STRUCTURAL STUDIES OF OLIGOMERIC ENZYMES

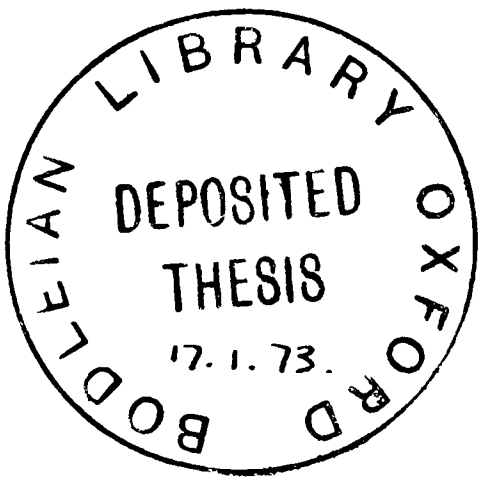
by

ANNE CHRISTINE BLOOMER

A thesis submitted for the degree of

Doctor of Philosophy

in the University of Oxford



Somerville College

September 1972

ABSTRACT

The functions of biological macromolecules are discussed in relation to their structure as linear polymers adopting specific conformations. If such molecules can be crystallised, their three dimensional structure can be determined at atomic resolution by the methods of X-ray crystallography whose applicability to the analysis of protein structures is considered. This is followed by an account of their application in the study of a glycolytic enzyme, triose phosphate isomerase (TIM), leading up to the calculation of an electron density map at 2.5 Å resolution.

The characteristics of crystals of chicken TIM are described and a detailed account is given of the measurement of X-ray diffracted intensities, the determination and refinement of the positions of heavy atom binding in isomorphous derivatives of TIM, and their subsequent use for calculating the phases of the protein structure factors. The search which was conducted for possible heavy atom derivatives produced several which have two sulphydryl sites in common. The problems resulting from these common sites and the systematic errors they can cause in the phase determination are examined in relation to the suitability of different methods of refinement in this situation.

The electron density map of TIM at 6Å Resolution is described and a preliminary dicussion of the 2.5 Å map is given.

TABLE OF CONTENTS

	Abstract	
	List of Figures	ii
	List of Tables	iv
	Acknowledgements	v
	Abbreviations and Notation	vi
Chapter I	Introduction	1
Chapter II	Preliminary Crystallographic Studies of Triose Phosphate Isomerase	45
Chapter III	Heavy Atom Isomorphous Derivatives	70
Chapter IV	Data Collection and Analysis at 2.5Å Resolution	106
Chapter V	Phase Determination and Further Refinement of Parameters for Isomorphous Derivatives	143
Chapter VI	Electron Density Maps of Chicken TIM	176
Appendix I	Computing Software	186
Appendix II	Relative Weighting of Coefficients in Fourier Syntheses	198
Appendix III	Derivative Scale Factor Determination Using Isomorphous and Anomalous Differences	208
	References	215

LIST OF FIGURES

1	Geometrical Conditions for Diffraction by a Crystal	11
2	Typical X-ray Spectrum	17
3	The Glycolytic Pathway	37
4	Reaction Catalysed by Triose Phosphate Isomerase (TIM)	38
5	Proposed Mechanism of the TIM-catalysed Reaction	38
6	Postulated Free Energy Profile for the TIM-catalysed Reaction	39
7	Diffraction Photographs of Chicken TIM	48
8	Molecular Weight of Chicken TIM Determined Crystallographically	52
9	Crystals of Chicken TIM	58
10	Symmetry Elements Present in Crystals of Chicken TIM	60
11	Two Dimensional Packing of Ellipsoids	61
12	Stereographic Projection of the Self Rotation Function	69
13	The Phase Triangle	73
14	Projections of 6Å Difference Patterson: MONO and DI	83
15	Projections and Harker Sections of 6Å Difference Patterson:PCMB	94
16	Projections of Best Sign ΔF and $\Delta\Delta F$ Maps: PCMB	95
17	Projections of 6Å Difference Patterson: EMP	97
18	Harker Sections of Difference Patterson and 2 site ΔF Map: EMP	98
19	Projections of 6Å Best Sign ΔF and $\Delta\Delta F$ Maps: EMP	99
20	Projections and Harker Sections of 6Å Difference Patterson:BAKER	100
21	Projections of 6Å Self Sign ΔF and $\Delta\Delta F$ Maps: BAKER	100
22	Projections of 6Å Best Sign ΔF and $\Delta\Delta F$ Maps: BAKER	101
23	Projections of 6Å Difference Patterson and Cross Sign Maps:$PtCl_4$	103
24	Projections of 6Å Best Sign ΔF and $\Delta\Delta F$ Maps: $PtCl_4$	104
25	Scheme for 2.5Å Intensity Measurement for Chicken TIM	107
26	Ratio of Friedel Equivalents for NATIVE Cl Levels of Reflections	115

27	Ratio of Friedel Equivalents for EMP Cl Levels of Reflections	115
28	Secondary Absorption and Radiation Damage Corrections	116
29	Scatter of Observations Within Data Sets	127
30	Mean Structure Amplitudes	129
31	Projections of 2.5Å Best Sign ΔF and $\Delta\Delta F$ Maps: MONO	134
32	Projections of 2.5Å Best Sign ΔF and $\Delta\Delta F$ Maps: EMP	134
33	Projections of 2.5Å Best Sign ΔF and $\Delta\Delta F$ Maps: BAKER	134
34	Wilson Plot of the Intensities for Native Crystals	137
35	Crick-Magdoff Plots of the Isomorphous Differences	138
36	Empirical Estimate of the Anomalous Scattering Ratio	140
37	Phase Triangle Illustrating Anomalous Scattering by Heavy Atoms	141
38	Phase Diagram Showing the Harker Construction	144
39	Phase Diagram Illustrating the Use of Anomalous Scattering Information	149
40	Effects of Errors in Heavy Atom Occupancy	158
41	Derivative Scale Factors Derived by the Method of Singh and Ramaseshan	163
42	Profiles of Atoms at Sulphydryl Sites in ΔF Maps	172
43	Phase Determination; Figures of Merit and Residual Errors	175
44	Electron Density Map at 6Å Resolution: A Complete Molecule	178
45	A Balsa Wood Model of a Molecule of Chicken TIM	179
46	Outlines of a Molecule of Chicken TIM	181
47	Molecular Packing in Crystals of Chicken TIM	182
48	Electron Density Map at 2.5Å Resolution; Three Turns of α-helix	183
A1	Data Processing Programs	190
A2	Difference Fourier Coefficients	204

LIST OF TABLES

1	Solutions of the Sulphydryl Derivative Difference Patterson	84
2	Contributions of Atoms at Sulphydryl Sites to Structure Factors	86
3	Survey of Potential Heavy Atom Derivatives	88
4	Changes in Unit Cell Parameters	92
5	Parameters of Data Collection and Processing at 6Å Resolution	93
6	Parameters After Refinement of Derivatives at 6Å Resolution	96
7	Structures Resulting From Refinement of Initial Sulphydryl Site Parameters	102
8	Linear Diffractometer Settings for Crystals of Chicken TIM	108
9	Primary processing of 2.5Å Resolution Intensities	113
10	Agreement Between Friedel Related Reflections	117
11	Variation of the Ratio of Equivalent Reflections	118
12	Level Scale Factors	121
13	Rescaling Levels of Measurements	124
14	Statistics of the Compiled Data Sets	126
15	Preliminary Scaling of Derivative Data	130
16	Incorporation of Low Resolution Data	131
17	Parameters After Initial Refinement of Derivatives at 2.5Å	133
18	Unit Cell Dimensions for Native and Derivative Crystals	135
19	Isomorphous and Anomalous Difference Statistics	139
20	Preliminary Phase Calculation	150
21	Effects of Errors In Heavy Atom Parameters	157
22	Derivative Scale Factors	164
23	Parameters After Further Refinement of Derivatives at 2.5Å	173
24	Calculation of Phases for the 6 - 2.5Å Reflections	174

ACKNOWLEDGEMENTS

I am grateful to my supervisor Professor David Phillips for the advice and assistance which he has given during many discussions about this work and I thank him for his part in its completion, when his interest and encouragement coupled with friendship and understanding have been deeply appreciated.

The work described in this thesis relies extensively on the parallel investigations of David Banner, Gregory Petsko and Ian Wilson to whom I am grateful for their collaboration. Related aspects of the structure of triose phosphate isomerase studied by them have included the methods of measuring diffracted intensities and refining the parameters of heavy atom isomorphous derivatives.

I thank all those members of Somerville College and the Laboratory of Molecular Biophysics who have made Oxford such a pleasant and stimulating place for me to live and work. I am also indebted to the friends and relatives whose kindness and hospitality have greatly eased my completion of this project during a year of difficult schedules and frequent commuting.

The award of a Medical Research Council Scholarship for Training in Research Methods (1968-71) and a Hertha Ayrton Research Fellowship from Girton College, Cambridge (1971-) are gratefully acknowledged.

Anne Bloomer

ABBREVIATIONS AND NOTATION

$\underline{a}\ \underline{b}\ \underline{c}$	Unit cell translations
d_{hkl}	interplanar spacing
*	refers to the reciprocal space cell
λ	wavelength
θ	Bragg angle
ρ	electron density
$\zeta\ \xi$	cylindrical coordinates in reciprocal space
$\phi \times \omega\ \theta$	setting angles for a 4-circle diffractometer
μ	angle of tilt for linear diffractometer
α	phase angle
$\beta\ \gamma\ \phi$	angles in the phase triangle
$\phi\ \psi$	spherical polar coordinates
σ	standard deviation
I	intensity
F f	observed structure factor or amplitude
Z	occupancy
B	temperature factor
k	scale factor
h k l	indices of reflections
x y z	real space coordinates
u v w	vector space coordinates
k'	ratio of real to anomalous scattering
E E'	RMS lack of closure for isomorphous and anomalous triangles
e e'	lack of closure for a single reflection
b a	intercept and slope of linear regression analysis
R	reliability index
m	figure of merit
TIM	Triose phosphate isomerase
DHAP	dihydroxyacetone phosphate
GAP	glyceraldehyde-3-phosphate
PCMB	p-chloromercuri-benzoate
MONO	2-chloromercuri-4-nitrophenol
DI	2,6-dichloromercuri-4-nitrophenol
EMP	ethyl mercury phosphate
BAKER	2,3-di-methoxytetramethylene)-bis acetate;mercury meso-
$PtCl_4$	potassium chloroplatinite
RMS	root mean square
HRS	Hamilton,Rollett & Sparks (1965).
CRT	cathode ray display tube
MRE	mean residual error

Chapter I INTRODUCTION

A) BIOLOGICAL MACROMOLECULES 2

B) X-RAY CRYSTALLOGRAPHY 9

- Geometry of Diffraction by Crystals 10
- Generation and Measurement of X rays 15
- Principles of Structural Analysis 18
- Phase Determination using Crystal Symmetry 20
- Phase Determination using Vector Methods 24
- Phase Determination using Direct Methods 28
- Phase Determination using Related Structures 31
- Protein Crystallography 32

C) MOLECULAR ENZYMOLOGY 34

- Structure of Proteins 34
- Function of Proteins 36
- Triose Phosphate Isomerase 38

D) CRYSTALLOGRAPHIC STUDIES OF TRIOSE PHOSPHATE ISOMERASE 40

A) BIOLOGICAL MACROMOLECULES

Living systems are capable of effecting such a variety and complexity of processes that they were long thought to lie outside the realm of the basic laws of physics and chemistry. The concept of vitalism, that is the existence in all living organisms of a "vital force" which clearly distinguishes animate from inanimate matter, was not completely discredited until the synthesis of urea from ammonium cyanate by Wohler in 1828. This was followed by syntheses of a variety of organic compounds from inorganic starting materials, proving beyond dispute that organic and inorganic chemistry were only different aspects of a single discipline.

The applicability of the law of conservation of energy to the respiration of animals, as demonstrated by Lavoisier and Laplace in 1785, was not yet sufficient to discredit the theory of spontaneous generation of micro-organisms. This survived for almost another century until the conclusive experiments of Pasteur in 1862 eventually disarmed all his critics. The apparent contradiction between these hypotheses, suggesting both a spontaneous origin of life and a constant vital force, illuminates the difficulty of distinguishing between living and non-living systems. Studies of the process of fermentation in yeasts led to their being assigned to both of these categories by different investigators, but the clarification of the nature of the fermentation process by Pasteur, followed by Buchner's demonstration in 1897 of the fermentation of sugars by a cell-free extract of yeast, provided the evidence which proved that yeast is a living cell, and which also destroyed any lingering vestige of vitalism.

However, the distinction between living and non-living matter has again been blurred, and remains so in the case of many viruses.

These do contain genetic information, in the form of nucleic acids, for the reproduction of like particles, and also a coat of proteins to encapsulate and protect the genes, but the chemical systems required for the process of reproduction are missing, and so this can only occur within a living cell which does possess this ability. Such cells contain both many more chemical species and also different types of proteins and nucleic acids, which have evolved with widely varying functions.

The two molecular species which are common to all forms of life or potential life, proteins and nucleic acids, are both linear polymers, consisting of amino acids and ribo- or deoxyribo-nucleosides respectively. A third major polymeric species is carbohydrate, whose functions include supplying a reserve of chemical energy and providing structural rigidity. The subunits of all of these polymers are poly-functional, but only in the cases of carbohydrates and fibrous proteins does this give rise to partial internal cross-linking. Nucleic acids and proteins are both synthesised as linear polymers but have been shown to exist in specific three dimensional conformation, the nature of which can profoundly influence their functional properties. Clearly the potential reactivity of any group is determined both by its local environment, in respect of such factors as steric hindrance, accessibility to reactants, the dielectric constant and conductivity of the medium, and also by its chemical state in respect of any charge, the disposition and availability of its electronic orbitals and the nature of any other reactants. For any particular conformation all of these factors will be precisely determined. The operation of natural selection will cause the evolution of that conformation which is best suited to a given function. The information specifying a particular conformation must be contained in the linear sequence of subunits, in the absence of any other mechanism,

and so indirectly it is this sequence which determines the characteristic properties of each polymer.

Nucleic acids are known to exist in both single- and double-stranded forms. The double helical structure which was proposed for genes of the latter type (Watson and Crick, 1953) explains their ability to replicate accurately, to provide information for protein synthesis by transcription and translation, and to mutate occasionally. This model leaves open the questions of the control of these processes by specific interactions between nucleic acids and proteins, and the extent to which specific sequences of nucleic acids may influence the basic structure. Single-stranded genes present a more difficult problem, but detailed analysis of one such gene from a bacterial virus has led to an equally definite three dimensional structure of considerable complexity being proposed for this molecule (Min Jou et al., 1972).

The relationship between the sequence and the structure of nucleic acids was illustrated by studies of a variety of synthetic polynucleotides in which it was observed that the two strands of the double helix are slightly further apart than normal for one particular sequence, that being a co-polymer of deoxyadenosine and deoxythymidine (Langridge, 1969). This observation supports the hypothesis that the points of initiation of the transcription process are associated with such sequences of nucleotides (Szybalski, Kubinski and Sheldrick, 1966), the increased separation of the strands of the helix possibly facilitating the insertion of initiating or other controlling molecules.

Interactions between nucleic acids and proteins must involve some kind of mutual recognition features in their respective conformations if the interactions are to be specific. Such features are postulated in a mechanism for the binding of a small protein, the lac repressor, to a

specific gene of double-stranded DNA (Adler et al., 1972) involving protuberances of the protein which are complementary to the wide groove of the DNA double helix, and specific amino acid and nucleoside residues interacting in precise relative orientations.

The details of protein conformations are most conveniently discussed in terms of the classification of protein structure which was originally proposed by Linderstrom-Lang. Primary structure is the covalent bonding of a protein. This includes the bonds within the amino acid monomers, the peptide linkages between them and also, whenever they are present, the disulphide bridges linking two parts of the polypeptide chain. Disulphide bridges are the only form of cross-linking which is found in globular proteins. They occur between two appropriately positioned cysteine residues but are not formed until after synthesis of the primary structure has been completed. The synthesis of all proteins is in the form of a single primary sequence, which can then fold itself into the correct conformation, although subsequent activation of some proteins involves the excision of certain residues from within this single chain.

The next two levels of this classification of protein structure concern the three dimensional conformation. Secondary structure is that imposed on the flexible polypeptide chain by hydrogen bonds between the carbonyl and amide groups of the backbone peptide bonds. The commonest forms observed are those predicted by Pauling, Corey and Branson (1951), the α-helix, a right handed helix with 3.6 amino acid residues per turn (left handed α-helices are also possible but have never been observed), and two β-pleated sheet structures in which the adjacent polypeptide chains run in parallel or anti-parallel directions.

Several correlations have been made between the primary sequence

and secondary structure of a protein, especially with respect to predicting the occurrence of α-helices. Considerations of the frequency of occurrence of individual amino acids in helical and non-helical regions have given rise to predictions having low statistical significance, except for the imino acid proline, which can not occur within an α-helix, but only at the amino terminal of one, because of stereochemical restrictions. However, analysis of the interactions between pairs of amino acids has produced significant conclusions, leading to the suggestion that it is the interactions between the side groups of residues and the polypeptide backbone which have the greatest influence in determining the secondary structure, and not those between different side chains (Robson and Pain, 1971; Finkelstein and Ptitsyn, 1972).

The tertiary structure of a protein is the sum of all the side chain interactions such as hydrogen bonds, salt linkages and van der Waals forces which, together with the primary and secondary structure stabilise the three dimensional conformation of a protein.

The variety of roles for which proteins have evolved is partly evident from their range of conformations. Fibrous proteins such as collagen, keratin and sclerotin have structural roles. Those such as actin and myosin combine a structural role with a capacity for dynamic changes. Globular proteins can act as carriers, as for example haemoglobin and transferrin do for oxygen and iron. The electron transport proteins combine this ability with a capacity to be oxidised or reduced in the presence of certain metabolites. Small proteins such as insulin and glucagon are found as hormones, integrating different aspects of metabolism. Proteins of all sizes are known to act as catalysts in many different reactions. The efficiency of such

proteins as biological catalysts or enzymes is such that all the chemical processes of life can occur in the absence of any extreme conditions of pH, temperature or pressure, that is in the conditions found within the living cell. Strong selective pressures have clearly been present to cause the evolution of such a variety of highly specialised enzymes.

A further level of organisation, the quaternary structure, is found in many of the globular proteins, whose active forms are aggregates of a small number of globular protein monomers. Most commonly the subunits of the aggregate are all identical, as in the tetrameric dehydrogenases, but an increasing number of mixed aggregates are being identified: tetrameric haemoglobins consist of two chains each of two very similar but not identical polypeptide chains; aspartate transcarbamylase is a dodecamer containing six units of both a smaller regulating protein and a larger catalytic protein. Almost invariably the disposition of the subunits in a protein with a quaternary structure is such that the symmetry of the oligomer is higher than that of the protomer, a feature which can result in the amplification of a signal controlling the activity of the protein.

The catalytic activity of an enzyme can be considered in terms of the binding of the reactants, the presence of groups on the enzyme able to participate in and facilitate the reaction, and the control of these processes by interaction with other molecules. The highest reaction rates will be achieved if the enzyme has a strong affinity for all its substrates, binding these in a favourable relative orientation whenever the binding is simultaneous, and a low affinity for its products ensuring their rapid release. The existence of groups on the enzyme having favourable properties such as ionisation and redox

potentials or nucleophilicity as required for any particular reaction, in positions appropriate to the orientation of the bound substrates will thus provide a considerable contribution to the catalytic power of an enzyme. Both the kinetic and the steric factors in enzymatic activity can be affected by interactions with other molecules in a variety of ways. The effects of such interactions could include blocking the substrate binding site, altering the conformation of that region of the protein, changing the character of important catalytic groups and interfering with charge relay systems in the protein. Such considerations illustrate the variety of control mechanisms which are possible, several of them arising from primary interactions at a considerable distance from the site of activity of the enzyme. A more radical method of exerting control over the rate of a metabolic reaction exists at the level of protein synthesis when transcription of the gene specifying the primary sequence of a particular enzyme can be induced or repressed.

The dependence of the activity of a protein on a specific conformation, can not preclude it from adopting a range of related conformations, as a protein must always be in a state of dynamic equilibrium with its surroundings, and so a conformational change may frequently enhance its activity. The theory of induced fit proposed by Koshland (1958) postulates that the binding of the substrate at the active site causes considerable alteration in the geometry of the protein, which induces the proper orientation of the catalytic groups. This conflicts with the allosteric theory of Monod, Wyman and Changeux (1965) which postulates that a conformational change in one subunit of an oligomeric protein, whether caused by a substrate at the active site or an effector at some distance from it, induces similar changes in all

the other subunits, which facilitate their subsequent rapid binding of substrates. Thus the allosteric theory requires conservation of the symmetry of the oligomeric conformations whereas the induced fit theory requires conservation of the relationship between conformation and ligand binding, and a continuous range of possibilities between these two extremes is possible. Individual proteins are thought to follow a path somewhere between those of the two idealised theories.

Elucidation of the mechanisms of these changes, and also of the remarkable catalytic ability of many proteins requires a detailed knowledge of the conformations of proteins. The prediction of conformations on the basis of primary sequence is not yet possible, and indeed the determination of the primary sequence of any large protein remains a problem fraught with technical difficulties. A variety of spectroscopic techniques can provide information about the local environment of selected groups, from which deductions about the dispositions of neighbouring groups are possible, but the only method which, if it can be applied successfully to a structure, produces the coordinates of all its constituent atoms, and thus a knowledge of the complete three dimensional conformation is the method of crystal structure determination, one application of which forms the major subject of this thesis.

B) X-RAY CRYSTALLOGRAPHY

Consideration of the external morphology of crystals provided the earliest basis for their study leading to the demonstration by Bravais of the possible existence of only fourteen regularly repeating space lattices. These can be assigned to the seven different crystal systems in which the unit cell is defined by a set of three vectors

whose generality decreases as the symmetry of the system increases; from the triclinic unit cell, defined by three vectors having undefined relative lengths and inclinations, to the cubic unit cell, defined by three equal but mutually perpendicular vectors. The law of rational indices postulated by Haüy, which relates the faces of the crystal to these unit translations of the lattice, states that the intercept which each plane face of the crystal makes with these three directions must be a rational fraction of the repeat distance of the lattice in that given direction. This restriction enables the relative lengths and inclinations of the edges of the morphological unit cell to be deduced.

Investigation of the arrangement of matter within the unit cell of a crystal was not thought to be generally possible until the discovery of X-ray diffraction in 1912. The diffraction photograph of a crystal of copper sulphate obtained by Friedrich, Knipping and Laue indicated firstly that X rays were indeed waves and not corpuscles, and secondly that they provided a very powerful tool, as was realised by W.L. Bragg, for studying the structure of regular crystalline solids, which act as three dimensional diffraction gratings for these rays whose wavelengths are comparable to interatomic distances.

Geometry of Diffraction by Crystals

The relationship between the orientation of a crystal with respect to the incident X-ray beam and the positions of the diffraction maxima was shown in two different ways. Laue used as his model an infinite three dimensional diffraction grating, analogous to the regular array of lattice points within the crystal, to show that the rays will interfere constructively only in those directions which satisfy the

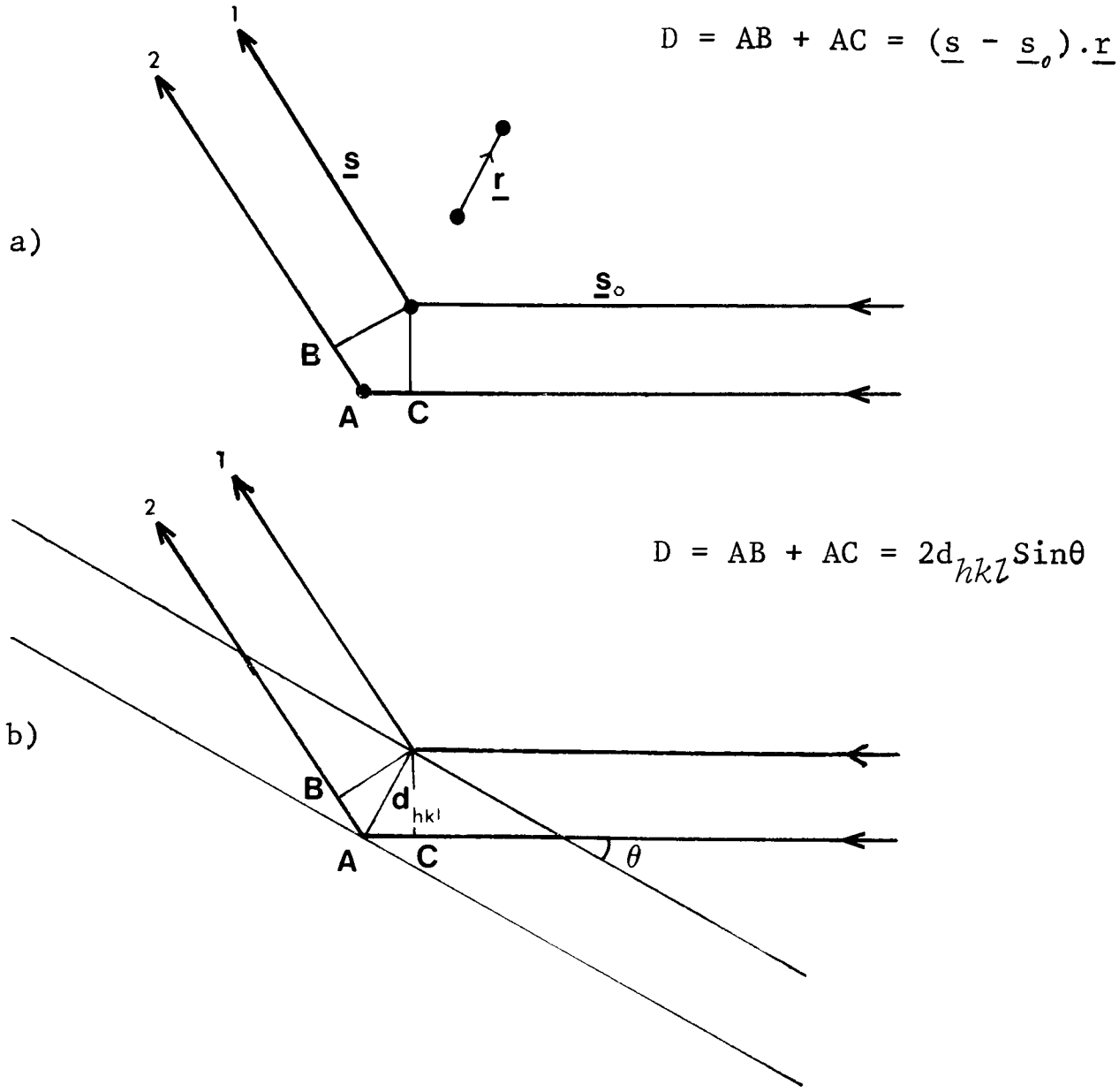


FIGURE 1 Geometrical Conditions for Diffraction by a Crystal. The path difference between rays 1 and 2 is denoted D, s and s_o are unit vectors.
a) Row of regularly spaced lattice points illustrating the derivation of Laue's equation; b) Traces of regularly spaced planes illustrating the derivation of Bragg's law; c) Superposition in reciprocal space of sections through the reciprocal lattice and Ewald sphere, illustrating the occurrence of diffraction only by those points which lie on the sphere.

equation

$$(\underline{S} - \underline{S}_o).\underline{r}_i = n_i\lambda \qquad i = 1,2,3 \tag{1}$$

where $\underline{S}_o$ and $\underline{S}$ are unit vectors parallel to the incident and diffracted beams of radiation having wavelength λ, n_i are integers and the three non-coplanar vectors $\underline{r}_i$ describe the nature of the periodicity of the diffracting object. The derivation of this equation can be seen from Figure 1(a) for the case of an object having one dimensional periodicity.

The set of vectors $\underline{r}_i^*$ which are reciprocal to the vectors $\underline{r}_i$, that is they obey the relationship

$$\underline{r}_i^* = \frac{\epsilon_{ijk}\ \underline{r}_j \wedge \underline{r}_k}{\underline{r}_i.(\underline{r}_j \wedge \underline{r}_k)} \tag{2}$$

where ϵ_{ijk} has the values +1,-1,0 if i,j,k are a cyclic or acyclic permutation of the integers 1,2,3, or any two of i,j,k are equal

define a lattice which is reciprocal to the space lattice of the crystal and which is said to exist in reciprocal space, a concept which greatly simplifies the understanding of the geometry of diffraction by crystals. Equations (1) can be rewritten in terms of the reciprocal vectors as

$$(\underline{S} - \underline{S}_o) = n_i \underline{r}_i^* \lambda . \tag{3}$$

Using the conventional crystallographic notation and replacing the vectors $\underline{r}_i$ which describe the unit cell translations by $\underline{a}$,$\underline{b}$ and $\underline{c}$ and the integers n_i by h,k and l, the reciprocal form of the Laue equation can be expressed

$$(\underline{S} - \underline{S}_o) = (h\underline{a}^* + k\underline{b}^* + l\underline{c}^*) \lambda . \tag{4}$$

If the vector joining the origin of reciprocal space to the point whose coordinates are the integers hkl is denoted $\underline{d}^*_{hkl}$ then

$$(\underline{S} - \underline{S}_o) = \underline{d}^*_{hkl}\lambda \ . \qquad (5)$$

W.L. Bragg obtained this result by a completely different method, considering the diffraction process in terms of reflection by a series of parallel planes in the crystal. The condition obtained from this model for the constructive interference between the diffracted rays gives rise to the well known equation of Bragg's Law

$$2d_{(hkl)} \sin\theta = n\lambda \qquad (6)$$

as is readily seen from Figure 1(b), where θ is the angle between the incident or diffracted ray and the planes of spacing $d_{(hkl)}$, and (hkl) are the Miller indices of the set of planes making intercepts 1/h, 1/k, 1/l with the unit cell translations $\underline{a},\underline{b},\underline{c}$ respectively. The integer n is necessary because Miller indices contain no common denominator, a convention arising from the morphological study of crystals. Ignoring this convention can be indicated by lack of parentheses surrounding the indices, and allows consideration of the diffraction by, for example, the set of planes denoted 222 instead of the equivalent second order diffraction from the planes (111).

It is readily shown that for a set of planes so defined, the relationship between the interplanar spacing d_{hkl} and the reciprocal lattice vectors is simply

$$\frac{1}{d_{hkl}} = \left| h\underline{a}^* + k\underline{b}^* + l\underline{c}^* \right| \qquad (7)$$

indicating that the vector $\underline{d}^*_{hkl}$ as already defined is reciprocal in magnitude to the spacing of the planes. It can also be proved to be perpendicular to them. The equation representing Bragg's Law may thus be cast into the form

$$2 \operatorname{Sin}\theta = \lambda \left| \underline{d}^*_{hk\ell} \right| \tag{8}$$

which closely resembles equation (5), the result of Laue's approach to the phenomenon, differing only in that one equation relates vectors and the other scalars, since it is obvious from Figure 1(b) that

$$\left| \underline{S} - \underline{S}_o \right| = 2 \operatorname{Sin}\theta \tag{9}$$

However, the orientation of the set of planes to the irradiating beam, is implicit in the model from which Bragg derived his law, so that the two approaches are formally identical. The real space approach of Bragg is more easily related to the setting and orientation of the irradiated crystal, whereas the reciprocal space model arising from the work of Laue is more readily related to the positions of the diffraction maxima.

It is clear from equation (6) that no diffraction can be observed from planes whose spacing is less than half of the irradiating wavelength. All other sets of planes can be seen to be represented by those points of the reciprocal lattice lying within a sphere of radius $(2/\lambda)$ centred on the origin of reciprocal space. A geometrical construction illustrating which of the many sets of planes capable of diffracting are actually in the correct orientation to do so, for any given setting of the crystal, was suggested by Ewald. The reflecting or Ewald sphere has a radius equal to half that of the limiting sphere, the origin of reciprocal space being one point on its surface and the path of the incident beam being the diameter of the sphere through that point. The applicability of this construction can be seen from Figure 1(c). In the case of a three dimensional crystal, the criterion for diffraction by a set of planes is that their corresponding reciprocal

lattice point lies on the surface of the sphere.

The radii of the Ewald and limiting spheres as already defined depend on the wavelength of the incident radiation. The validity of the two constructions is unaffected if both radii are multiplied by the wavelength, giving the dimensionless quantities unity and two for their respective radii, provided that the spheres are in this case superposed on a dimensionless reciprocal lattice. This is constructed by multiplying all previously defined reciprocal lattice distances by the wavelength, such that equations (1) and (7) become

$$\underline{r}^*_i = \lambda \in_{ijk} \frac{\underline{r}_j \wedge \underline{r}_k}{\underline{r}_i \cdot (\underline{r}_j \wedge \underline{r}_k)} \tag{10}$$

$$\frac{\lambda}{d_{hkl}} = |\underline{d}^*_{hkl}| \tag{11}$$

respectively. The main advantage of using these dimensionless definitions, in which the scale of the reciprocal lattice depends on the wavelength, is that variations in the incident wavelength can be regarded as transforming the reciprocal lattice points into radial streaks, leaving the Ewald sphere with a uniformly zero thickness. The reciprocal lattice points are only infinitely small in the case of an absolutely perfect and infinite crystal, a situation which never obtains in practice, all departures from strict regularity in the real space lattice of the crystal being transformed into reciprocal space as volumes surrounding the lattice points. It is thus always more realistic to consider reciprocal lattice spots, instead of points, and all variation within an experimental situation can be considered to contribute to the size of these reciprocal spots, provided that the dimensionless definitions are used. Reciprocal space will subsequently

always be considered to be defined in this way.

Generation and Measurement of X rays

The theory of diffraction by crystals is not specific to X rays, but is applicable to the scattering of all forms of radiation. The extent to which any atom scatters radiation, a theoretically calculable quantity denoted the atomic scattering or form factor, does depend on the nature of the radiation, such that the information contained in the diffraction pattern will vary accordingly. X rays are scattered by the electron shells surrounding the atoms, thus giving information about their electron densities. The scattering of neutrons depends upon both the potential of the nuclear forces and the magnetic moment of the electron shells, which yield information about the arrangement of atomic nuclei. Details about the electrical potential of an atom are provided by measurements of electron scattering. The majority of crystallographic investigations have been carried out with X rays, mainly because the generation of a suitable flux of radiation was technically possible for X rays many years earlier than for neutrons or electrons, and they remain the most convenient radiation for use, in addition to their providing the clearest distinction between the different types of atom present within a structure.

X-ray crystallography relies very heavily on technical advances in methods of both generating and detecting X rays and also data handling and processing. The first X rays observed by Röntgen in 1895 were produced on discharge of an induction coil within a glass tube. The many advantages to be gained from having a high intensity primary beam have now led to the development of generators in which a water cooled rotating anode can be bombarded by a high flux of electrons,

producing a powerful beam of X rays which can be very finely focussed. The detection of X rays was originally achieved by photographic methods, although the first quantitative measurement of intensities was achieved by W.H. Bragg using an ionisation spectrometer. Both film and counter methods of detection are now in common use, technical development over the past several years having produced highly automated systems employing each technique for the measurement under computer control of large numbers of diffracted intensities.

A diffraction spot always occupies a finite volume in reciprocal space unless the specimen is an ideally perfect crystal, with no irregularities of any sort, and it is irradiated by a point flux of X rays. Atomic thermal vibrations will always prevent such perfection, but it is found that in practice most crystals have a mosaic structure, that is they are composed of many small perfectly regular crystallites, which are slightly disoriented with respect to each other. The size of these mosaic blocks, relative to the size of the crystal governs the extent of the extinction suffered by the diffracted rays, this phenomena being a special case of the general absorption of X rays by crystals, which occurs when a set of planes are in the diffracting position.

The energy diffracted by a set of planes can be envisaged as arising throughout the volume of the reciprocal lattice spot, and thus the entire spot must pass through the surface of the reflecting sphere if the integrated intensity of the diffracted beam is to be used as a measure of this energy. The relation between these two quantities depends on the method and speed of moving the spot through the sphere, but a correction factor between them, known as the Lorentz factor can be calculated for any experimental situation. An additional geometrical

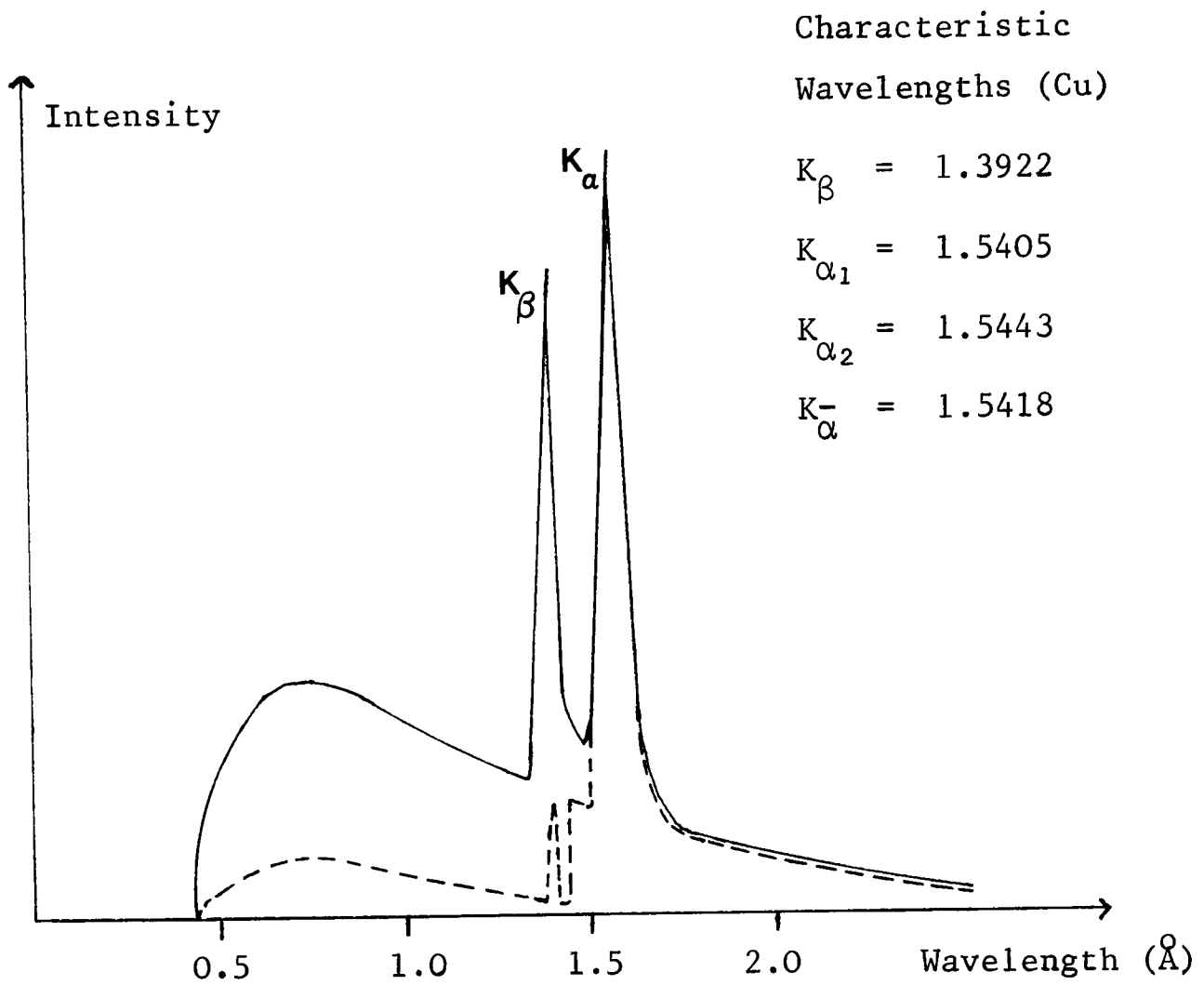


FIGURE 2 Typical X-ray spectrum emitted by a tube with a Copper target operating at about 35kV before (——) and after (- -) passage through a nickel filter.

correction factor is necessary because of the variation of the state of polarisation of the X rays for different angles of diffraction.

In order for Bragg's Law to be satisfied and there to be diffraction from a set of planes in any real crystal, there must be experimental provision for the variation of the wavelength or the angle of incidence of the radiation. The latter is most easily achieved by movement of the crystal within the incident beam, many mechanical systems having been devised for this purpose, and this allows the use of a single wavelength. The angle of diffraction of any reflection can be measured, so that if the wavelength is known, the interplanar spacing is calculable. The technique of Laue photography, however, uses a continuous range of wavelengths and a stationary crystal, revealing information about its symmetry.

Sources of X rays produce a continuous spectrum called white radiation, whose nature varies with the accelerating voltage within the X-ray tube, superimposed upon a small number of intense characteristic wavelengths, dependent on the metal used as the target for the electrons, the multiplet of characteristic lines arising from the transitions of electrons between different orbitals. The characteristic wavelength of the rays produced by any anode element decreases with increasing atomic number. A typical spectrum is shown in Figure 2 which also shows the spectrum after passage through a filter of an appropriate element, usually that with atomic number one less than that of the anode element, since the absorption edge of an atom occurs at a wavelength slightly shorter than that of its characteristic line. An alternative method of obtaining a monochromatic beam is to use that diffracted by a set of crystal planes in a specimen with a very low mosaic spread or even a perfect crystal.

The choice of wavelength to be used in an investigation depends on a variety of factors. The most common choice of crystallographers is that of the copper K_α doublet, combining relatively low absorption with high efficiencies of detection by photographic films and proportional and scintillation counters. Longer wavelengths such as iron and cobalt are useful for very large unit cells, giving greater separations between the reciprocal lattice spots, and for crystals containing metals which strongly absorb shorter radiations. The very short wavelength of molybdenum is useful for studies at high resolution, increasing the number of orders of diffraction observable within the maximum angle permitted by an instrument.

The range of Bragg angles for which diffracted intensities are measured determines the resolution of the structural detail which can result from subsequent analysis. Measurement of intensities at larger angles results in finer detail being apparent in the structure because of the reasoning of the classical theories of image formation and optical microscopy which are equally applicable to X-ray images. The focussing of light rays by lenses to form an image has no counterpart for X rays since appropriate lenses do not exist, and so that part of the image formation process has to be achieved mathematically for X-ray structural analyses.

Principles of Structural Analysis

The most useful technique for this purpose is that of Fourier series, as shown first by Bragg. The Fourier transform of the electron density within a crystal is zero except at the reciprocal lattice points, whose positions are determined by the size and shape of the

unit cell, and it varies between different reciprocal lattice points according to the arrangement of electrons within the unit cell. The value of the transform at any reciprocal lattice point with coordinates hkl is termed the structure factor, denoted F_{hkl}, a complex quantity whose product with its complex conjugate is proportional, for a mosaic crystal, to the intensity of the corresponding diffracted ray. The constant of proportionality includes such factors as the intensity of the incident beam, the volumes of the crystal and its unit cell and the intensity of scattering from a free unbound electron. The converse Fourier transform of the structure factors is proportional to the electron density distribution within the unit cell, and the determination of this is the basic problem of crystal structure analysis.

The structure factors are not measurable experimentally, but their amplitudes can be determined from the intensities of the diffraction maxima. The phase of every structure factor remains unknown, but this is the more important part of the information required for the Fourier synthesis of the electron density. A variety of methods have been developed for the determination of some or all of these phases, allowing a trial electron density synthesis to be computed. An assessment of the reliability of such a density map depends upon the extent to which it can be interpreted in terms of chemically reasonable groups, a factor which is difficult to express quantitatively, but analyses which are carried out to atomic resolution permit the calculation of the diffraction pattern which would be expected from the trial structure. The agreement between the calculated and observed structure amplitudes is expressed by the crystallographic reliability index or residual R, which is defined in terms of the structure amplitudes F as

$$R = \frac{\langle||F_{obs}| - |F_{calc}||\rangle}{\langle|F_{obs}|\rangle} \qquad (12)$$

Changes in the value of this index indicate progress in the refinement of the initial trial phases.

Methods of solving the phase problem may be considered in four main groups which are based on information deduced from restrictions imposed by the symmetry of the unit cell, the set of interatomic vectors present within the structure, direct mathematical relationships between the phases of certain structure factors, and a series of modified but related crystal structures respectively. The principles of each of these approaches are considered here in relation to the types of structure for which they are generally applicable for the determination of the phases of the structure factors.

Phase Determination Using Crystal Symmetry

Consideration of all the possible combinations of symmetry elements, inversion centres of symmetry, mirror planes of reflection, rotation axes and glide planes and screw axes, in which the plane or axis of symmetry is also associated with a translation parallel to itself, has shown that there exist only 230 of these which are consistent with a regularly repeating space-filling lattice. The properties of all these space groups are well known and have been tabulated in the International Tables for X-ray Crystallography (Volume I), where standard formulae are presented for each space group, relating the structure factor for any reflection to the positions of only those atoms contained within a single asymmetric unit of the translation unit cell. An inspection of these formulae, or more directly the

combination of symmetry elements from which they are derived, indicates that the presence of translational elements of symmetry is associated with the systematic absence of particular classes of reflections in specified lines or zones of reciprocal space. The systematic absence of certain classes of reflections throughout the volume of reciprocal space is associated with the presence of additional lattice points, points which are identically related to their environment, at the mid-points of the face or body diagonals of the unit cell. Recognition of these absences in the diffraction pattern of a crystal, a knowledge of the symmetry of this pattern and of the three basic vectors defining the repeating unit of the structure, provide the basis for assigning the structure to a particular space group. Additional information, such as whether or not the structure is centrosymmetric, which can often be deduced from chemical knowledge or statistical tests of the distribution of the diffracted intensities, usually provides an unambiguous assignment. Determination of the exact disposition of the symmetry elements reduces the size of the unknown structure to that of a single asymmetric unit, a simplification which in many cases is relevant to the phase problem.

In principle, the simplest examples of this problem, although sometimes involving additional ambiguity, are the centrosymmetric structures. The presence of this element of symmetry constrains the structure factor to be a real number, so that its phase may be regarded as 0 or π radians, corresponding to positive and negative amplitudes respectively. Most of the successful early structural analyses were either of this kind or of structures whose projections are centrosymmetric and are capable of solution by consideration only of two or more appropriate zones of reflections. The first crystal structure to

be solved by W.L. Bragg in 1912 was that of sodium chloride, whose structure is one of those few which contain no variable parameters and provide an example of the first class of methods for solving the phase problem.

The number of molecules of a compound which are present within a single unit cell of the crystal is determined from a knowledge of its molecular weight, the density of the crystal and the volume of the unit cell. Whenever the crystal lattice is non-primitive, as indicated by the existence of systematic absences, or contains elements of symmetry, then the number of molecules per asymmetric unit may be further determined. If this number is a rational fraction less than unity, then clearly the constituent atoms of the molecule must all occur at special positions within the unit cell, lying on elements of non-translational symmetry.

The structures of many simple inorganic compounds, containing equal numbers of each of two different types of atom, belong to the holosymmetric class of the cubic system having a face centred lattice with one atom of each type per lattice point. If the atoms of one type are placed at those points of intersection of the diad, triad and tetrad rotation axes which lie on a face centred lattice, then there are only two possible positions for the atoms of the second type. These are separated from the atoms of the first type either by half the unit translation along a tetrad axis parallel to the edge of the unit cell or by a quarter of the unit translation along a triad axis parallel to the body diagonal of the unit cell. These two structures are physically quite distinct, the atoms in each being surrounded by an octahedral or tetrahedral shell respectively of atoms of the opposite type. The very different pattern of relative intensities produced on

diffraction by these structures allows them to be readily distinguished. Sodium chloride has a structure with octahedral coordination, the tetrahedral coordination being found in zincblende and related structures. Eightfold cubic coordination can also be found in such simple structures, but this occurs in a primitive cell as is found for caesium chloride which has atoms of one type at the corners and the other type at the centre of the cubic unit cell.

The information obtained from symmetry considerations is usually less complete than in these simple cases, which appear simple only with the benefit of sixty years' experience, but can still appreciably reduce the number of unknown parameters, as for example in the case of many complex ions. If an ion having a tetrahedral structure, as exhibited by sulphates, chlorates, silicates and phosphates, can be shown to lie on a position of three-fold rotational symmetry, then clearly in the latter case, the phosphorus atom and one of the oxygen atoms must occur precisely on the triad axis, the remaining three atoms being symmetrically placed about it. If an approximate interatomic distance is known, these constraints reduce the problem to one of determining the axial position of the phosphorus atom, the direction of the axial oxygen atom with respect to it, and the azimuthal coordinate of one of the other oxygen atoms. These three parameters represent a two-fold reduction in the number required to specify any general position of this group of atoms within the unit cell, such reductions frequently making possible the solution of structures by a process of trial and error. This implies the comparison of the observed diffraction pattern with that calculated for every conceivable trial structure, such that any reduction in the number of unknown parameters considerably simplifies this process. The example discussed is clearly only one of many possible

groups and symmetry restrictions, the phosphate ion alone possessing molecular diad, triad and inverse tetrad axes, and two perpendicular mirror planes, any or all of which may be present as crystallographic elements of symmetry in the structure.

The number of trial structures which must be considered can sometimes be decreased if structural information can be deduced from the physical properties of the crystals. The orientation and extent of anisotropy of such properties as refractive index, thermal expansion, electrical conductivity, crystal morphology and cleavage can frequently indicate the alignment of groups within the unit cell. However, any method of solution of the phase problem which is based on the consideration of all the possible structures very rapidly becomes impossible as the complexity of the structure, and so the number of trial structures to be considered, increases and thus a different approach to the problem is required.

Phase Determination Using Vector Methods

The second category of methods of solving the phase problem is based upon the demonstration by Patterson in 1935 that the Fourier synthesis, in which the coefficients are the squares of the structure amplitudes, or intensities, contains peaks at positions relative to the origin corresponding to all the interatomic vectors within the structure. The relative height of a peak is equal to relative value of sums of the products of the number of electrons in each pair of atoms separated by the corresponding vector. A solution of the crystal structure on this basis requires firstly the identification of the set of interatomic vectors, and secondly the determination of the structure corresponding to this set of vectors. It is possible, but rarely found

in practice, for different structures to contain the same set of vectors and thus give identical diffraction patterns. Such structures are said to be homometric, and although no proof can be given that any structure which has been determined agrees uniquely with the observable diffraction pattern, it has been shown that a structure can not have a homomorph if it satisfies a certain set of conditions (Cochran, 1958; Karle and Hauptman, 1957).

Interpretation of a Patterson map in terms of a set of interatomic vectors can be a complicated process. The existence of any type of disorder within a crystal, including even thermal vibrations, results in a statistical distribution of the vectors and thus a broadening of the peaks in the Patterson map. Whenever the diffuseness of the peaks, or accidental identity of vectors within the structure, causes overlapping of the peaks on a vector map then the determination of a self-consistent set of vectors may be extremely difficult, especially if experimental errors are also significant. Knowledge of the molecular shape of part or all of a structure considerably helps the elucidation of its Patterson map. As an example, the presence of an aromatic ring in an organic compound must be associated with a particular pattern of vectors having hexagonal symmetry, and a detailed search of the Patterson function for such a vector set may reveal the orientation of this planar group within the unit cell.

The effect produced by the presence of symmetry elements in the cell was discussed by Harker in 1936 who showed that there would be a concentration of vectors on particular lines or sections through the Patterson function. All atoms which are related by a rotation axis are separated by a vector which has no component parallel to the axis

and all atoms which are related by a mirror plane are separated by a vector which is strictly perpendicular to that plane. Screw axes and glide planes can similarly be shown to give rise to Harker planes and lines respectively, so that in one of these sections alone there are vectors corresponding to every atom in the unique set and the atom related to it by the symmetry element. Only one or two of the three unknown coordinates can be determined from a single Harker section, but in many space groups, the combination of symmetry elements yields enough Harker sections to determine all the coordinates, and often also to test the consistency of different indications. However, two major difficulties can prevent the determination of the structure from the Harker sections alone. Firstly these sections may contain not only the self vectors between corresponding atoms but also any cross vectors between different atoms in the unique set, which are coincidentally separated by a vector lying in the Harker section. Such cross vectors can in principle be distinguished from the self vectors but this may be difficult if the Patterson peaks overlap to any significant extent. Secondly, the Patterson vectors are indeterminate with respect to the addition of multiples of the unit cell translations, which implies that atomic coordinates deduced from a Harker plane perpendicular to a diad axis, for example, are ambiguous to the extent of the addition of multiples of half of the unit cell translations. This ambiguity can usually be resolved by consideration of all the cross vectors, distributed throughout the Patterson map.

Various superposition methods have been suggested for obtaining the atomic structure in more difficult cases. If two copies of the Patterson function are superposed, after translating the origin of one map to either a suspected atomic site or the end point of a self

vector in the other, then the true atomic structure is revealed by the superposition, which can be most powerfully analysed by use of the minimum function. In principle this method will always work, but ambiguous results will arise if there are overlapping vectors and multiple superpositions will be necessary to reveal a single image of the structure.

The largest peak on any Patterson map will always be found at the origin, corresponding to the vector from every atom in the unit cell to itself. The next largest peak will represent those vectors involving the atoms having the largest scattering factors in the structure. The scattering factor of an element is directly related to its atomic number, which is loosely correlated with its atomic weight, and so those atoms with the largest scattering factors are generally referred to as heavy atoms. In an organic structure containing for example a single mercury atom, the weight of a mercury self vector will be one hundred times that of an oxygen self vector and ten times that of a mercury-oxygen cross-vector. This indicates the type of structure for which these vector methods are most powerful, those in which a small number of the atoms are very much heavier than the others.

Determination of the positions of the heavy atoms constitutes a trial structure, in which the remaining atoms are all simply ignored, but using phases calculated on this basis and the amplitudes observed in the diffraction pattern, the resultant Fourier synthesis will often contain peaks representing the remaining atoms at their true positions in the structure. This heavy atom method of solving the phase problem is clearly applicable in only a restricted number of cases, but has been very successful, as for example in the solution of the structure

of vitamin B_{12} which contains a single cobalt atom in a large organic molecule (Hodgkin et al., 1957).

The presence of heavy atoms within a structure results in observation of the phenomenon of anomalous scattering of the X-rays in which some of the scattered radiation is advanced in phase by one quarter of a cycle. All atoms do this to some extent, but it is usually only detectable for atoms heavier than sulphur. Although the existence of anomalous scattering, and especially its sensitivity to the wavelength of the incident radiation, can be utilised in a variety of ways to provide phase information its most important advantage is that it causes the breakdown of Friedel's law that the diffraction pattern is always centrosymmetric. Bijvoet showed in 1951 that it was possible to determine the correct enantiomorph for a non-centrosymmetric structure by means of this anomalous scattering effect. Determination of the correct enantiomer is always a difficulty with vector methods of analysis since the Patterson function is necessarily centrosymmetric.

Phase Determination using Direct Methods

The possibility of being able, in theory, to determine a crystal structure from a knowledge of the structure amplitudes alone, as shown by the success of the vector methods, led to considerations of methods for determining the phases directly from the set of structure amplitudes. The range of applicability of these direct methods is essentially complementary to that of the vector methods and so their development provides a third major instrument for attacking the phase problem.

The earliest results obtained by such methods were based on a

series of inequalities derived by Harker and Kasper who showed that the presence of axes of symmetry, rotation, inversion or screw axes, imposed constraints on some of the structure factors. The information in these inequalities is most useful for centric reflections in which only a sign has to be determined. They become more powerful if expressed in terms of the unitary structure factors, which are those structure factors resulting from an equivalent structure of point atoms without any thermal vibrations, and which are readily calculable.

Relationships of wider applicability were developed by Sayre in 1952 when he showed that for structures containing identical and discrete atoms the squares of all the structure factors were related to their self-convolutions, and that in practice these convolutions would be dominated by a relatively small number of terms. Thus if the signs of a few large reflections in a projection are known, and two such signs can usually be fixed arbitrarily as a means of defining the origin, then the signs of a few more reflections will be determined by them, and the process can be repeated, including more reflections at each stage.

Several procedures have been suggested for executing this analysis and systematically generating and examining all the relationships. The methods have been generalised to determine the phases of acentric reflections but the same basic principles are common to all of them. Methods of this type are especially prone to error, in that a single mistake in one cycle will be propagated throughout the succeeding cycles, and thus consideration of the probability of the correctness of any phase assignment becomes most important and formulae for determining this have been proposed by Cochran and Woolfson (1955)

and others.

The requirement for identical and discrete atoms in the derivation of these relationships can be considerably relaxed in practice, although the larger the departure from these conditions the less appropriate will be the estimated probabilities for the correctness of the assigned phaces. The types of structure for which this method is most appropriate are those of organic compounds containing only carbon, nitrogen, oxygen and hydrogen atoms, these being precisely the kind of structure for which interpretation of the Patterson function is the most difficult. However, the postulate of discrete atoms, together with the observation that the mean intensity of the diffracted rays, and thus their precision of measurement, decreases with increasing unit cell size, suggests that these methods are not very appropriate for the very large organic structures such as proteins and viruses. Many of these systems have been found to exhibit internal molecular symmetry and this provides an entirely different basis for the direct calculation of the phases of the structure factors.

The existence of non-crystallographic symmetry within a structure can only be used to provide phase information if the relative positions and orientations of the basic structural units are known. The rotation and translation functions, which were first developed in a series of publications by Rossmann and Blow, frequently provide this information which forms the basis of the various molecular replacement methods of phase determination. The power of this approach increases with the number of molecular subunits and so it is expected to become increasingly important as ever larger structures of this kind are investigated.

Phase Determination using Related Structures

Information which is provided by the study of a series of related crystal structures offers a further approach to the problem of the determination of phases and is considered as the fourth category of basic methods. Several series of inorganic compounds crystallise in the same form, differing only by a gradual change in the unit cell dimensions providing additional points of sampling of the molecular transform. This is essentially continuous and so all the signs of the centric structure factors can be determined if the transform can be sampled at a sufficient number of points to identify all the nodes. Application of this method to the protein haemoglobin was made possible by variation of the salt content of the mother liquor filling the interstices between molecules in the crystal, and the signs were successfully determined for reflections on some rows of the reciprocal lattice (Perutz, 1954).

Comparison of a series of isomorphous structures, whether naturally occurring series of minerals differing only by the substitution of one atomic species for another, or specifically prepared modifications, as can often be obtained by binding ligands to proteins, can provide the information to locate the points of difference between such structures by analysis of the Patterson function of the differences between the two sets of structure amplitudes. The demonstration by Green, Ingram and Perutz (1954) that the binding of a single heavy metal to a protein molecule could produce observable changes in the structure amplitudes was a turning point in protein crystallography. A combination of the information provided by two independent isomorphous derivatives, or a single derivative together with anomalous scattering information, is theoretically

sufficient to determine the phases of all the reflections of the native protein. In practice problems result from a lack of perfect isomorphism, experimental errors in recording the diffraction pattern, and inaccuracies in determination of the heavy atom parameters and relative scaling of the various sets of structure amplitudes, but this method of multiple isomorphous replacement remains the only proven method for solving the phase problem in the case of protein crystals.

Protein Crystallography

Crystallographic studies of proteins began in 1934 when Bernal and Crowfoot observed crystals of pepsin and discovered that they rapidly became disordered unless evaporation of the fluid filling the intermolecular interstices was prevented. This is now routinely achieved by mounting such crystals inside a fine capillary tube and sealing each end, frequently having left a drop of the mother liquor inside the tube, before orienting the crystal on a goniometer head prior to irradiation by the X rays. Radiation itself causes considerable damage to such crystals, altering the relative intensities of the reflections progressively with increasing doses of radiation suffered. The magnitude of these effects varies widely between different crystal species.

The eventual success of crystal structure analyses of proteins was accompanied by many doubts concerning the relevance of these results to biochemistry. The very existence of protein crystals was proof that these molecules did indeed possess well defined conformations, but not that the preferred conformation within the crystal was the same as that in solution, nor that several different conformations could not exist in dynamic equilibrium with each other and with the

solvent. Determination of the close structural homology between the monomeric sperm-whale myoglobin and the tetrameric horse haemoglobin, which crystallise in different space groups and have completely different arrangements of molecular packing, provided the first indication that intramolecular forces were stronger than crystal lattice forces in these situations. Subsequent studies of crystals of these proteins from many species, and in many different forms, have all supported this view. Physical properties of proteins, such as the α-helix content, the axial ratio of the equivalent ellipsoid, and chemical studies of the accessibility for reaction of specific amino acids in the protein, can all be determined from experiments in solution, but none of the many such results which are available are totally inconsistent with the known conformations of the protein molecules obtained by crystallographic studies. The not infrequent problems of reconciling results from such different techniques have so far all been accommodated by known difficulties in the interpretation of the experimental data.

Establishment of the general validity of the crystallographically determined conformations raises the question, when enzymes are being considered, of whether their catalytic activity is destroyed by the lattice constraints. A negative answer to this question was provided by the elegant experiments of Quiocho and Richards (1964), using crystals of ribonuclease and it is now generally accepted, though very difficult to prove experimentally in most cases, that proteins can be active in the crystalline state. Information about modes of binding of substrates, activators and inhibitors, which have been determined crystallographically now forms an important part of present knowledge about the molecular basis of enzymatic action.

C) MOLECULAR ENZYMOLOGY

The large amount of information which is now available concerning the molecular architecture and activity of enzymes is derived from studies of a very small sample of the known enzyme population. High resolution crystallographic studies, together with primary sequence determinations, are currently being completed at an accelerating rate, but of such results now available the majority are for monomeric, extracellular enzymes, which can in no sense be regarded as a typical sample. However, success in this field of crystallography has brought a change of emphasis away from an analysis of any protein whose crystals were both available and amenable to study, towards the choice of a system for study according to the interest of the results to be expected. This outline of present knowledge of the structural principles and mechanisms of activity of enzymes is followed by a consideration of the potential of this aspect of molecular biology and the reasons for choosing that system which forms the subject of the remainder of the present thesis.

Structure of Proteins

The first protein for which an electron density map was calculated, myoglobin, has eight lengths of α-helix which involve about 70% of the amino acid residues. The importance of this type of regular secondary structure was thus emphasised with the neglect of other possibilities until the conformations of several other proteins had been determined. No other group of globular proteins have been found to contain as large a proportion of helix as do the globins. The regularity of the helices observed in different structures varies considerably. The predominant forms of regular secondary structure

observed in some more recent protein structure analyses are the predicted β-pleated sheet structures. These have been observed in several proteins, there being extensive such regions in carbonic anhydrase, where adjacent chains in the sheet are anti-parallel, and in subtilisin and lactate dehydrogenase, where the chains are parallel. Both types of pleated sheet are found in carboxypeptidase.

Discretely organised units in the tertiary structure of a single polypeptide chain are observed in several monomeric enzymes, the chain being folded back on itself repeatedly within one region of a molecule before crossing into another region. The stability of these distinct regions is enhanced by hydrogen bonding or by disulphide bridges, three of which are observed in chymotrypsin. The relationship of such discrete regions to the manner in which proteins fold on adopting their specific conformation instead of that of a random coil remains unknown, but they suggest that centres of nucleation may represent the first stages of folding towards an ordered structure. The lack of significant progress towards an understanding of the folding problem in the light of the many protein structures which are now known, suggests that a less static technique than X ray crystallography is necessary for its elucidation. The interpretation of nuclear magnetic resonance spectra for staphylococcal nuclease, which had been selectively deuterated, and observed at various stages of its denaturation, has recently shed more light on this problem (Jardetzky et al., 1971).

The majority of enzymes are thought to be oligomeric, but few structural analyses for such systems have yet been completed. Crystals of the apoenzyme of the tetrameric lactate dehydrogenase molecule indicate that the molecular symmetry is 222, but on binding

the coenzyme there is a change of space group and the molecular symmetry is reduced to a single diad axis. Tetrameric haemoglobins contain a molecular diad axis which is conserved on binding of an oxygen molecule to the haem group despite relative movements of the subunits of several Ångstrom units. Significant differences may be found between related subunits as in the case of rhombohedral insulin, where the three dimers are related by a crystallographic triad axis to form a hexamer, but the diad axes are non-crystallographic and many differences between the subunits are apparent in the electron density map.

Function of Proteins

Crystallographic studies of the binding of substrates and their analogues have led to detailed proposals for the mechanism of action of several enzymes. That proposed for lysozyme was based on both binding studies and detailed model building of the substrate into the active site cleft. Confirmation of the mechanism by a variety of chemical experiments which were suggested by these proposals illustrates the value of the collaboration between chemists and crystallographers which becomes possible when a protein structure has been determined.

Enzymatic hydrolysis of peptide bonds by that class of enzymes which contain an especially reactive serine residue is thought to proceed by a common mechanism. Three of these enzymes, mammalian chymotrypsin, trypsin and elastase show considerable homology in their primary sequence, but the bacterial protease subtilisin has a completely unrelated sequence, although the same residues are present at the active site in similar dispositions, with a serine residue being

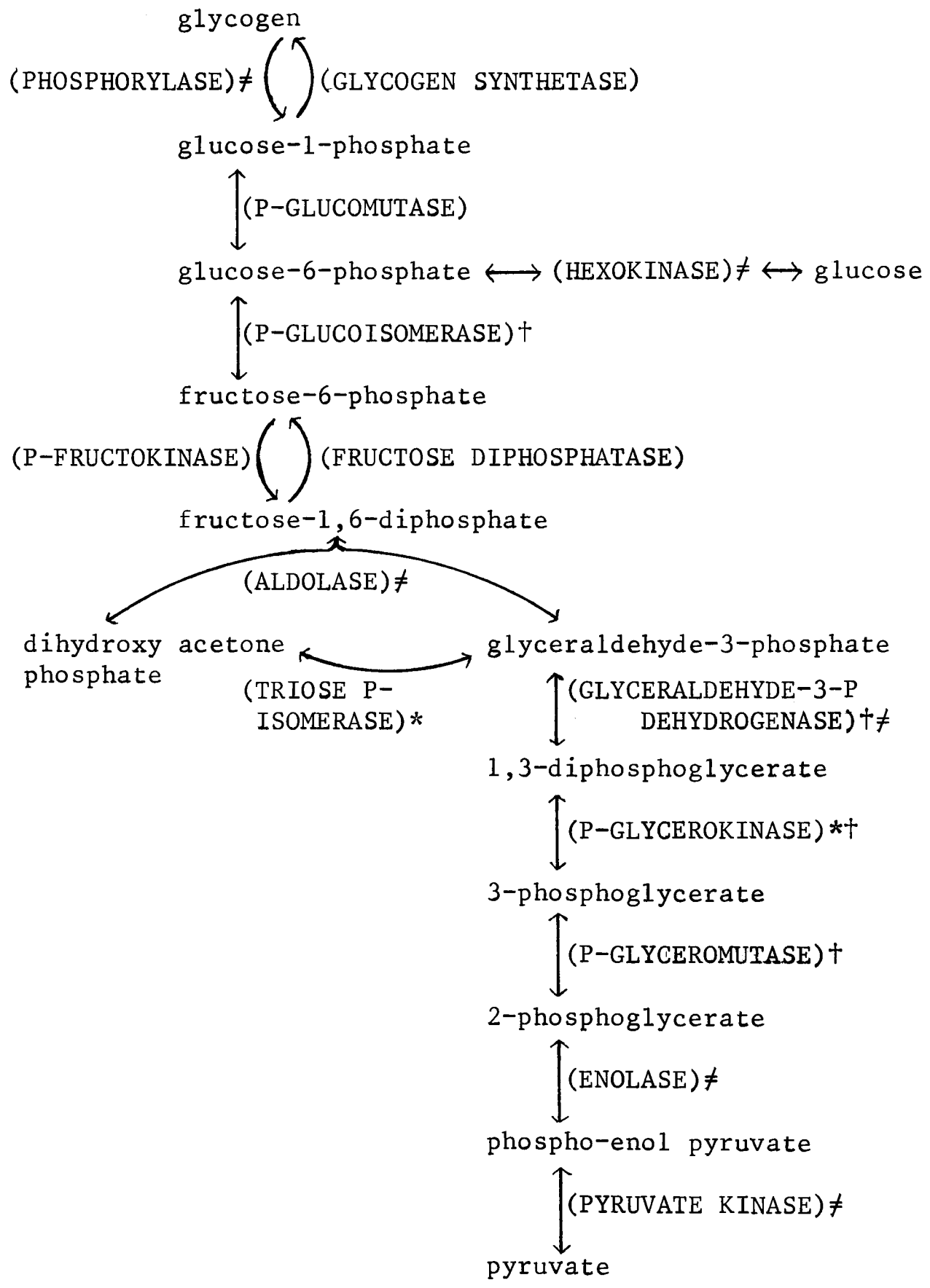
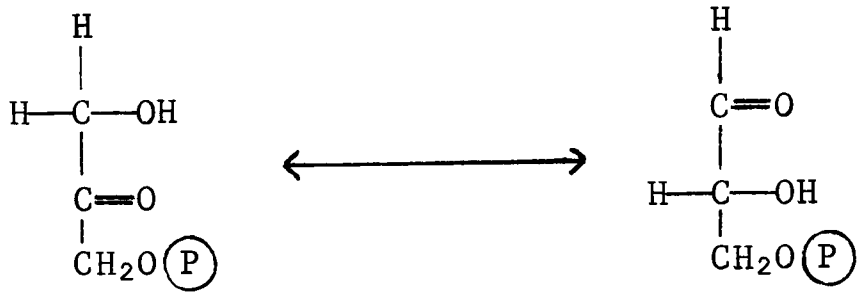


FIGURE 3 The Glycolytic Pathway.
The sequence of reactions by which glucose, glycogen and pyruvate are interconverted is shown, the enzymes involved being indicated in parentheses; low resolution electron density maps have been obtained for the enzymes marked * and † at Oxford and Bristol respectively; those marked ≠ are currently under investigation in these or other laboratories.

activated by a charge relay system as in these mammalian proteinases. This group of enzymes illustrates the two possibilities of molecular evolution, which can be divergent, for enzymes originating from a single precursor by gene duplication, or convergent, for enzymes adapting to fulfil a specific function, catalyzing a given reaction with the maximum efficiency.

Two major difficulties arise in any discussion of the kinetics of enzymatic reactions which is based on crystallographic results. Firstly the only crystallographic binding studies which are usually possible involve substrate analogues or inhibitors and not the true substrate, since the binding of this would be accompanied by reaction and consequent loss of the products, so that no ligand can be seen in the difference Fourier syntheses. The study of abortive complexes of an enzyme with various ligands avoids this problem, but yields results which may not be directly applicable to the true enzymatic reaction. Secondly it is clear that an understanding of the molecular architecture of an enzyme in the crystalline state can not provide a complete understanding of the importance of factors such as the localisation of an enzyme within a cell, its association with other cellular components, and the local concentrations of specific ligands, all of which may influence the catalytic efficiency of an enzyme.

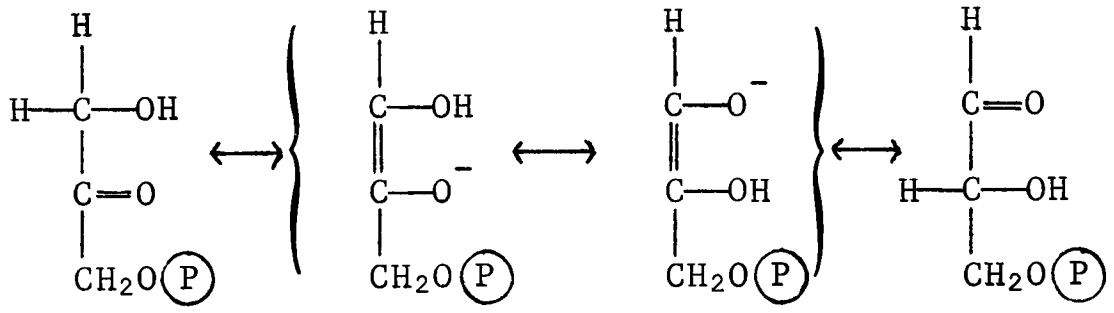
The latter difficulty can be partially overcome by the study of enzyme-enzyme complexes if these can be induced to crystallise. The cocrystallisation of a trypsin inhibitor protein with either chymotrypsin or trypsin (Rühlmann et al., 1971) indicates the possibilities of this approach. The study of a series of enzymes which are functionally related within a metabolic pathway and the investigation of their interactions with common ligands forms another approach towards



dihydroxy acetone phosphate

glyceraldehyde-3-phosphate

FIGURE 4 Reaction Catalysed by Triose Phosphate Isomerase (TIM).



DHAP Ene-diolate intermediates GAP

FIGURE 5 Proposed Mechanism of the TIM-catalysed Reaction.

the eventual elucidation of enzymatic relationships within a living cell.

The glycolytic pathway which is outlined in Figure 3 involves many enzymes in the conversion of the carbohydrates, glucose or glycogen, into smaller molecules which participate in other essential pathways. The degradation is accompanied by the synthesis of adenosine triphosphate, a molecule which provides chemical energy as necessary in other reactions by the cleavage of its terminal phosphate linkage. The importance of this pathway in the energy metabolism of a cell results in the glycolytic enzymes being present in the cells of vertebrate muscle tissue at relatively high concentrations so that the proteins can be extracted and purified with much greater ease than many other enzymes. Crystals of most of the enzymes have been obtained and are being studied as indicated in Figure 3.

