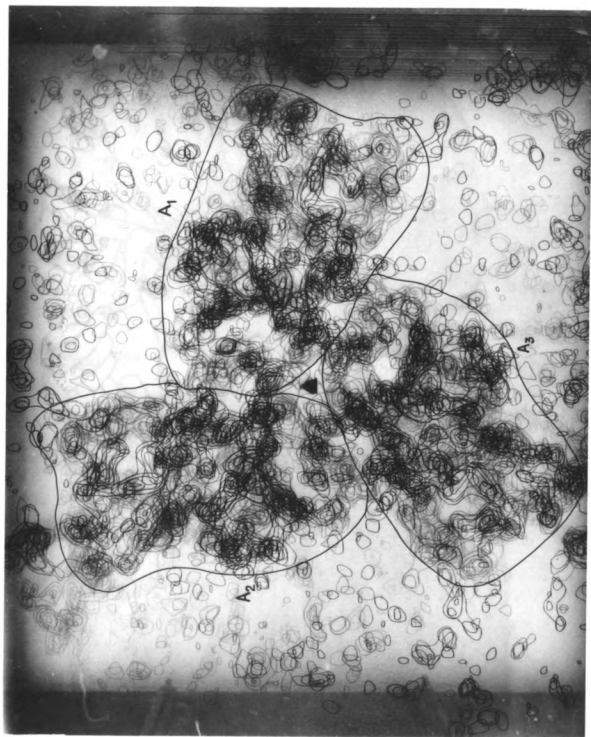
KDPG aldolase contains many of the strong helix former residues and there is a strong probability that they will cluster together to form α -helices. This is consistent with the results of optical rotatory dispersion of the protein which suggests a helical content of 31 - 33%. The presence of many proline and glycine residues will terminate helical

regions and will also impose many bends on the main chain. The presence of the many valine and isoleucine residues suggests the possibility of extended β -sheet regions but the simultaneous presence of the glutamates, which are strong β -sheet breakers would seem to prevent their formation.

The electron density map indicates some close intermolecular interactions between the subunits, especially of the second kind. Nevertheless looking along the three fold axis it is easy to isolate the protein subunits as indicated in Figure 20. The absence of disulfide bridges³, and thus main chain branching, facilitated the tracing for the polypeptide chain of KDPG aldolase in the electron density map. On careful inspection of the map, it was possible to follow most of the backbone of the protein. In doing so, it became inevitable that assumptions had to be made as to which path to follow at certain points when the contours are not connected and there presented more than one possibility. The general appearance of the folding of KDPG aldolase is shown schematically in Figure 21. The most obvious feature of the folding is its numerous bends. The regions denoted as 1 and 2 in Figure 21 are parts of the chain where the interpretation of the electron density was ambiguous and the indicated folding is mostly based on the intuition of the author.

Seven cylindrical regions of electron density significantly higher than the average and with a diameter of about $6\text{\AA}$ have been interpreted as helices. They are indicated by the numbers 3-9 in Figure 21. Two of the largest helical regions are found in the terminal parts of the polypeptide chain similar to other proteins (amino-terminal lysozyme; carboxyl terminal α -chymotrypsin). Measuring the length of these regions and assuming that they are α -helices (3.61 residues per $5.41\text{\AA}$), the total number of residues involved in helix formation is estimated to be about 55 or 25% of the total number of amino acids. There are

Figure 20. "Best" electron density along the three fold axis. The trimer of the second kind is indicated. The individual monomers are outlined and denoted as A_1 , A_2 and A_3 .



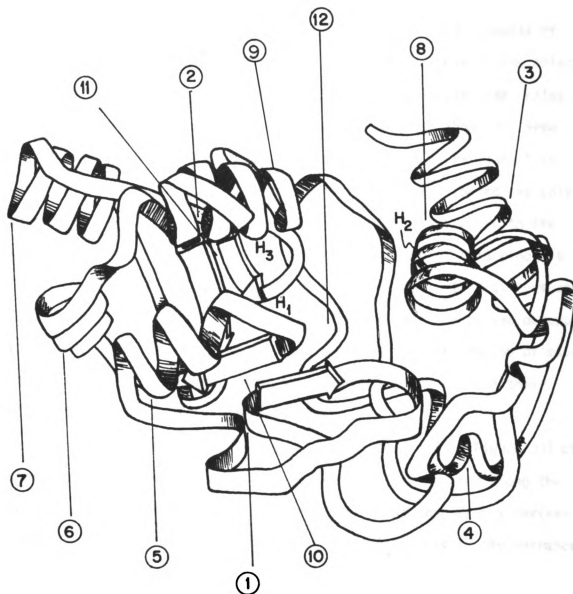


Figure 21. Schematic representation of folding of the KDPG aldolase monomer indicated as A₁ in Figure 20. View approximately perpendicular to the long direction of the central channel.

other short regions which might be helical but, since their peak heights are not especially high, they were not included in the foregoing assessment. The parts of the folding numbered 10 and 11 are the only regions which look like β -sheet structure; one is a parallel and one antiparallel.