Triose Phosphate Isomerase

The enzyme triose phosphate isomerase (E.C. 5.3.1.1), TIM occurs at the branch point in this pathway, catalysing the inter-conversion of dihydroxy acetone phosphate (DHAP) and D-glyceraldehyde-3-phosphate (GAP), the reaction shown in Figure 4. The mechanism proposed for the catalysis of this reaction by TIM involves an enediolate intermediate (Rieder and Rose, 1959) and the exchange of a proton with the surroundings as illustrated in Figure 5. A water molecule is not implicated, the enzyme interacting with a single substrate and product molecule. The equilibrium of the reaction (K_{eq} = 420 at 20°C) is strongly in favour of DHAP (Reynolds, Yates and Pogson, 1971) and so for this particular enzyme the problems arising from the necessity of using abortive complexes can be avoided, as the complex of TIM with the substrate DHAP would be expected to be

facing page 39

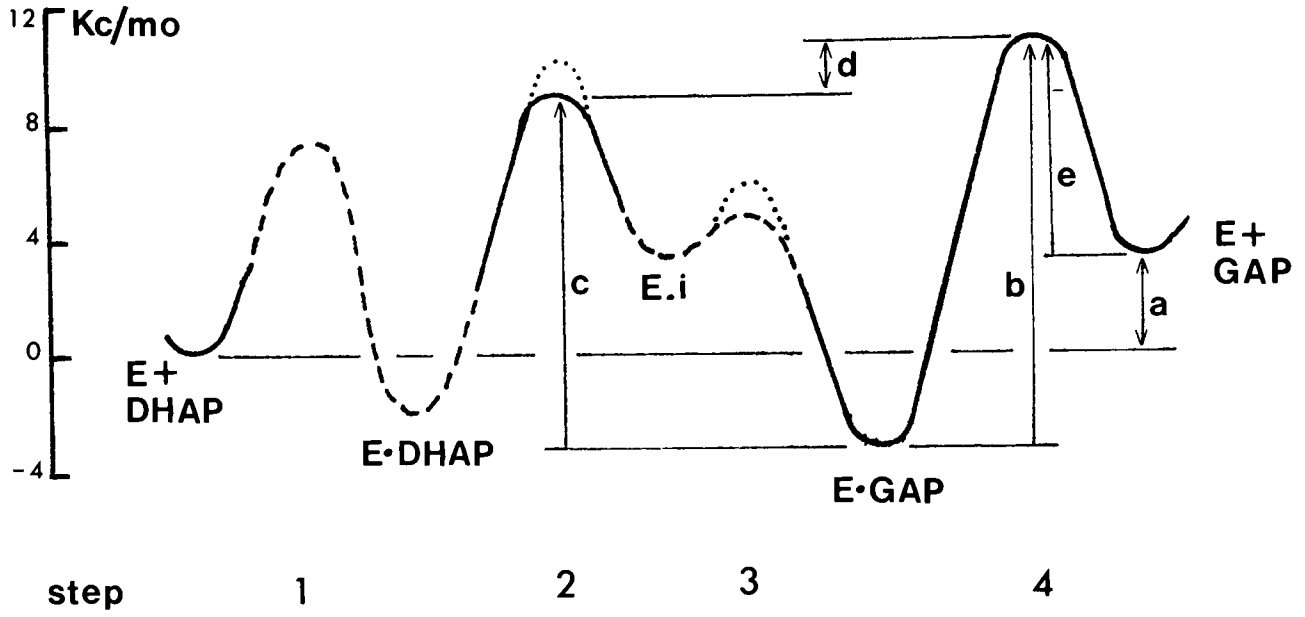


FIGURE 6 Postulated Free Energy Profile for the TIM-catalysed Reaction. Solid lines relate to experimentally quantified parts of the reaction scheme. Dashed lines relate to parts that are currently inaccessible. Dotted lines refer to the higher energy barriers for tritium (as distinct from hydrogen) abstraction processes. (For simplicity these are shown as transition state differences.)
Figure taken from Knowles, Leadley and Maister (1971).

observable crystallographically.

Preparations of the enzyme obtained from several species were studied electrophoretically by Scopes (1968), who demonstrated the existence of isozymes in the protein obtained from most species. The enzyme from chicken muscle exhibited only a single minor contaminant in addition to the major species and these preparations were considerably more homogeneous than those from rabbit muscle tissues. The preparation and properties of the enzyme from chicken muscle are described by Pogson (1972), together with details of its crystallisation on addition of ammonium sulphate to a solution of the enzyme in triethanolamine buffer at pH 7.4 in the presence of EDTA and mercaptoethanol at $4^{o}C$.

The earliest kinetic investigations of TIM used the enzyme obtained from rabbits, but have been accompanied more recently by those using the chicken enzyme. The simplicity of the reaction catalysed by TIM allows a very detailed analysis of its reaction kinetics. The postulated free energy profile for this reaction (Knowles, Leadlay and Maister, 1971) which is shown in Figure 6 indicates the progress which has been made in this direction.

The eventual determination of the structure of TIM at atomic resolution will provide a focus for the active collaboration of many who are involved in studying this enzyme and its mechanism, using a wide variety of techniques. Progress towards a knowledge of this structure forms the basis of this thesis, which includes the initial characterisation of crystals of chicken TIM and the subsequent analysis leading to the calculation of an electron density map at 2.5 Å resolution.

D) CRYSTALLOGRAPHIC STUDIES OF TRIOSE PHOSPHATE ISOMERASE

The overall strategy for any protein structure analysis has to be given detailed consideration as it almost invariably involves the energies of several people over an extended period of time, the intensive use of much costly scientific equipment and particularly because crystal structure determination is a technique in which the eventual success or failure of the analysis does not become apparent until the work is almost complete, although there are various indications concerning this at earlier stages. The most important factors in planning such a project are the nature and availability of the protein crystals and possible heavy atom isomorphous derivatives, the instruments available for the measurement of the diffraction pattern, the computing facilities for the subsequent analysis, and the methods of presenting and recording the structure when it has been determined.

All the crystals of chicken triose phosphate isomerase which were used in this work were a gift from Dr C.I. Pogson of the Department of Biochemistry, Bristol, who supplied additional crystals whenever they were requested, so that although suitable crystals were not locally available until quite recently, this did not present a problem. A further gift from Dr Pogson was the supply of two heavy atom derivatives which he had prepared by reacting TIM molecules in solution stoichiometrically with either a monomercurial or a dimercurial derivative of p-nitro-phenol, before crystallising the product. These derivatives are considered in more detail in Chapter III.

The very large number of diffraction intensities which must be measured for the analysis of macromolecules necessitates a highly automated system for the actual data collection. The last ten years have witnessed the commercial development of fully automatic diffrac-

tometers of two types, linear diffractometers, operating with a mechanical analogue to the reciprocal lattice, and four-circle diffractometers, controlled by a digital computer. These are both capable of measuring several hundred intensities or more without manual intervention. The speed and convenience of these machines eliminated the need for manual setting of either the earlier four-circle diffractometers or the optical densitometers then in use for measuring precession photographs. However, a major disadvantage of all diffractometers is that they can measure one, or at the most five, reflections at once, whereas many reciprocal lattice spots pass through the Ewald sphere almost simultaneously for structures having very large unit cells. This factor, coupled with the growing interest in studying ever larger molecules and the availability of computer-controlled optical microdensitometers has produced a renewed interest in photographic methods of data collection. The ability to index and record diffraction spots by computer has made the use of screenless precession or oscillation photographs a practical proposition for these structures. The design of automatic oscillation cameras has now made the efficiency of data collection by one of these methods comparable with that of multiple counter diffractometers for structures of about 50,000 Daltons per asymmetric unit. The photographic methods become more efficient for even larger structures (Arndt, 1968).

When the present structure analysis began in October 1968, the only fully automatic systems for data collection which were available in this laboratory were a linear diffractometer equipped with five counters (Arndt, North and Phillips, 1964) which is best suited for the collection of high resolution data, and a single counter four-circle diffractometer, which is the most efficient method of collecting low

resolution data, and also provides a convenient means of measuring selectively only the centrosymmetric projections. Diffraction photographs were used just for inspection and surveying of derivatives and crystals and so they were not densitometered.

The choice of automatic data collection by diffractometer was partly governed by an interest in maximising the extent of automation in the relatively routine stages of a protein structure analysis. The number of reflections involved in the case of fairly small proteins such as Lysozyme was such that reasonably close inspection of each measurement had been possible even if not desirable. An attempt was made during the present analysis to avoid the need for such inspection, which becomes quite impractical for such large numbers of reflections anyway, by incorporating a variety of statistical tests and editing facilities into the following computer analysis.

The Argus 500 computer which is available in this laboratory is designed for the on-line control of experiments in a time-sharing mode and for analyses requiring direct interaction with the operator. It is not particularly suitable for long calculations or handling large amounts of data, both of which are required in a crystal structure analysis. However, the University Computing Laboratory was expecting shortly to change its KDF-9 computer for a more powerful ICL 1906-A machine, a change which would involve extensive rewriting of all the programs. The futility of writing programs which would so soon be obsolete, coupled with the availability of sufficient time on the laboratory Argus led to the decision to use the latter machine for all the computing necessary for a low resolution analysis.

The Argus computer executes programs written in Fortran or lower level languages. It was originally supplied with only punched

paper tape input and output channels which made the handling of large quantities of data very time consuming. However, the subsequent addition of four magnetic tape decks, a graph plotter, a CRT visual display unit and additional blocks of store giving a total of 20,000 words of 24 bits each, has enabled all the high resolution analysis also to be carried out on this machine. This was not a desirable situation for the large amount of data involved but delays and initial problems at the University Computing Laboratory made it the most efficient way to complete the analysis at that particular time. The set of programs which were written for this were tailored to this machine as will be described especially in Chapter IV, and Appendix I.

The separate stages in the crystal structure determination of TIM at resolutions of 6Å and 2.5Å are described in the following pages. Initial characterisation of the crystals and measurement of the native intensities at low resolution comprise the next chapter. The various heavy atom derivatives which were used are described in Chapter III and the processing of all the high resolution data is recorded in Chapter IV. In the remaining chapters are considered various problems which arose in connection with the derivatives and the refinement of the heavy atom parameters. The final chapter contains an account of the results now available and some discussion of future possibilities and methods of presenting the three dimensional results which have been obtained.

In this introduction the background to, and the initial applications of, protein crystallography have been considered in relation to some of the major biochemical questions which it is hoped

will be eventually answered by the results of this and other techniques. No attempt has been made to cite all the original publications of the work referred to where these are readily available in text-books or reviews covering different aspects of molecular biophysics. A selection of such secondary sources is included at the beginning of the list of references.

Chapter II PRELIMINARY CRYSTALLOGRAPHIC STUDIES OF TRIOSE PHOSPHATE ISOMERASE

A) CRYSTALS OF RABBIT TIM 46

B) CHARACTERISATION OF CHICKEN TIM CRYSTALS 48

C) MORPHOLOGY OF CHICKEN TIM CRYSTALS 58

D) INTENSITY MEASUREMENT AT 6Å RESOLUTION 62

E) PROCESSING OF LOW RESOLUTION INTENSITIES 64

F) EVALUATION OF ROTATION FUNCTION OF CHICKEN TIM 67

A) CRYSTALS OF RABBIT TIM

The earliest X-ray diffraction photographs of TIM in this laboratory were taken by Dr L.N. Johnson using crystals of the enzyme obtained from rabbit muscle. The crystals were usually in the form of needles which were shown to belong to the space group $P2_1$ and estimated to contain about 50,000 Daltons per asymmetric unit, but some hexagonal bipyramid crystals were also obtained and shown to belong to the space group $P6_122$ or its enantiomorph $P6_522$. Measurement of the density and water content of these crystals indicated that the weight of protein per asymmetric unit was $26,203 \pm 3,000$ (Johnson and Waley, 1967). Estimates of the molecular weight of the enzyme by other techniques were all approximately twice this value. The hexagonal crystal form thus proves that the enzyme contains a two-fold axis of symmetry with two, or some multiple of two, identical subunits. The separation of five distinct peptides containing cysteine residues, coupled with an amino acid analysis indicating ten cysteine residues per enzyme molecule (Burton & Waley, 1968) showed that there could not be more than two identical subunits. The identification of only a single peptide species corresponding to each of the amino and carboxyl terminals of a polypeptide chain further suggested that the protein subunit having molecular weight approximately 26,000 Daltons contained a single polypeptide chain and that this enzyme is a dimer.

Crystal forms in which the asymmetric unit contains only a monomer of an oligomeric protein are in general the most suitable for analysis by the methods of X-ray diffraction, because they have a smaller unit cell with fewer diffracted intensities to be measured and the mean value of these is higher, so that they can be measured with greater precision. The only disadvantage of such forms of crystal is

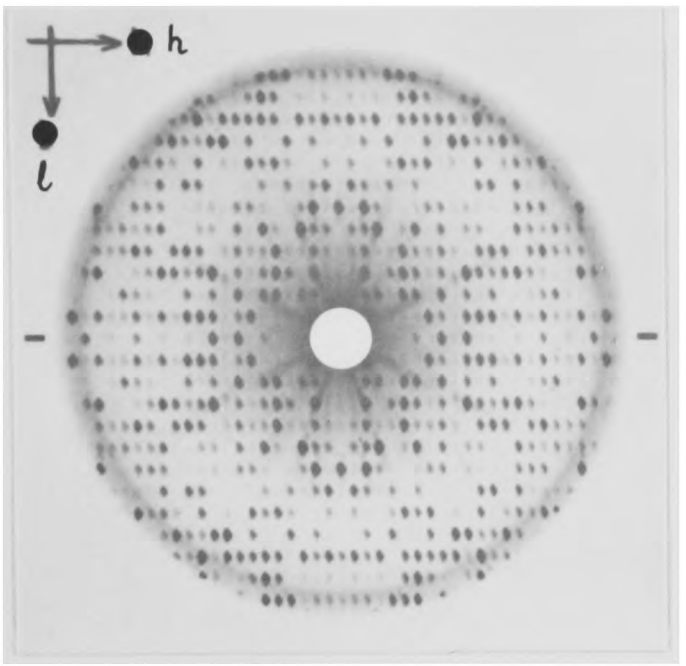
the impossibility of using the molecular symmetry to determine the phases of the structure factors, but the successful application of these methods has still to be demonstrated, and a dimeric protein is the case in which they are least powerful. The hexagonal form of the rabbit TIM crystals was also preferable to the needle-shaped monoclinic crystals for convenience of manipulating and mounting the crystals.

A survey of several possible heavy atom derivatives of these hexagonal crystals was carried out by Dr Johnson and Dr C.R. Beddell. Soaking the crystals in solutions of those reagents which would be expected to react with sulphydryl groups on the protein destroyed the crystals completely, even at very low concentrations. Potassium chloroplatinite produced large changes in the diffraction pattern at low angles, but considerable disordering at higher Bragg angles. Potassium chloroplatinate, cyanoplatinite and uranylfluoride all produced no intensity changes and gold chloride caused the crystals to turn orange, but again caused considerable disordering. The only derivative which looked even slightly promising was uranyl acetate, but an unambiguous and fully consistent interpretation of the differences which it produced was not possible for the two principal projections.

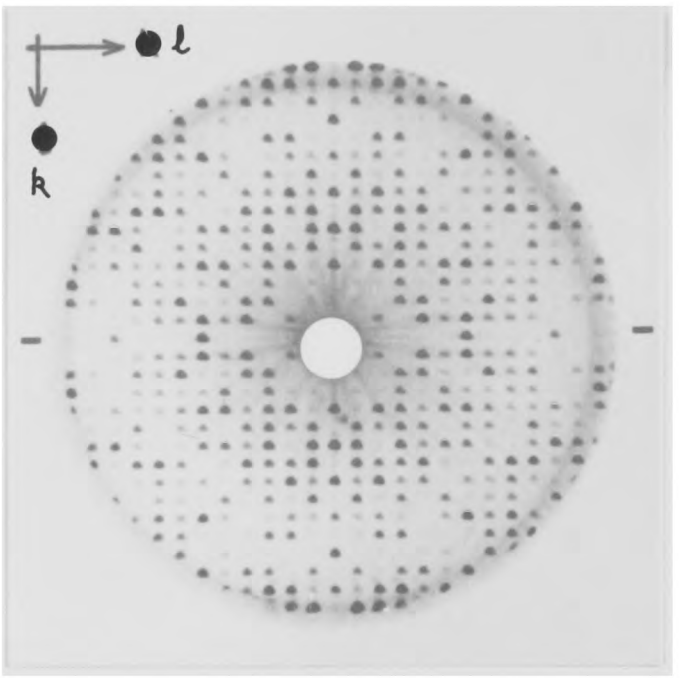
These difficulties and the electrophoretic inhomogeneity of preparations of rabbit TIM, which was mentioned in the previous chapter, suggested that attention should be concentrated on the crystals of chicken TIM which had been grown by Dr C.I. Pogson in Bristol. All of the crystallographic studies of TIM since the summer of 1968 have been concerned with these latter crystals, the transition marking the beginning of that stage of the project which forms the

facing page 48

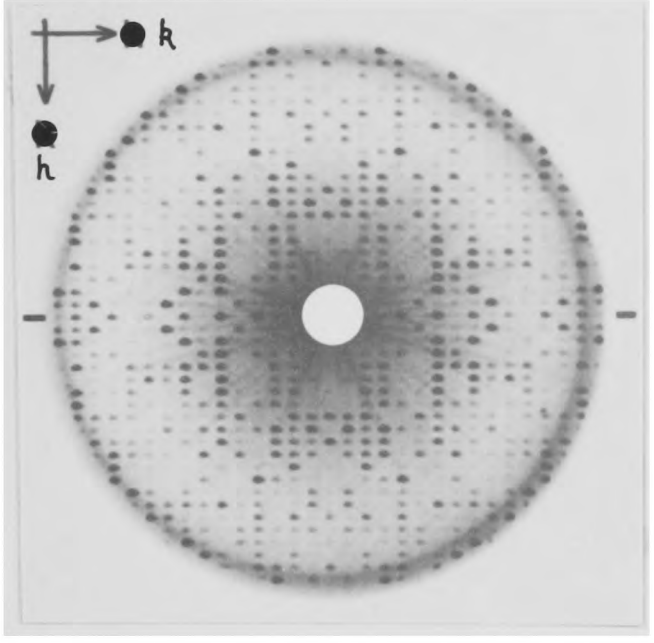
a) HOL



b) OKL



c) HKO



d) HKO (irradiated)

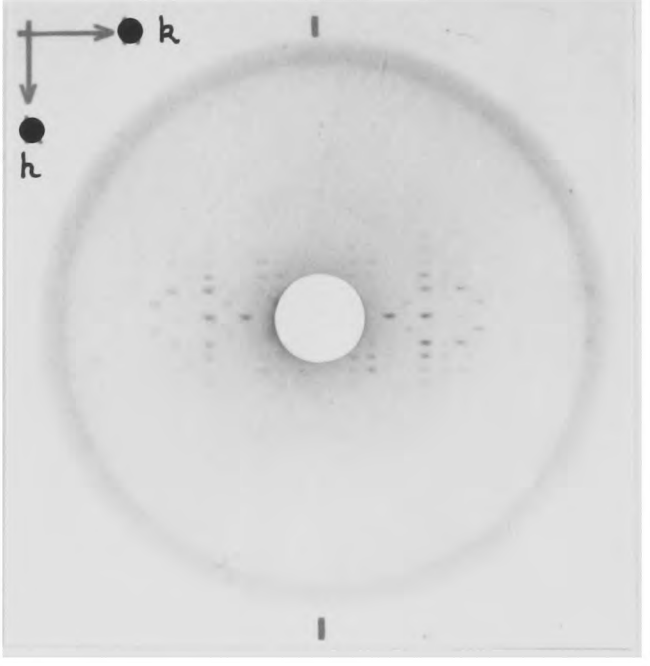


FIGURE 7 Diffraction Photographs of Chicken TIM. Precession photographs of the three principal projections taken with Cu K_α radiation with a precession angle of 9°; the horizontal axis of the crystal is indicated;
d) HKO projection after irradiation of the crystal for 140 hours by X rays from a fine focus tube running at 1200 watts.

subject of this thesis.

B) CHARACTERISATION OF CHICKEN TIM CRYSTALS

The earliest crystallisations of chicken TIM gave crystals without a regular morphology but of sufficient size (0.5 - 1.5 mm. in all directions) for them to be easily manipulated and mounted. Alignment of the crystals on the goniometer head was effected firstly on a polarising microscope by observing the extinction positions, and secondly on a precession camera, by means of both still and small angle screenless precession photographs.

The three principal projections of the diffraction pattern were recorded to a resolution of about 5Å in the 9^{o} precession photographs which are shown in Figure 7. The photographs of the hOl and hkO projections were taken by Dr Johnson who thence determined that the crystals belong to the orthorhombic system, having three mutually perpendicular but unequal axes, and the space group $P2_12_12_1$ in which there are two-fold screw axes of symmetry parallel to each of the crystallographic axes. This conclusion follows from the systematic absences which are present in Figure 7 where there is no intensity along any of the axial rows at any position corresponding to an odd value of the index. The dimensions of the unit cell were determined from these photographs but more accurate values were subsequently obtained by combining estimates from many different crystals to give the following results for the unit translations and volume of the cell in both real and reciprocal space, the estimated standard deviations being given in parenthesis:

a = 106.01(19) Å	a* = 0.014544(26) r.l.u.
b = 74.76(12)	b* = 0.020624(34)
c = 61.74(6)	c* = 0.024974(25)
V = 489308.5 Å	V* = $7.491.10^{-6}$

The density of the crystals was determined by means of a bromobenzene and xylene density gradient column (Low and Richards 1952). Quantities of bromobenzene, a 50:50 mixture, and xylene were pipetted in that order, in the ratio of 2:1:1 by volume into a glass boiling tube. A stirrer which fitted exactly inside the tube was plunged slowly in and out of the tube once to obtain a smoothly varying gradient. A crystal was transferred to the tip of a spatula and the accompanying salt solution quickly removed by touching the crystal with strips of filter paper before allowing it to fall slowly into the gradient column. This operation was completed in less than a minute in an attempt to minimise both any drying out of the crystal, which would increase its density, and also the amount of salt solution adhering to it when the crystal entered the column, which would have the converse effect. The crystal sank slowly through the column before coming to rest at a point whose vertical position was recorded by means of a scale attached to the glass tube. The column was then calibrated by similarly observing the equilibrium positions of a drop of each of several solutions of sodium bromide, whose specific gravities were separately determined and ranged from 1.20 to 1.34. Care was taken to form drops of these solutions of similar size to the crystals used, by choosing an appropriate syringe needle. The density gradient was usually very close to being linear and only such gradients were used for the actual measurements.

The best estimate of the density of these TIM crystals, obtained

from three crystals and two gradient columns, is

$$\rho = 1.245(3) \text{ gm.cm}^{-3}$$

The weight of protein which is contained within the asymmetric unit of the cell can not be calculated unless the amount of mother liquor or salt solution in the crystals is known. The volume occupied by such liquid is expected to remain constant, but its concentration can clearly fluctuate in different crystallisations, and so these measurements were obtained from the same batch of crystals as those used for the density determination.

After growth in a mother liquor of approximately 60% saturated ammonium sulphate, the crystals were routinely transferred to a fresh solution containing no dissolved protein, but with a higher salt content of 75% to prevent any dissolution of the crystals. They could be kept for many months in this solution and were normally so stored in a refrigerator at 4^{o}C for subsequent mounting and use on the X-ray sets.

The water content of the crystals was determined by weighing a crystal from which all the surface liquid had been removed, drying it to constant weight and then reweighing (North, 1959). The use of a micro-balance enabled this to be done with single crystals, which were left on the balance pan during drying. A small amount of phosphorus pentoxide was left in the balance cabinet as a dessicant, and the drying was complete after about five hours. Measurements on two crystals, the first being about ten times the volume of the second, gave values of 25.4(4)% and 20.6(1.0)% as the weight fraction of the original crystal which had been lost as water. The discrepancy between these values indicates the sensitivity of the result to the precise state of the crystal with respect to any residual surface liquid and

the onset of evaporation. The best estimate of the fraction of unbound water is thus 24.6 $\pm$ 3 %. The weight fraction of the water contained in the mother liquor was similarly determined, by drying to constant weight a drop of the liquid, the result of two independent measurements being 35.9(2)%.

Some of the water present in a protein crystal is firmly bound to the surface of the protein molecules and can not be removed by air drying (Haurowitz, 1950) as can that which moves freely in the interstitial channels of the crystal. Determination of the exact amount of such bound water is difficult, but the protein molecular weight can be simply determined as a function of this unknown quantity as follows:

$$\text{Wt. fraction of salt + unbound water} = \frac{\text{Wt. of unbound water}}{\text{Wt. fraction of salt in mother liquor}}$$

$$= \frac{24.6}{1-0.359} \ \%$$

$$= 38.4 \ \%$$

$$\text{Wt. fraction of protein + bound water} = 61.6 \pm 5.\%$$

If the weight fractions of the bound water and protein plus bound water are denoted α and β respectively, Avogadro's number N, the number of asymmetric units within the unit cell Z, the cell volume V and the crystal density ρ then the weight of protein M per asymmetric unit is given by

$$M = \frac{\rho \, N \, V \, \beta}{Z(1+\alpha)}$$

$$= \frac{1{\cdot}245. \ 6{\cdot}023. \ 10^{23} . \ 4{\cdot}89. \ 10^{-24} . \ 0{\cdot}616}{4(1+\alpha)}$$

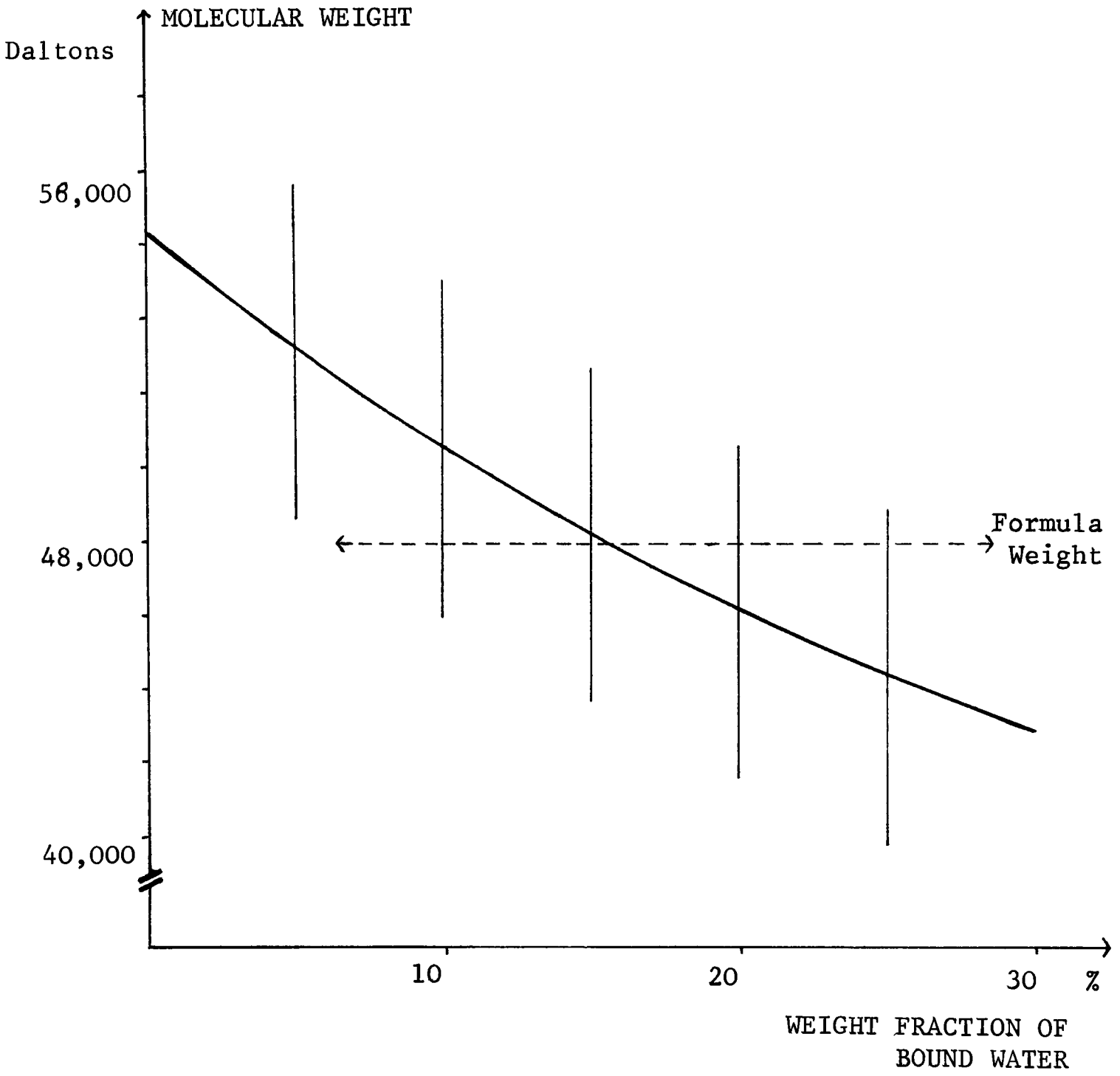


FIGURE 8 Molecular Weight of Chicken TIM as Determined Crystallographically. The Molecular Weight is expressed as a function of the amount of bound water

$$= \frac{5{\cdot}64 \;.\; 10^4}{(1 + \alpha)} \qquad \text{Daltons}.$$

The estimated standard deviation of this quantity is 4,500 Daltons. Comparison with the results of the previous section indicates that in these crystals there is one dimeric molecule of the enzyme per asymmetric unit of the crystal. The molecular weight of chicken TIM as determined from its amino acid composition is a multiple of 24,000 Daltons (Dr R.E. Offord *et al.*, unpublished results) suggesting that there are $(8.4 \pm 2.3).10^3$ Daltons of bound water per dimeric molecule representing $(17 \pm 8)\%$ of the weight of protein as can be seen from Figure 8. This corresponds to 220 molecules of water bound per protein monomer, although any number between 100 and 350 would not be inconsistent with these results. It remains to be seen how many such molecules can be identified on a high resolution electron density map of this molecule.

Structural differences between the two monomers will also be of especial interest in these crystals, since it is known from the rabbit crystals, that any such differences result from an equilibrium between the forces of the crystal lattice and the intramolecular forces themselves, and are not necessarily an inherent part of the structure of the dimer.

The feasibility of a high resolution structural analysis of these chicken TIM crystals depends largely on the extent of their diffracting power and their susceptibility to radiation damage. Inspection of the still photographs taken during alignment of the crystals indicated that the intensities were still significant at the edges of the films, these angles of diffraction corresponding to interplanar spacings of about 2Å. The diffraction spots were usually

fairly regular in shape, except from crystals which were obviously badly cracked. There has been no evidence of any form of twinning in these crystals.

Photographic methods of recording X-ray intensities provide a time-averaged result from which precise deductions about the nature of any radiation damage are impossible. The overall intensity diffracted by crystals which had been very heavily irradiated was observed to decrease in agreement with earlier results such as those obtained with sperm-whale myoglobin by Blake and Phillips (1962) who showed that an irradiated crystal could be regarded as containing a completely amorphous part, which gave no contribution to the crystalline diffraction pattern since all traces of an ordered structure had been destroyed, a strongly disordered part which contributed to the low angle diffraction pattern only and an unchanged part, where the molecules continued to diffract as if they were fully ordered. The relative sizes of these three volumes within the crystal were observed to change with increasing irradiation and there were also significant changes in the relative intensities of reflections. They found that these effects varied smoothly with the exposure, in contrast to the studies of Traub and Hirshfeld (1960) on crystals of insulin where discontinuous effects were observed. In both of these investigations the intensities were measured on a diffractometer, providing more information about temporal variations than is possible by photographic recording.

No significant loss of intensity was detected from crystals of chicken TIM which had been irradiated for only about 40 hours by X rays from a fine focus sealed tube running at a power of 1200 watts. Exposures larger than this produced noticeable effects. A precession

photograph of the hkO projection from a crystal which had already been irradiated for 140 hours by such an X-ray beam is shown in Figure 7(d). This shows that the loss of intensity is anisotropic, there being intensity along the b axis extending at least as far as the edge of the film, corresponding to spacings of 5Å, but no intensity for values of the h index greater than 7, which represents spacings of 15Å. A photograph of the Okl projection from the same crystal, after 100 hours exposure showed no intensity along the c axis for spacings less than 13Å, and a photograph of the [101] zone after 60 hours exposure had no intensity beyond the 12Å position in the direction perpendicular to the b axis. This orientation of the anisotropic effect suggests that the protein molecules are constrained much more strongly along the b axis than in the plane perpendicular to it.

In addition to this overall anisotropic loss of diffracting power, larger relative changes between structure factors are to be expected. These are most easily detected by repeated diffractometer measurements, since these do not produce a time-averaged result. Evaluation of a correction for this latter effect would need to be done for each reflection individually, extrapolating repeated measurements back to zero exposure, a procedure which would be extremely laborious. The alternative which is commonly adopted, and was followed throughout these studies, is to measure a set of reference reflections at intervals during the course of data collection on the diffractometer. The fall-off of these intensities is a monitor for that of the crystal as a whole. Data collection from any crystal was discontinued if the loss of intensity exceeded about 10%, an overall correction could be applied for such a loss, and any relative changes were thought to be minimised by keeping the crystal exposures below this level. An exposure

of about 40 hours per crystal is reasonably economic of crystals, so that the current availability of these, coupled with knowledge of their diffraction characteristics made the undertaking of a high resolution structural analysis a reasonable proposition.

The resolution to which a structure analysis is carried is defined by the extent of the data which is collected, and this determines the degree of detail which can be seen in the final image of the structure. It has been shown by James (1948) that the smallest detail able to be resolved in a three dimensional image, which is reconstructed from all the rays scattered by less than a given angle $2\theta_{max}$, corresponding to an interplanar spacing d_m, is given by

$$r = 0.71\ d_m = \frac{0.71\ \lambda}{2\mathrm{Sin}\,\theta_{max}}$$

which is a slightly lower resolution than the result of classical optics for two dimensional images, but the corresponding diffraction ripples are also smaller. However, in crystallographic analyses the effective resolution may be considerably less than the theoretical one, depending on the importance of errors in both the determination of the phases and the measurement of the structure amplitudes. The effective resolution is thus unknown until the electron density map is examined, but the convention normally adopted by crystallographers, that of describing the resolution in terms of the minimum inter-planar spacing d_m, will be adhered to here, amplified as necessary with comments on the apparent resolution.

The number of intensities which must be recorded, and consequently the subsequent data manipulation and processing, to include all those within the sphere of reciprocal space corresponding to a

particular resolution increases as the third power of the resolution, being approximately a quarter of a million for these crystals of chicken TIM at 1Å resolution. Protein structure analyses are thus usually completed in three successive stages at increasing resolutions, the first two of these being described in the current work.

The low resolution analysis of TIM extended to 6Å and involved 1370 independent reflections. The choice of this resolution was governed by both the information it was hoped to obtain at an early stage of the analysis and also the well known minimum in the angular distribution of the average intensities in this region (Bodo et al., 1959). This effectively minimises any ripples in the image which result from a sharp termination of the Fourier series without necessitating the application of an artificial temperature factor to the data to avoid them. At higher resolutions, beyond about 3.5Å, the fall-off in the average intensities causes such ripples to be less of a problem than the errors already mentioned as affecting the apparent resolution.

Experience of earlier analyses suggests that at 6Å resolution one can hope to delineate the molecular boundary, indicating the packing within the crystal, and to detect any helical regions of the polypeptide chain, but that one would not expect to be able to follow the course of chain in its entirety. The positions of heavy atom binding can be examined with respect to the molecular surface and contacts, and difference Fourier methods can be used successfully at this resolution to reveal ligand binding sites. However, the most important information to be derived from such a low resolution analysis is that concerning the quality of the phase determination and the prospect of extending the results to higher resolution.

The second stage of this analysis has produced an image of the structure at 2.5Å resolution involving 17,530 independent reflections. Although this is much less than atomic resolution it was hoped that this image would show clearly the course of the polypeptide backbone and in conjunction with a partial amino acid sequence allow the identification of individual residues. Until the analysis of this electron density map is complete, and the questions then raised by a knowledge of the structure have been formulated, the directions to be emphasised in any future work can not be decided. Possibilities at that stage would include increasing the clarity of the image, by the use of more or better isomorphous derivatives in the phase determination, extending the theoretical resolution by including more data, and analysing the binding of ligands at specific sites on the enzyme. High resolution analyses of protein structures, extending the results even beyond 2Å are necessary before the classical Fourier refinement techniques for small structures can be applied successfully, but the validity of such an approach in the case of protein structures has been demonstrated by Watenpaugh et al. (1971) in their analysis of rubredoxin.

The two electron density maps which have been calculated for chicken TIM at 6A and 2.5A resolution are considered together in Chapter VI. The data collection and processing was done in two distinct parts and so is discussed separately, consideration of the low resolution stage appearing later in this chapter and that of the high resolution data in Chapter IV.

facing page 58

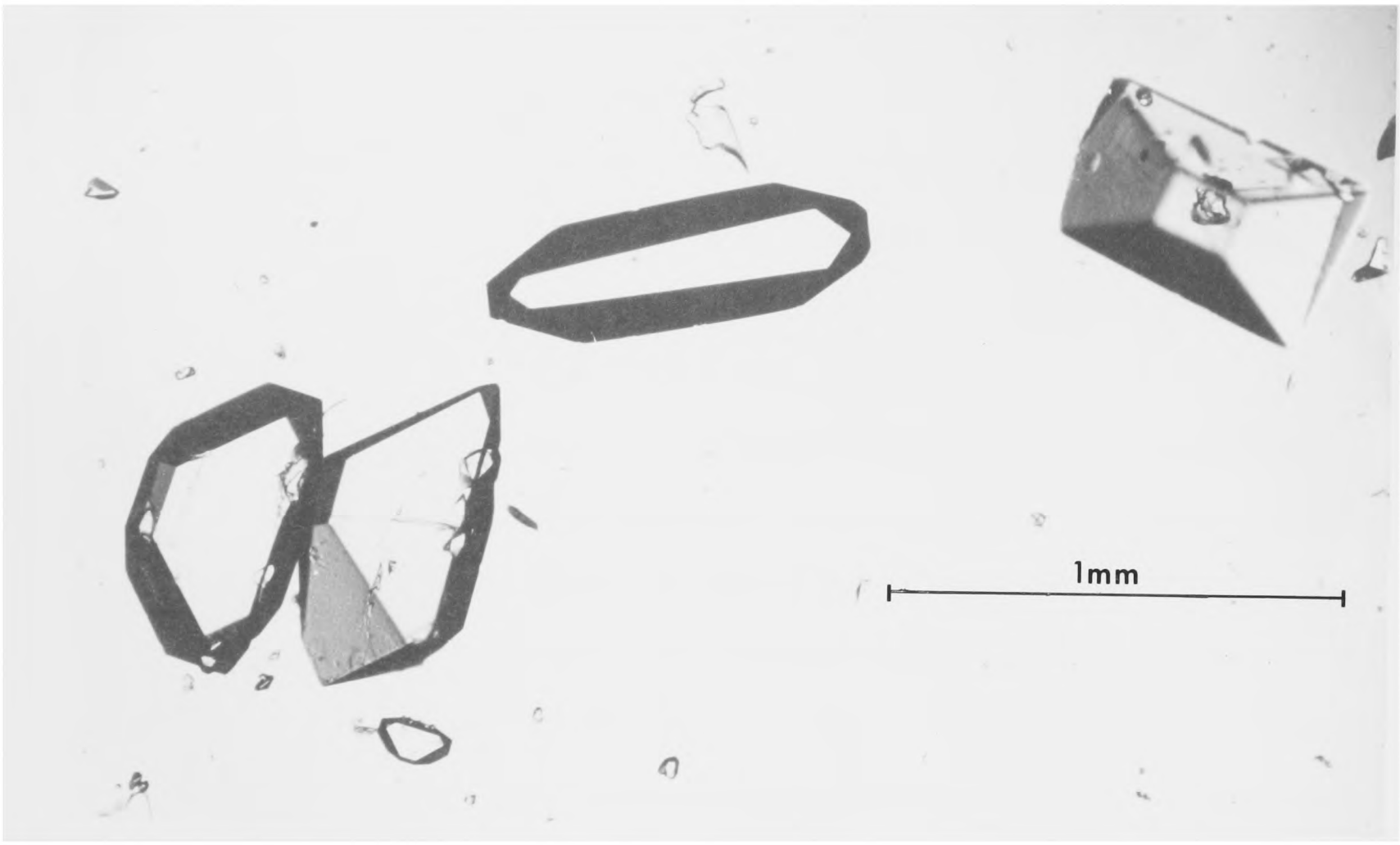


FIGURE 9 Crystals of Chicken TIM.

C) MORPHOLOGY OF CHICKEN TIM CRYSTALS

The morphology of different preparations of crystals varies considerably in quality. The earliest crystals had no identifiable faces, but more recently the habit has been recognisable as a series of standard forms. Growth on the sides or base of a glass vial produces distortions of these but the crystal faces which can be identified include those of the forms {hk0} and {h0l} which are always present, and {100} which is observed much less frequently.

The variety of shape and the optical quality of even carefully selected crystals can be seen from Figure 9. The longest dimension of the crystal may be parallel to its y or z axis, this tending to be consistent within a single preparation but varying between preparations and also between the native crystal and co-crystallised sulphydryl derivatives.

Identification of the faces present in the two prismatic forms requires accurate measurement of the interfacial angles. This is difficult because of the poor quality of these faces and the necessity of keeping the crystal inside a capillary tube, which can cause systematic errors in the determination of these angles. However, it is clear that the angles between the (hk0) and (h0l) faces and the (100) face are respectively greater than and less than 45^{o}. Estimates of these angles as 57^{o} and 33^{o} ($\pm$ 3^{o}) are to be compared with the values of 59.5^{o} and 34.8^{o} expected for faces of forms {560} and {502} respectively, although the presence of forms {110} and {301} would not be inconsistent with these results.

The morphology of a crystal must be related to its internal structure, but predictions about one of these properties from a knowledge of the other are very difficult. The existence of cleavage planes, and

their orientation, can usually be explained in terms of the underlying structure, but the theoretical determination of the geometrical form of a molecular crystal has only recently been achieved for the case of anthracene by Kitaigorodsky and Ahmed (1972), from detailed calculations of surface energies in terms of the atom-atom potentials in the known structure.

A problem of greater interest to protein crystallographers is the deduction of information concerning the structure from the observed morphology. This information can only be expected to concern the overall shape and molecular packing of a protein species, but the growing interest in polymeric systems, enzyme-enzyme interactions and other macromolecular complexes having well defined structures, suggests that even this type of structural information will prove useful. Detailed analysis of the morphology is possible as soon as the unit cell parameters and space group have been determined. This aspect of the chicken TIM crystals was not studied until after the low resolution electron density map was calculated, but the molecular packing is discussed here to enable the conclusions to be evaluated on their own merit without the support of other evidence which will later be shown to corroborate some but contradict others of these conclusions.

In a review of the whole subject of molecular packing Kitaigorodsky (1970) discusses both geometrical and physical models of such packing, the former being more applicable to macromolecules of unknown internal structure. His statement that "of all conceivable structures, the actual structure is the one most closely-packed" (loc. cit., p.177) supported by evidence that it is accurate to within 1% of the packing coefficient and that the space groups which allow

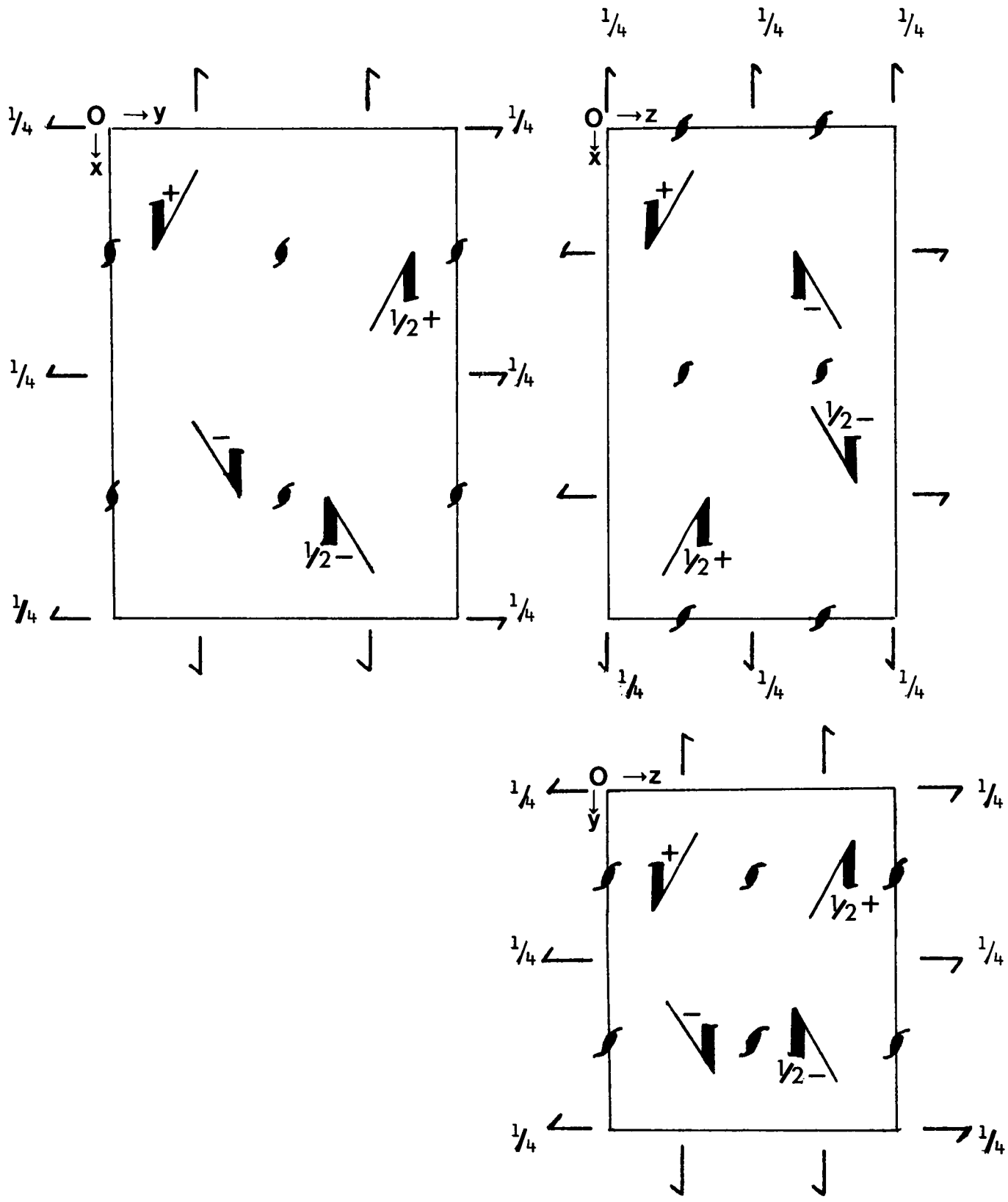


FIGURE 10 Symmetry Elements Present in Crystals of Chicken TIM. A complete unit cell and one set of general equivalent positions are shown for each of the principal projections; the horizontal screw diad axes are at heights 0 and 1/2 unless otherwise indicated as being at 1/4 and 3/4.

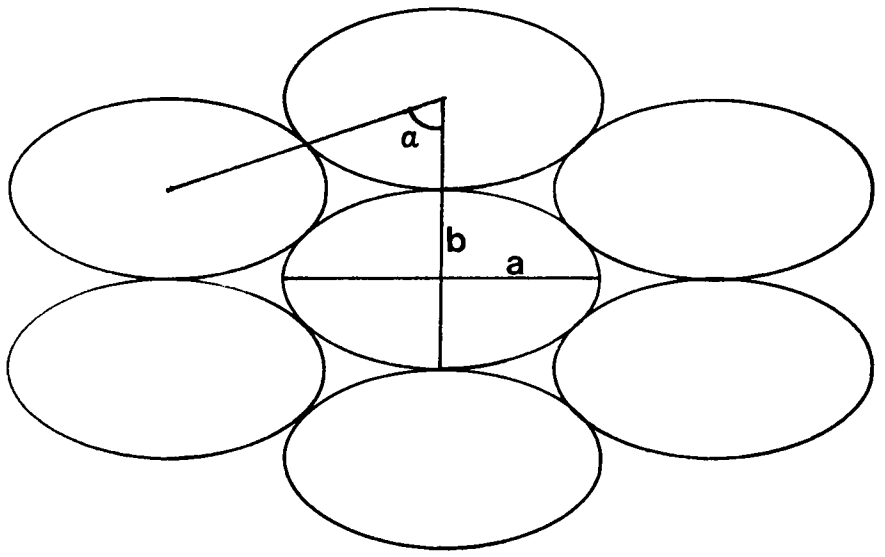
the closest approach to true close-packing are $P2_1/c$, $P2_1$ and $P2_12_12_1$ correlates well with the observed distribution between possible space groups of all the known protein crystals, $P2_1$ and $P2_12_12_1$ being by far the most commonly occurring, together with $P2_12_12$.

The principle enunciated by Harker and Donnay (1937) that the faces of a crystal correspond to the most densely packed planes in its structure was developed for atomic and ionic crystals. Its equivalent for molecular crystals would seem to be that the faces correspond to those planes which follow most closely the surface of the molecular envelopes as they are packed within the crystal lattice. This could be readily tested for any given crystal form once the molecular envelope is known, but the reverse calculation does not seem easily capable of general formulation, so this analysis of the chicken TIM crystals is necessarily superficial.

Assuming a value of 0.74 $ml.gm^{-1}$ for the partial specific volume of TIM determines the diameter of the sphere equivalent to a monomeric subunit of the enzyme as about 38Å. Thus a first approximation to the shape of the dimer would be a prolate ellipsoid of revolution having principal axes of 70Å and 38Å. If these are regarded as being packed in sheets parallel to the faces of the three forms observed in the crystals, some restrictions on the freedom of orientation of the ellipsoid become apparent.

The presence of the two faces of the form {100} suggests that the molecules are arranged in sheets parallel to the yz plane. The long axis of the ellipsoid can not be parallel to the z axis, since the unit translation in this direction is only 62Å, and if it is inclined to the y axis by more than about 20° then its position must be such that the screw diads parallel to the z axis do not pass through

facing page 61



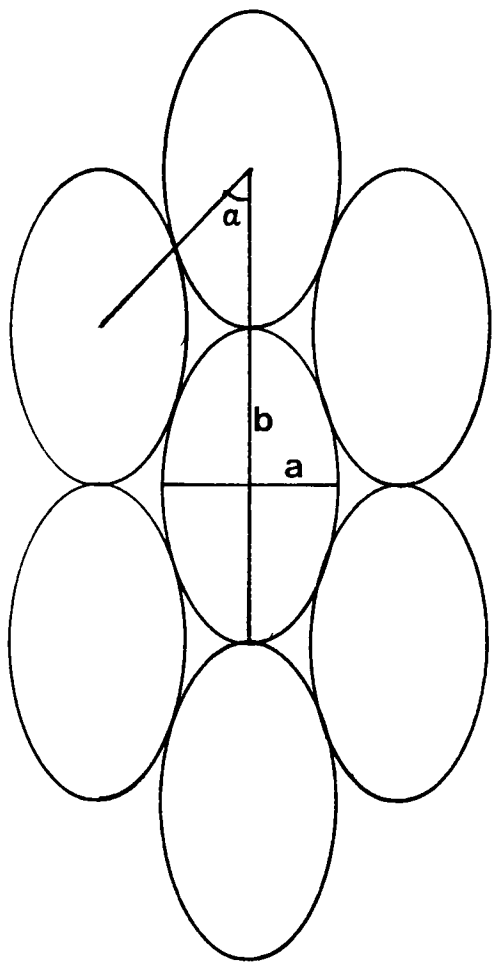


FIGURE 11 Two Dimensional Packing of Ellipsoids.
If the angle α is less than 60^{o} then the axial ratio of the allipse (b/a) must exceed unity.

its centre. Similarly if it is inclined to the y axis by less than about 70^{o} then it must not lie centrally on the screw diads parallel to the y axis. These conclusions follow from the disposition of symmetry elements within the unit cell as seen in Figure 10.

The presence of the form provisionally identified as {560} provides additional information. Since it is expected that the molecules are stacked in sheets perpendicular to x, then the observation that the angle between faces (560) and (100) is less than 60^{o} implies that the projection of the molecule on to the y axis is longer than that along x as is seen from Figure 11. This determines that the long axis of the molecule must be closer to the yz plane than to the x axis, and the earlier observations suggest that it lies closer to the y axis within this plane. The non-crystallographic diad axis of the molecule will thus lie closer to the xz plane than to either of the other principal planes.

Similar reasoning applied to the {502} form indicates that the projection along the z axis is larger than that along the x axis. However all the foregoing conclusions are based on an ellipsoidal molecule and are valid only to the extent that the molecule has approximately this shape. Further information about its structure requires the measurement of the diffracted intensities, since although the positions of these are determined by the size, shape and symmetry of the unit cell, the atomic arrangement within the cell is related to the distribution of intensity between the reflections.

D) INTENSITY MEASUREMENT AT 6Å RESOLUTION

All of the low resolution measurements which were used in the final analysis were collected on a Hilger-Watts four-circle diffractometer, which was then controlled by means of punched paper tapes specifying all the angular settings. These tapes were produced on the Argus computer by a program written by D.W. Banner and read by a control unit to which the diffractometer had been interfaced. This has now been replaced by an interface directly to the computer, and the diffractometer modified to incorporate a fifth circle with additional counters, allowing the simultaneous measurement of five reflections with on-line control by the computer (Banner, 1972).

The X rays were produced by a fine focus sealed tube with a copper anode which was operated at about 1200 watts. A nickel filter was used to obtain selectively the K_α rays of wavelength 1.5418Å. Reflections were measured in the equi-inclination setting by scanning in a series of discrete steps along the ω-circle thus moving the reciprocal lattice vector through the Ewald sphere within the plane containing both the incident and diffracted rays, and measuring the background at either end of the scan. The output signals from the scintillation counter were subjected to pulse height discrimination prior to conversion into digital intensities. The information recorded for each reflection, on punched tape and teleprinter output consisted of the three indices, two background measurements and the sum of the ordinates of the reflection intensity profile.

The crystals were aligned on the diffractometer with a reciprocal lattice axis parallel to the axis of the ϕ-circle, allowing an axial reflection to be used for the semi-empirical determination of the absorption correction (North, Phillips and Matthews, 1968).

Measurements of such a reflection were made at intervals of 15^{o} around the ϕ-circle, the complete absorption profile being obtained by interpolation. The absorption curve was measured at the beginning and end of the data collection. Three reference reflections were monitored after every fiftieth measurement cycle and the angular settings of the diffractometer were checked with respect to machine reference points at the same times. In no case did the reference reflections show a decrease in intensity sufficient to require the use of a correction for losses due to radiation damage. Other fluctuations in the intensities did not seem significant.

The information available in measurements of the anomalous scattering of the heavy atom derivatives (Bijvoet, 1954) and an improvement in the statistical quality of the data were obtained by routinely making two measurements of most of the reflections. This was accomplished by measuring two adjacent octants of the reciprocal sphere. The octant containing the indices hkl was always measured, the second one being either $h\bar{k}l$ or $hk\bar{l}$. If the negative index is that corresponding to the crystal axis about which it is mounted, then the members of the Bijvoet pair for each reflection are both subject to the same absorption correction. This was specifically chosen in the later sets of data but unfortunately not in the earlier ones. The order of measuring the reflections was such that the two equivalent reflections were separated by a short time interval. This was achieved by progressing along one row of reciprocal space from positive to negative values of the equivalence index, whilst keeping the other two indices constant, then incrementing whichever of the indices k and l was not the equivalence index and measuring the next row, and when that level of reflections had all been measured increasing the value of h by unity and beginning the next level.

Further details of the methods used are specific to a particular set of measurements. Those for the set of data from the native crystals are described here and those for each of the heavy atom derivatives are in the following chapter which also includes a tabular summary of all these particulars.

The native TIM crystal was aligned about the b* reciprocal axis and reflection 060 used for the absorption measurements. The highest point on the absorption curve had a value of 1.42. The sizes of collimator used were 1.0 mm. in the source aperture and 3.5 mm. in the detector aperture. For the 2558 measurements made, in the octants hkl and $h\bar{k}l$ of reciprocal space, each background was counted for 10 seconds. The 50 steps through each reflection, covering a range of 1^{o}, represented a total of 30 seconds counting. Reflections were recorded at the rate of about 50 per hour, the entire set being completed in about two days.

E) PROCESSING OF LOW RESOLUTION INTENSITIES

The reduction of a set of diffractometer measurements of background and integrated peak intensities to the corresponding relative structure factors, followed by the conversion of these to some fixed scale and an analysis of the quality of the data, requires a series of independent calculations. The division of these into particular programs and the individual authorship of these is listed in Appendix I, together with descriptions and acknowledgements of the programs used for the subsequent computations.

Primary processing of each reflection involved the following calculations. The net intensity I was determined by subtracting the

background counts I_1, I_1' from the integrated peak I_2 making due allowance for the particular counting times t_1, t_2. The standard deviation of this quantity σ_I was determined according to the theory of combination of errors assuming the counter to have obeyed Poisson statistics and neglecting all other sources of error. If

$$I = I_2 - \frac{t_2}{2t_1}\left(I_1 + I_1'\right)$$

then

$$\sigma_I^2 = I_2 + \frac{(t_2)^2}{(2t_1)^2}\left(I_1 + I_1'\right) \quad .$$

Whenever the net intensity of a reflection was negative it was set equal to zero. A measurement cycle was rejected if the difference between the two background measurements exceeded four times its standard deviation. The number of reflections falling into these two categories was recorded, as was the number of those where the intensity was less than its standard deviation. For the native data there were 147 measured intensities less than their standard deviations, 48 of them being negative, and 2 measurements were rejected because of unequal backgrounds out of a total of 2558.

The net intensity and its standard deviation were multiplied by the reciprocals of the Lorentz and polarisation factors and the transmission coefficient to give the corrected intensity. For the equi-inclination setting in which the diffractometer was operated the Lorentz factor can be shown to be

$$L = \frac{1}{\sin 2\theta}$$

and the polarisation factor is always given by

$$p = \frac{1 + \cos^2 2\theta}{2}$$

where θ is the Bragg angle (Arndt and Willis, 1966). The transmission

coefficient was calculated by the method of North, Phillips and Matthews (1968).

The square root of the corrected intensity, which is the relative structure amplitude, was punched on the output data tape, together with its standard deviation, after both of these quantities had been multiplied by ten to avoid serious rounding errors in treating these numbers as integers.

All of the above calculations, and sorting of the order of the reflections were done by a single program which was written by Dr A.C.T. North (North,1964).

Secondary processing of the data, by a program written by D.W. Banner, involved calculation of the weighted mean structure amplitude $\bar{F}$ from the two values of a pair of Friedel equivalent reflections, the difference between them ΔF, and the average values of these quantities for each row of reflections. The native data would not be expected to show any difference between the equivalents and the overall value for these 1358 reflections of which 1169 had valid measurements of both equivalents is

$$\frac{\langle |\Delta F| \rangle}{\langle \bar{F} \rangle} = 4.3(\pm\ 4.7)\ \%.$$

The relative structure factors for this set of native data were regarded as being on a standard scale, all sets of derivative data being scaled to it. However, for ease of correlation with the high resolution data, these native structure factors were subsequently scaled up by 1.32 and this revised standard scale will be the one quoted throughout this thesis. The derivation of that factor and the relation between the revised standard scale and an absolute scale is discussed with the high resolution data processing. The derivative

data were scaled to the native set by means of an overall scale factor, calculated such that the average derivative structure amplitude exceeded by 1.02 that for the native data. This procedure appeared to be satisfactory at low resolution, but much more serious consideration had to be given to it at high resolution.

F) EVALUATION OF ROTATION FUNCTION OF CHICKEN TIM

Evidence from crystals of rabbit TIM suggests that the asymmetric unit of the chicken TIM crystals contains two protein subunits having essentially identical electron density distributions $(\underline{x})$, $(\underline{x}^1)$ and being related by a two-fold rotation axis. This implies that the integral

$$\int_U \rho(\underline{x})\,\rho(\underline{x}^1)\,dU$$

over the volume of the molecule U will have a maximum value for the particular rotation matrix $[C]$ and translation vector $\underline{D}$ which relate corresponding points in the two subunits by

$$\underline{x}^1 = [C]\underline{x} + \underline{D}.$$

The rotation and translation parameters for which this overlap integral is a maximum will also give a maximum value of the corresponding convolution of the structure factors.

A convolution of the intensities is more readily calculable, since no information about the phases is necessary. Rossman and Blow (1962) have derived the precise form of this function and shown that its maximum value corresponds to the maximum overlap of the Patterson function with itself, after rotation by the matrix $[C]$, within the

sphere of integration centred at the origin of Patterson space. This rotation function as it is now generally referred to can be used for comparing a Patterson function with itself or with that of another crystal form or related structure. Only the first of these applications is considered here.

The volume of Patterson space over which the integral is to be evaluated requires careful consideration. Most of the features lying close to the origin will arise from intra-molecular, or in this case intra-subunit, vectors, but further from the origin the proportion of inter-molecular and inter-subunit vectors represented will increase. Clearly only the former set of vectors will contribute to maximising the overlap and so the volume of integration must be small enough for these to predominate, and yet large enough to contain sufficient features to determine the maximum.

The unit cell of chicken TIM crystals has dimensions 106.0, 74.8, 61.7 Å and so a radius of 35Å was chosen for the integration. All the reflections in the range 6-10Å were used to calculate the first Patterson function, but only the 250 strongest intensities were used for the second, interpolation being done on a 5x5x5 point grid. The symmetry of the rotation function is the same as that of the Patterson function (point group mmm) and so a complete search covered the range from 0^{o} to 90^{o} in both the spherical polar angles. The interval of search was 5^{o}, angles less than 2.5^{o} being insignificant for this data (Rossmann and Blow, 1962). It was thought that the inclusion of additional data at higher resolution would not clarify the results and no calculations with them have been attempted.

The program used for the calculation of the rotation function was written for the Argus computer by Dr M.A. Joynson. The results

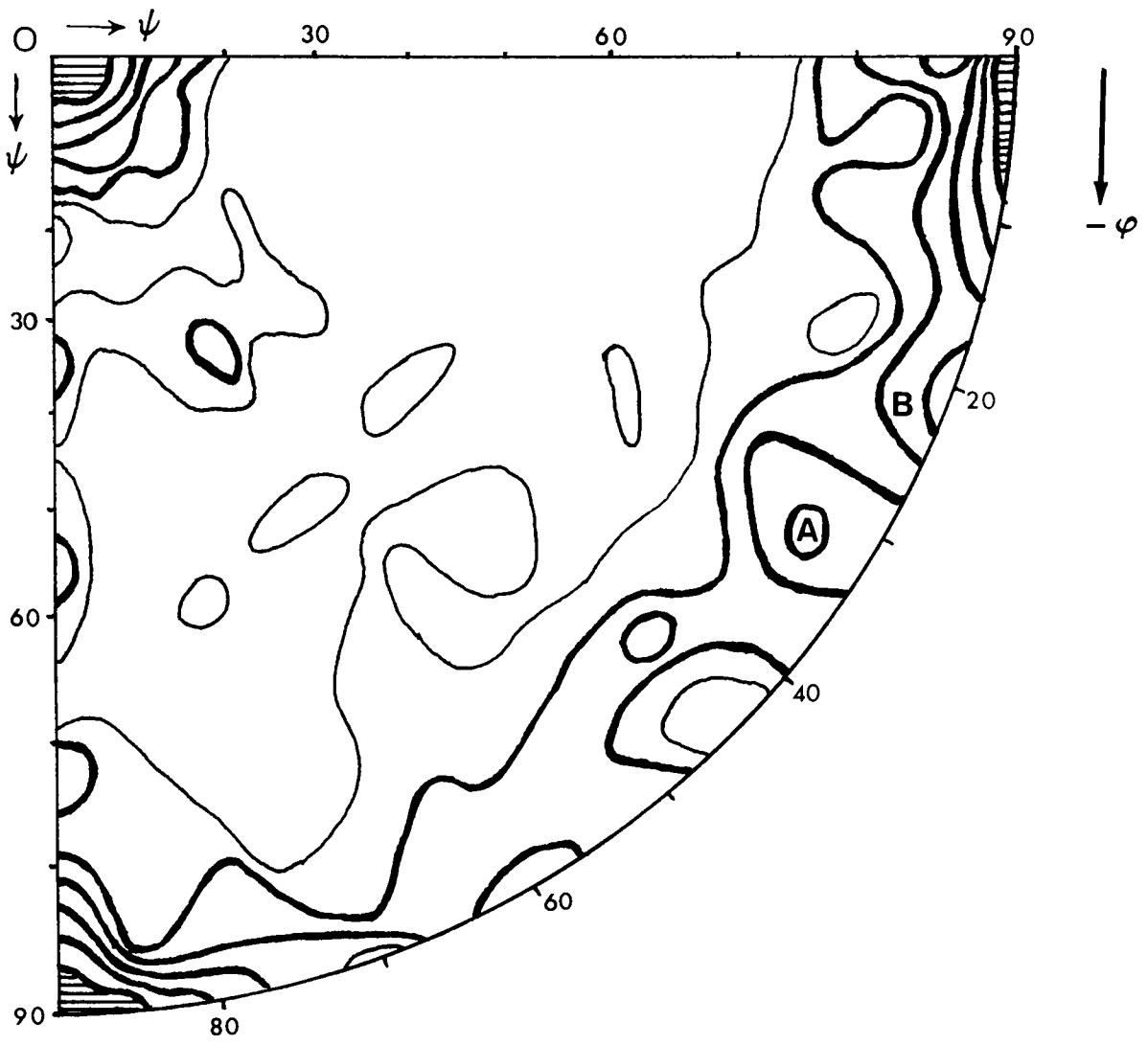


FIGURE 12 Stereographic Projection of the Self Rotation Function. The light contour represents the mean value of the function over the entire surface of the sphere; the heavy contours represent units of half the RMS deviation of the function from this mean value; the best estimate of the orientation of the molecular diad from these results is marked A, that subsequently determined from the 6Å electron density map by least squares refinement is marked B.

are seen in Figure 12 which is a stereographic projection of the value of the function over an octant of the sphere.

The maximum values of the function, excluding the origin, are all within about 10^{o} of the plane containing the x and z axes, and so the direction of the molecular diad axis may be assumed to be somewhere within this region. The determination of its orientation more precisely than this can be seen from the stereogram to have low significance. The best estimate of the coordinates of the axis from these results alone is $\phi = 32^{o}$ $\psi = 83^{o}$. The direction subsequently determined from the electron density map as described in Chapter 6 is $\phi = 21^{o}$ $\psi = 86^{o}$.

The low significance of these results is probably explained by subsequent knowledge of the molecular orientation within the unit cell and the location of the non-crystallographic axis. The use of the rotation function to locate a single diad axis is one of the applications for which it is least powerful, but in no other case does it appear to have produced such discouraging results as these, especially as there are reasons for expecting this molecule to have a very high degree of internal symmetry.

Chapter III HEAVY ATOM ISOMORPHOUS DERIVATIVES

A) METHODS OF SURVEYING POTENTIAL DERIVATIVES 71

Difference Patterson Functions and their Interpretation 71

Initial Refinement of Heavy Atom Parameters 76

Calculation of Difference Fourier Syntheses 79

B) CO-CRYSTALLISED MERCURIAL DERIVATIVES 82

Special Atomic Positions and their Implications 84

C) DIFFUSED DERIVATIVES 87

Changes in Unit Cell Parameters 92

D) INTENSITY MEASUREMENT FOR DERIVATIVES AT 6Å RESOLUTION 93

E) LOW RESOLUTION ANALYSIS OF THE PCMB DERIVATIVE 95

F) LOW RESOLUTION ANALYSIS OF THE EMP DERIVATIVE 97

G) LOW RESOLUTION ANALYSIS OF THE BAKER DERIVATIVE 99

H) CORRELATION OF ATOMIC POSITIONS IN THE SULPHYDRYL DERIVATIVES 100

I) LOW RESOLUTION ANALYSIS OF THE $PtCl_4$ DERIVATIVE 103

A) METHODS OF SURVEYING POTENTIAL DERIVATIVES

Crystals of potential heavy atom isomorphous derivatives were first examined by means of X-ray diffraction photographs to assess their quality and degree of crystallinity, subsequent intensity measurements being made quantitatively on a diffractometer. The methods of data collection were very similar to those already described for the native low resolution measurements, and so are not described separately except where they differ significantly from the previous outline or where the measurements were subsequently used in the determination of the phases of the native structure factors at 6Å resolution.

During the course of this survey of derivatives various undesirable features became apparent in the native data. A new set of such measurements was made and carefully checked against all the currently available derivative data sets. All the potential derivatives have thus been examined with respect to either the complete new set of native structure amplitudes, or a single projection of data obtained earlier, which was shown to be essentially identical to the corresponding part of the complete new set.

The parameters of the heavy atoms in an isomorphous derivative are determined from difference Patterson syntheses showing the vectors between the heavy atoms, or from modifications of this synthesis. The usefulness of any derivative, however, depends on the progress of refinement of these initial parameters and the magnitude of the residual errors.

Difference Patterson Functions and their Interpretation

The difference between the self Patterson functions of the

derivative and native protein structures is that synthesis in which the coefficients are $(|F_{PH}|^2 - |F_P|^2)$, and has features representing the vectors both between the heavy atoms themselves and between the heavy atoms and the atoms of the protein. This synthesis, which may be called the "ΔI Patterson", had been suggested as a method of locating the heavy atoms, but it was shown that the background of protein-heavy atom interactions considerably exceeded the interactions between the heavy atoms themselves for non-centrosymmetric structures (Blow, 1958). Blow proposed that a Patterson function using the coefficients $(|F_{PH}| - |F_P|)^2$, the "$(\Delta F)^2$ Patterson" would provide more information in such cases, it having been used previously only in centrosymmetric cases until Bragg (1958) used an analogous method in reciprocal space.

For centric reflections the structure factors of the protein, derivative and heavy atoms are all real quantities, so that unless the protein and derivative structure factors have opposite signs, a situation described as a crossover, the relationship

$$|f_H| = |\Delta F| \approx \big||F_{PH}| - |F_P|\big| \qquad (1)$$

between the heavy atom structure factor and the observed isomorphous difference is an exact one and the $(\Delta F)^2$ Patterson contains peaks representing only the interactions between the heavy atoms.

This is clearly not the case for acentric reflections where the relationship between the structure factors is one involving complex numbers as shown by the phase triangle in Figure 13. For such reflections the magnitude of the observed isomorphous difference ΔF is always less than the heavy atom structure factor. The features to be expected in a $(\Delta F)^2$ Patterson of a non-centrosymmetric structure

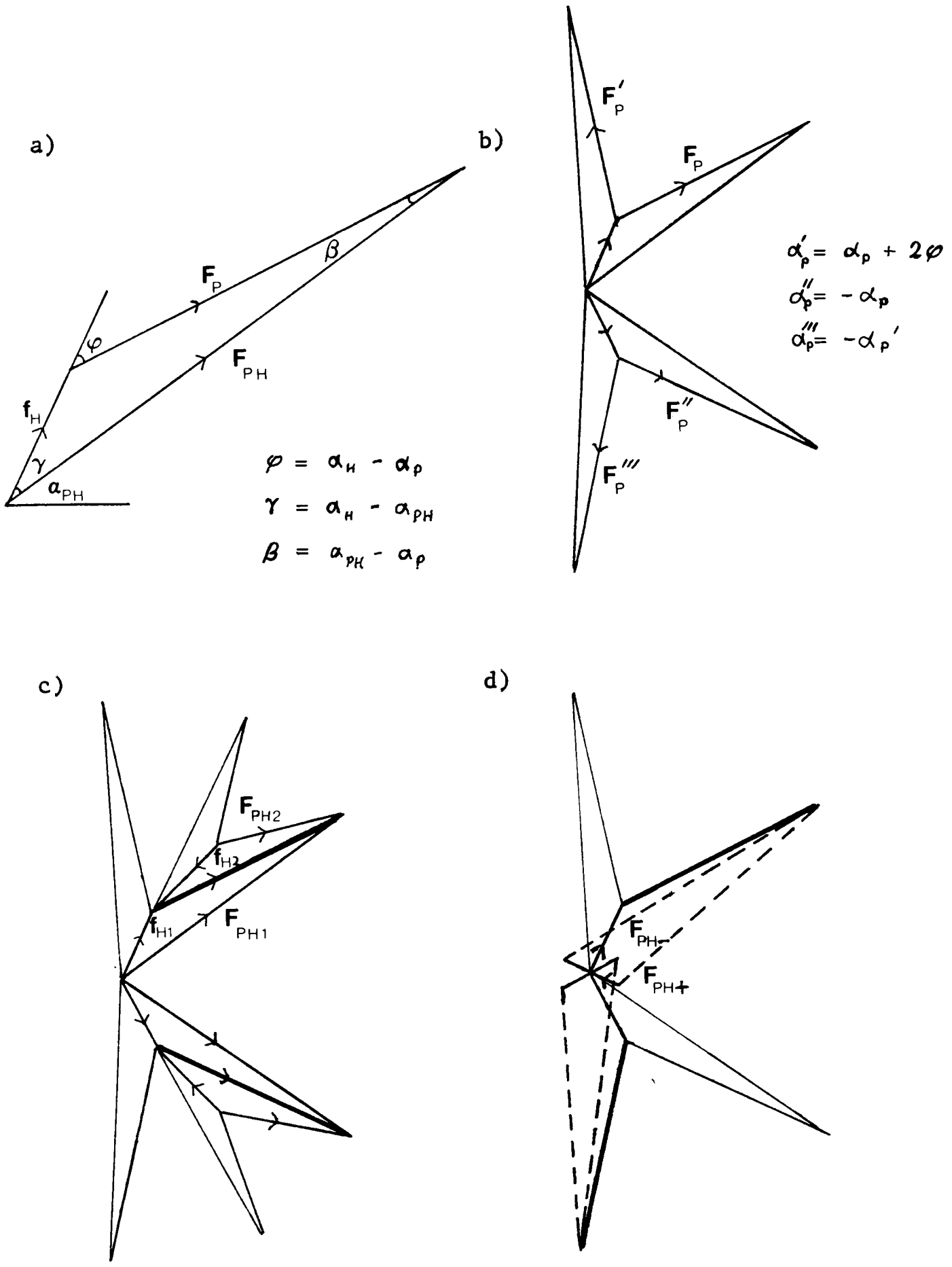


FIGURE 13 The Phase Triangle.
The relationship between the protein, derivative and heavy atom structure factors F_P, F_{PH}, f_H is shown (a) with the fourfold ambiguity in the phase of the protein structure factor α_P in (b); different pairs of these possibilities are eliminated by the use of a second isomorphous derivative (c) and by the use of anomalous scattering information (d), the reflection shown having F_{PH+} less than F_{PH-}.

are discussed in detail in the review by Phillips (1966) where it is shown that the interactions between heavy atoms will appear with only half their predicted weight, but there will also be features of low weight representing both protein-protein and protein-heavy atom interactions. However, the dominant features will be those to be expected in the self Patterson function of the heavy atom structure, and the relative ease of locating their positions by the $(\Delta F)^2$ Patterson, in comparison with other methods, has led to the widespread use of this synthesis which is subsequently referred to simply as the difference Patterson synthesis.

The centre of symmetry which is present in all Patterson syntheses means that there is always an ambiguity in the determination of atomic positions for non-centrosymmetric structures of $\pm(x,y,z)$. This implies that for heavy atom isomorphous derivatives only the real part of the heavy atom, and thus also the protein and derivative, structure factors can be determined. A second major ambiguity arises from the possibility of the phase triangle being reflected through the heavy atom structure factor, allowing only the component of the protein structure factor parallel to it to be determined. The second ambiguity can be resolved by the use of multiple isomorphous derivatives or anomalous scattering information, and the first ambiguity by consideration of the enantiomorph of the finally determined structure, it being known that proteins are composed of L-amino acids, or from the anomalous scattering information if more than one isomorphous derivative has been used. A geometrical illustration of these ambiguities is given in Figure 13.

The use of multiple isomorphous derivatives requires the resolution of any ambiguity in the choice of the origin with respect

to which the coordinates of the heavy atoms are expressed. This can be achieved by the use of difference Fourier series, once an interpretation of one of the derivatives is available, or it can be resolved in Patterson space by a correlation function. Several such functions have been proposed by various workers and their properties reviewed by Phillips (1966). The only such function which was used in this analysis is that function having coefficients

$$(|F_{PH2}| - |F_P|)(|F_{PH1}| - |F_P|) ,$$

the product of the isomorphous differences observed for two derivatives. The resultant map can be shown to have peaks representing the vectors between heavy atoms in the first derivative and those in the second. Such functions can be useful not only for correlating different derivatives but also the initial interpretation of the heavy atom structure in multiple site derivatives.

The existence of more than one asymmetric unit within the unit cell of a structure implies that certain related sets of vectors must be represented in the corresponding Patterson function. In the space group $P2_12_12_1$ the three screw diad axes generate three more atoms in the cell from any one atom, the fractional coordinates of the four atoms being related as

$$\begin{array}{ccccc} x & , & y & , & z \\ \frac{1}{2} - x & , & -y & , & \frac{1}{2} + z \\ \frac{1}{2} + x & , & \frac{1}{2} - y & , & -z \\ -x & , & \frac{1}{2} + y & , & \frac{1}{2} - z \end{array}$$

which are termed the general equivalent positions of the space group. Consideration of all the possible vectors between these positions shows that there are only four unique vectors, the remaining twelve being

generated by the mmm symmetry of the Patterson function. The unique set of self vectors, relating equivalent atoms, is seen for this space group to be those vectors with coordinates (uvw)

$$\begin{array}{ccccc} 0 & , & 0 & , & 0 \\ \frac{1}{2} - 2x & , & 2y & , & \frac{1}{2} \\ 2x & , & \frac{1}{2} & , & \frac{1}{2} - 2z \\ \frac{1}{2} & , & \frac{1}{2} - 2y & , & 2z \end{array}$$

which all occur either at the origin or on one of the Harker sections for this space group ($uv\frac{1}{2}$, $u\frac{1}{2}w$, $\frac{1}{2}vw$). The symmetry of the Patterson is seen to leave an uncertainty in the determination of atomic coordinates from these vector positions to the extent of the addition of one half of the unit translation to any or all of the atomic coordinates. This can be resolved whenever there are atoms in more than one set of general equivalent positions by the positions of the cross vectors between them. For atoms at (x_1, y_1, z_1) (x_2, y_2, z_2) the four independent cross vectors are those having coordinates

$$\begin{array}{ccccc} x_1 - x_2 & , & y_1 - y_2 & , & z_1 - z_2 \\ \frac{1}{2}-(x_1 - x_2) & , & \frac{1}{2}-(y_1 + y_2) & , & z_1 + z_2 \\ \frac{1}{2}-(x_1 + x_2) & , & y_1 + y_2 & , & \frac{1}{2}-(z_1 - z_2) \\ x_1 + x_2 & , & \frac{1}{2}-(y_1 - y_2) & , & \frac{1}{2}-(z_1 + z_2) \end{array}$$

These cross vectors can be found anywhere within the Patterson and are not restricted to particular lines or sections of it.

The complete interpretation of a general set of cross vectors in a difference Patterson map resolves all the ambiguities about relative origins to which the coordinates of individual sites within a single derivative are referred. Correlation function maps should

show the same pattern of vectors, since they also represent cross vectors between independent sites, albeit in different derivatives, and their interpretation thus allows correlation of the relative origins of the derivatives.

The initial coordinates determined for the heavy atom positions for each derivative were tested by refinement using the method described below and comparison of the predicted isomorphous differences with those that were observed. The parameters to be refined include a scale factor relating the structure amplitudes of the derivative to those of the protein. In this discussion of the difference Patterson method it has been assumed that these were on the same scale. Any errors in their relative scaling will clearly complicate the features appearing in the synthesis, increasing the weight of the protein-protein and protein-heavy atom interactions which will appear. A more detailed analysis of the problems involved in determining a scale factor is given in Chapter V. The initial choice of scale factor was always such that after scaling the average derivative intensity was slightly greater than that for the protein as must be the case for random structures of discrete atoms where Wilson statistics (Wilson, 1942) can be assumed to apply to the structure amplitudes.

Initial Refinement of Heavy Atom Parameters

The phase triangle reduces to a linear relationship for centric reflections, so that all three structure factors, for protein, heavy atom and derivative, are uniquely determined if any one of them and the other two structure amplitudes are specified. Thus the comparison of the experimentally determined amplitude of the isomorphous difference with the calculated value of the heavy atom structure factor,

using initial heavy atom parameters, allows both the refinement of these parameters and the determination of the signs of the protein and derivative structure factors. The determination of phases for acentric reflections requires additional information from anomalous scattering measurements or another isomorphous derivative in order to resolve the ambiguity of orientation of the phase triangle between the two possibilities which are related by reflection in the heavy atom vector. Refinement of parameters using acentric reflections also requires such additional information to provide a constraint on the relationships of the phase triangle.

The presence of three centrosymmetric projections allows all the heavy atom parameters to be refined without recourse to the acentric reflections and this method was adopted here. It has the advantages of both considering each potential derivative independently and being relatively economical in computing time, since only a fraction of the complete set of reflections are being considered. These are combined with the disadvantage of neglecting the anomalous scattering information, which is of intrinsically greater accuracy than measurements of isomorphous differences, although it has a lower statistical precision.

The program used for these refinements was originally coded in Fortran by Dr I.D.A. Swan but has since been extensively modified. The heavy atom structure factor f_H is calculated as the sum of the contributions from each of n atomic sites according to the formula

$$f_H(hkl) = \sum_n Z_n f_n(\theta) \exp(-B_n \sin^2\theta/\lambda^2) . T(hkl\ x_n y_n z_n)$$

where the geometrical factor T is a function of the indices hkl and the fractional coordinates x_n, y_n, z_n, the isotropic temperature factor B_n and the heavy atom occupancy Z_n are parameters to be refined for

each site, and the atomic form factor $f_n(\theta)$ was calculated as a power series

$$f_n(\theta) = -1609.\mathrm{Sin}^6\theta + 860.\mathrm{Sin}^4\theta - 211.\mathrm{Sin}^2\theta + 80.$$

the four coefficients being determined by the scattering expected from a bound mercury atom surrounded by a shell of light atoms displacing an equivalent sphere uniform density (Blake et al., 1965).

The signs of the protein and derivative structure factors for any reflection are those which minimise the lack of closure ϵ in the expression

$$| k F_{PH} - F_P - f_H | = \epsilon \qquad (2)$$

where k is the overall derivative scale factor. Refinement proceeds by minimising the RMS lack of closure error E for all the centric reflections with respect to independent changes in each parameter. The initial variation of each parameter has to be specified, it being changed in successive cycles according to the nature of the convergence or divergence of the refinement of that parameter. When convergence is complete, the parameter is artificially varied to test the possibility of its having refined to a false value giving only a local minimum of the lack of closure error.

The progress of the refinement is indicated by the values of the RMS lack of closure error E and the crystallographic reliability index R calculated for the isomorphous differences (Cullis et al., 1961) according to the equation

$$R = \frac{\langle | k F_{PH} - F_P - f_H | \rangle}{\langle | k F_{PH} - F_P | \rangle} . \qquad (3)$$

The sign appropriate to the isomorphous difference ΔF, which is the observable quantity equivalent to the heavy atom structure factor for a centric reflection, is determined by the signs of the protein and derivative structure factors. The relationship can be expressed in two ways

$$\Delta F = k F_{PH} - F_P \quad (4)$$

$$= (k|F_{PH}| - |F_P|)e^{i\alpha_P} \quad (5)$$

where α_P is the phase angle of the protein structure factor, which has the value 0 or π for a centric reflection. These two expressions are exactly equivalent for all reflections except those few which have crossovers, where the signs of the protein and derivative structure factors are different. Such crossovers can only occur if the sum of the protein and derivative structure amplitudes is less than the maximum possible heavy atom structure amplitude.

Both of these expressions for ΔF and the analogous expressions for the double difference $\Delta\Delta F$, which is the difference between the observed and calculated derivative structure factor,

$$\Delta\Delta F = k F_{PH} - (F_P + f_H) \quad (6)$$

$$= (k|F_{PH}| - |(F_P + f_H)|)e^{i\alpha_{PH}} \quad (7)$$

are used in the calculation of a variety of difference Fourier syntheses which provide useful information at many stages in the refinement of an isomorphous derivative.

Calculation of Difference Fourier Syntheses.

Fourier syntheses reveal the details of a structure and so difference Fourier syntheses can be used to examine differences between

structures, both for studying the structure giving rise to the differences between the observed and calculated structure amplitudes and also for studying closely related isomorphous structures when the phases of only one of them are known. The first application is the only one considered here, although the principles of the method are the same in either situation.

The coefficient of each reflection included in the synthesis must be specified in both amplitude and phase. The synthesis in which the coefficients are the observed isomorphous differences, the ΔF synthesis, will show the heavy atom structure. For centric reflections the signs of these coefficients, whose amplitudes are obtained directly from the observed structure amplitudes, can be obtained from equation (4) where the possibility of crossovers is considered or equation (5) where it is ignored. For acentric reflections equation (4) becomes a vectorial relationship, and so in the absence of knowledge of the derivative structure factor it can not be used, but equation (5) generates coefficients analogous to those in the centric case, all crossovers being neglected, although the reflections where this situation is possible can of course be excluded from the synthesis.

The phase of the protein structure factor which is used with the observed isomorphous difference in the ΔF synthesis for a given derivative may have been determined by that derivative alone or by a combination of one or more other derivatives but excluding that one. Such syntheses are termed self-sign and cross-sign difference syntheses respectively for centrosymmetric cases and self-phase or cross-phase as appropriate in other situations. A synthesis which was used frequently in this analysis involved protein phases which were determined jointly by several derivatives including that whose isomorphous

differences form the amplitudes of the coefficients. The need for this synthesis arose because of the lack of truly independent derivatives and it is referred to as a best-phase synthesis, by analogy with method of phase determination used, which is described in Chapter V.

Difference syntheses of all these types will indicate any additional heavy atom binding sites which may be present, but small changes in the known heavy atom sites are determined more easily from double difference syntheses, whose coefficients are evaluated according to equations (6) and (7) for centric reflections, only the second of these being useful for acentric reflections. The phase of these coefficients is that which is calculated for the derivative structure factor from the current estimate of the phases of the heavy atom and protein structure factors, the latter being self-phases, cross-phases or best-phases.

Difference and double difference syntheses calculated with coefficients defined in these two ways have been found the most useful in protein structure analysis (Phillips, 1966). The relative weights of features appearing in these syntheses can be considered in several ways and is analysed in more detail in Appendix II. The phases used in these difference syntheses are not necessarily those appropriate to the structures being examined, and the extent to which they differ from the true phases reduces the size of the features seen on the difference maps. This effect is rarely significant with centric reflections, except for the presence of undetected crossovers, but for syntheses using acentric reflections the atoms of the difference structure appear at no more than half of their true weight (Luzzatti, 1953). Atoms which were included erroneously in the phase determination

but do not correspond to features in the true difference structure will appear with almost the full weight expected for such atoms as demonstrated by Dickerson et al. (1967). Thus the use of self-phase and best-phase difference syntheses alone will not detect any incorrectly positioned atoms.

The use of these methods of locating heavy atoms and refining the parameters of the difference structure in the case of isomorphous derivatives or potential derivatives of chicken TIM crystals is reported in the following section. Those derivatives which were used in the 6Å phase determination are then considered individually in greater detail.

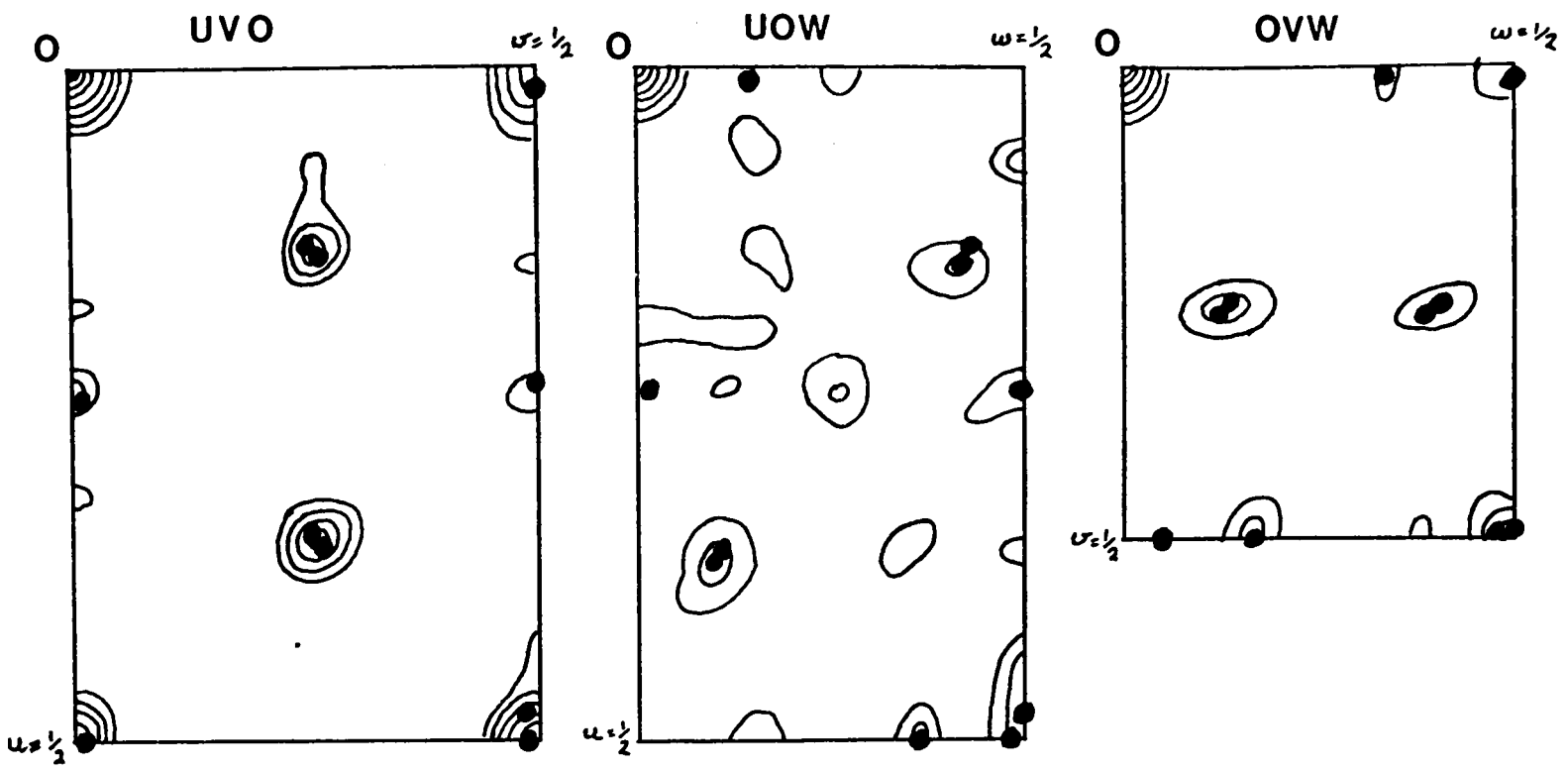
B) CO-CRYSTALLISED MERCURIAL DERIVATIVES

Chicken TIM contains only a single reactive sulphydryl group per subunit. This was used by Dr C.I. Pogson to obtain mercurial derivatives by addition of one molar equivalent per subunit of the Trentham mercurials (McMurray and Trentham, 1969) to a solution of the enzyme. Crystals were grown under the same conditions as for the native enzyme except for the exclusion of 2-mercaptoethanol (Pogson, 1972). The particular mercurial reagents were 2-chloromercuri-4-nitrophenol and 2,6-dichloromercuri-4-nitrophenol which are subsequently referred to as the mono- and di-mercurial derivatives respectively, with relation to the number of mercury atoms they contain.

The crystals were coloured yellow and had clearly defined morphologies. The MONO crystals were similar to those of the native enzyme, but all of the earliest batches of the DI crystals showed considerable elongation of the crystals parallel to the z axis, but the

facing page 83

MONO



DI

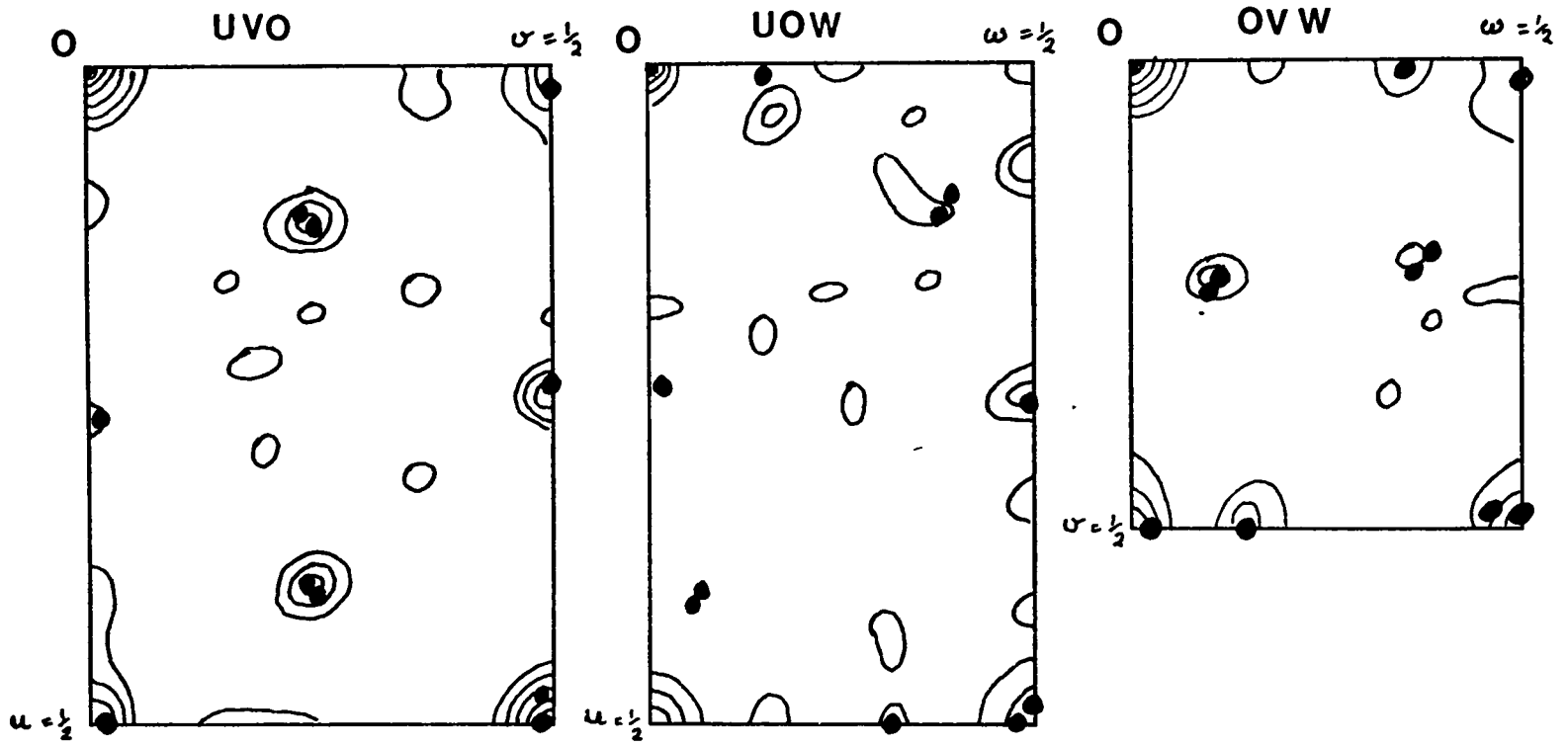


FIGURE 14 Projections of 6Å Difference Pattersons: MONO and DI.

two prismatic forms present appeared to be the same as for all the other crystals. X-ray precession photographs showed that both derivatives had the same unit cell as the native crystals and essentially similar diffraction patterns although there were very clear differences between them on inspection of individual reflections. The mercurial crystals appeared to be as well ordered as those of the native enzyme.

The three projections, which are shown in Figure 14, of the difference Patterson functions of these two derivatives indicate the extent of their similarity. The P(uvO) projection, which was obtained first, was readily interpreted in terms of two sites, although there was a three fold ambiguity in their coordinates as shown in Table 1 since the agreement between the predicted and observed peak heights was not good enough to allow any of these possible structures to be rejected. The ambiguity was subsequently resolved by the other projections of the vector map.

The first two of these possible structures form a homometric set having identical vectors, and thus structure amplitudes. Consideration of the structure factor equations shows that their structure factors are also identical so that they are not capable of distinction even by the signs of the isomorphous differences. The theoretical interest of such structures is accompanied by considerable practical difficulties, since the occurrence of atomic coordinates which are exact rational fractions of the unit cell edge, or values very close to such fractions, can complicate and reduce the precision of both the refinement of the atomic parameters and the phase determination.

facing page 84

TABLE 1 Solutions of the Sulphydryl Derivative Difference Patterson uv0 Projection.
The relative heights of the peaks for the three possible structures are listed, the differences between them being indicated with asterisks.

STRUCTURE				I	II	III	
ATOMIC COORDINATES				(1/8, 0,z) (0,1/4,z)	(1/8, 0,z) (1/4,1/4,z)	(1/8,1/4,z) (0, 0,z)	
VECTORS	0	0	0	8	8	8	
	1/4	0	0	2	2	0	*
	1/4	1/2	0	2	2	4	*
	0	1/2	0	4	4	4	
	1/2	0	0	4	4	8	*
	1/2	1/2	0	8	8	4	*
	1/8	1/4	0	4	4	4	
	3/8	1/4	0	4	4	4	

Special Atomic Positions and their Implications in Structure Analysis

The problems arise most seriously for the centric reflections for which the structure factor expression can be fully factorised into two parts, each depending on only a single coordinate and its associated index. The corresponding expression for acentric reflections always involves two indices in each factor, in addition to having both real and imaginary parts. The contribution to the structure factor of a centric reflection by an atom, whose coordinate is close to a rational fraction 1/4n of the unit cell, will be zero or very small whenever the corresponding index is an odd multiple of the integer n. The sign of this contribution will be very sensitive to the exact atomic coordinate relative to the special position, but the small isomorphous differences for these reflections are those which can be least accurately measured.

The reflections for which the index is an even multiple of the integer n are those for which the contribution to the structure factor is a maximum, and therefore those for which the most accurate measurements are to be expected, but they are precisely the reflections for which the contribution is least sensitive to changes in the atomic position. The refinement of atomic coordinates from such special starting positions thus relies on those intermediate reflections for which the isomorphous difference can be expected to be both reasonably well measured and fairly sensitive to the precise atomic position.

The term special position with reference to atomic sites in the unit cell can have several implications. In crystallographic literature it refers most frequently to those positions, usually associated with symmetry elements, producing a distribution of structure factors for one or more classes of reflections, which is

atypical of the distribution for all the reflections considered together. Such a situation does not arise in the space group $P2_12_12_1$ for any of the possible atomic positions considered here. However, the coordinates 1/4 and 1/8 are special in the sense that they give rise to nodes and antinodes in the structure factor graphs for particular classes of reflections. The coordinate 1/8 in structures I and II is special to the extent that, although they are homometric for any value of this x coordinate, the signs of the structure factors are only identical because of this coordinate. The coordinates 0, 1/4, 1/2, 3/4 are special in this space group for all three principal axes since they represent positions lying on or exactly midway between the screw diad axes, causing ambiguity in the determination of a unique origin or enantiomorph as will be discussed below for the correlation of parameters from different derivatives. It is this last sense which will be subsequently implied by use of the term special position.

The ambiguity which is implicit in the P(uvO) projection of the difference Patterson is completely resolved in favour of structure I by examination of the other two projections or the three dimensional function P(uvw). The longest projection P(Ovw) was interpreted in terms of two sites with coordinates.

(x, 0,$\pm$1/4) (x,$\pm$1/4, 1/6).

The projection P(uOw) was not capable of independent interpretation but was consistent with the atomic sites deduced from the other projections. The complete function P(uvw) independently gave rise to the same interpretation of two atomic sites. Indications of any departure of the atomic positions from the rational fractional

TABLE 2 Contributions of Atoms at the Sulphydryl Sites to the Structure Factors.

Values of the sum for two sites A , B at (1/8,0,3/4) and (0,1/4,1/6) respectively of the functions

$\cos(2\pi.\text{arg1}).\cos(2\pi.\text{arg2})$ for (h+k) even

$\sin(2\pi.\text{arg1}).\sin(2\pi.\text{arg2})$ for (h+k) odd

are listed for the three principal projections for different classes of reflections. Non-zero contributions are denoted A , B if they are the maximum for that site and a , b if they are not necessarily the maximum values; zero contributions are shown by blanks in the table (a); the arguments of the trigonometric functions are defined in (b) with the summary of these results.

a)

Index of arg2	4n	4n+1	4n+2	4n+3
(hk0)				
Index of arg1 4n	±A +B		-B	
4n+1		±a		±a
4n+2	±A -B		+B	
4n+3		±a		±a
(h01)				
4n	±A ±b		±A ±b	
4n+1		±a ±b		±a ±b
4n+2	±b		±b	
4n+3		±a ±b		±a ±b
(0k1)				
4n	A ±b		-A ±b	
4n+1		±b		±b
4n+2	-A ±b		A ±b	
4n+3		±b		±b

b)

Projection	hk0	h01	0k1
arg 1	h(x - 1/4)	l(z - 1/4)	k(y - 1/4)
arg 2	ky	hx	lz
No. of classes of reflections with contributions to F from: both sites	2	6	4
site A only	4	0	0
site B only	2	2	2
neither site	8	8	8

coordinates were never entirely consistent from all the available evidence, except for the x coordinate of the first site, which was found to be slightly closer to the origin than 1/8. The cross vectors between the two sites appeared at

(3/8, 1/4, 1/12) and (1/8, 1/4, 5/12)

as doubly weighted peaks with no trace of any splitting into separate components, so that there are no indications of the directions in which the atomic sites should be moved from the special positions, but the relative x and z coordinates are clearly fixed as

(.123, 0, 3/4) and (0, $\pm$ 1/4, .167)

The relative y coordinates can not be determined, but as the negative coordinate represents an enantiomorphic structure, it is not considered further here.

Within the space group $P2_12_12_1$ both sets of coordinates

(.123, 0, 3/4) (0, 1/4, .167)

give rise to atomic structures which are body-centred, so that for half of the reflections the calculated heavy atom structure factor is zero. The correlation between the two sites in the remaining reflections is very weak especially for the hkO projection as shown in Table 2 by the geometrical terms of the structure factors. The problems arising from this are discussed later in conjunction with information from other derivations.

After refinement of the parameters from these initial values the three projections of the ΔF and $\Delta\Delta F$ syntheses were calculated for both mercurial derivatives using their own signs. The dimercurial maps did not obviously show the positions of the expected additional mercury atoms, but there were consistent indications of a site of much lower occupancy about 6.5Å from each of the major sites. Refinement

of these four sites for the dimercurial derivative produced relative occupancies of the minor to major sites of 0.3, and the relative occupancies between the sets of positions, which were later shown to be related by the non-crystallographic diad axis, differed by 30% in opposite directions between the major and minor sites, as they did between the mono- and di-mercurial derivatives for the two major sites.

Reliability indices of about 45% were obtained for both derivatives at this stage. The difference Fourier syntheses still contained several uninterpretable features and the solution of the dimercurial structure was not as different from that of the mono-mercurial derivative as had been hoped. A later analysis by Dr Pogson (personal communication) of the mercury content of the crystals suggested that the dimercurial reagent may have partially decomposed in solution during the reaction with the sulphydryl groups of the enzyme and subsequent crystallisation.

At least one additional isomorphous derivative thus became necessary for the phase determination, since these two mercurial derivatives were not sufficiently different to provide independent information about the protein phases, although one of them could clearly be used in this way.

C) DIFFUSED DERIVATIVES

The availability of a single group on each protein subunit which is reactive towards the mercurial compounds discussed above suggested that other sulphydryl reagents would also produce isomorphous derivatives. Diffusion of p-chloromercuri-benzoate (PCMB) into crystals of both the native protein and the monomercurial derivative

TABLE 3 Survey of Potential Heavy Atom Derivatives.
The ligands were added to the mother liquor to give solutions of the concentrations listed, molar ratios M being expressed per protein monomer and spaces indicating uncertainty because of precipitation or insolubility. The extent of the data collected at 6Å resolution is shown as the complete three dimensional set,the three projections or the hk0 reflections only. An asterisk following this indicates the detection of intensity changes by the presence of a significant origin peak on the difference Patterson map. The number of sites N and the reliability index R are shown for those derivatives for which refinements were executed, asterisks indicating those which were used for the determination of phases for the protein.

Ligand	Abbreviation	Conc$^{n.}$	M	Data	N	R	Comments
p-chloromercuri benzoate	PCMB	0.3mM	15	3-D *	2	0.43	two major sulphydryl sites
2-mercuri-4-nitrophenol+PCMB	MP	0.3	15	3-D *	2	0.44	same as PCMB
2,6-dimercuri-4-nitrophenol	2,6-DI	0.1		3-D *			lower occupancy than if cocrystallised
2,4-dimercuri-6-nitrophenol	2,4-DI	0.1		3-D *			same as 2,6-DI
Uranyl acetate	UAc	1	50	hk0 *	2	0.57	cross sign Fouriers discouraging
Ceric ammonium sulphate	CAmS	10		proj*			no consistent solution for all projections
Potassium cyanoplatinite	CN	10	250	hk0 *	2	0.58	cross sign Fouriers discouraging
Potassium chloroplatinite(cocryst)	TC		2	hk0			higher concentrations tried by diffusion
Dimercury acetate	DMA	0.4	25	hk0			higher concentrations cause disordering
Potassium mercuri-iodide	MI	0.02	5	hk0			higher concentrations cause disordering
Uranyl nitrate	UN	1		hk0			
Uranyl fluoride	UF	1		hk0			
Thallium perchlorate	THAL	1	50	hk0			
Thorium nitrate	THOR	1		hk0			
Potassium gold cyanide	AUCY	1	50	proj*	2	0.56	two sulphydyl sites low occupancy
cis-di-amino-dichloroplatinum	PAC	1	250	hk0			
Platinum azo-benzene	PAZ	0.001	1	hk0			higher concentrations cause disordering
Palladium azo-benzene	PDAZ	0.01	10	hk0			crystals disordered
Potassium chloroplatinite	TC	1	100	proj*	2	0.55	sites approximately half occupied
Ethyl mercury phosphate	EMP	1	100	3-D *	4	0.45	sulphydryl sites & other major sites
Baker dimercurial	BAKER	1	100	3-D *	2	0.53	sulphydryl sites & two or more other sites
Iodoacetamide+EMP	IEMP	1	100	3-D *			same as EMP, should have used higher pH

(MONO) produced derivatives containing the same two heavy atom sites as in the MONO crystals, but refinements and difference Fourier maps calculated with these data suggested that the incorporation of the mercury atoms by diffusion was accompanied by smaller distortions of the remaining structure than was evident in the co-crystallisations.

Derivatives were also prepared by diffusion of two similar dimercurial compounds, 2,6-dichloromercuri-4-nitrophenol, as used in the co-crystallised derivative, and 2,4-di-acetoxymercuri-6-nitrophenol, in the hope of obtaining a derivative in which the second mercury atom of the ligand was also present in a specific position within the crystal. However the difference Pattersons calculated from both of these sets of measurements were very similar to each other and indicated a much lower occupancy of the heavy atom sites than was apparent in the co-crystallised derivative, so that these were not considered further.

The need for an independent derivative, with heavy atoms at sites other than the two sulphydryl sites, led to several non-specific ligands being investigated by diffusion into crystals of TIM. The results of this survey are summarised in Table 3.

Three uranyl compounds, the acetate, nitrate and fluoride were studied but only the first of these produced significant intensity changes in the diffraction pattern. Measurements of the hkO reflections to a resolution of 2.5Å were made for this derivative on the linear diffractometer, but the refinement of parameters in this projection for two uranyl binding sites was discouraging. The R factor of 0.57 calculated after a refinement at 6Å resolution did not increase significantly on extending the refinement to 2.5Å resolution as would be expected. The two binding sites determined from the difference Patterson map were inconsistent with the features on a cross-sign

difference Fourier projection calculated with signs from the MONO derivative.

Diffusion of ceric ammonium sulphate into the crystals produced intensity changes and two sites of heavy atom binding were indicated by the isomorphous differences of the hkO reflections, but the other two projections did not support these indications. No intensity changes were observed from crystals into which thallium perchlorate or thorium nitrate had been allowed to diffuse, and the latter compound caused complete degeneration of the crystals.

A series of platinum compounds was tried with varying degrees of success. Potassium chloroplatinite caused no intensity changes when co-crystallised with the enzyme, but changed the crystals to a deep red colour with a consequent loss of diffracting power when diffused into them at high concentrations for a long period of time. A promising derivative, with two platinum binding sites, was obtained on one occasion by soaking the protein crystal for 48 hours in a freshly made solution having 1mM. concentration of the complex ion in the ammonium sulphate mother liquor. Unfortunately this derivative has not yet proved reproducible. The complexing of platinum by the ammonium ions in the mother liquor is clearly a major source of this difficulty.

Potassium cyanoplatinite produced intensity changes and an interpretation of these in terms of two platinum sites from the results of a single projection, but the cross-sign difference Fourier map using signs from the MONO derivative was discouraging and inconsistent with such sites. The auricyanide ion was found to bind to the two sulphydryl sites but with much lower occupancy than was observed for the PCMB derivative. No intensity changes were produced by the platinum

complexes, cis-di-amino-dichloro-platinum and di-μ-chlorobis-(azobenzene-2C,N′) diplatinum (II), or by the corresponding palladium azobenzene complex, the first of these complexes being provided for us by Dr R.J.P. Williams and the other two by Dr Pogson.

Two further mercury compounds which were used, also produced no intensity changes, or if used at higher concentrations disordered the crystals. These compounds were dimercury acetate, which is a smaller molecule than the aromatic dimercurials already discussed and so it was hoped that this would react with additional cysteine residues on the enzyme, and potassium mercuri-iodide, which can dissociate to give several ions of different geometry, it being hoped that the enzyme possessed a binding site complementary to one of these.

Ethyl mercury phosphate EMP produced large intensity changes in the diffraction pattern which were analysed in terms of four major binding sites, two of these being the sulphydryl sites. Refinement of the parameters for these four sites gave encouraging results indicating that this was another useful derivative.

The dimercurial compound which was used successfully to give an isomorphous derivative in the analysis of haemoglobin (Cullis et al., 1961) and is commonly referred to as the BAKER dimercurial (mercury, meso-(2,3-di-methoxytetramethylene)-bis acetate) was given to the laboratory by Dr H.B. Wood of the U.S. National Institute of Health. This reagent diffused into crystals of TIM binding at the two sulphydryl sites as was clearly evident from the difference Patterson maps. Refinement of the parameters for one mercury atom at each site was encouraging, but even more so were the difference Fourier projections calculated using the signs from this initial refinement. One or two positions occupied by the second mercury atom

of the molecule were clearly identifiable close to each of the major sites, and refinement of these indicated that this mercury compound had also given a useful derivative.

An initial attempt at obtaining a derivative in which the sulphydryl groups were blocked by iodoacetamide, followed by EMP binding at two sites only, was unsuccessful but the possibilities of such a derivative have not yet been fully explored.

This survey had thus produced as promising heavy atom isomorphous derivatives PCMB with mercury atoms at the two sulphydryl sites, EMP with mercury atoms at the same two sites and also two other major binding sites, BAKER dimercurial with mercury atoms at the sulphydryl sites with significant minor sites close to these and lastly chloroplatinite with two independent platinum sites. The presence of the sulphydryl site in all three of the mercury derivatives was found to cause considerable problems in later stages of the analysis because of correlations between the heavy parameters in different derivatives, and so the lack of reproducibility of the chloroplatinite derivative was especially serious.

The presence of high concentrations of the ammonium ion in the mother liquor of the crystals undoubtedly contributes to this problem and probably also to lack of specific binding by several of the other reagents studied. Attempts to avoid this problem by transferring the crystals to a phosphate solution of equivalent ionic strength (Sigler and Blow, 1965) encountered a different problem of lack of isomorphism resulting from shrinkage of the unit cell and relative changes in the axial lengths.

TABLE 4 Changes in Unit Cell Parameters.
The dimensions observed for four types of crystals are listed with details of the ligands for which they are observed. These include the substrate dihydroxyacetone phosphate (DHAP), a possible transition state analogue 2-phosphoglycollate (PG) and two examples of heavy metal ligands binding at the sulphydryl sites (PCMB) and elsewhere ($Pt(CN)_4$).

STRUCTURE	I	II	III	IV
a	106.0Å	102.1Å	102.1Å	99.4Å
b	74.7	74.7	74.4	74.4
c	61.7	62.7	62.2	62.5

Crystals in sulphate

TIM
TIM+PCMB
TIM+(DHAP,PCMB)
TIM+$Pt(CN)_4$
TIM+DHAP
TIM+PCMB+PG
TIM+PG
TIM+PG+PCMB

Crystals in phosphate

TIM
TIM+PCMB
TIM+$Pt(CN)_4$
TIM+DHAP
TIM+PG
TIM+(DHAP,PCMB)
TIM+(PG,PCMB)

Changes in Unit Cell Parameters

Crystals of chicken TIM which have been transferred from ammonium sulphate into phosphate mother liquor give a diffraction pattern which is almost identical with that which is obtained from the crystals in sulphate. When the heavy atom reagents ceric ammonium sulphate, potassium auricyanide, potassium cyanoplatinite and cis-di-amino-dichloroplatinum were allowed to diffuse into the crystals after they had equilibrated with the phosphate solution it was found that the unit cell dimensions had changed from the standard values to those representing a unit cell whose volume was 3% smaller as shown in Table 4. No such change was observed with thorium nitrate, but the diffraction pattern of such a crystal appeared to be the same as that of a native crystal in phosphate, the origin peak of the difference Patterson map calculated using the hkO reflections being insignificant, and so there is no evidence that thorium binds specifically to the enzyme as the other ligands appear to do.

These changes in the unit cell parameters had to be interpreted in the light of earlier work by Wolfenden who had shown that the binding of 2-phosphoglycollate, a possible transition state analogue, was accompanied by a reversible contraction of the unit cell volume by a factor of 6%, and also that the inhibitor α-glycerophosphate caused the 3% contraction, as does the binding of 2-phosphoglycollate to crystals of the monomercurial derivative (Johnson and Wolfenden, 1970). Extension of this work by Petsko (Banner et al., 1971) has shown that the structure having a 3% decrease in the volume of the unit cell is that assumed when the substrate dihydroxy acetone phosphate is bound to the enzyme, although there is a slight difference in the shortest unit translation between this structure and that observed in

TABLE 5 Parameters of the Data Collection and Processing at 6Å Resolution.
Details of the sets of measurements for the native protein crystals and the isomorphous derivatives PCMB, EMP, BAKER and $Pt(Cl)_4$ as included in the final phase determination are listed.

Data Set	NATIVE	PCMB	EMP	BAKER	$Pt(Cl)_4$
Crystal Reference Number	CT32	P3	EMP1	BK2	TC13
Date of Data Collection	June1970	Nov.1969	July1970	Aug.1970	July1970
Crystal Mount	b*	b*	c*	b*	b*
Data Collected	hkl $h\bar{k}l$	hkl $hk\bar{l}$	hkl $h\bar{k}l$	hkl	hk0,h01 0kl
Absorption Curve:					
Maximum	1.417	1.510	1.480	1.510	1.720
Reflection	060	060	006	060	060
Axial Lengths:					
a* (r.l.u.)	0.01450	0.01437	0.01445	0.01452	0.01455
b*	0.02059	0.02067	0.02067	0.02062	0.02065
c*	0.02495	0.02498	0.02500	0.02496	0.o2495
Collimators:					
Source (mm.)	1.0	1.0	1.0	1.4	1.0
Detector	3.5	2.5	3.5	3.5	2.5
Peak Scan:					
Overall width	1.0°	0.7°	1.0°	1.0°	1.0°
Number of steps	50	35	50	50	50
Counting time	30secs.	35secs.	30secs.	40secs.	40secs.
Background time	10	7	15	15	15
Primary Processing:					
No. of measurements	2558	2205	2487	1360	793
No. with $F < \sigma$	147	237	295	112	107
No. with $F < 0$	48	170	142	60	43
No. with background difference $> 4\sigma$	2	305	10	2	0
Secondary Processing:					
Delta F Ratio $\langle \lvert\Delta F\rvert \rangle / \langle \bar{F} \rangle$	4.3(4.7)	7.9(4.2)	8.9(4.1)	—	5.9(8.5)
Scale Factor $1.02\langle F_P \rangle / \langle F_{PH} \rangle$		1.356	1.368	1.189	1.570
No. of Valid Observations:					
$\bar{F}$	1358	1328	1333	1290	405
ΔF	1169	832	1090	—	365

the presence of a heavy metal ligand in the phosphate mother liquor as shown in Table 4. The structure adopted on binding the substrate DHAP appears to be that adopted in the presence of phosphate when any ligand is bound at the active site.

The existence of these different structures means that enzyme-substrate interactions can not be studied by the use of difference Fourier methods and so recourse must be made to the more laborious methods of an independent analysis, or the possibilities of various direct methods of relating the diffraction patterns of these crystals for detailed comparison of their structures.

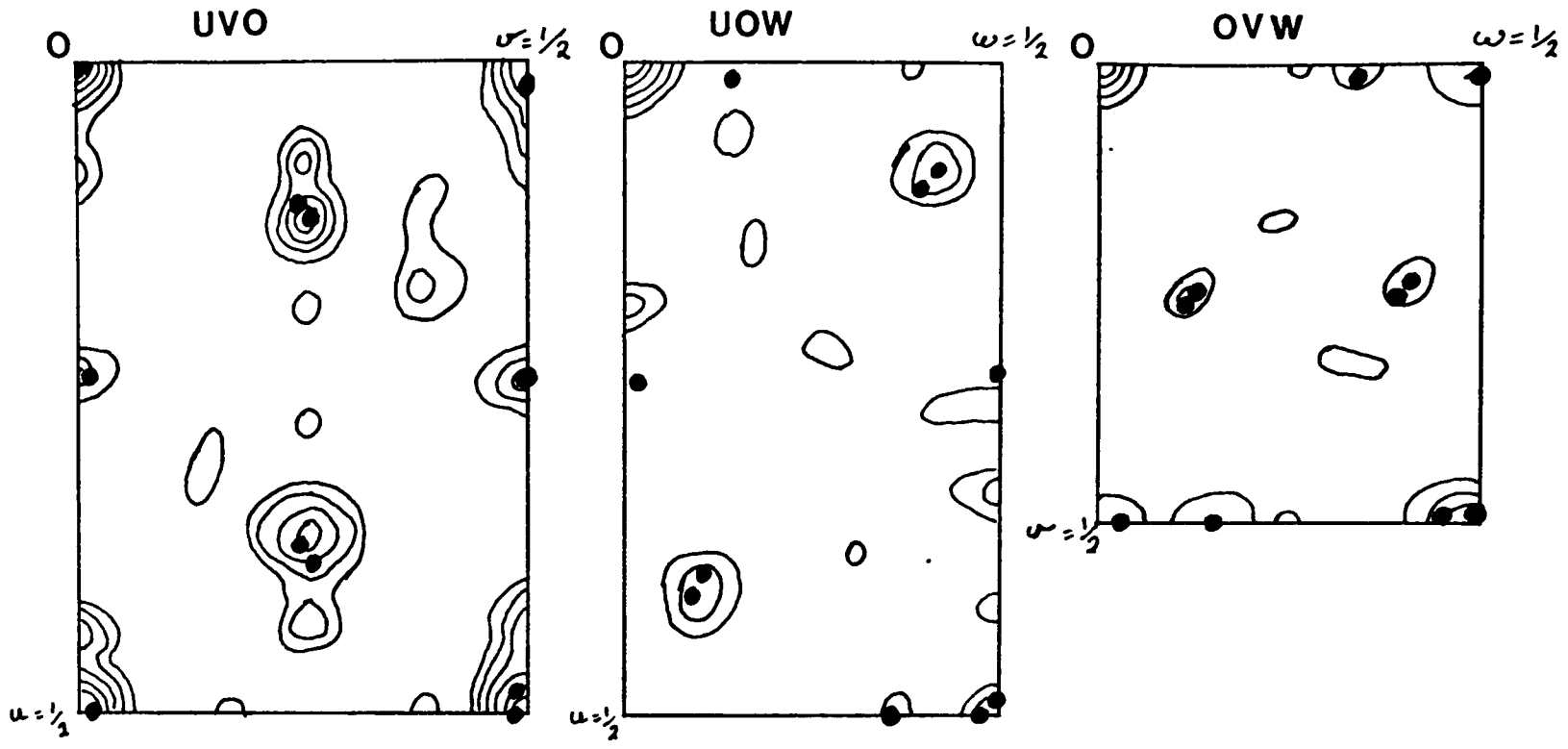
D) INTENSITY MEASUREMENT FOR DERIVATIVES AT 6Å RESOLUTION

The determination of phases at 6Å resolution included information from four isomorphous derivatives, p-chloromercuri-benzoate, ethyl mercury phosphate, Baker dimercurial and chloroplatinite. Intensity measurements for these crystals were made and subsequently processed by the methods described in the previous chapter. Parameters specific to each set of measurements are shown in Table 5.

The values shown for the reciprocal lattice translations are those measured on the particular crystal from which the intensity measurements were taken. Overall average values of these parameters for each derivative are given in the following chapter when the crystals used for the high resolution data collection are considered.

The data set for the PCMB derivative is lacking a significant fraction of the possible anomalous scattering measurements. This seems to be the result of having chosen to scan through too narrow an angle to include all of the reflections completely, as many of the

Projections



Harker Sections

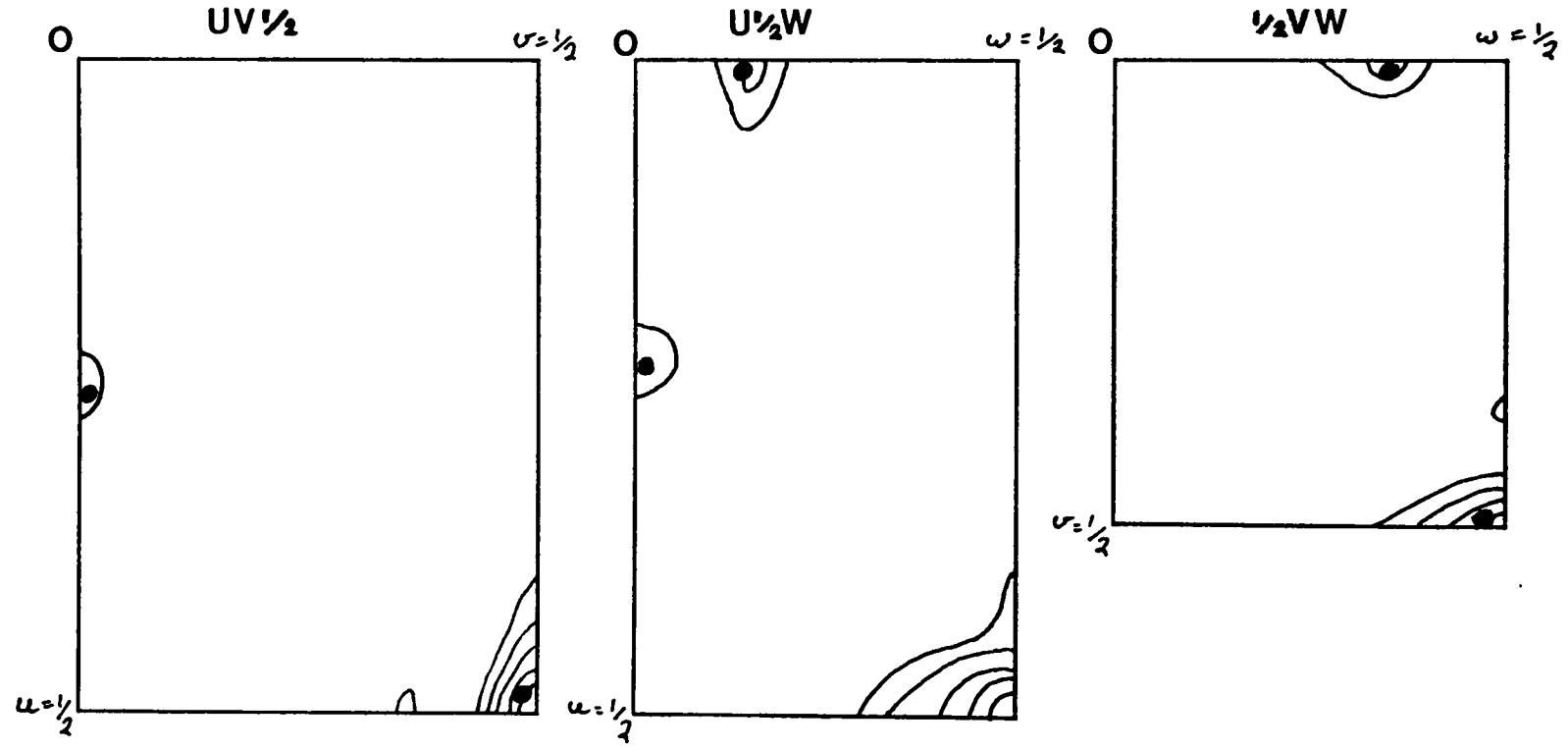


FIGURE 15 Projections and Harker Sections of 6 Å Difference Patterson: PCMB

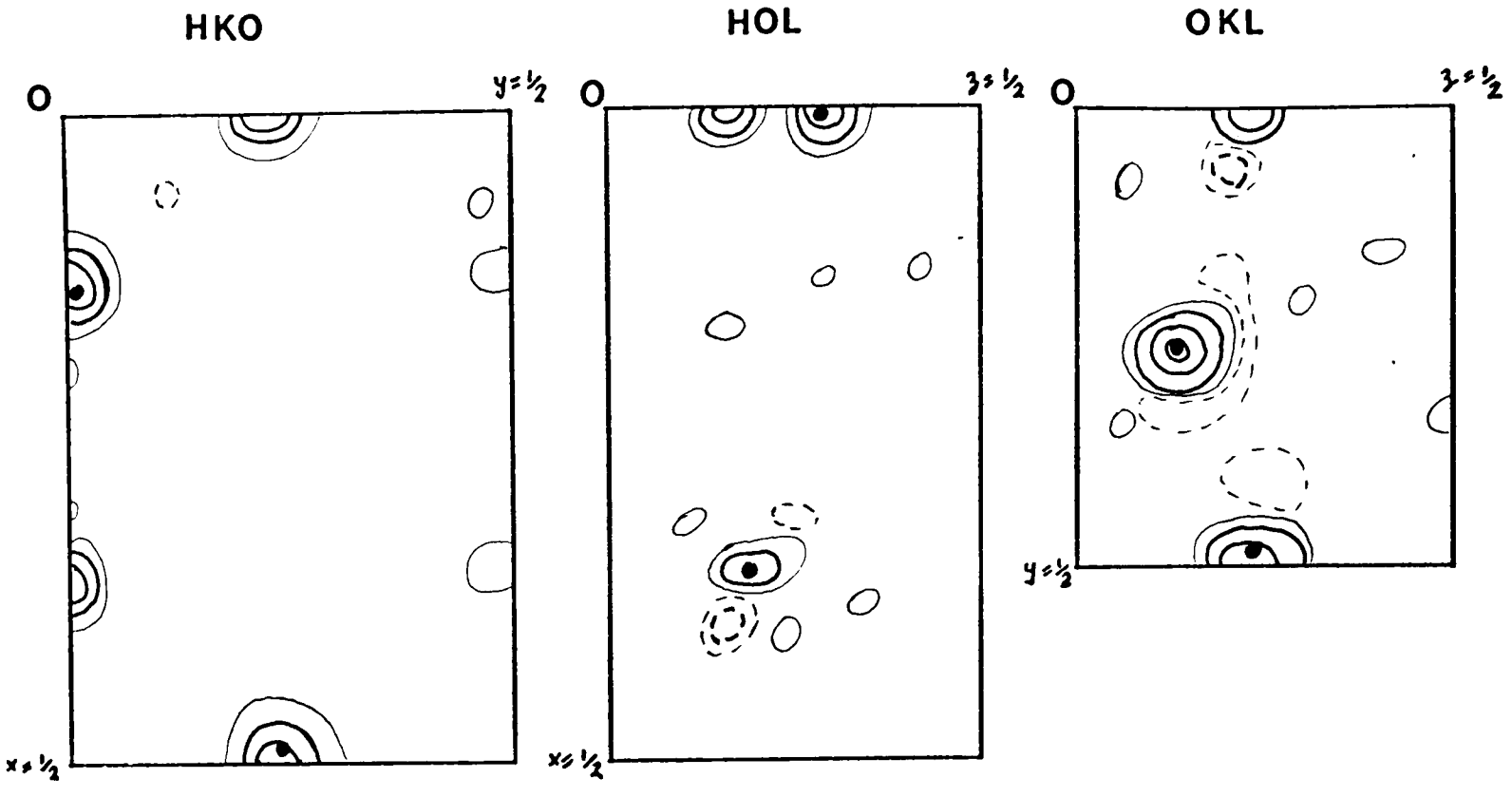
measurements were rejected as having significantly unequal backgrounds. The crystal used for the measurements of the EMP derivative had previously been used for the collection of intensities for the three centrosymmetric projections, but inspection of the reference reflections did not show any appreciable loss of intensity even at the end of the three dimensional data collection.

Difficulties encountered with the control system of the diffractometer during the collection of the intensities of the BAKER derivative meant that only a single Friedel equivalent was measured for each reflection, so that there are no anomalous scattering measurements. A complete failure of the control system involved a separation of one week between the measurement of the first half of the reflections, including all the centric ones, and the second half. These two sets of measurements were scaled separately to the native structure amplitudes. Analysis of the scale factor as a function of the index h showed a clearly defined discrepancy between the values appropriate to the sets of data. The structure amplitudes in the second set were multiplied by 1.072 to correct for this and allow all the measurements to be subsequently treated as a single set.

Measurements of the chloroplatinite derivative were made only for the centric reflections, and as discussed in the previous section this derivative can not yet be adequately reproduced.

The overall agreement between the Bijvoet pairs of reflections, as determined by the statistic $\langle|\Delta F|\rangle/\langle \bar{F} \rangle$ is essentially what would be expected for measurements having these estimated standard deviations, calculations based on Wilson statistics (Wilson, 1942) predicting values of about 4.5% and 7% for the PCMB and EMP derivatives respectively. The statistical accuracy of the measurements is lowest for

Difference Fourier ΔF



Double Difference Fourier ΔΔF

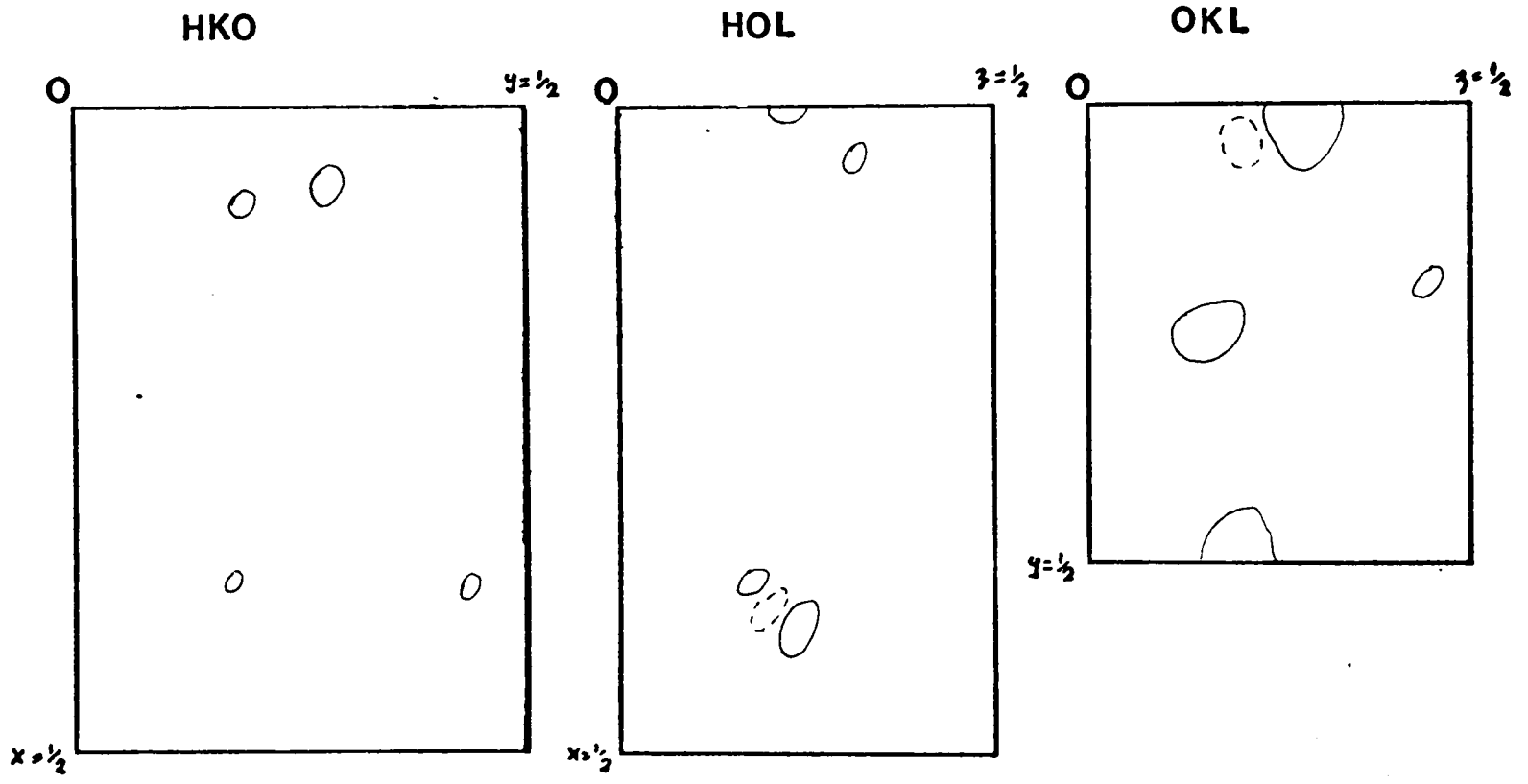


FIGURE 16 Projections of 6 Å Best Sign ΔF and ΔΔF Maps: PCMB Contours are drawn at intervals of 0·5 e.Å^{-3} above zero, the faint contours being at 0·25 e.Å^{-3}.

the chloroplatinite derivative. This arises from that crystal giving the weakest diffraction intensities as is reflected by the scale factor to the native measurements being largest for this derivative. In every case preliminary scaling made the ratio of the averages of the derivative and native structure amplitudes equal to 1.02, this scale factor then being refined together with the other heavy atom parameters using the centric reflections of each derivative. Values listed for the refined scale factors are relative to the preliminary scale factor.

For the refinement using the low resolution data, all the temperature factors were fixed at a value $B/\lambda^2 = 7.0$ until the occupancy and positional parameters of each site and the overall scale factor had been determined. The temperature factor was then refined whilst keeping all the other parameters constant. The results of these refinements are now considered for each derivative separately, following an analysis of the difference Patterson map and its interpretation in terms of specific sites of heavy atom binding.

E) LOW RESOLUTION ANALYSIS OF THE PCMB DERIVATIVE

Interpretation of the difference Patterson map for the PCMB derivative followed directly from the earlier solution of the co-crystallised mercurial derivatives. The projections of the PCMB difference Patterson map show the same principal features as those of the other two derivatives as can be seen by comparison of Figures 14 and 15. The Harker sections of the PCMB map are also shown, and the only features, apart from the origin peak, which do not lie on these sections, are the two doubly weighted cross vector peaks on the section

TABLE 6 Parameters After Refinement of Derivatives at 6Å Resolution. The first four columns show the parameters after independent refinement of each derivative and the final column the corrected parameters subsequently obtained for the BAKER derivative, after correlation of the results from all the sulphydryl derivatives. Occupancies Z and RMS lack of closures E should be multiplied by 3 to transform them to approximately an absolute scale.

Derivative		PCMB	EMP	BAKER	$PtCl_4$		BAKER
$\langle f_H \rangle$		45.	55.	36.	28.		39.
Reliability Index R		0.406	0.450	0.411	0.549		0.376
RMS lack of closure E		24.2	32.8	20.7	23.1		19.2
Number of sites		2	4	4	2		4
Refined Scale k		1.0023	1.0199	0.9951	0.9881		0.9939
Site 1	Z	.2385	.1933	.2285	.1559		.2237
	B/λ^2	7.0	14.0	7.0	7.0		7.0
	x	.1237	.1290	.1213	.1231		.1218
	y	.0120	.0152	-.0116	.0087		.0063
	z	.7348	.7443	.7702	.0164		.7666
Site 2	Z	.2714	.2065	.1720	.1615		.1749
	B/λ^2	7.0	2.0	7.0	7.0		15.0
	x	-.0085	-.0079	.0025	.4533		-.0022
	y	.2466	.2391	.2521	.2761		.2442
	z	.1713	.1681	.1803	.2165		.1797
Site 3	Z		.2390	.1253			.1527
	B/λ^2		7.0	7.0			8.5
	x		.0078	.0790			.0798
	y		.0481	.0238			-.0251
	z		.0446	.7793			.7870
Site 4	Z		.2420	.1042			.1268
	B/λ^2		7.0	7.0			21.0
	x		.0608	-.0402			.0412
	y		.1965	.2374			.2662
	z		.6704	.1540			.1620

v = 1/4, corresponding to the sulphydryl sites at

(.123, 0, 3/4) and (0, 1/4,.167).

Different sulphydryl derivatives show more variation in the secondary features of their difference Pattersons, but they all share a tendency for such features to show a higher symmetry than that of the Patterson space group mmm. All of the projections of the difference Patterson shown in Figures 14 and 15 appear to have an approximate centre of symmetry at the point having projected coordinates of (1/4, 1/4). This tendency is least apparent in the uOw projections, in correlation with the interpretation which places a mercury atom at a position (1/4, 1/4) relative to the centre of symmetry of the structure projected along the x and z axes but not along the y axis. The first and second of the above sites can be shown to satisfy this condition for the structure projected along the z and x axes respectively.

The uvO projections of the difference Patterson map show even higher symmetry with apparent lines of reflection such that the point (1/4, 1/4) appears to have the symmetry of the plane group mm. The three dimensional difference Patterson map shows evidence of most of these secondary features, especially those close to the origin along all three of the principal axes corresponding to vectors of about 10Å. These three peaks were those that seemed to be most sensitive to the relative weighting of the centric and acentric reflections in calculating the difference Patterson map, their height being considerably reduced when the amplitudes of the differences for the centric reflections were reduced by a factor of $2/\pi = 0.64$ as suggested by Moews and Bunn (1971). The theory justifying the use of

facing page 97

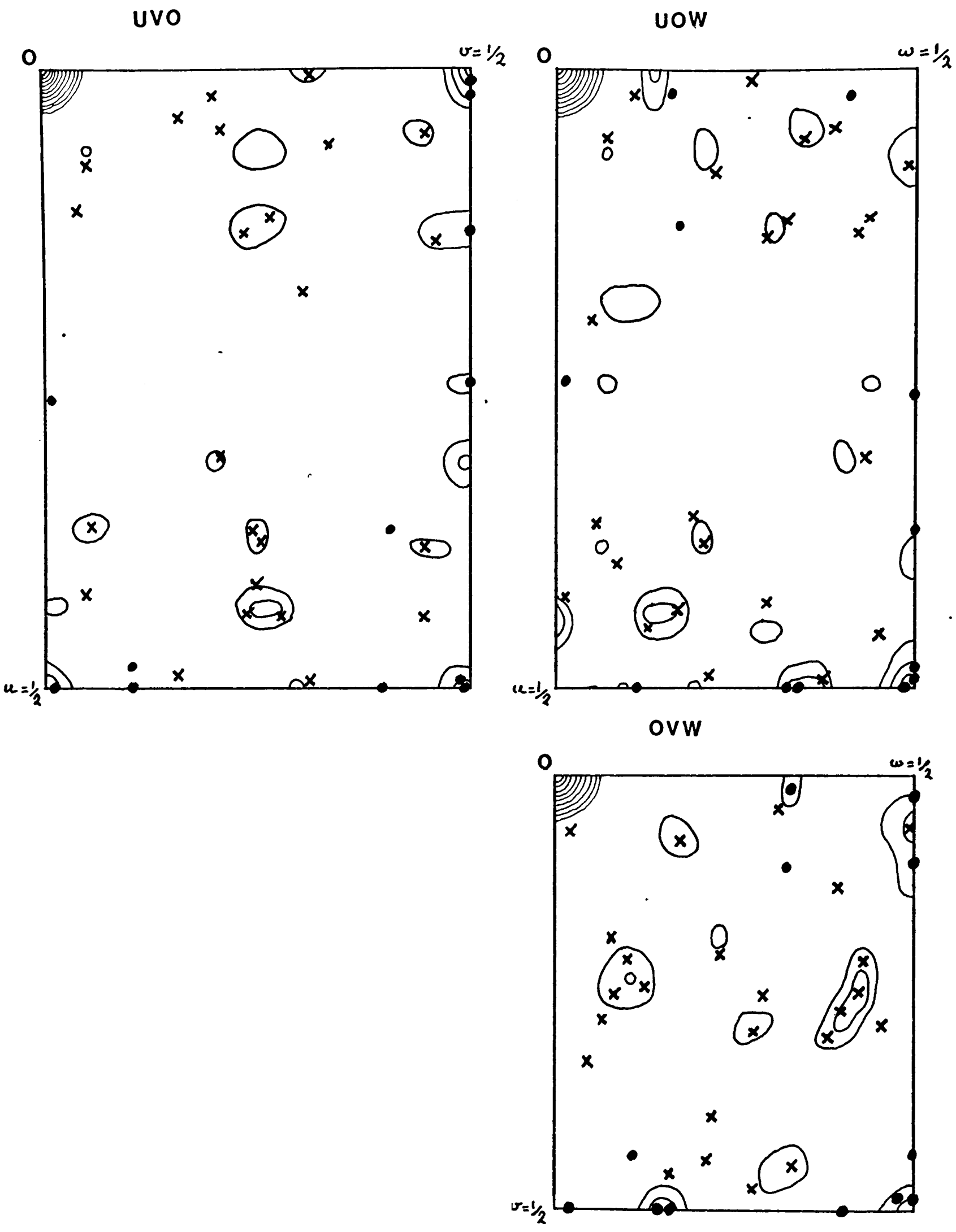


FIGURE 17 Projections of 6Å Difference Patterson: EMP

relative weights in various applications of Fourier series is considered in greater detail in Appendix II where it is shown that a more appropriate factor to use is $1/\sqrt{2} = 0.71$.

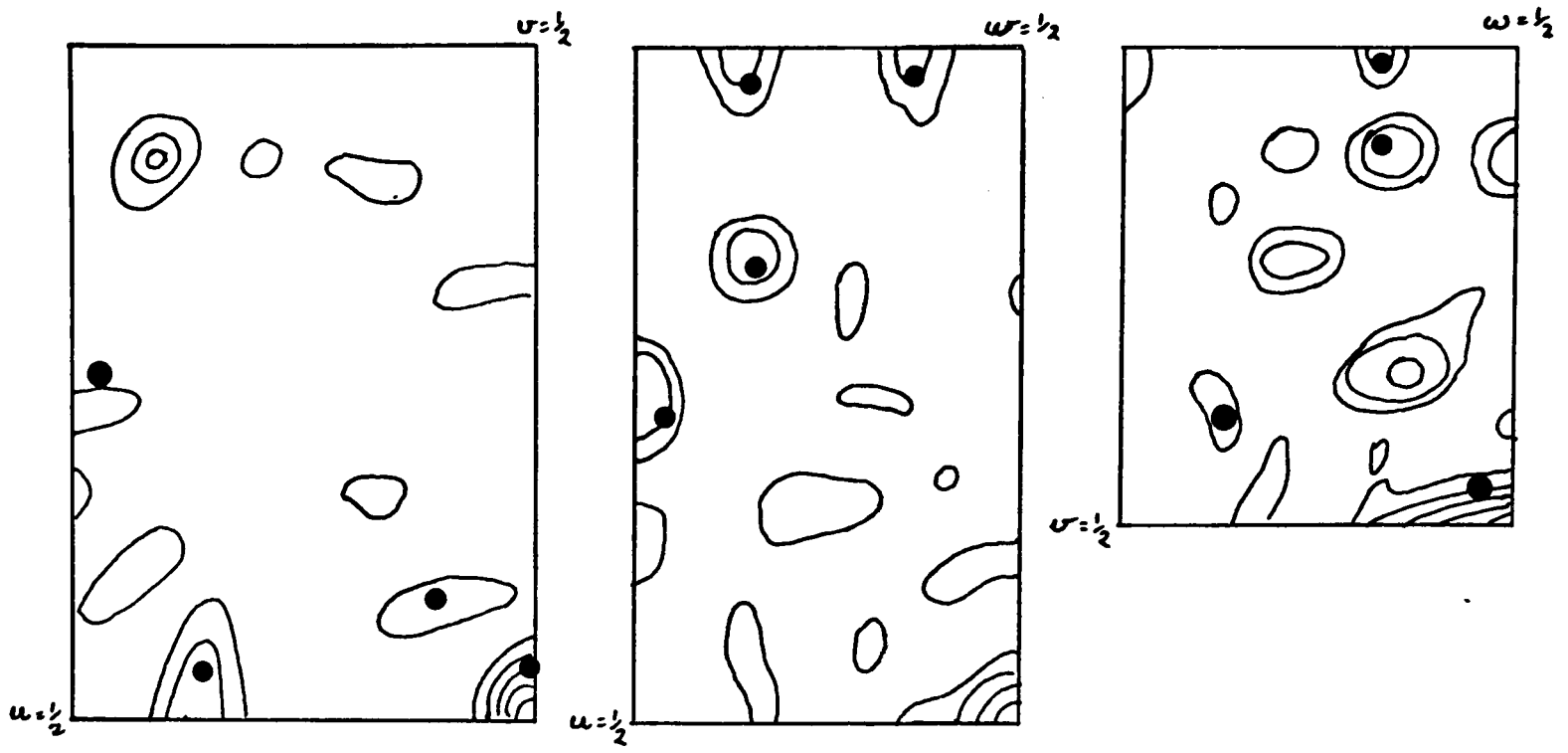
A convincing interpretation of these secondary features, which are of considerably lower significance than almost all of the features involved in the two site solution of the vector map, could not be made even with techniques such as the minimum superposition function which was applied to the uvO projection. Thus the parameters of the two sites of mercury binding were refined by the methods already described.

Difference Fourier maps were calculated in projection for self sign maps and in both three dimensions and projection for cross sign maps, the phases in the latter case being the best phases determined from all the four derivatives used at 6Å resolution. The projections of these best-sign difference and double difference maps are shown in Figure 16. No significant minor binding sites are apparent on these or on the three dimensional maps, and so the refinement for the PCMB derivative assuming two sites has produced a final set of parameters which are listed in Table 6 with those of the other derivatives, the reliability index being 0.406 for these parameters.