It is easy to distinguish about 25 relatively long segments of electron density which obviously extend from the backbone of the molecule toward the interior of the molecule; they are most likely side chains of residues like leucine and isoleucine. There are also about 23 large segments which are probably large polar side chains of residues like glutamic acid, arginine and lysine because they extend toward the solvent surrounding the molecule. Both observations are consistent with the large number of long side chain residues per subunit of KDPG aldolase (about 20 glutamic acids, 18 isoleucines, 22 leucines, 15 arginines and 7 lysines). The smaller number observed in the map is expected because these long side groups of the molecule should have more freedom of movement and should be expected to be less ordered, hence causing their electron density to be, in general, lower.

Another feature of the molecule is a channel (12 in Figure 21) of approximately dimensions $9 \times 9 \times 30 \text{ \AA}$ passing through its center along the long direction. Three heavy atom binding sites of the mercury derivatives are located in this channel. Moreover, site 2 is at the entrance site 3 further inside and site 1 is in the middle; the heavy atom sites are indicated as H2, H3 and H1, respectively in Figure 21. Compounds like HgS and EHgTS have been reported to bind at sulfhydryl groups. The subunit of KDPG aldolase contains four cysteine residues, two readily accessible and two buried. Therefore, the former may well be close to heavy atom sites 2 and 3, whereas H1 might be one of the buried cysteines. The antiparallel β -sheet structure is parallel to the axis of the channel.

4. Concluding Remarks

The determination of the structure of KDPG aldolase at a nominal resolution of $3.56\text{\AA}$ was based on phases determined using two mercury derivatives with similar substitution sites but different occupancies and intensity data extending over the whole resolution range, and one gold derivative with intensity data extending only to $5.1\text{\AA}$ resolution. The use of anomalous dispersion data of the mercury derivatives applied after the completion of the refinement of the phases improved considerably the statistics of the refinement in the resolution range where they contributed (about $6\text{\AA}$). The "best" electron density map calculated with these phases is of good quality; it is free of density in the solvent regions and the molecular boundaries are simple to establish except for some close contacts between the subunits (Figures 15 and 20). The electron density of a crystallographic asymmetric unit can be accounted for by a single continuous chain with only four small gaps of about $2\text{\AA}$ each.

At this stage, the location of the active site is yet unknown and no chemical evidence suggest if it is in the surface or near the center of the trimeric molecule. Since the substrates of the enzyme are hydrophilic, it may very well be exposed to the solvent or it is possible to be in a hydrophilic cavity. In the case of KDPG aldolase, there are three empty regions inside the molecule, the central channel and two cavities on each side of it, close to the surface of the subunit. However, it is not possible to attribute any significance to these features, or any others which might have gone unnoticed with respect to the biological function and activity of the enzyme. The first step toward this end will be the location of the active site region in the electron density of the protein from a derivative study of an enzyme-substrate

complex, where the substrate will be found by difference Fourier methods. The three dimensional arrangement of the amino acid residues and their side chains in the active site might explain the propensity of KDPG aldolase to form Schiff-bases with numerous carbonyl compounds, while it only cleaves specifically, 2-keto-3-deoxy-6-phosphogluconate. In particular, it will be possible to determine the nature of the second base which has been postulated to play an important role in the cleavage and whether one or two side chains can play that role²², if they are properly oriented around the active lysine residue. At the present resolution of the structure determination, it is not certain that the active site environment will be conclusively characterized, either because the electron density map is not well resolved at the region under study, or because the side chains of the amino acids surrounding the active site are poorly formed or even missing.

The completion of the amino acid sequence and the determination of the structure at higher resolution should permit more detailed studies of the interactions of substrates with the enzyme at pH 3.6, the pH of the solution where the structure was determined. Since the optimum pH for the enzymatic activity is 7.5, it must be determined, using difference Fourier methods^{69,70}, whether its conformation changes by changing the pH. Moreover, the complete characterization of the tertiary structure of the enzyme, particularly at the active site region, will permit the comparison of the folding to that of other carbohydrate binding proteins, since it has been established that certain structural arrangements are preserved in a variety of proteins that act on similar substrates⁷¹.