F) LOW RESOLUTION ANALYSIS OF THE EMP DERIVATIVE

The projections of the difference Patterson map for the EMP derivative which are shown in Figure 17 contain many similar features to those of the analogous PCMB maps, but have additional peaks indicative of other major binding sites, so that the EMP maps do not exhibit such a high degree of hyper-symmetry. These projection maps were interpreted

Harker Sections



Difference Fourier ΔF (2 site self sign)

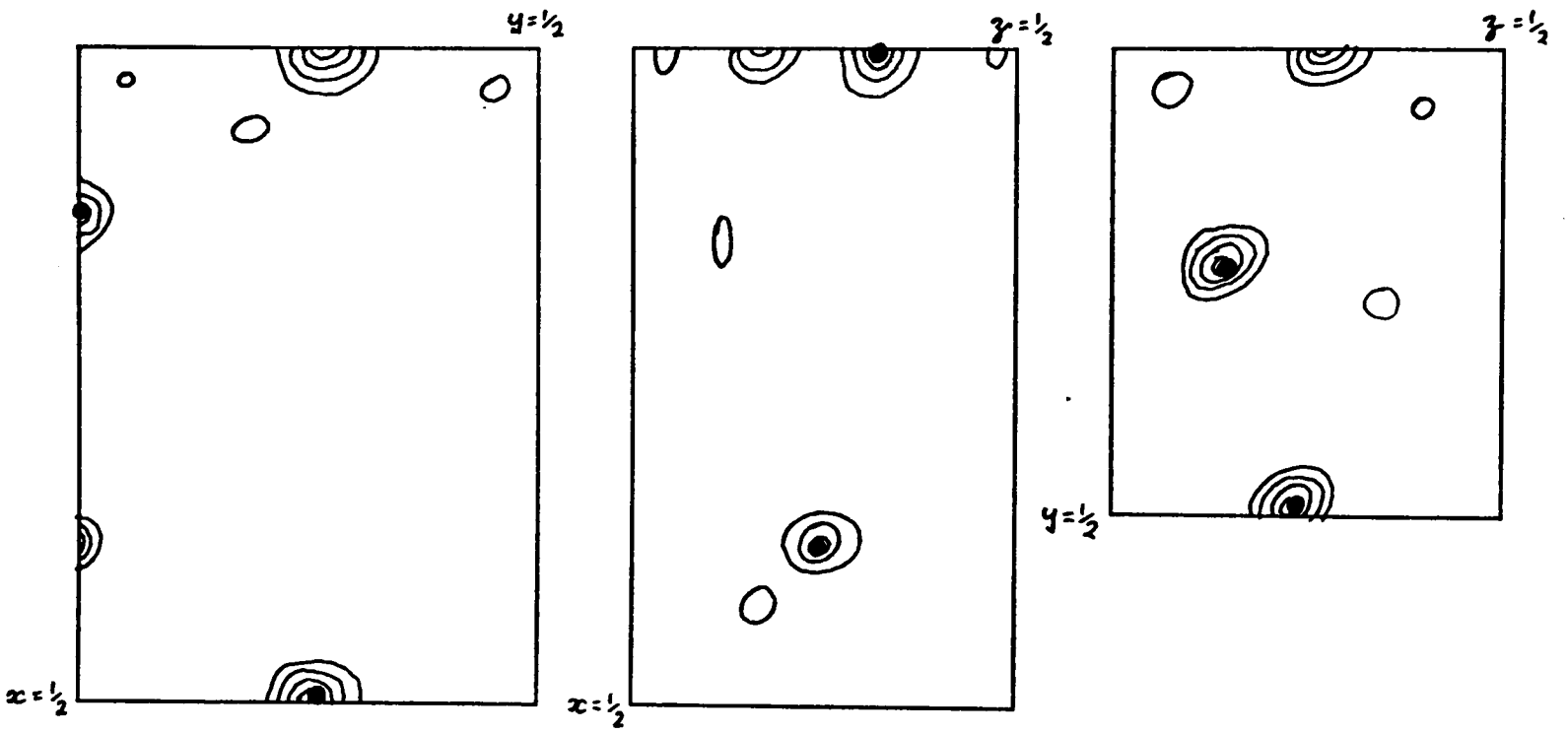


FIGURE 18 Harker Sections of 6Å Difference Patterson and Projections of 2 Site ΔF Map: EMP.

by D.W. Banner in terms of four major binding sites, the two sulphydryl sites and two other sites in more general positions. He also showed that this interpretation was consistent with the indications of correlation maps relating the isomorphous differences of this derivative to those observed for both PCMB and chloroplatinite.

This interpretation was confirmed by the three dimensional difference Patterson map, the Harker sections of this which are shown in Figure 18 illustrating the solution more clearly than the projection maps. The coordinates of the third and fourth sites were further confirmed by the difference Fourier projections, calculated using either the signs predicted by the PCMB derivative as cross-sign ΔF maps, or the signs predicted after refinement of the two sulphydryl sites alone for the EMP derivative as self-sign ΔF maps. The reliability index for these parameters was 0.58 and the self-sign ΔF maps are shown in Figure 18.

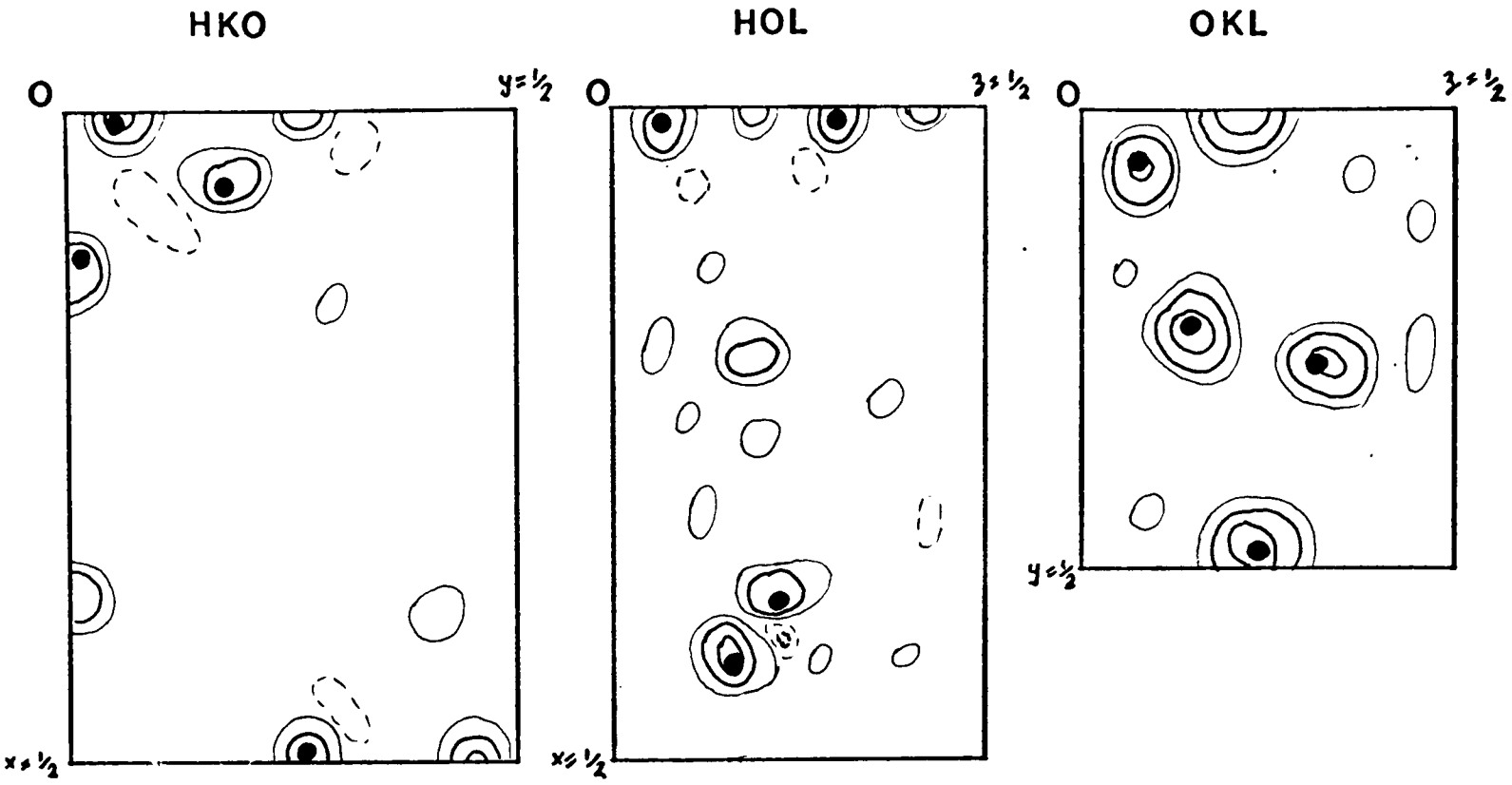
Refinement of the parameters for four sites proceeded from the initial coordinates

(.123, 0, 3/4) (0, 1/4, .167)

(.010, .040, .050) (.070, .200, .670),

similar occupancies being assumed for all the sites. Self-sign difference and double difference Fouriers calculated after the refinement were not as featureless as those for the PCMB derivative but no additional sites were found for which there were consistent indications from all three projections, and from three dimensional difference Fouriers when these were calculated using best phases from all four derivatives combined. The three projections of these best sign ΔF and $\Delta\Delta F$ maps are shown in Figure 19. The parameters for the

Difference Fourier ΔF



Double Difference Fourier ΔΔF

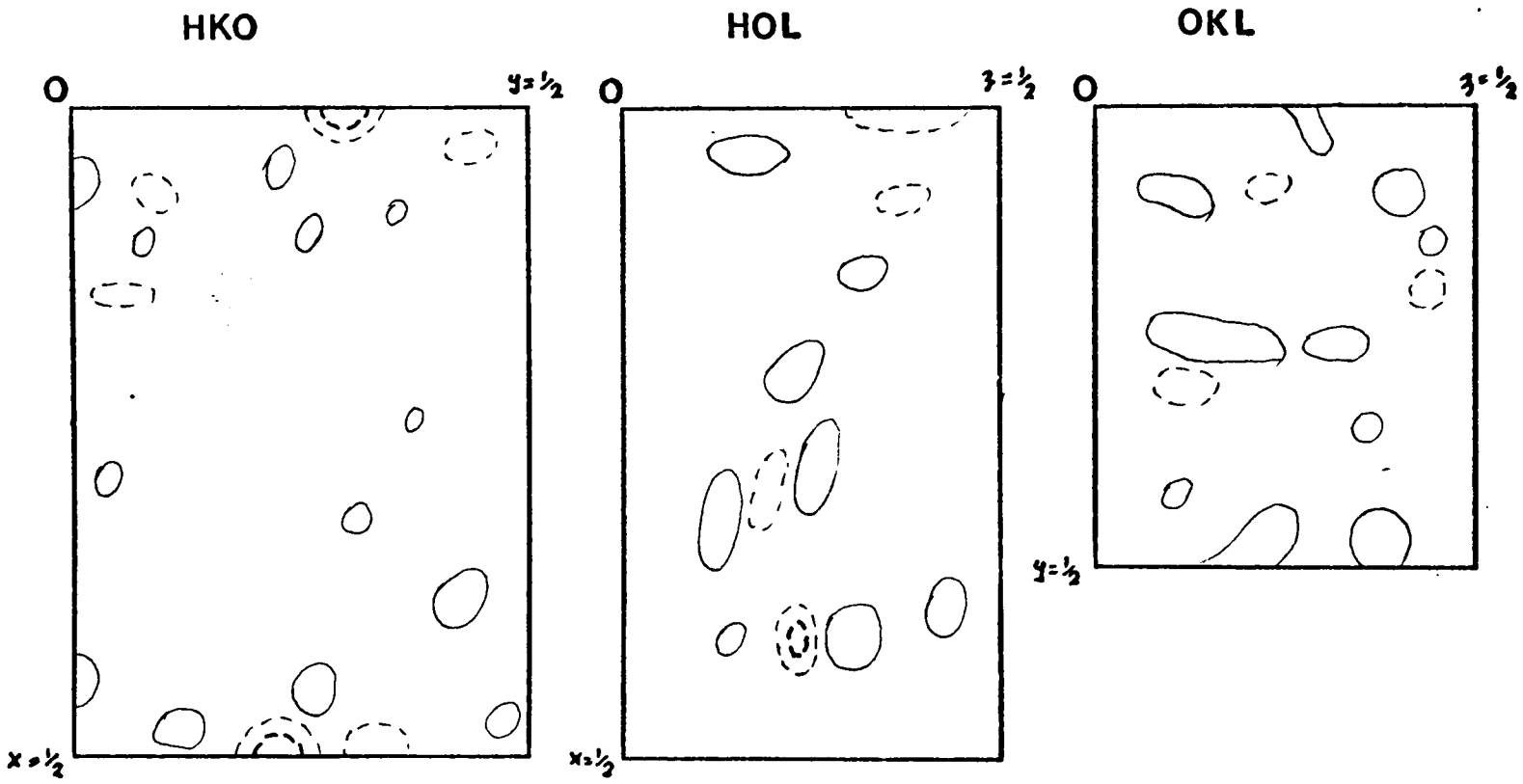


FIGURE 19 Projections of 6 Å Best Sign ΔF and ΔΔF Maps: EMP Contours are drawn at intervals of 0.5 e.Å^{-3} above zero, the faint contours being at 0.25 e.Å^{-3}.

EMP derivative after refinement for these four sites are listed in Table 6, the reliability index being 0.45 when the refinement was complete

G) LOW RESOLUTION ANALYSIS OF THE BAKER DERIVATIVE

Projections of the difference Patterson map for the BAKER derivative showed the familiar features, indicative of mercury binding at the two sulphydryl sites as seen in Figure 20. Some of the peaks are clearly elongated, especially along the direction parallel to the x axis of the structure, but there are very few additional peaks. The Harker sections of the three dimensional map, which are also shown in Figure 20, contain more features than are explicable by the sulphydryl sites alone, and the cross vectors are considerably extended as would be expected from their appearance in projection. These indications of the presence of additional mercury atoms close to those at the sulphydryl sites were not really sufficient to allow the determination of their coordinates, especially in consideration of the special positions occupied by the major sites.

A self-sign difference Fourier map was calculated in three projections after refinement of the parameters for two mercury binding sites. These projection maps showed clearly the presence of at least two subsidiary mercury sites as shown in Figure 21. Refinement of the parameters for four sites from the initial coordinates

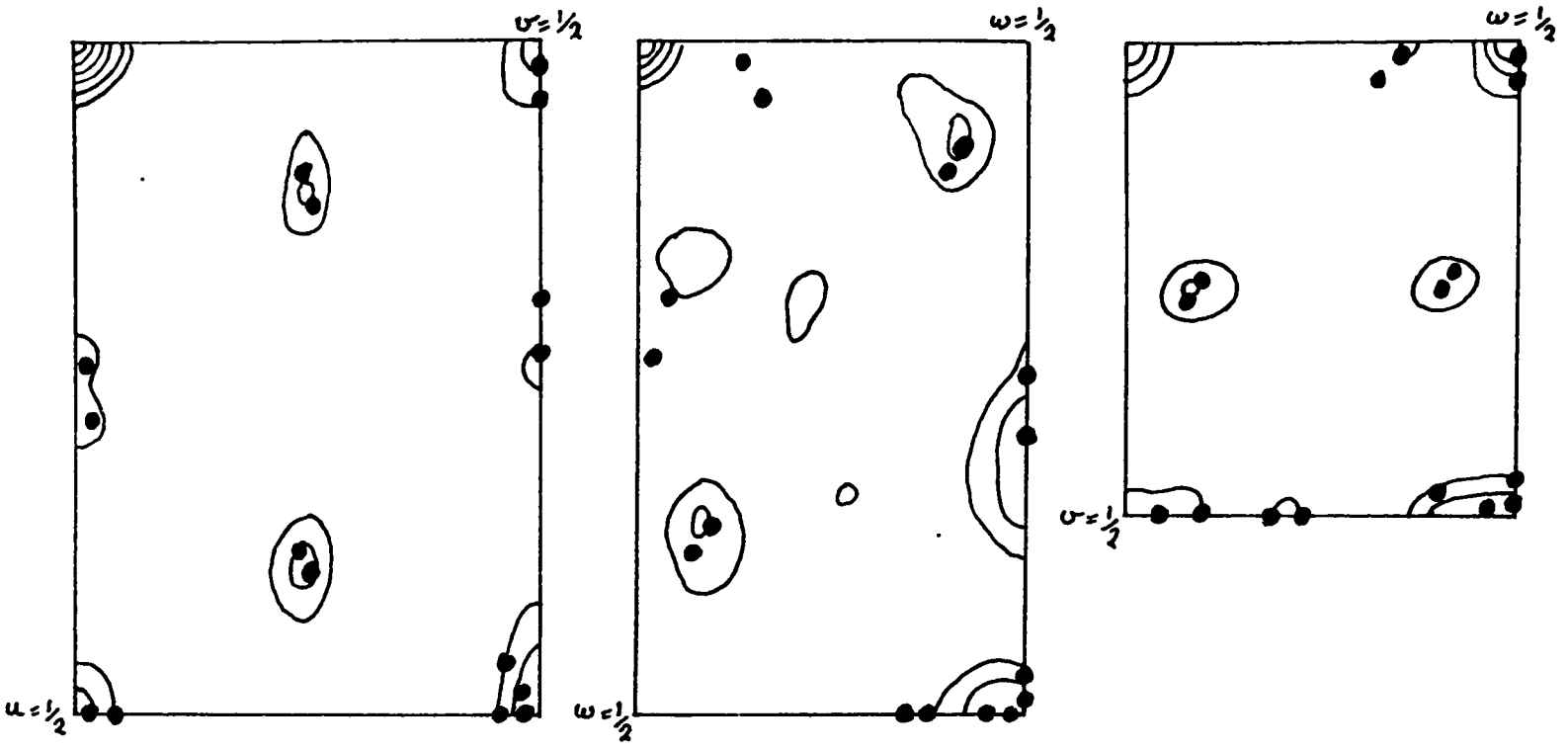
(.120,-.010, .770) (.002, .253, .178)

(.070, .020, .790) (-.040, .230, .170)

showed that the occupancy of the minor sites was about 60% of that of

facing page 100

Projections



Harker Sections

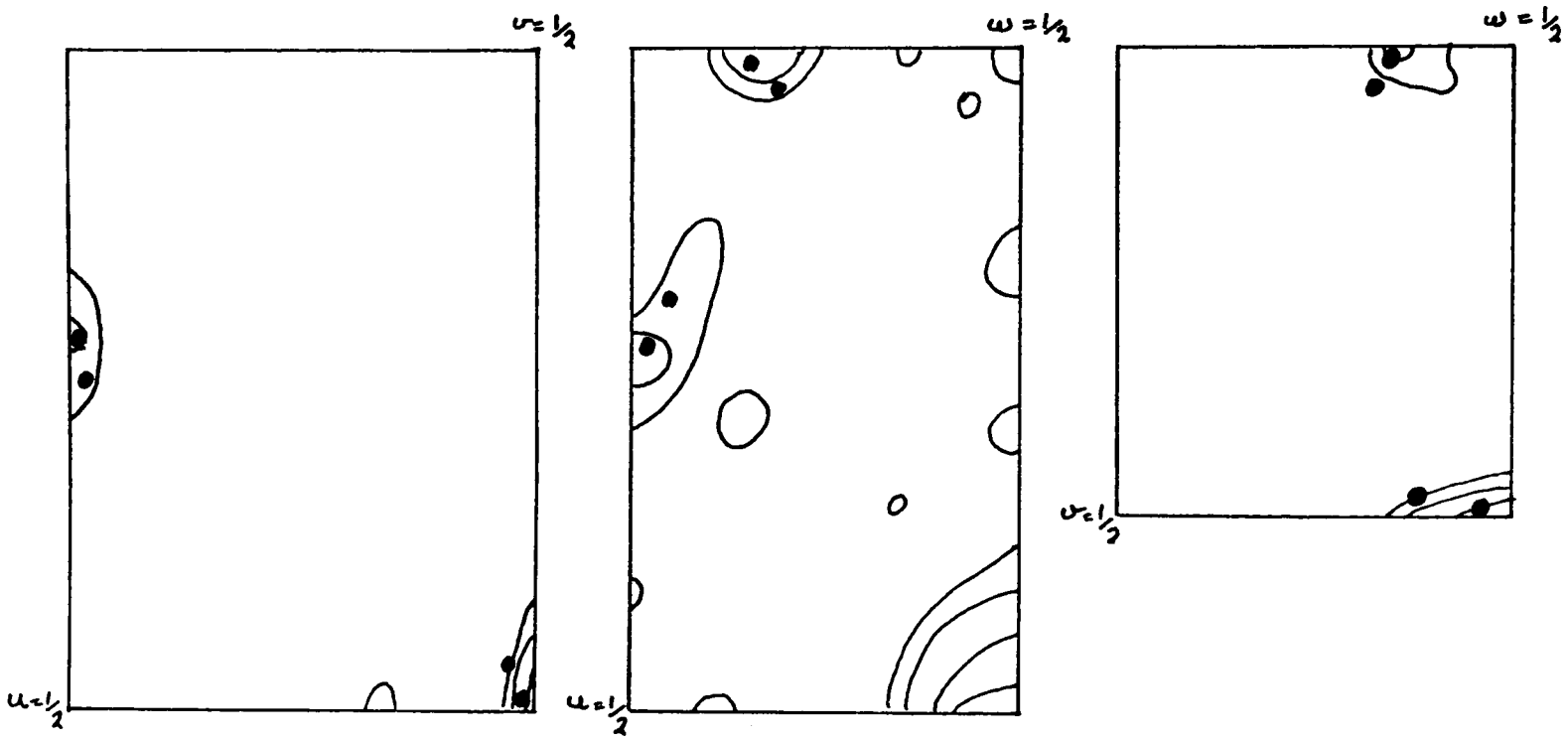


FIGURE 20 Projections and Harker Sections of 6Å Difference Patterson: BAKER

facing page 100

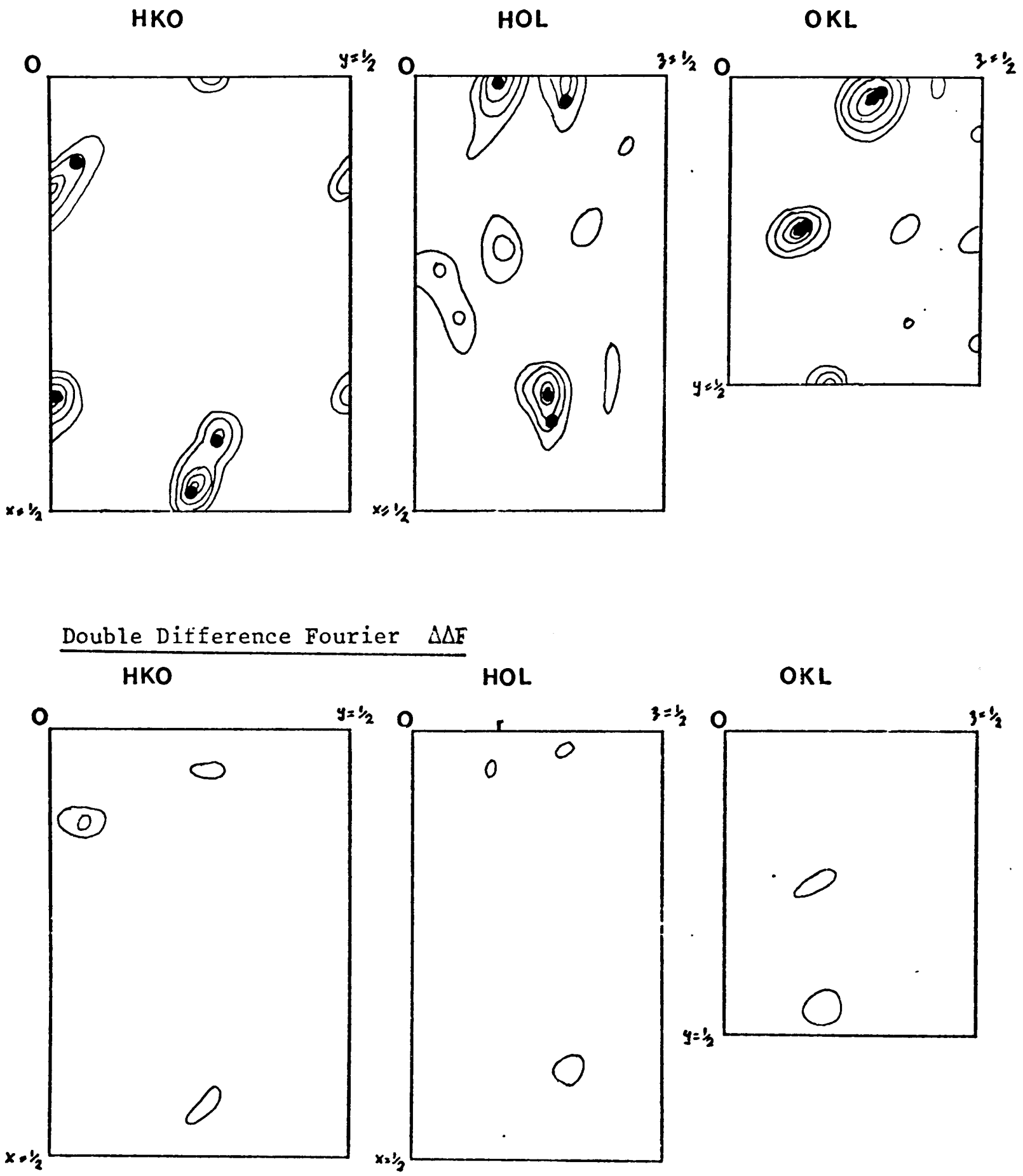


FIGURE 21 Projections of 6 Å Self Sign ΔF and ΔΔF Maps: BAKER
Contours are drawn at intervals of 1 e.Å^{-3} above zero,
the faint contours being at 0.5 e.Å^{-3}.
The signs were determined by the two main sulphydryl sites.

the major sites. This derivative would thus provide information for the determination of phases which is significantly different from that provided by the PCMB derivative.

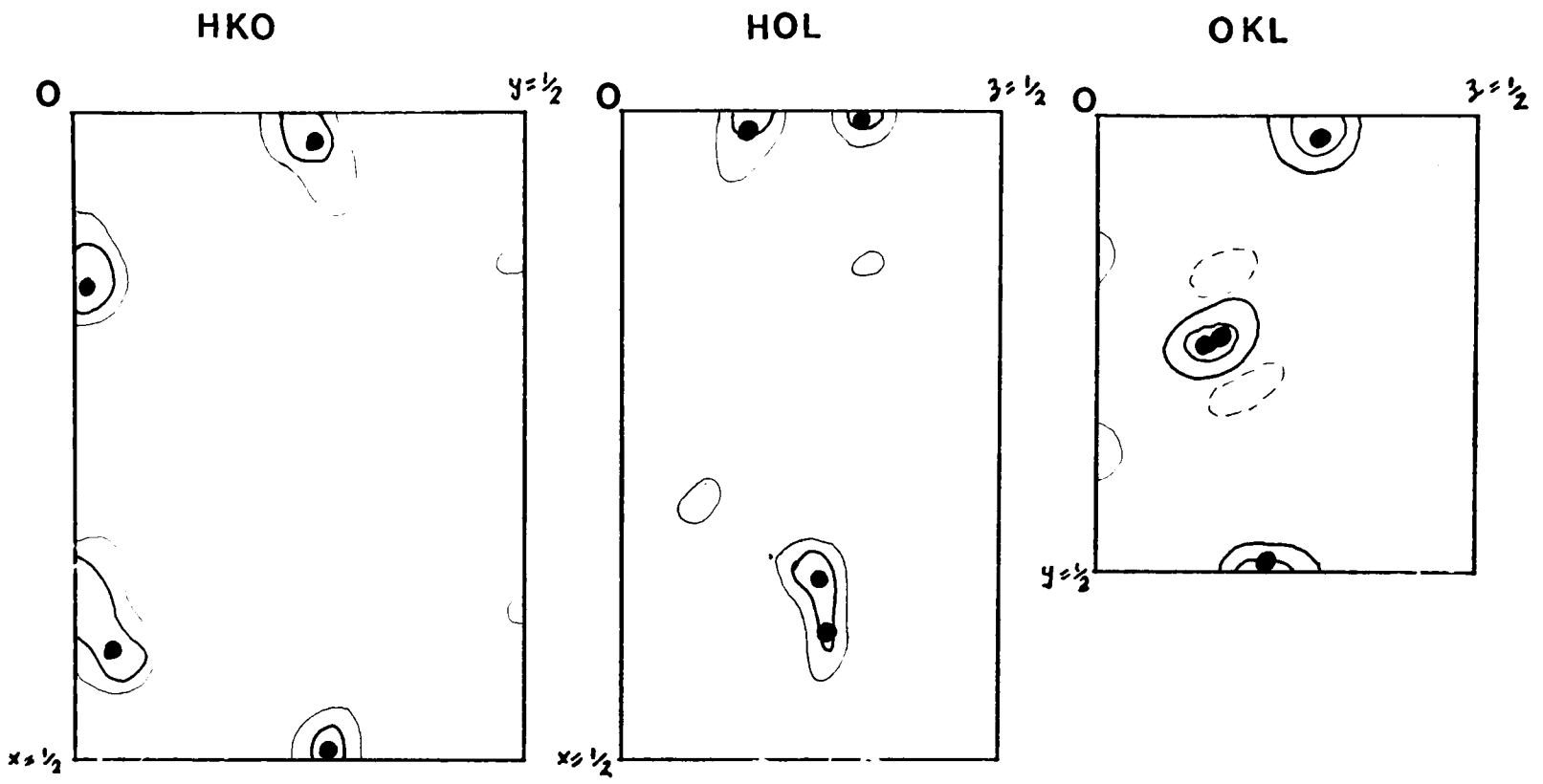
Refinement of the parameters for four mercury positions gave a reliability index of 0.411, which represents a significant improvement on the value of 0.530 obtained with only the two major mercury sites. These four mercury sites are indicated in the projections of the difference and double difference maps calculated with the best phases subsequently determined using all four derivatives as shown in Figure 22, no other sites being indicated.

The double difference maps of Figure 21 indicate a fifth site in a position close to the second sulphydryl site. Inclusion of this site in a refinement, and the calculation of a further set of self sign difference Fourier projections indicated a sixth site which was close to the first sulphydryl site. The positions of these two extra sites were closely related to those of the other two minor sites, by shifts directly across the special positions. The significance of this fact was not appreciated until the corresponding refinements of the 2.5Å resolution data were being considered, but it is now clear that the indications of these extra sites are a reflection of the weak correlation between the parameters of the main sulphydryl sites, which were subsequently shown to have refined away from the special positions in directions which were mutually inconsistent.

H) CORRELATION OF ATOMIC POSITIONS IN THE SULPHYDRYL DERIVATIVES

The parameters obtained after independent refinement of each derivatives are listed in Table 6, the coordinates of the sulphydryl

Difference Fourier ΔF



Double Difference Fourier ΔΔF

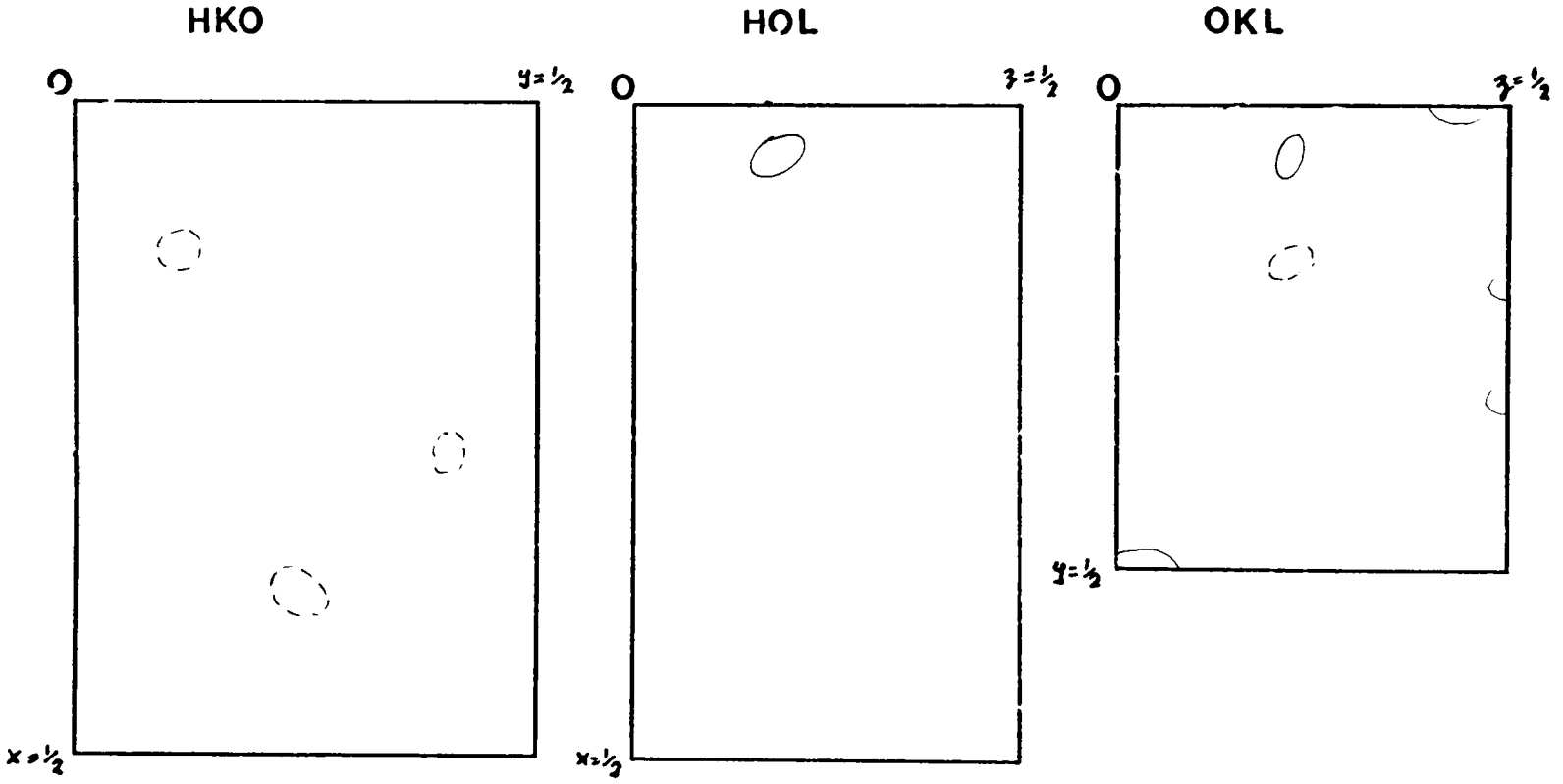


FIGURE 22 Projections of 6 Å Best Sign ΔF and ΔΔF Maps: BAKER Contours are drawn at intervals of 0.5 e.Å^{-3} above zero, the faint contours being at 0.25 e.Å^{-3}.

sites being similar in every case. However, the special positions of these sulphydryl sites allow sixteen possible directions of refinement away from the initial coordinates as listed in Table 7, but it can be shown that only eight of these represent different structures, the others simply representing an origin shift.

Closer inspection of the coordinates for the two major sites of the BAKER derivative

(.121,-.012, .770) and (.003, .252, .180)

revealed that transformation of these to a new origin at (1/2, 1/2, 1/2) gave coordinates

(.121, .012, .730) and (-.003, .248, .180)

which were in much closer agreement with those of the PCMB and EMP derivatives than were the original coordinates. Cross sign difference maps for the BAKER derivative were consistent with this change of coordinates in every case except that of the z coordinate of the first site, for which the indications were still that it should be slightly more than the exact value z = 3/4. This uncertainty was not clarified at all by the corresponding difference maps using 2.5Å resolution data.

The correlation between these two sulphydryl sites has been shown to be very weak (Table 2) and so the possibility of there being internal inconsistencies within any derivative could not be ignored even at this stage. The following procedure was adopted to detect these, relying on the third and fourth sites of the EMP derivative, which are in general positions, although the first of these is less than 5Å from the origin of the unit cell.

The parameters for the four sites of the EMP derivative were

TABLE 7 Structures Resulting from Refinement of Initial Sulphydryl Site Parameters.
The sixteen possibilities for all combinations of the directions of refinement away from the special positions are listed, the second set of eight giving rise to the same structures as the first set after transforming the origin to (1/2, 1/2, 1/2). Coordinates x,y,z are given relative to the special positions (0.124, 0, 3/4) and (o, 1/4, 0.165). Only the first of the sixteen enantiomorphic structures is shown.

Structure	Site 1	Site 2
I	.124 0+y 3/4-z	0-x 1/4-y .165
II	.124 0-y 3/4+z	0-x 1/4-y .165
III	.124 0-y 3/4-z	0-x 1/4+y .165
IV	.124 0+y 3/4+z	0-x 1/4+y .165
V	.124 0-y 3/4-z	0-x 1/4-y .165
VI	.124 0+y 3/4+z	0-x 1/4-y .165
VII	.124 0+y 3/4-z	0-x 1/4+y .165
VIII	.124 0-y 3/4+z	0-x 1/4+y .165
I'	.124 0-y 3/4+z	0+x 1/4+y .165
II'	.124 0+y 3/4-z	0+x 1/4+y .165
III'	.124 0+y 3/4+z	0+x 1/4-y .165
IV'	.124 0-y 3/4-z	0+x 1/4-y .165
V'	.124 0+y 3/4+z	0+x 1/4+y .165
VI'	.124 0-y 3/4-z	0+x 1/4+y .165
VII'	.124 0-y 3/4+z	0+x 1/4-y .165
VIII'	.124 0+y 3/4-z	0+x 1/4-y .165
I''	.124 0-y 3/4+z	0-x 3/4+y .165

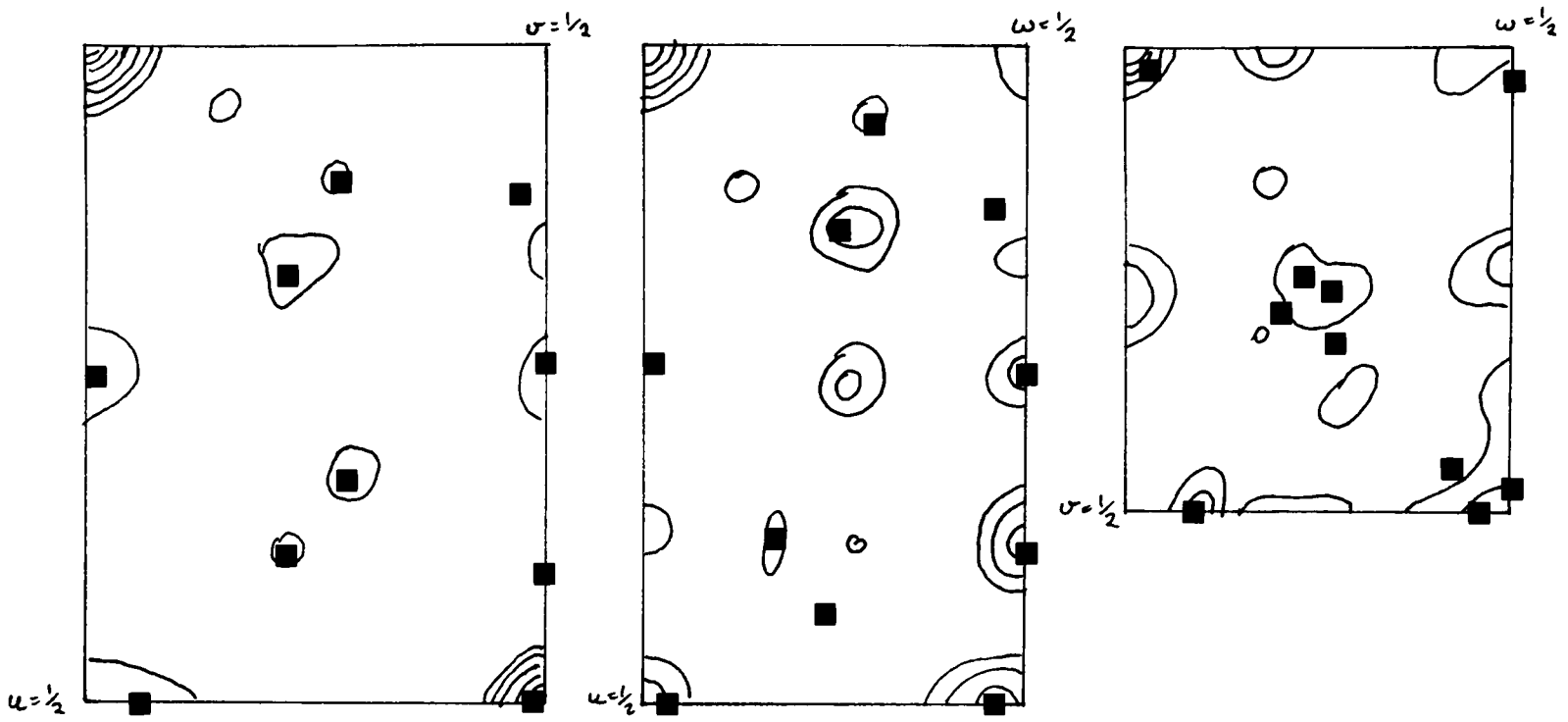
refined, subject to the coordinates of the sulphydryl sites being fixed on the exact special positions. Projections of a self sign difference Fourier map calculated at this stage indicated the directions in which the sulphydryl sites should be moved, and these were exactly as previously obtained and listed in Table 6. Cross sign difference Fourier projections calculated for the PCMB derivative, using signs determined by EMP, confirmed the directions of refinement which had been obtained previously for PCMB. A three dimensional difference Fourier map for the BAKER derivative, using the best phases calculated for PCMB and EMP, indicated four sites in positions which did not correspond exactly to either of the possibilities listed above, the coordinates of the major sites being approximately

(.123, .010, .770) and (-.005, .245, .180)

Refinement of the BAKER derivative using coordinates obtained from this three dimensional difference Fourier map gave the final parameters listed in Table 6, only four mercury sites being evident.

The three projections of this difference Fourier map indicated coordinates which were inconsistent with each other and with those from the three dimensional map. All the inconsistencies could, however, be explained in terms of the adverse effects of the superposition in projection of the major and minor sites. Thus the initial refinement for this derivative, using two sites only, produced shifts which compensated for the exclusion of the minor sites from consideration. The failure of subsequent refinement to reverse these shifts, moving the sites across the special positions, reflects the weak correlation between these sulphydryl sites. The inclusion of the fifth and sixth sites for the BAKER derivative on the evidence of the

Difference Patterson



Difference Fourier

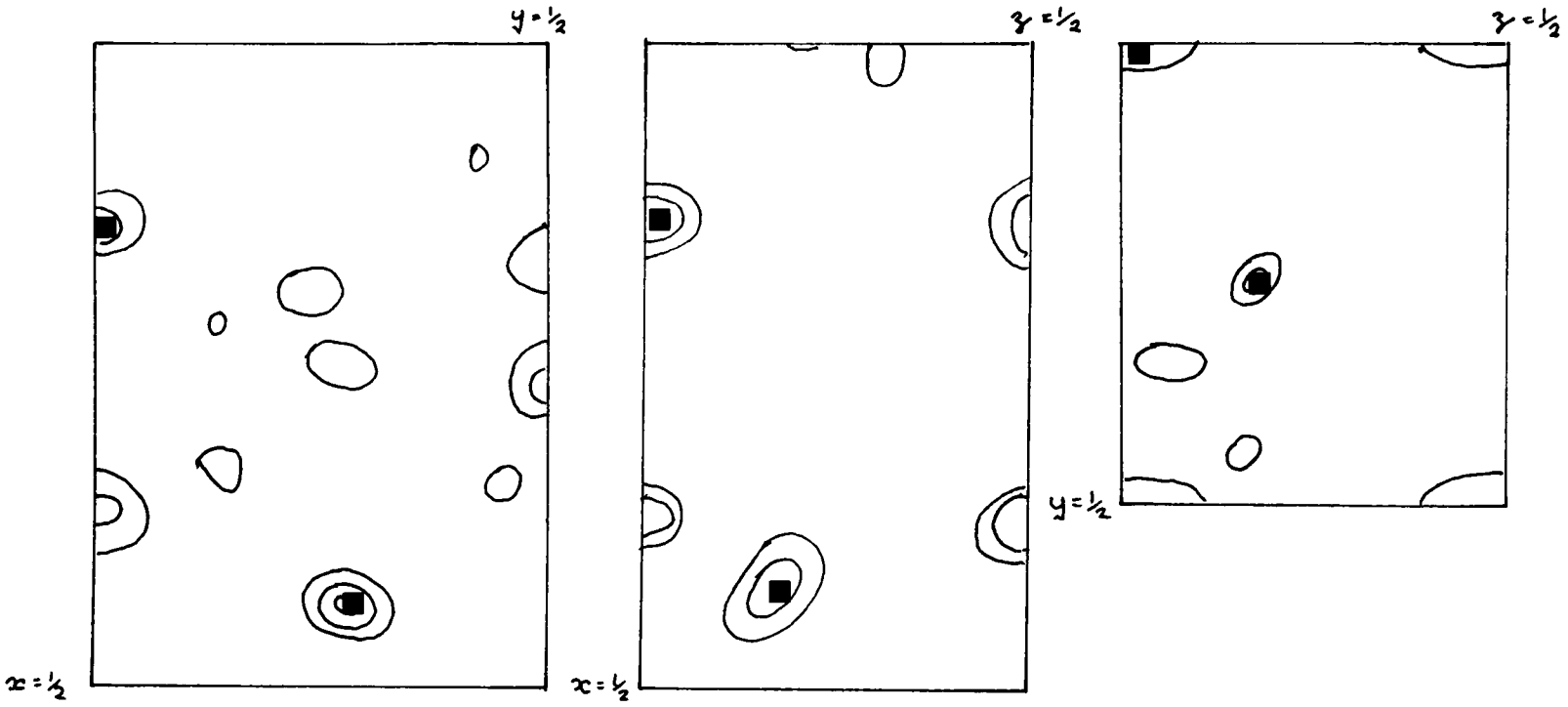


FIGURE 23 Projections of 6Å Difference Patterson and Cross Sign Fourier: $PtCl_4$.

projections of the self sign double difference map had confused the situation still further. However, the indications of the three dimensional cross sign difference Fourier map using the other two sulphydryl derivatives were completely unambiguous.

Refinement of the parameters for BAKER derivative from the starting positions indicated in this map are listed in Table 6, the final value of the reliability index being 0.376, lower than for either of the other derivatives.

I) LOW RESOLUTION ANALYSIS OF THE $PtCl_4$ DERIVATIVE

The projections of the difference Patterson map for the chloroplatinite derivative, which are shown in Figure 23 are remarkable in that the uvO projection again resembles that resulting from heavy atoms at the sulphydryl sites, even to the extent of showing the same pseudo-centre of symmetry relating the secondary features. However, the other two projections are seen to be very different from those of the sulphydryl derivatives. A consistent interpretation of all three projections involves two platinum sites with fractional coordinates

(.125, .007, .020) and (.458, .270, .223)

The assignment of the relative z coordinates of the two sites depends only on the cross vectors in the uOw projection, since those of the Ovw projection are all too close together, but the uOw projection of the difference Patterson appears more readily interpretable in this case than it is for the sulphydryl derivatives. However, the assignment of the z coordinate for the first site as .02 in preference to .52 would appear to be challenged by the h01 projection of the difference

facing page 104

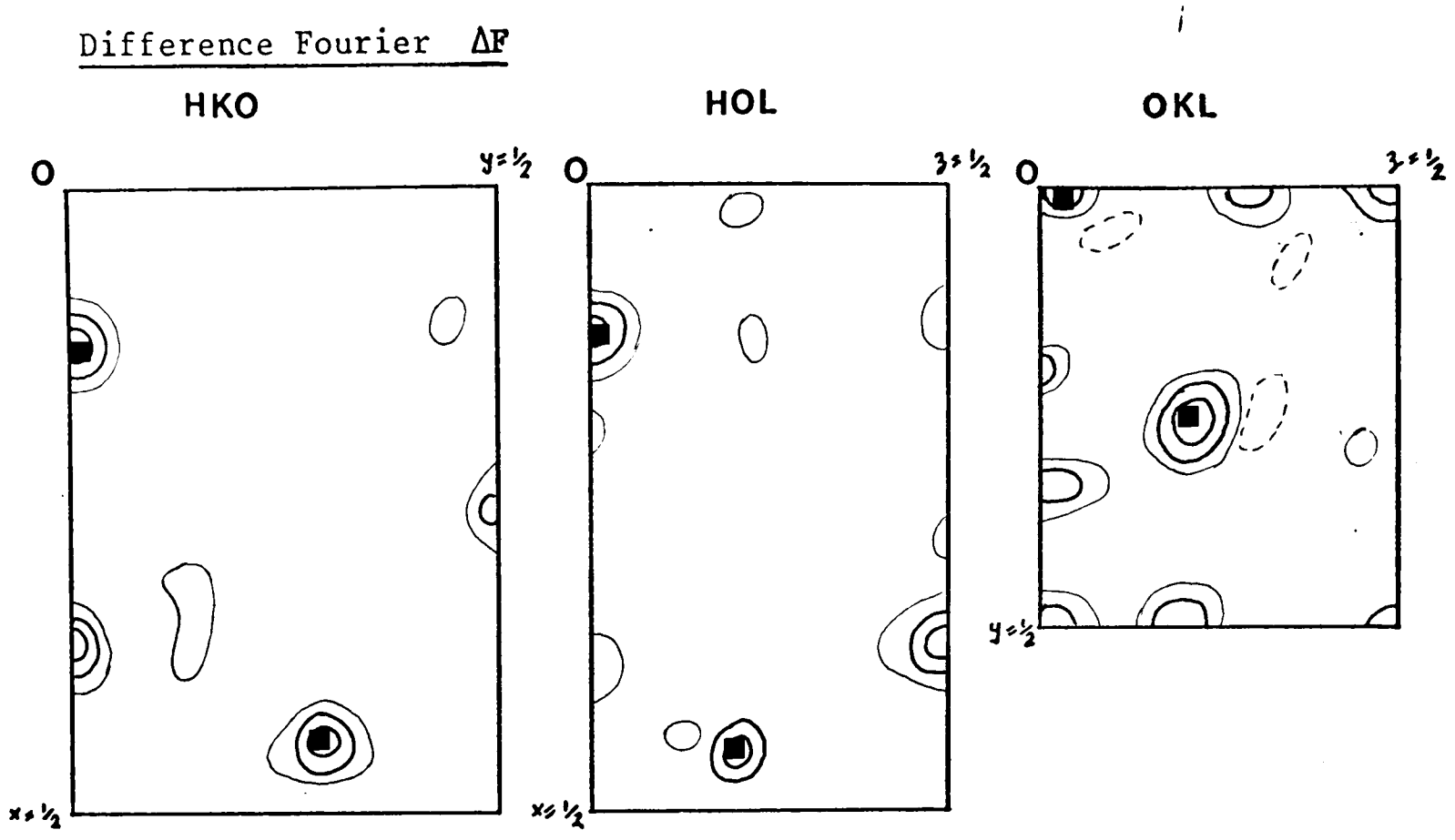


FIGURE 24 Projections of 6 Å Best Sign ΔF and ΔΔF Maps: $PtCl_4$
Contours are drawn at intervals of 0·3e.Å^{-3} above zero, the faint contours being at 0·15 e.Å^{-3}.

Fourier map using the best signs determined from the other three derivatives, seen in Figure 23, indicating that the signs of the hOl reflections having odd values of the index l are very poorly determined by the other derivatives.

Refinement of the parameters for the chloroplatinite derivative for the values z = .02 and .52, and also using both coordinates assuming three platinum sites, showed very poor discrimination between these three possibilities, as did the calculation of the best signs for all four derivatives together for each set of chloroplatinite parameters. The choice of those parameters indicated by the difference Patterson projection was later justified by the fact of the relationship between these two platinum sites by the molecular diad axis.

The reliability index after refinement of these two sites was 0.549, the final parameters being listed in Table 6. This value is much larger than that for any of the other derivatives, but the RMS lack of closure error is relatively small. Projections of the difference and double difference Fourier maps calculated with the best signs from all four derivatives are shown in Figure 24. There are no significant indications of any additional platinum sites.

Comparison of the difference Fourier projections using the best signs from three and four derivatives, seen in Figures 23 and 24 respectively for the chloroplatinite derivative, suggests that this derivative has had little effect on the overall quality of the sign determination. However, despite the continued presence of most of the residual features in the double difference maps for all the derivatives, the inclusion of the chloroplatinite derivative in the phase determination improved the internal consistency of the sign predictions, giving a 4% increase in the figure of merit, an estimate of precision which

is discussed in detail in Chapter V. The use of cross sign difference maps involving the other three derivatives in the confirmation of the positions of the platinum sites ensures that there can be no problems of correlation of these parameters with those determined for the sulphydryl derivatives.

The parameters listed in Table 6 were thus used for calculating the phases of the protein structure factors as described in Chapter V and an electron density map of the protein at 6Å resolution, whose features are considered in Chapter VI. Extension of these results to 2.5Å resolution involved fourteen times as many reflections. The measurement of their intensities and subsequent analysis is described in the following chapter.

Chapter IV DATA COLLECTION AND ANALYSIS AT 2.5Å RESOLUTION

A) INTENSITY MEASUREMENT ON THE LINEAR DIFFRACTOMETER 107

B) PRIMARY PROCESSING OF INTENSITIES 112

C) SECONDARY PROCESSING OF MEASUREMENTS FROM INDIVIDUAL CRYSTALS 115

D) COMBINATION OF MEASUREMENTS FROM DIFFERENT CRYSTALS 118

E) STATISTICS OF COMPILED DATA SETS 126

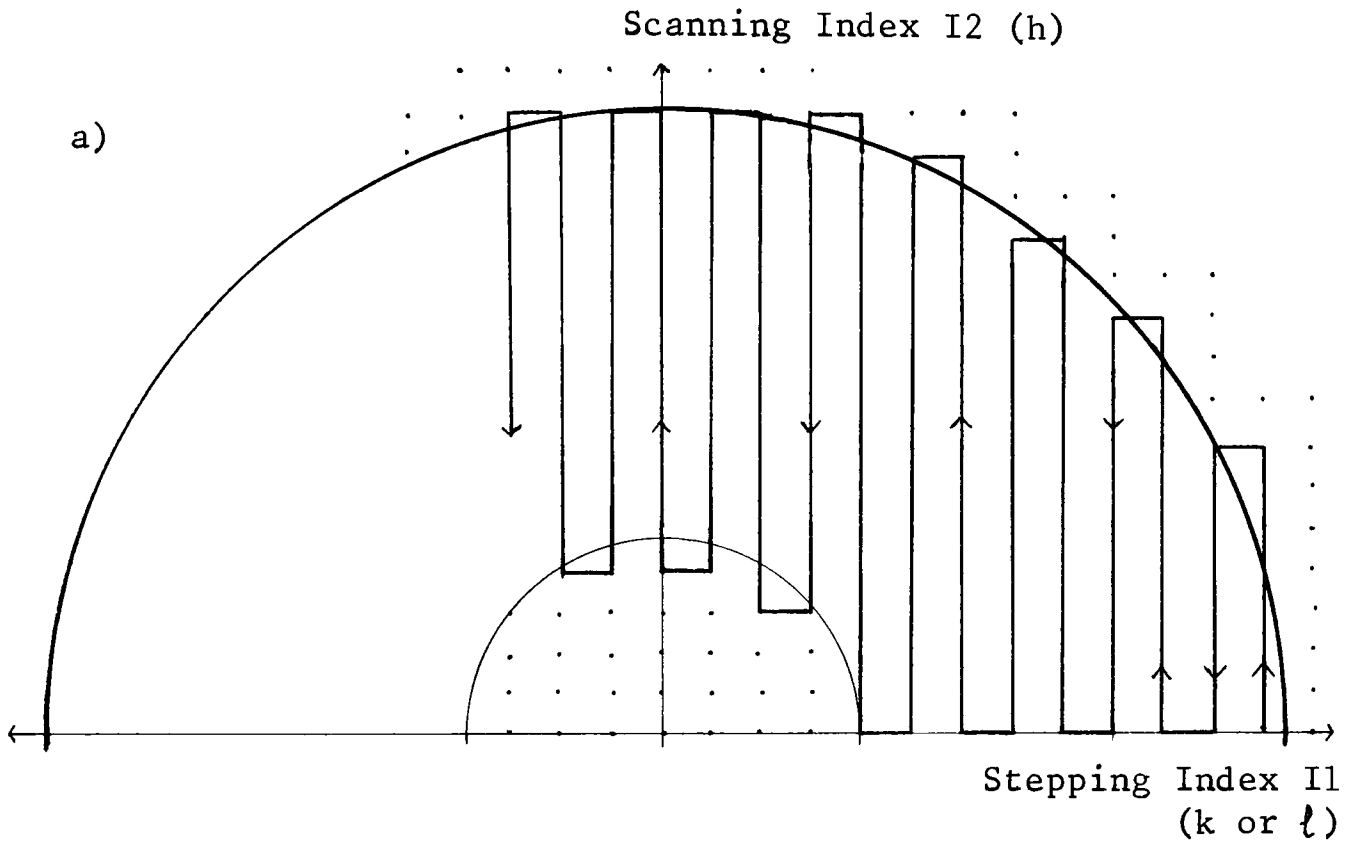
F) PRELIMINARY SCALING BETWEEN DATA SETS 129

G) INCORPORATION OF LOW RESOLUTION DATA 131

H) INITIAL REFINEMENT USING CENTRIC REFLECTIONS 132

I) STATISTICAL TESTS OF ISOMORPHOUS AND ANOMALOUS DIFFERENCES 135

Methods of Determining the Ratio f'/f'' Empirically 140



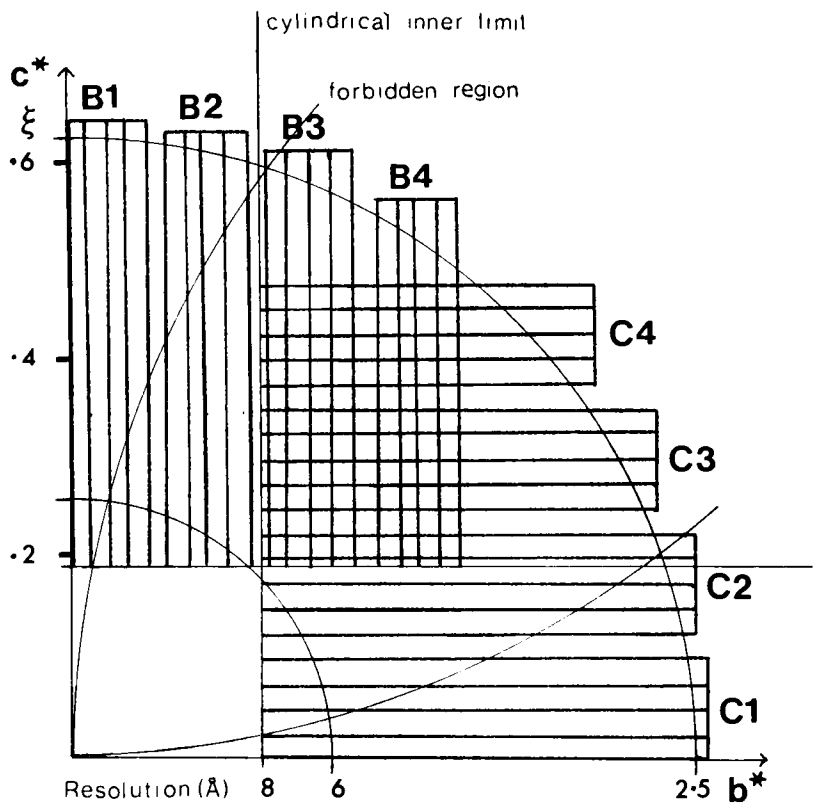


FIGURE 25 Scheme for 2.5Å Intensity Measurements for Chicken TIM.
a) Plan of a single level to show schematically the arrangement of stepping and scanning indices and the order of measuring the reflections;
b) Elevation of 2.5Å shere perpendicular to the a* axis to show the sets of levels in which the intensities were measured.

A) INTENSITY MEASUREMENT ON THE LINEAR DIFFRACTOMETER

The unit cell dimensions for crystals of chicken TIM are such that the positions of reflections in the range 6-2.5Å resolution, satisfy all the conditions necessary for the simultaneous measurement of several reflections (Phillips, 1964) and they occur in the optimum region of reciprocal space for measurement by an automatic linear diffractometer (Arndt and Phillips, 1961). A modification of this machine, which allows the quasi-simultaneous measurement of five reflections using the geometry of the flat cone setting (Arndt, North and Phillips, 1964), was used for the measurement of these intensities from crystals of chicken TIM and its derivatives.

The use of this machine necessitates several undesirable features in the strategy for the intensity measurement, but these are compensated by the efficiency of measuring 250 reflections an hour without manual intervention. The reflections are measured in levels of reciprocal space perpendicular to the real axis of the crystal which is aligned to be coincident with the vertical axis of the machine. Reciprocal space coordinates in this direction are denoted ζ, and the maximum value of this coordinate which is imposed by the construction of the instrument necessitates several sets of levels being collected using each of two different axial mountings as shown in Figure 25, in order to include all those reflections lying within an octant of the 2.5Å sphere. This involves a minimum of eight different sets of levels, providing far more information for scaling between them than is necessary, but this can not be avoided using the instrument in a fully automatic mode collecting an entire set of five levels in a single run. The only possible method of measuring the reflections automatically between inner and outer limits of the

TABLE 8 Linear Diffractometer Settings for Crystals of Chicken TIM.

a) Geometrical and instrumental parameters for the three principal axis mountings:

For each axial mounting are listed the crystal-counter distance D for measuring adjacent levels of reflections on counters separated by 0.75cm.; the maximum level index I_3 which can be obtained in the flat cone setting with a tilt of less than 30^o; the greatest angular separation between reflections measured on the centre and outermost counter Δ for these uppermost levels at their inner limit (ξ=0.2); the most convenient reflection for measuring an absorption curve; the width W at a distance D of a diffracted ray from the crystal with an angular divergence of 0.006 radians; the minimum separation of adjacent reflections in terms of the crystal orientation ($\delta\phi$) and counter position ($\delta 2\theta$) at the two extreme positions of the resiprocal lattice.

b) Appropriate Values for the eight settings used in these intensity measurements:

For the first four sets of levels about the b* and c* axes are listed the axial coordinate ζ in reciprocal space of the centre counter; the tilt μ relative to the incident beam for the flat cone setting; the approximate number N of unique reflections to be measured (8 d 2.6); the maximum value of the stepping index I_2; the three standard reflections used as references and the four used to check the derivatives, two sets of pairs the first of which always increases and the second decreases on changing from the native to derivative crystals.

a)

	D	I_3max	Δ	Absorption μ	Absorption *hkl*	Width	at I_2max δφ	at I_2max δ2θ	at I_1max δφ	at I_1max δ2θ
a*	51.6cm.	34	10'	3^o20'	8 0 0	6.5mm.	1^o56'	6'	37'	1^o15'
b*	36.2	24	21'	3^o33'	0 6 0	4.5	1^o22'	4'	26'	52'
c*	30.1	20	30'	4^o17'	0 0 6	3.8	1^o22'	4'	26'	52'

b)

	ζ	μ	N_{REF}	I_2max	References	Derivative Check Refs. EMP	Derivative Check Refs. BAKER
C1	.0499	2^o52'	940	29	9 3 0	16 0 0	16 0 0
C2	.1748	10^o 4'	910	29	6 6 0	14 1 0	13 2 0
C3	.2996	17^o27'	800	27	7 18 0	6 6 0	6 2 0
C4	.4245	25^o 9'	560	24		4 6 0	4 6 0
B1	.0412	2^o23'	790	24	10 0 3	16 0 0	16 0 0
B2	.1443	8^o20'	760	24	9 0 9	13 0 4	15 0 3
B3	.2474	14^o21'	700	23	3 0 5	2 0 8	3 0 6
B4	.3506	20^o33'	570	21		2 0 6	2 0 6

coordinate ξ, perpendicular to the axis of the instrument, is illustrated in Figure 25 from which it is seen that the measurements of a Bijvoet pair of reflections are separated by a considerable interval of time and that they are subject to completely different absorption corrections, which are both factors tending to reduce the accuracy of the data.

The decision to use b* and c* mountings of the crystals was governed by several factors. From the geometrical and instrumental parameters calculated for all three possible mountings and listed in Table 8 it is seen that the a* and b* mountings are to be preferred for reasons of firstly the spatial resolution of adjacent reflections which are more widely separated in these sets of levels, secondly a closer approximation to the conditions of true simultaneity since there is a smaller difference in the orientation of the crystal for measurements on the central and outermost counters for these mountings, and thirdly the larger crystal to counter distances for these mountings would be associated with lower background intensities because of the operation of the inverse square law. However, the a* mounting is also the least preferable one because firstly, the total number of different sets of levels which would require to be measured is at least ten, secondly the divergence of the diffracted beam is such that the aperture necessary to measure it completely at such large counter distances is too large for counters separated by a fixed vertical spacing of 7.5 mm., and thirdly the morphology of the crystals is such that they are extremely difficult to mount in this orientation. All the crystals used had maximum dimensions in the range 0.8-1.0 mm. for which these considerations were appropriate.

The choice of the b* and c* mountings which can be readily

obtained on mounting the crystals, with only four sets of levels about each axis being necessary, was the most acceptable compromise. The spatial resolution of adjacent reflections is adequate at 2.5Å resolution, as is the degree of quasi-simultaneity outside the inner limit ($\xi > 0.2$) and suitable detector apertures could be used to include the entire diffracted beam. The vertical positions for each set of five levels and the tilt of the crystal rotation axis relative to the incident X-ray beam are listed in Table 8, the nomenclature used to describe each set of levels being clear from Figure 25.

For almost all sets of levels the crystal was rotated by an angle of $1^{o}48'$ about the vertical axis (ϕ), detector apertures of 5 mm. diameter were used, values close to these being used for the remaining levels. A nickel filter was used to absorb the K_{β} radiation from the copper tube, air absorption being minimised by use of a tube containing helium between the crystal and detectors. The outer limit switch was set appropriately for each set of levels to include the reflections within the 2.5Å sphere and the inner switch was set to the value $\xi \sim 0.18$ corresponding to a resolution of 8Å in the central reciprocal lattice level.

The earlier studies of radiation damage dictated the use of a different crystal for each set of levels, requiring eight crystals for each complete set of data for native and derivative crystals. Differences in the occupancy of the heavy atom sites of the diffused derivatives within these eight crystals were minimised by soaking them all for similar lengths of time in separate aliquots of a single solution containing the heavy atom reagent. Any such differences which are present will clearly have a serious effect on the quality of the set of measurements for a derivative and the phase determining information

obtainable from it.

The availability of both the diffractometer and the appropriate crystals led to the intensities for the native and monomercurial crystals being measured several months before the diffused derivatives, which were used for the low resolution phase determination, had been investigated. The probability that the PCMB derivative would have provided equivalent phase determining information to that from the MONO derivative, but with greater precision, has to be balanced against the likelihood of eight crystals being far more homogeneous if prepared by cocrystallisation than if they are prepared by diffusion. More recently measurements were made for crystals of the EMP and BAKER derivatives, so that the 2.5Å resolution data contain sets of measurements for NATIVE, MONO, EMP and BAKER crystals.

The more recent series of measurements were made shortly after the installation of the diffractometer in a new laboratory, with a consequent series of failures of electrical components, especially in the circuitry of the proportional counters, which occurred frequently but erratically over a period of several months. The main cause of these must be assumed to be large fluctuations in the ambient temperature, which were unavoidable but very undesirable. The detection and elimination of spurious measurements resulting from such failures was a major feature of their subsequent computer processing. The intensity of the X-ray beam was much lower during the most recent measurements, than it had been earlier, thus contributing further to the statistical imprecision of these measured intensities.

Sets of levels were measured from the largest positive value of the stepping index, l and k for the b* and c* mounts respectively, scanning along successive rows in which the index h varies, throughout

the level. One set of measurements only, B1 of the BAKER derivative was made in the opposite direction, starting at negative 1 and finishing with the positive value.

The alignment of the crystals was checked immediately before and after each automatic run of measurements, and usually at least one interruption was made within a run for this purpose. The absorption curve was measured at these times, but the geometry of the linear diffractometer allows only the peak height of axial reflections in the equi-inclination setting to be measured, and not their integrated intensity. This was achieved for all the more recently used crystals by transferring them to a four circle diffractometer before or after measuring the intensities. The reference reflections which are listed in Table 8 were also monitored at these times, after checking the alignment, to give a general indication of the extent to which radiation damage was being suffered. For all crystals of the two diffused derivatives, four carefully selected reflections were measured whilst setting up the crystals to check on the occupancy of the heavy atoms, the reflections being chosen in closely adjacent pairs the members of which show significant but opposite isomorphous differences from the native crystals. The average loss of intensity of the three reference reflections during the period of the data collection is listed in Table 9 for each crystal.

The measured intensities and their backgrounds, which were measured at each limit of the crystal rotation for one half of the measurement time of the peak intensity, were recorded on punched paper tape and a teleprinter listing. A set of levels contained between five and ten thousand reflections, including both members of a Bijvoet pair, so that any close inspection of the teleprinter output was

impossible. The measurements were analysed by computer as described in the following sections, with the individual programs listed in Appendix I.

B) PRIMARY PROCESSING OF INTENSITIES

The program used for the primary processing of the measured intensities was written by Dr A.C.T. North and is a modification of his program for processing intensities measured on a four circle diffractometer, as described for the primary processing of the low resolution TIM measurements. The differences between these two versions of the program are essentially only the format expected on the input punched paper tape; the geometrical factors involved in the Lorentz correction, which can be shown (e.g. Arndt and Willis, 1966) to be

$$L = \frac{1}{\mathrm{Cos}\mu . \mathrm{Sin}\gamma} \qquad \text{where} \quad \mathrm{Sin}\mu = -\zeta$$

$$\cos\gamma = \frac{2 - (\xi^2 + \zeta^2)}{2\,(1 - \zeta^2)^{1/2}}$$

for the flat cone setting; and the calculation of the absorption correction appropriate to the incident and diffracted rays (North, Phillips and Matthews, 1968).

The statistical tests on the measurements were identical to those already described, the same parameters being chosen so that reflections whose structure amplitude was less than its standard derivation, or was negative, or for which the difference between the two background measurements exceeded four times its standard deviation were identified on a monitor record. The negative structure amplitudes

TABLE 9 Primary Processing of 2.5Å Resolution Intensities. Each set of measurements involved eight crystals for which are tabulated the crystal reference number, the date of the intensity measurements (Summer 1969 or Spring 1971), the maximum value of the absorption correction A_I, the relative scale factor k_F determined subsequently for scaling between crystals and derivative to native measurements, the statistics of the primary processing (number of measurements N, percentages of these less than their standard deviation SD, negative N-, or with significantly different backgrounds BG), and the percentage loss of intensity RD on average during the intensity measurements of the three reference reflections.

		Crystal	Date	A_I	k_F	Primary Processing N	SD%	N-%	BG%	RD%
NATIVE	C1	CT15	69	1.85	1.5	8372	23	11	1	-6
	C2	CT16	69	1.86	1.4	8167	21	11	3	-4
	C3	CT23	69	1.15	1.5	7989	24	17	11	-12
	C4	CT20	69	1.42	1.1	5675	23	11	5	-24
	B1	CT13	69	1.21	0.9	6583	12	9	8	-13
	B2	CT18	69	1.16	2.2	6488	31	15	1	-2
	B3	CT33	71	1.60	1.7	8024	29	16	3	-2
	B4	CT21	69	1.17	0.7	5442	9	5	3	-21
MONO	C1	M6	69	1.16	1.8	8441	19	9	1	-7
	C2	M8	69	1.32	1.2	7795	12	12	7	-21
	C3	M9	69	1.38	1.3	6792	16	9	8	-20
	C4	M20	71	1.32	2.2	6310	31	15	3	+1
	B1	M5	69	1.50	1.8	7044	19	9	2	-6
	B2	M10	69	1.60	1.4	6456	16	12	7	-12
	B3	M14	69	1.48	1.5	5969	17	9	3	-3
	B4	M19	71	1.38	2.5	6386	30	16	3	-3
EMP	E. C1	E10	71	1.47	1.9	9259	20	11	3	-2
	C2	E14	71	1.42	2.0	8672	21	14	6	-3
	C3	E16	71	1.32	2.4	7624	23	12	3	-8
	C4	E17	71	1.63	2.5	5142	31	20	4	-8
	B1	E3	71	1.18	2.3	7713	22	13	2	-5
	B2	E2	71	1.37	2.1	8457	23	12	4	-10
	B3	E5	71	1.75	2.7	6848	29	13	1	-6
	B4	E11	71	1.60	2.3	5953	25	13	4	-4
BAKER	C1	BK3	71	1.25	1.5	9587	19	13	1	-3
	C2	BK7	71	1.70	1.5	10298	21	11	5	-3
	C3	BK8	71	1.69	1.4	8961	22	12	3	+0
	C4	BK14	71	1.33	1.9	4945	29	14	2	-4
	B1	BK10	71	2.10	1.4	8834	21	12	2	-6
	B2	BK11	71	2.00	1.3	9540	19	12	4	-5
	B3	BK12	71	1.70	1.8	7665	25	13	2	-10
	B4	BK13	71	1.70	2.1	8024	29	15	2	-6

were set equal to zero, with a standard deviation of unity as an indicator that a valid measurement had been made. This indicator facility became available only for the processing of the more recent sets of measurements, for which the structure amplitudes and their standard deviations were recorded in five successive data files on magnetic tape, with about five hundred reflections in a file, this being the capacity of that version of the program. The earlier measurements were recorded on paper tape output, but they were subsequently transferred directly to magnetic tape data files, so that all the sets of measurements were analysed in an identical fashion for all processing beyond this primary stage.

A summary of the monitor output, combining the several passes through the program which were required for processing an entire set of five levels, is presented in Table 9. The maximum absorption correction was very rarely greater than 2.0, most of the curves being much more uniform than this. The scale factor which was subsequently applied to the structure amplitudes from each crystal is included here for comparison with the statistical analysis. The correlation between the larger values of this scale factor, and the percentage of the structure amplitudes in a set of levels which are less than their standard deviations, is quite good as would be expected.

The percentage of reflections in this category was used as one test of the acceptability of a set of intensities. Measurements from several crystals were discontinued and rejected if this analysis of the measurements made during the first overnight run showed that too many of the reflections were so weak that they could not be accurately measured. However, the level of acceptability varied between derivatives. The EMP crystals clearly diffracted less strongly than

the other crystals, as was also seen from the low resolution data, and the more recent MONO crystals were much smaller, having correspondingly weaker intensities, than the earlier ones.

The percentage of reflections for which the two estimates of the background were significantly different is an indication of the correctness of the orientation of a crystal throughout the intensity measurements. Many of these invalid measurements were repeated manually for the sake of completeness. However, such manual intervention frequently resulted in errors and incorrect indices being used for these measurements, which then had to be eliminated after being detected during the next stage of processing. Indexing errors which arose during the automatic measurement runs because of the failure of components in the control system were usually detected prior to the primary processing and could be corrected by using a modified input routine within the program so that such measurements need not be rejected. The only sets of levels which were seriously affected in this way were B2 and B3 of the EMP derivative.

The estimated standard derivatives of the structure amplitudes were subsequently modified, according to the formula

$$\sigma^2 = c^2 + e^2 . F^2 .$$

to include the contributions of both counting statistics (c) and other errors which would be expected to be some fraction (e) of the structure amplitude. For the native data this fraction was chosen as 1%. The derivative data were not processed until after a statistical analysis of the compiled native data set had been made, indicating that a more realistic value would have been 2%. This value was then used in determining the estimated standard deviations for the derivative sets

a)

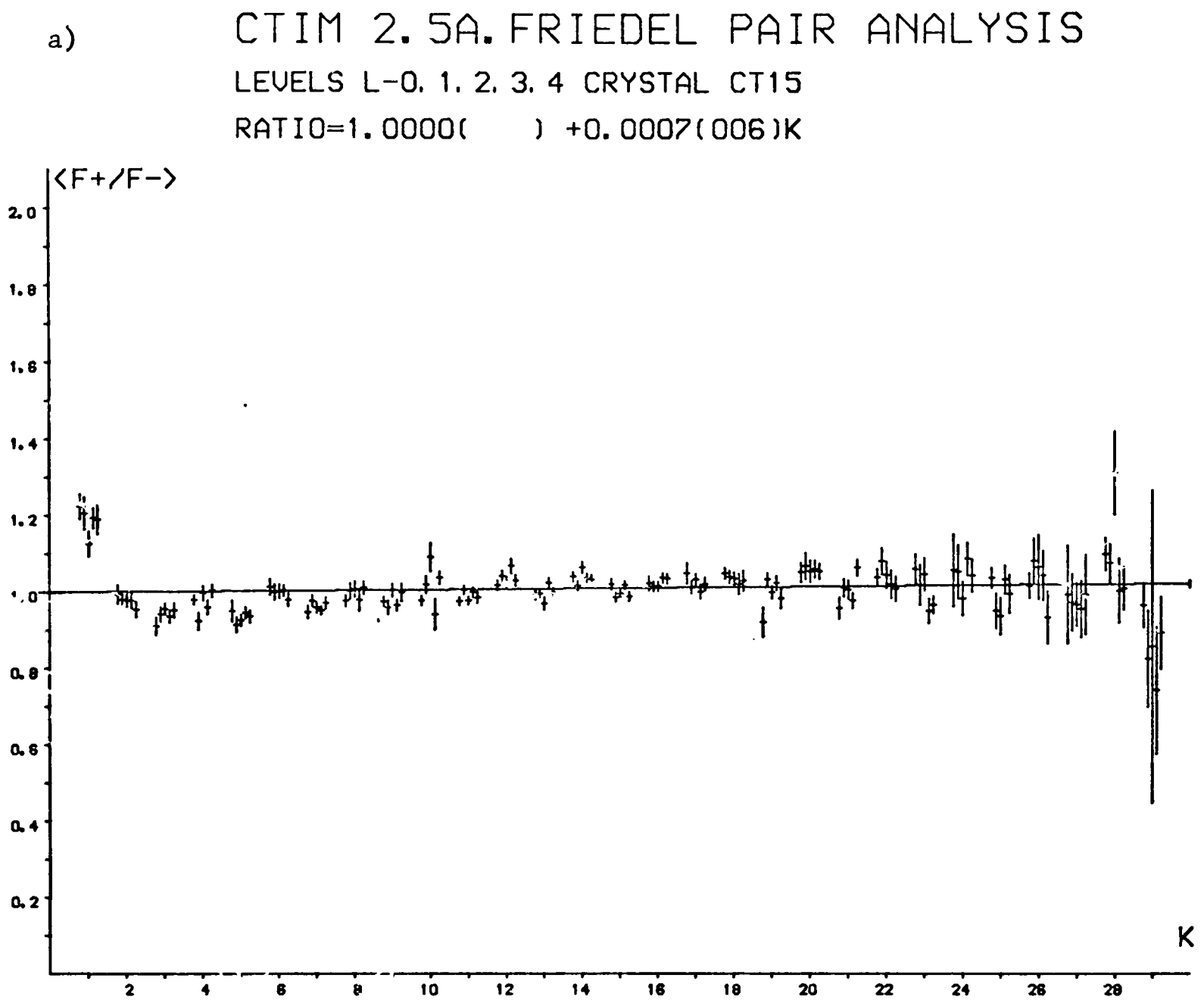
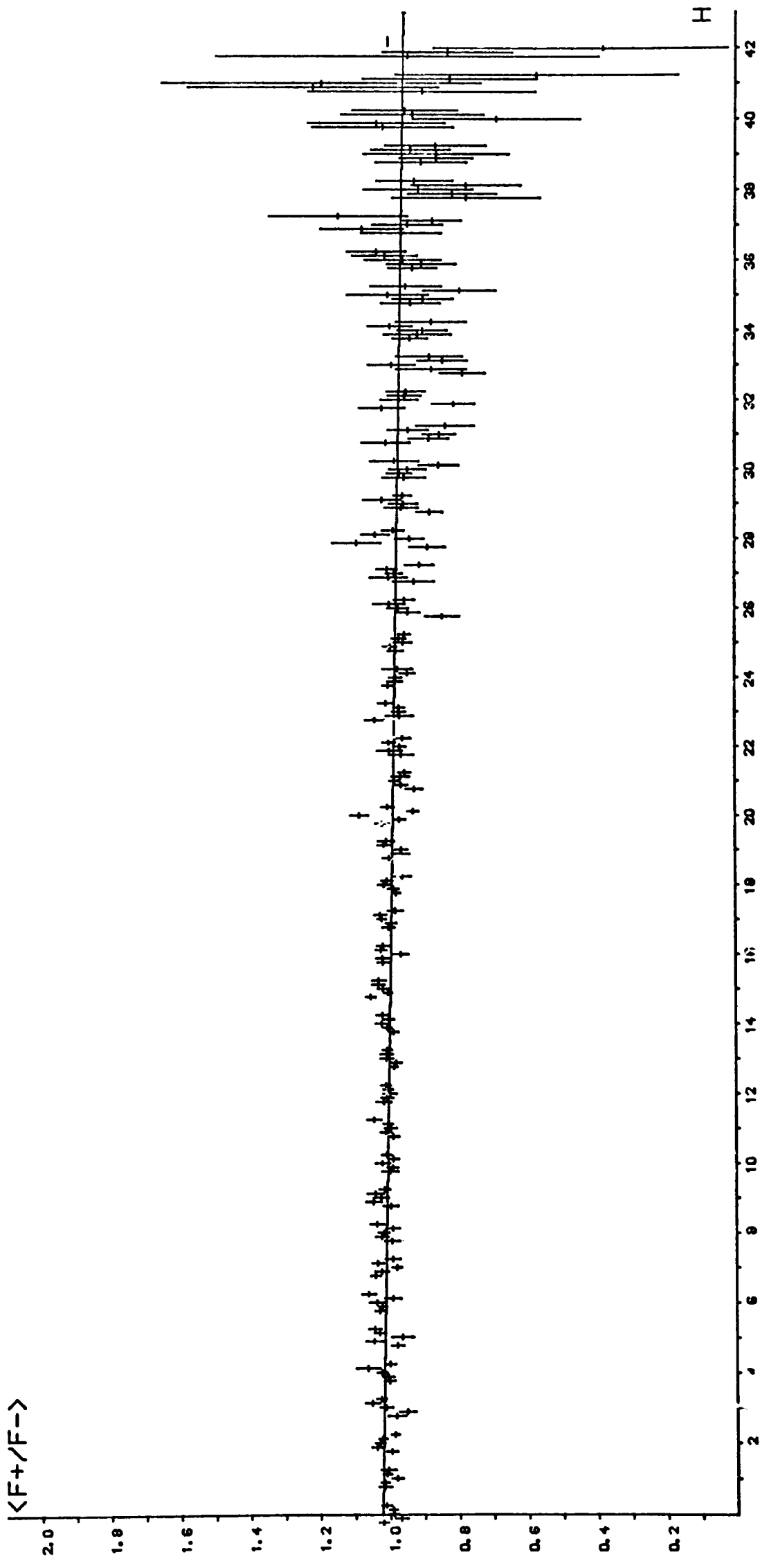


FIGURE 26 Ratio of Friedel Equivalents for NATIVE Cl Levels of Reflections. The levels $l=0 \rightarrow 4$ are shown from left to right; Analysis against (a) the stepping index k and (b) the scanning index h shows small positive and negative slopes respectively; the rows k=0,-1 were found to have been mis-set and so incorrectly indexed, and these reflections were completely eliminated from the file of measurements on editing; the reflections (20±1 2) measured on counter 3 was very badly affected by this error as is seen in (b).

FIGURE 26b (legend on previous page)



a)

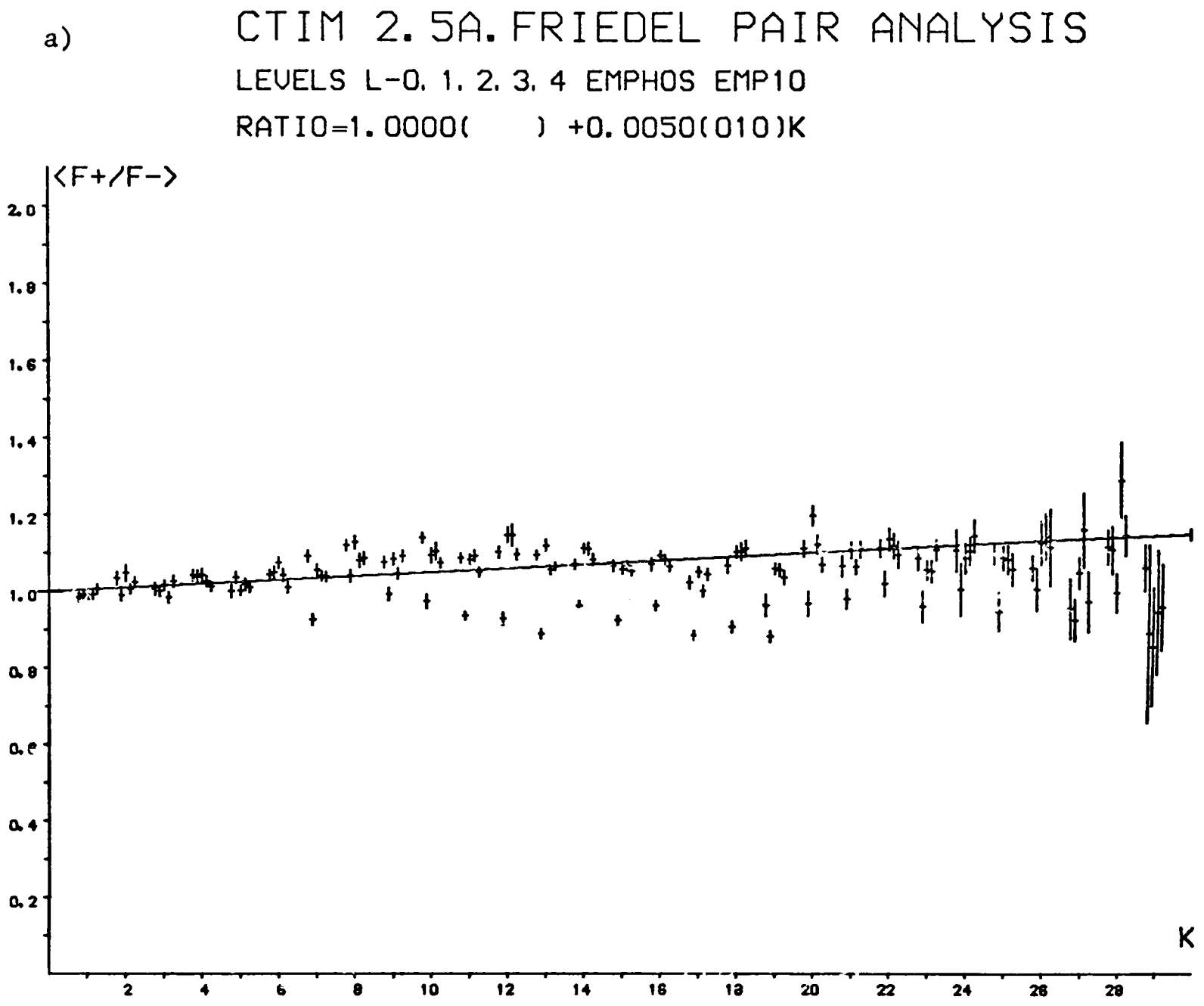
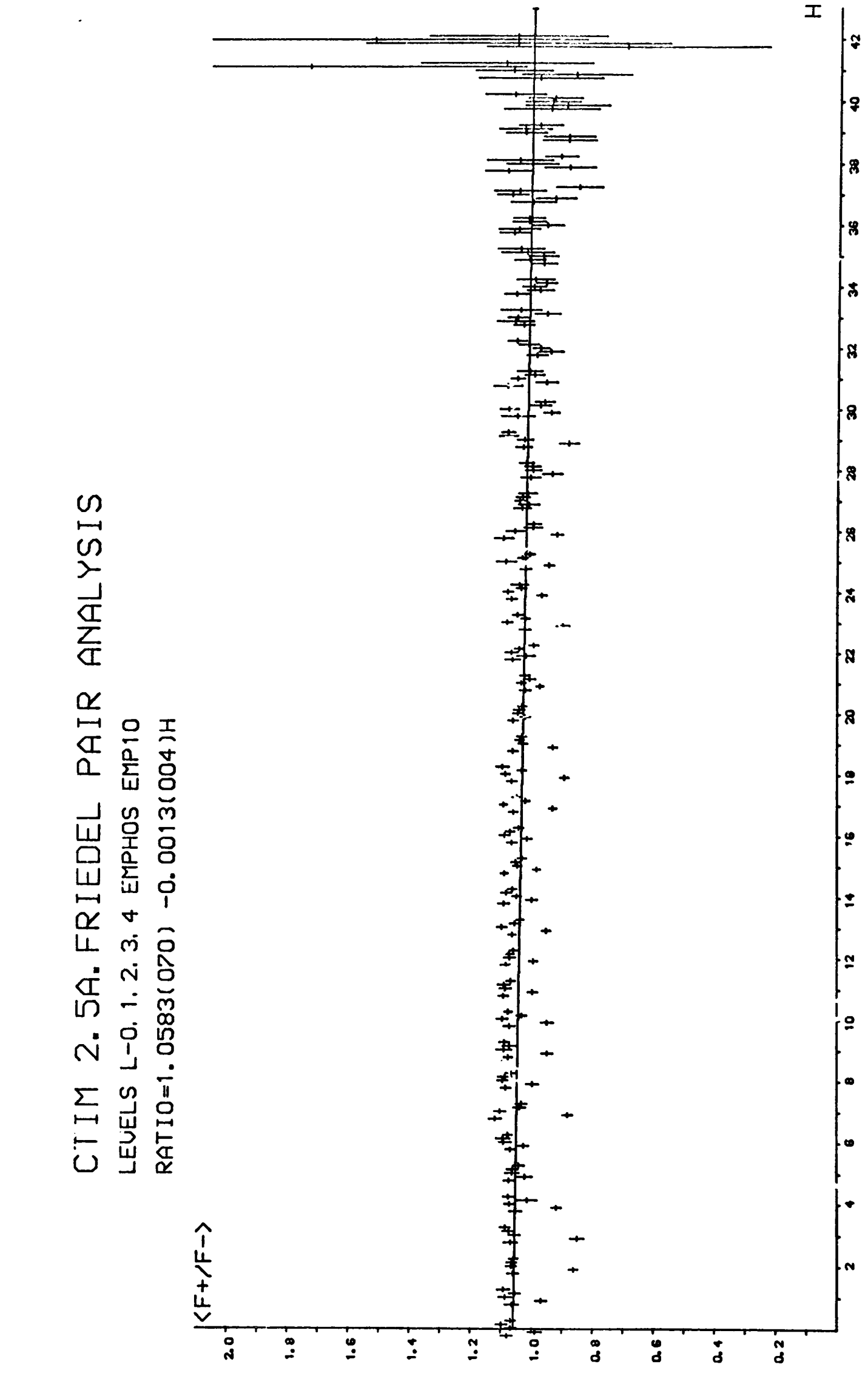


FIGURE 27 Ratio of Friedel Equivalents for EMP Cl Levels of Reflections. The levels $\ell=0 \rightarrow 4$ are shown from left to right; Analysis against (a) the stepping index k and (b) the scanning index h shows a fairly large positive and negative slope respectively; the counter for the level $\ell=1$ failed for $k \leqslant -7$; these reflections were removed from the file of measurements on editing.

FIGURE 27b (legend on previous page)



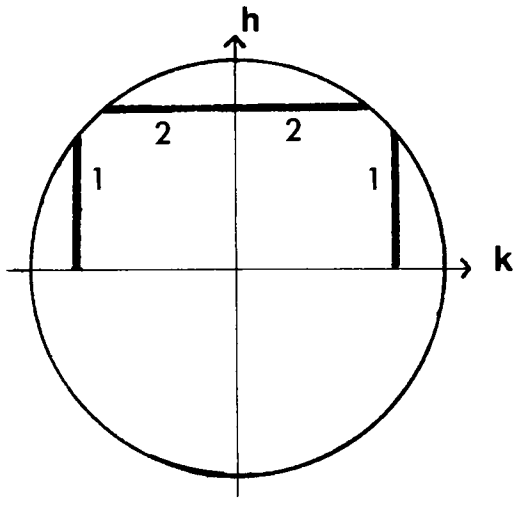
of data.

C) SECONDARY PROCESSING OF MEASUREMENTS FROM INDIVIDUAL CRYSTALS

The agreement between the Bijvoet pairs of reflections was analysed as a function of both the stepping and scanning indices within each level. For each row of reflections having either of these indices constant, the average values of the observed structure amplitudes and quantities derived from them were tabulated. The most informative quantity was the weighted mean value of the ratio (F_+/F_-), graphs of which were plotted for the five separate levels measured from each crystal. The lines of linear regression of these ratios against the appropriate index were calculated for each level separately, and for the five levels combined, the latter line being represented on the graphs. Typical pairs of graphs, one against each index, are shown in Figures 26 and 27.

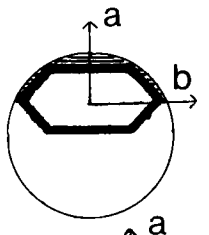
By inspection of these graphs several instances of instrumental failure were detected and the affected rows of measurements were then eliminated from the data files. The entire level $l=1$ of the C1 data for the BAKER derivative was removed, the proportional counter appearing to have been totally unreliable during this run. Other errors resulted in data for several individual reflections being removed, the causes of these including counters recording spuriously low intensities, indexing errors for reflections measured manually, and the failure to detect significantly unequal background intensities for very strong reflections, which resulted from an undetected computing overflow in the primary processing program as used for the earlier sets of measurements.

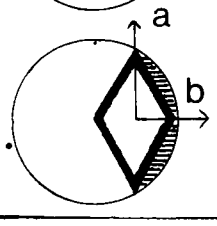
a)



b) Rows of reflections

of type :	1	2	
Ratio :	$\sum_h I(k) : \sum_h I(\bar{k})$	$\sum_k I(h) : \sum_k I(\bar{h})$	
	Ratio = 1 for all h.	Ratio =1 for all k.	A(hkℓ) =A(h̄kℓ) ≠A(hkℓ)
	Ratio > 1 → 1 as h → ∞.	Ratio > 1 → 1 as k → 0.	A(hkℓ) =A(hkℓ) ≠A(hkℓ)





c)

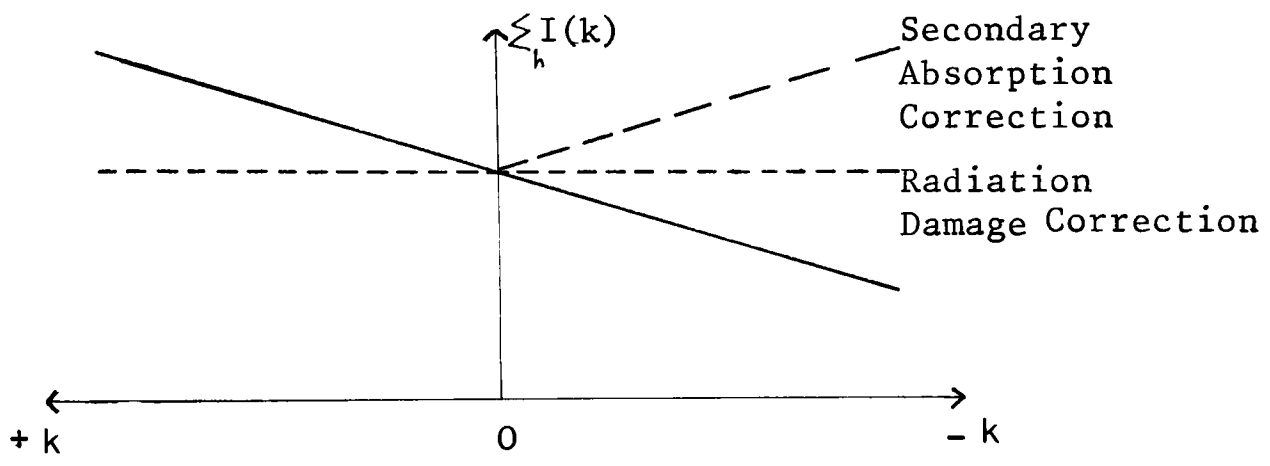


FIGURE 28 Secondary Absorption and Radiation Damage Corrections. The analysis of rows of reflections with respect to both indices within a level is shown in (a) for the octants hkℓ,hkℓ; the values expected for the ratios for two idealised cases of asymmetrically mounted crystals is shown in (b); the effects of corrections appropriate for radiation damage and secondary absorption are shown in (c) for a hypothetical plot of average measured intensities.

Errors affecting individual measurements were detected either from isolated points on the graphs or from the listing of all reflections for which the difference between two Friedel related reflections exceeded four times its standard deviation. Careful inspection of these lists together with the graphs enabled the extent of any necessary editing to be determined. This was then achieved very efficiently with visual display on the CRT of the computer of all the affected reflections.

The ratio of the mean value of the difference between the equivalents to the mean structure amplitude was calculated before and after editing and correcting for the slopes of the graphs, but the standard deviations of these quantities are so large that the variations between them are not very meaningful as can be seen from Tabel 10.

There are two possible methods of correcting for the slopes of these graphs, the choice depending on whether they are thought to arise from the effects of secondary absorption (North, Phillips and Matthews, 1968) or radiation damage. Secondary absorption may produce no effect at all as shown in Figure 28 or a slope such that the line of data points intersects the ordinate axis of the graphs at a ratio of exactly 1.0 when plotted against the stepping axis and the opposite slope when plotted against the scanning index (Figure 27), but tending to the value 1.0 for large values of that index. Radiation damage would produce effects of this second type, but the magnitude and direction of the slope in this case would be expected to be correlated with the loss of intensity observed in the reference reflections, which is summarised in Table 9.

The correction appropriate for secondary absorption is that

TABLE 10 Agreement Between Friedel Related Reflections.
For each set of levels are listed the number of reflections for which the difference between Friedel equivalent amplitudes exceeds 4σ before (N_1) and after (N_2) applying a secondary correction and editing, the corresponding values R_1 and R_2 of the ratio $\langle|\Delta F|\rangle/\langle\overline{F}\rangle$, and the standard deviation of the former, and the slopes observed and calculated on the basis of radiation damage losses of the weighted mean ratio $\langle w.F_+/F_-\rangle$ as a function of the stepping index.

		N_1	N_2	R_1%	R_2%	$\sigma(R_1)$	$10^4.S$	$10^4.S$	$10^4.\sigma$	
NATIVE	C1	17	2	14	11	(22)	10	7	(6)	
	C2	35	5	14	13	(20)	7	37	(5)	
	C3	46	9	13	13	(18)	23	48	(7)	
	C4	36	0	18	15	(21)	54	83	(11)	
	B1	96	38	10	10	(11)	26	41	(8)	
	B2	2	0	14	14	(28)	4	-15	(10)	
	B3	3	0	13	13	(22)	4	20	(7)	
	B4	191	9	12	7	(10)	57	91	(8)	
MONO	C1	10	7	12	12	(17)	12	21	(5)	
	C2	110	12	15	12	(13)	43	67	(6)	
	C3	56	11	14	12	(14)	40	48	(10)	
	C4	11	4	16	18	(25)	-2	-32*	(22)	(-40)
	B1	2	2	12	11	(18)	12	10	(5)	
	B2	23	3	15	12	(15)	25	61	(7)	
	B3	23	4	13	12	(16)	6	0*	(17)	(-11)
	B4	3	1	15	15	(24)	8	-59	(10)	
EMP	C1	45	8	12	11	(16)	3	32*	(8)	(50)
	C2	37	18	13	12	(16)	5	17*	(21)	(25)
	C3	12	8	13	13	(18)	15	11	(6)	
	C4	43	2	25	17	(23)	17	129*	(13)	(121)
	B1	47	5	12	11	(10)	10	42*	(10)	(32)
	B2	96	1	13	11	(16)	21	64*	(12)	(51)
	B3	63	2	18	15	(23)	15	14	(18)	
	B4	6	2	12	12	(18)	10	24	(7)	
BAKER	C1	3†	6	12†	10	(15)	5	-12*	(7)	(-8)
	C2	43	39	12	12	(18)	5	23*	(9)	(6)
	C3	2	2	12	11	(18)	0	4	(6)	
	C4	0	1	16	14	(26)	8	-17	(8)	
	B1	3	2	10	10	(15)	12	-24≠*	(5)	(-20)
	B2	5	1	10	9	(17)	10	6	(5)	
	B3	5	0	12	11	(17)	22	19	(6)	
	B4	1	0	14	14	(23)	12	-21	(10)	

* after editing the slope was re-determined as shown in parentheses
† excluding level $\mathbf{l}$=1 BAKER which was completely removed on editing
≠ this set of levels B1 BAKER was measured in the reverse order

making the more highly absorbed octant of reflections equal to the others. The correction for radiation damage involves both reducing the stronger reflections and also increasing the weaker ones, by a factor related to half of the slope of the graphs against the stepping index.

The parameters of the linear regression determined for these graphs against the stepping index, with an origin constraint are listed in Table 10 and against both indices without an origin constraint in Table 11. They provide no conclusive evidence as to the cause of these slopes. The correction applied in every case was that appropriate to the secondary absorption effect, a linear correction with respect to the stepping index being applied to all the structure amplitudes and standard deviations. This correction was applied prior to the editing and the calculation of new discrepancy ratios listed in Table 10.

All of the analysis described in this section, and subsequent processing and refinement up to the stage of calculating an electron density map has been repeated following the discovery by S.H. Banyard of an error in the application of the absorption correction in the primary processing program. This had produced systematic curvatures in the graphs analogous to those of Figure 27 which became increasingly obvious for the upper sets of levels measured, but had been ascribed to general inadequacies in the semi-empirical method of absorption correction used (North, Phillips and Matthews, 1968). However in his analysis of measurements from human lysozyme, Banyard encountered much larger effects than had been met in this work, leading to the discovery of the true cause of these features.

TABLE 11 Variation of the Ratio of Equivalent Reflections.
The linear regression parameters for each set of levels are listed for an analysis of the ratio $R = \langle w.F_+/F_- \rangle$ according to the expression
$R = A.i + B$
for the scanning index h and the stepping index which is k or ℓ for the c* and b* sets of levels respectively.

		Stepping index				Scanning index			
		A	σ	B	σ	A	σ	B	σ
NATIVE	C1	.0028	5	.969	7	-.0016	2	1.023	3
	C2	.0036	5	1.002	6	-.0041	3	1.098	4
	C3	.0030	7	1.020	7	-.0008	3	1.056	4
	C4	.0089	11	.993	12	-.0058	7	1.158	10
	B1	.0053	8	.986	8	-.0028	3	1.074	5
	B2	.0003	10	.980	10	-.0002	5	.985	8
	B3	.0041	6	.979	6	-.0021	3	1.036	4
	B4	.0093	8	.998	7	-.0037	3	1.122	4
MONO	C1	.0027	5	.991	6	-.0011	2	1.037	4
	C2	.0081	6	.981	7	-.0017	4	1.087	6
	C3	.0050	10	.998	13	-.0002	6	1.053	9
	C4	-.0034	22	1.002	24	+.0030	10	.936	13
	B1	.0028	5	.979	5	-.0003	2	1.009	4
	B2	.0087	6	.970	6	-.0041	3	1.110	5
	B3	.0057	16	.936	16	-.0067	7	1.078	11
	B4	-.0016	8	.957	7	-.0019	4	.967	6
EMP	C1	.0018	8	1.020	10	-.0013	4	1.058	7
	C2	.0080	20	.912	24	-.0044	10	1.062	18
	C3	.0019	6	.989	6	-.0020	3	1.035	5
	C4	.0089	12	1.051	14	-.0061	8	1.219	12
	B1	.0056	10	.984	10	-.0016	4	1.054	7
	B2	.0046	12	1.021	12	-.0011	5	1.075	8
	B3	-.0008	17	1.024	17	+.0018	6	.996	9
	B4	.0049	6	.975	5	-.0026	4	1.046	5
BAKER	C1	.0047	7	.988	8	-.0000	3	.984	5
	C2	.0018	9	1.007	11	-.0011	5	1.041	8
	C3	-.0016	6	1.025	7	-.0000	3	1.009	5
	C4	.0009	8	.970	8	-.0024	6	1.010	8
	B1	-.0040	5	1.020	5	+.0023	2	.951	4
	B2	.0038	4	.963	4	-.0018	3	1.023	3
	B3	.0048	5	1.015	4	-.0005	3	1.015	4
	B4	.0036	7	1.006	8	-.0033	4	1.006	6

A subsidiary program was used to correct the error, without having to repeat the actual primary processing with all the handling of punched paper tapes which that involves. The experience which had been gained with the further processing of the uncorrected sets of measurements, was used to improve certain aspects of the data compilation, especially the inter-level scaling and the refinement of derivative scale factors, as will be described in later sections.

D) COMBINATION OF MEASUREMENTS FROM DIFFERENT CRYSTALS

The levels of structure factors which were measured from different crystals or using different counters were scaled together using the method of Hamilton, Rollett and Sparks (1965) in which an iterative procedure is used to minimise the weighted sum for all observations and all reflections of the residual

$$(F_k^2 - G_k^2 . \bar{F})^2$$

where F_k is the estimate from level k, for which the scale factor is the reciprocal of G_k, of that structure factor for which the best estimate from all observations is $\bar{F}$. The factors G_k are refined in successive cycles of the HRS procedure. For each set of chicken TIM measurements at 2.5Å the twenty levels measured about each of two axes required the determination of forty scale factors and there were approximately 10,000 reflections common to both the b* and c* sets of measurements. Consideration of these reflections in groups with constant values of k and ℓ and summation along h to obtain average values for the rows of reflections common to both sets of levels, reduces the number of observations to 400, a number which

can be much more easily managed. All the level scale factors were determined using such common rows of reflections, which contained an average of about 25 reflections, for many of which there were measurements of both members of a Bijvoet pair.

Averaging along the common rows raises the problem of whether the average should be that for the structure factors or the intensities, and also whether or not equal weighting of the reflections is the most appropriate procedure. The directly observed quantities were the intensities, but the structure factors are the quantities for which the most accurate values are required for the subsequent analysis. Some type of statistical weighting appears to be necessary, since earlier consideration of the agreement between equivalent measurements within a level had shown the likelihood of there being reflections for which the corresponding measurements in the b* and c* sets of levels could be

$$F_c = 100(10) \qquad F_b = 10(100)$$

when scaled relative to one another. The corresponding intensities would be

$$I_c = 10{,}000(2{,}000) \quad I_b = 100(2{,}000)$$

so that the presence of one such reflection in a common row of 50 measurements could affect the relative average intensities by 2%. The scale factor for any single level is determined by the intersections with all of the twenty levels about the other axis, so that this single reflection would bias the scale factor for that level of intensities by 0.1%. This is clearly unacceptable since many such poorly measured intensities are present in all the sets of data.

The simple expedient of excluding from the common row averages all those measurements which were less than their own standard deviation applies a very sharp, but arbitrary truncation. A weighting scheme defined according to the relationship

$$w = \frac{1}{\sigma_c^2} + \frac{1}{\sigma_b^2}$$

was used in the determination of the common row averages, the standard deviations being those for the structure amplitudes or intensities as appropriate. This scheme does not have any obvious disadvantages, but as a precautionary investigation, the set of forty scale factors for the native measurements were determined using four different sets of common row averages. These were for weighted and unweighted sums of both structure factors and intensities, the corresponding sets of scale factors being listed in Table 12, after appropriate normalisation.

The residuals after the refinement of the HRS scale factors were lower for the unweighted sets of common row averages than for the weighted ones, but tests of the agreement between the b* and c* measurements after scaling were unable to distinguish between them. The use of the weighted intensity averages was thought to be the most desirable method, and all further consideration of HRS level scale factors was in terms of such common row observations.

Whenever both reflections of a Bijvoet pair were common to two sets of levels they were included in the common row averages as two independent observations. Only one observation was included if there were fewer than four estimates of the reflection, and whenever the only two measurements were F_+ in one set of levels and F_- in the other, then the contribution to the common row totals was weighted

TABLE 12 Level Scale Factors.
The forty scale factors appropriate to the levels of structure amplitudes are listed for four different sets of common row averages calculated for the NATIVE data as described in the text. For the derivatives only the scale factors determined from the common row averages of weighted intensities are listed. For each set of data the number of reflections N for which the discrepancy $|\overline{F}_c - \overline{F}_b|$ exceeded 4 times its standard deviation is also listed

		NATIVE				MONO	EMP	BAKER				
		F	wF	I	wI	wI	wI	wI	N_N	N_M	N_E	N_B
C1	ℓ=0	1.558	1.568	1.579	1.582	1.143	0.807	1.082	2	7	5	20
	1	1.568	1.548	1.565	1.556	1.128	0.839	0.0	1	6	3	
	2	1.573	1.548	1.568	1.557	1.178	0.763	0.975	0	9	7	15
	3	1.504	1.511	1.521	1.512	1.184	0.803	1.094	1	7	3	11
	4	1.529	1.520	1.537	1.522	1.173	0.764	1.032	0	2	10	17
C2	5	1.381	1.380	1.389	1.382	0.747	0.860	1.196	1	6	13	18*
	6	1.410	1.416	1.413	1.412	0.756	0.945	1.051	1	4	18	13
	7	1.442	1.427	1.444	1.428	0.791	0.855	1.034	4	5	13	12
	8	1.384	1.409	1.407	1.404	0.825	0.951	1.200	3	8	15	12
	9	1.398	1.427	1.417	1.422	0.738	0.888	1.120	4	24	9	11
C3	10	1.511	1.500	1.497	1.499	0.762	0.983	0.923	5	9	10	12
	11	1.581	1.568	1.569	1.567	0.799	1.093	0.956	2	5	30	25
	12	1.579	1.593	1.574	1.581	0.832	1.000	0.936	6	3	12	10
	13	1.559	1.548	1.543	1.538	0.825	1.096	1.079	3	7	7	13
	14	1.550	1.546	1.541	1.531	0.805	1.037	1.045	2	5	8	12
C4	15	1.048	1.055	1.053	1.054	1.376	1.091	1.331	3	5	19	9
	16	1.081	1.087	1.083	1.090	1.371	1.160	1.346	2	5	8	3
	17	1.109	1.102	1.096	1.104	1.375	1.030	1.352	1	4	13	4
	18	1.094	1.101	1.097	1.103	1.357	1.088	1.359	0	0	9	2
	19	1.070	1.083	1.077	1.077	1.340	1.055	1.281	1	1	1	1
B1	k=0	0.948	0.885	0.903	0.902	1.110	1.155	0.955	2	4	2	30*
	1	0.926	0.899	0.913	0.906	1.109	0.999	0.986	4	0	6	31*
	2	0.966	0.944	0.953	0.951	1.147	0.994	1.007	10	1	7	26*
	3	0.940	0.920	0.934	0.928	1.114	0.974	0.967	7	2	9	34*
	4	0.914	0.915	0.916	0.909	1.116	1.056	0.936	3	2	11	35*
B2	5	2.114	2.132	2.133	2.104	0.860	0.883	0.816	0	4	8	5
	6	2.110	2.175	2.160	2.153	0.890	0.953	0.876	2	6	11	4
	7	2.207	2.248	2.248	2.241	0.908	0.914	0.875	1	8	16	2
	8	2.083	2.138	2.142	2.123	0.905	0.893	0.843	1	8	4	1
	9	2.059	2.090	2.087	2.080	0.871	0.894	0.848	2	14	10	0
B3	10	1.756	1.740	1.754	1.745	0.911	1.112	1.276	42*	7	12	12
	11	1.709	1.723	1.729	1.713	0.944	1.155	1.254	35*	5	18	9
	12	1.716	1.734	1.734	1.719	0.961	1.162	1.248	47*	7	9	2
	13	1.720	1.736	1.734	1.739	0.938	1.178	1.244	42*	7	6	2
	14	1.706	1.720	1.716	1.715	0.967	1.168	1.222	46*	5	10	2
B4	15	0.715	0.710	0.709	0.711	1.595	0.995	1.416	3	13	20	5
	16	0.752	0.739	0.739	0.740	1.651	0.995	1.449	1	3	23	4
	17	0.775	0.765	0.762	0.763	1.581	1.031	1.458	3	10	22	2
	18	0.742	0.738	0.736	0.738	1.635	1.014	1.443	2	7	8	1
	19	0.749	0.749	0.743	0.746	1.594	0.990	1.443	1	10	8	2

* all reflections in these levels eliminated from data sets on editing.

down in the case of all the heavy atom derivatives by 0.5. The sets of HRS scale factors calculated and applied to the native and derivative measurements are shown in Table 12.

The agreement between the measurements from different crystals was analysed in a manner similar to that for agreement of Friedel equivalents within a single level. In the present case, the disagreement is considered in terms of the discrepancy

$$\Delta F_{cb} = \overline{F}_c - \overline{F}_b$$

$$\overline{F}_c = \langle F_{c+} , F_{c-} \rangle$$

$$\overline{F}_+ = \mathrm{RMS}(F_{c+} , F_{b+})$$

$$\overline{F} = \langle \overline{F}_+ , \overline{F}_- \rangle$$

and the mean values were determined using the appropriate least squares weights. Averages involving the discrepancy were calculated for each common row of reflections, individual reflections being monitored whenever the discrepancy exceeded four times its standard deviation. The number of such reflections, and their distribution amongst the twenty levels of each index is shown in Table 12 for all the sets of data.

It was immediately clear that the B3 levels of the NATIVE set were inconsistent with the other levels. The B3 measurements had been recollected at the same time as the derivative data since the original B3 NATIVE measurements had had very large standard deviations. This had necessitated the use of a crystal from a different preparation of the protein, from that from which all the other crystals used for the 2.5Å NATIVE intensity measurements had

been obtained. No significant differences were detected between the new and old sets of B3 NATIVE data, but the significance of all the appropriate tests was necessarily low. The discrepancy between the new B3 measurements and all the other sets of levels was ascribed to innate differences in the structure of different batches of crystals. In order to obtain a consistent set of NATIVE data, all the measurements from the B3 levels having the index ℓ less than 20 were excluded from the set of data. Further consideration was given to these measurements only to the extent necessary for the determination of the rescale factors for the few remaining reflections in this set of levels, for which there was no overlap with the c* sets of levels. The differences between the two sets of B3 measurements were not large enough for these few reflections ($20 \leqslant \ell \leqslant 24$) to be entirely rejected.

Rejection of most of the measurements from this crystal was felt to be justified in the interests of the internal consistency of the NATIVE data set, but since this crystal was from the same batch as all those used for the 2.5Å intensity measurements for the derivatives, it is unfortunate that the possibility of there being different types of crystal can not be excluded.

The B1 levels of the BAKER derivative were also clearly inconsistent with the c* sets of levels, and so they were completely excluded from the data set for this derivative. No reasonable explanation can be proposed for this poor agreement. The level ℓ=5 of the C2 BAKER measurements was also rejected at this stage, because the internal agreement of the Bijvoet pairs in this level was poor, and there was no indication from the overlapping b* measurements of which regions, if any, of the ℓ=5 level contained valid measurements. The cause of these inadequate measurements was an unreliable proportional

counter on the diffractometer during this run.

The sets of measurements for the MONO and EMP derivatives were accepted without any levels having to be rejected.

All four sets of data were then edited with respect to individual reflections for which the discrepancy between equivalents from the same or different crystals, or the deviation of any individual measurement from a derived weighted mean, exceeded specified absolute values or multiples of the standard deviation. The amplitudes corresponding to the observed intensities, and all their means, differences and standard deviations were displayed on the CRT of the computer whenever the data for a reflection failed to pass any of these statistical tests. Direct input from a keyboard allowed modification of the data whenever necessary. One or more of the amplitudes were set to zero, depending on the consistency of the remaining measurements, whenever the disagreement for a reflection was serious.

The best values for the mean structure amplitude and the anomalous difference between Friedel equivalents were also determined in the course of these editing checks. For those reflections which had four measurements, these quantities were defined as

$$\overline{F} = \langle \overline{F}_+ , \overline{F}_- \rangle$$

$$\Delta F = \overline{F}_+ - \overline{F}_-$$

using the mean values of the positive and negative equivalents already defined. For reflections with only three measurements, the one resulting from a different crystal was excluded from the calculation of the anomalous difference and it was also excluded from the calculation of the mean amplitude if it differed from the mean of the other

facing page 124

TABLE 13 Rescaling Levels of Measurements.
The HRS residual, the number of common reflections N, the average $\overline{F}$, four parameters involving $|\Delta F| = |\overline{F}_c - \overline{F}_b|$ and its σ indicating the overall agreement between the c* and b* sets of measurements and the number of reflections having $|\Delta F| > 4\sigma$ are listed. The two columns for each set set of data show the values after initial HRS scaling and after editing and HRS rescaling. Any levels of data rejected are listed, as are any HRS rescale factors differing from unity by more than 1%. Block averages of a discrepancy index are tabulated, with the final rescale factors which were derived from them as described in the text and the resultant measures of agreement.

	NATIVE		MONO		EMP		BAKER	
Residual	2370	1014	1400	1255	2968	1977	2638	1358
N(Refs.)	10023	7073	9363	9304	9772	9684	9680	7018
$\overline{F}$	145.6	146.1	117.	116.1	100.	99.31	124.1	124.1
$\langle\|\Delta F\|\rangle/\langle\overline{F}\rangle$ %	14.	13.9	14.	13.4	15.	14.5	13.6	12.4
$\langle \Delta F \rangle/\langle\overline{F}\rangle$ %	0.05	0.37		-0.47		-0.04		-0.02
$\langle\|\Delta F\|/\sigma\rangle$	1.0	0.92	1.0	0.99	1.1	1.07	1.1	0.99
$\langle \Delta F /\sigma\rangle$	0.96	0.05		-0.04		-0.01		0.02
N4σ	42 (+212*)	13	122	60	220	151	220	47
Data Rejected	*B3 1 20 1=0, k 1						B1 1=5	
Rescale diffs.	1=0,0.969;1=14,1.011 1=1,0.983;1=15,0.986 1=3,0.985;1=16,0.985 1=7,1.010						1=0,1.017 1=17,1.017	

$\langle\Delta F\rangle/\langle\overline{F}\rangle$ averaged over blocks of 5 counters (B1 → B4) after rescaling; %

C1	+4·8	−1·5	0	+6·6	−0·4	+1·8	−1·7	−2·6	+0·6	+0·5	−1·6	−0·0	0	+1·0	−1·1	+2·0
C2	+3·5	−2·4	0	+3·3	−1·6	+0·4	+0·8	−2·8	+0·3	+1·1	−1·4	+0·5	0	−0·3	−0·6	+1·6
C3	+1·0	−5·8	0	+2·8	−1·8	+2·0	+0·9	−4·9	−2·0	+0·5	+0·2	+0·4	0	−0·4	+0·2	−4·0
C4	+3·3	−10·2	0	+5·2	+1·0	+2·7	+3·4	+2·2	−2·7	−1·6	−3·1	+3·0	0	+4·0	+1·1	−3·3

Final Rescale Factors

C1	B1	0.971	1.028	1.007	0.993	1.001	0.994	0.993	0
C2	B2	0.990	0.947	1.007	1.018	0.998	1.004	0.998	1.011
C3	B3	1.012	(1.017)*	1.010	1.008	1.002	0.988	1.014	0.999
C4	B4	1.011	1.041	0.976	0.978	1.011	1.013	0.994	0.991

Agreement after final rescaling

$\langle\|\Delta F\|\rangle/\langle\overline{F}\rangle$ %	14.3	13.5	14.5	12.4
$\langle \Delta F \rangle/\langle\overline{F}\rangle$ %	+0.01	-0.17	+0.26	-0.01
$\langle\|\Delta F\|/\sigma\rangle$	1.05	1.015	1.077	0.985
$\langle\|\Delta F /\sigma\rangle$	-0.09	-0.02	+0.01	+0.01
N4σ	119	74	139	54

two measurements by more than four times its standard deviation. The anomalous difference for reflections having only two measurements, each from a different crystal, was calculated normally but its standard deviation was artificially doubled.

The files of edited data were then subjected to a further cycle of determining the level scale factors using HRS refinement. These rescale factors were determined in order to ascertain whether either the weighting scheme used for the common row averages, where the weights depend on the standard deviations and thus the initial scale of the measurements, or the inclusion of the sets of inconsistent levels which were subsequently eliminated, had biased the original set of scale factors. The few rescale factors which differed from unity by more than 1% are listed in Table 13, indicating that the exclusion of the inconsistent levels was primarily responsible for their variation. Most of the rescale factors differed from unity by less than 0.2%.

The residuals of the HRS refinement, the number of common measurements and the average value of the mean structure amplitude are also shown in Table 13 for both the initial scale factors and the rescale values, together with the corresponding values of statistics representing the overall agreement between the b* and c* sets of measurements. It is clear from these that for the NATIVE and MONO sets of data the average value of $\bar{F}_c$ differed from that of $\bar{F}_b$ by 0.4%. For the NATIVE data this resulted in a discontinuity in the average value of $\bar{F}$ when considered as a function of the index k. The value for k less than 5, corresponding to the B1 measurements was significantly less than would have been expected. Consideration of the direction of the discrepancy between the b* and c* measurements

showed that this resulted from the level scale factors being too low for B1 and too high for B2 for the NATIVE data. The reason for this is not known.

This disagreement was calculated for the individual common rows of intersections, but the results showed clearly that the variations appeared predominantly between the groups of levels from different crystals and not within them, and so only the averages over blocks of 25 intersections have been listed in Table 13, corresponding to the sixteen possible intersections between the four sets of levels about two axes.

A set of eight final rescale factors were calculated from these average values of $\langle \Delta F_{cb} \rangle / \langle \bar{F} \rangle$ by the following procedure. The average value of each column of the 4x4 matrix of numbers was taken as an approximation to the scale factors for levels B1,B2,B3,B4 and a new matrix, which would be expected to result from the application of these scale factors, was generated. The average value of each row of the new matrix was taken as an approximation to the scale factors appropriate for levels C1,C2,C3,C4. By considering averages first along the rows of the matrix, and then along the columns of a new matrix, a different set of eight scale factors was determined, although the resultant matrix to be expected after application of either set of eight scale factors was essentially the same in both cases. It showed no significant features for the native or any of the derivative sets of data.

The final rescale factors applied to each set of data were the averages, after normalisation, of the two sets determined as described in the previous paragraph. These factors differed significantly from unity for only the NATIVE and MONO sets of measurements as

TABLE 14 Statistics of Compiled Data Sets.
The average structure amplitude and the number of reflections having 0,1,2,3,4 observations and 0,1,2,X estimates of the anomalous difference, where X denotes an estimate involving two crystals, are listed for each data set together with averages involving the scatter and deviation between measurements. An asterisk indicates that for the derivative sets of data the scatter and deviations are those relative to $\overline{F}_+$ and $\overline{F}_-$ as appropriate and the average $\langle|\overline{F}_+ - \overline{F}_-|\rangle$ involves only the acentric reflections.

		NATIVE	MONO	EMP	BAKER
Nref	0	1026	972	1220	3185
	1	2187	1880	2636	1750
	2	7653	5968	5070	5990
	3	2151	2917	3243	1887
	4	4534	5814	5382	4739
Ndif	0	3404	3125	4576	5113
	1	9418	8312	7255	7485
	2	4534	5814	5382	4739
	X	195	300	338	214
$\langle\overline{F}\rangle$		144	114	95	119
$\langle w.\|\Delta F_3\|/\overline{F}_3\rangle$ %		3.9(1)	4.4(1)	4.5(1)	4.3(1)
$\langle w.\|\Delta F_2\|/\overline{F}_2\rangle$ %		3.1(1)	4.0(1)	4.4(1)	3.9(1)
$\langle\|\Delta F_3\|\rangle/\langle\overline{F}\rangle$ %		13.0	11.4	10.8	11.9
$\langle\|\Delta F_2\|\rangle/\langle\overline{F}\rangle$ %		10.9	12.3	10.8	11.2
*$\langle$Scatter/$\overline{F}\rangle$ %		18.1	16.0	16.0	14.7
*$\langle\|F_i - \overline{F}\|\rangle/\langle\overline{F}\rangle$ %		9.3	8.6	9.0	8.1
*$\langle$Scatter/$\sigma\rangle$ %		188.	151.	149.	140.
$\langle\|\overline{F}_+ - \overline{F}_-\|\rangle/\langle 2\overline{F}\rangle$ %		7.4	6.9	7.2	6.8
*$\langle\|\overline{F}_+ - \overline{F}_-\|\rangle/\langle 2\overline{F}\rangle$ %			6.8	7.0	6.7
Limiting Values for large $\overline{F}$					
*$\langle$Scatter / $\overline{F}\rangle$ %		1.9	1.7	2.5	2.7
*$\langle$Scatter / $\sigma\rangle$ %		2.8	1.7	1.8	1.8

can be seen from Table 13. Application of these final rescale factors gave agreement between the b* and c* sets of measurements which was improved according to some of the criteria listed in Table 13 but not according to others. However, the resulting set of compiled data was much more homogeneous for the NATIVE measurements, and there was no reason for scaling the derivatives differently, and this procedure was thought to be equally valid, though less necessary, for these sets of measurements.

Overall statistics for the compiled data sets were determined using these structure amplitudes and the means and differences derived from them.

E) STATISTICS OF COMPILED DATA SETS

The agreement between equivalent measurements was considered in terms of groups of reflections having constant values of the indices k and l, to allow rapid identification of any anomalous rows or levels of reflections, and also in terms of those having similar values of $Sin^2\theta$ and the structure amplitude F, to allow the nature of the variations to be analysed. However, there were no significant departures from the trends which would be expected. The overall agreement can be seen from averages in Table 14.

The levels of reflections measured perpendicular to the b* axis appear slightly better than those about the c* axis as judged by both the weighted mean ratio of the difference between a pair of Bijvoet reflections to their mean amplitude, and also the ratio of the mean difference to the mean amplitude for all the reflections in the level. The significance of this difference between the b* and c*

facing page 127

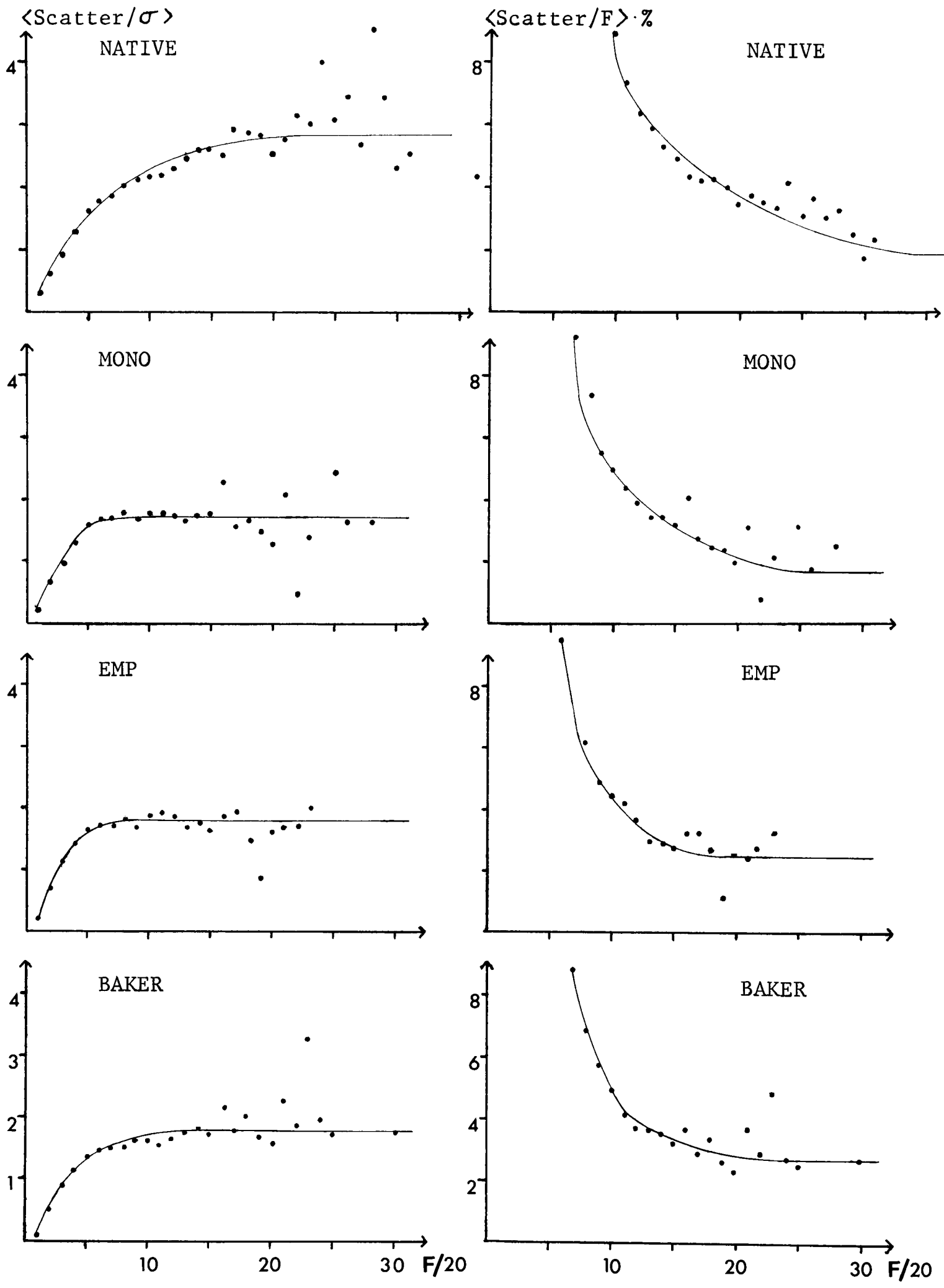


FIGURE 29 Scatter of Observations Within Data Sets.
The mean ratios <Scatter/σ> and <Scatter/F> are shown as a function of the structure amplitude F.

levels is low, but it is consistent with earlier conclusions concerning the optimum strategy for the data collection. In all cases the value of the weighted mean ratio is about 4.2%, and the ratio of the means is approximately 11.5%.

The scatter of all the measurements of a reflection was calculated as the square root of

$$s^2 = \sum_{i=1}^{n} \frac{(F_i - \overline{F})^2}{n - 1}$$

for n measurements, and the deviation was determined as the direct mean

$$dev^{n.} = \sum_{i=1}^{n} \frac{(| F_i - \overline{F} |)}{n}$$

For the derivative sets of data, the contribution of the anomalous scattering was excluded from these quantities, considering the differences as being

$$| F_i - \overline{F}_+ | \qquad \text{or} \qquad | F_i - \overline{F}_- |$$

as appropriate.

The average deviation is about 9% of mean structure amplitude in this case and the average scatter is almost twice this value. However, the variation of the scatter with the mean structure amplitude as shown in Figure 29 provides more information than the overall values. For large amplitudes the scatter (σ_{obs}) tends to a constant value of about 2% of the mean amplitude $\overline{F}$ as shown by the mean ratio of ($\sigma_{obs}/\overline{F}$). The ratio of the scatter to calculated

standard deviation (σ_{calc}) tends to a constant minimum at large amplitudes as can be seen from the variation of the mean ratio ($\sigma_{obs}/\sigma_{calc}$), suggesting that the calculated standard deviation should have included a larger percentage of the structure amplitude than that used, 1% for the native, 2% for the derivative sets of data. The ratio of the scatter to the standard deviation indicates the values

2.8 3.4 3.6 3.6

whereas the ratio of scatter to the amplitude indicates

1.9 1.7 2.5 2.7

which are lower in every case. The discrepancy between these indications must be assumed to result from the simplified treatment of the errors neglecting other sources of variation. For the very weak reflections, the scatter and deviations were anomalously small, as a result of having eliminated all the bad disagreements in which the integrated peak intensity was recorded as being less than the background.

Calculation of the average value of the difference between a Bijvoet pair of reflections, considering both the acentric reflections alone and all the reflections together, indicated the surprising fact that this difference is larger for the centric than the acentric reflections. This difference is no smaller than that resulting from the previous processing of the measurements, which had involved incorrect application of the absorption correction as already described, although it might be expected to be one of the most sensitive indicators of this error. The systematic difference it caused between a Bijvoet pair of reflections was in opposite directions for the b* and c* sets of levels. More than a quarter of the reflections in each set of measurements were recorded four times as can be seen

facing page 129

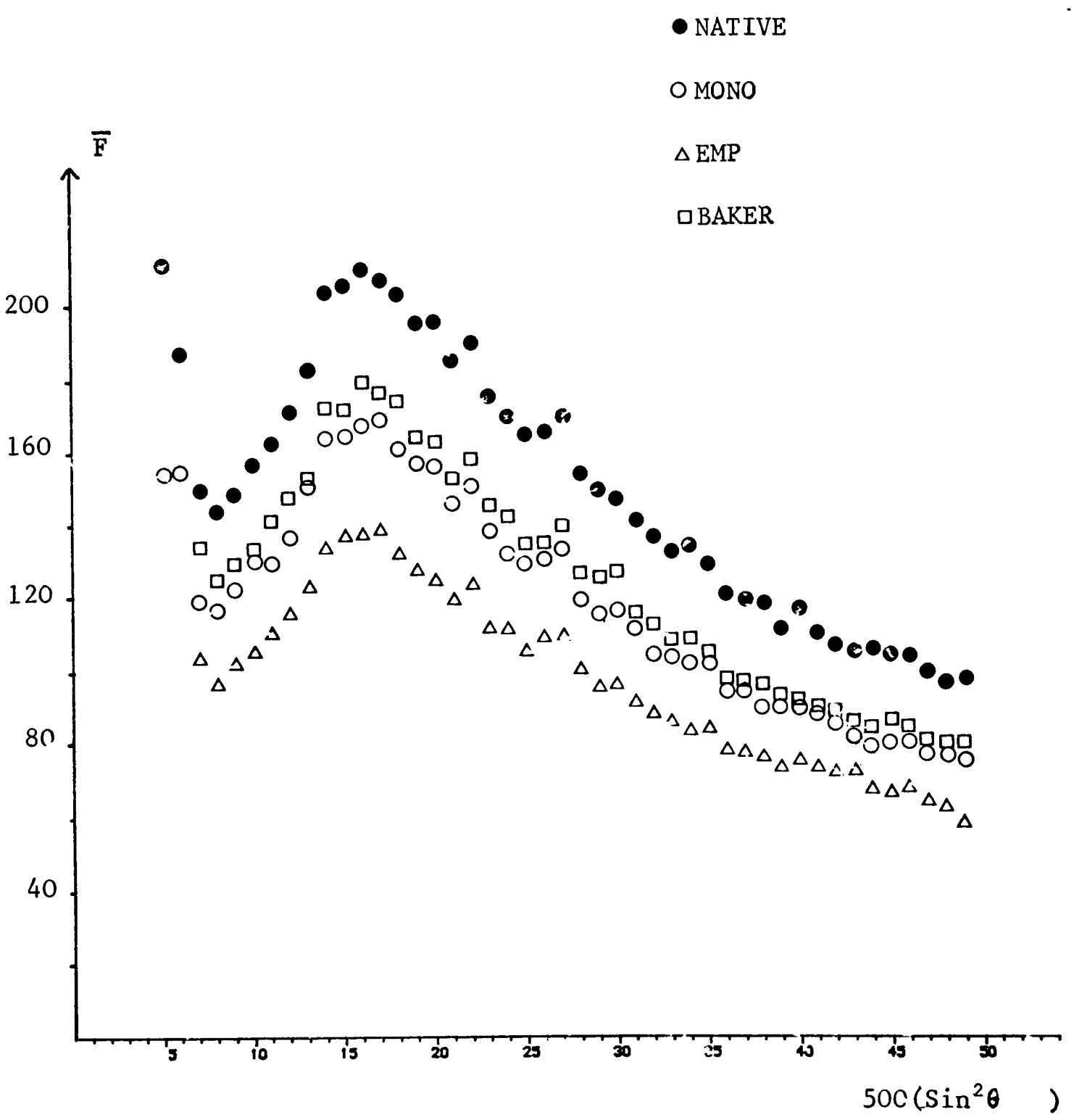


FIGURE 30 Mean Structure Amplitudes.
For each set of measurements the variation of the mean amplitude $\overline{F}$ is shown as a function of (Sin$^2\theta$).

from Table 14 and so the errors would tend to cancel out in those cases. Any other effects of this error are not readily predictable. The overall statistics of the data have not changed much on re-processing them, but experience of previous difficulties of scaling between derivative and native sets of measurements, and anomalous variations of the mean structure amplitude with the indices for the NATIVE reflections had enabled much more detailed analyses of the level scaling to be made on re-processing the measurements. The mean structure amplitudes are seen in Figure 30, before scaling them together. The NATIVE amplitudes should be multiplied by three to convert them approximately to an absolute scale.

F) PRELIMINARY SCALING BETWEEN DATA SETS

The scale factors which needed to be applied to the intensities to make their average values equal to those for the NATIVE measurements, were calculated for all reflections having common integral values of h,k,l,($\bar{F}$/20) and (500 $\mathrm{Sin}^2\theta$). The variation with this function of the Bragg angle was logarithmic, the parameters of the least squares regression line of the logarithm of the scale factor k being shown in Table 15. The relative temperature factor B between the derivative and native data was readily determined from the slope of this line which equals ($2B/\lambda^2$)/500. The largest B value of 3.2 is for the BAKER derivative. The exponential scale factors were applied to the derivative measurements and the resulting rescale factors determined as listed in Table 15.

Lines of regression of these rescale factors against the three indices were calculated, but detailed inspection of the variations in

TABLE 15 Preliminary Scaling of Derivative Data.
The scale factors appropriate for the derivative intensities, making their mean value equal to that of the native intensities are listed as an overall constant, and an exponential function from which the temperature factors were deduced. After application of the scale factor, rescale factors were determined for the principal projections, in terms of both intensities and amplitudes, and for all the reflections in terms of blocks with constant h,k,l, $S=50(\sin^2\theta/\lambda^2)$, for which the linear regression parameters are listed with their standard deviations.

		MONO	EMP	BAKER
Scale Factors:				
k_I= constant		1.5880(0)	2.3273(0)	1.4410(0)
k_I= exp(aS + b)	b	0.3925(542)	0.7806(523)	0.2387(160)
	a	+0.0029(26)	+0.0028(26)	+0.0053(48)
Temperature Factor	B	1.7	1.7	3.2
Rescale Factors:				
Intensities				
k_I= ah + b	b	0.9961(72)	1.0140(67)	1.0359(86)
	a	+0.0005(5)	-0.0010(10)	-0.0023(5)
k_I= ak + b	b	1.0541(143)	0.9986(123)	0.9974(120)
	a	-0.0048(12)	-0.0001(10)	+0.0003(10)
k_I= al + b	b	0.9607(137)	0.9747(133)	0.9501(149)
	a	+0.0054(14)	+0.0032(14)	+0.0061(16)
k_I= aS + b	b	0.9963(136)	0.9945(155)	1.0017(119)
	a	+0.0002(7)	+0.0002(8)	+0.0000(6)
k_I(hk0)		0.9169(64)	0.9377(66)	0.8534(57)
k_I(h01)		1.0716(59)	1.0245(72)	1.0428(106)
k_I(0k1)		0.9579(69)	1.0010(77)	1.0124(82)
Amplitudes				
k_F(hk0)		0.9705(76)	1.0021(86)	0.9480(73)
k_F(h01)		1.0033(66)	1.0097(95)	1.0122(130)
k_F(0k1)		0.9829(85)	1.0151(102)	1.0147(108)

the rescale factors showed that these were not linear. The largest variations for each derivative were with respect to l, the slope of the regression line being positive in all three cases. This results from the scale factors being less than unity for the first five blocks ($0 \leqslant l \leqslant 4$) and greater than unity for the second five ($5 \leqslant l \leqslant 9$). The similarity of the trends in these scale factors for all three derivatives suggests that the NATIVE data set is responsible for the variations observed and these can clearly be correlated with the final rescaling factors which were applied to the levels of NATIVE data. The variation with respect to k of the rescale factor between the MONO and NATIVE sets of intensities can be correlated with the final level rescaling factors applied to the MONO measurements. The variation with respect to h of the intensity rescaling factors for the BAKER derivative have no apparent explanation.

The rescale factors for the principal projections, which are listed in Table 15 with their standard deviations, show some highly significant variations from unity. Statistical considerations which are discussed in the next chapter suggest that values less than unity would be expected in these projections for the rescale factors k_I to equalise the average intensities, and values of unity for the factors k_F to equalise the average structure amplitudes. The values calculated for k_I for the hO1 reflections of each derivative and for k_F for the hkO reflections of MONO and EMP thus indicate inhomogeneity within the sets of structure amplitudes for which there is no apparent reason. Such variations between the centric and acentric reflections were also apparent in the low resolution data, in which all the measurements for the native or derivative reflections were obtained from a single crystal.

facing page 131

TABLE 16 Incorporation of Low Resolution Data.
The mean difference between the common reflections of the 6Å set and the 2.5Å set is listed with its standard deviation, the number of common reflections in the two sets of measurements and the number of those for which the difference between the two estimates of the structure amplitude exceeded four times its standard deviation.

	NATIVE	MONO/PCMB	EMP	BAKER
$\lvert\Delta F\rvert$	11.0	19.8	19.4	16.4
σ	13.6	15.6	20.3	20.2
$\langle F \rangle$	155.	153.	154.	152.
No. of common reflections	696	655	723	700
No. of large differences ($N4\sigma$)	28	66	37	23

G) INCORPORATION OF LOW RESOLUTION DATA

The low resolution reflections represent most of those for which no measurement was made on the linear diffractometer. The two sets of data contained about 700 reflections in common. The scale factor applied to the low resolution structure amplitudes to make the average intensities of the common reflections equal in both sets of native data was 1.318(11). The agreement between the two sets of measurements was satisfactory, as shown by the values of the mean differences and the numbers of individual reflections differing significantly which are listed in Table 16.

The differences between the sets of derivative data have not been adjusted to allow for the preliminary scaling of the low resolution measurements made on the four circle diffractometer to an average amplitude 2% greater than that of the native reflections whereas these averages were made equal for the linear diffractometer measurements. The only mean difference which exceeds its standard deviation is that between the PCMB derivative in the 6Å data and the MONO derivative in the 2.5Å data. These are expected to be similar derivatives, but are clearly not identical. In all other cases less than 5% of the reflections differed by more than four standard deviations.

After scaling the two sets of measurements together, they were again divided into the 6Å set, and a 2.5Å set from which all reflections corresponding to spacings greater than 6Å had been removed. The refinements and phase determinations were done separately for these two sets of reflections because of the problems which would otherwise arise both from the use of different solutions for preparing the diffused derivatives, which could give rise to different absolute and relative occupancies, and also from the use of PCMB in one set

and MONO in the other as the first derivative.

H) INITIAL REFINEMENT USING CENTRIC REFLECTIONS

The heavy atom parameters indicated by the reflections in the range of 6-2.5Å resolution were refined initially by the same methods as those already described for the low resolution data, using the program which is based on a modification of the method of Hart (1961) for consideration of only the centric reflections. The initial parameters for the MONO and EMP derivatives were those previously obtained from the low resolution reflections. Difficulties in determining the exact positions of the sulphydryl sites in the BAKER derivative, which have been described in the previous chapter, led to the choice of several sets of initial parameters, for this derivative, all of which were refined independently.

For reflections in this range of resolutions, variations in the occupancy, the isotropic temperature factor and the atomic scattering factor need to be considered separately, whereas at low resolution refinement of the occupancy is best achieved by fixing the other two variables whenever the occupancy is allowed to vary. For the higher resolution reflections the temperature factor and occupancy were always allowed to refine together. In some instances this produced negative temperature factors, which were re-set to $B/\lambda^2 = 7.0$ for the beginning of any further refinement and to $B/\lambda^2 = 0.5$ for use in phase determination.

Analysis of the atomic scattering factor used for the mercury atoms

$$f_n(\theta) = -1609.\mathrm{Sin}^6\theta + 860.\mathrm{Sin}^4\theta - 211.\mathrm{Sin}^2\theta + 80.$$

TABLE 17 Parameters After Initial Refinement at 2.5Å Resolution. The parameters listed after refinement of two sites for the MONO and four sites for the EMP derivatives were obtained using as initial parameters those from the 6Å data. The initial coordinates for the sulphydryl sites of the BAKER derivative were taken from those of the MONO derivative, the remaining parameters being those of the low resolution data. The results of refining the MONO derivative in terms of the four sites of the EMP derivative are also shown.

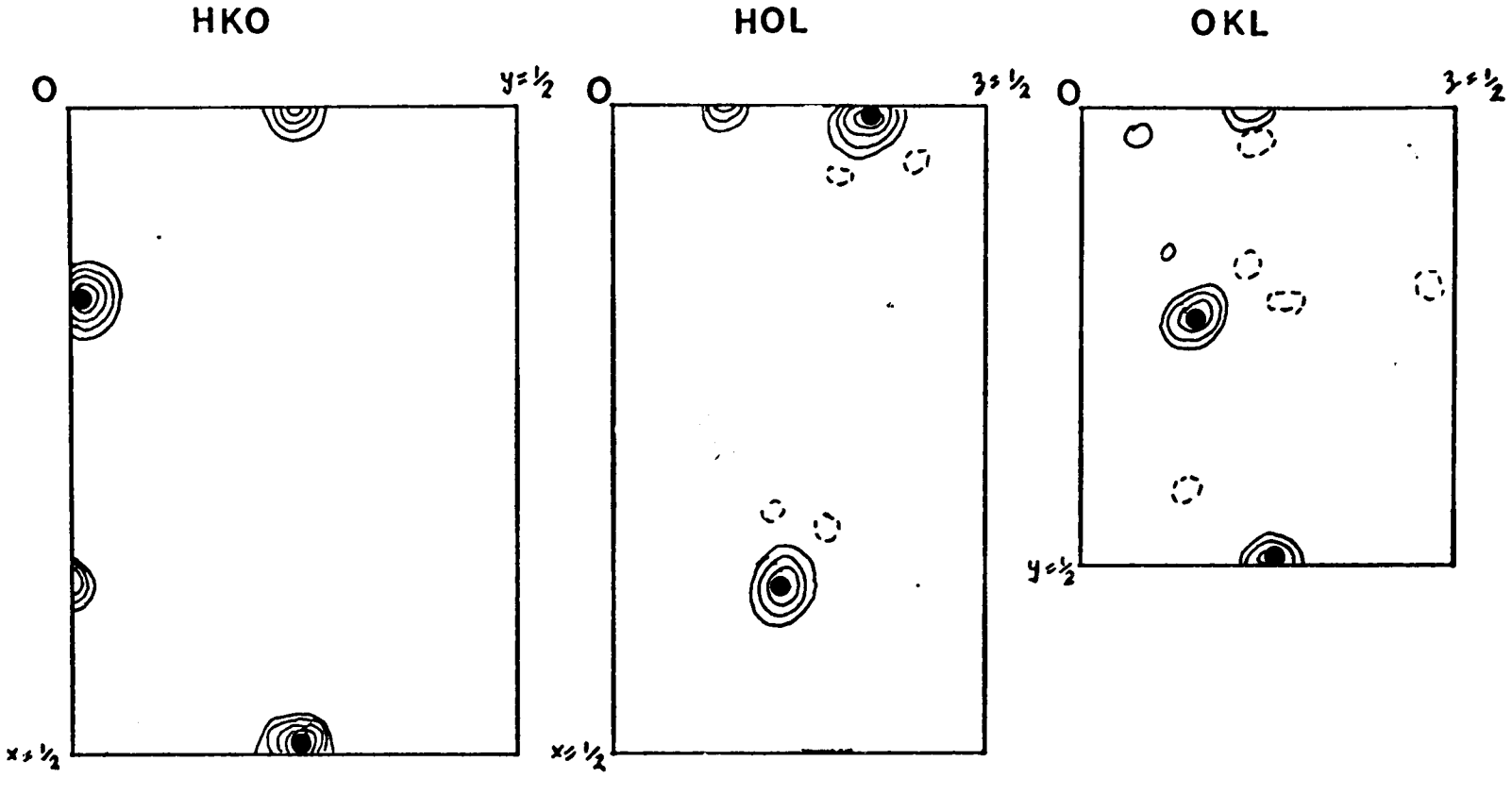
Derivative		MONO	EMP	BAKER		MONO
$\langle f_H \rangle$		32.	33.	25.		33.
Reliability Index R		0.619	0.616	0.623		0.611
RMS lack of closure E		34.0	31.5	24.8		33.6
Number of sites		2	4	4		4
Refined Scale k		0.9762	0.9794	0.9528		0.9774
Site 1	Z	.2569	.2792	.1775		.2432
	B/λ^2	3.74	13.5	-0.25		4.73
	x	.1249	.1239	.1229		.1248
	y	.0101	.0110	.0071		.0102
	z	.7321	.7310	.7638		.7339
Site 2	Z	.3109	.2076	.1518		.3281
	B/λ^2	8.9	3.3	13.35		-0.01
	x	-.0054	-.0066	-.0024		-.0056
	y	.2467	.2443	.2455		.2460
	z	.1639	.1622	.1682		.1635
Site 3	Z		.2069	.1299		.0415
	B/λ^2		10.2	12.3		-2.2
	x		.0007	.0788		-.0026
	y		.0470	-.0257		.0477
	z		.0512	.7801		.0716
Site 4	Z		.1423	.0793		.0478
	B/λ^2		7.8	11.31		2.31
	x		.0621	.0430		.0632
	y		.1901	.2709		.1812
	z		.6747	.1554		.6742

was performed by considering separately the reflections in successive shells of $Sin^2\theta$ within the 2.5Å sphere in reciprocal space. Ten groups of reflections were used, each containing about 190 reflections. For each group the occupancy of each site was refined, all other parameters being fixed at the values determined from consideration of all the centric reflections. Inadequacies in the parametric form of the atomic scattering factor would be revealed by variations of the refined occupancies from different shells of reflections.