Preliminary results of enzyme substrate complexes where pyruvate or 2-keto-3-deoxy-6-phosphogalactonate have been used as substrates, in

order to locate the active site, are not encouraging. The diffraction pattern along the directions (hhh), (hh0) and (h00) did not show any changes. In addition, the crystals are not stable for more than fifteen days. When the complex is reduced with sodium borohydride, the crystals become very fragile and they do not diffract X-rays. A possible explanation of this behaviour is that the active site is partially blocked in the crystalline state and the enzyme does not react readily so that the diffraction pattern remains unchanged. However, upon prolonged soaking or by shifting the equilibrium of the reaction to the right by reduction, the formation of the complex takes place accompanied by conformational changes of the enzyme incompatible with the lattice that force the crystal to shatter as in the case of lactate dehydrogenase⁷². In that case, the enzyme-substrate complex should crystallize in a different space group and it will be necessary to determine its structure independently.

REFERENCES

1. J.C. Kendrew, Science 139, 1259 (1963); M.F. Perutz, Science 140, 863 (1963).
2. R.H. Hammerstedt, H. Möhler, K.A. Decker, and W.A. Wood, J. Biol. Chem. 246, 2069 (1971).
3. D.C. Robertson, R.H. Hammerstedt, and W.A. Wood, J. Biol. Chem. 246, 2075 (1971).
4. R.L. Vandlen, D.L. Ersfeld, A. Tulinsky, and W.A. Wood, J. Biol. Chem. 248, 2251 (1973).
5. D.C. Robertson, W.W. Altekarr, and W.A. Wood, J. Biol. Chem. 246, 2084 (1971).
6. L.R. Barran, and W.A. Wood, J. Biol. Chem. 246, 4028 (1971).
7. H. Möhler, K.A. Decker, and W.A. Wood, Arch. Biochem. Biophys. 151, 251 (1972).
8. E. Grazi, H.P. Meloche, G. Martinez, W.A. Wood, and B.L. Horecker, Biochem. Biophys. Res. Commun. 10, 4 (1963).
9. H.P. Meloche, and W.A. Wood, J. Biol. Chem. 239, 3511 (1964).
10. H.P. Meloche, and W.A. Wood, J. Biol. Chem. 239, 3515 (1964).
11. J.M. Ingram, and W.A. Wood, J. Biol. Chem. 240, 4146 (1965).
12. J.M. Ingram, and W.A. Wood, J. Biol. Chem. 241, 3256 (1966).
13. H.P. Meloche, Biochem. Biophys. Res. Commun. 18, 277 (1965).
14. H.P. Meloche, Biochemistry 9, 5050 (1970).
15. I.A. Rose, and E.L. O'Connell, Arch. Biochem. Biophys. 118, 758 (1967).
16. H.P. Meloche, Biochemistry 6, 2273 (1967).

17. H.R. Mahler and E.H. Cordes, "Biological Chemistry", p. 423-426, Harper International Edition (1966).
18. W.A. Wood in "The Enzymes", 3rd ed., 7, 281 (1972).
19. E.W. Frampton, and W.A. Wood, J. Biol. Chem. 236, 2571 (1961).
20. R. Kovachevich and W.A. Wood, J. Biol. Chem. 213, 757 (1955).
21. M. A. Roseman, Ph.D. Thesis, Michigan State University (1970).
22. H.P. Meloche, J. Biol. Chem. 248, 6945 (1973).
23. A.J.C. Wilson, "Elements of X-ray Chrystallography", Addison-Wesley Publishing Co. (1970).
24. H. Lipson, and W. Cochran, "The Determination of Cyrstal Structures", G. Bell and Sons, Ltd., London (1953).
25. B.E. Warren, "X-ray Diffraction," Addison-Wesley Publishing Co. (1969).
26. J.M. Robertson, J. Chem. Soc. 1195 (1936); J.M. Robertson and J. Woodward, J. Chem. Soc. 219 (1937).
27. C. Bokhoven, J.C. Schoone, and J.M. Bijvoet, Acta Cryst. 4, 275 (1951).
28. D. Harker, Acta Cryst. 9, 1 (1956).
29. D.W. Green, V.M. Ingram, and M.F. Perutz, Proc. Roy. Soc. A225, 287 (1954).
30. W.L. Bragg, and M.F. Perutz, Proc. Roy. Soc. A225, 315 (1954).
31. M.M. Bluhm, G. Bodo, H.M. Dintzis, and J.C. Kendrew, Proc. Roy. Soc. A246, 369 (1958).
32. D.M. Blow, Proc. Roy. Soc. A247, 302 (1958).
33. G. Bodo, H.M. Dintzis, J.C. Kendrew, and H. Wyckoff, Proc. Roy. Soc. A253, 70 (1959).
34. D.C. Phillips in "Advances in Structure Research by Diffraction Methods" 2, 75 (1966).

35. R.E. Dickerson, J.C. Kendrew, and B.E. Strandberg in "Computing Methods and the Phase Problem in X-ray Crystal Analysis" p. 236-251, Oxford, Pergamon Press (1961).
36. M.G. Rossmann, Acta Cryst. 13, 221 (1960).
37. E. Dodson and M. Vijayan, Acta Cryst. B27, 2402 (1971).
38. W.L. Bragg, Acta Cryst. 11, 70 (1958).
39. M.F. Perutz, Acta Cryst. 2, 867 (1956).
40. G. Kartha and R. Parthasarathy, Acta Cryst. 18, 749 (1965).
41. L.K. Steinrauf, Acta Cryst. 16, 317 (1963).
42. J.M. Bijvoet, Nature Lond., 173, 888 (1954).
43. A.C.T. North, Acta Cryst. 18, 212 (1965).
44. D.M. Blow and F.H.C. Crick, Acta Cryst. 12, 794 (1959).
45. R.E. Dickerson, J.C. Kendrew, and B.E. Strandberg, Acta Cryst. 14, 1188, (1961).
46. R.G. Hart, Acta Cryst. 14, 1194 (1961).
47. C.C.F. Blake, R.H. Fenn, A.C.T. North, D.C. Phillips, and R.J. Poljak, Acta Cryst. 16, A77 (1963).
48. J. Kraut, L.C. Sieker, D.F. High, and S.T. Freer, 48, 1417 (1962).
49. M. Zeppezauer, H. Eklund, and E.S. Zeppezauer, Arch. Biochem. Biophys. 126, 564 (1968).
50. M.V. King, Acta Cryst. 7, 601 (1954).
51. H.W. Wyckoff, M. Doscher, D. Tsernoglou, T. Inagami, L.N. Johnson, K.D. Hardman, N.M. Allewel, D.M. Kelly and F.M. Richards, J. Mol. Biol. 27, 563 (1967).
52. R.L. Vandlen, and A. Tulinsky, Acta Cryst. B27, 437 (1971).

53. A.C.T. North, D.C. Phillips, and F.S. Mathews, Acta Cryst. A24, 351 (1968).
54. A.J.C. Wilson, Nature 150, 152 (1942).
55. G. Germain, P. Main, and M.M. Woolfson, Acta Cryst. A27, 368 (1971).
56. H. Hauptman, and J. Karle, "Solution of the Phase Problem I. The centrosymmetric Crystal", A.C.A. Monograph No. 3 Brooklyn: Polycrystal Book Service (1953).
57. "Least Squares Program to Refine Heavy Atom Parameters in an Isomorphous Series", written by M.G. Rossmann and modified by A. Wonacott, N. Adams, and G. Ford.
58. D.M. Blow, and M.G. Rossmann, Acta Cryst. 15, 1060 (1962).
59. B.W. Matthews, Acta Cryst. 20, 230 (1966).
60. J. Drenth, J.N. Jansonius, and B.G. Wolthers, J. Mol. Biol. 24, 449 (1967).
61. W.N. Lipscomb, J.C. Coppola, J.A. Hartsuck, M.L. Ludwig, H. Muirhead, J. Searl, and T.A. Steitz, J. Mol. Biol. 19, 423 (1966).
62. D.D. Tsay, Ph.D. Thesis, Michigan State University (1975).
63. R.E. Dickerson and I. Geis "The Structure and Action of Proteins", New York: Harper and Row (1969).
64. A.V. Guzzo, Biophys. J. 5, 809 (1965).
65. D. Kotelchuck, and H.A. Sheraga, Proc. Nat. Acad. Sci. U.S. 62, 14 (1969).
66. T.T. Wu, and E.A. Kabat, Proc. Nat. Acad. Sci. U.S. 68, 1501 (1971);
T.T. Wu, and E.A. Kabat, J. Mol. Biol. 75, 13 (1973).
67. P.Y. Chou, and G.D. Fasman, Biochemistry 13, 11 (1974).
68. P.Y. Chou, and G.D. Fasman, Biochemistry 13, 222 (1974).

69. R.L. Vandlen and A. Tulinsky, Biochemistry 12, 4983 (1973).
70. A. Mavridis, A. Tulinsky and M.N. Liebman, Biochemistry 13, 3661 (1974).
71. A. Liljas and M.G. Rossmann, Ann. Rev. Biochem. 43, 475 (1974).
72. M.G. Rossmann, M.J. Adams, M. Buehner, G.C. Ford, M.L. Hackert, P.J.Jr. Lentz, A. Jr. McPherson, R.W. Schevitz and I.E. Smiley, Cold Spring Harbor Symp. Quant. Biol. 36, 179, (1971).

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03145 2570