The scatter of the resulting values for each site made a detailed analysis and the fitting of a power series curve to the data points impossible, but in most cases there was no trend indicative of any changes required in the scattering factor. The refined occupancies for the third and fourth sites of the EMP derivative increased and decreased respectively at higher angles, but these variations were eliminated by decreasing and increasing their respective temperature factors. It became clear that this method of refining parameters was not always capable of detecting and correcting those parameters, which produced only a local minimum in the lack of closure. However, errors of this sort affecting the occupancy and temperature factor would be detected by the procedure of refining shells of reflections separately, and those affecting the coordinates of the heavy atoms would be detected by examination of difference Fourier maps.

The parameters obtained by such initial refinements are listed in Table 17, and the corresponding self sign difference and double difference map projections are shown in Figures 31-33 for the three derivatives. The most significant features on the three dimensional maps were residual peaks on the $\Delta\Delta F$ maps at all the atomic sites, their

Difference Fourier ΔF



Double Difference Fourier ΔΔF

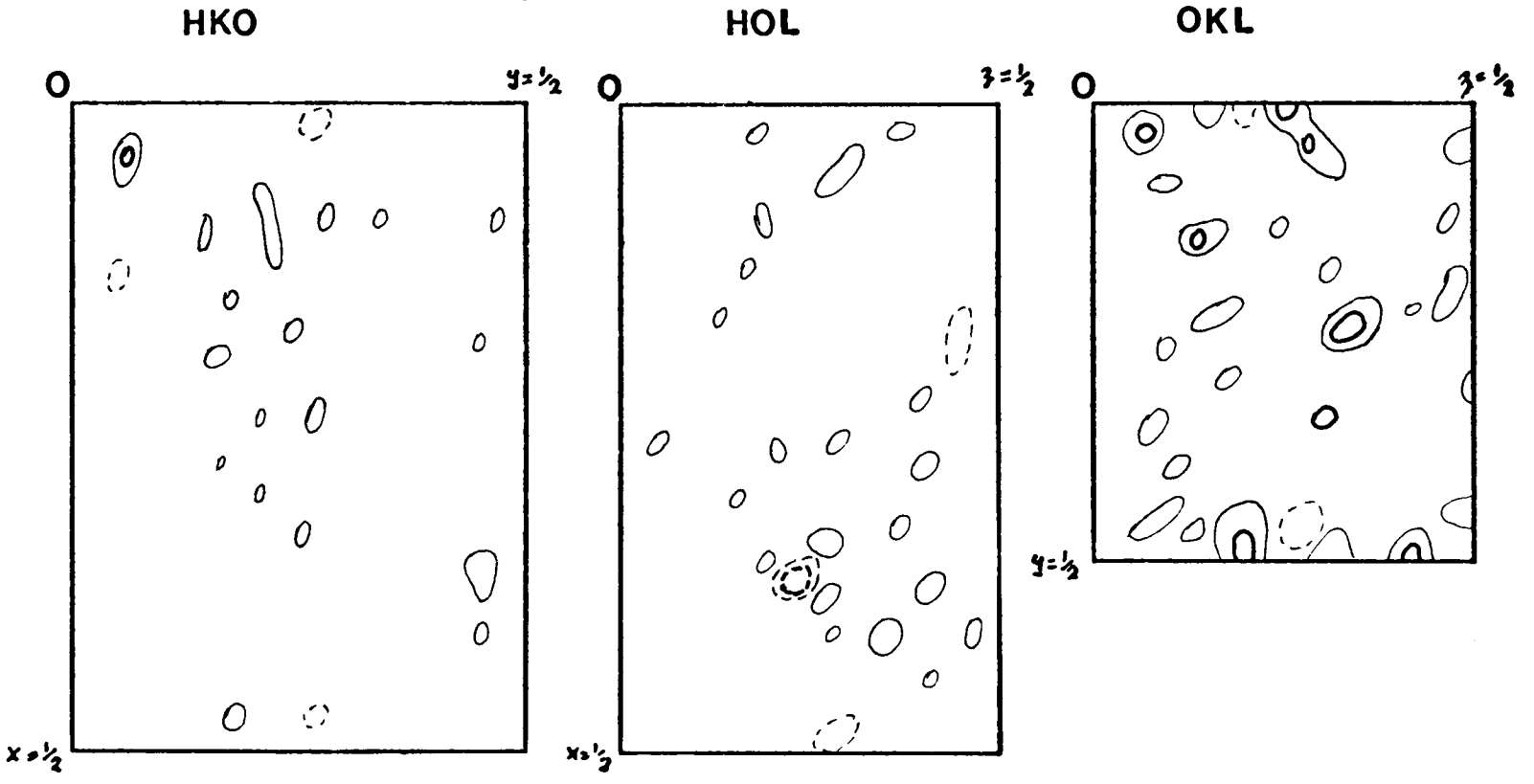
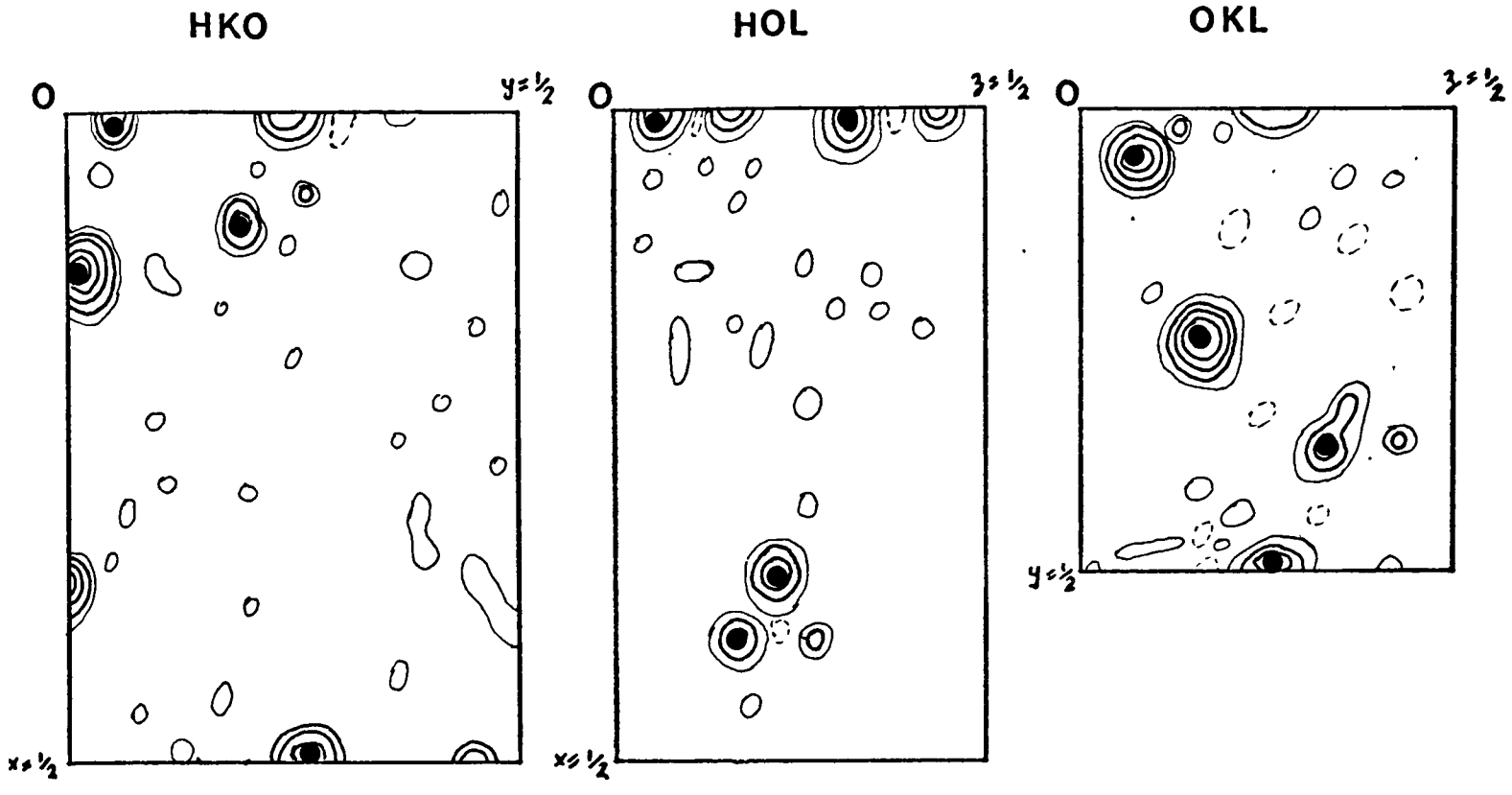


FIGURE 31 Projections of 2.5Å Best Sign ΔF and ΔΔF Maps: MONO Contours are drawn at intervals of 1 e.Å^{-3} above zero, the faint contours being at 0.5 e.Å^{-3}.

Difference Fourier ΔF



Double Difference Fourier ΔΔF

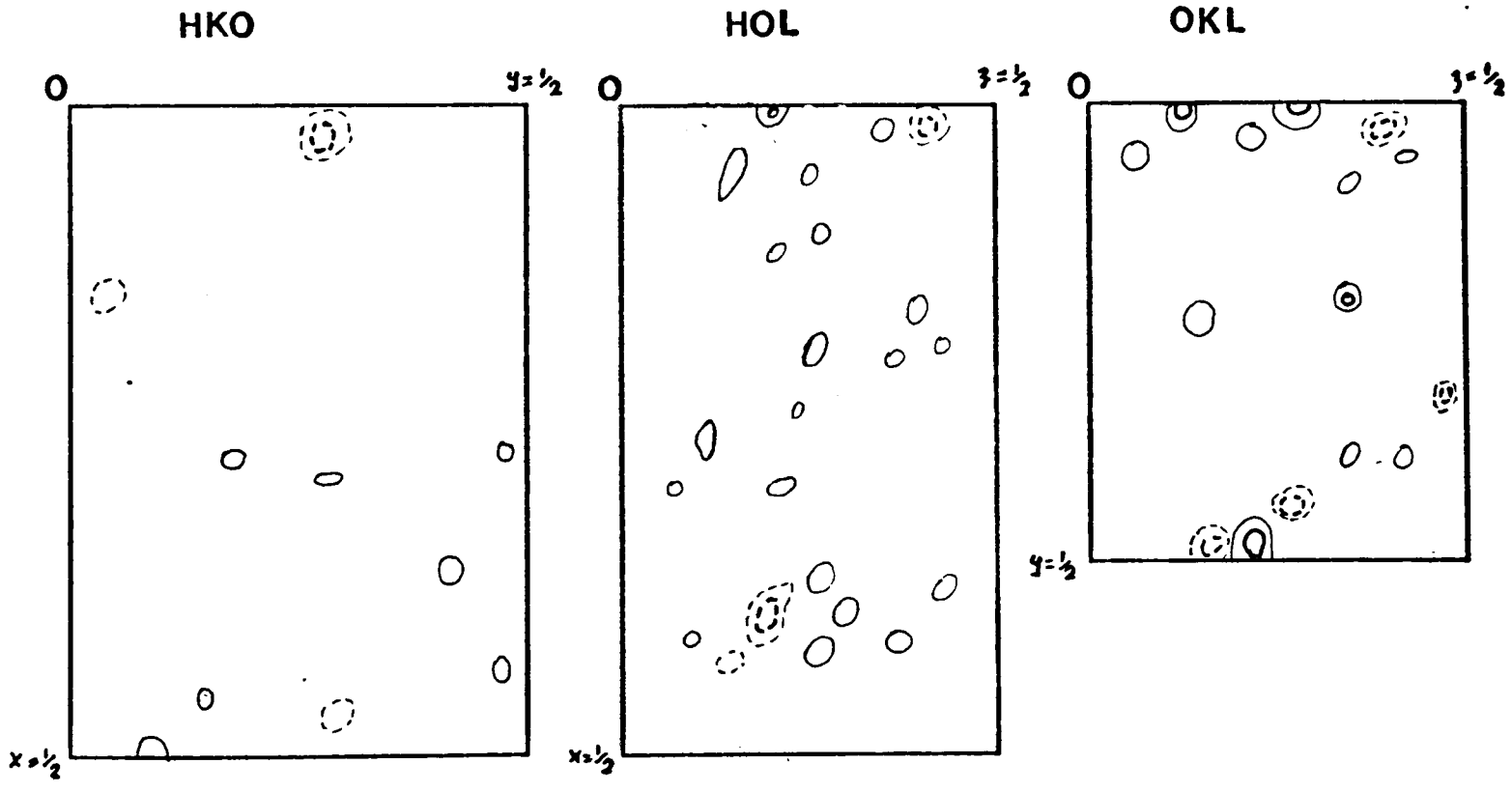
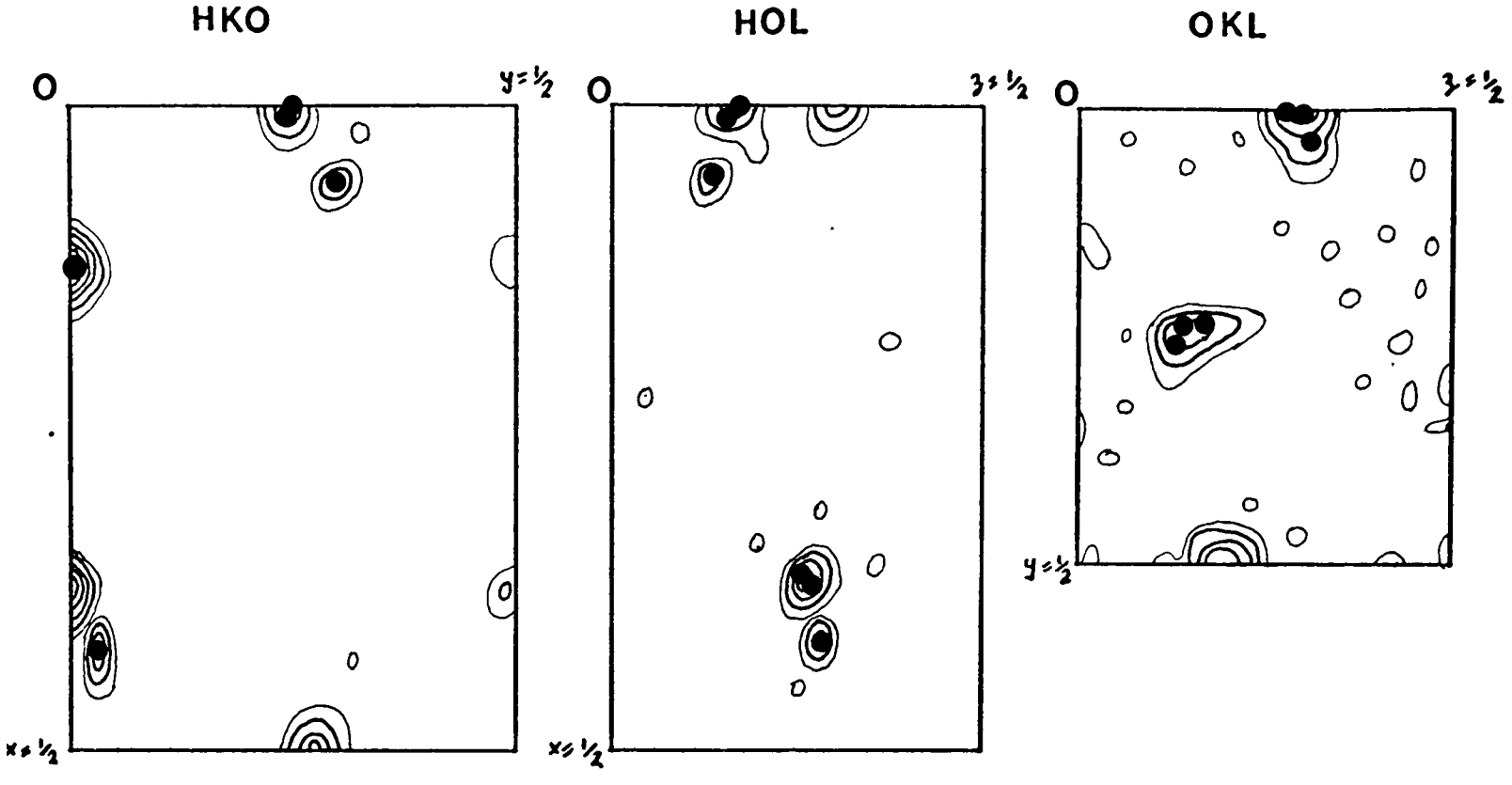


FIGURE 32 Projections of 2.5 Å Best Sign ΔF and ΔΔF Maps: EMP Contours are drawn at intervals of 1 e.Å^{-3} above zero, the faint contours being at 0.5 e.Å^{-3}.

Difference Fourier ΔF



Double Difference Fourier ΔΔF

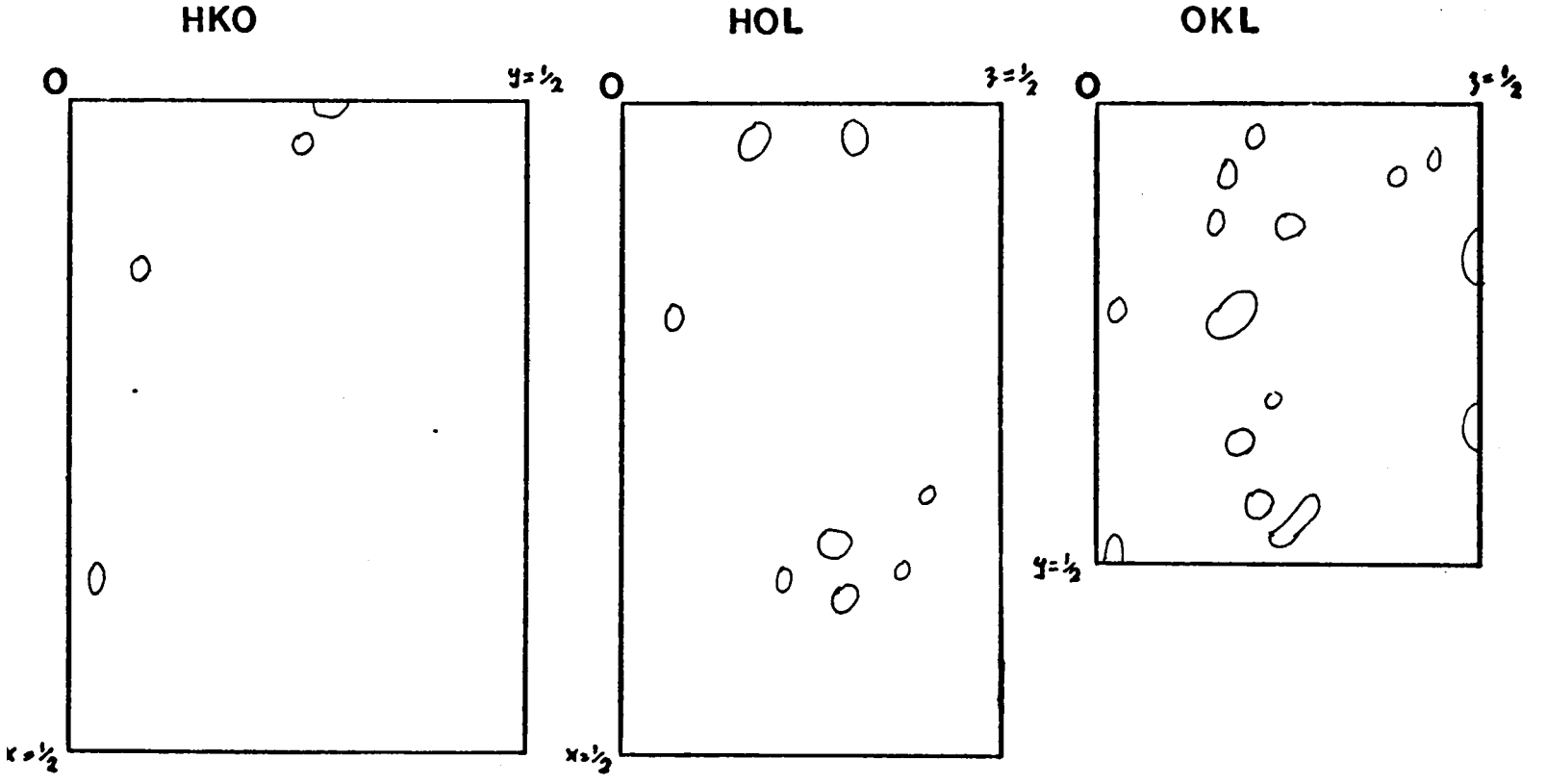


FIGURE 33 Projections of 2.5 Å Best Sign ΔF and ΔΔF Maps: BAKER Contours are drawn at intervals of 1 e.Å^{-3} above zero, the faint contours being at 0.5 e.Å^{-3}.

peak height being about 15-20% of those on the ΔF maps. Smaller peaks were also observed on the ΔF and $\Delta\Delta F$ maps for the MONO derivatives in the positions of both sites 3 and 4 of the EMP derivative, and also sites 3 and 4 of the BAKER derivative. There were indications of anisotropic thermal motion at several of the mercury binding sites.

No new atomic sites were indicated consistently by the three dimensional maps and the projection maps for any derivative. The indications of mercury sites in the MONO derivative in the positions of sites 3 and 4 of the EMP derivative were tested by refining four atomic sites against the isomorphous differences observed for the MONO derivative. The refined parameters are listed in Table 17, from which it appeared that these additional sites probably were significant.

Further refinement of the parameters for these derivatives involved detailed consideration of the methods of phase determination and the ways in which errors in the heavy atom parameters affect these phases, the electron density map of the protein and any difference Fourier maps or refinements for which the phases were subsequently used. Refinement of the parameters is especially complicated for these derivatives of chicken TIM because of the close correlation between the derivatives, all of which have two sites in common. These difficulties and other complications caused by the use of common site derivatives are considered in the next chapter following an analysis of the observed isomorphous and anomalous differences, and their implications with respect to the usefulness of the phase determining information to be expected from the separate derivatives.

TABLE 18 Unit Cell Dimensions for Native and Derivative Crystals. The average values of the reciprocal cell dimensions with their standard deviations are listed in r.l.u., together with the number of crystals for which each dimension was measured, the fractional change relative to the values for the native crystals and the derived values of the real unit cell translations in Å.

Reciprocal Cell:	a*	b*	c*
NATIVE	.014544(26)	.020624(34)	.024974(25)
MONO	.014470(27)	.020743(20)	.025030(60)
EMP	.014460(14)	.020650(20)	.024980(60)
BAKER	.014500(22)	.020622(20)	.024970(38)
Fractional Change in d*			
MONO	-.005	+.008	-.004
EMP	-.004	+.001	+.000
BAKER	-.003	-.000	-.000
Number of crystals			
NATIVE	12	8	8
MONO	10	6	5
EMP	5	2	4
BAKER	7	2	5

Real Cell:	a	b	c
NATIVE	106.01	74.76	61.74
MONO	106.55	74.33	61.60
EMP	106.63	74.66	61.72
BAKER	106.33	74.76	61.75

I) STATISTICAL TESTS OF ISOMORPHOUS AND ANOMALOUS DIFFERENCES

Crystals are strictly isomorphous only if they satisfy the two necessary conditions that their unit cells and also the positions of corresponding atoms within the unit cells are both identical. Protein crystals can show degrees of isomorphism ranging from that of a perfect derivative, differing from the native crystals only by the presence of a small number of ligands which have caused no distortion of the protein structure, to that shown by crystals of chicken TIM in the presence and absence of the substrate DHAP. These two structures are clearly related, but they can be regarded as isomorphous only at very low resolution, becoming significantly different above a resolution of about 8Å. The dimensions of the unit cells of two crystals can be readily compared, but the differences between the contents of their unit cells must be known in order to compare the observed differences in their structure amplitudes with those that would be expected for isomorphous structures.

The mean values of the unit cell translations in crystals of chicken TIM and its three derivatives, for which intensity measurements at 2.5Å resolution have been made, are listed in Table 18 with the standard deviations estimated from measurements of several crystals. The longest dimension a is the most variable, the other two differing only in crystals of the MONO derivative, but all the changes are less than 1%. None of the dimensions of the unit cell in BAKER crystals differ significantly from those of the native protein. The relatively large variations in the longest unit translation would be expected in view of the much larger changes in this dimension on diffusing substrates or inhibitors into the crystals as described in the previous chapter.

The relation between the measured intensities I_o and the intensities expressed on an absolute scale I_{abs} must be known before the expected isomorphous differences can be calculated. The required scale factor k can be determined by consideration of the average values of the measured intensities, whose variation with $Sin^2\theta$ depends on the overall temperature factor B of the atoms in the crystal. The relationship between these quantities and the atomic scattering factors

$$\sum f_i^2(\theta) = \langle I_{abs}(\theta) \rangle = k \langle I_o(\theta) \rangle \exp\left(-2B \, Sin^2\theta/\lambda^2\right)$$

was shown by Wilson (1942) to be appropriate for diffraction data at atomic resolution. This probability distribution of the diffracted intensities is known to be a poor approximation to that observed from protein crystals (Harker, 1953 and Luzzatti, 1955) for which the Wilson plots of $\log(\langle I_o \rangle / \Sigma f_i^2)$ against $(Sin^2\theta/\lambda^2)$ are not linear. However they approximate to a linear relationship for the high resolution data. The scale factor determined from a Wilson plot is likely to be an underestimate, because of the effects of poorly resolved atoms or, in the case of plots using projection data only, the overlapping of atoms in the projected structure. For protein crystals, the plots are usually calculated by considering only the atoms of the protein molecule and neglecting those of the bound water and mother liquor (Cullis et al., 1961).

A single molecule of TIM contains 1700 atoms, excluding hydrogen, as determined from the amino acid composition. The crystallographic determination of the molecular weight indicates the presence of about 1750 atoms of equivalent atomic weight to nitrogen. All of

Selection of measurements	k_{F_p}	$\lambda^2 B$	
A) all reflections, 6 - 2.5Å.	2.3	83.	
B) most reflections, 5 - 2.5Å	2.8	58.	(BEST)
C) high resolution, 3 - 2.5Å	3.4	46.	

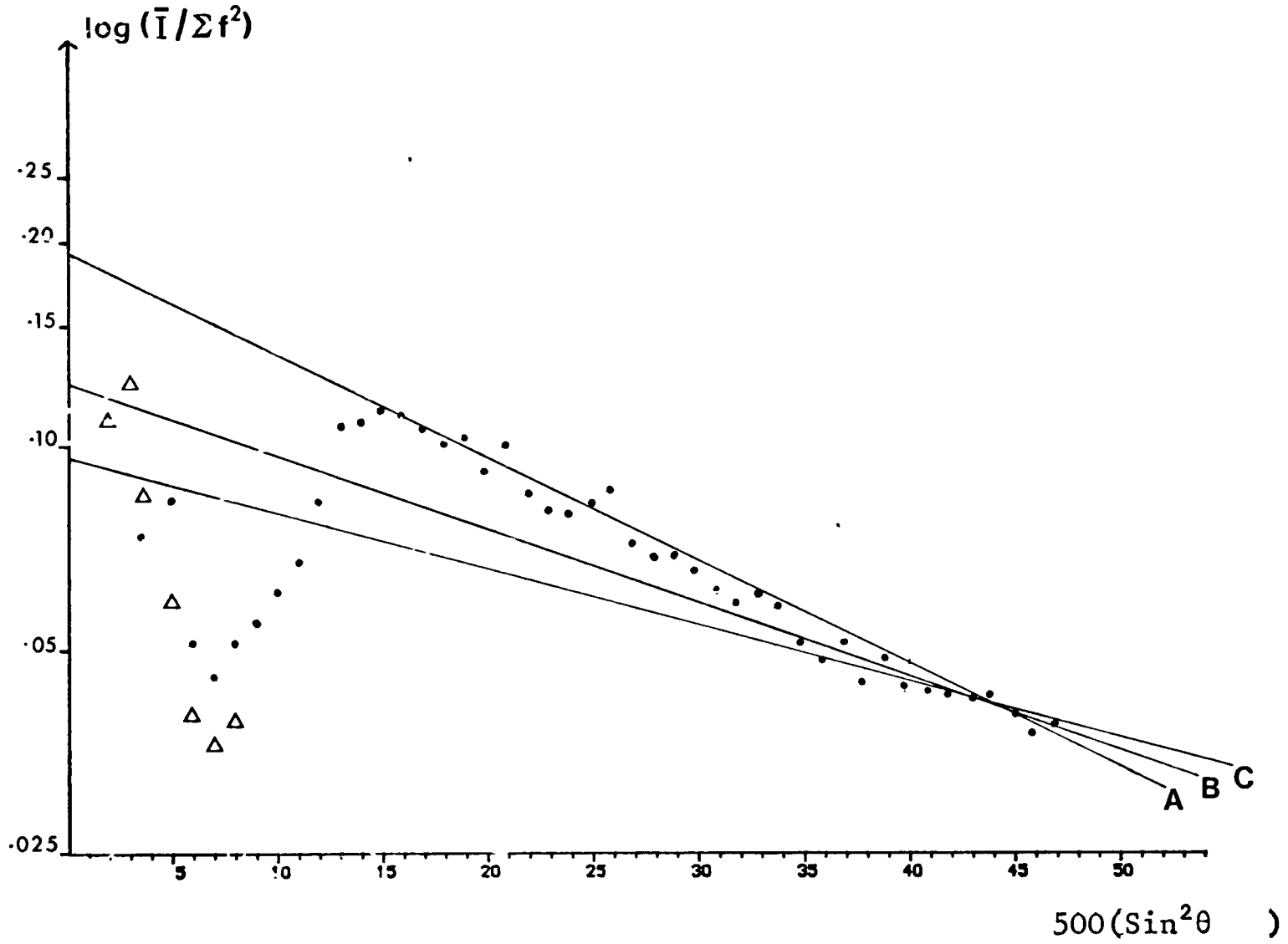


FIGURE 34 Wilson Plot of the Intensities for NATIVE Crystals. Measurements from the 6Å and 2.5Å sets of data are shown Δ and • respectively.

the non-hydrogen atoms in a protein have scattering curves which are similar to that of nitrogen, differing in relation to their exact atomic number, and so the value of the structure amplitude F(000) of the undiffracted ray is approximately that resulting from a structure having four molecules of 3400 nitrogen atoms per unit cell, which equals $9{\cdot}5 \ .10^4$.

The average intensity expressed on an absolute scale as determined for the same hypothetical structure equivalent to the crystals of chicken TIM is $16{\cdot}7 \ .10^4$, corresponding to an RMS structure amplitude of 408. These values are approximate only for reflections having very small Bragg angles. The RMS structure amplitude for all the reflections within the 2.5Å sphere in the compiled set of measurements for the native crystals is $2{\cdot}8 \ .10^4$. The variation of their amplitude with $\mathrm{Sin}^2\theta$ is shown in Figure 30 and the variation of the ratio of their intensity to the sum of the squares of the atomic scattering factors is shown as a Wilson plot in Figure 34 from which the scale factor necessary to convert the amplitudes to an absolute scale is seen to be 2.8 ± 0.2 and the overall temperature factor of the crystal 25 ± 6 .

Multiplication of the heavy atom occupancies by this scale factor of 2.8 converts them into fractional values on an absolute scale. The most highly occupied site, the second sulphydryl site in the MONO derivative is thus 90% occupied, implying that this site contains an average of 72 electrons out of a total of 80 for a single mercury atom. The occupancies in the diffused derivatives are somewhat lower, the weakest site identified representing about 10 electrons.

Knowledge of the heavy atom occupancies on an absolute scale enables the absolute ratio of the mean derivative to native intensities

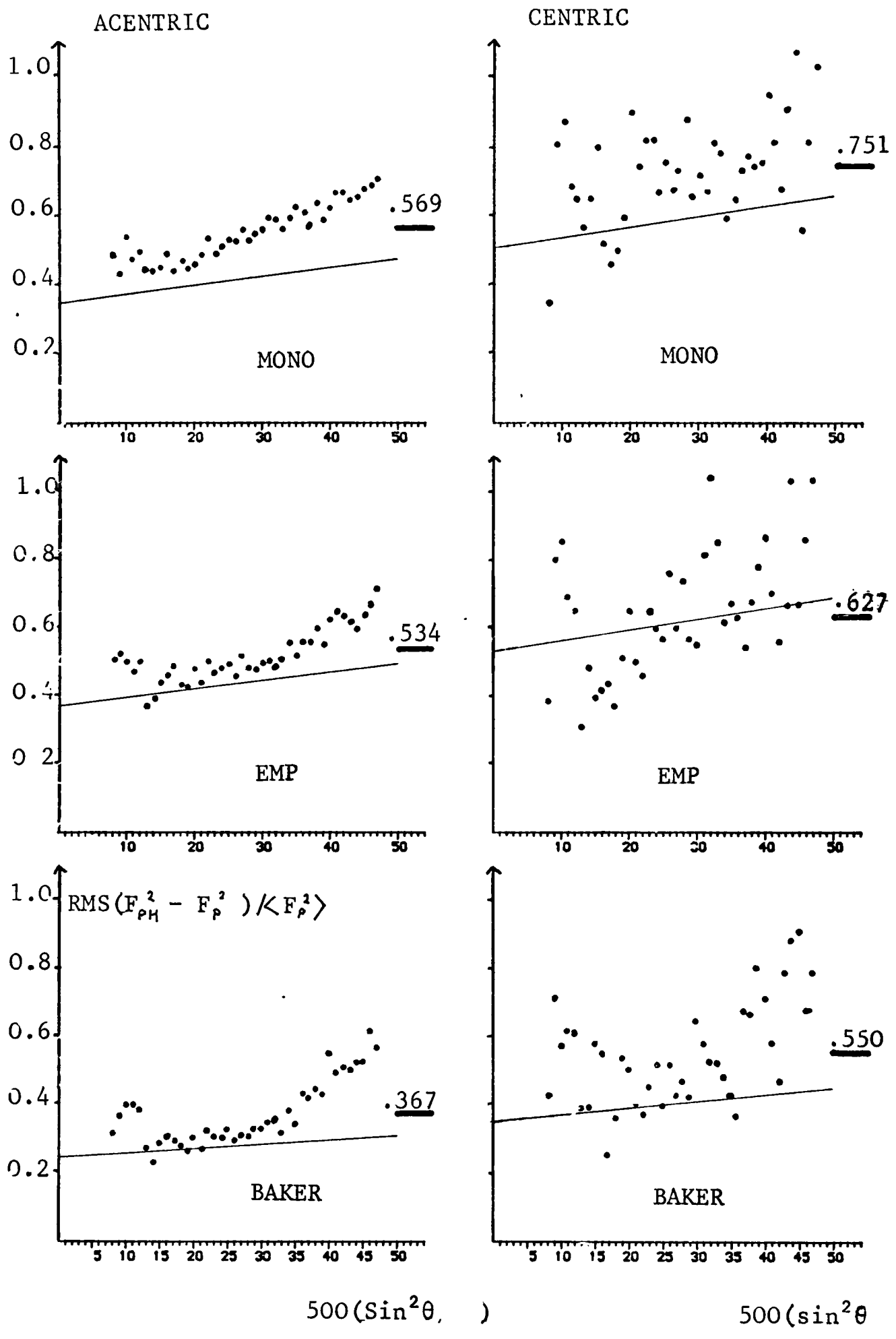


FIGURE 35 Crick-Magdoff Plots of the Isomorphous Differences. The observed (•) and calculated (-) values are shown for the acentric and centric reflections separately as a function of ($Sin^2\theta$) for each derivative. The overall observed values are marked (▬).

to be estimated, since if Wilson statistics are obeyed

$$\langle I_{PH} \rangle = \sum_{protein} f_N^2 + \sum_{\substack{heavy \\ atoms}} f_{Hg}^2 = \langle I_p \rangle + \sum_{\substack{heavy \\ atoms}} f_{Hg}^2$$

The expected ratios $\langle I_{PH} \rangle / \langle I_p \rangle$ for the three derivatives are listed in Table 19, as calculated both for the initially determined heavy atom occupancies using only the centric reflections in the refinement and also those determined by the methods described in the following chapter.

The mean isomorphous difference in the intensities was shown by Crick and Magdoff (1956) to be related to the extent of heavy atom substitution by the equation

$$\phi\Delta I = \frac{RMS(I_{PH} - I_P)}{\langle I_P \rangle} = n \cdot \left(\frac{N_H}{N_P} \right)^{1/2} \cdot \left(\frac{f_H}{f_P} \right)$$

where N is the number of atoms, subscripts H and p refer to the heavy atoms and the protein, and n is a constant which is 2 for centric and $\sqrt{2}$ for acentric reflections. The observed values of this function are plotted in Figure 35 and compared with those determined by the above formula.

In every case the observed differences are larger than would be expected if the derivatives were perfectly isomorphous, the discrepancy being greatest for the MONO derivative. The differences observed for the BAKER intensities exceed the expected value only for reflections close to the 2.5Å limit.

The predicted values of the Crick and Magdoff function $\Phi\Delta I$ are not very sensitive to the derivative scale factor but they are proportional to the heavy atom occupancy. The occupancy and scale

facing page 139

TABLE 19 Isomorphous and Anomalous Difference Statistics.
Mean and RMS values for the observed intensities, structure amplitudes, isomorphous and anomalous differences for centric and acentric reflections are listed for each derivative. The different values listed for the native measurements arise partly from the presence of different reflections in the derivative sets of data and partly from the effects of rounding errors. All the amplitudes should be multiplied by 3 to obtain an approximately absolute scale. The derivative scale factors calculated according to Wilson statistics are listed and also the empirical estimate of the anomalous scattering ratio.

		MONO		EMP		BAKER	
		Acent.	Centric	Acent.	Centric	Acent.	Centric
F_P	Mean	144	147	145	149	144	147
	RMS	26509	30443	26776	31348	26491	29960
F_{PH}	Mean	145	150	145	149	144	150
	RMS	26416	31456	26805	32168	26309	32192
Isomorphous Differences:							
$\lvert F_{PH} - F_P \rvert$	Mean	33	41	32	38	23	31
	RMS	42	53	40	49	30	40
$\lvert F_{PH}^2 - F_P^2 \rvert$	Mean	10079	13919	9579	12311	6695	9983
	RMS	15088	22867	14302	19660	9730	16475
Anomalous Differences:							
$\lvert F_{PH+} - F_{PH^-} \rvert$	Mean	20	22	20	21	19	21
	RMS	26	29	27	28	26	27
$\lvert F_{PH+} - F_{PH-} \rvert$	Mean	5289	6246	5629	6208	5164	6579
	RMS	7146	8326	7545	8769	6967	9957
Scale Factor $k_I = 1 + \langle f_{H\,calc}^2 \rangle / \langle F_P^2 \rangle$							
Initial occupancies (Table 17)							
$(\sin\theta/\lambda)=0.0$			1.060		1.063		1.026
0.2			1.095		1.103		1.045
Final occupancies (Table 23)							
$(\sin\theta/\lambda)=0.0$			1.066		1.072		1.030
0.2			1.105		1.117		1.051
Empirical Ratio $k' = f_H'/f_H''$ (Overall value)			3.38		3.09		2.41

factor are highly correlated and the determination of the derivative scale factor by these statistical methods is possible only to the extent that the estimates of heavy atom occupancy are valid. All of the mean observed differences listed in Table 19 involved the amplitude of these differences and not their sign and so they are not expected to differ significantly from their values after more appropriate scaling of the derivative data.

The ratio of the observed isomorphous to anomalous differences was shown by Matthews (1966b) to provide an empirical estimate of the ratio

$$k' = f'/f'' = 2\langle |F_{PH} - F_P| \rangle / \langle |F_{PH+} - F_{PH-}| \rangle$$

where f', f'' are the real and imaginary components of the heavy atom scattering factor. The theoretical value for mercury is 8.3 increasing negligibly with the angle of diffraction, and the observed values are shown in Table 19 and Figure 36. The low value of k' and its decrease at higher angles reflect the relative accuracy with which the isomorphous and anomalous differences were measured.

Random errors in the measured intensities will cause an apparent increase in the heavy atom occupancy (Dodson and Vijayan, 1971), since the RMS values of both

$$|F_{PH} - F_P| \qquad \text{and} \qquad |F_{PH+} - F_{PH-}|$$

are increased by errors in the appropriate structure amplitudes if these errors are assumed to be independent. The isomorphous differences, being larger, are expected to have smaller percentage errors than the anomalous differences unless there are significant errors in the scaling between native and derivative crystals or a serious lack of

facing page 140

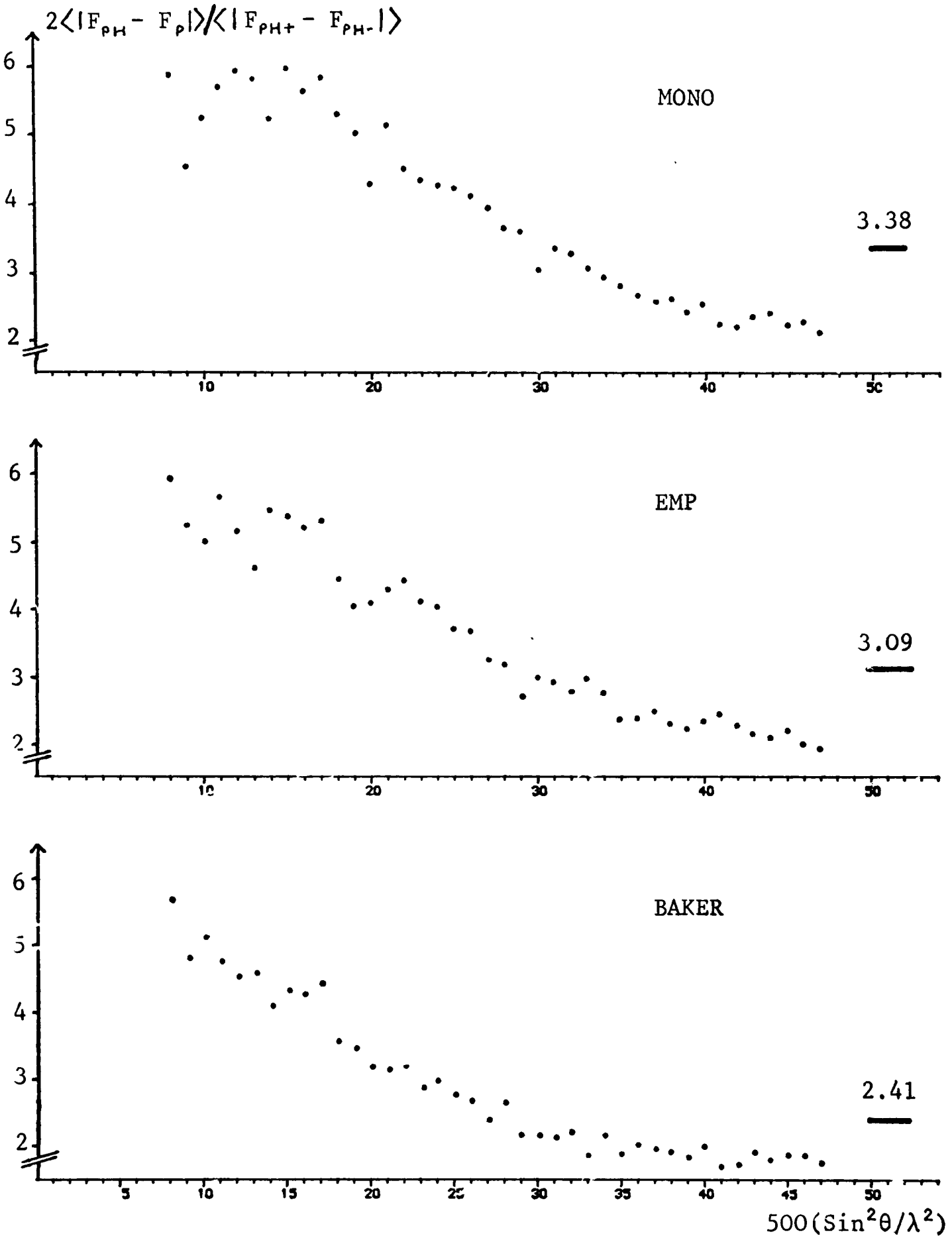


FIGURE 36 Empirical Estimate of the Anomalous Scattering Ratio. The values calculated for each derivative using the formula of Matthews (1966b) after preliminary scaling of the derivative data are shown for the acentric reflections as a function of $(\sin^2\theta/\lambda^2)$.

isomorphism. The relative accuracy of the observed isomorphous and anomalous differences influences the precision of the phase determination and the calculation of heavy atom structure factors for the reasons discussed in the next chapter.

Methods of Determining the Ratio f'/f'' Empirically

The formula proposed by Matthews (1966b) for the determination of the ratio f'/f'' is derived as follows, the notation used being shown in Figure 37, from which it is clear that for all reflections having f'' small compared with F_{PH} then

$$F_{PH+} - F_{PH-} = 2f'' \sin \gamma \qquad (1)$$

Matthews states that for these reflections it will always be a good approximation that

$$|F_{PH} - F_P| = |f' \cos \gamma| \qquad (2)$$

so that averaging over many reflections gives

$$\frac{\langle |F_{PH+} - F_{PH-}| \rangle}{2\langle f'' \rangle} = \frac{2}{\pi} = \frac{\langle |F_{PH} - F_P| \rangle}{\langle f' \rangle}$$

and thus

$$k' = \langle f'/f'' \rangle \sim \frac{2\langle |F_{PH} - F_P| \rangle}{\langle |F_{PH+} - F_{PH-}| \rangle} \qquad (3)$$

which relationship can be used as described above. Exclusion from the averages of all the weak reflections is said to ensure that the approximations of equations (1) and (2) are valid.

However, the approximation (2) is also bad whenever γ

facing page 141

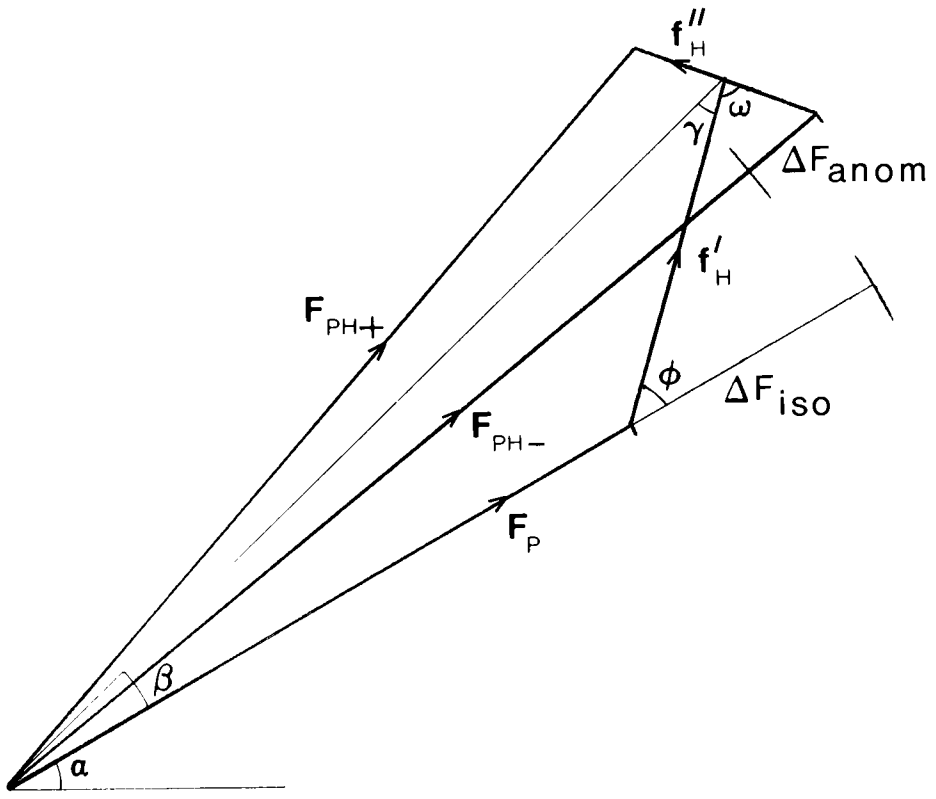


FIGURE 37 Phase Triangle Illustrating Anomalous Scattering by Heavy Atoms.
If the anomalous scattering arises from only a single species of heavy atom then the angle ω must be 90^{o}.

approaches $\pi/2$. Since the protein and heavy atom structure factors are uncorrelated, the angle φ is distributed randomly, β is usually small except for the weak reflections, and so γ would also be expected to be randomly distributed. Analyses of the phase triangles for chicken TIM indicate that this is the case.

It is proposed that the empirical ratio is better determined as

$$k'' = \left[\frac{2 \langle F_{PH}^2 - F_P^2 \rangle}{\langle (F_+ - F_-)^2 \rangle} \right]^{1/2} \qquad (4)$$

which may be derived as follows. The relationship

$$F_{PH}^2 = F_P^2 + f'^2 - 2F_P f' \cos \phi \qquad (5)$$

is true for every reflection, and if the protein and heavy atom structure factors are uncorrelated

$$\langle f'^2 \rangle = \langle F_{PH}^2 - F_P^2 \rangle . \qquad (6)$$

From equation (1) it is clear that

$$\langle f''^2 \sin^2 \gamma \rangle = \langle (F_{PH+} - F_{PH-})^2 \rangle / 4 \qquad (7)$$

and thus if γ is not correlated with the heavy atom structure factor

$$k''^2 = \frac{2 \langle F_{PH}^2 - F_P^2 \rangle}{\langle (F_{PH+} - F_{PH-})^2 \rangle} . \qquad (8)$$

Calculation of the empirical ratio for the 2.5Å measurements of chicken TIM using this formula (4) gave even lower results than those obtained by the use of Matthews' formula (3) as seen in Figure 36. A serious disadvantage of the use of equation (4) is the

sensitivity of the ratio to any small error in the derivative scale factor, since the numerator of this equation is the average of the absolute intensity difference. This problem does not arise with Matthews' formula which must thus be used whenever the derivative scale factor is uncertain. The measurements for chicken TIM do not provide a very good test of these alternative formulae since any refinement of the derivative scale factors is complicated as described in the next chapter by the sylphydryl binding sites being common to all the derivatives.

Chapter V PHASE DETERMINATION AND FURTHER REFINEMENT OF PARAMETERS FOR ISOMORPHOUS DERIVATIVES

A) METHODS OF PHASE DETERMINATION 144

Use of Anomalous Scattering Information 148

B) DETERMINATION OF PHASES FOR CHICKEN TIM 150

Mean Residual Errors 151

C) EFFECT OF ERRORS IN HEAVY ATOM PARAMETERS 155

D) DETERMINATION OF DERIVATIVE SCALE FACTORS 160

E) METHODS OF REFINING HEAVY ATOM PARAMETERS 165

F) FURTHER REFINEMENT OF PARAMETERS FOR CHICKEN TIM 167

Refinement of the Derivative Scale Factors 169

Heavy Atom Sites in the BAKER Derivative 172

G) CALCULATION OF ELECTRON DENSITY MAPS FOR CHICKEN TIM 174

facing page 144

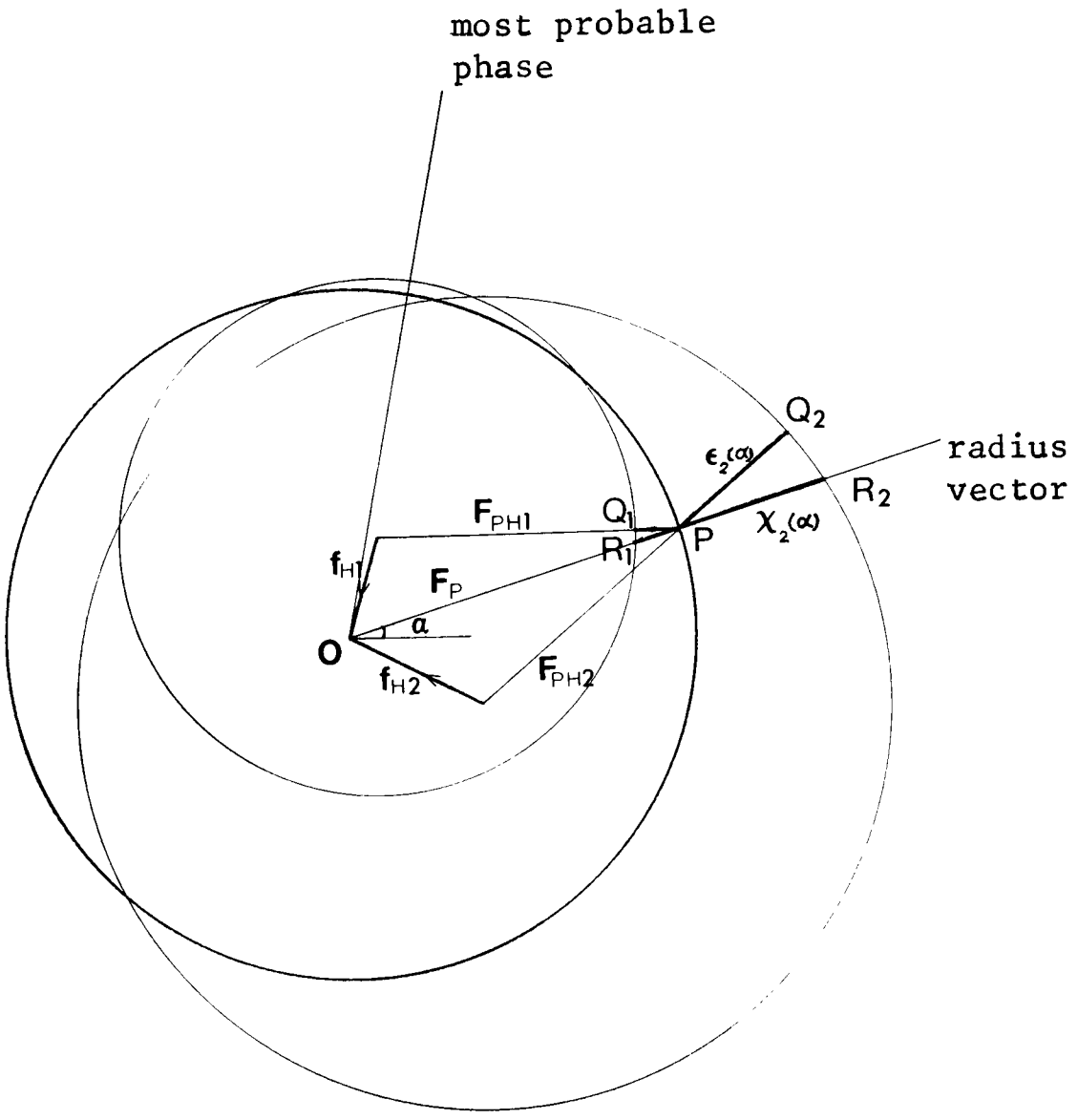


FIGURE 38 Phase Diagram Showing the Harker Construction. For the i^{th} derivative and protein phase angle α PQ_i represents the lack of closure $\epsilon_i(\alpha)$ of Blow and Crick (1959) and PR_i the radius vector $\chi_i(\alpha)$ used by Cullis et al. (1961).

A) METHODS OF PHASE DETERMINATION

Phase calculation by computer techniques has now superseded the use of graphical methods, but the analysis of the effects of errors in the isomorphous replacement method by Blow and Crick (1959) can be most easily considered in terms of the graphical construction suggested by Harker (1956) which is illustrated in Figure 38. The phase circle for the protein, which has its centre at the origin and radius equal to the structure amplitude, intersects the circle for an isomorphous derivative in two places. Ideally this ambiguity can be completely resolved by the use of a second derivative having different heavy atom structure factors. However, in practice the series of circles usually do not all intersect at a single point because of errors in the observed structure amplitude, the calculated heavy atom model and lack of perfect isomorphism. In this situation the phase of the protein structure factor can not be precisely defined, but possible phase angles can be ascribed different probabilities according to the corresponding lack of closures of the phase triangle.

The lack of closure E for any reflection has components ϵ, arising from inadequacies in the model assumed for the heavy atom substitution, and σ, arising from the error in the observed isomorphous difference. These will be related as

$$\langle E^2 \rangle = \langle \epsilon^2 \rangle + \langle \sigma^2 \rangle$$

The average value of σ^2 can be estimated from the data, and for centrosymmetric reflections the average value of E^2 can also be determined. However, for acentric reflections the lack of closure can be taken as the magnitude of PQ in Figure 38 if the error is

thought to lie in the derivative amplitude, or PR if in the protein amplitude. Correlation of the information from different derivatives becomes complicated if the error is ascribed to the protein amplitude and so for simplicity Blow and Crick assumed all the errors to lie in the derivative amplitude, and that the errors for each derivative are not correlated. They showed that a Gaussian distribution of the errors in F_{PH}, which is consistent with their observations, determines the form of the phase probability distribution as

$$P(\alpha) \propto \exp(-\phi(\alpha)^2/2E^2)$$

where P is the probability of the phase angle α being correct, and E is the RMS lack of closure $\phi(\alpha)$ which equals the magnitude of PQ as determined for the appropriate phase angle. Information from j derivatives can be combined as

$$P(\alpha) \propto \Pi_j \exp(-\phi_j(\alpha)^2/2E_j^2)$$

since the phase probability distributions of each derivative are assumed to be independent.

Blow and Crick showed that the RMS error in the electron density map of a protein is a minimum when the centroid of the phase probability distribution is used for the phase angle, if there is assumed to be no error in the structure amplitude. Weighting the structure amplitude down, by the radius of gyration of the phase probability distribution for that reflection, minimises the error in the electron density map. They showed that if the probability of a phase being correct is less than 75% it is better to omit the reflection from the synthesis than to include it with its full weight. This analysis led Blow and Crick to distinguish between the most

probable phase for a structure factor and what they termed the best phase. This is the centroid of the distribution, since it gives rise to the best Fourier, that is the synthesis having minimum RMS error. Dickerson, Kendrew and Strandberg (1961) defined the figure of merit m for a reflection as the mean value of the cosine of the error in phase angle and showed that this quantity was equal to the radius of gyration of the phase probability distribution. The coefficients used in a best Fourier synthesis may thus be described by their polar coordinates (mF_p, α_p) where α_p denotes the best phase angle. These authors related the figure of merit to the RMS error in the electron density of the best Fourier synthesis, showing that the result

$$\langle \Delta\rho^2 \rangle = {}^{2}/_{V^2} \cdot \sum_{hkl} F_{hkl}^2 \left(1 - m_{hkl}^2\right)$$

was equivalent to that previously obtained by Blow and Crick.

The methods proposed by Blow and Crick for analysis of the errors involved in the isomorphous replacement method have formed the basis of those used in most of the protein structure analyses. Aspects of their treatment which are inherently undesirable include the undue reliance placed on the accuracy of the protein structure amplitude, the difficulty of combining phase information obtained in this way with that from other sources or an additional derivative, and the assumption that the probability distributions of the derivatives are independent. Avoidance of these problems is possible by the methods outlined below.

Errors in the structure amplitudes of the protein were considered in the analysis of haemoglobin (Cullis et al., 1961). The lack of closure for a reflection was defined in terms of the

points of intersection $\mathbf{x}_i$ of the phase circles with the radius vector OR_1PR_2 in Figure 38 as $\phi_i = (x_i - \bar{x})^2$.

This RMS deviation was evaluated for the protein and derivative circles. Initially an overall value for the RMS lack of closure E was chosen, but the use of separate values E_i allows relative weights to be attached to the protein and derivative amplitudes in the expression for the phase probability distribution. The appropriate weighting for determining the mean radius vector $\bar{\mathbf{x}}$ is correlated with the values of E_i, whose determination however is more complicated than that of the corresponding quantities in the method of Blow and Crick since estimates of E from the centrosymmetric projections of a single derivative necessarily also involve the contribution of E value for the protein (H. Muirhead, see Phillips, 1966).

Inclusion of information from an additional derivative by this method requires all the previous calculations to be repeated, since the lack of closures change on determination of a new mean radius vector. The analysis of the Blow and Crick also must be repeated since the determination of the best phase angle

$$e^{i\alpha_{BEST}} \propto \int_0^{2\pi} e^{i\alpha} \cdot \Pi_j \, e^{-\phi_j^2/2E_j^2} \, d\alpha$$

does not readily allow the form of the probability distribution to be stored for further use. By assuming a Gaussian distribution of errors in the intensity rather than in the amplitude of the derivative structure factor Hendrickson and Lattman (1971) showed that the probability distribution could be formulated as

$$P(\alpha) \propto \exp(A.\cos\alpha + B.\sin\alpha + C.\cos 2\alpha + D.\sin 2\alpha).$$

The form of this distribution may thus be completely specified by four parameters and additional information incorporated by summation of the corresponding coefficients. This applies to independent phase probability distributions which may arise from direct methods, non-crystallographic symmetry, partial structure information, anomalous scattering or further isomorphous derivatives.

The correlations which do exist between the probability distributions arising from different isomorphous derivatives must not be very important in practice, since so many protein structures have now been successfully interpreted from maps in which phases determined by these methods were used. The assumption of independent distributions, which is not valid, can be avoided by a method which treats all the phase circles equally, but determines the phase angles by Fourier transform methods as has been proposed by Moffat (1969). However, his method has not yet been developed in computer programs or applied in a protein structure analysis.

Use of Anomalous Scattering Information

The information available in the anomalous scattering of X-rays was shown by Bijvoet (1951) to be useful in phase determination. The combination of the phase predictions from anomalous scattering of a derivative with those from isomorphous replacement has been discussed by Blow and Rossman (1961), North (1965) and Matthews (1966a). The calculated anomalous scattering component for any phase angle α can be seen from Figure 37 to be

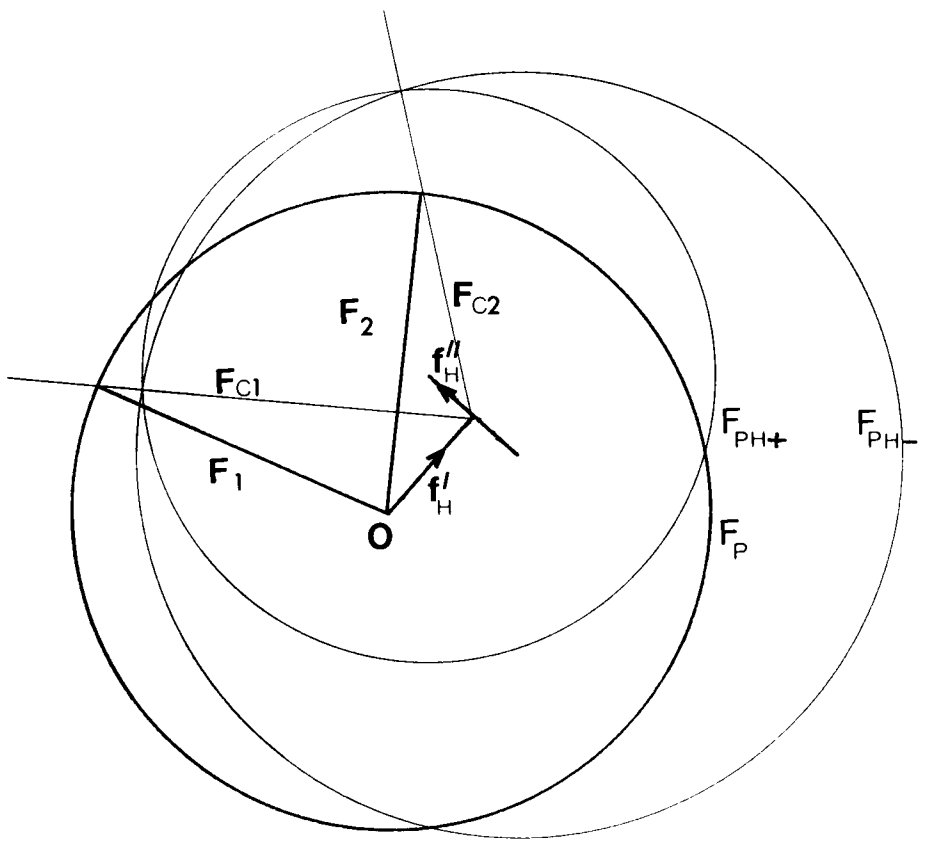


FIGURE 39 Phase Diagram Illustrating Use of Anomalous Scattering Information.
The diagram uses the notation of Matthews (1966a) showing the most probable phases of the derivative amplitude F_c as F_{c1} and F_{c2} from which the most probable phases of the protein are seen to be F_1 and F_2.

$$\Delta F_{calc} = -2 f_H'' \sin\gamma$$
$$= -2 (F_p f_H'' / (F_{PH} f_H))(b_H \cos\alpha - a_H \sin\alpha)$$

where a_H b_H are the real and imaginary components of the heavy atom structure factor f_H and so the lack of closure for the phase determination from anomalous scattering data is

$$\phi_j'(\alpha) = (F_{PH+} - F_{PH-}) - \Delta F_{calc} .$$

North suggests that the value of F_{PH} used in the determination of ΔF_{calc} may be taken as the constant observed value

$$F_{PH\ obs} = (F_{PH+} + F_{PH-})/2$$

but does not distinguish between the probability distributions resulting from the use of this or the calculated value

$$F_{PH\ calc} = F_p . \cos\beta + f_H . \cos\gamma .$$

Matthews shows that a distribution of errors similar to that proposed by Blow and Crick is assumed by a choice of the calculated derivative structure amplitude as illustrated by the construction of Figure 39.

The RMS lack of closure E' for the anomalous scattering contribution is used in an expression for the probability distribution

$$P_{anom}(\alpha) \propto \prod_j \exp(-\phi_j'^2 / 2E_j'^2)$$

analogous to that resulting from the isomorphous differences. The

facing page 150

TABLE 20 Preliminary Phase Calculation.
The parameters specified for this calculation and the resulting figures of merit m and RMS lack of closure errors E and E' are tabulated. The values of the mean residual error MRE are shown separately for the isomorphous and anomalous errors and for those for which one or two members of a Bijvoet pair of reflections were measured for the derivatives.

Resolution	Reference	m	Overall	Acentric	Centric	MRE(Iso)
∞ - 6 Å	(Phases 15-)		.836	.799	.932	.66
6 - 2.5 Å	(Phases 20-)		.616	.608	.791	.55

∞ - 6 Å Phase Calculation (15-)			PCMB	EMP	BAKER	$PtCl_4$
Input parameters:	Scale		1.0023	1.0199	0.9939	0.9881
	E		20.27	24.10	18.90	19.30
	E'		16.54	17.56	—	—
Average errors: (Acen. Centric)	MRE	Iso(1)	.43 .65	.42 .85	.47 .71	0 .91
		Iso(2)	.56 1.69	.51 1.32	0 0	0 1.00
		Anom	.65	.67	0	0
	RMS E	acent.	20.59	24.12	18.39	0
		centric	30.98	40.75	22.46	27.60
	RMS E'		18.81	20.21	—	—

6 - 2.5 Å Phase Calculation (20-)			MONO	EMP	BAKER
Input parameters:	Scale		0.9774	0.9794	0.9528
	E		27.4	25.5	21.8
	E'		27.8	26.9	26.6
Average errors: (Acen. Centric)	MRE	Iso(1)	.38 .46	.40 .62	.54 .64
		Iso(2)	.55 1.00	.53 1.02	.52 .83
		Anom	.40	.45	.43
	RMS E	acent.	28.5	26.1	22.8
		centric	35.5	36.8	30.2
	RMS E' *	b	14.50	12.92	11.53
		a	+.923	+.995	+1.003

* linear regression analysis of $E' = a.(250.\mathrm{Sin}^2\theta/\lambda^2) + b.$

value of E' may be estimated from the discrepancy between measurements of equivalent reflections in the centrosymmetric zones.

B) DETERMINATION OF PHASES FOR CHICKEN TIM

The methods of Blow and Crick (1959) and North (1965) described in the previous section were employed for determining the phases of the protein structure factors from isomorphous and anomalous differences respectively. The derivative scale factors and heavy atom parameters listed in Tables 6 and 17 for the 6Å and 2.5Å resolution sets of reflections were used, having been determined by refinements considering only the centric reflections. Initially the RMS lack of closures E_j, as determined in these refinements, and E_j', as determined from the centric reflections $\langle (F_{PH+} - F_{PH-}) \rangle$, were specified but analysis of these phase determinations allowed more appropriate values to be used subsequently. The choice of E_j, E_j' for each derivative determines the relative importance of the derivatives and the anomalous and isomorphous differences. Since the breadth of the probability distribution determines the position of its centroid, this choice also influences the relative weighting of the structure factors in the symthesis of the best electron density map.

Calculation of a set of phases allows the RMS values of E, E' resulting from the phase probability distribution to be determined. The values obtained are shown in Table 20 with the results of a further phase determination for which they were the parameters. The anomalous scattering ratio was specified as the theoretical value -8.3, the negative sign indicating that the heavy atom coordinates

used represented an inversion of the true structure. This became clear when phase calculations for the low resolution measurements in which the ratio was specified as being positive, or the anomalous scattering not considered, gave average figures of merit which were 2% and 4% lower respectively than that using the negative ratio.

The figure of merit decreases fairly uniformly with increasing $\mathrm{Sin}^2\theta$. The average value for the centric reflections is consistently higher than that for the acentric ones as would be expected since the sign probability distribution in the former case is more precise than the acentric phase distribution. Analysis of the lack of closure errors indicates that different values of E are appropriate for the acentric and centric distributions.

Mean Residual Errors

The mean residual error (MRE) was suggested by Dickerson, Weinzierl and Palmer (1968) as an indication of the extent to which appropriate E values have been used. They define an overall value

$$MRE_{iso} = \langle \phi_j^2(\alpha_\beta) / 2 E_j^2 \rangle$$

where the lack of closure is that for the best phase and the averages are taken over all reflections and each isomorphous derivative. The overall value of this quantity and also the individual values of the MRE for each derivative are listed in Table 20, with the analogous quantities for the anomalous differences

$$MRE_{anom} = \langle \phi_j'^2(\alpha_\beta) / 2 E_j'^2 \rangle$$

as suggested by North and Phillips, 1969. The isomorphous MRE was determined for acentric and centric reflections separately and also

for those where one or both of the structure amplitudes F_{PH+} and F_{PH-} had been measured.

The phase program which was used reduces the value of E to $E/\sqrt{2}$ for those reflections having measurements of both amplitudes. Since the data files do not contain F_{PH+} and F_{PH-}, but only F_{PH} the weighted mean amplitude and ΔF_{+-} the anomalous difference, the number of measurements is determined by whether ΔF_{+-} has a zero or positive standard deviation, these clearly corresponding to one and two measurements respectively. The validity of such variation of the E values depends upon the distribution of errors assumed by Blow and Crick in their analysis. The use of appropriate E values is reflected by a value of 0.5 for the MRE.

Inspection of Table 20 indicates that the overall E values used for the 6-2.5Å reflections were essentially correct (MRE=0.55) but that the ratio of the E values appropriate for one or two measurements should be slightly less than $\sqrt{2}$ for the acentric reflections. The values of E' used were all too large. For the centric reflections the appropriate E values should be larger than for the acentric, especially for those reflections having two measurements where the use of $E/\sqrt{2}$ appears to be less valid than for the acentric reflections. Similar conclusions are drawn from the results for the low resolution measurements but in this case the overall value of E and the individual values of E' were clearly too small, so that the figures of merit for these reflections are too large. Their contribution to a 2.5Å resolution Fourier synthesis will thus be weighted too heavily relative to that of the higher resolution reflections.

The expected ratio between the E values for acentric and

centric reflections has not been considered theoretically, but following the treatment of Blow and Crick it is clear that the only component of E which should depend on whether one or two measurements of F_{PH} are made, is that part which is dependent on observational errors and not on errors in the heavy atom model or lack of isomorphism. A more appropriate treatment than the use of E and $E/\sqrt{2}$ would thus be the use of the standard deviation σ of the observed isomorphous difference and the residual error so that for each reflection

$$E^2_{hkl} = \epsilon^2 + \sigma^2_{hkl}$$

where ϵ is that lack of closure error which is not due to inaccurate measurements and

$$\epsilon^2 = \langle E^2_{hkl} \rangle - \langle \sigma^2_{hkl} \rangle .$$

Preliminary investigations with the chicken TIM measurements to 2.5Å resolution indicate that $\langle \epsilon^2 \rangle$ defined in this way is not constant but varies with both E and σ, so that this procedure can not usefully be incorporated into the phase determination in this case.

The RMS value of E calculated for any derivative using the determined set of best phases represents a compromise between the values appropriate for the n_1 reflections having one measurement and the n_2 having both measurements of F_{PH+} and F_{PH-}. This value of E was thus multiplied within the phase program by the constant

$$(n_1 + n_2) / (n_1 + n_2/\sqrt{2})$$

for each derivative set of measurements to give the value of E appropriate for a single measurement. Analysis of the RMS values showed that E is essentially constant with respect to $Sin^2\theta$ but E'

increases linearly, reflecting the increasing inaccuracy of the anomalous measurements at higher resolution. For subsequent phase calculations E′ was specified as being a linear function of $Sin^2\theta$.

The calculated phases and figures of merit were used for Fourier syntheses having the following coefficients:

best electron density map $m F_p e^{i\alpha_p}$

difference syntheses $\Delta F = m(F_{pH} - F_p) e^{i\alpha_p}$

double difference syntheses $\Delta\Delta F = m E e^{i\alpha_{pH\,calc}}$
$= m\left(F_{pH} - |(\underline{F}_p + \underline{f}_H)|\right) e^{i(\alpha_p+\beta)}$

For the low resolution synthesis the ΔΔF maps are featureless for each derivative and the ΔF maps contain significant peaks only at the known heavy atom sites, as discussed in Chapter III, providing no indication of any systematic errors in the electron density map of the protein. Several features of this map are discussed in detail in Chapter VI. The higher resolution syntheses showed significant features in the ΔΔF maps as mentioned in the preceding chapter and large regions of negative density in the protein map at the sites of heavy atom binding. The greatest depth of the holes represents about 3 electrons at the sulphydryl sites and 2 electrons at the additional two EMP sites. Such systematic features in the protein map must result from errors in the calculated phases. The derivative scale factor is the parameter which can exert the most obvious influence (Blow, Rossman and Jeffrey, 1964) in biasing the phases to produce such features, but the possible effects of errors in the other heavy atom parameters are also considered in the following section.

C) EFFECT OF ERRORS IN THE HEAVY ATOM PARAMETERS

The parameters for a single heavy atom derivative are correlated with each other and with those for other derivatives, since initial phases are necessary for the refinement of parameters, and vice versa, for acentric reflections which form the subject of this discussion. A rigorous treatment of these correlations has not been attempted, this analysis being initially restricted to consideration of an error in a single parameter for one heavy atom derivative, all other errors, both observational and in the heavy atom model, being assumed to be negligible. One justification for such a simplified treatment is the possibility of investigating the parameters of the Nth derivative with respect to phases determined by a separate set of N-1 derivatives, a procedure which does eliminate some of the undesirable correlations. Difference and double difference maps calculated for the Nth derivative using these phases (α_p^{N-1}) reflect errors in its parameters differently from those using the best phases determined by all the derivatives (α_p^N). Difference maps for any of the N-1 derivatives will reflect errors in the parameters of the Nth derivative only if the phases α_p^N to which it contributed are used.

The correlation between the scale factor k and the heavy atom occupancy z for a derivative can be seen by including these parameters implicitly in the relationship of the phase triangle (Figure 37)

$$k^2 F_{PH}^2 = F_P^2 + z^2 f_H^2 + 2 \underline{F_P} . z \underline{f_H} . \quad (1)$$

The assumption that only a single occupancy factor is involved, as in a single site derivative, does not affect the underlying basis of

this analysis. Assuming that the protein and heavy atom structure factors are uncorrelated, which is equivalent to assuming the validity of Wilson statistics, implies

$$k^2 \langle F_{PH}^2 \rangle = \langle F_P^2 \rangle + Z^2 \langle f_H^2 \rangle \qquad (2)$$

so that the occupancy and scale factor will tend to increase or decrease together. If this relation is valid then clearly the derivative and heavy atom structures must be correlated, however loosely, since otherwise the phase triangle relationship

$$F_P^2 = k^2 F_{PH}^2 + Z^2 f^2 - 2k \underline{F_{PH}} \cdot Z \underline{f_H} \qquad (3)$$

would imply equality between

$$k^2 \langle F_{PH}^2 \rangle \quad \text{and} \quad \langle F_P^2 \rangle - Z^2 \langle f_H^2 \rangle$$

which can not be true if (2) is true, but the difference between them must be fairly small.

If the scale factor k^N for a derivative is assumed to be larger than its true value, then the angle ϕ in the phase triangle for every reflection will be decreased, so that the phase determined for the protein structure factor α_P^N will be distorted from its true value towards the value of α_H. The coefficient in the Fourier synthesis may then be regarded as $(F_P e^{i\alpha_P} + \delta f_H e^{i\alpha_H})$ where δ is a small positive amplitude. The magnitude of the erroneous component need not be related to the amplitude of f_H for the reflection, but as their phase angles are equal, the Fourier synthesis will give peaks at the sites of the heavy atom structure in addition to an image of the protein structure.

Difference maps calculated for this derivative using

TABLE 21 Effects of Errors in the Heavy Atom Parameters. The features expected in difference, double difference and protein Fourier maps when the scale factor or occupancy parameters are incorrect for the N^{th} derivative are summarised, the reasons for these features being discussed in the text. The features will be localised at the sites of the heavy atoms in this N^{th} derivative.

Parameter which is too large for the N^{th} derivative	Scale Factor	Occupancy
Phases calculated with N-1 derivatives;		
F_P	–	–
$(\Delta F)_N$	–	–
$(\Delta\Delta F)_N$	–	hole
Phases calculated with N derivatives;		
F_P	peak	hole
$(\Delta F)_{N-1}$	very small peak	very small hole
$(\Delta\Delta F)_{N-1}$	–	–
$(\Delta F)_N$	large peak	smaller peak
$(\Delta\Delta F)_N$	peak	hole

coefficients

$$\left(k^N F_{PH}^N - F_P\right) e^{i\alpha_P^N}$$

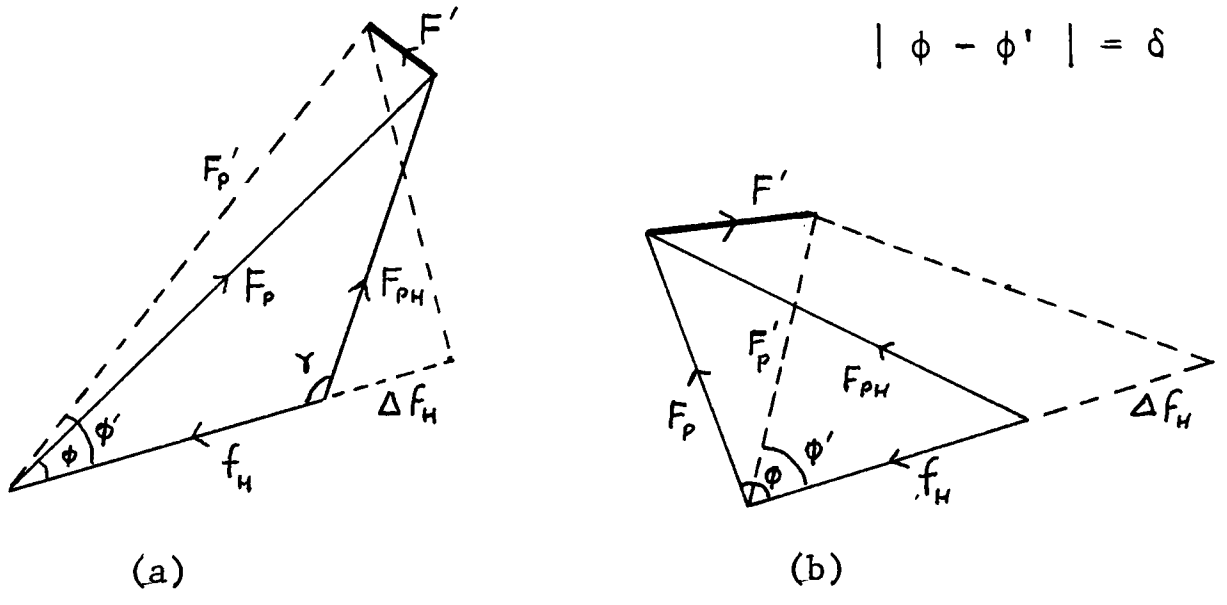
will show excessively large peaks at the heavy atom sites since the average value of the amplitude is positive and the phase angle contains a component in the direction of α_H. Similarly the double difference maps

$$\left(k^N F_{PH}^N - |(F_P e^{i\alpha_P^N} + Zf_H e^{i\alpha_H})|\right) e^{i\alpha_{PH}^N}$$

will be expected to show peaks at the heavy atom positions since the value of α_{PH}^N will be closer to α_H than is the true value of α_{PH} and the average amplitude of the coefficients is positive.

Difference and double difference maps for any of the N-1 derivatives whose parameters are not in error will not show these features since the average amplitude of the coefficients is essentially zero and the contributions of different reflections to the density at the heavy atom sites of the Nth derivative consequently tend to cancel each other out. Any features present will be very small. Similarly few features would be expected in the difference maps calculated for the Nth derivative using the phases α_P^{N-1} to which it did not contribute, since although the coefficients of the difference syntheses have a positive amplitude on average, the phase angle has no component related to the direction of α_H so that the contributions of different reflections again tend to cancel out. These conclusions are summarised in Table 21 together with the effects expected from a heavy atom occupancy which is incorrectly assumed to be too large.

Determination of the extent of uncertainty in the derivative scale factor which is permissible before a derivative ceases to provide



$$F_{PH}^2 = F_P^2 + f_H^2 - 2F_P f_H \cos\phi$$
$$= F_P^2 + f_H^2(1 + \Delta)^2 - 2F_P f_H \cos\phi'$$

Component parrallel to f_H of the error in the structure factor F_P caused by the extra occupancy Δf_H is given by $F'\sin\phi$

$$F'\sin\phi = 2F_P \sin\phi . \sin\delta/2.$$

But $$\cos\phi - \cos\phi' = \frac{F_P^2 + f_H^2 - F_{PH}^2}{2F_P f_H} - \frac{F_P^2 + f_H^2(1+\Delta)^2 - F_{PH}^2}{2F_P f_H(1+\Delta)}$$

$$2\sin(\phi+\delta/2)\sin(\delta/2) = \frac{(-f_H^2\Delta^2 - 2f_H^2\Delta + F_P^2\Delta + \Delta f_H^2(1+\Delta)^2 - F_{PH}^2\Delta)}{2F_P f_H}$$

Thus $$4F_P f_H \sin(\phi + \delta/2)\sin(\delta/2) = \Delta(F_P^2 - F_{PH}^2 - f_H^2)$$

and assuming that δ is small then

$$F'\sin\phi \sim \Delta(F_P^2 - F_{PH}^2 - f_H^2)/(2f_H)$$

FIGURE 40 Effects of Errors in the Heavy Atom Occupancy. Increasing the occupancy modifies the protein structure factor by a component which is (a) parallel and (b) anti-parallelto the heavy atom structure factor. The analysis of the magnitude of this component is due to Professor D.C. Phillips.

useful information has been considered by Blow (1957). He showed that a fractional error ϵ in the scale factor altered the angle between the protein and heavy atom structure factors ϕ to ϕ' where

$$\cos\phi - \cos\phi' = \frac{F_{PH}^2 - F_P^2 - f_H^2}{2 F_P f_H} - \frac{F_{PH}^2(1+\epsilon)^2 - F_P^2 - f_H^2}{2 F_P f_H}$$

as can be seen from Figure 40. Thus

$$\cos\phi - \cos\phi' \sim \epsilon F_{PH} / F_P f_H$$

Denoting the RMS value of ϵ by E and considering the distribution of ϕ' to have a standard deviation σ about the mean value ϕ it is clear that

$$\langle | \cos\phi - \cos(\phi \pm \sigma) | \rangle \sim \langle F_{PH}/F_P \rangle . \langle E/f_H \rangle$$

The ratio of derivative to native structure amplitudes is essentially unity and so the usefulness of a derivative for phase determination may be considered in terms of the ratio $\langle E/f_H \rangle$. This result is applicable to all sources of error which may be regarded as being in the magnitude of the derivative structure amplitude. All errors were regarded in this way by Blow and Crick (1959).

Matthews (1969) states that the distribution of the probability of the phase determination being correct is

$$P = \tfrac{1}{2}\left(1 + \operatorname{erf}(\sqrt{2}\, f_H/E)\right)$$

for a centric reflection, where erf denotes the error function. Their experience with α-chymotrypsin led Blow and Matthews to suggest that when the ratio

$$R_{EF} = \langle E/f_H \rangle$$

and the analogous quantity for the anomalous differences

$$R'_{EF} = \langle E'/f'' \rangle$$

exceed unity the phase information is of only marginal benefit. The values of these ratios will be expected to increase with $\mathrm{Sin}^2\theta$ as the effective heavy atom occupancy decreases.

Errors in the occupancy of a heavy atom bias the protein phase determination differently for different reflections as is seen in Figure 40, whereas errors in the scale factor have been shown always to bias the phase angle in the same direction. It is seen from Figure 40 and the accompanying analysis (D.C. Phillips, personal communication) that for a fractional change Δ in the occupancy the component parallel to f_H of the error introduced into the protein structure factor is

$$\begin{aligned} F' \mathrm{Sin}\,\phi &\sim \Delta (F_P^2 - F_{PH}^2 - f_H^2)/2f_H \\ &= \Delta (-2F_{PH} f_H \mathrm{Cos}\,\gamma)/2f_H. \end{aligned}$$

The average value of the term in parentheses must be negative so that an increase in occupancy is associated with a hole in the protein electron density map. The low correlation between the derivative and heavy atom structures predicts that this average value will be small.

A very small hole will be present at these heavy atom sites in the difference maps for other derivatives, the coefficients of these syntheses having positive average amplitudes. No features would be expected in their double difference maps. The derivative for which the occupancy is too large would be expected to have smaller peaks than expected in its difference maps and holes of negative density in its double difference maps. These $\Delta\Delta F$ maps would appear to be very

similar irrespective of whether or not this derivative contributed to the phase determination.

Errors in the positional parameters xyz can not be considered in this way, except in terms of a decreased occupancy at the true position and increased occupancy at the false position. Errors in the temperature factor can also be considered in terms of changes in the distribution of occupancy in the region of the heavy atom. The remaining parameters which are specified in any phase determination are the anomalous scattering ratio, which affects the relative weighting of the contributions of the isomorphous and anomalous differences, as do the relative values of the lack of closure estimates E and E′. The absolute values of these estimates affect the distribution of the figure of merit and thus, in best Fourier syntheses, the relative weighting of different reflections, a topic which is considered in more general terms in Appendix II.

D) DETERMINATION OF DERIVATIVE SCALE FACTORS

A variety of methods has been suggested for determining the relative scale factor between native and derivative sets of data. Blow (1958) discusses experience with several of these indicating that they predict slightly different results and that none of them is entirely satisfactory. All of the methods which are now commonly used rely on the distribution of intensities being consistent with Wilson statistics (Wilson, 1942).

Calculation of the relative Wilson plots described in the previous chapter requires the heavy atom occupancy to be known, which is a serious disadvantage of that method. The scale factor and

occupancy are closely correlated, but the values calculated for chicken TIM in Table 19 showed that the scale factor is relatively insensitive to changes in the occupancy. Increasing the occupancy by 20% will result in a scale factor of 1.05 being predicted as approximately 1.06. However, experience with chicken TIM indicates that even a 1% error in the scale factor is associated with systematic features in the electron density map at the heavy atom sites.

The necessary condition for the applicability of Wilson statistics to these intensity distributions can be shown to be the presence of all the heavy atoms at positions of zero electron density in the protein (A. Klug, personal communication). If this condition is satisfied then the convolution of the electron densities of the protein and heavy atom is zero, and thus the product of the Fourier transforms of the two electron density distributions is also zero. The relationship of the phase triangle

$$F_{PH}^{\,2} = (\underline{F_P}^{\,2} + \underline{f_H}^{\,2})^2$$

$$= F_P^{\,2} + f_H^{\,2} + 2\,\underline{F_P} \cdot \underline{f_H}$$

reduces to the basic equation of Wilson statistics whenever the average value of

$$2\,\underline{F_P} \cdot \underline{f_H} = 2\,F_P\,f_H \cos\phi \qquad (1)$$

is equal to zero. This will only occur when the average is taken over enough reflections and to a sufficient resolution so that the heavy atom is clearly resolved from the electron density of the protein. At 6Å resolution the heavy atoms are unlikely to be resolved in this way and so the average value of (1) will be positive or negative depending upon whether the heavy atom site is a positive or negative

region in the truncated Fourier synthesis. At 2.5Å resolution the heavy atom probably is separately resolved and so the average value of (1) would be expected to be zero, implying that statistical methods are appropriate in this case for a determination of scale factors.

Singh and Ramaseshan (1966) have proposed an analytical method for the determination of derivative scale factors, which implicitly assumes that Wilson statistics are obeyed, but involves only the observed intensities without any knowledge of the heavy atom occupancy being necessary. They show that for all reflections

$$\Delta I = 4 F_{PH} f_H'' \sin\gamma = 4 F_P f_H \sin\phi / k'$$

$$f_H^2 = F_P^2 + F_m^2 \pm \left[4 F_m^2 . F_P^2 - k'^2 \Delta I^2 / 4 \right]^{1/2}$$

where $$F_m^2 = \left(F_{PH+}^2 + F_{PH-}^2 \right) / 2$$

and $$\Delta I = F_{PH+}^2 - F_{PH-}^2$$

although the difference between F_m and F_{PH} is shown in Appendix III not to be significant. Substitution of this expression into the equation of the phase triangle, with explicit consideration of the scale factors S_P , S_{PH} by which the protein and derivative amplitudes should be multiplied, and consideration of averages within the set of reflections gives their formula

$$S_A^2 = \left(S_P / S_{PH} \right)^2 = \frac{\langle F_m^2 . F_P^2 \rangle \mp \left\{ \langle F_m^2 . F_P^2 \rangle^2 \quad k'^2 \langle \Delta I^2 \rangle \langle F_P^4 \rangle / 2 \right\}^{1/2}}{2 \langle F_P^4 \rangle}$$

by which the relative scale factors may be directly determined. The derivation of this formula is considered in detail in Appendix III where it is shown that an alternative method of averaging between reflections produces the result

facing page 163

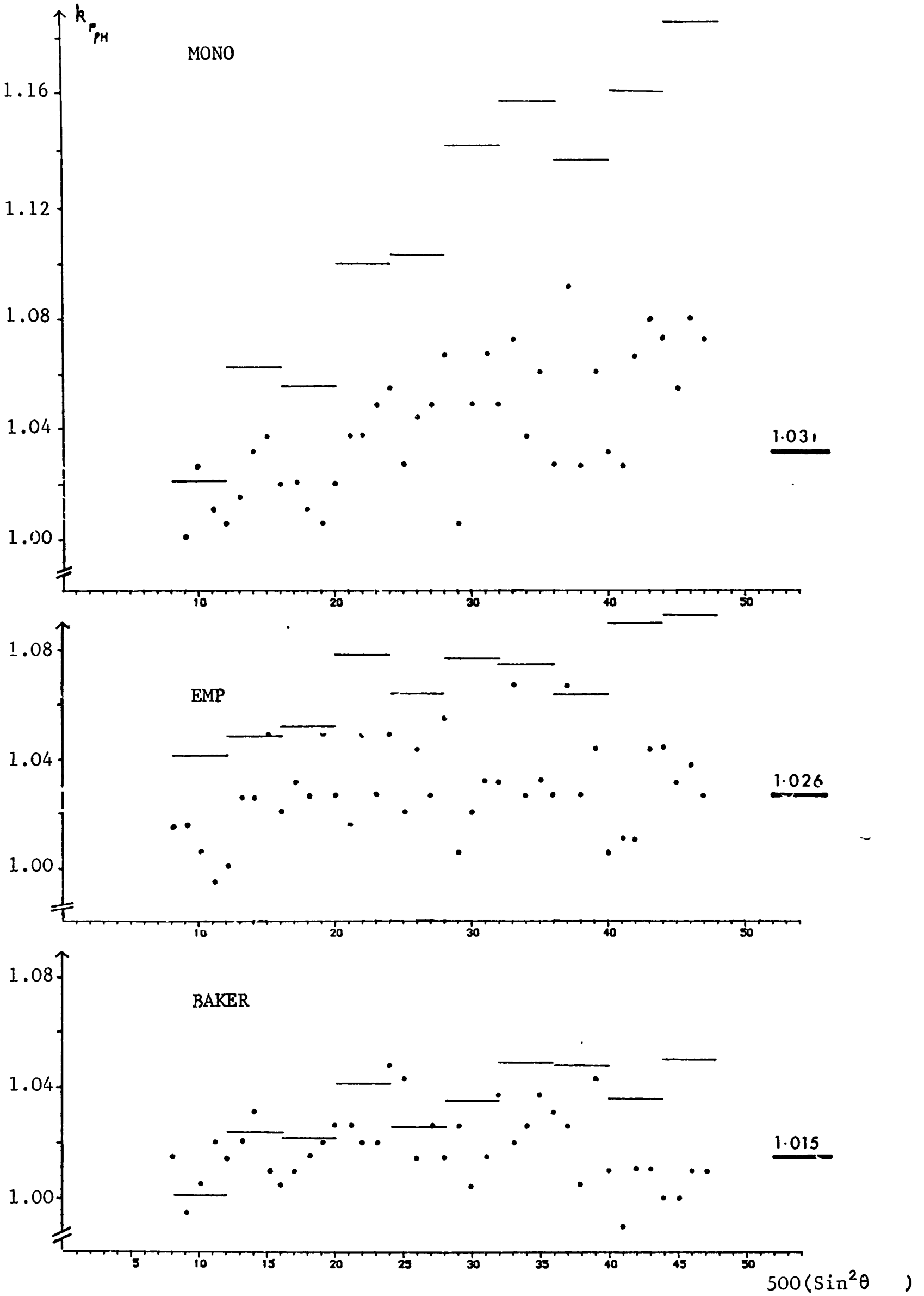


FIGURE 41 Derivative Scale Factors Determined By the Method of Singh and Ramaseshan. For each derivative the values of $k_{F_{PH}}$ determined by the published expression (—) and a modification of it (•) described in the text and Appendix III are shown as functions of ($Sin^2\theta$), the overall value from the modified expression being shown (▬).

$$S_R'^2 = \left[1 \mp \left\{1 - k'^2 \Delta F^2 / F_{PH}^2\right\}^{1/2}\right](F_{PH}^2/F_P^2)/2.$$

Evaluation of these expressions for the chicken TIM data gave the results shown in Figure 41.

A method suggested by Kraut (see Dickerson et al., 1967) has been used by Dickerson et al. for centrosymmetric projections and by Arnone et al. (1971) for three dimensional data. This is based on the equality expected in the origin peaks in centrosymmetric projections of the ΔI and $(\Delta F)^2$ Patterson functions, which represent only the heavy atom self vectors in each case. For centric reflections, assuming no correlation between protein and heavy atom

$$F_{PH} = kF_P \pm f_H$$

$$\langle k F_P \rangle = \langle F_{PH} \rangle$$

and thus the required scale factor for the derivative amplitudes is

$$k = \langle F_P^2 \rangle / \langle F_P . F_{PH} \rangle$$

For acentric reflections the $(\Delta F)^2$ Patterson includes the heavy atom vectors at only half weight (Phillips, 1966) and so a general application of this method requires separate determination of the heavy atom structure amplitudes. Arnone et al. use an expression for f_H^2 which is analogous to that given in the previous section. Implicit consideration of the derivative scale factor in this equation allows it to be refined by an iterative procedure. The assumptions underlying this method are the same as those of Singh and Rameseshan, the two methods differing in the manner of determining the averages. This is discussed further in Appendix III.

A modification of this Kraut method is based on the statistical

TABLE 22 Derivative Scale Factors.
The results of a variety of methods referred to in the text which have been applied to the derivatives of chicken TIM for the 6-2.5Å reflections are summarised.
Scale factors calculated as a function of $(\sin^2\theta/\lambda^2)$ are expressed by the linear regression parameters $k_{F_{PH}} = b + a.500(\sin^2\theta/\lambda^2)$.

		MONO	EMP	BAKER
Centric refinement		0.9774	0.9794	0.9528
3-D Refined from preliminary phase calculation	b	0.9954	1.0089	1.0092
	a	-.0007	-.0012	-.0011
Singh and Ramaseshan	b	1.0186	1.0410	1.0103
	a	+.0034	+.0010	+.0008
Modified Singh and Ramaseshan	b	0.996	1.107	1.021
	a	+.0016	+.0004	-.0001
Wilson statistics	b	1.033	1.036	1.015
	a	+.0004	+.0004	+.0002
Kraut				
Modified Kraut	acentric	1.066	1.060	1.034
	centric	1.025	1.045	1.030
Scale Factors used in phase calculation (30-) for electron density map		1.0419	1.0497	1.0316
3-D Refined from these phases (30-)	b	1.0364	1.0512	1.0418
	a	-.0012	-.0016	-.0013

relation for acentric reflections

$$\langle F_{PH}^2 - F_P^2 \rangle = 2\langle (F_{PH} - F_P)^2 \rangle$$

The scale factor will influence the left hand side of this equation to a much greater extent than the right hand side and thus if

$$\langle F_{PH}^2 \rangle = k^2 \langle F_P^2 \rangle$$

then to a first approximation the scale factor may be evaluated from

$$k^2 = \frac{2\langle (F_{PH} - F_P)^2 \rangle}{\langle F_P^2 \rangle} + 1$$

by considering either the direct ratio or a linear regression analysis. Preliminary calculations of the scale factor by this method indicate that it gives useful results (Table 22). This modification has the advantage of not requiring any measurements of the anomalous differences.

Methods of determining derivative scale factors which do not assume the validity of Wilson statistics must involve knowledge of the angles in the phase triangle, and thus of the protein phases. The scale factors determined from the centrosymmetric projections of chicken TIM are clearly inapplicable to the acentric set of reflections, because of the large holes already described in the resulting electron density map. The normal methods of refining any of the heavy atom parameters for acentric reflections, using estimated phases, are considered below and shown to be not readily applicable to the derivatives of chicken TIM, since these all contain the two common sulphydryl sites of mercury binding.

E) METHODS OF REFINING HEAVY ATOM PARAMETERS

Methods of refinement which are applicable only to centric reflections were discussed in Chapter III and so the present summary is restricted to those methods which are also applicable to acentric data. These are based on minimising the discrepancy between either the observed and calculated heavy atom structure amplitudes for a single derivative with anomalous scattering measurements or the observed and calculated derivative structure amplitudes using estimated protein phases.

The first method is discussed by Dodson and Vijayan (1971) who use the equation analogous to that given in the preceding section

$$f_H^2 = F_P^2 + F_{PH}^2 \pm 2 F_{PH} F_P \left[1 - (k\Delta F / 2F_P)^2 \right]^{1/2}$$

to obtain upper and lower estimates of the heavy atom structure amplitude denoted f_{HUE} and f_{HLE}. The upper estimate can only be the correct value of f_H if it is less than the maximum possible value of f_H. Excluding those reflections where this is true, the value of f_{HLE} can be identified with f_{Hobs}. Calculation of the heavy atom structure factor in terms of parameters specifying the occupancy, temperature factor and coordinates of each site enables these parameters to be refined by a least squares procedure minimising

$$\left(f_{H\,obs} - f_{H\,calc} \right)^2$$

as a function of all of the parameters. However, the derivative scale factor is involved only in the calculation of f_{Hobs} and so can not be refined by this method.

Refinement methods using preliminary phases for the protein structure factor are reviewed by Dickerson, Weinzierl and Palmer

(1968). The quantity minimised by a least square procedure is

$$E^2 = \left(k F_{PH} - (F_p \cos\beta + f_H \cos\gamma) \right)^2$$

where the angles β and γ in the phase triangle are determined by the phase angles α_p and α_H. In this case the derivative scale factor can be refined in the same way as any other parameter. The earliest applications of this method involved alternate cycles of phase determination and least squares refinement, but it is computationally more efficient to repeat the refinement cycles until convergence is reached before determining another set of phases, and this procedure is now usually adopted.

The undesirability of refining parameters against a set of phases, which have been partly determined by these parameters, led to a further extension of this method. Each derivative can be excluded from the phase determination in turn and the parameters for this Nth derivative refined against the set of phases determined by the remaining N-1 derivatives. The disadvantages, which remain even in this modification of the phase and refinement procedure are most easily visualised by consideration of the properties of double difference maps, since the lack of closure error which is being minimised is identical to the amplitude of the coefficient in a $\Delta\Delta F$ map. Thus least squares refinement in reciprocal space is equivalent to minimising the features in a $\Delta\Delta F$ map.

The features to be expected in double difference maps when an incorrect value is assumed for a heavy atom parameter are summarised in Table 21. Combinations of these features will clearly result if several parameters, involving more than one derivative, are incorrect. The existence of common sites in the derivatives will make it impossible

to distinguish an incorrect parameter for the Nth derivative, which is being refined, from one for any of the N-1 derivatives, which were used for the phase determination. Experience of this problem for the derivatives of chicken TIM is discussed in the following section.

The most successful approach to determining parameters for the common sites was found to be the inspection in the region of the Nth heavy atom site of $\Delta\Delta F$ maps which were based on protein phases to which this site had not contributed in any of the derivatives. The use of only N-1 sites to determine the phases, which was suggested by Professor D.C. Phillips, inevitably increases the lack of closure error for any derivative in which the Nth site is substituted, but the errors in these protein phases did not produce any systematic features in the protein density map at the Nth atomic site. This indicates that occupancy errors can introduce only small systematic errors into the protein phases, since the complete exclusion of a site may be regarded as a gross under-estimation of its heavy atom occupancy. The lack of features in the protein density map at this Nth heavy atom site allows the features at this position in the ΔF and $\Delta\Delta F$ maps to be attributed entirely to the heavy atom at that site and not to the protein phases.

F) FURTHER REFINEMENT OF PARAMETERS FOR CHICKEN TIM

The features observed in the difference and double difference maps calculated using the preliminary set of phases for the 6-2.5Å reflections can be explained in the light of the results discussed earlier in this chapter. Since all the scale factors were too small, peaks were observed in the $\Delta\Delta F$ maps at all the heavy atom sites used

in the phase calculation. These arise since the average value of ($kF_{PH\,obs} - F_{PH\,calc}$) is negative if the scale factor k is too small and the bias in the phases introduces a component which is anti-parallel to f_H in the protein structure factor. The third and fourth sites of the EMP and BAKER derivatives were the least affected, each of these sites being present in only one of the three isomorphous derivatives. However, any set of phases to whose determination these EMP sites contributed, gave rise to peaks which appeared to be significant in the ΔF and $\Delta\Delta F$ maps for the MONO derivative. Inclusion of these sites in centric refinements of this derivative gave reasonable values for the occupancies and temperature factors, associated with small improvements in the reliability index R and lack of closure E for the centric reflections. Difference maps for the MONO derivative, using phases to which there had been no contribution from sites in these positions, showed peaks which were not significant in comparison with those observed on $\Delta\Delta F$ maps of the EMP derivative using phases determined by the best set of available parameters.

The component of the overall residual

$$\psi = k F_{PH} - |(F_P + f_H)|$$

which contributes to these peaks in $\Delta\Delta F$ maps for the MONO derivative can not be reduced by variation of any of the parameters in a least squares refinement for this derivative, since they arise from an error in the EMP scale factor. The peaks at the sulphydryl sites arise from errors in both the MONO scale factor, which appears explicitly in the above expression for the residual, and also other scale factors, for derivatives in which these sites contributed to the phase deter-

mination. These peaks can only be interpreted by a least squares refinement of the MONO derivative as indications of occupancies which are too small. Clearly the small scale factor of this derivative will probably be associated with occupancies which are also too small and in the case of the chicken TIM data it is not known which of these two parameters was primarily poorly determined by the centric reflections.

Convergence of the phase determination and least squares refinement procedure in this situation would be expected to be very slow, irrespective of whether N or N-1 derivatives were used for calculating the protein phases. Refinement of the scale factors alone avoids the problems of some of these correlations. Since these are the parameters in which inaccuracies generate the largest systematic features in the protein electron density map, it would seem that these are the parameters of major concern in a situation such as the present analysis of chicken TIM where correlations between parameters are making a more general refinement extremely difficult. Progress towards an acceptable set of scale factors was checked by determination of further sets of protein phases and calculation of electron density maps. Throughout these further refinements the protein, ΔF and $\Delta\Delta F$ Fourier syntheses were calculated only for the regions of space surrounding the heavy atom sites. The complete asymmetric unit, which is one quarter of the unit cell, was calculated only for the first ΔF and $\Delta\Delta F$ maps when the possibility of there being further minor binding sites for heavy atoms was thoroughly investigated.

Refinement of the Derivative Scale Factors

Refinement of the derivative scale factor whilst keeping all

the other parameters constant can be readily achieved since a least squares minimisation of

$$\left(k F_{PH\ obs} - F_{PH\ calc}\right)^2$$

is equivalent to a linear regression analysis and the required scale factor is

$$k = \langle F_{PH\ obs} \cdot F_{PH\ calc} \rangle / \langle F_{PH\ obs}^2 \rangle$$

where $F_{PH\ calc}$ is determined from the protein structure factor, using its estimated phase, and the calculated heavy atom parameter, using estimates of all the heavy atom parameters except the scale factor. This method is equally valid for both centric and acentric reflections if the structure factors are regarded as real and complex quantities respectively and so was applied to the 6-2.5Å resolution reflections of chicken TIM without discrimination between these two categories.

Determination of the variation of the scale factor with respect to $Sin^2\theta$ showed that it always decreased by 3-4% between the limits of the resolution range considered. This negative slope must be attributed to an accumulation of errors since the scale factor calculations using the methods of Wilson statistics, Singh and Ramaseshan and the simplified Kraut procedure, which have been described above, all give rise to a positive slope. The results obtained by all these methods may be compared in Table 22.

The negative slope of the refined scale factor as a function of $Sin^2\theta$ will arise from observational errors in the intensities. The exact relationship in the phase triangle may be stated as

$$F_{PH\,calc}^2 - F_P^2 = f_H^2\left(1 - 2 F_P \cos\phi / f_H\right)$$

and in terms of the refined scale factor as

$$k^2 F_{PH\,obs}^2 - F_P^2 = f_H^2 (1 - 2F_P \cos\phi / f_H)$$

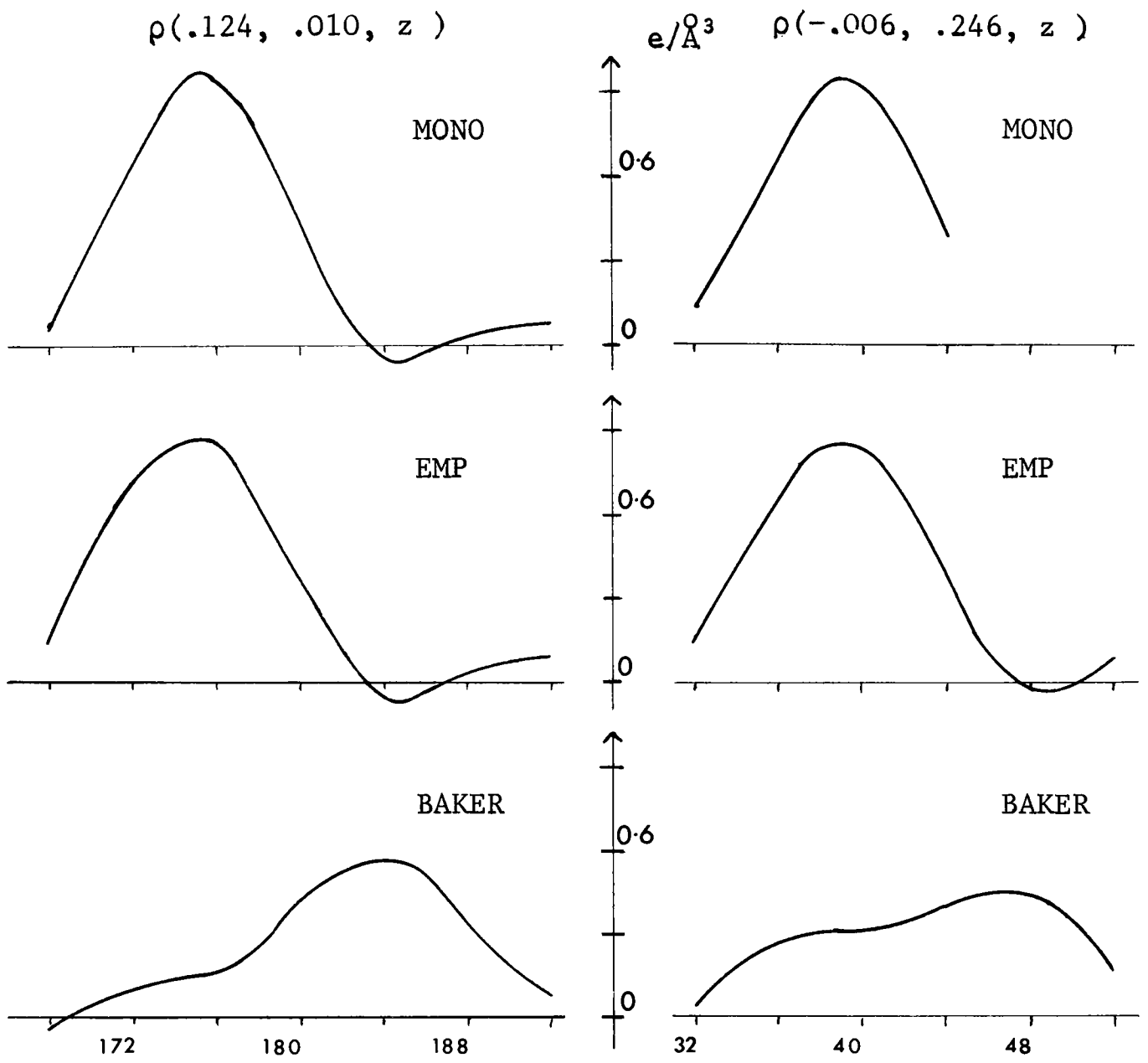
Rearrangement of the left hand side gives

$$(k^2 - 1) F_{PH\,obs}^2 + (F_{PH\,obs}^2 - F_P^2) = f_H^2 (1 - 2F_P \cos\phi / f_H)$$

Errors in the observed intensities result in the RMS value of $(F_{PH\,obs}^2 - F_P^2)$ being too large (Dodson and Vijayan, 1971) and these errors are known to increase at higher angles. If the right hand side of the equation is assumed to be constant, for a good fit to the heavy atom model, then the increasing value of $(F_{PH\,obs}^2 - F_P^2)$ must be accompanied by a decreasing value of the scale factor.

It is seen from the results in Table 22 that the 3-D refined scale factors differed only slightly from those obtained in the centric refinements and used for the phase determination. An iterative procedure was developed to refine the scale factors towards the values found to be appropriate in the protein electron density map. The heavy atom occupancies were refined, whilst fixing the scale factor at its new value, considering the centric reflections only. The phase determination was repeated using the new occupancies and scale factors, and a further refinement of the derivative scale factors was then executed. Several cycles of this procedure were necessary before appropriate scale factors were obtained.

It is now clear that this process could have been accomplished more quickly either by repeating the occupancy and scale factor refinement alternately several times for a single phase calculation, or by recognising that the bias in the protein phases would result in an artificially small change in the scale factor on refinement and



Phase Calc$^{n.}$ Reference No. (27-) (28-)

Excluding all sites at (.124, .010, ∿.75) (-.006, .246, ∿.17)

FIGURE 42 Profiles of Heavy Atoms at Sulphydryl Sites in ΔF Maps. The line profile along the z axis through the peak on the difference Fourier maps is shown for each derivative, these sites having been excluded from the phase determination.

counteracting this by doubling the change in the scale factor before using it for refining the occupancies.

The scale factors which were used for determining the phases with which a complete 2.5Å resolution electron density map of the protein was calculated are listed in Table 22. They were obtained by a combination of this method and those already described.

Heavy Atom Sites in the BAKER derivative

Determination of an appropriate set of scale factors allowed the other heavy atom parameters to be considered in more detail. No serious errors were detectable in the positions or occupancies of the heavy atom sites from difference or double difference maps, using phases to which every heavy atom site had contributed. Exclusion of the sulphydryl sites, one at a time from the phase determination, gave ΔF and $\Delta\Delta F$ maps in which the properties of these sites could be observed without interference from those of other derivatives. These maps enabled the positions of the binding sites in the BAKER derivative to be clarified.

Line profiles through the peaks at the sulphydryl sites, as they appear in these ΔF maps, are shown in Figure 42. It is clear that the first site in the BAKER derivative lies on the opposite side of the line $z = 3/4$ to the corresponding site in the other derivatives. The clearly defined shoulder on this peak was interpreted as an additional mercury binding site and refined accordingly. The occupancy of this site, denoted site 1A, is about 35% of that of site 1 for this derivative.

A more surprising observation was that the z coordinate of the second site in the BAKER derivative is also clearly different from

TABLE 23 Parameters After Further Refinement of Derivatives at 2.5Å Resolution.
The first three columns show the current set of best parameters for each derivative and the fourth column shows those parameters for the BAKER derivative which were actually used for the calculation of phases (30-) and the electron density map of the protein, which have the relative occupancies of sites 2 and 2A reversed from their true values.

Derivative		MONO	EMP	BAKER		BAKER
$\langle f_H \rangle$		37.	35.	27.		28.
Reliability Index R		0.617	0.598	0.599		0.605
RMS lack of closure E		35.31	32.70	25.90		26.14
Number of sites		2	4	6		6
*Refined scale k		1.0419	1.0497	1.0316		1.0316
Site 1	Z	.2754	.3193	.1700		.2125
	B/λ^2	4.7492	13.8687	0.5000		2.6000
	x	.1248	.1240	.1215		.1218
	y	.0103	.0112	.0082		.0078
	z	.7339	.7308	.7664		.7658
Site 2	Z	.3345	.2150	.1069		.0606
	B/λ^2	8.1062	2.9968	9.7000		8.9000
	x	-.0056	-.0066	.0029		.0037
	y	.2460	.2445	.2427		.2413
	z	.1617	.1624	.1890		.1966
Site 3	Z		.1972	.1217		.1245
	B/λ^2		8.8875	13.2625		10.0625
	x		.0010	.0788		.0791
	y		.0471	-.0269		-.0264
	z		.0515	.7732		.7784
Site 4	Z		.1477	.1523		.1339
	B/λ^2		7.5625	14.8625		14.0125
	x		.0617	.0424		.0417
	y		.1901	.2718		.2725
	z		.6746	.1490		.1548
Site 1A	Z			.0828		.0762
	B/λ^2			0.0000		0.5000
	x			.1294		.1315
	y			.0041		.0042
	z			.7220		.7235
Site 2A	Z			.0525		.1193
	B/λ^2			3.1000		14.1953
	x			-.0069		-.0034
	y			.2474		.2467
	z			.1593		.1623

* Scale factor refined from 3-D phases, not from the centric refinement.

that in the other derivatives. This had been observed at low resolution but its significance was not noticed when the earliest refinements of the 2.5Å centric reflections were carried out and so the initial coordinates of the main sulphydryl sites had been set equal to those of the other derivatives. The refinement had not been able to change the coordinate z_2 from ~.16 to ~.19 any more than it had changed z_1 from ~.73 to ~.77. Refinements in terms of sites at z_2 ~ .16 and ~.19, denoted sites 2A and 2 respectively showed significant occupancy in each site. Lower values of the reliability index were obtained when the occupancy of site 2 was specified as being larger than that of site 2A.

The positions of these six binding sites are entirely consistent with those observed at low resolution and discussed in Chapter III. The sites 1A and 2A are only slightly more than 2Å from the major sites and so would not be resolved in the low resolution maps. The indications of these sites are fairly strong at 2.5Å resolution and so they must be assumed to represent a second preferential position for the mercury atom which binds covalently to the sulphydryl group since they are too strong to represent a protein feature which has been displaced by the ligand.

The best parameters available at this stage for all the derivatives are listed in Table 23. Comparison with Table 17 shows that increasing the scale factor has caused negligible changes in the atomic coordinates, but has affected both the absolute occupancies and the relative values between different sites. The temperature factors have also changed in many cases. Phases for the protein structure factors were calculated from these parameters, but unfortunately this was done before the relative occupancies of sites 2 and

TABLE 24 Calculation of Phases for 6-2.5Å Reflections.
The parameters specified for this calculation and the resulting figures of merit m and RMS lack of closure errors E and E' are tabulated. The values of the mean residual error MRE are shown separately for the isomorphous and anomalous errors and for those for which one or two members of a Bijvoet pair of reflections were measured for the derivatives. The RMS error in the electron density map of the protein calculated using these phases and those of the low resolution reflections (Table 20) are shown for the cases where the anomalous ratio is specified as being negative, positive and zero.

Figure of merit	Overall	Acentric	Centric	MRE(Iso)
6 - 2.5Å (Phases 30-)	.646	.640	.807	.57

6 - 2.5Å Phase Calculation (30-)			MONO	EMP	BAKER
Input parameters: Scale			1.0419	1.0497	1.0316
	E		27.95	25.88	22.20
	E' *	b	7.37	8.71	4.76
		a	1.131	1.142	1.288
Average errors: (Acen. Centric)	MRE	Iso(1)	.40 .50	.42 .65	.51 .61
		Iso(2)	.56 1.10	.55 1.17	.48 1.05
		Anom	.55	.54	.56
	RMS E	acent.	27.80	25.96	21.79
		centric	37.09	35.86	30.80
	RMS E'		25.45	26.70	25.92
		b	7.42	8.52	4.59
		a	1.131	1.150	1.315
Fraction of refs. measured once and twice:					
$(n_1 + n_2)/(n_1 + n_2.\sqrt{2})$ {see p.153}			1.353	1.331	1.359

Choice of correct enantiomer:		m(acentric)	RMS(Δρ)
∞ - 6 Å Phases 15	-ve	.791	.0102
	+ve	.774	.0109
	zero	.735	.0119
6 - 2.5 Å Phases 30	-ve	.640	.0388
	+ve	.631	.0396
	zero	.590	.0416

* linear regression analysis of $E' = a.(250.\sin^2\theta/\lambda^2) + b.$

2A for the BAKER derivative had been correctly determined.

G) CALCULATION OF A ELECTRON DENSITY MAPS FOR CHICKEN TIM

The methods used for determination of the protein phases have already been described. Results of the low resolution determination are listed in Table 20 and for the remaining 2.5Å reflections are shown in Table 24. Their mean residual errors are seen to be essentially correct. The overall figure of merit is 0.655, its variation with $Sin^2\theta$ being shown in Figure 43.

The electron density map of the protein at 2.5Å resolution was calculated using the Fourier program of the X-RAY-70-SYSTEM of programs at the Atlas Computing Laboratory, Harwell. The synthesis was scaled such that an electron density of $1e.Å^{-3}$ appears on the map as the integer 100, assuming a scale factor of 3 between the observed amplitudes and an absolute scale. The density was sampled at intervals of approximately 1/2 Å but the actual intervals were chosen to provide output on a lineprinter at a physical scale of exactly 2 $cm.Å^{-1}$ for sections perpendicular to the z axis.

The 6Å electron density map was calculated on the laboratory Argus computer. It was scaled to the same values as the 2.5Å resolution Fourier synthesis but in this case the density was only sampled at intervals of approximately 2Å.

The F(000) term was not included in either Fourier synthesis. The protein molecules contribute $9 \cdot 5.10^4$ towards its amplitude, that of the water and salt molecules being estimated as 4.10^4 since such small molecules represent about 40% of the weight of the unit cell. An amplitude of $13 \cdot 5.10^4$ for this F(000) term represents an average

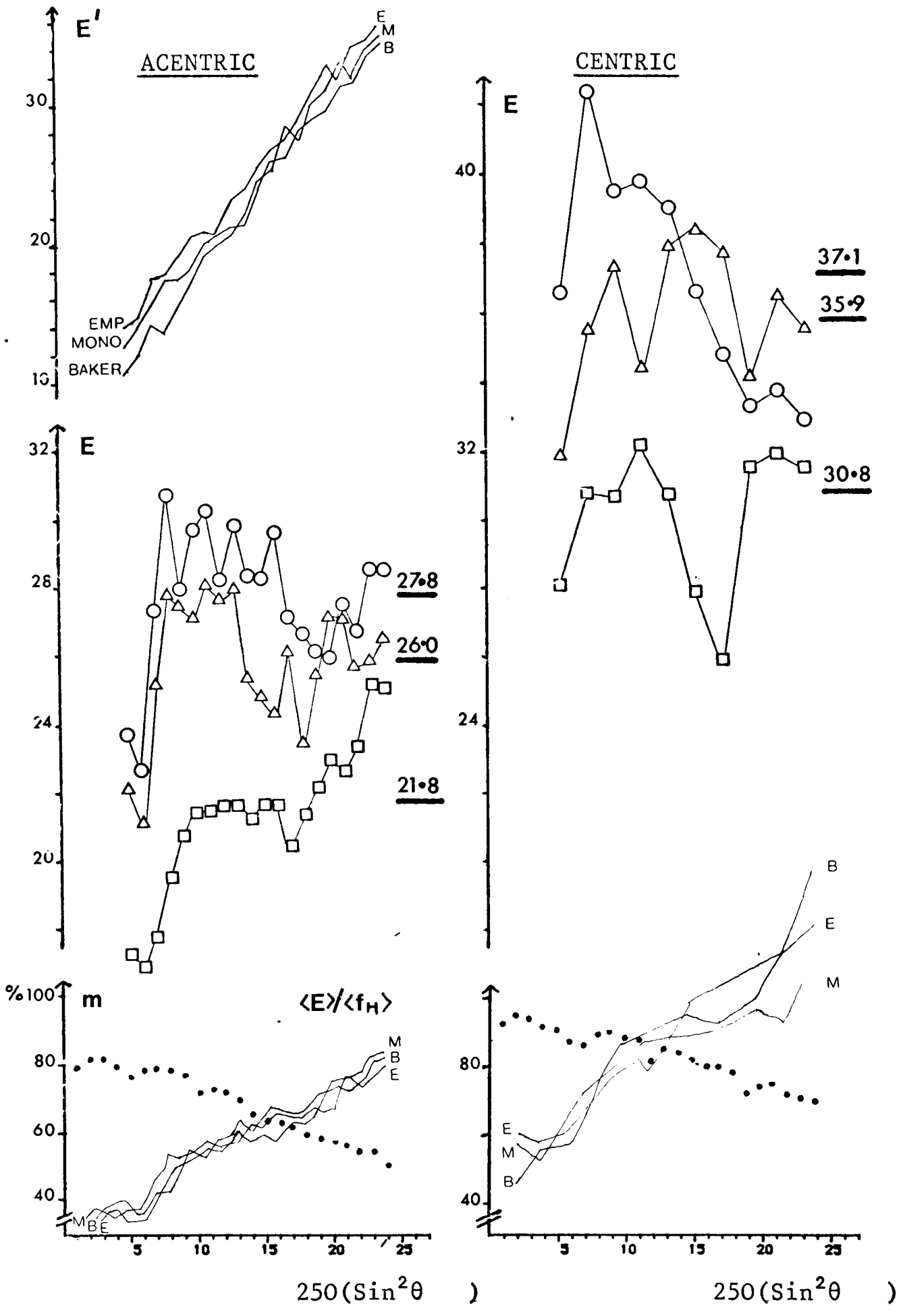


FIGURE 43 Phase Determination; Figures of Merit and Residual Errors. The figure of merit (m) and the RMS lack of closure errors of the isomorphous and anomalous difference triangles (E and E') for each derivative are shown for acentric and centric reflections separately as a function of ($Sin^2\theta$).

○ MONO △ EMP □ BAKER

density of 0.27 e.$\mathring{A}^{-3}$. This must be considered in relation to the features of the protein map, a preliminary description of which is given in the following chapter.

The structure factors which were used to calculate these maps are available on magnetic tape in the Laboratory of Molecular Biophysics.

Chapter VI ELECTRON DENSITY MAPS OF CHICKEN TIM

A) SOURCES OF ERROR IN ELECTRON DENSITY MAPS 177

B) 6Å RESOLUTION ELECTRON DENSITY MAP 179

C) MOLECULAR DIAD AXIS 181

D) 2.5Å RESOLUTION ELECTRON DENSITY MAP 183

A) SOURCES OF ERROR IN ELECTRON DENSITY MAPS

Electron density maps have the minimum RMS error when the coefficients used for their Fourier synthesis are the centroids of the structure factor probability distribution (Blow and Crick, 1959). In this case the RMS error is

$$\langle \Delta\rho^2 \rangle^{1/2} = \left[\sum_{hkl} F_p^2 (1 - m^2) . 2/V \right]^{1/2}$$

(Dickerson, Kendrew and Strandberg, 1961) a quantity which is readily evaluated. However, systematic errors in the map clearly may exceed this value. The effects of such errors in the parameters of the isomorphous derivatives have been discussed in the previous chapter. Errors arising from a sharp truncation of the Fourier series are minimised at low resolution by including only those reflections within the 6Å sphere, as can be seen from the radial distribution of structure amplitudes in Figure 30. At 2.5Å resolution an abrupt truncation of the series is not likely to be important since the amplitudes are decreasing with increasing $\mathrm{Sin}^2\theta$ in this region. James (1948) has shown that the magnitude of the diffraction ripple caused by sharp truncation of the series is less significant anyway for three dimensional syntheses than for those calculated only in projection.

Relative weighting of groups of reflections will introduce systematic features into the map if the weighting is inappropriate. It is clear from their respective figures of merit that the centric reflections are being weighted more heavily than the acentric ones. This is usual for a best Fourier synthesis. However the MRE values for the centric reflections indicate that the E values used for the phase probability distributions were too small for these reflections.

facing page 178

(a)



(b)

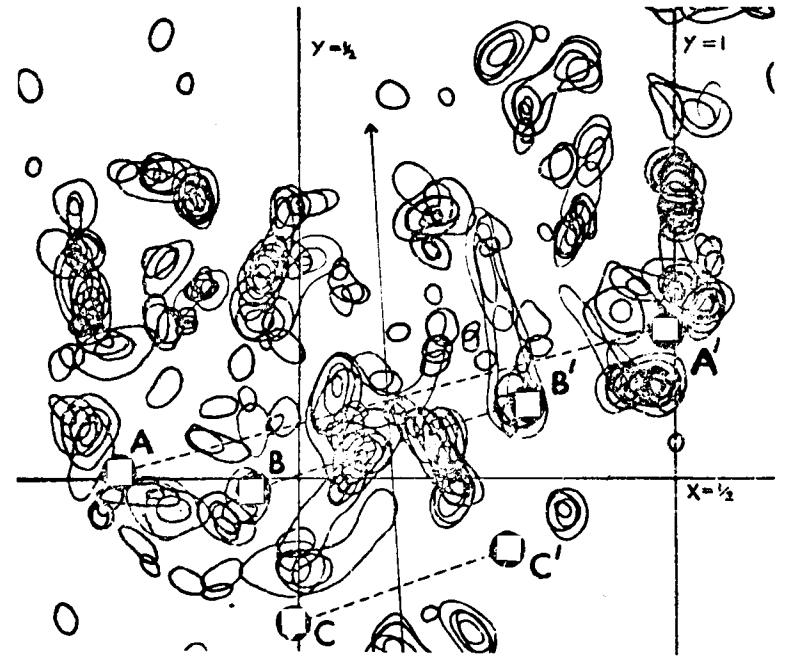


FIGURE 44 Electron Density Map at 6Å Resolution: A Complete Molecule.
The density is viewed along [001] and shows a single molecule. Contours are at intervals of approximately 0.1 e.Å^{-3} omitting the zero and negative contours(a) and also the 0.1 contour(b). The positions of the sulphydryl (A), EMP (B) and platinum (C) sites are marked.

Small E values produce high figures of merit and so the centric reflections are excessively weighted in the Fourier synthesis. The true protein structure will thus appear to be superposed on an image of its enantiomer, although the latter will be of very low weight and probably not detected.

The decrease of the average figure of merit with increasing resolution implies that the high resolution terms are contributing with only low weight. This inevitably decreases the effective resolution of the map. An artificial temperature factor to increase the weight given to reflections according to the gradual decrease of the average figure of merit, was applied in the analysis of sperm-whale myoglobin. Small but significant improvements were observed in the resultant electron density map (D.C. Phillips, see Dickerson, Kendrew and Strandberg, 1961). This procedure has not been adopted since then, but it can be supported theoretically by the analysis of Appendix II where it is shown that the errors arising from uncertainty in the exact phase of a reflection will cancel out when many reflections are considered.

A sensitive criterion is needed to distinguish between electron density maps which have been calculated with slightly modified coefficients in this way. Henderson and Moffatt (1971) suggest that the RMS peak to background ratios in difference syntheses calculated using a set of phases are a more sensitive indicator of their correctness than is the interpretability of an electron density map in terms of an atomic structure. Other possible tests include the comparison of regions of density which are related by non-crystallographic symmetry, and the comparison of the modified sets of phases with calculated phases. This becomes possible after refine-

facing page 179

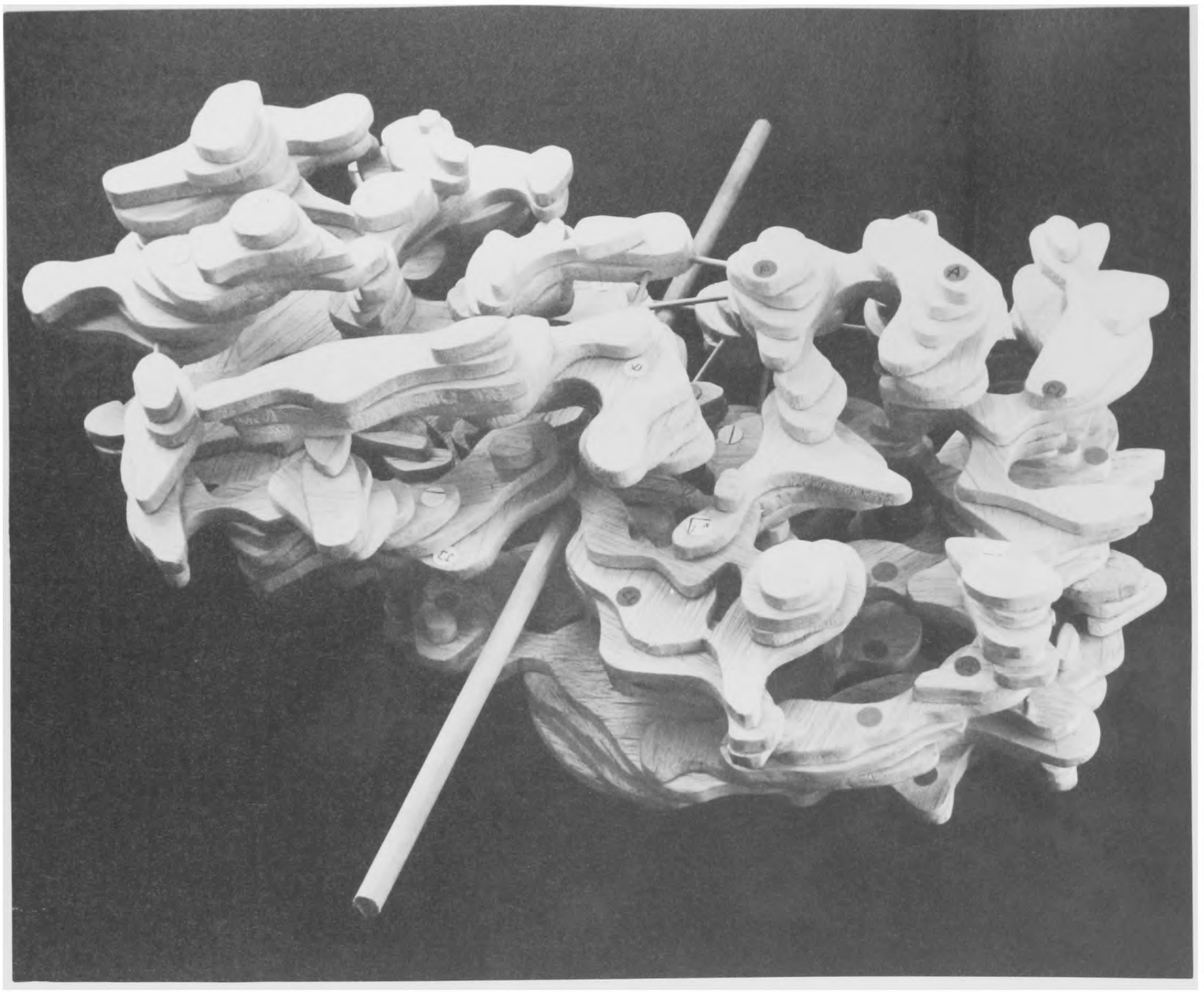


FIGURE 45 A Balsa Wood Model of a Molecule of Chicken TIM.

ment of all the atomic coordinates for a protein molecule to such high precision as has been achieved by Watenpaugh et al. (1971) in their analysis of rubredoxin. However, the application of these ideas to the structure of chicken TIM must await a more detailed interpretation of its atomic structure than that which can be presented below.

B) 6Å RESOLUTION ELECTRON DENSITY MAP

A single molecule of TIM can be clearly distinguished from its neighbours in the electron density map which is represented in Figure 44a. These contour lines indicate all regions of density greater than 0.1 e.Å^{-3} above the mean density, or approximately 0.37 e.Å^{-3} above an absolute origin. A balsa wood model as shown in Figure 45 was constructed to represent the density exceeding 0.1 e.Å^{-3} above the mean density. Hardly any of these contour lines continue across the intermolecular boundaries and so very few artificial cuts had to be made in the density in order to define a single molecule.

The balsa wood model provides the best means of visualising the continuity of the density, but details within the molecule are seen more easily by consideration of the higher levels of density. These are shown in Figure 44b. The extent of the regions of density whose features resemble those of α-helices indicates that these crystallographic results are consistent with those of Jergensons (1966) derived from measurements of optical rotatory dispersion. He predicted a helical content of about 35%. No attempt was made to trace the course of any parts of the polypeptide chain from this low

resolution map.

The silhouettes of the molecule as viewed along the three principal axes are shown in Figure 46. Preliminary studies of the packing of molecules within the crystal are possible in terms of these two dimensional outlines, but the nature of intermolecular contacts can not be studied until a complete atomic model in three dimensions is available. The molecular packing is shown by the projections in Figure 47, from which the most interesting conclusions relate to the crystal morphology.

The dimeric molecules are arranged approximately in sheets perpendicular to the z axis so that the faces $\{001\}$, a form which has never been observed, might be expected to be well developed in these crystals. The form $\{100\}$, which is the least frequent of the three forms observed, appears from packing considerations alone to be much less likely to develope. Prismatic forms of the types $\{hk0\}$ and $\{h01\}$ which are observed can be seen to be more probable from considerations of the molecular packing than the forms $\{0kl\}$ which have not been observed. Traces of the possible faces are indicated in Figure 47 from which it is seen that those which are most probable are $\{110\}$ and $\{201\}$.

The conclusions in Chapter II which were based on the crystal morphology predicted the directions of the major and minor axes of the molecule as viewed down the y axis incorrectly. It is now clear that the projection of the molecule on to the z axis is slightly longer than that on to the x axis, but this does not invalidate any of the other conclusions of that discussion. The mistake arose because the presence of the form $\{100\}$ gave a misleading indication of the nature of the molecular packing into layers.

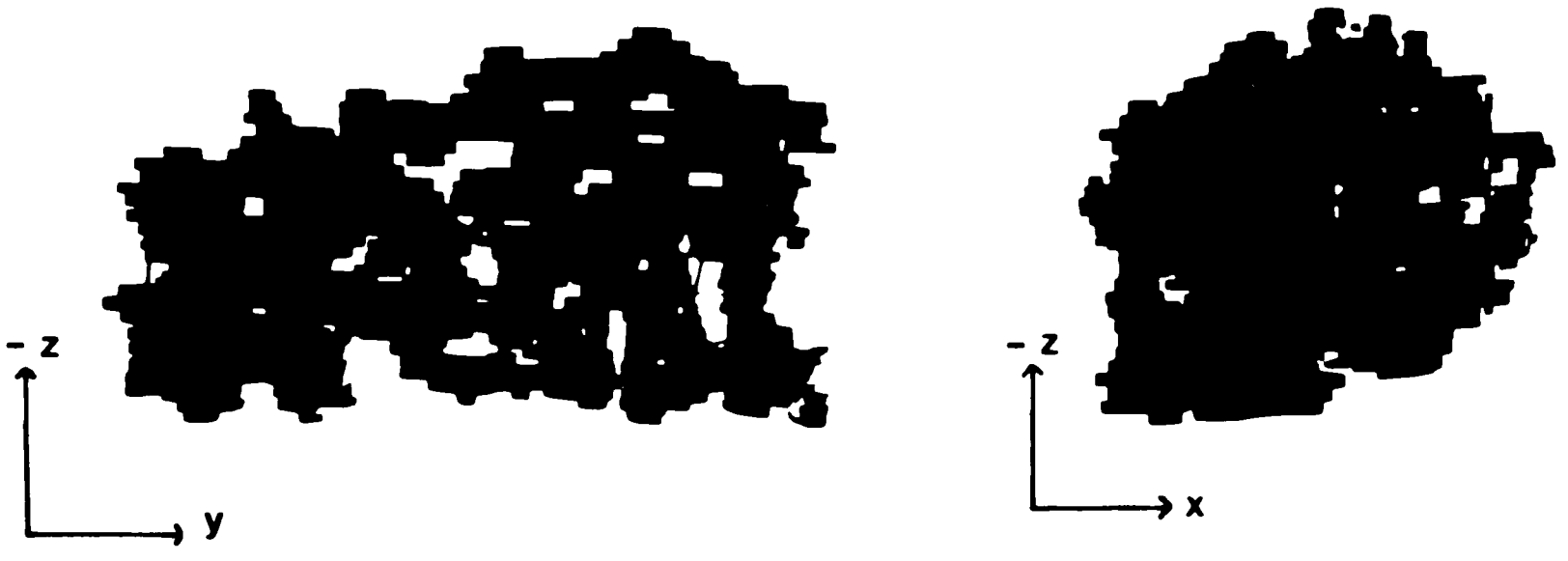


FIGURE 46 Outlines of a Molecule of Chicken TIM.
The silhouettes of the balsa wood model are shown as viewed along the three principal axes.

The molecules of TIM are clearly able to rotate within the crystals more easily about their longest axis than about any other direction. Their arrangement within the crystal indicates that rotations about the y axis are more likely than those about either of the other two principal axes. This is consistent with both the observations concerning radiation damage, which was much more pronounced in the xz plane than along the y axis, and also the changes in the unit cell dimensions, which are significant for both the x and z axes but the y axis is almost invariant. The exact orientation of the longest axis of the molecule, having direction cosines which are approximately (-0.22, 0.86, 0.45), determines that as was observed any changes in the x and z dimensions will be in opposite directions.

C) MOLECULAR DIAD AXIS

The approximate position of the non-crystallographic diad axis can be seen by inspection of the electron density maps. The projection of this axis is indicated in Figure 44b which also shows the positions of the heavy atom binding sites. The sulphydryl sites are separated by 62Å and lie fairly close to the surface of the molecule whereas the additional EMP sites, separated by 31Å appear to be more deeply buried. There was less reason for expecting the chloroplatinite sites to be related by the molecular diad, than for the other sites, but it is found that they are related in this way, their separation being 29Å.

The orientation of the diad which is determined by the mid-points of the lines joining these three pairs of heavy atom sites is

$$\phi = -13^{\circ} \qquad \psi = 88^{\circ}$$

facing page 182

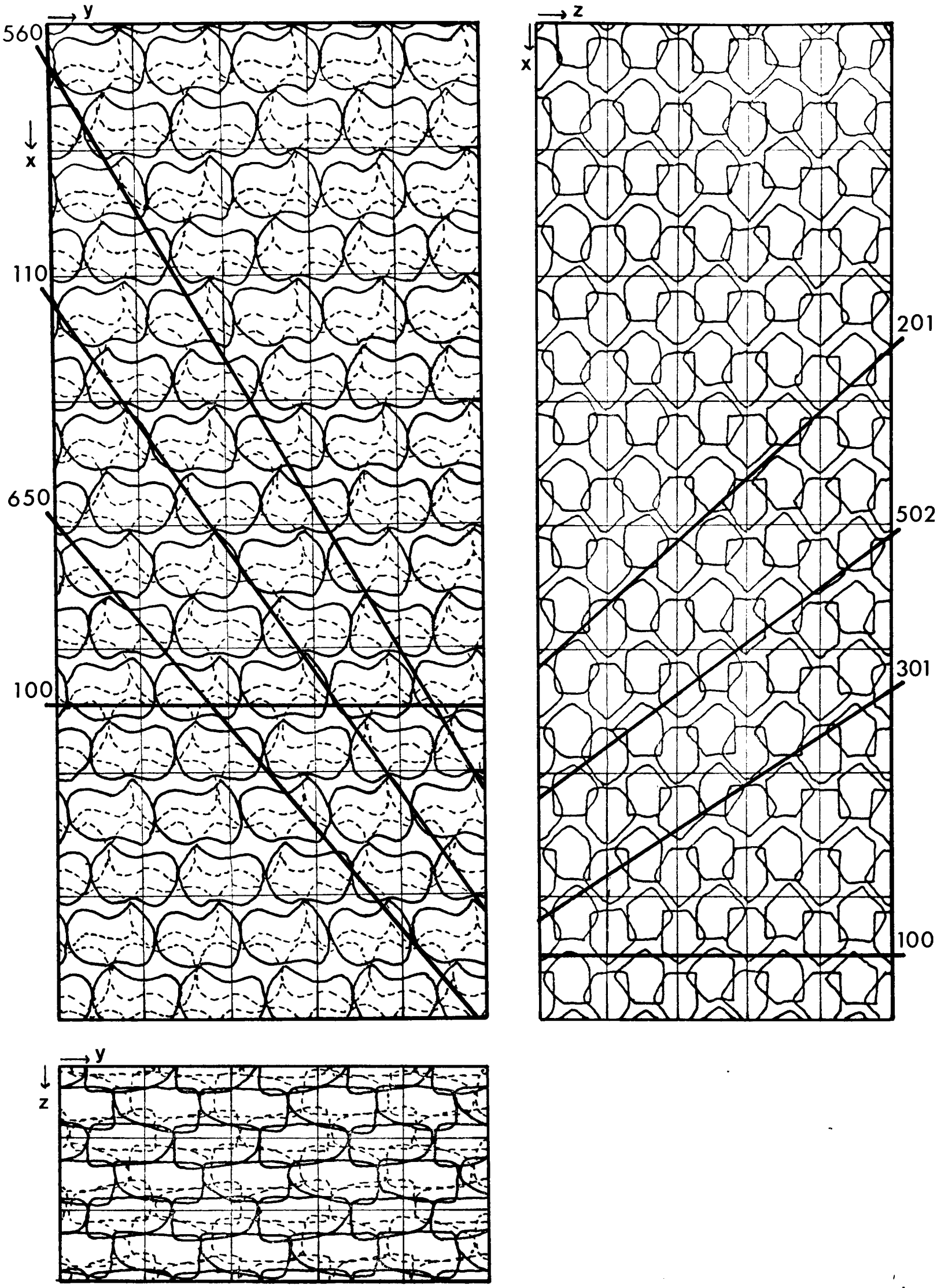


FIGURE 47 Molecular Packing in Crystals of Chicken TIM. The three projections are shown with the traces of several possible crystal faces.

using spherical polar coordinates defined in the manner shown in Figure 12. Consideration of the planes perpendicular to the lines between related binding sites, and their intersections indicate that the orientation of the molecular diad is

$$\phi = -18^{\circ} \qquad \psi = 86^{\circ}$$

It is clear from the arrangement of the heavy atom sites that these angles will not be very accurately determined in this way.

Identification of related features in the electron density map is easier if the map can be sectioned perpendicular to the diad axis of the molecule. The orientation of this axis can then be refined by inspecting the positions of pairs of related features and resectioning the map in a direction that gives better agreement. Regions of the map studied in this way by Banner (1972) enabled him to predict the orientation of the diad as

$$\phi = -20^{\circ} \qquad \psi = 87^{\circ}$$

The direction of the axis was also refined by a least squares analysis in which the correlation of the electron density with itself, after rotation by 180° about this axis, was maximised within the boundary of the molecule. This refinement was carried out using a modification of a program obtained from Mrs J.M. Baldwin (Cox, 1967). The orientation determined by this method is

$$\phi = -21.17^{\circ} \qquad \psi = 86.60^{\circ}$$

for a diad axis passing through the point whose coordinates are

$$(46.04,\ 46.74,\ 27.78)\ \text{Å}.$$

This differs from the mid-point of the line joining the sulphydryl sites by only 1/2Å in the y direction. The correlation of the electron density is 0.65 and the residual 0.43, the latter quantity being defined as



FIGURE 48 Electron Density Map at 2.5Å Resolution.
Three turns of α-helix.

$$R = \langle |\rho_1 - \rho_2| \rangle / \langle |\rho_1 + \rho_2| \rangle$$

where subscripts 1 and 2 refer to the original and the rotated densities.

This approach can be extended to consider variations of the correlation between different regions of the molecule. Definition of suitable regions will become easier when the atomic structure of a monomer of the enzyme is known and a higher resolution map of the molecule available to allow more detailed interpretation of the results than is possible at this stage.

D) 2.5Å RESOLUTION ELECTRON DENSITY MAP

A region of the best 2.5Å resolution map which has yet been calculated is shown in Figure 48. A polypeptide chain in the conformation of an α-helix can be readily fitted to this density. The helix is right-handed, an observation which confirms that the correct enantiomer had been identified from the contribution of the anomalous scattering to the phase determination. The fact that the anomalous scattering ratio was specified for the phase calculation as being negative means that all the heavy atom coordinates listed here represent the enantiomorph of the true heavy atom structure.

The electron density map is being interpreted by means of an optical comparator (Richards, 1968), a device which allows superposition of sections of the electron density map with parts of a skeletal model representing the amino acids. The density corresponding to the polypeptide chain has been identified by I.A. Wilson in many regions of the molecule and the model of these peptide groups has

been built. However, a discussion of the conformation of this enzyme and details of the geometry of its active site must be deferred until the interpretation has been completed and the atomic coordinates determined. The primary sequence of most of the amino acids in chicken TIM is already known, having been determined in this laboratory by A.J. Furth, J. Milman, R.E. Offord and J.D. Priddle. It is expected that parts of their sequence can be independently confirmed by the evidence of the electron density map but that there will be many residues where this is not possible.

Refinement of the atomic coordinates of the enzyme can follow an initial interpretation of the electron density map. For this purpose a better map may have to be calculated because, although the current map is capable of interpretation, there are many ways in which it could be improved. The precision of the intensity measurements could now be greatly increased by the use of both stronger sources of X rays and a reduction in the temperature of the crystals, which will minimise the extent of any radiation damage. Uncertainty in the anomalous differences would be minimised if the measurements of a Bijvoet pair of reflections were separated by only a short time interval. The precision of the phase determination would be improved if more accurate measurements of the isomorphous and anomalous differences were obtained in this way and further improvement could also result from the use of an additional derivative. The problems which have been raised by the use of derivatives having common sites suggest that a derivative in which the sulphydryl sites were not substituted would be most suitable for this purpose. A search for possible derivatives is in progress. Recent experiments with potassium chloroplatinite have produced derivatives in which platinum was bound

not only at the two sites already described but also at the sulphydryl sites (G.A. Petsko and I.A. Wilson, personal communication). The most promising approach thus seems to be the use of a blocking reagent to react with the sulphydryl group before preparing a derivative.

These possibilities for extending the crystallographic work beyond the present stage are, however, of less immediate importance than the complete interpretation of the existing 2.5Å resolution density map and determination of the atomic structure of TIM. Correlation of this structure with results obtained from a wide variety of techniques by members of the Oxford Enzyme Group and many others in their study of this enzyme is a prospect which will shortly be realised, enabling a deeper insight into the function of this enzyme and its importance in metabolism.

Appendix I COMPUTING SOFTWARE 187

A) GENERAL PROGRAMS 187

B) PROGRAMS FOR PAPER TAPE DATA (LOW RESOLUTION) 188

C) PROGRAMS FOR MAGNETIC TAPE DATA 189

Appendix II RELATIVE WEIGHTING OF COEFFICIENTS IN FOURIER SYNTHESES 198

A) APPROXIMATIONS USING STATISTICAL AVERAGES 199

B) COSINE APPROXIMATION 199

C) POWER SERIES APPROXIMATION 200

D) EXACT ANALYSES 202

E) SINGLE DERIVATIVE PROTEIN MAPS 204

Appendix III DERIVATIVE SCALE FACTOR DETERMINATION USING ISOMORPHOUS AND ANOMALOUS DIFFERENCES 208

Appendix I COMPUTING SOFTWARE

A) GENERAL PROGRAMS

Programs which were used infrequently are referred to individually in the text. Extensive use was made of the following programs which are generally available in the laboratory:

1) Primary Data Processing: A program written by Dr. A.C.T. North to produce relative structure factors from the diffractometer intensity measurements. Versions are available for both the linear diffractometer (flat-cone setting) and the four circle diffractometers (equi-inclination setting) with paper tape and magnetic tape options for the output data,(North, 1964).

2) Fourier Series Calculation: A program written by Dr. S.D. Dover for all space group symmetries up to and including those of the orthorhombic system. Versions are available for both paper and magnetic tape input and output channels.

3) Refinement of Heavy Atom Parameters: A program "PANGLOSS" to use the centric data for refinement of these parameters was written in Fortran by Dr. I.D.A. Swan following the logic of an earlier version in Autocode. Subsequent modifications to this program have been made by individual users.

4) Phase Determination using Multiple Isomorphous Derivatives: A program to do this by the methods described in Chapter 5 was written in Fortran by Dr. G.J. Stubbs with a version for space group $P2_12_12_1$. This has been modified by the present author to include a more detailed analysis of the errors and to allow the use of magnetic tapes for the input and output of data.

5) Contouring of Electron Density Maps: A program to produce contours on the graph plotter which is controlled by the Argus

computer was written by Dr. S.D. Dover. It has subsequently been modified by G.A. Petsko and other users and a second version, to display contours on the CRT has been developed by Dr. C.D. Barry and L.O. Ford.

B) PROGRAMS FOR PAPER TAPE DATA (LOW RESOLUTION)

A subroutine written in machine code by Dr. A.C.T. North was used in all the following programs for flexibility in reading the data tapes.

Secondary data processing was done with a Fortran program written by D.W. Banner to determine the mean value of equivalent reflections, the difference between them and to analyse the agreement between the equivalents.

All the following programs were written by the author in Fortran. They were used in the original analysis but all of the relevant data is now available on magnetic tape and the computation is done more efficiently using the programs described in the following section. Those programs which require the storage of many reflections are limited to indices not exceeding those of the 6Å chicken TIM data. The package of programs includes those identified by the following titles:

1) SCALER: determines and applies an overall scale factor between two sets of data.

2) DSA: derivative scale analyser, to determine this scale as a function of the indices h,k,l and $Sin^2\theta$ with the parameters of the weighted linear regression analyses. This was based extensively on a program written by Dr. L.N. Johnson.

3) AMALGAMATE: combines onto a single tape the data for native and four derivative crystals.

4) SIGN CROSSOVER: monitors those centre reflections for which the signs determined on refinement of the heavy atom parameters of a single derivative differ from those determined by multiple derivatives in the phase program.

5) APPLICATOR: applies a set of calculated phases to the observed isomorphous differences for any derivative, after rescaling the derivative data.

6) DOUBLE DIFFERENCE: determines the real and imaginary parts of the difference between the observed and calculated derivative structure factors, assuming the calculated phase; determines the standard residuals and analyses the distribution of angles within the phase triangle.

7) SEPARATOR: separates centric or acentric data from a mixed tape.

8) MULTIPLIER: performs multiplication or division by an integer or floating point number to all the data items on a tape, as needed for example in the relative weighting of successive lists of reflections.

9) PTAPE-MAGT: transfers data on paper tapes of standard format to magnetic tape standard format.

C) PROGRAMS FOR MAGNETIC TAPE DATA

The magnetic tape system and the basic Fortran subroutines for using these tapes were developed by Dr. S.D. Dover. The program giving visual display on the CRT of data stored on magnetic tape was written by Dr. A.C.T. North. Two programs written by the author for general use allow the transfer of data between tapes, and the output on punched paper tape of the data stored for any selected group of reflections when a teleprinter record of this is required.

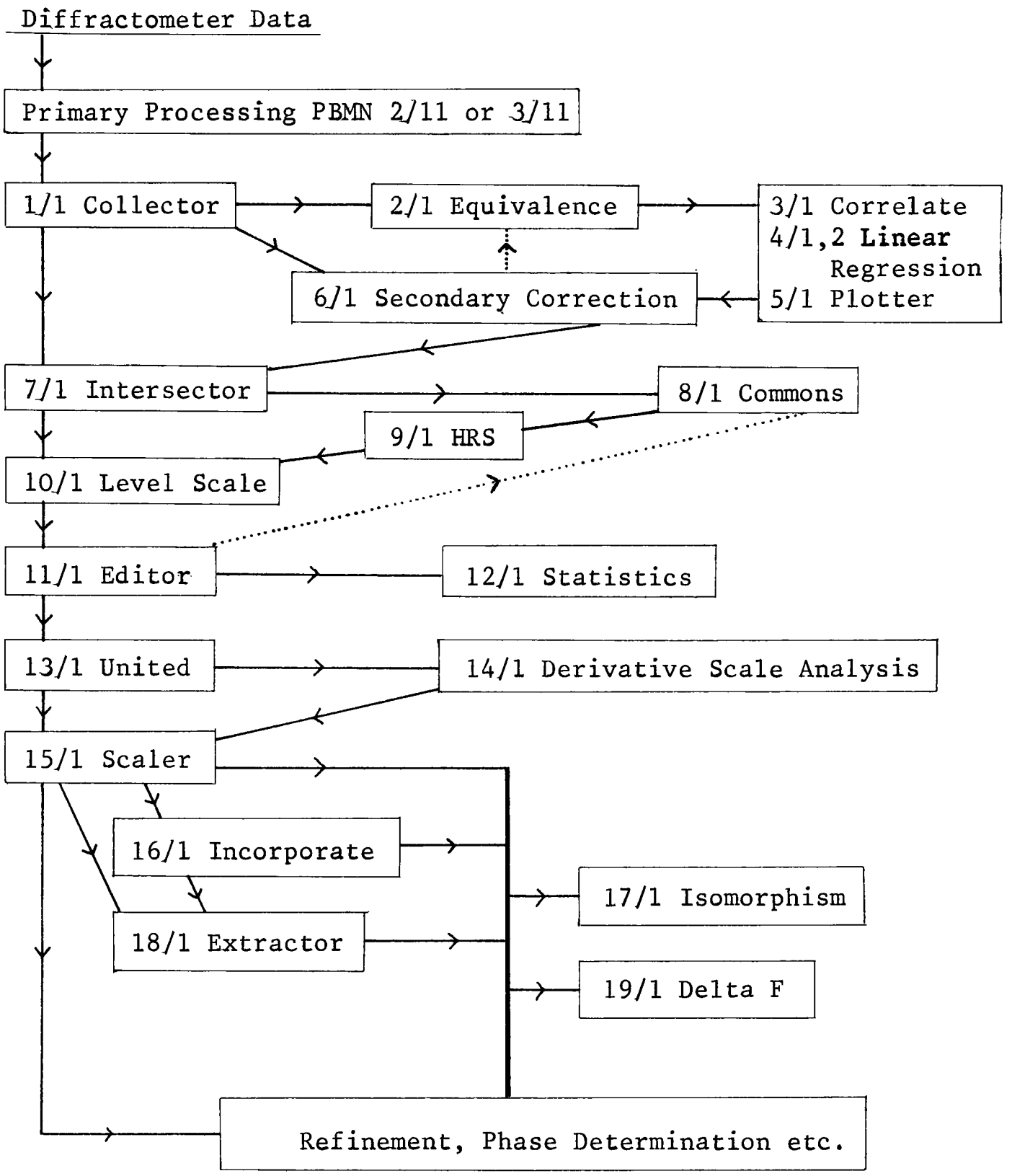


FIGURE A1 Data Processing Programs.
The Sequence of programs is shown by the flow diagram where the vertical arrows represent the essential steps which involve format changes within the data files. The facilities for optional analysis of the data are indicated and the dotted lines show the possibilities for re-cycling and repeating stages of scaling, editing etc. if necessary.

The following set of programs written by the author, which are mainly in Fortran, were designed to handle data for the symmetries up to that of the orthorhombic system. Their data storage capacity is limited by the core store of the computer but exceeds by a factor of two that which was required for this work. The individual programs are identified in the conventional laboratory code by the symbols PBFBm/n followed by a single word description, the significance of the code being Program Binary Fortran Bloomer where m,n are integers referring to the program number and the current version of the program respectively.

All programs which read magnetic tapes monitor the title of every file, and those that write magnetic tapes contain provision for revising the title, the identifier being automatically changed. Data is stored on magnetic tape in the laboratory standard format of 10 or 20 words per reflection, the first containing the three indices, with 200 words per block. All files have a block counter which is routinely checked whenever a file is read. Standard deviations, which are shown as σ in the tape formats, are those appropriate to the quantity above them on the preceding line, or the quantity to their left in the formats with 20 and 10 words respectively. Weighted values are used for all the determinations of mean structure factors of a single reflection, but are only used when indicated by ω for overall averages of groups of reflections.

PBFB1/1 COLLECTOR: collects together from multiple files on tape all data for a single level of reflections; writes a single file of output with data for both reflections of a Bijvoet pair, after sorting on both indices; removes systematic absences; applies

specific outer resolution cut-off; monitors all missing data between this and a cylindrical inner resolution limit.

PBFB2/1 EQUIVALENCE: compares Bijvoet pairs of reflections; monitors all those where $|\Delta F|/\sigma$ or σ exceeds a fixed value; calculates and lists for all values of each index within the level the mean values and standard deviations of

$\overline{F}$, ΔF, $|\Delta F|$, F_+, F_-, $(\omega.F_+/F_-)$;

determines overall value for the level of $\langle|\Delta F|\rangle/\langle F\rangle$ and its standard deviation; if $F < \sigma$ for any measurement it is excluded from all averages involving ΔF.

PBFB3/1 CORRELATE: correlates results for the five levels from a single crystal; reads paper tape output of previous program for five different levels; punches new tape containing $\langle\omega.F_+/F_-\rangle$ and σ for all values of both indices in a compact format for visual comparison of five levels.

PBFB4/1 and 2 LINEAR REGRESSION: two versions without and with an origin constraint; determines the parameters and standard deviations of the weighted linear regression of the data from the previous program, for each level separately and overall values for the five together.

PBFB5/1 PLOTTER: plots graphically the data of programs 3/1 and 4/1 or 2 for ease of visual inspection.

PBFB6/1 SECONDARY CORRECTION: secondary absorption correction or radiation damage correction, as determined by the analysis of previous programs is applied to the file of data written by 1/1; writes new file in same format; provides editing facility with CRT display and handswitch control for both specified reflections identified from 2/1 monitor output and also those where $|\Delta F|/\sigma$

or σ exceed certain values; determines the new value, after applying the correction and editing, of $\langle|\Delta F|\rangle/\langle\overline{F}\rangle$, its standard deviation and the reflection count for 0,1,2 equivalents measured.

PBFB7/1 INTERSECTOR: intersecting levels of data are merged from two sets of files of levels about the different indices (2,3) being replaced by a single set of files of levels about one of them (3) in the format

hkl F_{+3} σ F_{-3} σ F_{+2} σ F_{-2} σ 0

PBFB8/1 COMMONS: common row totals are extracted for inter-level scaling; for every row common to both sets of levels the following totals are determined and punched on paper tape

$\sum\omega.F^{n}_{\pm3}$ σ $\sum\omega.F^{n}_{\pm2}$ σ N_{REF} $\sum\omega$

where subscripts 2,3 refer to the sets of data, ± refer to separate Friedel equivalents, the contribution to both sums being weighted down if there is no true equivalance, superscript n is 1 or 2 for structure amplitudes or intensities, weights ω are unity or statistical mean weights.

PBFB9/1 HRS: HAMILTON, ROLLETT and SPARKS method used for the determination of the best scale factors, the iterative procedure continuing until convergence is complete. This program was based extensively on a translation into Fortran by Dr. L.N.Johnson of an earlier autocode version of the HRS analysis.

PBFB 10/1 LEVEL SCALE: HRS level scale factors are applied; reflections where $|\overline{F}_3 - \overline{F}_2|/\sigma$ exceeds a specified value are monitored; a new tape is written with the data in the format

hkl F_{3+} F_{3-} F_{2+} F_{2-} $\overline{F}_3$ $\overline{F}_2$ $\overline{F}_3-\overline{F}_2$ $\overline{F}$ $\overline{F}_+$

0 σ σ σ σ σ σ σ σ $\overline{F}_-$;

for the grid of intersecting data the mean values summed along the common rows and the overall values are printed of

$\langle \overline{F}_3 - \overline{F}_2 \rangle / \langle \overline{F} \rangle$, $\langle (\overline{F}_3 - \overline{F}_2)/\sigma \rangle$, $\overline{F}$

together with the number of common reflections for each intersection; mean values involving $|\overline{F}_3 - \overline{F}_2|$ instead of the true difference can also be determined.

PBFB 11/1 EDITOR: editing facility using CRT display and direct keyboard input, for reading data files as written by the previous program; reflections displayed are determined by their indices or the discrepancy between any two of their observed or derived structure amplitudes exceeding specified numbers of standard deviations; best values of the mean structure amplitude and anomalous difference are determined, a new tape being written in the format

hkl	F_{3+}	F_{3-}	F_{2+}	F_{2-}	F_3	F_2	ΔF	F	F_+
Scatter	σ	σ	σ	σ	σ	σ	σ	σ	$\overline{F}_-$

where the scatter is calculated from the spread of all the measurements.

PBFB 12/1 STATISTICS: statistical analysis of the internal consistency and completeness of the edited data files; these are analysed in terms of $Sin^2\theta$. $\overline{F}$, and any two of the three indices, together with overall values, of

$\langle \omega . \lvert \Delta F_3 \rvert / \overline{F}_3 \rangle$	or	$\langle \lvert \Delta F_3 \rvert \rangle / \langle \overline{F} \rangle$
$\langle \omega . \lvert \Delta F_2 \rvert / \overline{F}_2 \rangle$		$\langle \lvert \Delta F_2 \rvert \rangle / \langle \overline{F} \rangle$
$\langle$Scatter/ $\overline{F}\rangle$ *		$\langle \lvert F_i - \overline{F} \rvert \rangle / \langle \overline{F} \rangle$ *
$\langle$Scatter/ $\sigma\rangle$ *		$\langle \lvert \Delta F/2 \rangle \rvert \rangle / \langle \overline{F} \rangle$
$\langle \overline{F} \rangle$		

where those quantities indicated * include only the scatters or

deviations relative to the appropriate value $\overline{F}_+$, $\overline{F}_-$ for acentric reflections; the number of reflections for which there are 0,1,2,3,4 estimates of the structure amplitude and 0,1,2 estimates of the anomalous difference are also listed for each variable.

PBFB 13/1 UNITED: unites into a single set of files, one for each level, the data for native crystals and up to four isomorphous derivatives identified by numerical subscripts on a new tape of format

hkl	F_P	$\overline{F}_{PH1}$	ΔF_1	$\overline{F}_{PH2}$	ΔF_{PH2}	ΔF_2	$\overline{F}_{PH3}$	ΔF_3	$\overline{F}_{PH4}$	ΔF_4
0	σ	σ	σ	σ	σ	σ	σ	σ	σ	σ

PBFB 14/1 DSA: derivative scale analysis for any single derivative, determining the intensity or structure amplitude scale factor as a function of the indices h,k,l, F_P and $\mathrm{Sin}^2\theta$ with the parameters of the weighted linear regression analyses.

PBFB 15/1 SCALER: applies a scale factor to each set of derivative data as an overall constant or a linear function of h,k,l, $\mathrm{Sin}^2\theta$ or an exponential function of $\mathrm{Sin}^2\theta$.

PBFB 16/1 INCORPORATE: incorporates a sphere of low resolution data, monitoring any discrepancy between those values and any overlap in the high resolution data files; determines the overall values for the overlapping reflections for the native and each derivative set of data of the difference, and its absolute value, its standard deviation and the number of overlapping reflections.

PBFB 17/1 ISOMORPHISM: isomorphism criteria, absolute scale and derivative to native scaling can be investigated by the values printed as a function of $\mathrm{Sin}^2\theta$ for centric and acentric reflections and the overall values of the RMS and mean of the moduli of

F_P^2 F_{PH}^2 F_P F_{PH} $(F_{PH}^2-F_P^2)$ $(F_{PH}-F_P)$ $(F_+^2-F_-^2)$ (F_+-F_-)

and ratios subsequently derived from these quantities. Special functions are calculated from these mean values for consideration of the absoltue scale factor and overall temperature factor (Wilson, 1942), the degree of isomorphism (Crick and Magdoff, 1956), the empirical ratio of the real to imaginary parts of the heavy atom scattering (Matthews, 1966b) and the scale factor relating derivative and native measurements (Singh and Ramaseshan, 1966).

PBFB 18/1 EXTRACTOR: extracts from a set of files any specified set of projectio nand acentric data, between inner and outer resolution limits, multiplies or divides all data items by a constant, writes a new tape with data in the same format in separate files or combined into a single master data file.

PBFB 19/1 DELTA F: mean value or RMS of the modulus of the anomalous difference ΔF is determined as a function of $Sin^2\theta$ and overall for the three principal projections and the acentric data.

PBFB 30/1 PANGLOSS: modification of the refinement program already described.

PBFB 31/1 PHASES: modification of standard laboratory phase program already described, additional analysis of lack of closure errors e, e' for isomorphous and anomalous differences being included; data output files in the format

hkl m A_P B_P e_1 σ_1 e_1' σ_1' e_2 σ_2

e'_2 σ_2' e_3 σ_3 e_3' σ_3' e_4 σ_4 e_4' σ_4'

where the subscripts refer to four derivatives and

σ = $RMS(\sigma_{FPH}, \sigma_{FP})$ isomorphous s.d.

σ' = $RMS(\sigma_{FPH+}, \sigma_{FPH})$ anomalous s.d.

PBFB 32/1 APPLICATOR: analogous to the program for low resolution data; generates data files suitable for ΔF or $\Delta\Delta F$ maps, analysis of residuals and errors, correlations etc.; phases are applied to the observed differences for a single derivative, the format of the data tape written being

hkl	A_P	B_P	A_{PH}	B_{PH}	a_H	b_H	$A_{\Delta F}$	$B_{\Delta F}$	m
$\alpha_P\alpha_{PH}\alpha_H$	F_P	F_{PH}	f_H	ΔF_i	ΔF_a	e	σ_i	e'	σ_a'

where the standard deviations are defined as above, and

$$\Delta F_i = F_{PH} - F_P$$

$$\Delta F_a = F_{PH+} - F_{PH-}$$

all real and imaginary parts A, a, B, b being weighted down by the figure of merit, as in

$$A_{\Delta F} = m\Delta F_i . \cos\alpha_P$$

PBFB 33/1 E-EPRIME : determines the mean or RMS values of E, E' from the values of e,e' for individual reflections; analysis for each derivative, acentric and each projection of centric reflection separately, for ranges of $\sin^2\theta$.

PBFB 34/1 REFINED 3-D SCALE: uses the data files written by PBFB 32/1 and analyses the derivative scale factor determined as a function of h,k, l, F_p and $\sin^2\theta$ by the linear regression relationship

$$k = \frac{\langle F_{PH\ obs.}\ F_{PHcalc}\rangle}{\langle F_{PH\ obs}{}^2\rangle}$$

calculating the parameters of the least squares line for the variation of k using both unit weights and weights proportional to $\langle F_{PHobs}\rangle^2$.

PBFB 35/1 ERROR-SIGMA ANALYSIS: analyses the correlation between e, e', σ, σ', m, F_{PH} by considering the variation with respect to

e, e', m, F_{PH} of

$(e^2 - \sigma^2)$, $(e' - \sigma'^2)$, σ^2, σ'^2

PBFB 36/1 RESIDUALS: reads the data files written by 32/1 and analyses as a function of $Sin^2\theta$ for centric and acentric reflections separately the values of

the reliability index $R = <\Delta\Delta F>/<\Delta F>$

the Kraut R factor $R_K = <\Delta\Delta F>/<\Delta F>$

the mean fractional difference $D = <\Delta F>/<F_{PH}>$

for the observed isomorphous differences, and the calculated values and ratios of

$<e>$, $<f_H>$, $<e'>$, $<f_H'>$.

Appendix II RELATIVE WEIGHTING OF COEFFICIENTS IN FOURIER SYNTHESES

Difference Fourier syntheses involving acentric reflections contain features representing not only the difference structure, but also the structure from which the phases were calculated, and the convolution of this with the difference structure. Luzzatti (1953) showed that in such syntheses the peaks representing the difference structure appear with a weight which is less than or equal to one half of that which would otherwise be expected. This result has since been derived by several different methods. Various approximations are always involved in these analyses, but may be considered in two main categories; firstly the statistical averages over many reflections, and secondly the algebraic approximations simplifying the representation of the coefficient for a single reflection.

The present analysis considers the validity of these approximations; suggests a method for avoiding any algebraic approximations in consideration of the difference Fourier synthesis having coefficients ΔF, the analogous result for $(\Delta F)^2$ syntheses having been derived by Phillips (1966); considers the effect which including or neglecting crossover terms has on the features in the synthesis; and finally applies these considerations more generally to the analysis of Fourier syntheses in which the phases are uncertain, as is the case with "best" electron-density maps of a protein. The terminology employed throughout this discussion is that of protein, derivative and heavy atom structure factors, using the notation of Figure 13a, but the results are clearly relevant to other forms of difference synthesis.

A) APPROXIMATIONS USING STATISTICAL AVERAGES

The basic result of Wilson statistics (Wilson, 1942) may be expressed as

$$< F_{PH}^2 > \quad = \quad < F_P^2 > \; + \; < f_H^2 >$$

which thus implies, as is readily seen from the phase triangle

$$< \underline{F}_P \cdot \underline{f}_H > \quad = < F_P f_H \, \mathrm{Cos}\phi > \; = 0 \qquad (1)$$

If this is true, then since

$$F_P^2 \; = \; F_{PH}^2 + f_H^2 - 2F_{PH}f_H \mathrm{Cos}\gamma$$

it follows that

$$< \underline{F}_{PH} \cdot \underline{f}_H > \; = \; < F_{PH} f_H \mathrm{Cos}\gamma > \; = \; < f_H^2 > \qquad (2)$$

It can be shown experimentally that ϕ and γ are essentially randomly distributed angles, but that β is usually small. However the derivative and heavy atom structure factors are clearly loosely correlated so the average shown in equation (2) does not vanish. Similarly it can be shown that

$$< \underline{F}_{PH} \cdot \underline{F}_P > \; = \; < F_P^2 > \; = \; < F_{PH} F_P \, \mathrm{Cos}\beta >$$

The use of these average approximations appears to be unavoidable in consideration of the features shown by differences syntheses. As shown in Chapter V they are more likely to be applicable for high resolution sets of measurements.

B) COSINE APPROXIMATION

The difference between the observed isomorphous difference and the projected heavy atom structure factor

$$F_{PH} \; - \; F_P \; - \; f_H \mathrm{Cos}\phi \; = \; \delta$$

is a small positive quantity which can be neglected if f_H is small compared with F_P, so that

$$\Delta F \quad \sim \quad f_H \mathrm{Cos}\, \phi$$

This led Moews and Bunn (1971) to suggest that since for acentric reflections

$$< \Delta F > \sim < f_H > . 2/\pi$$

the observed isomorphous differences for centric reflections should be multiplied by $(2/\pi)$ before squaring them for use in difference Patterson syntheses. The exact result obtained in the following section shows that this is incorrect, and that the quantity more closely related to the difference Patterson syntheses

$$< \Delta F^2 > \sim < fH^2 > < Cos^2\phi >$$

indicates correctly the factor of $(1/\sqrt{2})$ to be applied to the observed differences.

Extension of this result to difference Fourier syntheses by D.M. Blow (personal communication) shows that

$$\Delta F\, e^{i\alpha p} \sim f_H Cos\phi\, e^{i\alpha p}$$

$$= f_H Cos\phi(Cos\phi - iSin\phi)e^{i(\alpha p+\phi)}$$

$$= f_H e^{i\alpha H}(Cos^2\phi - i\ Sin\phi Cos\phi)$$

since $\alpha_h = \alpha_P + \phi$

Thus if ϕ is random and not correlated with f_H then

$$< \Delta F e^{i\alpha p} > = < f_H e^{i\alpha H} > (\tfrac{1}{2} - i.0)$$

so that the contribution of acentric reflections is weighted down on average by one half for both difference Fourier and Patterson syntheses. However, the origin of the factor $Cos^2\phi$ has been shown to be quite different in these two cases.

C) POWER SERIES APPROXIMATION

This analysis of the features expected in a difference Fourier synthesis is by E. Dodson (see Moult, (1970), p. 33) and is included here because it provides a physical interpretation of

the additional features

$$F_{PH}^2 = F_P^2 + f_H^2 + 2F_P f_H \cos\phi$$

$$F_{PH} = (F_P + f_H)\{1 + 2F_P f_H(\cos\phi-1)/(F_P + f_H)^2\}^{\frac{1}{2}}$$

$$\sim (F_P + f_H)\{1 + F_P f_H(\cos\phi-1)/(F_P + f_H)^2\}$$

using a power series expansion and a first order approximation. This will introduce more serious errors as

$$f_H/F_P \rightarrow 1 \qquad \text{and } \cos\phi \rightarrow -1$$

when the term in the square brackets tends to zero. However it is usually applicable and thus

$$F_{PH} - F_P = f_H + F_P f_H(\cos\phi-1)/(F_P + f_H)$$

$$\Delta F e^{i\alpha p} = \{f_H(F_P + f_H) + F_P f_H(\cos\phi-1)\}e^{i\alpha p}/(F_P + f_H)$$

$$= \{f_H^2 + \tfrac{1}{2}F_P f_H(e^{i\phi} + e^{-i\phi})\}e^{i\alpha p}/(F_P + f_H)$$

But $\phi = \alpha_H - \alpha_P$

and thus

$$\Delta F e^{i\alpha p} = \frac{f_H^2}{Fp+f_H} e^{i\alpha p} + \frac{Fpf_H}{Fp+f_H} \cdot \frac{e^{i\alpha_H}}{2} + \frac{Fpf_H}{Fp+f_H} \cdot \frac{e^{i(2\alpha p-\alpha_H)}}{2}$$

Of the three terms on the right of this equation the first represents the native protein structure with a weight that is usually fairly small, but approaches half when f_H/F_P approaches unity. The second term represents the heavy atom structure which has a maximum weight of one half when f_H/Fp is very small, decreasing to a quarter as this ratio approaches unity. The third term represents randomly distributed features provided that F_H and f_H are not correlated and ϕ is random.

D) EXACT ANALYSES

The analysis of Phillips (1966) showed that

$$F_{PH} - F_P = f_H\cos\gamma - 2F_P\sin^2\frac{\beta}{2}$$

and thus the coefficients of the difference Patterson synthesis are

$$\Delta F^2 = f_H{}^2\cos^2\gamma + 4F_P{}^2\sin^4\frac{\beta}{2} - 4F_Pf_H\cos\gamma\sin^2\frac{\beta}{2}$$

Statistical averages suggest that γ is essentially randomly distributed and β is usually small. Using the notation of Phillips where P, H refer to protein and heavy atom respectively, then the three terms on the right of this equation represent firstly the H.H. interactions at half weight, secondly the P.P interactions at low weight and thirdly the H.P. interactions which will also have low weight.

The recognition of reflections where there is a crossover, i.e. $\beta > 90^\circ$, by the use of the sum instead of the difference of the protein and derivative amplitudes, implies

$$F_{PH} + F_P = f_H\cos\gamma + 2F_P\cos^2\frac{\beta}{2}$$

and thus

$$\Delta F_{sum}{}^2 = f_H{}^2\cos^2\gamma + 4F_P{}^2\cos^4\frac{\beta}{2} + 4F_Pf_H\cos\gamma\cos^2\frac{\beta}{2}$$

The use of this sum synthesis whenever possible for reflections for which $\beta > 90$ thus reduces the size of any spurious background features representing P.P or H.P interactions but it has no effect on the weight of the features representing H.H. interactions. These appear with half weight on average, from acentric reflections irrespective of whether or not crossover terms are included in the synthesis.

Extension of this analysis to the case of difference Fourier syntheses becomes possible by using the method of Blow which was

described above. Thus

$$\Delta F e^{i\alpha p} = (f_H \cos\gamma - 2F_P \sin^2\tfrac{\beta}{2})\, e^{i\alpha p}$$
$$= f_H e^{i\alpha_H}(\cos\gamma . e^{-i\phi}) - 2F_P e^{i\alpha p}(\sin^2\tfrac{\beta}{2})$$

Consideration of crossover terms will change this result to

$$\Delta F_{sum} e^{i\alpha p} = f_H e^{i\alpha_H}(\cos\gamma . e^{-i\phi}) + 2F_P e^{i\alpha p}(\cos^2\tfrac{\beta}{2})$$

The protein features normally have negative weight, this being reversed, and the amplitude of the term reduced by recognition of appropriate crossover terms. The weight of the heavy atom features can be considered from

$$< \cos\gamma e^{-i\phi} > = < \cos(\phi-\beta) . e^{-i\phi} >$$
$$= <(\cos^2\phi\cos\beta + \sin\phi\cos\phi\sin\beta)$$
$$- i(\sin^2\phi\sin\beta + \sin\phi\cos\phi\cos\beta)>$$
$$\sim \tfrac{1}{2}\{<\cos\beta> - i<\sin\beta>\}$$

assuming that statistically ϕ and β are independent and also that ϕ is random. The real term inside these square brackets will be almost unity, giving rise to heavy atom features of slightly less than half weight. The imaginary term will be small and its contribution will be expected to cancel out on average if the heavy atom structure factor and β are not correlated.

The recognition of crossover terms does not alter the weight of the contribution of these reflections to the heavy atom synthesis. However, it is clear that whenever $\beta > 90^o$ the coefficient

$$f_H e^{i\alpha_H} . \tfrac{1}{2}\cos\beta$$

is negative. Elimination of errors of this sort in a difference syntheses can be achieved by using the coefficient

$$\Delta F\ e^{i(\alpha p+\pi)}$$

for those reflections where $\beta > 90^o$. This has the further effect

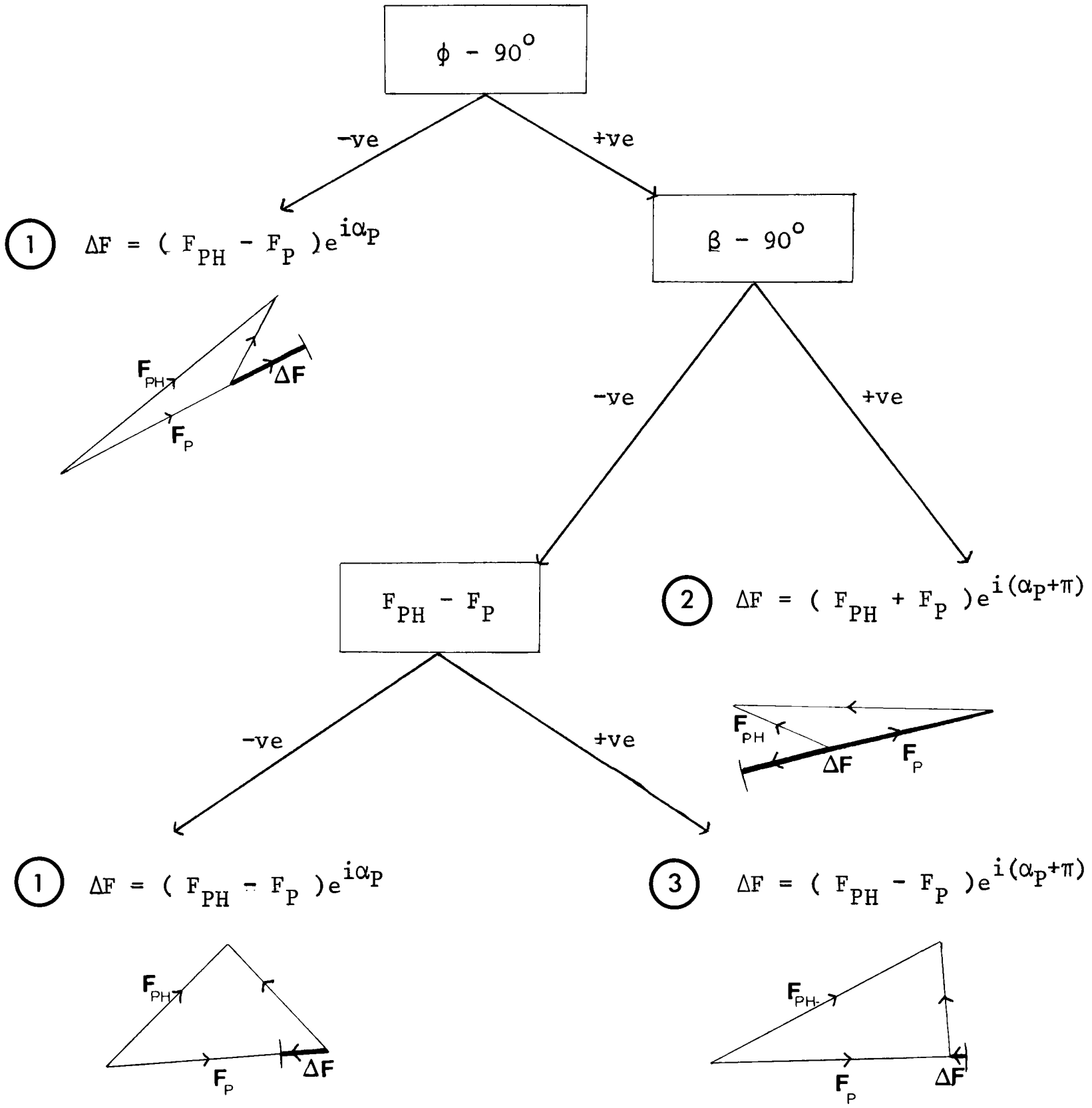


FIGURE A2 Difference Fourier Coefficients.
If the geometry of the phase triangle is known then minimisation of the spurious background errors in difference Fourier maps suggests the use of coefficients defined as shown above for four different geometries and described in the text.

of making the weight of the protein features contributed by these reflections negative, as it is already for the other reflections. The net effect of increasing the spurious negative background in this way whilst improving the quality of the heavy atom synthesis can not be readily assessed statistically. A geometrical illustration of the coefficients expected to generate the clearest representation of the heavy atom structure is given in Figure A2 for possible configurations of the phase triangle.

E) SINGLE DERIVATIVE PROTEIN MAPS

The only information obtainable about the protein phases from a single isomorphous derivative is that

$$\alpha_P = \alpha_H \pm \phi$$

and so the best Fourier synthesis (Blow and Crick, 1959) is calculated using the coefficients

$$F_P \cos\phi \; e^{i\alpha_H}$$

From the results of the previous section it is seen that

$$F_P \cos\phi \; e^{i\alpha_H} = F_P . e^{i\alpha_P} (\cos^2\phi \pm i \sin\phi\cos\phi)$$

Thus the acentric reflections, for which the angle ϕ may be assumed to be randomly distributed, contribute with only half the expected weight to the protein density. If the sign of ϕ is also random the spurious background contributed by the imaginary factor in this expression is expected to be low. However, unless the contribution of the acentric reflections is weighted up to make it equal, on average, to that of the centric reflections, then the image of the protein will be confused by a weak image of its enantiomer. An increase in the clarity of such a protein map would be expected for precisely the same reasons as for difference maps which are weighted in this manner.

Similar considerations apply to all "best" Fourier maps for which the coefficients are weighted down by the figure of merit m which can be shown to be the weighted mean value of the cosine of the error in the best phase angle αp (Dickerson, Kendrew and Strandberg, 1961)

$$m = \langle P_i \cos(\alpha_i - \alpha p)\rangle / \langle P_i \rangle$$

Thus if the actual error in the phase angle is denoted ε where α_T is the true phase and

$$n = \cos\varepsilon = \cos(\alpha p - \alpha_T)$$

it can be seen that the coefficient in a best Fourier synthesis may be rewritten as

$$mF_P e^{i\alpha p} = F_P e^{i\alpha_T}\{m.e^{i(\alpha p - \alpha_T)}\}$$

$$= F_P e^{i\alpha_T}\{m(\cos\varepsilon + i\sin\varepsilon)\}$$

$$= F_P e^{i\alpha_T}\{m.n + i.m\sin\varepsilon\}$$

On average $m \sim n$

so that $\langle m.n\rangle \sim m^2$

The imaginary term in the above expression will be expected to be cancelled out on average and not produce systematic features, so that the best Fourier synthesis is essentially that whose coefficients are

$$m^2 F_P . e^{i\alpha_T}$$

This has three implications which may be observable in electron density maps of proteins which are calculated from these coefficients. Firstly the overall level of electron density will be reduced by $\langle m^2\rangle$ from that expected in a calcualted F_P synthesis. Secondly the ratio of the contributions of centric and acentric reflections will be in the ratio of the appropriate values of $\langle m^2\rangle$. For the chicken TIM data this is approximately

$$\langle m\rangle^2_{centric} / \langle m\rangle^2_{acentric} = (0.83/0.65)^2 = 1.62$$

Thirdly the higher resolution terms will contribute proportionately less than the low resolution terms since m always decreases at higher resolution. For the chicken TIM data the extent of this is seen approximately from

$$\langle m \rangle^2_{\infty-6\text{Å}} / \langle m \rangle^2_{3.5-2.5\text{Å}} = (.84/.61)^2 = 1.90$$

These results suggest that relative weighting factors between groups of reflections could improve the quality of the map, since any increase in the weight of the residual terms whose phase is advanced by $\pi/2$ would still be expected to cancel out on average and not generate systematic features in the density. The significance of such weighting factors and their effect with respect to the clarity of a protein electron density map and its effective resolution have still to be fully investigated. A sensitive analytical test of the quality of an electron density map is necessary for this purpose.

The theoretical basis for applying an artificial negative temperature factor to the coefficients, which compensates for the average radial changes in the figure of merit, is evident from this discussion. An electron density map of the haem group of sperm-whale myoglobin, which was calculated using coefficients modified in this way, was found by Phillips to agree with an ideal haem group slightly better than did the unmodified best Fourier synthesis (see Dickerson, Kendrew and Strandberg, 1961), as would be predicted by the results derived above.

The overall conclusion of this analysis is that the significance of features in any Fourier synthesis, which arise from random errors in the coefficients, is reduced by the appropriate weighting of those groups of reflections for which there are systematic differences in

the estimated errors of their coefficients. For difference Patterson and difference Fourier syntheses the coefficients of the acentric reflections have been shown to require increased weighting by a factor of 2, relative to those of the centric reflections. For "best" Fourier syntheses of the electron density in which the structure amplitudes of reflections are reduced according to their figures of merit, it suggested that centric and acentric reflections should be weighted according to the ratio of the appropriate values of $\langle m^2 \rangle$. Relative weighting of low and high resolution reflections should compensate for the decrease of $\langle m^2 \rangle$ with increasing resolution by means of a smoothly varying function such as an exponential or a power series. Difference Fourier maps in which the coefficients are reduced according to the figure of merit of each reflection would seem to require relative weighting with respect to both variations in $\langle m^2 \rangle$ and also the factor of 2 between centric and acentric reflections. However, the importance of these weighting schemes and their effect on the degree of interpretability of a Fourier synthesis, as determined by some analytical criterion, has yet to be established.

Appendix III DERIVATIVE SCALE FACTOR DETERMINATION USING ISOMORPHOUS AND ANOMALOUS DIFFERENCES

The expressions proposed by Singh and Ramaseshan (1966) and Arnone et al.(1971) for determining this scale factor are compared and shown to involve exactly the same assumptions about the statistical distribution of the amplitudes, but to differ in the methods of using the statistical averages. An alternative method of considering the average quantities in the expression of Singh and Ramaseshan is suggested.

Both groups of authors assume the basic equations of Wilson statistics as discussed in Appendix II and determine the calculated heavy atom structure amplitude according to the formula of Matthews (1966b). This involves the ratio k' of the isomorphous to anomalous contributions of the heavy atom scattering, and also a weighting factor ω to reduce the importance assigned to the observed anomalous differences. Arnone et al. use a value $\omega = 0.75$ in their expression

$$f_H^2 = F_{PH}^2 + F_P^2 \pm 2F_PF_{PH}\{1 - (\omega k'\Delta F/2F_P)^2\}^{\frac{1}{2}} \qquad (1)$$

Since $2F_{PH}.\Delta F = 2F_{PH}(F_{PH+} - F_{PH-})$

$$= (F_{PH+}^2 - F_{PH-}^2)$$

$$= \Delta I$$

Their expression can be rewritten as

$$f_H^2 = F_{PH}^2 + F_P^2 \pm 2\{F_PF_{PH} - (\omega k'\Delta I/4)^2\}^{\frac{1}{2}}$$

Except for the use of F_{PH} instead of F_m defined by Singh and Ramaseshan as

$$F_m^2 = \tfrac{1}{2}(F_{PH+}^2 + F_{PH-}^2) = F_{PH}^2 + \frac{1}{k',2}.f_H^2$$

this expression is common to both methods. Use of F_m by Singh and Ramaseshan leads them to the expression

$$F_m^2 = F_P^2 + f_H^2(1 + \frac{1}{k'^2}) + 2F_P f_H \cos\phi \qquad (2)$$

but since they consider that

$$(1 + \frac{1}{k'^2}) \sim 1$$

later in their analysis, the relevance of distinguishing between F_{PH} and F_m in this problem is not obvious. Their analysis is thus considered here in terms of F_{PH} without any assumption about the size of k' being necessary.

Exclusion of those reflections having small protein and derivative amplitudes permits an exact calculation of the heavy atom structure amplitude since the square root of the term in the square brackets will always be negative for the remaining reflections. This corresponds to the angle β being acute. Arnone et al. evaluate $\langle f_H^2 \rangle$ in this way and use the suggestion of Kraut that

$$\langle f_H^2 \rangle = \langle k^2 F_{PH}^2 \rangle - \langle F_P^2 \rangle$$

to relate this quantity to the scale factor k by which the derivative amplitudes should be multiplied. This equation is simply the basic expression of Wilson statistics.

The use of F_{PH} in the determination of $\langle f_H^2 \rangle$ requires an iterative procedure for evaluating the best scale factor. Arnone et al. suggest such a procedure. It is not possible to determine k analytically from such expressions involving $\langle f_H^2 \rangle$. Singh and Ramaseshan obtain an analytical expression for k by considering the explicit estimate of f_H for each reflection as provided by the anomalous and isomorphous differences separately.

From the anomalous difference for each reflection it may be seen,

for example from Figure 37, that

$$F_{PH+}^2 = F_{PH}^2 + f_H''^2 - 2F_{PH}.f_H''\mathrm{Cos}(\gamma+\tfrac{\pi}{2})$$

$$F_{PH-}^2 = F_{PH}^2 + f_H''^2 + 2F_{PH}.f_H''\mathrm{Cos}(\gamma+\tfrac{\pi}{2})$$

Therefore $\Delta I = -4F_{PH}f_H''\mathrm{Cos}(\gamma+\frac{\pi}{2})$

$$= (4F_{PH}f_H\mathrm{Sin}\gamma)/k'$$

$$= (4F_Pf_H\mathrm{Sin}\phi)/k'$$

The expression resulting from the isomorphous difference

$$F_{PH}^2 = F_P^2 + f_H^2 + 2F_Pf_H\mathrm{Cos}\phi \qquad (3)$$

may be combined with

$$\Delta I = 4F_Pf_H\mathrm{Sin}\phi/k' \qquad (4)$$

by eliminating ϕ to give the expression for the upper and lower estimates of f_H as used by Arnone et al. following the definition of Matthews. Alternatively by eliminating f_H an exact equation for each reflection is obtained

$$F_{PH}^2 - F_P^2 = \{k'\Delta I/(4F_P\mathrm{Sin}\phi)\}^2 + k'\Delta I\mathrm{Cos}\phi/(2\mathrm{Sin}\phi)$$

Therefore $(F_{PH}^2 - F_P^2)\mathrm{Sin}^2\phi = \{k'\Delta I/(4F_P)\}^2 + (k'\Delta I/4)\mathrm{Sin}2\phi$

Explicit consideration of the scale factors S_P and S_{PH}, by which the native and derivative amplitudes should be multiplied to place them on an absolute scale, involves rewriting this equation as

$$(S_{PH}^2F_{PH}^2 - S_P^2F_P^2)\mathrm{Sin}^2\phi = \{k'S_{PH}^2\Delta I/(4S_PF_P)\}^2 + (k'S_{PH}^2\Delta I/4)\mathrm{Sin}2\phi$$

The reciprocal of the relative scale factor k between derivative and native sets of amplitudes

$$k = S_{PH}/S_P = S_R^{-1}$$

is specified by a quadratic relationship on manipulation of the

above expression

$$(F_{PH}^2 - S_R^2F_P^2)\mathrm{Sin}^2\phi = \{k'\Delta I/(4S_RF_P)\}^2 + (k'\Delta I/4)\mathrm{Sin}2\phi$$

Multiplying throughout by $(S_RF_P)^2$ and rearranging gives

$$S_R^4F_P^4\mathrm{Sin}^2\phi + S_R^2((k'\Delta I/4)F_P^2\mathrm{Sin}2\phi - F_{PH}^2F_P^2\mathrm{Sin}^2\phi) + (k'\Delta I/4)^2 = 0 \qquad (5)$$

This is an exact equation for each reflection and allows explicit determination of the upper and lower values of S_R since

$$S_R^2 = \frac{(F_{PH}^2F_P^2\mathrm{Sin}^2\phi - KF_P^2\mathrm{Sin}2\phi) \pm \sqrt{(F_{PH}^2F_P^2\mathrm{Sin}^2\phi - KF_P^2\mathrm{Sin}2\phi)^2 - 4F_P^4K^2\mathrm{Sin}^2\phi}}{2F_P^4\mathrm{Sin}^2\phi} \qquad (6)$$

where $K = (k'\Delta I/4)$

Except for the use of F_{PH} instead of F_m this result is exactly that obtained by Singh and Ramaseshan. They then substitute the statistical averages

$$\langle \mathrm{Sin}^2\phi \rangle = \tfrac{1}{2} \qquad \langle \mathrm{Sin}2\phi \rangle = 0$$

so that

$$S_R^2 = \frac{\langle F_{PH}^2F_P^2 \rangle \pm \{ \langle F_{PH}^2F_P^2 \rangle \langle k^2\Delta I^2/2 \rangle \langle F_P^4 \rangle \}^{\frac{1}{2}}}{2 \langle F_P^4 \rangle} \qquad (7)$$

However, this expression using the higher moments of the structure amplitudes will be more sensitive to experimental error in their measurement than expressions involving only the first or second moments, as does the use of ΔI, since this avoids the use of the weighted mean of F_{PH} in the expression

$$F_{PH+}^2 - F_{PH-}^2 = \Delta I = 2F_{PH}.\Delta F$$

which is an exact relation in the absence of errors. It does not seem necessary to take averages at this stage and equation (6) may be further manipulated, whilst maintaining the exact relationship

for each reflection, to give

$$S_R^2 = \frac{(F_{PH}^2 \mathrm{Sin}^2\phi - K\mathrm{Sin}2\phi) \pm \{(F_{PH}^2 \mathrm{Sin}^2\phi - K\mathrm{sin}2\phi) - 4K^2\mathrm{Sin}^2\phi\}^{\frac{1}{2}}}{2F_P^2 \mathrm{Sin}^2\phi}$$

Therefore

$$S_R^2 = \frac{F_{PH}^2\{\mathrm{Sin}^2\phi - \mathrm{Sin}2\phi.k'\Delta F/2F_{PH} \pm \{(\mathrm{Sin}^2\phi - \mathrm{Sin}2\phi.k'\Delta F/2F_{PH})^2 - k'^2\Delta F^2/(4F_{PH}^2)\}^{\frac{1}{2}}\}}{2F_P^2.\mathrm{Sin}^2\phi}$$

Thus on taking statistical averages within this exact equation

$$S_R^2 = \frac{\langle F_{PH}^2\rangle\{\frac{1}{2} \pm \{(\frac{1}{2})^2 - \langle k'^2\Delta F^2/(4F_{PH}^2)\rangle\}^{\frac{1}{2}}\}}{\langle F_P^2\rangle}$$

$$= \tfrac{1}{2}\{1 \pm \{1 - \langle(k'\Delta F/F_{PH})^2\rangle\}^{\frac{1}{2}}\} \langle F_{PH}^2\rangle / \langle F_P^2\rangle \qquad (9)$$

The positive sign must be taken for the square root with no ambiguity. This is most easily seen by considering the situation where

$$\langle F_{PH}^2\rangle = \langle F_P^2\rangle$$

that is when a preliminary scale factor has already been applied. The appropriate scale factor k, the reciprocal of S_R must exceed unity by an amount which depends, according to Wilson statistics, on the total heavy atom occupancy. In this case

$$S_R^2 = k^{-2} = \tfrac{1}{2}\{1 \pm \{1 - \langle(k'\Delta F/F_{PH})^2\rangle\}^{\frac{1}{2}}\}$$

and it is clear that as the heavy atom occupancy, and thus the anomalous differences tend to zero, the scale factor k decreases to unity or increases to infinity by use of the positive and negative signs respectively. Clearly only the former result can be valid.

Use of equation (9) to evaluate derivative rescaling factors for the chicken TIM measurements to 2.5Å resolution gave the results presented in Figure 41. These are compared with those calculated by

the original formula published by Singh and Ramaseshan. Their formula determines that the scale factor for the MONO derivative increases rapidly with increasing resolution, indicating the sensitivity of their expression to errors, especially those of decreasing isomorphism. This is observed for the MONO derivative at increasing resolution. In the absence of large errors in the data both formulae appear to give comparable results. However, the use of the expression derived here appears to be justified both theoretically and in practice.

The existence of errors is most easily taken into consideration by use of a weighting factor to multiply the anomalous ratio k', whether an empirical or absolute value is used for this quantity. The relation between the scale factor and the degree of isomorphism of a derivative is seen by expanding the ratio

$$< (k'\Delta F/F_{PH})^2 >$$

Substituting the value of Matthews (1966b) for k' gives

$$< (k'\Delta F/F_{PH})^2 > \;=\; <\left(\frac{<|F_{PH}-F_P|>.\ 2\Delta F}{<\Delta F>\ F_{PH}}\right)^2>$$

and use of the formula suggested in Chapter 5 for the empirical ratio gives

$$< \frac{2<F_{PH}^2-F_P^2>.\ \Delta F^2}{<\Delta F^2>\ F_{PH}^2}>$$

This latter expression is clearly closely related to the quantity

$$\phi\Delta I \;=\; \frac{\{RMS(F_{PH}^2 - F_P^2\}}{<F_P^2>}$$

which Crick and Magdoff (1956) show to be related to the degree of isomorphism of a derivative.

This comparison of two published methods of determining the scale factor has shown that they are more similar than is immediately

apparent. The method of Kraut, as applied by Arnone et al. requires an iterative procedure but is not especially sensitive to errors in the measurements. The modification of Kraut's method, which was discussed in Chapter V, relying only on measurements of the isomorphous differences gave useful results from the first cycle and iteration would clearly be expected to improve the scale factors. The original formulation of their method by Singh and Ramaseshan gives an expression for the scale factor which is very sensitive to errors in the measurements or lack of isomorphism. The simplified expression derived here has been shown to be considerably less sensitive to errors. Variations of the scale factor with the resolution of the measurements can indicate lack of isomorphism if the influence of other sources of error has been minimised.

REFERENCES

General text books and reviews referred to in the introduction:

MAHLER,H.R. & CORDES,E.H.,(1971). Biological Chemistry, 2nd edn., Harper & Row, London.

INTERNATIONAL TABLES FOR X-RAY CRYSTALLOGRAPHY:
Vol.I Symmetry Groups,(1952). HENRY,N.F.M. & LONSDALE,K. Eds.
Vol.II Mathematical Tables,(1959). KASPER,J.S. & LONSDALE,K. Eds.
Vol.III Physical & Chemical Tables,(1962). MacGILLAVRY,C.H. & RIECK,G.D. Eds.
The Kynoch Press, Birmingham.

LIPSON,H. & COCHRAN,W.,(1966). The Determination of Crystal Structures, 3rd edn., Bell & Sons, London.

STOUT,G.H. & JENSEN,L.H.,(1968). X-Ray Structure Determination. A Practical Guide, Macmillan, New York.

HOLMES,K.C. & BLOW,D.M.,(1966). The Use of X-ray Diffraction in the Study of Protein and Nucleic Acid Structure, Interscience, New York.

PHILLIPS,D.C.,(1966). Advances in Protein Crystallography, in Adv. Str. Res. by Diffraction Methods, 2:75-140, BRILL,R. & MASON,R. Eds.

NORTH,A.C.T. & PHILLIPS,D.C.,(1969). X-ray Studies of Crystalline Proteins in Prog. in Biophys. and Mol.Biol. 19:1-132.

BLUNDELL,T.L. & JOHNSON,L.N.,(1972). Protein Crystallography, in M.T.P.Biannual Review of Science, Vol. Chemical Crystallography, ROBERTSON,J.M. Ed., Butterworths, London.

COLD SPRING HARBOR SYMPOSIUM IN QUANTITATIVE BIOLOGY,(1971). Vol.36: Structure and Function of Proteins at the Three Dimensional Level.

ADLER,K., BEYREUTHER,K., FANNING,E., GEISLER,N., GRONENBORN,B., KLEMM,A., MULLER-HILL,B., PFAHL,M. & SCHMITZ,A.,(1972). Nature 237:322-7.
How lac Repressor Binds to DNA.

ARNDT,U.W.,(1968). Acta Cryst.B24:1355-7.
The Optimum Strategy in Measuring Structure Factors.

ARNDT,U.W., NORTH,A.C.T. & PHILLIPS,D.C.,(1964). J. Sci. Instr.41:421-5.
Adaptation of a Linear Diffractometer to Measure Three Reflections Quasi-simultaneously.

ARNDT,U.W. & PHILLIPS,D.C.,(1961). Acta Cryst. 14:807-18.
The Linear Diffractometer.

ARNDT,U.W. & WILLIS,B.T.M.,(1966). Single Crystal Diffractometry, Cambridge University Press.

ARNONE,A., BIER,C.J., COTTON,F.A., DAY,V.W., HAZEN,E.E., RICHARDSON,D.C., RICHARDSON,J.S. & YONATH,A.A.,(1971). J. Biol. Chem. 246;2302-16. A High Resolution Study of an Inhibitor Complex of the Extracellular Nuclease of Staphylococcus aureus.

BANNER,D.W.,(1972). D.Phil.Thesis, University of Oxford. X-ray Diffraction Studies of Molecules of Large Molecular Weight.

BANNER,D.W., BLOOMER,A.C., PETSKO,G.A., PHILLIPS,D.C. & POGSON,C.I.,(1971). Cold Spring Harbor Symp. Quant. Biol. 36:151-6. Crystallographic Studies of Triose Phosphate Isomerase.

BIJVOET,J.M.,(1954). Nature 173:888-91. Structure of Optically Active Compounds in the Solid State.

BLAKE,C.C.F., KOENIG,D.F., MAIR,G.A., NORTH,A.C.T., PHILLIPS,D.C. & SARMA,V.R.,(1965). Nature 206:757-63. Structure of Hen Egg-white Lysozyme.

BLAKE,C.C.F. & PHILLIPS,D.C.,(1962). in Biological Effects of Ionising Radiations at the Molecular Level, I.A.E.A., Vienna. Effects of X-irradiation on Single Crystals of Myoglobin.

BLOW,D.M.,(1957). Ph.D.Thesis, University of Cambridge. X-ray Analysis of Haemoglobin: Determination of Phase Angles by Isomorphous Replacement.

BLOW,D.M.,(1958). Proc. Roy. Soc. A247:302-36. The Structure of Haemoglobin VII. Determination of Phase Angles in the Non-centrosymmetric [001] Zone.

BLOW,D.M. & CRICK,F.H.C.,(1959). Acta Cryst. 12:794-802. The Treatment of Errors in the Isomorphous Replacement Method.

BLOW,D.M. & ROSSMANN,M.G.,(1961). Acta Cryst. 14:1195-202. The Single Isomorphous Replacement Method. see Corrigendum(1962). Acta Cryst. 15:1060.

BLOW,D.M., ROSSMANN,M.G. & JEFFREY,B.A.,(1964). J. Mol. Biol. 8:65-78. The Arrangement of α-Chymotrypsin Molecules in the Monclinic Crystal Form.

BODO,G., DINTZIS,H.M., KENDREW,J.C. & WYCKOFF,H.W.,(1959). Proc. Roy. Soc. A253:70-103. The Structure of Myoglobin V. A Low Resolution Three-dimensional Fourier Synthesis of Sperm-whale Myoglobin Crystals.

BRAGG,W.L.,(1958). Acta Cryst. 11:70-75. The Determination of the Coordinates of Heavy Atoms in Protein Crystals.

BURTON,P.M. & WALEY,S.G.,(1968). Biochem. J. 107:737-44. Studies on the Subunits of Triose Phosphate Isomerase.

COCHRAN,W.,(1958). Acta Cryst. 11:579-85.
Structure Factor Relations and the Phase Problem.

COCHRAN,W. & WOOLFSON,M.M.,(1955). Acta Cryst. 8:1-12.
The Theory of Sign Relations Between Structure Factors.

COX,J.M.,(1967). J. Mol. Biol. 28:151-6.
Structure and Function of Haemoglobin: Mathematical Methods Used in the Comparison of Quaternary Structures.

CRICK,F.H.C. & MAGDOFF,B.S.,(1956). Acta Cryst. 9:901-8.
The Theory of the Method of Isomorphous Replacement for Crystals,I.

CULLIS,A.F., MUIRHEAD,H., PERUTZ,M.F., ROSSMANN,M.G. & NORTH,A.C.T., (1961). Proc. Roy. Soc. A265:15-160.
The Structure of Haemoglobin VIII. A Three-dimensional Fourier Synthesis at 5.5Å Resolution: Determination of Phase Angles.

DICKERSON,R.E., KENDREW,J.C. & STRANDBERG,B.E.,(1961). Acta Cryst. 14: 1188-95. The Crystal Structure of Myoglobin: Phase Determination to a Resolution of 2Å by the Method of Isomorphous Replacement.

DICKERSON,R.E., KOPKA,M.L., VARNUM,J.C. & WEINZIERL,J.E.,(1967). Acta Cryst. 23:511-22.
Bias, Feedback and Reliability in Isomorphous Phase Analysis.

DICKERSON,R.E., WEINZIERL,J.E. & PALMER,R.A.,(1968). Acta Cryst. B24: 997-1003. A Least Squares Method for Isomorphous Replacement.

DODSON,E. & VIJAYAN,M.,(1971). Acta Cryst. B27:2402-11.
The Determination and Refinement of Heavy Atom Parameters in Protein Heavy-atom Derivatives. Some Model Calculations Using Acentric Reflections.

DONNAY,J.D.H. & HARKER,D.,(1937). Amer. Min. 22:446-67.
A New Law of Crystal Morphology Extending the Law of Bravais.

FINKELSTEIN,A.V. & PTITSYN,O.B.,(1972). J. Mol. Biol. 62:613-25.
Statistical Analysis of the Correlation Among Amino Acid Residues in Helical, β-Structural and Non-regular Regions of Globular Proteins.

GREEN,D.W., INGRAM,V.M. & PERUTZ,M.F.,(1954). Proc. Roy. Soc. A225:287-307. The Structure of Haemoglobin IV. Sign Determination by the Isomorphous Replacement Method.

HARKER,D.,(1953). Acta Cryst. 6:731-6.
The Meaning of the Average of $|F|^2$ for Large Values of the Interplanar Spacing

HARKER,D.,(1956). Acta Cryst. 9:1-9.
The Determination of the Phases of the Structure Factors of Non-centrosymmetric Crystals by the Method of Double Isomorphous Replacment.

HAMILTON,W.C., ROLLETT,J.S. & SPARKS,R.A.,(1965). Acta Cryst. 18:129-30.
On the Relative Scaling of X-ray Photographs.

HART,R.G.,(1961). Acta Cryst. 14:1194-5.
The Crystal Structure of Myoglobin. Refinement of Heavy Atom Parameters Other than Relative y's.

HAUROWITZ,F.,(1950). Chemistry and Biology of Proteins, p.91, Academic Press, New York.

HENDERSON,R. & MOFFATT,J.K.,(1971). Acta Cryst. B27:1414-20.
The Difference Fourier Technique in Protein Crystallography: Errors and their Treatment.

HENDRIKSON,W.A. & LATTMAN,E.E.,(1970). Acta Cryst. B26:136-43.
Representation of Phase Probability Distributions for Simplified Combination of Independent Phase Information.

HODGKIN,D.C., KAMPER,J., LINDSEY,J., MACKAY,M., PICKWORTH,J., ROBERTSON,J.H., SHOEMAKER,C.B., WHITE,J.G., PROSEN,R.J. & TRUEBLOOD,K.N.,(1957). Proc. Roy. Soc. A242:228-63.
The Structure of Vitamin B_{12} I. An Outline of the Crystallographic Investigation of Vitamin B_{12}.

JAMES,R.W.,(1948). Acta Cryst. 1:132-4.
False Detail in Three-dimensional Fourier Representations of Crystal Structures.

JARDETZKY,O., MARKLEY,J.L., THIELMANN,H., ARATA,Y. & WILLIAMS,M.N., (1971). Cold Spring Harbor Symp. Quant. Biol. 36:257-62.
Tentative Sequential Model for the Unfolding and Refolding of Staphylococcal Nuclease at High pH.

JERGENSONS,B.,(1966). Makromol. Chem. 91:74-86.
Classification of Proteins According to Conformation.

JOHNSON,L.N. & WALEY,S.G.,(1967). J. Mol. Biol. 29:321-2.
Preliminary Crystallographic Data for Triose Phosphate Isomerase.

JOHNSON,L.N. & WOLFENDEN,R.W.,(1970). J. Mol. Biol. 47:93-100.
Changes in the Absorption Spectrum and Crystal Structure of Triose Phosphate Isomerase Brought about by 2-phosphoglycollate, a Potential Transition State Analogue.

KARLE,J. & HAUPTMAN,H.,(1957). Acta Cryst. 10:515-24.
A Unified Algebraic Approach to the Phase Problem II. Space Group P1.

KITAIGORODSKY,A.I.,(1970). Adv. Str. Res. by Diffraction Methods 3:173-247. General View on Molecular Packing.

KITAIGORODSKY,A.I. & AHMED,N.A.,(1972). Acta Cryst. A28:207-10.
Theoretical Determination of the Geometrical Form of the Crystals of Anthracene.

KNOWLES,J.R., LEADLEY,P.F. & MAISTER,S.G.,(1971). Cold Spring Harbor Symp. Quant. Biol. 36:157-63.
Triose Phosphate Isomerase: Isotope Studies on the Mechanistic Pathway.

KOSHLAND,D.E.,(1958). Proc. Nat Acad. Sci, Wash. 44:98-104.
Application of a Theory of Enzyme Specificity to Protein Synthesis.

LANGRIDGE,R.,(1969). J. Cell. Physiol. 74(Sup.1):1-20.
Nucleic Acids and Polynucleotides.

LOW,B.W. & RICHARDS,F.M.,(1952). J. Amer. Chem. Soc. 74:1660-6.
The Use of the Gradient Tube for the Determination of Crystal Densities.

LUZZATTI,V.,(1953). Acta Cryst. 6:142-53.
Resolution d'une Structure Cristalline Lorsque les Positions d'une Partie des Atomes sont Connues: Traitement Statistique.

LUZZATTI,V.,(1955). Acta Cryst. 8:795-806.
Sur l'Emploi des Méthodes Statistiques dans l'Étude de la Structure Critalline des Proteines.

MATTHEWS,B.W.,(1966a). Acta Cryst. 20:82-6.
The Extension of the Isomorphous Replacement Method to Include Anomalous Scattering Measurements.

MATTHEWS,B.W.,(1966b). Acta Cryst. 20:230-9.
The Determination of the Position of Anomalously Scattering Heavy Atom Groups in Protein Crystals.

MATTHEWS,B.W.,(1969). in Crystallographic Computing, AHMED,F.R., Ed. Munksgaard, Copenhagen.
Determination and Refinement of Phases for Proteins.

McMURRAY,C.H. & TRENTHAM,D.R.,(1969). Biochem. J. 115:913-21.
A New Class of Chromophoric Organomercurials and their Reactions with D-glyceraldehyde-3-phosphate Dehydrogenase.

MIN JOU,W., HAEGMAN,G., YSEBAERT,M. & FIERS,W.,(1972). Nature 237:82-8.
Nucleotide Sequence of the Gene Coding for the Bacteriophage MS2 Coat Protein.

MOFFAT,J.K.,(1969). Ph.D.Thesis, University of Cambridge.
X-ray Studies of Chemically Modified Horse Haemoglobin.

MONOD,J., WYMAN,J. & CHANGEUX,J.P.,(1965). J.Mol. Biol. 12:88-118.
On the Nature of Allosteric Transitions: A Plausible Model.

MOEWS,P.C. & BUNN,C.W.,(1971). Acta Cryst. B27:1780-3.
Multiple Heavy Atom Sites in Protein Crystal having Centrosymmetric Projections: Interpretation of Vector Maps I Modification of the three dimensional Patterson Difference Synthesis.

MOULT,J.,(1970). D.Phil.Thesis, University of Oxford.
The Structure of Some Sugars and their Relationship to the Protein Lysozyme.

NORTH,A.C.T.,(1959). Acta Cryst. 12:512-4.
A Cubic Form of Ox Haemoglobin.

NORTH,A.C.T.,(1964). J. Sci. Instr. 41:42-65.
Computer Processing of Automatic Diffractometer Data.

NORTH,A.C.T.,(1965). Acta Cryst. 18:212-6.
The Combination of Isomorphous Replacement and Anomalous Scattering Data in Phase Determination of Non-centrosymmetric Reflexions.

NORTH,A.C.T., PHILLIPS,D.C. & MATTHEWS,B.W.,(1968). Acta Cryst. A24:351-9.
A Semi- Empirical Method of Absorption Correction.

PAULING,L., COREY,R.B. & BRANSON,H.R.,(1951). Proc. Nat. Acad. Sci. Wash. 37:205-11. The Structure of Proteins: Two Hydrogen Bonded Helical Conformations of the Polypeptide Chain.

PERUTZ,M.F.,(1954). Proc. Roy. Soc. 225:264-86.
The Structure of Haemoglobin III. Direct Determination of the Molecular Transform.

PHILLIPS,D.C.,(1964). J.Sci. Instr. 41:123-9.
On the Design of Single Crystal Diffractometers to Measure a Number of Reflections Simultaneously.

PHILLIPS,D.C.,(1966). Adv. Str. Res. by Diffraction Methods 2:75-140.
Advances in Protein Crystallography.

POGSON,C.I.,(1972) Manuscript in preparation.
Preparation of Triose Phosphate Isomerase from Chicken Muscle; Procedure for Crystallisation and Co-crystallisation with Organomercurial Ligands.

QUIOCHO,F.A. & RICHARDS,F.M.,(1964). Proc. Nat. Acad. Sci. Wash. 52:833-9.
Intermolecular Cross linking of a Protein in the Crystalline State: Carboxypeptidase A.

REYNOLDS,S.J., YATES,D.W. & POGSON,C.I.,(1971). Biochem. J. 122:285-97.
Dihydroxy acetone phosphate:its Structure and Reactivity with α-glycero phosphate dehydrogenase, Aldolase and Triose Phosphate Isomerase and Some Possible Metabolic Implications.

RICHARDS,F.M.(1968). J.Mol. Biol. 37:225-30.
The Matching of Physical Models to Three-dimensional Electron Density Maps: A Simple Optical Device.

RIEDER,S.V. & ROSE,I.A.,(1959). J Biol Chem. 234:1007-10.
The Mechanism of the Triose Phosphate Isomerase Reaction.

ROBSON,B. & PAIN,R.H.,(1971). J Mol. Biol. 58:237-259.
Analysis of the Code Relating Sequence to Conformation in Proteins: Possible Implications for the Mechanism of Formation of Helical Regions.

ROSSMANN,M.G. & BLOW,D.M.,(1962). Acta Cryst. 15:24-31.
Detection of Subunits Within the Crystallographic Asymmetric Unit.

RÜHLMANN,A., SCHRAMM,H.J., KUKLA,D. & HUBER,R.,(1971). Cold Spring Harbor Symp. Quant. Biol. 36:148-50.
Pancreatic Trypsin Inhibitor (Kunitz). PartII. Complexes with Proteinases.

SCOPES,R.K.,(1968). Biochem. J. 107:139-50.
Methods for Starch-Gel Electrophoresis of Sarcoplasmic Proteins.
An Investigation of the Relative Mobilities of the Glycolytic Enzymes From the Muscles of a Variety of Species.

SIGLER,P.B. & BLOW,D.M.,(1965). J.Mol. Biol. 14:640-4.
A Means of Promoting Heavy-Atom Binding in Protein Crystals.

SINGH,A. & RAMASESHAN,S.,(1966). Acta Cryst. 21:279-80.
The Determination of Heavy Atom Positions in Protein Derivatives.

SZYBALSKI,W., KUBINSKI,H. & SHELDRICK,P.,(1966). Cold Spring Harbor Symp. Quant. Biol. 31:123-7.
Pyrimidine Clusters on the Transcribing Strand of DNA and Their Possible Role in the Initiation of RNA Synthesis.

TRAUB,W. & HIRSHFELD,F.L.,(1960). Acta Cryst. 13:753-60.
Intensity Measurements from Insulin Crystals on the Goniostat; Sources of Error and Accuracy of Data.

WATENPAUGH,K.D., SIEKER,L.C., HERRIOTT,J.R. & JENSEN,L.H.,(1971). Cold Spring Harbor Symp. Quant. Biol. 36:359-68.
The Structure of a Non-Heme Iron Protein: Rubredoxin at 1.5Å Resolution.

WATSON,J.D. & CRICK,F.H.C.,(1953). Nature 171:737-8.
A Structure for Deoxyribose Nucleic Acid.

WILSON,A.J.C.,(1948). Nature 150:151-2.
Determination of Absolute from Relative X-ray Intensity.